

UNIVERSITY OF LATVIA

Faculty of Biology



GINTS KALNIŅŠ

**STRUCTURAL AND FUNCTIONAL STUDIES OF BACTERIAL  
TRIMETHYLAMINE PRODUCING ENZYMES**

DOCTORAL THESIS

Submitted for the degree of Doctor of Biology,

subfield: Molecular Biology

Rīga, 2020

This work of doctoral thesis was carried out at the Latvian Biomedical Research and Study Centre from year 2013 to 2019.



Latvian Biomedical  
Research and Study Centre  
research and education in biomedicine from genes to human

This work has been supported by the state research program „Biomedicīna sabiedrības veselībai (BIOMEDICINE) and University of Latvia Foundation grant “Bacterial microcompartments as synthetic nanoreactors” funded by SIA Mikrotik. CIISB research infrastructure project LM2015043 funded by MEYS CR is gratefully acknowledged for the financial support of the measurements at the CF Cryo-electron Microscopy and Tomography CEITEC MU.



*MikroTik*

The form of this thesis is a collection of research papers in biology, subfield – molecular biology.

Supervisor: Prof., Dr. biol. Kaspars Tārs

Reviewers:

- 1) Dr. biol. Andris Zeltiņš, Latvian Biomedical Research and Study Centre
- 2) Dr. chem. Kristaps Jaundzems, Latvian Institute of Organic Synthesis
- 3) Prof., Cheryl A. Kerfeld, Ph.D., Lawrence Berkeley National Laboratory (Berkeley, CA, USA) and Michigan State University (East Lansing, MI, USA)

The thesis will be defended at the public session of the Doctoral Committee of Biology, University of Latvia, on 27<sup>th</sup> March 2020 at Latvian Biomedical Research and Study Centre, Ratsupites Str. 1, k-1.

The thesis is available at the Library of the University of Latvia, Jelgavas street 1, LV-1004, Riga.

Chairman of the Doctoral Committee  
Kloviņš/

\_\_\_\_\_ / Dr. biol. Jānis

Secretary of the Doctoral Committee

\_\_\_\_\_ / Daina Eze /

© University of Latvia, 2020

© Gints Kalniņš, 2020

## ABSTRACT

Microbiome is a significant contributor to the homeostasis of the human organism and it is responsible for the biosynthesis of several compounds that the human organism cannot by itself. The impact of such metabolites is various – there are beneficial, neutral and harmful examples. Trimethylamine (TMA) has been identified as one such bacterial metabolite with a potential negative impact on health. High levels of trimethylamine-N-oxide (TMAO), an oxidized form of TMA, have been associated with cardiovascular and other pathological conditions. Modifying the activity of bacterial TMA producing enzymes could have value in medicine. Therefore more detailed functional and structural characterization of such TMA producing enzymes is necessary.

Several bacterial TMA producing enzymes has been identified so far – CutC choline lyase and CntA carnitine oxygenase being among ones known best. The aim of this study was to characterize these enzymes by solving 3D structure, determining specific activities and studying the properties of associated supra-molecular structures. In this study we solved the 3D crystal structure of CutC choline lyase from *Klebsiella pneumoniae* and discovered that it exists in two different conformations dependent on the presence or absence of substrate, a feature not described for other closely related enzymes. *Klebsiella pneumoniae* CutC is a part of a GRM2 type bacterial microcompartment (BMC) system polyhedral or quasi-icosahedral bacterial organelles consisting of a protein shell and encapsulated enzymatic core. We therefore continued with a closer investigation of the assembly and encapsulation mechanisms of this BMC. We discovered that the core enzymes are encapsulated in a hierachial manner, with the CutC choline lyase playing a central adaptor role in this process. We present a cryo-EM structure of a pT=4 quasi symmetric shell particle at 3.3 Å resolution and study the requirements for recombinant shell formation. We also analyze the substrate panels of CntA oxygenase from four different hosts and confirm that in general the CntA enzymatic activity is high for carnitine and gamma-butyrobetaine, medium for betaine and very low for choline, thus clearing some previous controversies. We also demonstrated that the TMA could be produced from carnitine by *Providencia rettgeri* bacteria and by purified CntA enzyme at very low oxygen concentrations, thus suggesting that CntA, despite being an aerobic enzyme, could still give a contribution in the total pool of TMA production in anaerobic gastrointestinal system.

## ABSTRAKTS

Mikrobioms ir svarīgs cilvēka homeostāzes uzturēšanai un tas ir spējīgs producēt vairākus savienojumus, kurus cilvēka organisms pats par sevi nespēj. Šādu metabolītu ietekme ir daudzveidīga – tie var būt labvēlīgi, neitrāli un kaitīgi. Trimetilamīns (TMA) ir ticis identificēts kā viens no baktēriju metabolītiem ar potenciāli negatīvu ietekmi uz cilvēka veselību. Augsts trimetilamīna oksīda (TMAO) līmenis organismā ir ticis saistīts ar augstāku risku saslimt ar sirds-asinsvadu un citām saslimšanām. Bakteriālo trimetilamīnu producējošo enzīmu aktivitātes izmainīšanai varētu būt nozīmīgs pielietojums medicīnā. Tādējādi ir nepieciešams detalizētāks funkcionāls un strukturāls šādu enzīmu raksturojums.

Vairāki bakteriāli TMA producējoši enzīmi ir tikuši identificēti – CutC holīna liāze un CntA karnitīna oksigenāze ir vislabāk pazīstamās. Šī darba mērķis bija raksturot šos enzīmus, noteikt to trīsdimensionālās struktūras, detektēt specifisko aktivitāti un pētīt asociētu lielmolekulāru kompleksu īpašības. Šajā darbā tika noteikta *Klebsiella pneumoniae* CutC holīna liāzes trīsdimensionālā kristālstruktūra un tika atklāts, ka, atkarībā no substrāta klātbūtnes, šis enzīms eksistē divās dažādās konformācijās – unikāla īpašība starp radniecīgiem enzīmiem. *Klebsiella pneumoniae* CutC ir daļa no GRM2 tipa bakteriālā mikrokompartimenta (BMC) sistēmas, kvazi-ikosahedrālas vai ikosahedrālas baktēriju organellas, kura sastāv no proteīnu čaulas un enzimatiskas serdes. Pētījums atklāja, ka enzimatiskā serde tiek iepakota hierarhiālā veidā, kurā CutC holīna liāze funkcionē kā centrālais adaptorproteīns. Mēs prezentējam ar krio-elektronmikroskopiju iegūtu pT=4 kvazisimetrisku bakteriāla mikrokompartimenta čaulas daļiņas struktūru 3.3 Å izšķirtspējā un pētījām rekombinantu čaulu izveidošanai nepieciešamos priekšnosacījumus. Mēs arī testējām substrātu profilu CntA oksigenāzēm no četriem dažādiem saimniekorganismiem un secinājām, ka CntA oksigenāze ir visaktīvākā ar karnitīna un gamma-butirobetaīna substrātiem, vidēji aktīva ar betaīna substrātu un mazaktīva ar holīna substrātu. Mēs arī demonstrējam, ka *Providencia rettgeri* baktēriju kultūra un attīrīts CntA enzīms spēj producēt TMA zemās skābekļa koncentrācijās, kas norāda, ka CntA enzīms, par spīti tam, ka tas ir aerobs, nelielā mērā varētu piedalīties TMA producēšanā anaerobiskā vidē zarnu traktā.



## TABLE OF CONTENTS

ABSTRACT.....	3
ABSTRAKTS .....	4
TABLE OF CONTENTS.....	5
ABBREVIATIONS .....	7
INTRODUCTION .....	8
1. LITERATURE OVERVIEW.....	10
1.1 TMA and TMAO in <i>Animalia</i> organisms .....	10
1.2 TMA and TMAO role in pathological processes .....	12
1.2.1 TMA and trimethylaminouria.....	12
1.2.2 TMAO and associated conditions.....	12
1.2.3 Managing the amounts of TMA and TMAO in organism.....	14
1.2.4 Criticism of TMA and TMAO as pathological agents .....	14
1.3 Bacterial TMA producing enzymes .....	15
1.3.1 CutC choline lyase.....	15
1.3.2 CutC associated bacterial microcompartments.....	17
1.3.3 CntA Rieske non-heme oxygenase.....	19
1.3.4 Other TMA producing bacterial enzymes .....	21
2. MATERIALS AND METHODS.....	22
2.1 Plasmid construction and protein expression .....	22
2.2 Protein purification.....	23
2.3 MALDI-TOF-MS analyses .....	24
2.4 X-ray crystallography analysis of CutC .....	24
2.5 Transmission electron microscopy analysis .....	25
2.4 Cryo-EM single particle analysis of BDPs .....	25
2.5 Enzymatic <i>in vitro</i> activity assays of CutC and CntA.....	28
3. RESULTS .....	30
3.1 Structure and function of CutC choline lyase .....	30
3.2 Encapsulation mechanisms and structural studies of GRM2 BMC shell particles .....	40
3.3 CntA oxygenase substrate profile comparison and oxygen dependency of TMA production in <i>Providencia rettgeri</i> .....	54
4. DISCUSSION .....	63
4.1 Structure of CutC choline lyase .....	63

4.2 CutC-associated GRM2 type bacterial microcompartment structure and encapsulation mechanisms .....	64
4.3 CntA oxygenase substrate profile and oxygen dependency.....	65
5. CONCLUSIONS.....	67
6. MAIN THESIS OF DEFENSE.....	68
7. PUBLICATIONS.....	69
8. APPROBATION OF RESEARCH.....	70
9. ACKNOWLEDGEMENTS .....	71
10. REFERENCES .....	72

## ABBREVIATIONS

ATP – adenosine triphosphate	MOPS – 3-(N-morpholino) propanesulfonic acid
BDP – bacterial microcompartment shell derived particle	NADH – nicotinamide adenine dinucleotide, reduced
BMC – bacterial microcompartments	NAD <sup>+</sup> – nicotinamide adenine dinucleotide, oxidized
cryo-EM – cryogenic electron microscopy	PCR – polymerase chain reaction
coA – coenzyme A	PEG – polyethylenglycol
DNA – deoxyribonucleic acid	PTAC – phosphotransacetylase
DTT – dithiothreitol	SDS-PAGE – sodium dodecyl sulfate polyacrylamide gel electrophoresis
EDTA – ethylenediaminetetraacetic acid	SPA – single particle analysis
FMO3 – flavin containing monooxygenase 3	TAAR5 – trace amine-associated receptor 5
GCPR – G protein-coupled receptor	TEM – transmission electron microscopy
CTF – contrast transfer function	TFA – trifluoroacetic acid
GBB – gamma butyrobetaine	TMA – trimethylamine
GRE – glycy radical enzyme	TMAO – trimethylamine-N-oxide
GRE-AE – glycy radical enzyme activating enzyme	UPLC-MS/MS – ultra performance liquid chromatography - tandem mass spectrometry
GRM – glycy-radical enzyme associated bacterial microcompartment	
IPTG – isopropyl $\beta$ -D-1- thiogalactopyranoside	
MALDI-TOF-MS – matrix-assisted laser desorption/ionization time-of-flight mass spectrometry	

## INTRODUCTION

Microbes play significant roles in the life of all higher organisms, including humans. They are everywhere in biosphere, and we are in a constant and uninterrupted contact with them every day. Human organism is a host for a wide spectrum of bacteria, archaea, fungi and viruses, generally designated microbiota. The microbiota has various impacts on human organism. There are mutualistic examples, such as microbiota synthesizing important nutrients or vitamins in exchange for living environment. Many species of microbiota can also be commensals without any direct positive or negative influence on its host. A certain fraction of microbiota can also be pathogens in certain situations. All roles of microbiome are not fully understood yet and its influence may have far reaching consequences in immunological processes, brain development, pathogenesis of diseases and other areas.

One way how the microbiota can exercise its influence is by secreting unique metabolites fulfilling particular functions in human organism. Such bacterial metabolites as short chain fatty acids, secondary bile acids and trimethylamine (TMA) has been identified as potential causes of pathological conditions. TMA is a simple organic compound produced by bacterial degradation of organic compounds, and this is done by both free environmental bacteria and microbiota bacteria in digestive systems. TMA and its metabolite trimethylamine-N-oxide (TMAO) has been an object of a notable interest in recent years, because the later has been positively associated with the development of cardiovascular diseases. Therefore development of methods for TMA/TMAO reduction or elimination could have value in therapy and prevention. If the aim would be reducing the bacterial production of TMA/TMAO, more information about bacterial TMA producing pathways would be valuable.

At the time of the start of this study several bacterial enzymes capable of producing trimethylamine from nutrients such as carnitine, gamma-butyrobetaine (GBB) and choline were already identified. Two most relevant of these are CutC choline lyase producing TMA from choline and CntA carnitine oxygenase producing TMA from carnitine and GBB. Characterization of CutC and CntA enzymes may give new fundamental insights into functioning of glycy radical enzymes (GRE) and Rieske type oxygenases and also could be valuable for design of specific targeted inhibitors and evaluation of the contribution that these enzymes give to the total pool of TMA/TMAO in organism. Since CutC is a bacterial microcompartment (BMC) associated enzyme, a characterization of this supra-molecular

system would give new fundamental insights in the functioning of bacterial organelles that could be useful for synthetic biology and biotechnological applications.

The aim of this study was to investigate two most relevant and well-known TMA producing enzymes, CutC and CntA, and to give deeper fundamental insight into their properties, functioning and structure. The specific tasks for achieving this were:

- 1) Perform functional and structural characterization of CutC choline lyase from *Klebsiella pneumoniae*;
- 2) Examine the properties, assembly and structure of GRM2 type CutC associated bacterial microcompartments;
- 3) Examine the substrate specificity and oxygen dependency of CntA oxygenase.

## 1. LITERATURE OVERVIEW

### 1.1 TMA and TMAO in *Animalia* organisms

Trimethylamine (TMA) is a simple organic compound, a tri-methylated derivative of ammonia. It was predicted and chemically synthesized in 1851 (Hofmann, 1851). A gas in room temperature, it is volatile and therefore has to be stored under pressure as liquid or as a 40% solution in water. The most notable property of TMA is its intensive, strong odor, similar to one created by fish and decomposing biomass. Indeed, TMA has one of the lowest human odor detection thresholds found in nature, with studies reporting thresholds as low as 0.21 ppb (Leonardos et al., 1969, Amoore & Hautala et al., 1983, Stephens, 1971). These thresholds for related chemicals such as methylamine (32000 ppb), dimethylamine (340 ppb) or triethylamine (480 ppb) are several orders of magnitude higher (Amoore & Hautala, 1983). This high detection sensitivity is ensured by trace amine-associated receptor 5 (TAAR5), a GPCR receptor functioning as a vertebrate olfactory receptor with a high specific affinity for TMA and N-methylpiperidine (Liberles et al., 2006). TMA can induce different responses among different species: while for humans and rats TMA is repellent at any concentrations, for mouse it is an attractive pheromone at low and repellent at large concentrations (Liberles et al., 2006; Li et al., 2013).

Not surprisingly, after TMA chemical synthesis and characterization, it was soon discovered in herring-brine (Hofmann, 1853). It was later identified in other biological fluids such as urine (Dessaignes, 1856), blood and cerebrospinal fluid (Dorée and Golla 1911) and bile (von Zeynek 1899). TMA presence has also been detected in fungi (Hanna et al., 1932) and plants (Cromwell et al., 1966). Despite TMA presence in such diverse group of living organisms, all TMA producing enzymatic pathways discovered so far are bacterial, although for marine animals there are some evidence about endogenous production of TMA (Goldstein & Funkhouser, 1972; Seibel & Walsh, 2002).

Since TMA is a small molecule, soluble both in water and fat, it can easily cross biological membranes and spread into organism. So far there is no evidence about mammalian enzymes being capable of producing TMA, so it is assumed that in mammals it comes either directly from food or it is synthesized by microbiota bacteria from precursors in gastrointestinal system. Choline, L-carnitine and  $\gamma$ -butyrobetaine have been demonstrated as such precursors in humans (Koeth et al., 2013, Tang et al., 2013, Koeth et al., 2014). Products rich in choline such as eggs have been showed to increase TMAO amounts in human

organism (Miller et al., 2014) and the presence of TMA-producing microbiota species correlate with higher TMAO amounts in mice (Romano et al., 2015).

In mammals, TMA is further metabolized by hepatic flavin containing monooxygenase 3 (FMO3) into trimethylamine oxide (TMAO), an odorless compound (Lang et al., 1998). On average in Western diet about 50 mg of TMA is generated in human organism daily and almost all is eventually excreted with urine. More than 90% of this excreted TMA is usually in its metabolized TMAO form (Al-Waiz et al., 1987a, Zhang et al., 1996). Not every mammalian species oxidize TMA into TMAO with equal efficiency. For example, mice have significant FMO3 expression differences between male and female livers – while in females FMO3 expression is high, and most of the TMA is converted into TMAO (like in humans), in males FMO3 expression is repressed by testosterone (Falls et al., 1997) and is 1000-fold lower if compared to females (Li et al., 2013). As a result, most of the TMA in male mice remains unconverted and in urine reach milimolar concentrations. It has been recently demonstrated that a human molybdenum-containing mitochondrial reductase can reduce TMAO into TMA (Schneider et al., 2018). This suggests that a counter reduction of TMAO could also contribute to the overall TMA/TMAO balance in human organism.

For marine fish and cephalopods accumulation of TMAO in the tissue is thought to be a vital adaptation for protein stabilization against denaturation from hydrostatic pressure and osmolytes (Yancey et al., 1982; Yancey et al., 2001; Treberg et al., 2002). There is a positive correlation of the depth of the fish habitation with the amount of TMAO in tissue – in the shallow-living marine species the content of TMAO is as low as 40 mM/kg, while in deep-living fish species it can increase almost tenfold up to 386 mM/kg (Yancey et al., 2014). This is also supported by *in vitro* studies, demonstrating protein-stabilizing capabilities of TMAO (Yancey et al., 1982; Yancey et al., 1999; Zou et al., 2002; Sarma et al., 2013). The source of TMAO in marine animals is unknown. While capability to oxidize TMA to TMAO has been showed in various animals, this still gives no answer about the source of the precursory TMA. A food chain accumulation of TMA/TMAO from plankton to higher organisms could be an explanation, but there are some evidences that fish can keep homeostasis of TMAO in check even without nutritional supplementation of TMA or TMAO (Cohen et al., 1958; Goldstein & Palatt, 1974). Considering TMAO importance, relying only on external TMA/TMAO sources would be a risky strategy for keeping homeostasis. This points to a possibility of endogenous TMA production in fish, although direct evidences to support this are few (Goldstein & Funkhouser, 1972; Seibel & Walsh, 2002).

## 1.2 TMA and TMAO role in pathological processes

### 1.2.1 TMA and trimethylaminouria

The majority of TMA in human organism is usually oxidized into TMAO by hepatic FMO3 (Fig 1). However, insufficient activity of FMO3, caused by genetic mutations in FMO3 gene, can upset this balance (Dolphin et al., 1997; Treacy et al., 1998). Impaired oxidation of TMA results in its accumulation in the organism, and, since TMA can cross skin easily (Kenyon et al., 2004), the excess is excreted through skin. This is the cause of a genetic condition known as trimethylaminouria or fish odor syndrome, manifesting itself as a very unpleasant, fishy aroma of the body. Although this condition is not life-threatening and seems to have no direct negative impact on health, the psychological and social impacts can be major – those affected are often depressed, socially isolated and even suicidal (Christodoulou, 2012). True prevalence of trimethylaminouria is largely unknown, since the symptoms may vary among individuals and many, perhaps, even most cases are not reported; one estimate of the amount of heterozygous carriers in white British population is 0.5-1%, resulting in a frequency of 1 per 40000 births (Al-Waiz et al., 1987b). Trimethylaminouria is a polymorphic trait – more than 30 different missense and nonsense variations in FMO3 gene responsible for this condition have been described (Shephard et al., 2015). This accounts for the large difference of severity of symptoms among individual cases.

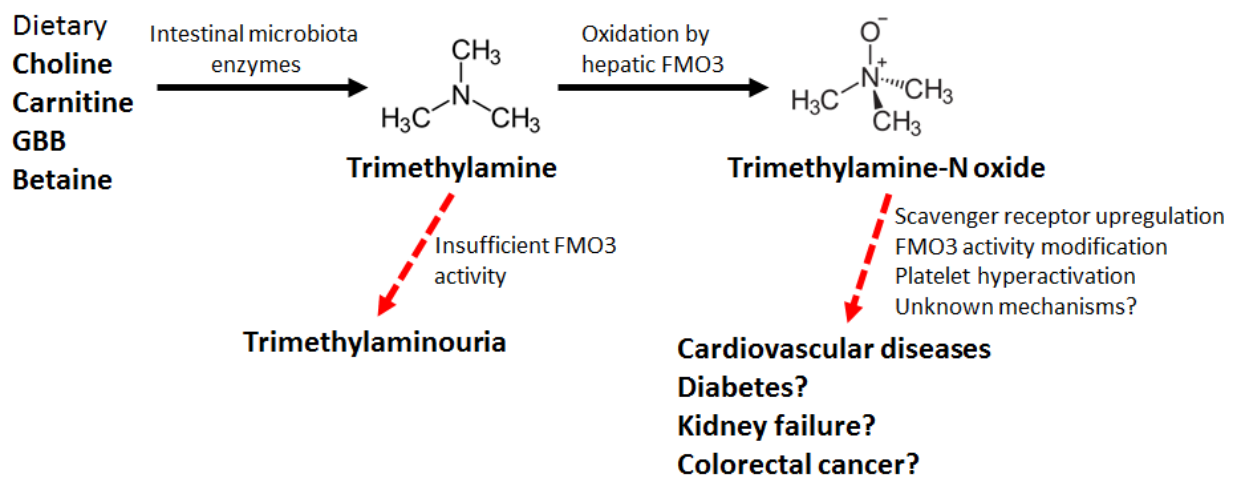


Fig 1. TMA and TMAO production and impact in animal and human organisms.

### 1.2.2 TMAO and associated conditions

TMA and especially its metabolite TMAO has been in intense spotlight during the last few years since a metabolomics study of clinical data revealed an association of elevated



TMAO with cardiovascular diseases (Wang et. al., 2011). Since then, numerous studies of clinical data have associated TMAO with several conditions (Fig 1) such as pathogenesis of cardiovascular diseases (Tang et. al., 2013; Trøseid et. al., 2015; Organ et. al., 2016; Randrianarisoa et. al., 2016), kidney failure (Tang et. al., 2015; Kim et. al., 2016; Missailidis, 2016), insulin resistance (Dambrova et. al., 2016; Tang et. al., 2017) and colorectal cancer (Bae et. al., 2014). Meta-studies have confirmed consistent correlation between TMAO and cardiovascular diseases, demonstrating its usefulness as a biomarker (Heianza et. al., 2017; Schiattarella et. al., 2017). Animal studies in mice (Wang et. al., 2011; Koeth et. al., 2013; Chen et. al., 2017; Geng et. al., 2018) and rat (Li et. al., 2017) have demonstrated development of atherosclerosis, endothelial damage and cholesterol accumulation in endothelium caused by TMAO. There are also data from mice studies suggesting TMAO involvement in development of diabetes (Gao et. al., 2014; Dambrova et. al., 2016). pyruvate and fatty acid oxidation impaired by TMAO has also been reported in mice cardiac mitochondria (Makrecka-Kuka M. et. al., 2017).

There are several suggested mechanisms for how TMAO could exercise its influence on mammalian organism. A direct upregulation of atherogenic macrophage scavenger receptors and stimulation of foam cell formation by TMAO was historically first reported pathway (Wang et. al., 2011). In this study an association between dietary choline degradation by microbiota to TMA, subsequent TMA conversion into TMAO by FMO3 in liver and atherosclerotic lesion development in mice was demonstrated. Analogous results were observed in another study with another TMA precursor – L-carnitine, where TMAO produced from L-carnitine in mice also promoted atherosclerosis and inhibited the reverse cholesterol transport from foam cells (Koeth et. al., 2013). Since then, other studies have also showed TMAO as being capable of inducing scavenger receptors (Geng et. al., 2018; Mohammadi et. al., 2016) and supporting the dysfunction of endothelium (Li et. al., 2017; Ma et. al, 2017).

Another suggested mechanism would be TMA/TMAO influence on the activity of FMO3, the TMA metabolizing enzyme and a key regulator of cholesterol metabolism. Upregulation of FMO3 cause increased reabsorption of cholesterol in intestines (Warrier et. al., 2015). Furthermore, FMO3 knockdown impairs glucose tolerance and prevents hypercholesterolemia and atherosclerosis in mice model (Miao et. al., 2015). Since TMA is a substrate of FMO3, it could be possible that TMA/TMAO increase the expression or activity of FMO3 and therefore could be both pro-atherogenic and pro-diabetic agent.

TMAO has been also observed as capable of causing calcium release in platelets, creating hyperreactivity and increasing the risk of thrombosis (Zhu et. al., 2016). This mechanism has been showed to be dependable on the composition of microbiota species in mice – transplantation of species capable of choline degradation into TMA had the observable effect of transferring the increased platelet reactivity as well (Skye et. al., 2018). It is possible that FMO3 can also influence the platelet reactivity (Zhu et. al., 2018), suggesting the possibility of two different TMA/TMAO effects exercised through FMO3.

### 1.2.3 Managing the amounts of TMA and TMAO in organism

Since high amounts TMA/TMAO has been involved in development of trimethylaminouria and possibly other medical conditions as well, creating strategies for lowering the amount of TMA/TMAO have medical significance. There are several potential strategies. Dietary management (Busby et. al., 2004) and low dosage wide-spectrum antibiotic treatment (Treacy et. al., 1995; Chalmers et. al., 2006) has been used in management of the symptoms of trimethylaminouria. They both have their drawbacks – limiting dietary choline and carnitine can have repercussions, since they are important nutrients, and antibiotics can be used only in short term in order to avoid dysbiosis. Another proposed strategy is targeting microbiota in a non-lethal way – either by remodeling the composition of gastrointestinal microbiota species or targeting particular bacterial enzymes with specific inhibitors. Resveratrol has been showed to modify the composition of gut microbiota and lowering the produced TMA amounts (Chen et. al., 2016). Specific inhibitors targeting choline degrading enzyme CutC have been showed as capable of preventing TMA production in bacterial cultures and mouse models (Wang et. al., 2015; Roberts et. al., 2018). In these studies this inhibitory effect was also protective against development of atherosclerotic lesions in mice. Meldonium, a chemical analogue of carnitine and GBB, has also been showed as capable of influencing production of TMA/TMAO in humans and rats (Dambrova et. al., 2013; Kuka et. al., 2014).

### 1.2.4 Criticism of TMA and TMAO as pathological agents

A certain amount of criticism has also been directed towards hypothesis of TMAO as a pathogenic agent (Fennema et. al., 2016, Cho & Caudill, 2016). A major missing part in this hypothesis is the fact that the exact effector receptor for TMA/TMAO has not been properly identified and confirmed, so it is unknown what pathways are underneath the observed TMAO effects. Meta-analysis of clinical data does not confirm direct causation between

TMAO and cardiovascular diseases, since higher TMAO levels in these studies were associated with generally increased mortality independently of the cause (Heianza et. al., 2017; Schiattarella et. al., 2017). Therefore, higher TMAO levels could be merely a consequence and just a marker of general health complications. Also, if TMA/TMAO is regarded as a cause of cardiovascular diseases, then there is a paradox of the generally accepted beneficiary effect of marine fish on human health and the high levels of TMAO in such foods. There is also a study showing TMAO having a beneficial effect against atherosclerosis (Collins et. al., 2016) and other studies showing that higher consumption of such TMA precursors as choline and betaine are not associated with elevated risks of cardiovascular diseases (Nagata et. al., 2015, Bidulescu et. al., 2007; Dalmeijer et. al, 2008). So whether TMA/TMAO in real-world, low dosage, long exposure conditions can substantially affect the development of any pathological condition in humans is still not proven or disproven convincingly.

### **1.3 Bacterial TMA producing enzymes**

#### **1.3.1 CutC choline lyase**

CutC choline lyase is a typical member of glyceryl radical enzyme (GRE) class, with a common trait of having a glyceryl radical in the active site. GRE are metabolically diverse. Most are catabolic – pyruvate formate-lyase (Knappe & Wagner, 1995; Becker et. al., 1999), glycerol dehydratase (O'Brien et. al., 2004), choline lyase (Craciun & Balskus, 2012), propane-1,2-diol dehydratase (LaMattina et. al., 2016), trans-4-hydroxy-L-proline dehydratase (Levin et. al., 2017), anaerobic ribonucleotide reductase (Sun et. al, 1996, Logan et. al., 1999) and isethionate sulfite lyase (Peck et. al., 2019). There is also an example of an anabolic GRE – benzylsuccinate synthase (Leuthner et. al., 1998; Funk et. al., 2014). Glyceryl radical enzymes are dimeric proteins (Fig 2A), having a 10- $\beta/\alpha$  barrel fold – the monomer core is made of a  $\beta$  barrel consisting of 10 strands and it is surrounded by outside  $\alpha$  helices.

Anaerobic bacterial choline degradation into TMA and acetaldehyde has been demonstrated already in the 1960s (Hayward & Stadtman, 1960), but the exact enzyme behind this process was identified only in 2012, when Craciun and Balskus (2012) demonstrated that this reaction in *Desulfovibrio desulfuricans* is performed by a GRE, designated as CutC (EC 4.3.99.4). This discovery was later confirmed with *in vitro* studies of purified CutC enzyme and mutagenic active site-mapping (Craciun & Balskus, 2012; Craciun et. al., 2014). CutC is a relatively widespread bacterial enzyme, with representatives of *Proteobacteria*, *Firmicutes* and *Actinobacteria* having the corresponding gene (Rath et. al., 2017).

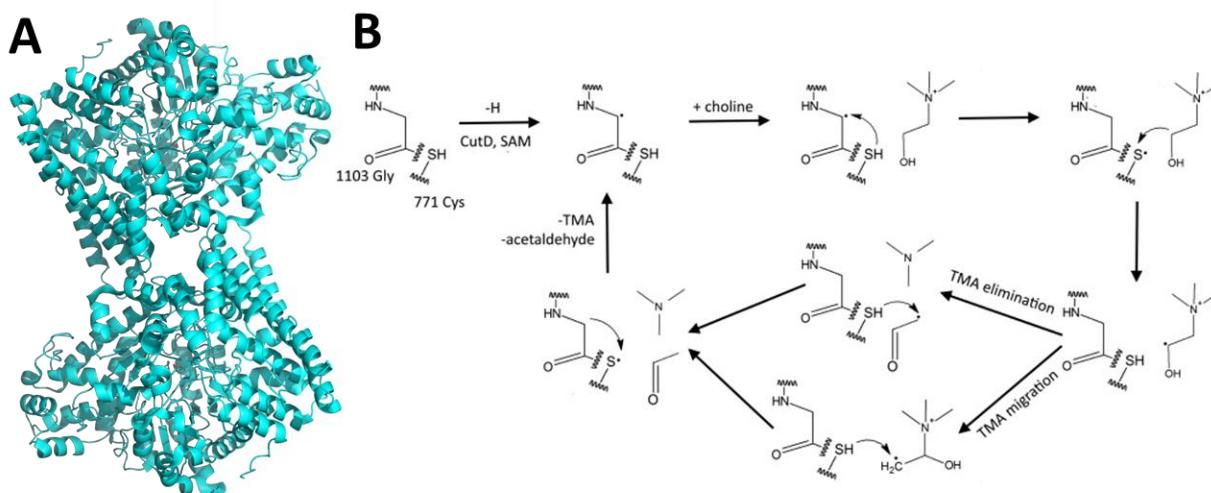


Fig 2. A: Crystal structure of glycerol dehydratase dimer (PDB 1R9D), a closely related enzyme of CutC. B: Proposed catalytic mechanism of CutC (according to Thibodeaux & van der Donk, 2012).

Like all GRE, CutC requires an activating enzyme for the initial generation of the radical in the active site. This reaction is performed by a special class of enzymes – glyxyl radical enzyme activating enzymes (GRE-AE), containing a 4Fe-4S cluster, coordinated via three cysteines (Shisler & Boderick, 2014). The  $[4\text{Fe}-4\text{S}]^{2+}$  cluster can be reduced by NADH and the fourth uncoordinated iron atom in the cluster then becomes capable of coordinating the S-adenosylmethionine molecule and cleaving it into methionine and the highly reactive 5'-deoxyadenosyl radical. This radical then further abstracts hydrogen atom from glyxyl radical enzyme glycine C2 atom, generating the glyxyl radical (Fig 2B). The glyxyl radical itself then further abstracts hydrogen atom from active site cysteine sulfur atom, creating a thiyl radical. The thiyl radical then abstracts the hydrogen atom from the substrate molecule, causing a molecular rearrangement of the substrate molecule. The nature of this rearrangement is not exactly clear, but there are evidence indicating that the TMA in CutC is eliminated directly, since there are several amino acids serving as proton donors and acceptors in the active site and TMA elimination pathway requires proton transport (Thibodeaux & van der Donk, 2012; Bodea et. al., 2016). After this step the glyxyl radical is then regenerated and the reaction cycle can start again without any additional steps. It has to be noted that the catalysis is strictly anaerobic – both GREs and GRE-AE are inactivated by the presence of oxygen, because oxygen degrades the partially coordinated GRE-AE 4Fe-4S cluster and by interacting with GRE glyxyl-radical it cuts the main protein chain in GRE (Wagner et. al., 1992; Yang et. al., 2009).

### 1.3.2 CutC associated bacterial microcompartments

CutC enzyme is not a free cytoplasmic enzyme, but rather a bacterial microcompartment (BMC) associated enzyme. Bacterial microcompartments are prokaryotic organelles consisting of a protein shell and enzymatic core performing concerted metabolic functions (Cheng et. al., 2008; Jorda et. al., 2013; Kerfeld et. al., 2018). According to their functions BMC can be divided in two large groups: metabolosomes and carboxysomes. The carboxysomes contain carbonic anhydrase and RuBisCo and their function is to perform carbon fixation. Metabolosomes are more diverse and are involved in breakdown of such compounds as choline, glycerine, 1,2-propanediol and ethanolamine. The metabolosomes typically contain signature enzyme performing the breakdown of the initial substrate and producing an aldehyde intermediate. This aldehyde intermediate is then further processed by an aldehyde dehydrogenase and alcohol dehydrogenase pair, producing alcohol and R-CoA products. The R-CoA is then finally processed by phosphotransacetylase (PTAC), producing free CoA and phosphorylated compound (acyl-P or propionyl-P). The latter is then transported to cytosol and converted by acyl kinase (AcK) into one carboxylic acid and one ATP molecule (Fig 3A; Zarzycki et. al., 2015).

The BMC is formed of three types of BMC proteins (Fig 3B): hexameric BMC-H (Kerfeld et. al., 2005; Sutter et. al., 2016), trimeric BMC-T (Heldt et. al., 2009; Klein et. al., 2009) and pentameric BMC-P (or BMV; Tanaka et. al., 2009; Wheatley et. al., 2013). The BMC shells are quasi icosahedral or icosahedral, with the BMC-H and BMC-T proteins forming the surfaces of BMC shells (Sinha et. al., 2014; Sutter et. al., 2016; Sutter et. al., 2017) and pentameric BMC-P capping the 5-fold icosahedron vertices (Wheatley et. al., 2013). The transport of metabolites and cofactors across the shell are thought to be ensured by pores formed inside the centers of BMC-H hexamers and BMC-T trimers (Klein et. al., 2009; Chowdhury et. al., 2015). The pore sizes vary – the BMC-H hexamers have smaller and consistently open 4-7 Å large pores, while for BMC-T trimers up to 14 Å larger pores both in open and closed states have been observed (Takenoya et. al., 2010; Thompson et. al., 2015).

The enzymatic core is encapsulated by encapsulations peptides (EP) – short, N or C-terminal amphiphilic helices ensuring both shell-core interaction and also crosslinking and stabilizing the core itself (Fan et. al., 2010; Fan et. al., 2012; Lawrence et. al., 2014; Jakobson et. al., 2017). The encapsulation of enzymatic pathways inside the BMC shell has several benefits. First, the intermediate substrate local concentrations are higher inside the BMC and this

speeds up the whole enzymatic pathway as a result. The enzyme encapsulation inside shell also has the benefit of helping to contain volatile or toxic metabolites, such as CO<sub>2</sub> and aldehydes, respectively. It is also possible that the shell gives additional protection against the oxygen damage in some anaerobic BMCs.

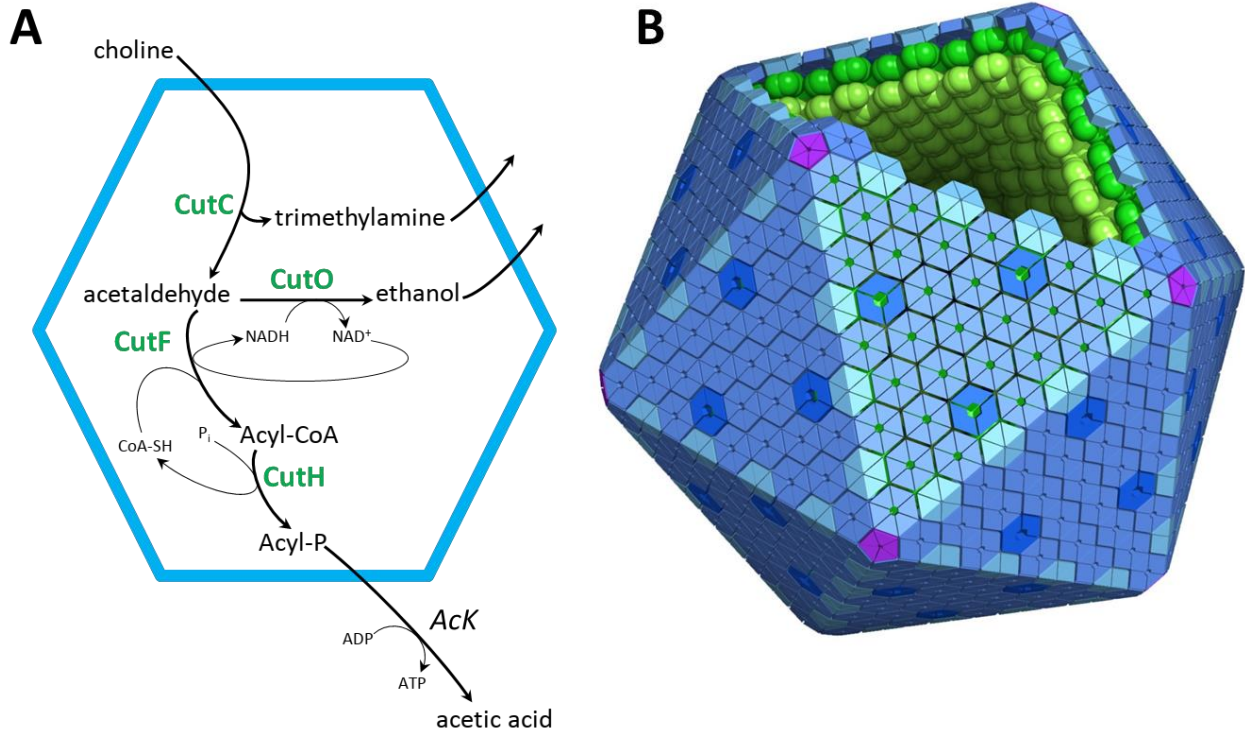


Fig 3. Structure and functioning of bacterial microcompartment. A: Scheme of general mechanism of a choline-metabolizing metabolosome. Blue – shell components, green – core enzymes. Image adapted from Zarzycki (2015). B: A generalized model of BMC. In dark blue – trimeric BMC-T proteins, in blue and light blue – hexameric BMC-H proteins, in purple – pentameric BMC-P proteins, in green – encapsulated enzymatic core proteins. Image adapted from Yeates (2013).

Several glycol-radical enzymes, including CutC, are signature enzymes of BMC. Such BMC are designated glycol-radical enzyme associated bacterial microcompartments (GRM) (Axen et. al., 2014; Zarzycki et. al., 2015). The GRM loci can be divided into subgroups based on the structure of the BMC loci and the identity of the signature enzyme. CutC choline lyase is a signature enzyme of two different GRM subgroups – GRM1 and GRM2. The GRM1 type BMC loci are contained by *Actinobacteria*, *Firmicutes* and *Deltaproteobacteria* while GRM2 type can be found almost exclusively in *Gammaproteobacteria* (Fig 4, Axen et. al., 2014).

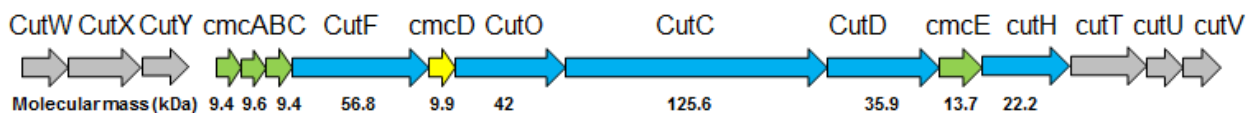


Fig 4. *Klebsiella pneumoniae* GRM2 locus. Structural shell BMC-H proteins cmcA, cmcB, cmcC and cmcE are colored in green, BMC-P protein cmcD is colored in yellow. Core enzymes CutF (aldehyde dehydrogenase), CutO (alcohol dehydrogenase), CutC (choline lyase), CutD (glycyl-radical activating enzyme) and CutH (phosphotransacylase) are colored in blue. Regulatory and transporter genes are colored in gray.

There are some notable differences between these two BMC loci types, with the GRM2 type having only five shell proteins versus eight for GRM1, GRM2 type CutC having an unusual, around 340 amino acids long N-terminal extension of unknown function and GRM1 locus having a BMC-T while GRM2 lacking it. The CutC N-terminal extension is unique among BMC signature enzymes and it has been proposed that its function is associated with either ensuring the encapsulation of enzymatic core or playing a role in core multimerization (Zarzycki et. al., 2015). The formation of GRM2 type BMC has been demonstrated with TEM analysis of cross-sections of *Escherichia coli* and *Proteus mirabilis* cells (Jameson et. al., 2016, Herring et. al., 2018).

### 1.3.3 CntA Rieske non-heme oxygenase

Rieske non-heme oxygenases are iron containing metalloenzymes with a coordinated iron atom in the active site and an additional 2Fe-2S cluster. Rieske oxygenases are mostly known for the capability of breaking aromatic rings (Barry & Challis, 2013), but there is also a variety of other substrates such as cholesterol (Van der Geize, 2007), stacchyrine (Daughtry et. al., 2012) and caffeine (Summers et. al., 2011). Not all Rieske oxygenases are prokaryotic – the choline monooxygenase can be found in plants (Rathinasabapathi et. al., 1997) and DAF-36/Nvd cholesterol oxygenase can be found in nematodes and insects (Yoshiyama-Yanagawa et. al., 2011). Rieske non-heme oxygenases form trimeric assemblies (Fig 5A). The 2Fe-2S clusters and mononuclear iron coordination sites are at the opposite ends of the monomers, and the catalytic site is formed between two adjacent monomers.

A particular two component Rieske oxygenase/reductase system, designated CntA (EC 1.14.13.239) and CntB was identified in *Acinetobacter baumannii* and demonstrated as capable of cleaving L-carnitine and GBB into trimethylamine and semialdehyde products



(Zhu et. al, 2014). In this study a lower TMA-producing CntA activity was also detected for glycine betaine and none was detected for choline.

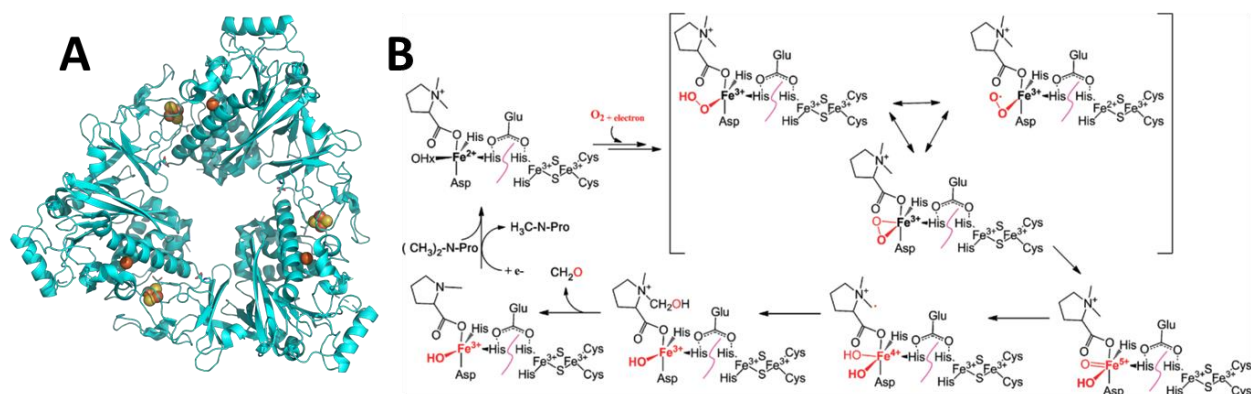


Fig 5. A: Crystal structure of Stc2 stacchydrene demethylase trimer (PDB 3VCA), a closely related enzyme of CntA. B: catalytic mechanism of Stc2 stacchydrene demethylase proposed by Daughtry (2012). Figure adapted from Daughtry (2012).

In contrast to CutC, the CntA/CntB genes are not widespread among different taxons of bacteria, and are mostly limited to *Gammaproteobacteria*, especially *Escherichia coli* strains (Rath et. al., 2017). Interestingly, this CntA/CntB oxygenase-reductase pair was already discovered earlier and designated as YeaX/YeaW, but without identifying the substrates (Boxhammer et. al., 2008). Another study analyzed CntA activity from *Escherichia coli* and discovered that the substrate panel was different, with a high activity for choline being the main contrast if compared with studies of *Acinetobacter baumannii* CntA (Koeth et. al., 2014). So the information about the exact substrate panel of CntA is somewhat controversial, and it could be also possible that it varies among different hosts. Since choline is an important contributor to the pool of TMA precursors, it is important to know how much CntA can contribute to its production from choline.

The exact catalytic mechanism of CntA is not known, but it is probably similar to other Rieske oxygenases. Stc2 stacchydrene demethylase can be mentioned as an example (Daughtry et. al., 2012). Stc2 is a Rieske oxygenase that is capable of performing a breaking of C-N bond by oxidative demethylation of stacchydrene (Figure 5B). For the catalytic cycle to start, the Rieske 2Fe-2S cluster of oxygenase has to be reduced by the reductase component that uses NADH as an electron donor. The reaction cycle is started when the substrate bounds to the Fe<sup>2+</sup> active site. In the next step oxygen is bound to the active site iron atom, forming a high-valent iron-oxo species. Then the substrate carbon atom in the cleavable C-N atom pair undergoes oxidative hydroxylation and the resulting product spontaneously decomposes, producing tertiary amine and aldehyde products.



#### 1.3.4 Other TMA producing bacterial enzymes

In the literature there are two other TMA producing bacterial enzymes described: TMAO reductase and ergothionase. TMAO reductase (EC 1.7.2.3) or TorA is a molybdenium containing metalloenzyme that is a part of bacterial TMAO respiratory (Tor) system (Méjean et. al., 1994; Dos Santos et. al., 1998). TorA uses TMAO as a terminal acceptor of electron transport chain and reduces it to TMA in anaerobic conditions when oxygen is not available. Another TMA producing enzyme is ergothionase which has been isolated from *Burkholderi* sp. HME13. This enzyme cleaves ergothioneine, a thiourea derivative of histidine, into TMA and thiolurocanic acid (Muramatsu et. al., 2013). Ergothioneine has antioxidant properties, and this is thought to be its main function in living organisms.

The contributions of these two enzymes to the total pool of TMA/TMAO are probably not major, because the TorA affects only the proportion of TMA and TMAO in the gastrointestinal system, but not the combined amount of these two compounds; and the ergothioneine is a relatively uncommon metabolite produced by only few bacterial and fungal species. There are also suggestions that there could be other bacterial enzymes responsible for TMA production not discovered yet, so the range of TMA-producing enzymes may be extended in the future (Rath et. al., 2017).

## 2. MATERIALS AND METHODS

### 2.1 Plasmid construction and protein expression

The DNA coding sequences for recombinant proteins used in this study were initially amplified with PCR from genomic DNA extracted from bacterial cultures. The CutC, CutD, CutF, CutO, cmcA-E and CntA genes were amplified with Phusion high-fidelity DNA polymerase and Pfu DNA polymerase (Thermo Fisher Scientific) according to manufacturer protocols. *Escherichia coli* BL21-DE3 were purchased from Sigma. *Klebsiella pneumoniae* MSCL535, *Serratia marcescens* MSCL1476 and *Providencia rettgeri* MSCL730 cultures were obtained from The Microbial Strain Collection of Latvia. GenBank entries ARRZ01000032.1 and CP018676.1 (*Klebsiella pneumoniae*), CP017671.1 (*Providencia rettgeri*), CP018930.1 (*Serratia marcescens*) and CP001509 (*Escherichia coli* BL21-DE3) were used for primer design. The amplified genes were cloned in pET-Duet 1 and pRSF-Duet 1 vectors in NcoI/HindIII and NdeI/XhoI restriction enzyme sites.

pET-Duet 1 and pRSF-Duet 1 vectors were chosen for our experiments because these are high copy plasmids and have two T7, isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG)-inducible promoters with multiple cloning sites. Since these two vectors contain different antibiotic resistance genes and origin regions, they can also be co-expressed, increasing the amount of available promoters to four. If more promoters were necessary, it was possible to amplify with PCR the whole Duet region with the particular proteins and clone it into the desirable vector using XhoI site, therefore increasing the total amount of available promoters from two to four in one plasmid.

The plasmids and amplified PCR fragments were digested with appropriate restriction enzymes, and analyzed by agarose gel electrophoresis in 1% agarose gel in standard TAE buffer. The obtained vector DNA was also additionally treated with FastAP thermosensitive alkaline phosphatase (Thermo Fisher Scientific) according to manufacturer protocol. The digested fragments were illuminated in agarose gel using ethidium bromide and ultraviolet light, cut out from gel and eluted by using GeneJET Gel Extraction Kit (Thermo Fisher Scientific) according to manufacturer protocol. The digested fragments were then ligated into vectors by using T4 DNA Ligase (Thermo Fisher Scientific) according to manufacturer protocol.

The ligated plasmids were then transformed in chemically competent *Escherichia coli* XL-1 Blue cells and seeded on agarized plates containing antibiotics (50  $\mu$ g/ml kanamycin or 100  $\mu$ g/ml ampicillin). Individual colonies were seeded in 5 ml of LB medium with 25  $\mu$ g/ml

kanamycin for pRSF-Duet 1 and 50 µg/ml ampicillin for pET-Duet 1 vectors and grown overnight with shaking at +37°C. The plasmid DNA was then extracted by using the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) according to manufacturer protocol. The correct plasmids were selected by agarose gel electrophoresis after analytical digestion and confirmed by Sanger sequencing using BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) according to manufacturer protocol.

The proteins were typically expressed in *Escherichia coli* BL21-DE3 cells. The seeding stock was made by seeding several colonies from agarized plates into LB medium containing antibiotics and incubating it overnight at +37°C. The seeding stock was then added to 2xTY medium in volume ratio of 1:20. The 2xTY medium was supplemented with 25 µg/ml kanamycin and/or 50 µg/ml ampicillin for corresponding plasmids. The expression was induced with addition of 1 mM IPTG upon reaching mid-log phase (typically 0.6-0.7 OD<sub>540</sub>). The temperature was lowered to 20°C and proteins were expressed for 16-18 hours with shaking at 200 rpm.

## 2.2 Protein purification

For the purification of 6xHis tagged proteins (CutC, CutD, CntA, CntB, CutO, CutF and CutH) a conventional Ni<sup>2+</sup> affinity chromatography was used. The cells were lysed with ultrasound in lysis buffer typically containing 40-100 mM Tris-HCl (pH 8.0), 200 mM NaCl, 1% Triton X-100, 1 mM PMSF and 1 mM dithiothreitol (DTT). The lysate was cleared after centrifugation at 12000-14000g and passed through 1 or 5 ml HisTrap columns (GE Healthcare). The proteins were eluted with 300 mM imidazole. The purity of eluted proteins was confirmed with sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) stained with Coomassie blue. For additional purification proteins were purified on Superdex 200 gel filtration columns. CutD was purified in denaturizing conditions in presence of 7 M urea and refolded further.

For the purification of bacterial microcompartment derived shell particles (BDPs) a sedimentation based method was developed according to other cases mentioned in literature (Havemann & Bobik, 2003; Cai et. al., 2016). Cell lysis was performed enzymatically in a lysis buffer containing 20-40 mM Tris-HCl (pH 8.0), 300 mM NaCl, 20 mM MgCl<sub>2</sub>, 1 mM PMSF, 1 mM DTT, 1 mg/ml lysozyme and 0.1 mg/ml deoxyribonuclease. The suspension was incubated at +4-7 C° with shaking for 1 hour. The lysate was then purified from larger debris by centrifugation at 10000g for 10 minutes. The supernatant was then collected and centrifuged for 3 hours at 50000g. The glassy pellet containing BDPs were then suspended in

a buffer containing 20 mM Tris-HCl (pH 8.0) and 300 mM NaCl, centrifuged at 10000g for 10 minutes and then loaded on 16/900 Superose 6 gel filtration column equilibrated in 20 mM Tris-HCl (pH 8.0) and 300 mM NaCl. 2 ml fractions were collected and analyzed with SDS-PAGE and transmission electron microscopy (TEM). The BDPs were usually present in fractions from 60 to 100 ml.

### **2.3 MALDI-TOF-MS analyses**

For analysis of CutC degradation, confirmation of CutO and CutF identities and identification of protein content in GRM2 BDPs matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) was used. This method enables to determine exact molecular masses of proteins and positively identify unknown proteins by peptide mass fingerprint.

For analysis of the suspected zones of CutO and CutF the corresponding protein band was cut out of PAGE gel. The gel fragments were then washed twice for 1 hour with 500  $\mu$ l of 0.2 M ammonium bicarbonate and 50% acetonitrile to remove the Coomassie stain. Afterwards the gel fragments were washed twice with 200  $\mu$ l of 100% acetonitrile for 15 minutes and incubated with trypsin (Sigma-Aldrich, cat. No T6567) solution in 40 mM ammonium bicarbonate and 10% acetonitrile for 3 hours at +37 °C.

The peptide or protein solution (1  $\mu$ l) was then mixed with 1  $\mu$ l of 0.1% TFA and 1  $\mu$ l of matrix solution containing 15 mg/ml 2,5-dihydroxyacetophenone in 20 mM diammonium citrate and 75% ethanol. 1  $\mu$ l of the obtained mixture was loaded on the target plate and dried. Samples were analysed using a Bruker Daltonics Autoflex mass spectrometer.

### **2.4 X-ray crystallography analysis of CutC**

X-ray crystallography is the most popular method for protein 3D structure determination. Because in earlier studies several GRE has been successfully crystallized (Becker et. al., 1999; O'Brien et. al., 2004; Funk et. al., 2014), it was reasonable to assume that CutC could also be crystallized.

The initial screening of CutC crystallization conditions were performed with commercial JCSG+, Structure screen 1&2 and Pact Premier screens (Molecular Dimensions). We used standard MRC plates with sitting drop setup. The drops were made by mixing 0.4  $\mu$ l or 1  $\mu$ l mother liquid and CutC solution drops. For crystallization trials CutC was concentrated to 16 mg/ml in 20 mM Tris-HCl (pH 8.0). Small crystals grew within a week. The best results for CutC were from PactPremier screen: H9 well for CutC with an addition

of 5 mM choline and F9 well for CutC without choline. The crystallization conditions were further optimized by adjusting the concentrations of PEG and salts, and the final crystallization conditions for CutC in the presence of choline were 20% v/w PEG 3350, 20–60 mM potassium/sodium tartrate, and 100 mM Bis-Tris (pH 8.5). Most optimal crystallization conditions for choline-free CutC were 20% (v/w) PEG 3350, 100–160 mM potassium/sodium tartrate and 100 mM Bis-Tris (pH 6.5).

For cryoprotection the crystals were soaked in 30% v/w glycerol in mother liquid and frozen in cryoloops in liquid nitrogen. The diffraction data were collected on MAX-lab MAXII synchrotron beamline I911-3 in Lund, Sweden.

The diffraction data were indexed with MOSFLM (Leslie, 1992) and scaled with SCALA (Evans, 1997) in CCP4 pipeline. The phases were determined with molecular replacement in MOLREP (Vagin & Teplyakov, 1997) by using a CutC homology model built from glycerol dehydratase (Protein Data Bank ID 1R9D, O'Brien et. al., 2004). The CutC choline bound and choline free crystals formed different crystals – the former had an orthorhombic lattice and space group P212121 with 8 molecules in the asymmetric unit and the later had a monoclinic lattice and space group P21 with 4 molecules in the asymmetric unit. Models were further built and refined with COOT (Emsley & Cowtan, 2004) and REFMAC5 (Murshudov et. al., 1997).

## **2.5 Transmission electron microscopy analysis**

A negative staining method with uranyl acetate was used for the visualization of purified BDPs particles. Formvar coated Cu 200 grids were used. 5 µl of sample were put on the grid and incubated for 3 minutes. The grids were dried, briefly washed with 1 mM EDTA solution and stained with 1% uranyl acetate for 1 minute. Grids were dried and analysed by using JEM-1230 TEM electron microscope at 100 kV.

## **2.4 Cryo-EM single particle analysis of BDPs**

BDPs containing encapsulated CutC<sub>336-1128</sub> C-terminal domain were analyzed by cryogenic electron microscopy (cryo-EM) single particle analysis (SPA). In last few years a significant improvement of cryo-EM resolution limits has been achieved by introduction of direct electron detectors and phase plates, thus enabling to achieve 3D reconstructions in near-atomic resolution range. In SPA analysis the sample is vitrified and the 3D reconstruction is calculated from 2D projections of different particles in TEM images. The

cryo-EM method can illuminate asymmetric features impossible to see in crystal structures. An additional benefit of cryo-EM SPA is the possibility of sorting the particles in 3D or 2D classes, thus enabling to illustrate different states or objects. BDPs are especially suitable for cryo-EM analysis since these particles are large and easy to pick and they have icosahedral or quasi-icosahedral symmetries, enabling to make 3D reconstructions in high symmetry point groups.

The BDPs were concentrated with ultrafiltration to 1 mg/ml in a buffer solution containing 100 mM NaCl and 20 mM Tris-HCl (pH 8.0). Vitrobot Mark IV (Thermo Fisher Scientific) was used for automated sample freezing. 4  $\mu$ l of BMC sample were applied to a Cu 200 R2/1 grid (Quantifoil), blotted for 4 seconds at 18°C temperature and plunge-frozen in liquid ethane-propane. The samples were stored in liquid nitrogen until further use.

Cryo-EM data were collected with Talos Arctica microscope (Thermo Scientific) equipped with Falcon 3EC direct electron detector (Thermo Scientific) operating at 200 kV. A total amount of 1316 images were collected in automated manner using EPU software (Thermo Scientific). The micrographs were recorded at 120000 $\times$  magnification with defocus range between -1.4 and -3.0  $\mu$ m and the pixel size of 1.23  $\text{\AA}$ /px. The images were stored as a stack of 40 frames. The exposure of one frame stack was 60  $e/\text{\AA}^2$  and the exposure time was 1 second.

The motion-correction and dose-weighting of frames were done with MotionCor2 (Zheng et. al., 2017) and the contrast transfer function (CTF) were evaluated with Gctf (Zhang, 2016). The whole single particle analysis process was performed in RELION 3.0 pipeline (Zivanov et. al., 2018). RELION 3.0 offers a single pipeline for all basic cryo-EM applications and the most resource intensive tasks are optimized for GPU computing, thus enabling to use for calculations a high end home computer instead of a server cluster. Log- and reference based autopicking and manual inspection resulted in 62533 selected particles. A subsequent 2D classification revealed significant heterogeneity among particles and, accordingly, five different initial groups of 2D classes could be distinguished (Fig 6). Two initial 3D models were made, one in I symmetry and second in D5 symmetry.

Using the initial I model two 3D classifications were performed – one in I symmetry and one in C1 symmetry. The I symmetry classes were very similar and both were used for final high resolution refinement with additional Bayesian polishing and CTF refinement steps, resulting in a map with a resolution of 3.3  $\text{\AA}$ . The C1 3D classification revealed a class with one missing pentameric unit, for whom a high resolution map could be refined to 8.8  $\text{\AA}$  resolution. Second initial model in D5 symmetry was used for 3D classification and 3D

refinement steps in D5 symmetry, resulting in a map with a resolution of 9.6 Å (Fig 6).

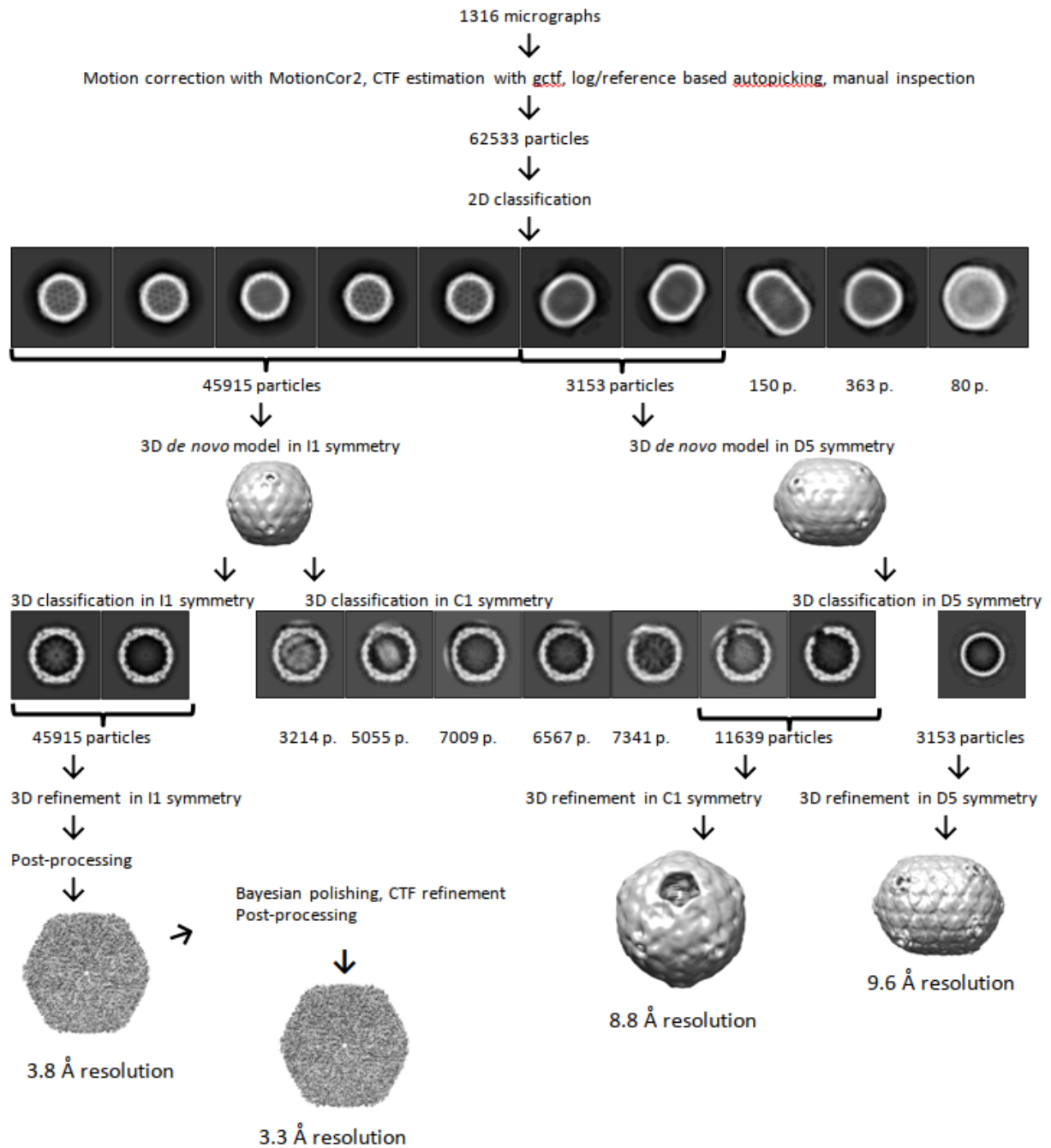


Fig 6. Scheme of cryo-EM analysis of BDPs.

Since high similarity X-ray crystallographic structures were available both for cmcD (PDB ID 4N8X) and cmcC (PDB ID 4QIV), we used homology models as initial models. Initial fitting of models into maps were done in UCSF-Chimera (Pettersen et. al., 2004). The models were then built further using Coot (Emsley & Cowtan, 2004) and refined with REFMAC5 in CCP-EM package (Burnley et. al., 2017) and PHENIX (Adams et. al., 2011).

## 2.5 Enzymatic *in vitro* activity assays of CutC and CntA

In order to confirm the activity of the purified TMA producing enzymes, methods for determination of CutC and CntA enzymatic activities were developed.

CutC activating enzyme CutD is necessary for the generation of a glyceryl radical in the active site of CutC; it is indispensable for CutC activity. The CutD, as often is the case for GRE-AE, is insoluble when overexpressed and it is also oxygen sensitive, thus requiring a reactivation procedure (O'Brien, 2004). We developed a method for reactivation of CutD where it was solubilized in 7 M urea, purified in denatured state by a Ni<sup>2+</sup> affinity chromatography and refolded anaerobically by dialyzing it against a refolding buffer containing 50 mM KCl, 25 mM MOPS/KOH (pH 7.5), 5 mM dithiothreitol, 0.2 mM Na<sub>2</sub>S, and 0.25 mM (NH<sub>4</sub>)<sub>2</sub>Fe(SO<sub>4</sub>)<sub>2</sub>. Then CutC and CutD activity assay was performed in an argone atmosphere in a buffer containing 25 mM Tris-HCl (pH 8.0) and 50 mM NaCl. The final concentration of proteins was 5 mg/ml for CutC and 4 mg/ml for CutD, and the reaction mix was supplemented with 3 mM S-adenosyl methionine, 1 mM choline and 5 mM sodium dithionite. The mixture was incubated for 15 hours at room temperature and the reaction was stopped by addition of formic acid to 50% v/v concentration. The amount of the produced TMA was measured by using UPLC-MS/MS.

CntA is a Rieske oxygenase and for its activation it requires for its 2Fe-2S cluster to be in a reduced state. This can be achieved both by the reductase component that uses NADH as an electron donor or by a direct reduction of CntA 2Fe-2S cluster by sodium dithionite (Daughtry et. al., 2012). The activities of CntA were tested with both methods. CntA was concentrated to a concentration of 0.1 mg/ml and transferred to a buffer containing 20 mM Tris-HCl (pH 8.0) and 50 mM NaCl. Substrates were added to 2 mM concentrations. Sodium dithionite was added to 0.5 mM concentration. Alternatively, instead of sodium dithionite, CntB was added to a concentration of 0.1 mg/ml and NADH was added to a concentration of 0.5 mM. The reaction was stopped by addition of formic acid to 5% v/v concentration. The amount of the produced TMA was measured by using ultra performance liquid chromatography - tandem mass spectrometer (UPLC-MS/MS).

The activities of CutC/CutD and CntA were tested also by expression of the particular enzymes in *Escherichia coli* BL21-DE3 cells. Night cultures of transformed cells harboring pRSF-CutC, pET-CutD and pRSF-CntA plasmids were seeded in small volumes (1-15 ml) of LB or 2xTY medium, supplemented with 1 mM choline (for detection of CutC activity) or 5 mM carnitine (for detection of CntA activity). The cultures were induced with 0.5-1 mM IPTG after reaching A<sub>540</sub> > 0.3 overnight. The bacterial culture was centrifuged at 5000g for



10 minutes and the supernatant was collected. Formic acid was added to the obtained solution to 5% v/v concentration. The amount of the produced TMA was measured by using ultra performance liquid chromatography - tandem mass spectrometer (UPLC-MS/MS).

## 3. RESULTS

### 3.1 Structure and function of CutC choline lyase

- **Isolation, recombinant expression and purification of CutC from *Klebsiella pneumoniae***
- ***In vitro* and *in vivo* tests of CutC activity**
- **Solving of CutC crystal structure**
- **Two conformations of CutC in choline free and choline bound states reveal conformational changes triggered by binding of choline to its active site**

# Structure and Function of CutC Choline Lyase from Human Microbiota Bacterium *Klebsiella pneumoniae*\*

Received for publication, June 10, 2015, and in revised form, July 16, 2015. Published, JBC Papers in Press, July 17, 2015, DOI 10.1074/jbc.M115.670471

Gints Kalnins<sup>†1</sup>, Janis Kuka<sup>§</sup>, Solveiga Grinberga<sup>§</sup>, Marina Makrecka-Kuka<sup>§</sup>, Edgars Liepinsh<sup>§</sup>, Maija Dambrova<sup>§</sup>, and Kaspars Tars<sup>†‡¶1</sup>

From the <sup>†</sup>Latvian Biomedical Research and Study Center, LV-1067 Riga, the <sup>§</sup>Latvian Institute of Organic Synthesis, LV-1006 Riga, and the <sup>¶</sup>University of Latvia, LV-1586 Riga, Latvia

**Background:** The bacterial glycol radical enzyme CutC converts choline to trimethylamine, a metabolite involved in pathogenesis of several diseases.

**Results:** The structures of substrate-bound and substrate-free CutC revealed significant differences.

**Conclusion:** Choline binding to the active site triggers a conformational change from the open to closed form.

**Significance:** A novel substrate-driven conformational mechanism and a potential target for drug design have been identified.

CutC choline trimethylamine-lyase is an anaerobic bacterial glycol radical enzyme (GRE) that cleaves choline to produce trimethylamine (TMA) and acetaldehyde. In humans, TMA is produced exclusively by the intestinal microbiota, and its metabolite, trimethylamine oxide, has been associated with a higher risk of cardiovascular diseases. Therefore, information about the three-dimensional structures of TMA-producing enzymes is important for microbiota-targeted drug discovery. We have cloned, expressed, and purified the CutC GRE and the activating enzyme CutD from *Klebsiella pneumoniae*, a representative of the human microbiota. We have determined the first crystal structures of both the choline-bound and choline-free forms of CutC and have discovered that binding of choline at the ligand-binding site triggers conformational changes in the enzyme structure, a feature that has not been observed for any other characterized GRE.

Trimethylamine (TMA)<sup>2</sup> is a tertiary amine that is produced by bacteria from the quaternary amine compounds choline (1, 2) and carnitine (3), which are present in food. The human organism is incapable of synthesizing TMA, and therefore, the bacteria of the microbiota are the only source of this metabolite. After absorption in the bloodstream, TMA is further metabolized to trimethylamine oxide (TMAO) by FMO3 (flavin-containing monooxygenase 3) in the liver. Defective or insufficient amounts of FMO3 cause TMA accumulation in the blood, which manifests as trimethylaminuria, or fish odor syndrome (4). Recent studies have linked the increased availability of ter-

tiary amines with a higher risk of cardiovascular disease via an intestinal microbiota-dependent pathway (5). High levels of TMAO, which are produced from choline (6) and carnitine (3) with the assistance of the intestinal microflora, have been shown to be a cause of atherosclerosis in mice. Studies with human volunteers (3, 7) have confirmed TMA production from carnitine and choline by the intestinal microflora, and clinical data cross-examinations have associated high levels of TMAO with an elevated risk of atherosclerosis (7), severe heart failure (8), and renal failure (9). Lowering TMA and TMAO levels in humans could therefore have potentially preventive and therapeutic effects on trimethylaminuria and cardiovascular diseases. TMAO also increased glucose tolerance and adipose tissue inflammation in mice fed a high-fat diet (10), suggesting involvement in the development of diabetes. Structural biology studies of microbiota enzymes are important to guide the discovery of novel inhibitor compounds.

CutC choline lyase, a TMA-producing glycol radical enzyme (GRE), has recently been identified and characterized (11, 12). A common feature of all GREs is a glycol radical in the active site. The GRE-activating enzyme, a 4Fe-4S cluster protein, cleaves *S*-adenosylmethionine (SAM) into a radical (13) that further generates a glycol radical on the active site glycine. The structures of numerous GREs that catalyze different reactions, such as pyruvate formate-lyase (PFL) (14), ribonucleotide reductase (15), glycerol dehydratase (GD) (16), and benzylsuccinate synthase (BSS) (17), have been determined. Along with the activating enzyme CutD, CutC cleaves choline, forming TMA and acetaldehyde under anaerobic conditions. Under aerobic conditions, catalysis is impossible because the glycol radical readily reacts with oxygen, cleaves the main chain, and irreversibly inactivates the enzyme (18).

The *cutC/cutD* gene cluster is possessed by several representatives of the human intestinal microbiota, including bacteria in the genera *Klebsiella*, *Escherichia*, *Streptococcus*, and *Enterobacter* (11). The mean abundance of representatives of these genera in the stool microbiota is low, but in certain individuals, it can rise to 0.56% for *Klebsiella pneumoniae* and to 96% for *Escherichia coli* (19). Not all bacteria of the microbiota are capable of producing TMA, including *Bacteroides*, the most abun-

\* This work was supported by State Research Program BIOMEDICINE and Bio-Struct-X Grant Proposal ID 7869. The authors declare that they have no conflicts of interest with the contents of this article.

The atomic coordinates and structure factors (codes 5A0U and 5A0Z) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

<sup>1</sup> To whom correspondence should be addressed: Latvian Biomedical Research and Study Center, Ratsupites 1, LV-1067 Riga, Latvia. Tel: 371-6780-8200; Fax: 371-6744-2407; E-mail: gints@biomed.lu.lv.

<sup>2</sup> The abbreviations used are: TMA, trimethylamine; TMAO, trimethylamine oxide; GRE, glycol radical enzyme; SAM, *S*-adenosylmethionine; PFL, pyruvate formate-lyase; GD, glycerol dehydratase; BSS, benzylsuccinate synthase; Bistris propane, 1,3-bis[tris(hydroxymethyl)methylamino]propane; UPLC, ultra performance liquid chromatography.



dant genus in stool (19, 20). Therefore, even a small increase in the amount of TMA-producing microorganisms could substantially affect the amounts of TMA and TMAO in an individual. These findings suggest that the CutC/CutD metabolic pathway could be accountable for a significant portion of TMA production in human intestines.

It has been shown that the production of TMAO can be suppressed by shifting the microbial degradation pattern of supplemental/dietary tertiary amines (21). In addition, the cardioprotective drug meldonium, an inhibitor of L-carnitine biosynthesis and transport (22), has been shown to lower TMAO levels in both rats and humans (21, 23) through inhibition of TMA production by the intestinal microbiota.

CutC/CutD enzymatic activity in *Desulfovibrio alaskensis* has previously been demonstrated both *in vivo* and *in vitro* (11, 12). Our model organism, *K. pneumoniae*, has previously been shown to be capable of producing TMA from both choline and carnitine (21), but the activity of the CutC/CutD enzymes in this organism has not been demonstrated. To our knowledge, no crystal structure of CutC has been determined until now. Here, we report the crystal structure of the CutC enzyme from *K. pneumoniae* in both its choline-free and choline-bound forms.

### Experimental Procedures

**Plasmid Construction and Protein Expression**—Genomic DNA was extracted from *K. pneumoniae* culture (The Microbial Strain Collection of Latvia) by proteinase K treatment and precipitation with ethanol. CutC (GenBank™ accession number EPO20241.1) and CutD (accession number EPO20361.1) sequences from the GenBank™ accession number ARRZ01000032.1 entry were used for primer design (CutC, ATATTCATGACGGCACACTACAACTTAACGCCGC (forward) and AATTAAGCTTTTAGAACTTCTCAATCACCGTACGGC (reverse)); and CutD, TATAGGCCTCATCGCAAAAACAAGAATTAACGGG (forward) and AATTCCATGGTTAATGGCGGACTAAGCGAATATC (reverse)). The *cutC* and *cutD* genes were amplified by PCR and cloned into pRSFDuet and pET1 vectors containing N-terminal His<sub>6</sub> tags with tobacco etch virus protease cleavage sites. The constructs were sequenced for confirmation. The proteins were expressed in *E. coli* BL21(DE3) cells. Upon reaching  $A_{540} > 0.3$ , induction was performed with 1 mM isopropyl  $\beta$ -D-thiogalactopyranoside in 2 $\times$  Tryptone/yeast extract medium for 3 h at +37 °C.

**CutC Purification**—Frozen cell paste was suspended in lysis buffer (100 mM Tris-HCl (pH 8.0), 200 mM NaCl, 1% Triton X-100, 20 mM MgSO<sub>4</sub>, 0.1 mg/ml DNase, 1 mg/ml lysozyme, 1 mM PMSF, and 2 mM DTT). Cells were lysed by ultrasound, and the lysate was centrifuged at 14,000  $\times$  g for 40 min. CutC protein was purified by nickel affinity chromatography using a 1-ml HisTrap column (GE Healthcare). For this step, 20 mM imidazole in 40 mM Tris-HCl (pH 8.0) and 300 mM NaCl was used as a washing buffer, and 300 mM imidazole in 40 mM Tris-HCl (pH 8.0) and 300 mM NaCl was used as an elution buffer.

**Limited Proteolysis**—For preparative purposes, the 124-kDa CutC protein obtained in HisTrap elution buffer was immediately digested with chymotrypsin using 1 mg of chymotrypsin/100 mg of CutC protein for 30 min at room temperature. The

90–92-kDa CutC fragment was separated from the digestion products and chymotrypsin by gel filtration on a Superdex 200 column in 20 mM Tris-HCl (pH 8.0). The protein was concentrated by ultrafiltration and stored in 50% glycerol at –20 °C.

For analytic purposes, after purification from the HisTrap column, the 124-kDa CutC protein was transferred to 20 mM Bistris propane buffer at pH 6.5 and 8.5 via ultrafiltration. Trypsin was added at 1 mg of chymotrypsin/100 mg of CutC protein. Digestion was performed for 6 h at room temperature. The CutC molecular mass after cleavage was characterized by both SDS-PAGE and MALDI-TOF-MS.

**CutD Purification and Reconstitution**—Cell lysis was performed as described for CutC. After lysate centrifugation, the pellet was resuspended in 7 M urea with 5 mM DTT. CutD was purified by nickel affinity chromatography using a 1-ml HisTrap column as described for CutC, except all of the buffers were supplemented with 7 M urea. After purification, the CutD solution was sparged with argon and dialyzed overnight in anaerobic solution containing 150 mM KCl, 25 mM MOPS/KOH (pH 7.5), 5 mM DTT, 20 mM Na<sub>2</sub>S, and 25 mM (NH<sub>4</sub>)<sub>2</sub>Fe(SO<sub>4</sub>)<sub>2</sub>. After dialysis, the CutD solution was desalted and purified from precipitates on a 1-ml Zeba desalting column, equilibrated with 25 mM Tris-HCl (pH 7.5) and 50 mM NaCl. All steps after CutD purification with the HisTrap column were performed in an argon atmosphere in a disposable glove bag (Sigma-Aldrich).

**In Vitro Activity Assay**—The activity assay was performed in 25 mM Tris-HCl (pH 8.0), 50 mM NaCl, 3 mM SAM, 1 mM choline, 5 mM sodium dithionite, 5 mg/ml CutC (untreated or treated with chymotrypsin), and 4 mg/ml CutD. All stock solutions were sparged beforehand with argon for 10–15 min in a disposable glove bag. Samples with all components except SAM, CutC, or CutD were incubated as negative controls. Incubation was performed in an argon atmosphere at room temperature for 15 h. After incubation, the solution was mixed with 100% (50:50, v/v) formic acid and stored at –20 °C. The TMA quantity was analyzed by ultra performance liquid chromatography (UPLC)-MS/MS.

**In Vivo Activity Assay**—For the *in vivo* activity assay, 0.5 ml of *E. coli* BL21(DE3) night cultures harboring pRSF-CutC, pET-CutD, empty pET vector, and cotransfected pRSF-CutC and pET-CutD were inoculated in 15 ml of LB medium in Hungate tubes (Chemglass). The LB medium was sparged beforehand with argon for 30 min and supplemented with 0.4 M NaCl, 1 mM choline chloride, and 0.069 mg/ml ferric ammonium citrate as described previously (11). The cultures were induced with 0.5 mM isopropyl  $\beta$ -D-thiogalactopyranoside after reaching  $A_{540} > 0.3$ . Incubation was performed for an additional 18 h at 37 °C with shaking. The cell culture was centrifuged, the supernatant was mixed with 100% (50:50, v/v) formic acid, and the TMA quantity was analyzed by UPLC-MS/MS.

**CutC Crystallization and Data Collection**—The protein was concentrated by ultrafiltration to 16 mg/ml in 20 mM Tris-HCl (pH 8.0). To obtain the substrate-bound form, choline chloride was added to a final concentration of 5 mM. CutC was crystallized using the sitting drop method, mixing 1  $\mu$ l of protein solution with 1  $\mu$ l of precipitant (20% (v/w) PEG 3350, 20–60 mM potassium/sodium tartrate, and 100 mM Bistris (pH 8.5)). CutC



## Structure and Function of CutC Choline Lyase

crystals without choline were obtained in slightly different conditions (20% (v/w) PEG 3350, 100–160 mM potassium/sodium tartrate, and 100 mM Bistris (pH 6.5). Crystals appeared overnight. Crystals were soaked in cryoprotectant (30% glycerol in mother liquid) and frozen in liquid nitrogen. Data were collected at MAX-lab Synchrotron beamline I911-3 (Lund, Sweden). The best diffracting crystals were of the choline-bound form and diffracted up to 2.4 Å resolution. The best crystals without choline diffracted to 3.0 Å resolution.

**MALDI-TOF-MS**—Crystallization drop solution (1 μl) was mixed with 1 μl of 0.1% TFA and 1 μl of matrix solution containing 15 mg/ml 2,5-dihydroxyacetophenone in 20 mM ammonium citrate and 75% ethanol. Then, 1 μl of the obtained mixture was loaded on the target plate, dried, and analyzed using a Bruker Daltonics Autoflex mass spectrometer.

**UPLC-MS/MS**—Amount of TMA was measured by UPLC-MS/MS. A Micromass Quattro Micro tandem mass spectrometer in positive ionization electrospray mode with a Waters Acquity UPLC system was used to perform the analysis. Chromatographic separation was achieved on an Acquity UPLC BEH HILIC column (2.1 × 50 mm, 1.7 μm) in gradient mode with 10 mM ammonium acetate buffer (pH 4.0) as mobile phase A and acetonitrile as mobile phase B. The flow rate was 0.4 ml/min with a column temperature of 50 °C. The gradient program was 0 min at 8% A, 3.5 min at 8% A, 4.5 min at 60% A, 10 min at 60% A, 10.2 min at 8% A, and 12 min at 8% A. The ion source parameters were a capillary voltage of 2.8 kV, source temperature of 120 °C, and desolvation gas temperature of 350 °C at a flow rate of 700 liter/h. The cone voltage was 30 V, and the collision energy was 15 eV. Quantification was performed by integration of the multiple-reaction monitoring trace of TMA (60.1 ≫ 45.0 Da).

**Structure Determination**—Data were indexed with MOS-FLM (24) and scaled with SCALA (25) from the CCP4 suite (26). Choline-bound CutC crystals exhibited an orthorhombic lattice and space group  $P2_12_12_1$  with 8 molecules in the asymmetric unit. CutC crystals without choline exhibited a monoclinic lattice and space group  $P2_1$  with 4 molecules in the asymmetric unit. The choline-bound structure was determined by molecular replacement in *MOLREP* (27), using a CutC homology model built from GD (Protein Data Bank ID 1R9D) (16). The choline-free structure was determined in *MOLREP* using coordinates from the determined choline-bound structure. Models were further built manually with Coot (28) and refined with REFMAC5 (29). Data processing, refinement, and validation statistics are shown in Table 1.

## Results

**Spontaneous and Induced Cleavage of the N-terminal Domain**—During full-length CutC crystallization trials, protein crystals appeared after what turned out to be a spontaneous partial degradation. Mass spectrometry analysis of the crystallization drops (Fig. 1A) proved that the molecular mass of the protein of interest was reduced from the expected 124 kDa to 90–92 kDa, corresponding to a loss of 305–325 N-terminal amino acids. The 124-kDa full-length CutC protein proved to be very unstable, with the N-terminal part starting to degrade almost immediately after purification by nickel affinity chro-

**TABLE 1**

### Data processing and refinement statistics

Values in parentheses are for the highest resolution bin. PDB, Protein Data Bank; r.m.s.d., root mean square deviation.

	CutC	
	Choline-bound	Choline-free
PDB ID	5A0U	5A0Z
Space group	$P2_12_12_1$	$P2_1$
Wavelength (Å)	1.000	0.984
Resolution (Å)	2.4	3.0
Unit cell dimensions (Å)	$a = 89.4, b = 221.9,$ $c = 419.5$	$a = 103.6, b = 154.6,$ $c = 120.8$
Chains in asymmetric unit	8	4
Highest resolution bin (Å)	2.40–2.46	3.00–3.08
$R_{\text{merge}}$	0.173 (0.512)	0.108 (0.378)
Total No. of observations	923,068	117,615
Total No. unique	304,571	66,881
$I/\sigma$	4.7 (1.8)	5.2 (1.9)
Completeness (%)	93.57 (87.81)	88.47 (74.82)
Multiplicity	3.0 (2.7)	1.8 (1.8)
$R$ -factor	0.19 (0.294)	0.23 (0.312)
$R_{\text{free}}$	0.25 (0.341)	0.28 (0.335)
Protein atoms	50,144	23,203
Ligand atoms	56	
Solvent molecules	1976	
Average Wilson $B$ -factor	15.8	26.6
Atomic $B$ -factor for protein	28.4	37.6
Atomic $B$ -factor for ligand	27.0	
Atomic $B$ -factor for solvent	22.4	
r.m.s.d. from ideal		
Bond lengths (Å)	0.012	0.014
Bond angles	1.565°	1.721°
Ramachandran outliers (%)	0.41	0.86

matography. To ensure the reproducibility and homogeneity of the protein for obtaining better crystals, in further experiments, full-length CutC was treated with chymotrypsin and purified by gel filtration on a Superdex 200 column. Chymotrypsin treatment resulted in a protein that had the same length as after spontaneous degradation. The exact biological role of the N-terminal domain, which contains ~305–325 residues, remains unknown because the truncated protein contains the active site, has the same oligomerization state, and exhibits observable (although reduced) TMA-producing activity, as demonstrated below. Analogously, N-terminal truncations of up to 52 amino acids had no effect on the activity and oligomerization state of *D. alaskensis* CutC (12). The truncated C-terminal part of CutC also aligns well with other full-length GRES. Residues 334–1128 in the BLASTP alignment (Fig. 1B) show 93% coverage with full-length CutC from *D. alaskensis* (11, 12), 98% coverage with full-length GD (16), and 96% coverage with full-length BSS (17), which indicates that the N-terminal domain is disposable for the catalytic function of the enzyme.

**Enzymatic Activity**—CutC/CutD from our study was tested for activity both *in vivo* and *in vitro*. The *in vivo* experiments consisted of anaerobic coexpression of CutC and CutD in *E. coli* BL21(DE3) cells in the presence of choline (Fig. 2A). The amount of TMA was higher when the CutC and CutD enzymes were expressed together rather than separately. Enzymatic activity was also tested *in vitro* with purified CutC and CutD proteins (Fig. 2B). Both untreated and chymotrypsin-treated CutC enzymes showed higher TMA-producing activities in the presence of SAM compared with the control without SAM, although the activity for chymotrypsin-treated protein was reduced. Dependence on SAM is a fundamental attribute of GRE catalysis because it is crucial for the generation of glycol

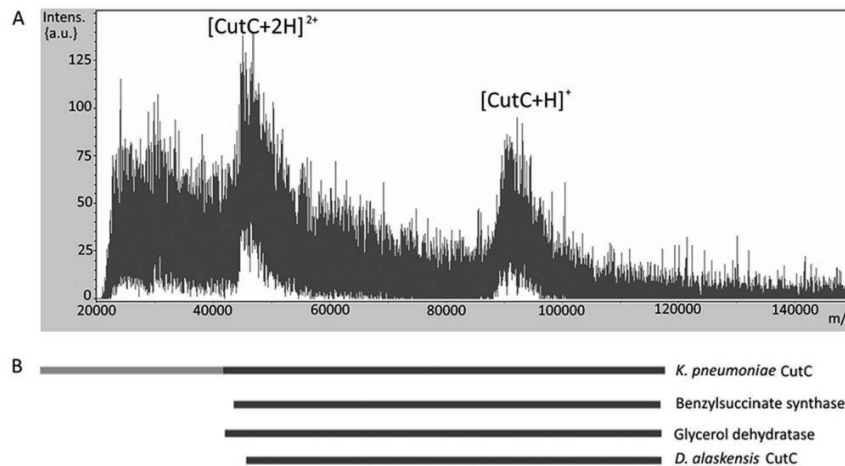


FIGURE 1. A, MALDI-TOF-MS analysis of CutC from a crystallization drop. Approximately 45,000–47,000 and 90,000–92,000  $m/z$  large peaks can be distinguished. These peaks correspond to the double- and single-protonated CutC states, respectively. B, BLASTP alignment of *K. pneumoniae* CutC with other full-length GRES. The CutC N-terminal domain (residues 1–334), not visible in our crystal structures, is shown in light gray. Intens. (a.u.), intensity (absorbance units).

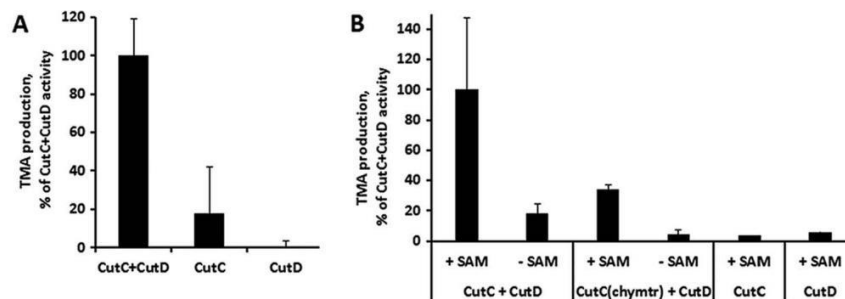


FIGURE 2. A, CutC enzymatic activity *in vivo*. Data are normalized to the empty vector and show elevated TMA production when *cutC* and *cutD* genes are coexpressed compared with separate *cutC* and *cutD* expression. B, CutC enzymatic activity *in vitro*. Data show TMA production in the presence of SAM for both full-length and chymotrypsin-treated (*chymtr*) CutC, but only when both SAM and CutD are present. Results are means  $\pm$  S.D. of two independent experiments.

radical. The data also showed that both CutC and CutD are essential for catalysis, thus confirming that these two enzymes perform catalysis together.

**Quality of the Model, Overall Structure, and Oligomerization State**—Electron density for the protein chain is visible starting from residue 334 in one of the eight monomers in the asymmetric unit, whereas an additional 3–4 residues are invisible in other molecules. As judged from mass spectrometry, for both the choline-bound and choline-free forms, an additional 10–30 N-terminal amino acids are present in the crystal in a disordered state and are not visible in the electron density map. For the choline-free form, there are several additional disordered regions that are clearly visible in the choline-bound form. These flexible regions are somewhat different among different molecules in the asymmetric unit, but there are several common regions around residues 937–946, 980–984, 1006–1017, and 1028–1040, for which no interpretable electron density for any chain could be observed.

The 10- $\beta/\alpha$  barrel structure, which is typical of other GRES, is also characteristic of CutC (Fig. 3). The monomer can be divided into three subdomains: N-terminal half-barrel (residues 334–771), C-terminal half-barrel (residues 771–1079), and glycine loop domain (residues 1079–1128). Gel filtration

analysis (Fig. 4) suggested that CutC in solution both before and after treatment with chymotrypsin exists in oligomers. In the crystal structures of both the choline-bound and choline-free forms, a dimer with 2-fold symmetry can be distinguished (Fig. 3). Dimerization would be expected because similar dimers with 2-fold symmetry have been characterized for the majority of GRES (12, 14–17). For this dimer, AREAIMOL calculated a total contact surface of 14,525.5  $\text{\AA}^2$  (for the choline-bound form), thus confirming that it is likely the biological CutC dimer.

**Active Site**—CutC Cys-771 and Gly-1103 align almost perfectly with the GD catalytic dyad Cys-433 and Gly-763, thus strongly suggesting that these 2 residues form glycy and thiyl radicals that are essential for catalysis (Fig. 5). Theoretical and practical studies have suggested previously that, in CutC, the glycy radical abstracts the hydrogen atom from cysteine to create a thiyl radical. This thiyl radical then further abstracts the hydrogen atom from the choline C1 atom, causing molecular rearrangement and TMA elimination either directly or through intermediate TMA migration to the generated radical site at C1 (Fig. 6) (12, 30). Cys-771 in the active site is positioned appropriately for thiyl radical hydrogen abstraction from the choline C1 atom, thus confirming the previous studies. Additionally, the choline C1 atom aligns with the glycerol carbon atom, from



## Structure and Function of CutC Choline Lyase

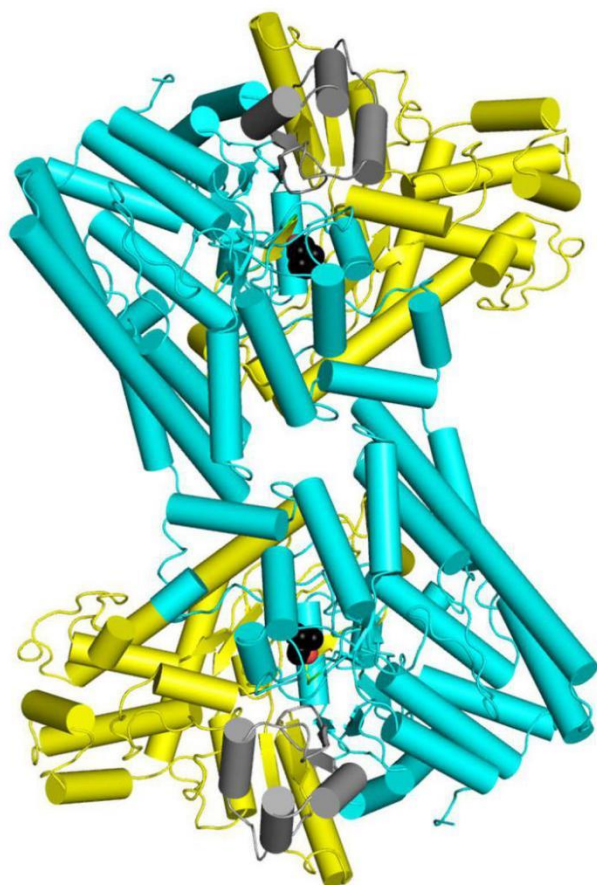


FIGURE 3. Structure of the CutC 2-fold dimer. Cyan, N-terminal half-barrel; yellow, C-terminal half-barrel; gray, glycine loop domain. Choline is displayed as a sphere model, with black carbon atoms and red oxygen atoms. This image was created using PyMOL.

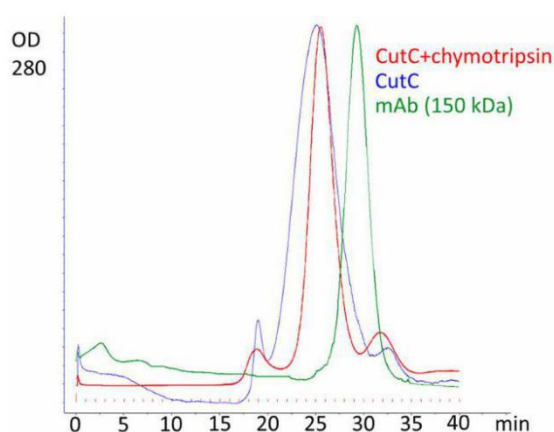


FIGURE 4. Gel filtration of untreated and chymotrypsin-treated CutC on a Superdex 200 column. Human monoclonal IgG antibody, larger than a CutC monomer but smaller than a CutC dimer, was used as a marker.

which the hydrogen abstraction is shown to occur (16). The choline OH group is located at the appropriate distance for hydrogen bond formation with the acid group of Glu-773. The

corresponding Glu-435 of GD aligns nearly identically with Glu-773 of CutC, and the former analogously forms a hydrogen bond with the hydroxyl group of glycerol. The positively charged quaternary amine nitrogen of choline is located at the appropriate distance to interact electrostatically with the side chain of Asp-498 and to form a cation- $\pi$  interaction with the nearby Phe-677.

The active site pocket in the choline-bound form is completely closed and mostly fits the choline surface (Fig. 7A). Therefore, in this form, there is very little space left for designing substrate-like inhibitors with additional chemical groups. However, in the choline-free form, there is a possible tunnel to the outside that is formed by the movement of flexible regions, as discussed below (Fig. 7B). The C1 atom is the closest choline carbon atom to this tunnel, and it is thus the potential site for designing analog inhibitors. Such an inhibitor with additional atoms at C1 could potentially bind to the active site and prevent the CutC enzyme from transitioning to the active state.

**Choline-bound and Choline-free Forms of CutC**—The choline-bound and choline-free CutC forms have noticeable conformational differences (Fig. 8A). In the choline-free form, the CutC conformational change caused the N-terminal half-barrel and glycine loop domain to move away in opposite directions from a gap in which flexible regions, fixed in the choline-bound form, are located. In the choline-free form, the C-terminal half-barrel and glycine loop domain ( $\alpha$  helices 20, 21, 23–26, 28, and 30–32 and  $\beta$  strands 7 and 8) are shifted up to 4 Å, but in the N-terminal half-barrel, they are shifted only up to 1.5 Å (Fig. 8A). A very similar conformational change has been observed for BSS, another GRE (17). Two forms of BSS, BSS $\alpha\beta\gamma$  and BSS $\alpha\gamma$ , have been crystallized, and they differ in a very similar manner to the choline-bound and choline-free forms of CutC, respectively (17). BSS $\alpha\beta\gamma$  consists of the main  $\alpha$  subunit and two small  $\beta$  and  $\gamma$  subunits, whereas BSS $\alpha\gamma$  contains only the main  $\alpha$  subunit and a small  $\gamma$  subunit. Similar to choline-free CutC, BSS $\alpha\gamma$  has a more noticeable conformational shift in the C-terminal half-barrel and glycine loop domain compared with the N-terminal half-barrel (Fig. 8B).

In the choline-free form, the flexible regions expose the active site, including Gly-1103 (Fig. 9A). In the choline-bound form, the active site is completely shielded from the outer environment via ordering of flexible regions (Fig. 9B), which suggests that conformational changes must be associated with increasing availability of the active site. The glycine loop is shifted 2.4 Å toward the top of the protein (Fig. 9C). Similar movement was also observed for BSS and thought to be associated with making glycine accessible to the activating enzyme (17).

Resistance to chymotrypsin treatment was observed for the BSS $\alpha\beta\gamma$  complex, but not for the BSS $\alpha\gamma$  complex. Analogously, CutC resistance to 6 h of chymotrypsin treatment was noticeable in the presence of choline, but in its absence (Fig. 10). The chymotrypsin resistance that was observable at pH 6.5 and 8.5 also excluded the possible influence of pH under different crystallization conditions. The CutC structure in the absence of choline appears to be open in both solution and the crystal, and the binding of choline to the active site appears to change the conformation to a stable, closed, more structured, and chymotrypsin-resistant form.

## Structure and Function of CutC Choline Lyase

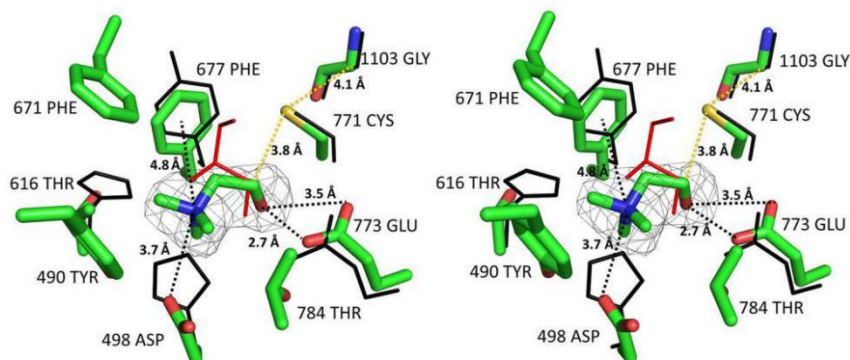


FIGURE 5. Stereoview alignment of CutC and GD (Protein Data Bank ID 1R9D) active sites. CutC and choline are shown as stick models, and GD (black) and glycerol (red) are shown as line models. An  $F_0 - F_c$  map for choline is displayed at the  $3\sigma$  level. Images were created with PyMOL.

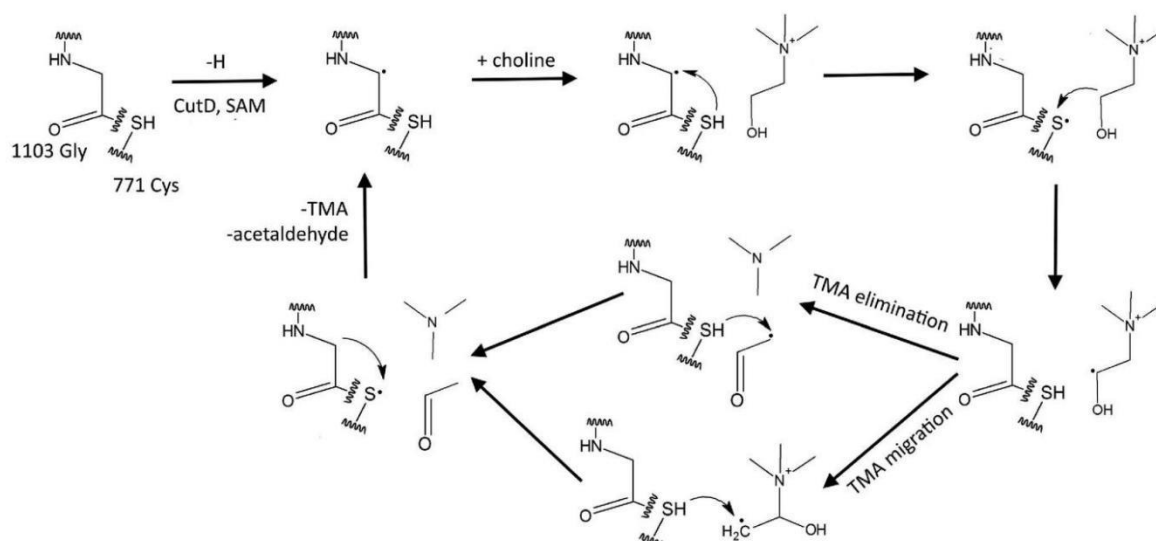


FIGURE 6. Current proposed reaction mechanism of CutC choline lyase. Dots indicate radical formation.

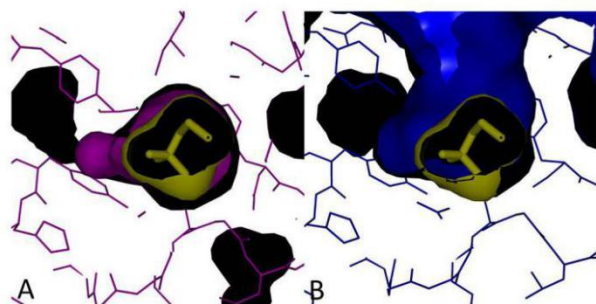


FIGURE 7. Surface models of active site pockets. A, choline-bound form. The choline-bound CutC form is displayed in purple, and choline is displayed in yellow (stick and surface models). B, choline-free form. The choline-free CutC form is displayed in purple, and choline is displayed in yellow (stick and surface models). Choline in the active site was modeled using data from the choline-bound form. Images were created with PyMOL.

### Discussion

Recent studies indicate that the influence of the microbiota on human organism functions is very diverse. The association of the microbiota with diseases such as cancer (31), diabetes

(32), and cardiovascular diseases (33) has been demonstrated. TMAO is a metabolite created in the host liver from microbiota-produced TMA, and it has been associated with several diseases (3–10). TMA production in the intestine can be suppressed by antibiotics (34, 35), but this effect is not complete, and long-term antibiotic treatment may lead to an imbalance in the microbiota (36). An alternative solution would be to develop specific enzymatic inhibitors that block TMA production but do not cause as much of an imbalance in the microbiota as antibiotics do. More information about the structures of TMA-producing enzymes from bacteria in the human microbiota therefore could be significant for finding ways to block TMA/TMAO formation, thus providing new microbiota-linked drug targets to treat cardiovascular diseases and trimethylaminuria. CutC choline lyase is very appropriate as a target of such studies, as it produces TMA from a common nutrient, choline, and can be found in the genomes of several microbiota representatives (11, 12, 19). Although there is too little free space in the choline-bound form of CutC to design substrate analog inhibitors, the choline-free form may contain a possible



## Structure and Function of CutC Choline Lyase

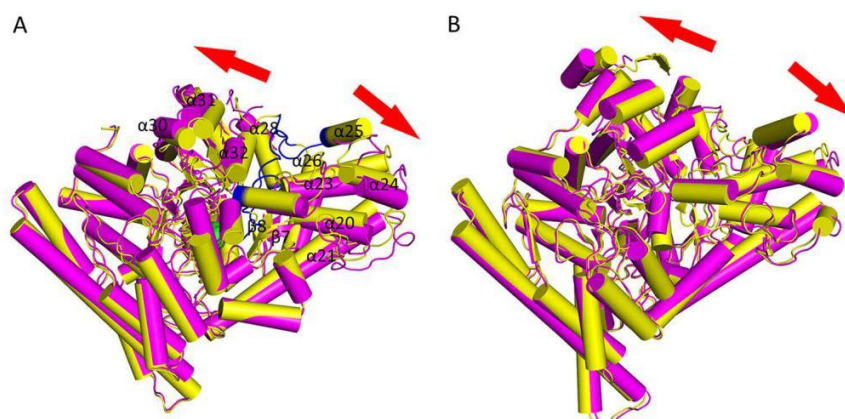


FIGURE 8. *A*, choline-bound (yellow) and native (purple) CutC forms. Regions present in the choline-bound form but absent in the native form electron density map are colored in blue. Choline is displayed as a sphere model, with green carbon atoms and red oxygen atoms. *B*, BSS  $\alpha$  subunit, BSS  $\alpha\beta\gamma$  (yellow), and BSS  $\alpha\gamma$  (purple) (13). Images were created with PyMOL.

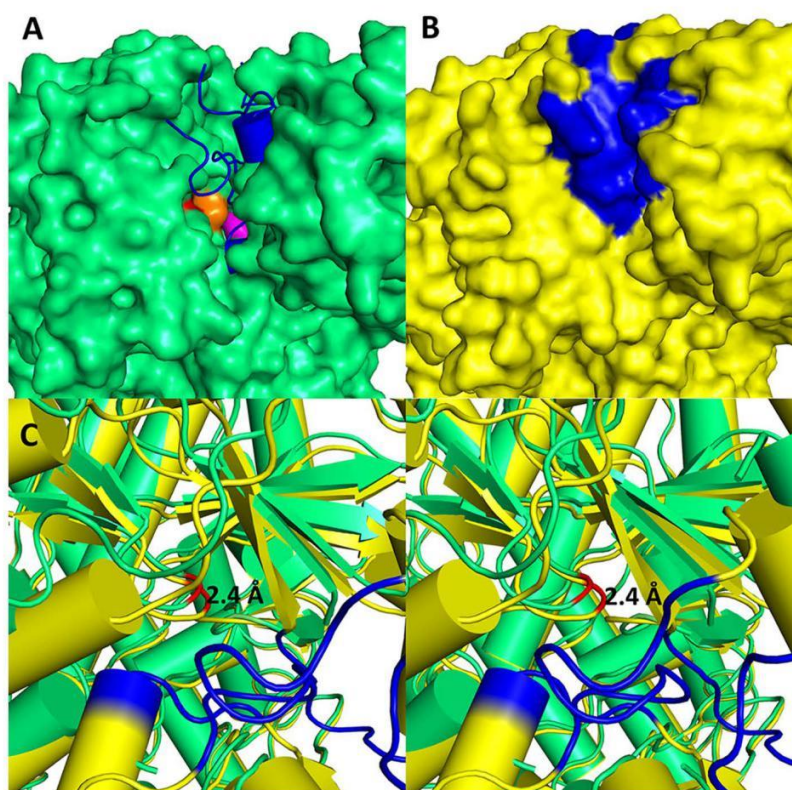


FIGURE 9. *A*, surface model of the choline-free CutC form (green). The catalytic dyad Gly-1103 is shown as a red surface, and Cys-771 as an orange surface. Choline in the active site (purple surface) and disordered regions (blue) were modeled using data from the choline-bound form. *B*, surface model of the choline-bound CutC form (yellow). Regions disordered in the choline-free form are displayed in blue. *C*, stereoview alignment of glycine loops for the choline-bound (yellow) and choline-free (green) CutC forms. Gly-1103 in both forms is displayed in red. Regions disordered in the choline-free form are displayed in blue. Images were created with PyMOL.

tunnel made by movement of disordered regions that could be used for this purpose.

The nature of the catalysis of CutC and other GREs is strictly anaerobic because oxygen reacts with glycy radicals, inactivating the enzymes. However, the glycy radical in GREs has proven to be very stable, lasting up to several days *in vitro* (37).

Several protective strategies have been developed to prevent oxidative enzyme inactivation. One such strategy is using enzymes that act similarly to “spare parts”, reattaching the oxygen-cleaved C-terminal domains (38). The determined structures of certain GREs indicate another protective mechanism: shielding glycine from the outer environment by burying it in



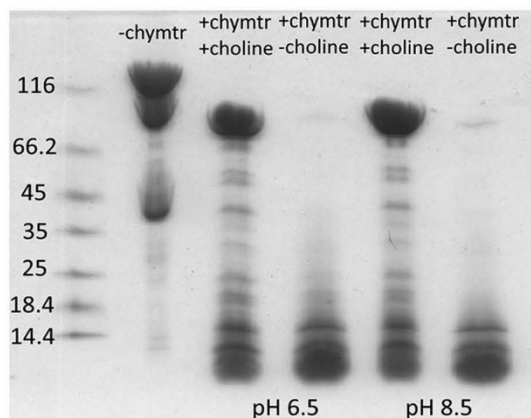


FIGURE 10. Chymotrypsin treatment of CutC for 6 h at room temperature and SDS-PAGE analysis. Marker sizes are shown on the left in kilodaltons. Partial degradation was noticeable for CutC even without the addition of chymotrypsin (*chymtr*).

the protein core. However, it is unclear how the activating enzyme generates the glycy radical and how the ligand gains access to the ligand-binding site. Co-crystallization studies with the activating enzyme of PFL and a PFL glycine loop-mimicking peptide have suggested that the active site glycine interacts with the activating enzyme directly (39). This means that a local movement of the glycine loop out of the protein core, a global structural change, or a combination of both must happen to achieve radical formation. PFL has been shown to exist in solution in both open and closed forms, and the equilibrium between these forms is thought to be modulated by the activating enzyme (40). Direct structural evidence for conformational changes, similar to the changes observed for *K. pneumoniae* CutC, has also been observed for BSS (17). For BSS, this conformational shift is regulated by an accessory  $\beta$  subunit. However, no accessory subunit or activating enzyme was present during the crystallization and chymotrypsin treatment of *K. pneumoniae* CutC, showing that, at least for CutC, no help from other proteins is required for conformational changes to occur. Chymotrypsin tolerance in the presence of choline also demonstrated that conformational changes occur in solution as well as in crystalline form. These results suggest that, in the open form, the active site is freely accessible to choline and that binding of choline is the only critical factor that transforms the structure from the open to closed form.

Interestingly, CutC is the only GRE characterized so far for which ligand binding to the ligand-binding site causes such an effect on the conformation in the crystal structure. Two other GREs, PFL and GD, have been crystallized in both ligand-bound and ligand-free states (14, 16), but no significant conformational changes in crystal structure between these two states have ever been observed. Additionally, both the ligand-bound and ligand-free GD states align better with the closed choline-bound CutC state. This raises the question of whether there is a correlation between the observed choline-induced CutC conformational changes and the accessibility of the active site glycine to the activating enzyme. It would be an inefficient strategy for a GRE to enter a stable and closed conformation indepen-

dently of the presence or absence of a glycy radical in the active site. An alternative explanation would be that the activating enzyme is able to induce CutC conformational changes partially or completely independently of the presence or absence of choline. Choline-driven conformational changes in this case could serve only as an additional regulatory mechanism, perhaps increasing affinity for the substrate.

Other results regarding the CutC active site, the mechanism of catalysis, and the oligomerization state are consistent with data from studies on other GREs. CutC has a typical  $10\text{-}\beta/\alpha$  barrel fold, and dimerization is observed. A catalytic dyad is formed by glycine and cysteine, which are located appropriately for hydrogen abstraction from the choline C1 atom, and the choline in the ligand-binding site is coordinated by glutamic acid, aspartic acid, and phenylalanine residues that were previously mapped with site-directed mutagenesis in *D. alaskensis* CutC and shown to be essential for catalysis (12, 30).

**Author Contributions**—G. K. designed and constructed vectors for protein expression, expressed and purified proteins, characterized enzymatic activity, crystallized protein, determined protein x-ray structure, and wrote the paper. J. K., S. G., M. M.-K., E. L., and M. D. characterized enzymatic activity. K. T. conceived and coordinated the study, determined protein x-ray structure, and wrote the paper. All authors analyzed the results and approved the final version of the manuscript.

**Acknowledgments**—We thank the MAX-lab staff for support.

#### References

- Hayward, H. R., and Stadtman, T. C. (1960) Anaerobic degradation of choline. II. Preparation and properties of cell-free extracts of *Vibrio cholerae*. *J. Biol. Chem.* **235**, 538–543
- Sandhu, S. S., and Chase, T., Jr. (1986) Aerobic degradation of choline by *Proteus mirabilis*: enzymatic requirements and pathway. *Can. J. Microbiol.* **32**, 743–750
- Koeth, R. A., Wang, Z., Levison, B. S., Buffa, J. A., Org, E., Sheehy, B. T., Britt, E. B., Fu, X., Wu, Y., Li, L., Smith, J. D., DiDonato, J. A., Chen, J., Li, H., Wu, G. D., Lewis, J. D., Warrier, M., Brown, J. M., Krauss, R. M., Tang, W. H., Bushman, F. D., Lusk, A. J., and Hazen, S. L. (2013) Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat. Med.* **19**, 576–585
- Treacy, E. P., Akerman, B. R., Chow, L. M., Youil, R., Bibeau, C., Lin, J., Bruce, A. G., Knight, M., Danks, D. M., Cashman, J. R., and Forrest, S. M. (1998) Mutations of the flavin-containing monooxygenase gene (*FMO3*) cause trimethylaminuria, a defect in detoxication. *Hum. Mol. Genet.* **7**, 839–845
- Brown, J. M., and Hazen, S. L. (2014) Metaorganismal nutrient metabolism as a basis of cardiovascular disease. *Curr. Opin. Lipidol.* **25**, 48–53
- Wang, Z., Klipfell, E., Bennett, B. J., Koeth, R., Levison, B. S., Dugar, B., Feldstein, A. E., Britt, E. B., Fu, X., Chung, Y. M., Wu, Y., Schauer, P., Smith, J. D., Allayee, H., Tang, W. H., DiDonato, J. A., Lusk, A. J., and Hazen, S. L. (2011) Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63
- Tang, W. H., Wang, Z., Levison, B. S., Koeth, R. A., Britt, E. B., Fu, X., Wu, Y., and Hazen, S. L. (2013) Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N. Engl. J. Med.* **368**, 1575–1584
- Trosetid, M., Ueland, T., Hov, J. R., Svardsdal, A., Gregersen, I., Dahl, C. P., Aakhus, S., Gude, E., Bjørndal, B., Halvorsen, B., Karlsen, T. H., Aukrust, P., Gullestad, L., Berge, R. K., and Yndestad, A. (2015) Microbiota-dependent metabolite trimethylamine-N-oxide is associated with disease severity and survival of patients with chronic heart failure. *J. Intern. Med.* **277**, 717–726



## Structure and Function of CutC Choline Lyase

- Tang, W. H., Wang, Z., Kennedy, D. J., Wu, Y., Buffa, J. A., Agatista-Boyle, B., Li, X. S., Levison, B. S., and Hazen, S. L. (2015) Gut microbiota-dependent trimethylamine *N*-oxide (TMAO) pathway contributes to both development of renal insufficiency and mortality risk in chronic kidney disease. *Circ. Res.* **116**, 448–455
- Gao, X., Liu, X., Xu, J., Xue, C., Xue, Y., and Wang, Y. (2014) Dietary trimethylamine *N*-oxide exacerbates impaired glucose tolerance in mice fed a high fat diet. *J. Biosci. Bioeng.* **118**, 476–481
- Craciun, S., and Balskus, E. P. (2012) Microbial conversion of choline to trimethylamine requires a glycol radical enzyme. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 21307–21312
- Craciun, S., Marks, J. A., and Balskus, E. P. (2014) Characterization of choline trimethylamine-lyase expands the chemistry of glycol radical enzymes. *ACS Chem. Biol.* **9**, 1408–1413
- Shisler, K. A., and Broderick, J. B. (2014) Glycol radical activating enzymes: structure, mechanism, and substrate interactions. *Arch. Biochem. Biophys.* **546**, 64–71
- Becker, A., Fritz-Wolf, K., Kabsch, W., Knappe, J., Schultz, S., and Volker Wagner, A. F. (1999) Structure and mechanism of the glycol radical enzyme pyruvate formate-lyase. *Nat. Struct. Biol.* **6**, 969–975
- Logan, D. T., Andersson, J., Sjöberg, B. M., and Nordlund, P. (1999) A glycol radical site in the crystal structure of a class III ribonucleotide reductase. *Science* **283**, 1499–1504
- O'Brien, J. R., Raynaud, C., Croux, C., Girbal, L., Soucaille, P., and Lanzilotta, W. N. (2004) Insight into the mechanism of the B<sub>12</sub>-independent glycerol dehydratase from *Clostridium butyricum*: preliminary biochemical and structural characterization. *Biochemistry* **43**, 4635–4645
- Funk, M. A., Judd, E. T., Marsh, E. N., Elliott, S. J., and Drennan, C. L. (2014) Structures of benzylsuccinate synthase elucidate roles of accessory subunits in glycol radical enzyme activation and activity. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 10161–10166
- Wagner, A. F., Frey, M., Neugebauer, F. A., Schäfer, W., and Knappe, J. (1992) The free radical in pyruvate formate-lyase is located on glycine-734. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 996–1000
- Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214
- Romano, K. A., Vivas, E. I., Amador-Noguez, D., and Rey, F. E. (2015) Intestinal microbiota composition modulates choline bioavailability from diet and accumulation of the proatherogenic metabolite trimethylamine-*N*-oxide. *mBio* **6**, e02481-4
- Kuka, J., Liepinsh, E., Makreka-Kuka, M., Liepins, J., Cirule, H., Gustina, D., Loza, E., Zharkova-Malkova, O., Grinberga, S., Pugovics, O., and Dambrova, M. (2014) Suppression of intestinal microbiota-dependent production of pro-atherogenic trimethylamine *N*-oxide by shifting *L*-carnitine microbial degradation. *Life Sci.* **117**, 84–92
- Dambrova, M., Liepinsh, E., and Kalvinsh, I. (2002) Mildronate: cardioprotective action through carnitine lowering effect. *Trends Cardiovasc. Med.* **12**, 275–279
- Dambrova, M., Skapare-Makarova, E., Konrade, I., Pugovics, O., Grinberga, S., Tirzite, D., Petrovska, R., Kalvins, I., and Liepins, E. (2013) Meldonium decreases the diet-increased plasma levels of trimethylamine *N*-oxide, a metabolite associated with atherosclerosis. *J. Clin. Pharmacol.* **53**, 1095–1098
- Leslie, A. G. W. (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, No. 26
- Evans, P. R. (1997) SCALA. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, No. 33
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242
- Vagin, A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022–1025
- Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
- Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255
- Thibodeaux, C. J., and van der Donk, W. A. (2012) Converging on a mechanism for choline degradation. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 21184–21185
- Zitvogel, L., Galluzzi, L., Viaud, S., Vétizou, M., Daillère, R., Merad, M., and Kroemer, G. (2015) Cancer and the gut microbiota: an unexpected link. *Sci. Transl. Med.* **7**, 271ps1
- He, C., Shan, Y., and Song, W. (2015) Targeting gut microbiota as a possible therapy for diabetes. *Nutr. Res.* **35**, 361–367
- Howitt, M. R., and Garrett, W. S. (2012) A complex microworld in the gut: gut microbiota and cardiovascular disease connectivity. *Nat. Med.* **18**, 1188–1189
- Fraser-Andrews, E. A., Manning, N. J., Ashton, G. H. S., Eldridge, P., McGrath, J., and Menagé Hdu, P. (2003) Fish odour syndrome with features of both primary and secondary trimethylaminuria. *Clin. Exp. Dermatol.* **28**, 203–205
- Treacy, E., Johnson, D., Pitt, J. J., and Danks, D. M. (1995) Trimethylaminuria, fish odour syndrome: a new method of detection and response to treatment with metronidazole. *J. Inher. Metab. Dis.* **18**, 306–312
- Jernberg, C., Löfmark, S., Edlund, C., and Jansson, J. K. (2010) Long-term impacts of antibiotic exposure on the human intestinal microbiota. *Microbiology* **156**, 3216–3223
- Walsby, C. J., Ortillo, D., Yang, J., Nnyepi, M. R., Broderick, W. E., Hoffman, B. M., and Broderick, J. B. (2005) Spectroscopic approaches to elucidating novel iron-sulfur chemistry in the “radical-Sam” protein superfamily. *Inorg. Chem.* **44**, 727–741
- Wagner, A. F., Schultz, S., Bomke, J., Pils, T., Lehmann, W. D., and Knappe, J. (2001) YfiD of *Escherichia coli* and Y06I of bacteriophage T4 as autonomous glycol radical cofactors reconstituting the catalytic center of oxygen-fragmented pyruvate formate-lyase. *Biochem. Biophys. Res. Commun.* **285**, 456–462
- Vey, J. L., Yang, J., Li, M., Broderick, W. E., Broderick, J. B., and Drennan, C. L. (2008) Structural basis for glycol radical formation by pyruvate formate lyase activating enzyme. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 16137–16141
- Peng, Y., Veneziano, S. E., Gillispie, G. D., and Broderick, J. B. (2010) Pyruvate formate-lyase, evidence for an open conformation favored in the presence of its activating enzyme. *J. Biol. Chem.* **285**, 27224–27231

### **3.2 Encapsulation mechanisms and structural studies of GRM2 BMC shell particles**

- **Isolation, recombinant expression and purification of GRM2 core enzymes and shell proteins**
- **Testing the influence of particular shell gene sets on the formation of bacterial microcompartment shell derived particles (BDPs) and their size distribution**
- **The encapsulation mechanisms of GRM2 core is hierarchical and CutC choline lyase most likely plays a secondary role as an adaptor protein**
- **Calculated cryo-EM structure of smaller, pT=4 BDP particles at 3.3 Å resolution**
- **Various minor BDP forms with additional hexameric rings, larger triangulation numbers or fusion of smaller icosahedrons could be identified in the cryo-EM analysis**

ARTICLE

<https://doi.org/10.1038/s41467-019-14205-y>

OPEN

# Encapsulation mechanisms and structural studies of GRM2 bacterial microcompartment particles

Gints Kalnins <sup>1\*</sup>, Eva-Emilija Cesle<sup>1</sup>, Juris Jansons<sup>1</sup>, Janis Liepins <sup>2</sup>, Anatolij Filimonenko<sup>3</sup> & Kaspars Tars<sup>1,4</sup>

Bacterial microcompartments (BMCs) are prokaryotic organelles consisting of a protein shell and an encapsulated enzymatic core. BMCs are involved in several biochemical processes, such as choline, glycerol and ethanolamine degradation and carbon fixation. Since non-native enzymes can also be encapsulated in BMCs, an improved understanding of BMC shell assembly and encapsulation processes could be useful for synthetic biology applications. Here we report the isolation and recombinant expression of BMC structural genes from the *Klebsiella pneumoniae* GRM2 locus, the investigation of mechanisms behind encapsulation of the core enzymes, and the characterization of shell particles by cryo-EM. We conclude that the enzymatic core is encapsulated in a hierarchical manner and that the CutC choline lyase may play a secondary role as an adaptor protein. We also present a cryo-EM structure of a  $pT=4$  quasi-symmetric icosahedral shell particle at 3.3 Å resolution, and demonstrate variability among the minor shell forms.

<sup>1</sup>Latvian Biomedical Research and Study Centre, Ratsupites 1, Riga 1067, Latvia. <sup>2</sup>Institute of Microbiology and Biotechnology, University of Latvia, Jelgavas 1, Riga 1004, Latvia. <sup>3</sup>Central European Institute of Technology, Masaryk University, Kamenice 753/5, 62500 Brno, Czech Republic. <sup>4</sup>University of Latvia, Jelgavas 1, Riga 1004, Latvia. \*email: [gints@biomed.lu.lv](mailto:gints@biomed.lu.lv)



**B**acterial microcompartments (BMCs) are organelle-like structures consisting of a large quasi-icosahedral or polyhedral protein shell up to 200–300 nm in size and an encapsulated enzymatic core<sup>1–6</sup>. It is estimated that approximately 25% of bacterial taxa have genomes that contain a BMC locus of some sort that performs various functions<sup>7–9</sup>. The most well-studied and well-known types of BMCs are the carboxysomes because they are widespread among cyanobacteria and play an important role in carbon fixation<sup>5,10,11</sup>. Another, more diverse group of BMCs are the metabolosomes. Metabolosomes are specialized BMCs that breakdown various compounds, such as propanediol<sup>12,13</sup>, ethanolamine<sup>14</sup>, choline<sup>15</sup>, and rhamnose/fucose<sup>16,17</sup>. In metabolosomes, the catabolic breakdown of the substrate mostly involves an aldehyde intermediate, which may be toxic to the cells and/or volatile and hence is sequestered within the BMC protein shell<sup>18,19</sup>. Encapsulation of the enzymatic pathway also has the benefit of increasing the local substrate concentrations and increasing the overall efficiency of the pathway. The metabolosomes generally contain at least four or five different encapsulated core enzymes, including signature enzymes that perform the initial breakdown of the substrate, alcohol dehydrogenase, aldehyde dehydrogenase, and phosphotransacylase<sup>5,20</sup>.

The BMC shell consists of three types of BMC shell proteins: the BMC-H, BMC-T, and BMC-P proteins<sup>5,6</sup>. BMC-H is a hexameric BMC protein consisting of one Pfam00936 domain, and it is capable of forming uniform flat sheets<sup>2,21–24</sup>. The Pfam00936 domain consists of an  $\alpha$ - $\beta$  double sandwich that contains a four-stranded antiparallel beta sheet and is flanked by two  $\alpha$ -helices on one side. The BMC-T protein is somewhat similar to BMC-H according to its sequence, but instead of one Pfam00936 domain, as in BMC-H, it contains two fused domains and as a consequence is trimeric instead of hexameric<sup>25–28</sup>. BMC-H and BMC-T hexamers and trimers have large pores that can be up to 14 Å in size in their centers that are thought to ensure metabolite flow across the BMC shell<sup>19,29,30</sup>. The BMC-P (or BMV)<sup>8,31,32</sup> monomer consists of a single Pfam03319 domain, which is structurally unrelated to the Pfam00936 domain and contains a 5-stranded  $\beta$ -barrel. BMC-P proteins are pentameric, and their function is thought to be limited to capping the vertices of icosahedrons<sup>28,32</sup>. In the formation of BMCs, a key role is played by encapsulation peptide (EP) sequences<sup>33–36</sup>. EPs are small, 10–20 residue-long amphipathic  $\alpha$ -helices that are attached to the core enzymes N- or C-terminally or inside the flexible surface loops. Genetic fusions of such sequences have been demonstrated to be sufficient for the encapsulation of non-native proteins in propanediol utilization (Pdu)<sup>35,37</sup>, ethanolamine utilization (Eut)<sup>38</sup>, and beta-carboxysome<sup>39</sup> BMC shell particles. BMC-T and BMC-H proteins, which are encoded in the Pdu locus<sup>40,41</sup>, and a BMC-H protein, which is encoded in the Eut locus<sup>38</sup>, have been demonstrated to serve as EP targets, so it seems that the specific shell partner may vary among different types of BMCs. In addition to binding the core to the shell, EPs are thought to have a crosslinking influence in the enzymatic core as well<sup>5,33,42–44</sup>.

The glycyl-radical associated microcompartment group 2 (GRM2) type BMC locus encodes the glycyl-radical enzyme-associated microcompartment (GRM) subgroup that includes the choline utilizing locus C (CutC) choline lyase as its signature enzyme<sup>8,9,20</sup>. CutC cleaves its initial substrate, choline, into trimethylamine (TMA) and acetaldehyde<sup>45,46</sup>. TMA itself has been under intense scrutiny as a bacterial metabolite with potential therapeutic importance since its oxidized form, trimethylamine-N-oxide (TMAO), has been identified as a likely contributor to cardiovascular diseases<sup>47,48</sup>. Like most BMC loci, the GRM2 locus contains regulatory, shell, core enzyme, and transporter genes. There are two unusual traits of the GRM2 locus. First, there is a

unique, ~340 residue-long N-terminal extension of CutC, which is somewhat homologous to the subsequent 340 residues of CutC and probably originated as an N-terminal duplication<sup>20,46</sup>. The exact function of this extension is unknown, but its involvement in core multimerization or encapsulation processes has been proposed earlier<sup>20</sup>. Another unusual trait of GRM2 is the lack of BMC-T genes in the locus, which is a trait shared with only a few other BMC loci<sup>8,9</sup>. The GRM2 locus therefore contains five structural shell genes encoding four BMC-H proteins and one BMC-P protein.

The versatility of BMCs and their capacity to encapsulate large cargos of various sizes have made them appealing targets for synthetic biology applications. These organelles are generated with the specific goal of packing entire enzymatic pathways to increase their efficiency and lower the effects of intermediate toxicity, making them attractive platforms for the construction of recombinant metabolic pathways. There have been some successful examples, including a two-component ethanol-producing<sup>35,49</sup> BMC-based recombinant particle and a polyphosphate-accumulating<sup>50</sup> compartments. A major obstacle for more extensive research in this area is the lack of a robust encapsulation system. An ideal platform must be universal enough to successfully encapsulate various enzymes of different sizes, solubility, and oligomerization states. Native EPs have been successfully used for this purpose<sup>34,35,37–39,48–50</sup>, however, since they are amphipathic, insolubility or excessive aggregation is a potential outcome. These limits have been circumvented in some recent cases by creating artificial encapsulation mechanisms, either by introducing non-native binding partners in target and shell proteins<sup>51,52</sup> or postponing assembly<sup>53,54</sup>. There are several BMC types that have not yet been studied in detail, and those could reveal themselves as universal and robust biotechnological platforms.

In this study, we demonstrate a practical production system for recombinant GRM2 shell particles from *Klebsiella pneumoniae* and the requirements for recombinant shell formation. We present a 3.3 Å resolution cryo-EM structure of pT = 4 BMC particle, demonstrate the presence of variable minor shell types, and identify the potential roles of the particular core enzymes in the core encapsulation process.

## Results

**Formation of shell particles and effects of BMC-H variants.** In numerous reported cases, the recombinant expression of structural BMC shell genes has successfully yielded stable empty shell particles without enzymatic cores<sup>28,38,39,53</sup>. We extensively investigated what kind of minimal gene set is essential for *Klebsiella pneumoniae* GRM2 shell particle formation (Fig. 1, Table 1, Supplementary Figs. 1–3). We were able to obtain particles which we designated BMC shell-derived particles (BDPs) due to the fact that these are only partially representative of native BMCs—they lack full enzymatic core, are produced in a non-native expression system, and are smaller and more regular than native BMCs. In all cases we purposefully purified BDPs from equivalent amounts of biomass for the results to be comparable. We observed that the minimal requirement for BDP formation is the cmcC + D protein pair, which forms predominantly small type BDPs eluting between 90 and 105 ml on Superose 6 column (Table 1, Supplementary Fig. 3). Curiously, neither for cmcA + D nor cmcB + D we were able to observe formation of BDPs, despite the high similarity between cmcA, cmcB, and cmcC (Fig. 1a). We reasoned that cmcABC should probably be co-expressed from one promoter since these genes in the *Klebsiella pneumoniae* genome are separated by only 8–10 bp-long sequences. Such a construct (cmcABC + D) also resulted predominantly in small type BDPs,



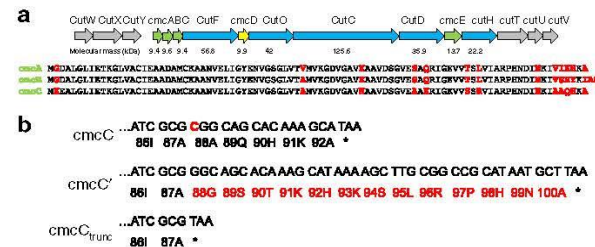
although the yield was much lower than that of the cmcC + D variant. Remarkably, cmcAB + D construct was able to form low amounts of small type particles despite cmcA + D and cmcB + D unable to do so (Supplementary Fig. 3). It is possible that some kind of a synergistic effect between cmcA and cmcB is responsible for this ability to form BDPs.

Due to a defective oligonucleotide used in the PCR, we accidentally created a mutant cmcC variant, designated cmcC', containing a frameshift mutation that affected the last five cmcC residues and created an additional elongation containing eight residues (Fig. 1b). We observed that the yield of the cmcABC' + D BDPs was greatly increased when compared to that of cmcABC + D (Fig. 2a, Supplementary Fig. 1b). Based on the sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) analysis of gel filtration fractions (Fig. 2a), we also concluded that cmcABC' + D particles are not uniform in size, as some eluted immediately after the void volume of 60 ml (designated large type particles), while some material formed a discrete peak at 90–100 ml (designated small type particles), and the rest were spread out in the intermediate zone between these two peaks (designated intermediate particles). Negative staining transmission electron microscopy (TEM) analysis of the particles

in these zones (Fig. 2b, Supplementary Figs. 4–7) confirmed that purified cmcABC' + D BDPs are indeed different in size and are partially sorted during gel filtration according to their sizes; the 90–100 ml peak contained predominantly 20–30 nm particles, and the BDP size increased in the direction of the large type particle peak, which contained particles up to 200 nm in size. This confirmed that the wide distribution of the shell proteins in the gel filtration is a result of the presence of different sizes of particles and not merely because of the aggregation of small particles. Surprisingly, truncation of cmcC in cmcABC<sub>trunc</sub> + D also resulted in the formation of large type BDPs (Table 1, Supplementary Fig. 2a), while this truncation did not have any observable effect on the formation of small type particles in cmcC<sub>trunc</sub> + D BDPs (Supplementary Fig. 3b).

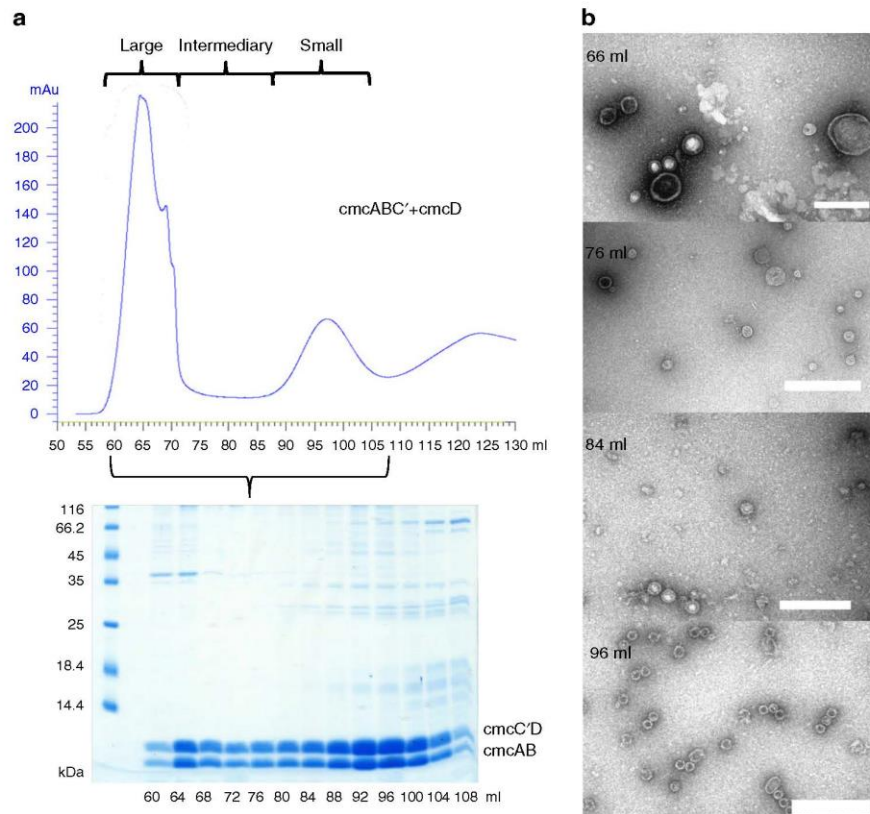
In our subsequent experiments, we co-expressed the full structural gene sets cmcABC + D + E and cmcABC' + D + E. cmcE contained an around 30 amino acid C-terminal elongation if compared to cmcA, cmcB, or cmcC (Supplementary Fig. 8), and we observed its co-migration with the large type BDPs (Supplementary Fig. 1d, e). While cmcE had no influence on the size distribution or yield of the cmcABC' + D BDPs, for the native gene set cmcABC + D cmcE inclusion caused the formation of the large type particles in a pattern similar to that of cmcABC' + D (Supplementary Fig. 1d). We also tested both mutant and C-terminally truncated cmcC' + D and cmcC<sub>trunc</sub> + D BDP variants (Fig. 1b, Supplementary Figs. 3a, b) and observed that there were no differences in the yield or size distribution of BDPs between all three cmcC variants (Supplementary Fig. 3). An obvious common trait of cmcE and mutant cmcC' is the presence of a C-terminal elongation that consists of 8 residues for cmcC' and 40 residues for cmcE.

When the cmcABC and cmcABC' were expressed in the absence of BMC-P cmcD, we could purify almost none of the BMC proteins in the case of cmcABC; however, for cmcABC', it was possible to observe an even “smear” of BMC proteins at the gel filtration volumes of 60–100 ml (Supplementary Fig. 1c). A very similar “smear” was also observed for cmcE particles during cmcE + D co-expression (Supplementary Fig. 3f). In this latter case, cmcD was not observed in the gel filtration profile, suggesting a lack of interaction between these two proteins. When analyzing such “smeared” material of the cmcABC gel filtration 40–42 ml and cmcE + D 94–96 ml fractions with TEM (Supplementary Fig. 9), both rounded and elongated nanotubule-like particles could be observed. Curiously, although both were C-terminally elongated, cmcC' and cmcE had different abilities to



**Fig. 1** *Klebsiella pneumoniae* GRM2 locus and variants of cmcC. **a** *Klebsiella pneumoniae* GRM2 locus. Structural shell BMC-H proteins cmcA, cmcB, cmcC, and cmcE are colored in green, and BMC-P protein cmcD is colored in yellow. Core enzymes CutF (aldehyde dehydrogenase), CutO (alcohol dehydrogenase), CutC (choline lyase), CutD (glycyl-radical activating enzyme), and CutH (phosphotransacylase) are colored in blue. Regulatory and transporter genes are colored in gray. The genes have been named according to previous research<sup>15</sup>. **b** C-terminal amino acid sequences of three cmcC variants—cmcC (native), cmcC' (mutated), and cmcC<sub>trunc</sub> (truncated).

pET-Duet-1 T7-1	pET-Duet-1 T7-2	pRSF-Duet T7-1	Results
cmcABC	–	–	No purified particles
cmcABC	cmcD	–	Predominantly small type particles
cmcABC	cmcD	cmcE	Large, intermediary, and small type particles
cmcABC'	cmcD	cmcC'	Large, intermediary, and small type particles
cmcABC'	cmcD	cmcAB	Large, intermediary, and small type particles
cmcABC'	–	–	Irregular “smeared” material
cmcABC'	cmcD	–	Large, intermediary, and small type particles
cmcABC'	cmcD	cmcE	Large, intermediary, and small type particles
cmcABC <sub>trunc</sub>	cmcD	–	Large, intermediary, and small type particles
cmcAB	cmcD	–	Predominantly small type particles
cmcA	cmcD	–	No purified particles
cmcB	cmcD	–	No purified particles
cmcC	cmcD	–	Predominantly small type particles
cmcC	cmcD	cmcE	Predominantly small type particles
cmcE	cmcD	–	Irregular “smeared” material
cmcC'	cmcD	–	Predominantly small type particles
cmcC <sub>trunc</sub>	cmcD	–	Predominantly small type particles



**Fig. 2** Characterization of *cmcABC'* + *cmcD* BDPs. **a** Gel filtration of sedimented BDPs and SDS-PAGE analysis of the fractions. Two noticeable BDP peaks were formed: one immediately after the empty volume of 60 ml (large type particle zone) and one at approximately 90–100 ml (small type particle zone). An intermediary zone with smaller BDP protein amounts was formed between these two zones. **b** Examples of TEM analysis of the BDP samples in the large particle zone (66 ml), intermediary zone (76 and 84 ml), and small particle zone (96 ml). Scale bar: 200 nm.

interact with the pentameric units; the former is able to interact with them, while the latter is unable to interact with them.

Mass spectrometry analysis identified peaks matching all four shell proteins in the *cmcABC* + D, *cmcABC'* + D and *cmcABC<sub>trunc</sub>* + D BDPs (Supplementary Fig. 10a–e). *cmcC'* was identified as the major BMC-H protein in *cmcABC'* + D BDPs. Curiously, in the case of *cmcABC<sub>trunc</sub>* + D BDPs the major protein was *cmcB*. It was virtually impossible to distinguish between *cmcB* and *cmcC* in *cmcABC* + D small type BDPs, so both could be the major shell components. While *cmcD* could be detected in small type particles, it was practically undetectable in large type particles (Supplementary Fig. 10a, b and 10d, e). This would be expected, since increased triangulation numbers in larger particles would significantly reduce the proportion of fivefold vertices occupied by *cmcD*. In *cmcABC'* + D particles the proportion of BMC-H proteins among large and small type BDPs were largely similar; however, the proportion of *cmcC<sub>trunc</sub>* was significantly increased in large type *cmcABC<sub>trunc</sub>* + D BDPs (Supplementary Fig. 10d, e). Unlike *cmcC'*, *cmcC<sub>trunc</sub>* is not the major protein in *cmcABC<sub>trunc</sub>* + D BDPs but instead is present in minor amounts in large type particles. Also, we identified several unexpected *m/z* peaks in several cases (Supplementary Fig. 10a, b, d, e, i). The identities of these peaks could be degradation products of BMC proteins. Since we did not observe any such peaks in *cmcC'* + D, *cmcC<sub>trunc</sub>* + D, or *cmcC* + D small type BDPs (Supplementary Fig. 10f–h), and they appear in *cmcAB* + D particles (Supplementary Fig. 10i), source of these peaks are *cmcA*

and/or *cmcB* proteins. It is unclear whether these products have any significant impact on assembly process.

We compared expression levels of BMC-H components of *cmcAD*, *cmcBD*, *cmcCD*, *cmcC'D*, and *cmcC<sub>trunc</sub>D* constructs and observed in SDS-PAGE gel that the expression levels of these proteins are fairly similar (Supplementary Fig. 11), with *cmcC<sub>trunc</sub>* exhibiting slightly lower expression levels than others. It could be possible that these lower expression levels are responsible for its low content in *cmcABC<sub>trunc</sub>* + D BDPs. All tested proteins were also soluble in similar amounts, except in the case of *cmcA*. *cmcA* is much more insoluble than other BMC-H proteins, although mass spectrometry analysis confirmed its inclusion in soluble composite *cmcABC'* + D, *cmcABC* + D, and *cmcABC<sub>trunc</sub>* + D BDPs (Supplementary Fig. 10a–b and c–d). Thus, the solubility and availability of BMC-H proteins could be dependable on the composition of other shell components. We tested the influence of expressing additional *cmcC'* and *cmcAB* from another promoter on the BDP size and yield (Supplementary Fig. 12); however, we did not observe any dramatic differences in yield or particle size distributions in gel filtration, just a very minor improvement of overall yield in case of additional *cmcC'*. It is possible that additional *cmcC'* or *cmcAB* expressed from a different promoter do not take part in the complex interplay between *cmcA*, *cmcB*, and *cmcC* genes located so closely to each other during their translation; perhaps formation of particular heterohexamers consisting of more than one type of BMC-H protein is the key to this process.



**Table 2 Summary of BDP encapsulation experiments (Supplementary Figs. 13–16).**

pET-Duet-1 T7-1	pET-Duet-1 T7-2	pRSF-Duet-1 T7-1 <sub>1</sub>	pRSF-Duet-1 T7-2 <sub>1</sub>	pRSF-Duet-1 T7-1 <sub>2</sub>	pRSF-Duet-1 T7-2 <sub>2</sub>	Results
cmcABC'	cmcD	CutC	-	-	-	CutC in all zones
cmcABC'	cmcD	CutF	-	-	-	No encapsulation observed
cmcABC'	cmcD	CutO	-	-	-	No encapsulation observed
cmcABC'	cmcD	CutH	-	-	-	No encapsulation observed
cmcABC'	cmcD	CutC	cmcE	-	-	CutC in all zones
cmcABC'	cmcD	CutF	cmcE	-	-	No encapsulation observed
cmcABC'	cmcD	CutO	cmcE	-	-	No encapsulation observed
cmcABC'	cmcD	CutH	cmcE	-	-	No encapsulation observed
cmcABC'	cmcD	CutC <sub>1-325</sub>	-	-	-	No encapsulation observed
cmcABC'	cmcD	CutC <sub>336-1128</sub>	-	-	-	CutC <sub>336-1128</sub> in all zones
cmcABC'	cmcD	CutC <sub>336-1128</sub>	CutO	-	-	CutO and CutC <sub>336-1128</sub> in all zones
cmcABC'	cmcD	CutC <sub>336-1128</sub>	CutF	-	-	CutC <sub>336-1128</sub> in all zones
cmcABC'	cmcD	CutC	CutO	-	-	CutO and CutC in all zones
cmcABC'	cmcD	CutC	CutF	-	-	CutC in all zones, CutF predominantly in large type particle zone
cmcABC'	cmcD	CutC	CutH	-	-	CutC in all zones
cmcABC'	cmcD	CutC	CutF	CutO	-	CutC, CutF, and CutO predominantly in large type particle zone
cmcABC'	cmcD	CutC	CutF	CutO	CutH	CutC, CutF, and CutO predominantly in large type particle zone
cmcC'	cmcD	CutC	-	-	-	CutC predominantly in small type particle zone
cmcC'	cmcD	CutC	CutF	CutO	-	No encapsulation observed
cmcC	cmcD	CutC	-	-	-	CutC predominantly in small type particle zone
cmcC	cmcD	CutC	CutF	CutO	-	No encapsulation observed

**Hierarchy of the GRM2 core encapsulation mechanism.** Since the GRM2 locus has several unusual traits, we wanted to test whether it is possible to encapsulate some of the core enzymes by the recombinant co-expression of our BDPs. We demonstrate our experiments with three core enzymes: CutC (signature enzyme choline lyase), CutO (alcohol dehydrogenase), and CutF (aldehyde dehydrogenase). CutD (CutC-activating enzyme) is an insoluble protein when produced recombinantly<sup>46</sup>, and we were not able to detect encapsulation of CutH (phosphotransacylase) at all in any of tested BDPs (Supplementary Fig. 14c–e). We selected the cmcABC' + D pET-Duet1 construct as the BDP platform for encapsulation experiments since it offered a more convenient co-expression setup in a two-plasmid expression system and a greater yield than that of native cmcABC + D particles. Our results are summarized in Table 2 and Supplementary Figs. 13–16.

To illuminate the role of the unique N-terminal extension of CutC, we created two new constructs by cutting CutC into two parts containing either the N-terminal 326 amino acids or the C-terminal 792 amino acids, which were visible within the electron density of our previously reported crystal structure<sup>46</sup>. We performed control experiments to test whether Superose 6 gel filtration can properly separate core enzymes from BDPs (Supplementary Figs. 17 and 18). For these experiments we used purified His6x-tagged core enzymes and as a size marker representing BDPs we chose 29 nm diameter bacteriophage AP 205 virus-like particles<sup>55</sup>. We concluded that gel filtration on the Superose 6 column is able to separate CutC (116–132 ml), CutO (116–140 ml), and CutH (120–144 ml) from AP 205 virus-like particles migrating as small type BDPs (Supplementary Fig. 11b). This confirmed that these three core enzymes as such generally do not overlap with BDP elution fractions. However, CutF eluted in a pattern both consistent with a tetrameric oligomerization state (roughly the same as a dimeric CutC) and also as a larger complex starting as early as 64 ml (Supplementary Figs. 17a, c and 18a), thus partially overlapping with BDP elution zones. The presence of CutF caused shift of elution of full-length CutC

towards larger-sized aggregates as early as 60 ml, but not N-terminally truncated CutC<sub>336-1128</sub>, indicating a crosslinking influence of CutF on full-length CutC. We also tested whether there is some direct interactions between formed shells and core enzymes or any such interactions mediated by *Escherichia coli* proteins. We expressed core enzymes (CutC + CutO or CutC + CutF + CutO) and BDPs (cmcABC' + D) in separate batches, then mixed the biomasses in equivalent amounts, and proceeded further with cell lysis, ultracentrifugation, and gel filtration as usually (Supplementary Figs. 18c, e). CutC + CutO and CutC + CutF + CutO biomasses without addition of cmcABC' + D BDPs were treated in a similar fashion as controls. Only in the case of CutC + CutF + CutO mixed with cmcABC' + D biomass we were able to detect low amounts of CutC core enzyme (Supplementary Fig. 12e), suggesting some association with BDPs. However, neither CutC nor CutO without BDPs, nor CutC or CutO, when mixed with cmcABC' + D BDPs, could be observed in fractions before 108 ml (Supplementary Figs. 12b–d), confirming the absence of association with already formed BDPs. This strongly suggests that the observed co-migration of CutC and CutO with BDPs after co-expression described below is almost certainly due to encapsulation. In gel filtration on Superose 6 column CutO is generally eluted slightly later than CutC (Supplementary Fig. 17) and in 6xHis tagged CutC pull-down assays CutO also failed to associate with it (Supplementary Fig. 11b). It could be that the chromatography is a too aggressive method disassembling unencapsulated CutC–CutO complexes. Alternatively, it is also possible that CutC and BMC shell forms a composite binding site for CutO. This, however, would result in CutO being localized exclusively on the inner shell surface and thus being encapsulated in relatively small amounts, especially, when the larger volume/surface ratios of native-type BMCs are considered. Such mechanism would result in a very uneven distribution of CutO in BMC.

Of all three tested core enzymes, CutC was the only enzyme capable of co-migration with BDPs by itself when co-expressed

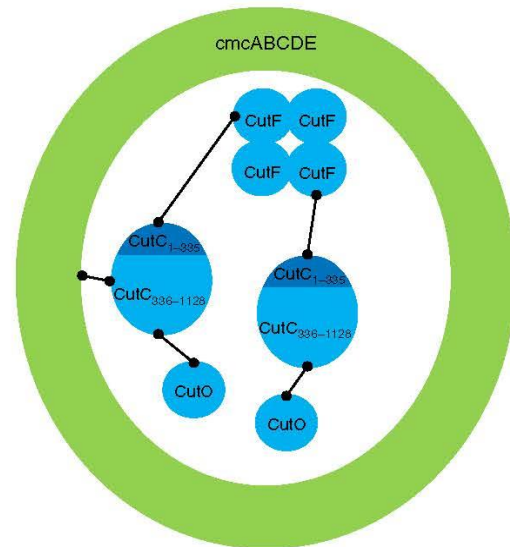


with *cmcABC'* + D shell genes (Supplementary Fig. 13a). The encapsulation of CutC was further confirmed by His6x tag capture experiments. Both large and small type BDPs containing encapsulated His6x-tagged CutC could not bind to the HisTrap column, although free CutC could be efficiently purified by this method<sup>45</sup>, suggesting that the CutC along with the His6x tag is sequestered inside the BDP lumen (Supplementary Fig. 19a, b). This was true for both small type and large type BDP materials. The presence of 6xHis tag in BDP-associated CutC was confirmed by western blot (Supplementary Fig. 19c). Neither CutO nor CutF was able to co-migrate with BDPs in detectable amounts. This was somewhat unexpected, considering that CutF (but not CutO) has an EP-like elongation (NEQNVERVIRQVLERLG) at the C-terminal end<sup>56</sup>. CutH contains an N-terminal EP-like elongation (MIDTLVREKIAARL)<sup>56</sup>, but we failed to detect any CutH co-migration with BDPs as well. The presence or absence of *cmcE* had no influence on co-migration—CutC was present, and CutO and CutF absent irrespective of the presence or absence of *cmcE* (Table 2, Supplementary Fig. 13d–f).

An unexpected observation was that both CutO and CutF co-migration with BDPs could be observed if they were co-expressed with CutC (Table 2, Supplementary Fig. 14). Testing the N- and C-terminal domains of CutC revealed that the N-terminal domain was not essential for the encapsulation of CutC C-terminal part and the encapsulation of the N-terminal domain as such could not be detected either (Supplementary Fig. 15a, b). Thus, the first 335 amino acids of CutC are not necessary for its encapsulation. The presence of the CutC N-terminal domain was also not needed for the encapsulation of CutO, which could be encapsulated by both full-length CutC and truncated CutC<sub>336–1128</sub>. However, the N-terminal domain was essential for the CutC-mediated co-migration of CutF with BDPs, which further highlighted the role of the N-terminal CutC domain (Supplementary Figs. 14 and 15). The presence of CutF changed the co-migration pattern; while encapsulated CutC and CutC + CutO were evenly spread throughout all BDP-containing fractions after gel filtration, indicating that there was no preference for a certain particle size, CutF co-migrated predominantly with large type BDPs (Supplementary Fig. 15b, c). Remarkably, the presence of CutF even shifted a portion of the CutC and CutO proteins from the small type BDP zone into large type BDP particle zone (Supplementary Fig. 15c). When the CutC + CutF + CutO proteins were co-expressed with the *cmcC'* + D and *cmcC* + D constructs, capable of forming only small type BDPs, no co-migration of any protein was detected, suggesting that the core size could be too large for encapsulation in small type particles. However, CutC alone could be encapsulated in *cmcC* + D particles very efficiently (Supplementary Fig. 16). To ensure proper identification, the identity of the encapsulated CutO and CutF bands were additionally confirmed by peptide mass fingerprinting analysis (Supplementary Fig. 20).

Our observations of CutF migration in co-expression with BDPs are consistent with the observations for purified core enzymes (Supplementary Fig. 17a)—in both cases CutC and CutF interaction had a size-increasing effect. We also concede that some proportion of CutC + CutF and CutC + CutF + CutO could be present in a free form intermixed with BDPs in the case of coexpressions with BDPs, since the presence of CutC could be observed in small amounts mixed with BDPs as well (Supplementary Fig. 12e). Thus, encapsulation of the core enzymes in the presence of CutF is not certain.

Nevertheless, our data strongly suggest that CutC acts as a mediator of the encapsulation of the enzymatic core. It must be noted that our recombinant BDP system may not completely accurately represent native encapsulation system of *K. pneumoniae*, but it is very likely that CutC plays the central encapsulation



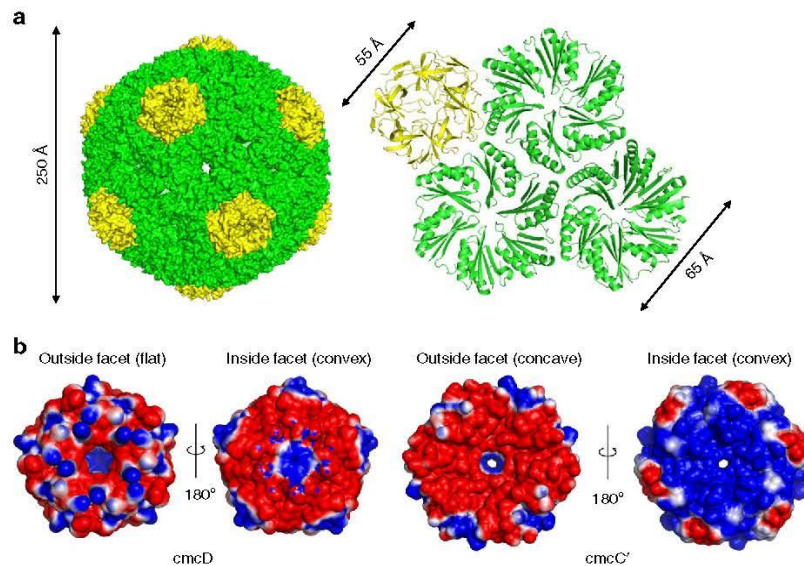
**Fig. 3 Proposed enzymatic core encapsulation mechanism of CutC, CutF, and CutO in GRM2 BDPs.** CutC is serving as an adaptor for the encapsulation of other enzymes. CutC C-terminal domain is responsible for encapsulation and also for interaction with CutO. The CutC N-terminal domain is responsible for CutC interaction with CutF. CutF together with CutC N-terminal domain crosslinks the enzymatic core and increases its size.

role in native conditions as well and the interaction between the CutC N-terminal portion and CutF is most likely necessary for the assembly of the native, large-sized enzymatic cores. An analog of CutF from a different BMC locus has been demonstrated to form a tetramer in a crystal structure<sup>35</sup>; therefore, it could serve as a multivalent cross-linker in the BMC core. Our proposed enzymatic core encapsulation mechanism is summarized in Fig. 3. The C-terminal part of CutC anchors itself and the entire enzymatic core to the shell and simultaneously ensures the encapsulation of CutO, while the N-terminal part of CutC ensures the encapsulation of CutF, which then further crosslinks the enzymatic core and increases its size.

**Cryo-EM characterization of BDPs.** We analyzed *cmcABC'D* + CutC<sub>336–1128</sub> BDPs with cryo-EM. The peak containing the small type BDPs in the gel filtration volumes from 90 to 100 ml was chosen for analysis because this material appeared to be the most uniform in TEM analysis (Fig. 2) and contained encapsulated CutC<sub>336–1128</sub> (Supplementary Fig. 6B).

We calculated a near-atomic (3.3 Å) resolution map for the icosahedral particles and built an atomic model (Fig. 4). The particles had a  $pT = 4$  quasi-symmetry with 12 *cmcD* pentamers occupying the vertices of the icosahedron and 180 BMC-H monomers that were arranged in 30 hexamers within the facets. The *cmcABC'* + D small type particle peak contains all three BMC-H proteins; most likely, they all contribute to the averaged electron density of BMC-H positions. The map is not of sufficiently high resolution to distinguish between these three types of proteins, so we chose to model *cmcC'* in the model, since the mass spectrometry data identified it as the most abundant BMC-H protein in *cmcABC'* + D small type particles (Supplementary Fig. 10a). Electron density for *cmcC'* chain in pentameric–hexameric contacts was interpretable from residues 3 to 88, for the rest of *cmcC'* chains in hexameric–hexameric contacts it was interpretable from residues 3 to 83. The resulting





**Fig. 4** Cryo-EM structure of pT=4 quasi-icosahedral BDP and its pentameric and hexameric components. **a** Surface model of pT = 4 quasi-icosahedral BDP particle, displayed on the left side. A ribbon model of a cmcD pentamer and three cmcC' hexamers is displayed on the right side. Pentameric cmcD protein is colored in yellow and hexameric cmcC' is colored in green. Note that the fivefold symmetry axis is located at the center of cmcD pentamer and threefold axis is located in the middle between three cmcC' hexamers. **b** Electrostatic surface potential of pentameric cmcD and hexameric cmcC'. Note the pores in the centers of pentamers and hexamers. The surface contour levels were set to  $-1$  kT/e (red) and  $+1$  kT/e (blue).

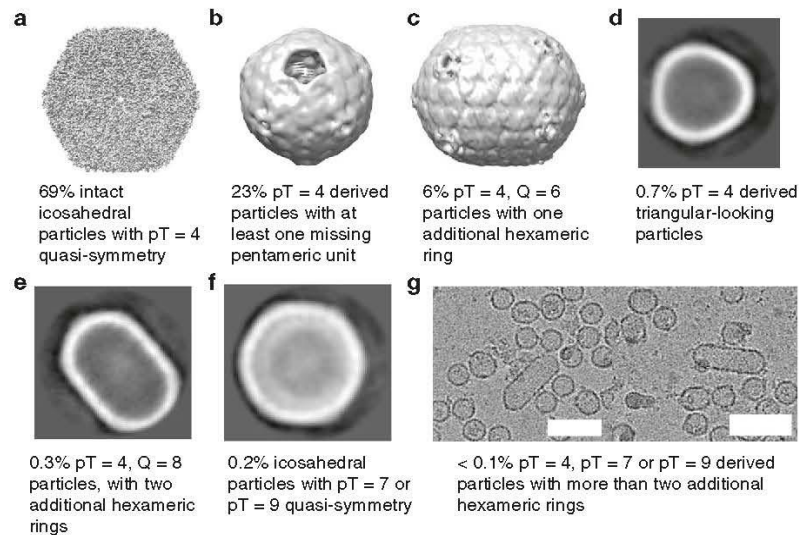
pT = 4 BDP particles are 250 Å in diameter. These particles are formed similarly to previously reported larger pT = 9 pseudosymmetric particles from *Haliangium ochraceum*<sup>28</sup>, pT = 4 particles comprised double-fused *Haliangium ochraceum* BMC-H proteins<sup>57</sup>, mixed icosahedral and elongated pT = 4 and pT = 3 particles from *Halotheca* sp<sup>58</sup>, and smaller T = 1 particles comprised circularly permuted BMC-H proteins<sup>59</sup>.

In the pT = 4 BDP shell, there are three pores: one is in the center of the cmcC' hexamer, the second is at the threefold axis between the cmcC' hexamers, and the third is in the center of the cmcD pentamer (Fig. 4). The pore in the center of the hexamer is approximately 7 Å in diameter as judged after subtraction of atomic Van der Waals diameters and contains typical GSG pore motifs on the rim (Fig. 4b, Supplementary Fig. 8). The pore in the center of the cmcD pentamer has a funnel shape and is approximately 16 Å wide on the outer surface but is only approximately 4 Å in diameter on the inner surface as judged by the Van der Waals diameters of atoms forming it (Fig. 4b). In electrostatic surface potential map it is visible that the cmcC' hexamer inner facet is charged more positively than the outside facet (Fig. 4b). The convex facets of cmcC' and cmcD are directed towards the lumen, a feature observed in earlier studies of other types of BMCs<sup>28,57,58</sup>. The narrowest part of the pentameric pore is very hydrophobic in nature, as it is lined by tyrosines and phenylalanines (Supplementary Fig. 21). Such a ring-like structure of aromatic amino acids in the BMC-P pore is unusual and seems to be a characteristic of the GRM2-type (Supplementary Fig. 8). This feature is not completely conserved, as *Aeromonas hydrophila* GRM2-type BMC-Ps have only one aromatic amino acid instead of two in the matching central pore forming motif (Supplementary Fig. 8). Most BMC-P proteins encoded by other loci contain a small helix between the β5 and β6 strands, and the rim of the pore is lined by GS motifs to form a pore of comparable size. Beta carboxysomal shell protein CcmL can be illustrated as an example of this typical fold (Supplementary Fig. 21). Compared to BMC-P proteins from other hosts, the

GRM2-type BMC-P sequence contains a deletion of five amino acids downstream from the pore motif, which thus transforms the small helix between the β5 and β6 strands into a shorter loop. In our structure the tip of this loop is disordered—residues K66, D67, and R68 and side chains of L64, N65, Y69, and K70 are invisible in electron density and are not modeled. The bulky aromatic amino acids seem to compensate for the shorter loop in terms of closing the pore. Remarkably, the GRM3-type BMC-P contains the same five amino acid deletion but retains the typical GS pore motif, so the GRM3 BMC-P pore would be significantly wider than that of any other BMC-P (Supplementary Fig. 8). The third pore at the BDP threefold axis is 3 Å in diameter and mostly featureless, as it is formed by the nearby main chain of glycine residues. Substrate and cofactor transport across the shell is probably mediated by the largest cmcC' pore since the other two types of pores are too small or hydrophobic to be involved in this process, if it is assumed that the pores are mostly static in their conformations.

The cryo-EM analysis revealed significant heterogeneity among the particles (Fig. 5). While the majority of particles were revealed to be pT = 4 pseudosymmetric icosahedrons, roughly a quarter of these pT = 4 particles were missing at least one pentameric unit, resulting in an incomplete particle. There were also smaller quantities of elongated pT = 4, Q = 6 and pT = 4, Q = 8 particles, and there were more elongated filamentous particles and icosahedrons with larger pseudosymmetry numbers (possibly pT = 7 or pT = 9). A very unusual subpopulation consisted of smaller triangular-shaped particles, which, judging by their size, could be derivatives of three fused pT = 4 icosahedrons (Fig. 5d). Several asymmetric C1 classes of icosahedral particles with visible electron density inside the BDP particle could also be identified (Supplementary Fig. 22), which we hoped contained encapsulated CutC<sub>336</sub><sup>1128</sup>. However, despite extensive efforts, including particle subtraction, we failed to generate any meaningful reconstructions beyond 20–30 Å resolution of the object inside the particles. There are several possible reasons for this result.





**Fig. 5 Cryo-EM classes of BDP subtypes.** **a** 3D class at 3.3 Å resolution and atomic model of whole intact pT = 4 quasi-symmetric icosahedral particles, in total 69%. **b** 8.8 Å resolution 3D class of pT = 4 derived particles, with at least one missing pentameric unit, in total 23%. **c** 9.6 Å resolution 3D class of pT = 4, Q = 6 quasi-symmetric particles, in total 6%. **d** 2D class of triangular-looking particles, judging from the size probably derived from three fused pT = 4 particles, in total 0.7%. **e** 2D class of pT = 4, Q = 8 particles, in total 0.3%. **f** 2D class of pT = 7 or pT = 9 quasi-icosahedral particles, in total 0.2%. **g** Separate micrographs of tubular particles elongated with more than two hexameric rings, in total less than 0.1%. Scale bar: 60 nm.

First, the number of such particles is quite small, with only 3–5 thousand particles in each class. Additionally, choline-free CutC is partially disordered, as demonstrated in our previous research<sup>46</sup>, and this could complicate the process of the sorting of proper two-dimensional (2D) classes. In addition, it is also possible that encapsulated CutC is not rigidly bound at a particular position within the BDP shell. If this is indeed the case, the CutC protein acts as a nucleator of shell assembly. Still, the fact that the unidentified electron density was located in the inside of the particles suggests that some kind of cargo is encapsulated, and it can reasonably be suspected as CutC.

The minor BDP shell subpopulations illustrate the flexibility of BMC-H proteins in creating different shapes, including the ability to form curved surfaces in small tubular particles, more planar surfaces in larger icosahedral particles and semi-regular polyhedrons by fusing separate icosahedrons into more complex formations. This is not surprising since native metabolosomes, when visualized in TEM, have large, irregular shapes<sup>12,13,15</sup>. The multiple building modes are a result of the flexible contacts between BMC-H hexamers that range from 150° in pT = 4 icosahedrons to 160° in the elongated pT = 4, Q = 6 particles (Supplementary Fig. 23) to a planar 180° (ref. <sup>21</sup>), as has been previously demonstrated in the crystal structure of a BMC-H protein homolog.

The main hexamer–hexamer contacts between cmcC' are formed by the conservative K–R–X triad (Fig. 6a, Supplementary Fig. 24), which has already been demonstrated in previous reports<sup>21,22</sup>. A hydrogen bond is formed between the K25 side chain nitrogen and the main chain of R78 in the adjacent monomer. The R78 side chain can in turn form a hydrogen bond with the main chains of K25 and A27 in the opposite monomer. These interactions result in a ring structure of four linked cmcC' monomers. On the inside surface, there is also a probable network of salt bridges formed by the E62, R66, E16, and R74 residues. The side chain electron density is very poor for this network, and it is impossible to visualize the exact contacts between the particular residues, but their positions in the main chain suggest

that interactions between them do occur. These residues are not completely conserved, but most cmcC BMC-H homologs contain several basic and acidic residues in the same regions (Supplementary Fig. 8). This salt bridge network could play a significant role in controlling the planarity of the contacts between the hexameric units; for example, the distances between the amino acids at the E62 and R66 positions in opposite monomers can vary significantly from 8 Å in a planar crystal contact, as shown previously<sup>21</sup>, to almost 3 Å in the 150° contact found in our pT = 4 BDP structure.

There are also several specific polar pentameric–hexameric cmcC'–cmcD interactions (Fig. 6b). There are hydrogen bonds between the cmcD G49 main chain and the cmcC' K25 side chain and between cmcD E79 side chain and cmcC' G51 side chain. There are also two possible salt bridges between the cmcC' R78 and cmcD D54 and the cmcC' D49 and cmcD K80 side chains. The solvent-accessible buried surface areas between the hexameric–hexameric and the pentameric–hexameric contacts are very similar (1050 and 1010 Å<sup>2</sup>, respectively). Nevertheless, in our case, roughly a quarter of the pT = 4 BDPs were missing at least one pentameric subunit, a somewhat surprising fact when the numerous pentameric–hexameric contacts and comparable buried surface areas are considered. In another case, a pT = 4 BMC shell particle was formed by a double-fused BMC-H protein even in the absence of any BMC-P proteins<sup>57</sup>. This demonstrates that BMC-P proteins are much more loosely integrated into GRM2 BDPs than BMC-H proteins and has potential implications for designing synthetic BMC particles with encapsulated cargo.

## Discussion

Our results demonstrate that even highly similar BMC-H proteins may have significantly different assembly properties. A computational study has shown that shell proteins with high spontaneous curvature are more efficient in forming empty particles<sup>60</sup>. High spontaneous curvature components would form nanotube-like or







first noticed in tomato bushy stunt virus (TBSV)<sup>64</sup>. Whether this actually contributes to the contact angles between BMC-H hexamers remains unclear.

The assembly of the enzymatic core is a crucial step in BMC formation. Carboxysomes contain specific adaptor proteins, the purpose of which is to ensure the crosslinking and encapsulation of the enzymatic core<sup>33,42–44</sup>. This is not the case for metabolosomes, where there are no specific additional adaptor proteins and encapsulation must be carried out by EP that is directly attached to enzymes. However, not all core enzymes contain identifiable EP, and it has been hypothesized previously that a piggybacking mechanism ensures the encapsulation of the core components without EP<sup>56</sup>. Experimental observations supports this as well—activating enzyme of GRM3 signature enzyme binds to the signature enzyme and is probably encapsulated in such manner<sup>36</sup>. There is also some evidence of EP-independent interaction of core enzymes with the BMC shell—for example, PduS interacts with one of the *Citrobacter freundii* pdu shell component PduT without mediation of an identifiable EP<sup>65</sup>. There is a benefit to a piggybacking strategy, as it would provide an opportunity to control the stoichiometry of encapsulated enzymes and ensure that no component is present in such low quantities as to bottleneck the enzymatic cascade.

Our experimental results show that this mechanism could be true for GRM2 enzymatic core components as well, as the interactions between some GRM2 core components are strictly hierarchical. Encapsulation is dependent on sequential interactions between particular enzymes, where CutC is responsible for core encapsulation in the shell and, together with CutF, is most likely responsible for crosslinking the core into a larger assembly (Fig. 3). It is interesting that CutC<sub>336–1128</sub> has no identifiable EP sequence and therefore results in encapsulation that seems to be different from that resulting from the canonic EP-shell interaction. A reason for this could be the lack of BMC-T proteins in the GRM2 locus. BMC-T PduB has been shown to be an essential component for EP-mediated encapsulation of cargo<sup>40</sup>, because of this, an entirely different BMC encapsulation system may have evolved in the GRM2 locus to compensate for the absence of BMC-T. Interestingly, CutF still contains an EP-like sequence at the C-terminal end and CutH has an analogous sequence at the N-terminal end, but their functions in our recombinant BDP system, if any, are limited purely to crosslinking. We failed to detect any CutH encapsulation or co-migration with BDPs or core enzymes, although it is vital for regeneration of coenzyme A inside the lumen and most likely is encapsulated in native BMCs<sup>66</sup>. The observations are not necessarily true for native GRM2-type BMCs—it is possible that EPs of CutF and CutH become functional only in native conditions. Nevertheless, the role of the CutC as a central adaptor stays strongly suspected in native GRM2 BMCs as well. The encapsulation pattern could point to another important consideration for metabolosome organization. An independent encapsulation of CutO and CutF mediated by CutC could be a good way to ensure that both alcohol and aldehyde dehydrogenase are in close proximity to the signature enzyme. Proper NAD<sup>+</sup>/NADH recycling inside the enclosed compartment is based on the requirement that the aldehyde intermediate is more or less equally divided between the alcohol and aldehyde dehydrogenase components<sup>9,20</sup>, and this could be a good mechanism for ensuring this.

## Methods

**Construct design.** GenBank™ entry ARRZ01000032.1 was used for PCR primer design, with the following entries that correspond to named proteins: CutC (EPO20241.1), CutO (EPO20327.1), CutF (EPO20363.1), cmcA (EPO20272.1), cmcB (EPO20357.1), cmcC (EPO20328.1), cmcE (EPO20271.1), and cmcD (EPO20293.1). cmcB had to be obtained via separate gene synthesis by General

Biosystems (USA), since it was not possible to amplify it due to interference from the cmcA and cmcC genes. Two additional constructs were made from full-length CutC: CutC<sub>1–325</sub>, corresponding to the first 325 amino acids of CutC, and CutC<sub>336–1128</sub>, corresponding to amino acids 336–1128. Constructs containing cmcABC and cmcABC' were generated by amplifying and cloning the whole cmcABC region with forward cmcA and reverse cmcC primers. Genomic DNA obtained from a *Klebsiella pneumoniae* strain MSCL535 (Microbial Strain Collection of Latvia) was used as a PCR template. The primer pairs used are listed in Supplementary Table 1. To ensure maximal simplicity in the design and co-expression, the proteins were expressed in a pET-Duet1/pRSP-Duet1 two-plasmid system. These plasmid vectors contain identical Duet regions that contain dual promoters and can also be cotransformed. DNA transcribed under the T7-1 promoter was inserted using *NcoI* and *HindIII* sites, and for insertion after the T7-2 promoter, *NdeI* and *XhoI* sites were used. To increase the number of available promoters from four to six for cmcABC' + D + CutC + CutF + CutO and cmcABC' + D + CutC + CutF + CutO + CutH constructs, we amplified the whole Duet region containing CutO or CutO + CutH and inserted it at the end of the pRSP-Duet1 vector at the *XhoI* site.

**Protein expression and purification.** pET-Duet1 plasmids containing cmcABC + cmcD and, optionally, pRSP-Duet1 plasmids containing CutC/CutO/CutF/cmcE were transformed into *Escherichia coli* BL21-DE3 chemically competent cells (Sigma-Aldrich, cat. No CMC0014). Cells were grown in 2xTY medium containing 50 µg/ml ampicillin and (if pRSP-Duet1 vector was used) 30 µg/ml kanamycin. Cells were grown at +37 °C to OD<sub>590</sub> 0.7 and shaken at 200 r.p.m., cooled at 20 °C for 30 min and induced with 1 mM IPTG. Induction was performed overnight for approximately 16 h at 20 °C with shaking at 200 rpm. The biomass was then collected, centrifuged, and frozen at –20 °C.

Cell lysis was performed in a buffer containing 100 mM Tris-HCl (pH 8.0), 300 mM NaCl, 0.1% Triton X-100, 20 mM MgSO<sub>4</sub>, 0.1 mg/ml DNase, 1 mg/ml lysozyme, 1 mM PMSF, and 2 mM DTT. The lysate was incubated at +6 °C with shaking for 1 h and then centrifuged at 10,000g for 10 min. The supernatant was then collected and centrifuged at 50,000g for 3 h. The supernatant was discarded, and the pellet was suspended in a small volume of 300 mM NaCl and 20 mM Tris-HCl (pH 8.0), maintaining a proportion of 4 ml per 5–10 g of the initial cell biomass. The suspension was then centrifuged at 10,000g for 10 min, and the supernatant was collected. The solution was then loaded on a 16/900 Superose 6 gel filtration column (GE Healthcare) equilibrated in 300 mM NaCl and 20 mM Tris-HCl (pH 8.0), and 2 ml fractions were collected. Fractions 30–60 (corresponding to 60–120 ml), which contained the BMC proteins, were analyzed with 15% SDS-PAGE and TEM.

His6x-CutC, His6x-CutC<sub>336–1128</sub>, His6x-CutH, His6x-CutF, and His6x-CutH purification and BDP capture tests were performed by Ni<sup>2+</sup> affinity chromatography. The frozen biomass was suspended in lysis buffer containing 100 mM Tris-HCl (pH 8.0), 200 mM NaCl, 1% Triton X-100, 1 mM PMSF, and 2 mM DTT. Cells were lysed by ultrasound, and the lysate was centrifuged at 14,000g for 40 min. CmcE and CutO were purified on a 1-ml HisTrap column (GE Healthcare). For this step, 20 mM imidazole in 40 mM Tris-HCl (pH 8.0) and 300 mM NaCl was used as a washing buffer, and 300 mM imidazole in 40 mM Tris-HCl (pH 8.0) and 300 mM NaCl were used for the elution buffer. The cmcABC' + D + CutC and cmcABC' + D + CutC + CutF + CutO capture tests were performed in the same manner; all fractions were equalized by volume and the volume of the sample loaded on the SDS-PAGE gel was equalized as well. Gel filtration experiments were repeated at least twice to confirm the observations.

Western blot analysis of His6x-CutO and BDP-encapsulated CutC were performed by using His-Tag Antibody HRP Conjugate Kit (Novagen, cat. No 71840-3). The samples were loaded on a 8% SDS-PAGE gel and transferred afterwards to a nitrocellulose membrane in a semi-dry fashion. The blot was visualized with ECL Prime chemiluminescent detection reagent kit (GE Healthcare, cat. No RPN2232).

**Mass spectrometry analysis.** The SDS-PAGE bands of CutO, CutF, and the individual components of BDPs were identified by the peptide mass fingerprinting method. The protein band was cut out from the Coomassie blue-stained polyacrylamide gel and washed twice for 1 h with 500 µl of 0.2 M ammonium bicarbonate and 50% acetonitrile. Then, the gel fragments were washed twice with 200 µl of 100% acetonitrile and incubated with a trypsin (Sigma-Aldrich, cat. No T6567) solution containing 40 mM ammonium bicarbonate and 10% acetonitrile for 3 h at +37 °C. A total of 1 µl of the obtained peptide solution was mixed with 1 µl of 0.1% TFA and 1 µl of matrix solution containing 15 mg/ml 2,5-dihydroxyacetophenone in 20 mM ammonium citrate, and 75% ethanol. For BDP analyses, samples from particular fractions of gel filtration on Superose 6 were used directly instead of peptide solution. Then, 1 µl of the obtained mixture was loaded on the target plate, dried, and analyzed using a Bruker Daltonics Autoflex MALDI-TOF mass spectrometer.

**TEM analysis.** BDPs were visualized by TEM with uranyl acetate negative staining. A 5 µl sample drop was placed on a formvar-coated TEM 200 copper grid (Sigma-Aldrich) and incubated for 3 min. The grids were dried, briefly washed with 1 mM



EDTA solution, and negatively stained with 1% uranyl acetate for 1 min. The grids were then dried and analyzed on a JEM-1230 TEM electron microscope at 100 kV.

**Cryo-EM analysis of BDPs and model building.** A total of 4  $\mu$ l of purified cmcABC + cmcD + CutC<sub>335</sub> BDPs with a concentration of 1 mg/ml in 100 mM NaCl and 20 mM Tris-HCl (pH 8.0) were applied to the EM grid. The grids (Quantifoil, Cu grids, 200 mesh, R2/1) were blotted for 4 s using a Vitrobot (Mark IV, Thermo Fisher) at 18 °C and in 100% humidity, plunge-frozen in liquid ethane-propane and stored in liquid nitrogen until further use. Cryo-EM data were collected with a 200 kV Talos Arctic microscope (Thermo Scientific) equipped with a Falcon 3EC direct electron detector (Thermo Scientific). A total of 1316 images (Supplementary Fig. 26) was collected in an automated manner using EPU software (Thermo Scientific). The data were collected at the nominal magnification of  $\times 120,000$ , corresponding to a calibrated pixel size of 1.23  $\text{\AA}$  px<sup>-1</sup>, with an underfocus in the range of  $-1.4$  to  $-3.0$   $\mu$ m and an exposure time of 1.0 s, comprising an overall dose of 60 e<sup>-</sup>  $\text{\AA}^2$  for each specimen. Data from a single exposure were stored as a set of 40 movie frames.

Motion-correction and dose-weighting of the frames were performed in MotionCor2 (ref. 67), and CTF correction was performed with Gctf<sup>68</sup>. The single particle analysis was performed with the RELION 3.0 pipeline<sup>69</sup>. The general scheme used for the analysis is given in Supplementary Fig. 13. Approximately 900 particles were picked by log-based autopicking. After several rounds of reference-based autopicking and a final manual inspection, a total of 62,533 particles were selected. A subsequent 2D classification was performed that illuminated a noticeable heterogeneity in the BDP particle morphology. Five 2D classes with a total amount of 45,915 particles were selected for three-dimensional (3D) de novo model generation in I point group symmetry. The obtained icosahedral model was then used in two 3D classifications—in I and C1 point group symmetries. The icosahedral model was 3D refined and postprocessed, and the refined map reached 3.8  $\text{\AA}$  resolution. After further CTF refinement and Bayesian polishing cycles, the refined particles were used for another high-resolution 3D refinement, reaching a final resolution of 3.3  $\text{\AA}$ . The final map was sharpened by applying a *b*-factor of  $-149.892$   $\text{\AA}^2$ . Two additional lower resolution maps were made; one was in C1 point group symmetry based on two 3D classes of C1 classification and the other was in D5 point group symmetry based on two 2D classes of initial 2D classification (Supplementary Fig. 13). The C1 reconstruction was obtained from 11,639 particles and the D5 reconstruction was obtained from 3153 particles. The final resolutions after high-resolution 3D refinement were 8.8  $\text{\AA}$  for the C1 map and 9.6  $\text{\AA}$  for the D5 map. For all final 3D refinements, the resolutions were estimated using the 0.143 cut-off criterion with gold-standard Fourier shell correlation between the two independently refined half-maps.

High similarity X-ray crystallographic structures were available for both cmcD (PDB ID 4N8X) and cmcC' (PDB ID 4QIV), so homology models built with SWISS-MODEL<sup>70</sup> were used as the initial models. The initial fitting of the models in the map was performed manually by using UCSF-Chimera<sup>71</sup>. The models were then further built manually using Coot<sup>72</sup> and refined with REFMAC5 in the CCP-EM package<sup>73</sup> and PHENIX<sup>74</sup>. Residues K66, D67, and R68 and side chains of L64, N65, Y69, and K70 of cmcD are invisible in electron density and are not modeled. Images were generated in PyMol. The overall statistics of the maps and the atomic model are listed in Supplementary Table 2. cmcC' chain in pentameric-hexameric contacts was interpretable from residues 3 to 88 (out of 100 total residues), for the rest of cmcC' chains in hexameric-hexameric contacts it was interpretable from residues 3 to 83 (out of 100 total residues). cmcD chain was interpretable from residues 1 to 84 (out of 88 total residues); also, residues K66, D67, and R68 and side chains of L64, N65, Y69, and K70 are invisible in electron density and are not modeled.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The associated macromolecular structural data are available in the Protein Data Bank repository as 6QN1 entry and in the Electron Microscopy Data Bank repository as EMD-4595, EMD-4596, and EMD-4597 entries. All other relevant data are included in the paper and its Supplementary Information files.

Received: 25 April 2019; Accepted: 18 December 2019;

Published online: 20 January 2020

## References

- Drews, G. & Niklowitz, W. Beiträge zur Cytologie der Blaualgen. II. Zentroplasma und granulare Einschlüsse von Phormidium uncinatum. *Arch. Mikrobiol.* **24**, 147–162 (1956).
- Kerfeld, C. A. et al. Protein structures forming the shell of primitive bacterial organelles. *Science* **309**, 936–938 (2005).
- Yeates, T. O., Kerfeld, C. A., Heinhorst, S., Cannon, G. C. & Shively, J. M. Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nat. Rev. Microbiol.* **6**, 681–691 (2008).
- Kerfeld, C. A. & Melnicki, M. R. Assembly, function and evolution of cyanobacterial carboxysomes. *Curr. Opin. Plant Biol.* **31**, 66–75 (2016).
- Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F. & Sutter, M. Bacterial microcompartments. *Nat. Rev. Microbiol.* **16**, 277–290 (2018).
- Lee, M. J., Palmer, D. J. & Warren, M. J. Biotechnological advances in bacterial microcompartment technology. *Trends Biotechnol.* **37**, 325–336 (2018).
- Cheng, S., Liu, Y., Crowley, C. S., Yeates, T. O. & Bobik, T. A. Bacterial microcompartments: their properties and paradoxes. *BioEssays* **30**, 1084–1095 (2008).
- Jorda, J., Lopez, D., Wheatley, N. M. & Yeates, T. O. Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci.* **22**, 179–195 (2013).
- Axen, S. D., Erbilgin, O. & Kerfeld, C. A. A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Comput. Biol.* **10**, e1003898 (2014).
- Cannon, G. C., Heinhorst, S. & Kerfeld, C. A. Carboxysomal carbonic anhydrases: structure and role in microbial CO<sub>2</sub> fixation. *Biochim. Biophys. Acta* **1804**, 382–392 (2010).
- Turmo, A., Gonzalez-Esquer, C. R. & Kerfeld, C. A. Carboxysomes: metabolic modules for CO<sub>2</sub> fixation. *FEMS Microbiol. Lett.* **364**, fnx176 (2017).
- Bobik, T. A., Havemann, G. D., Busch, R. J., Williams, D. S. & Aldrich, H. C. The propanediol utilization (*pdu*) operon of *Salmonella enterica* serovar Typhimurium LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme B12-dependent 1,2-propanediol degradation. *J. Bacteriol.* **181**, 5967–5975 (1999).
- Havemann, G. D. & Bobik, T. A. Protein content of polyhedral organelles involved in coenzyme B12-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar Typhimurium LT2. *J. Bacteriol.* **185**, 5086–5095 (2003).
- Kofoid, E., Rappleye, C., Stojiljkovic, I. & Roth, J. The 17-gene ethanolamine (*eut*) operon of *Salmonella typhimurium* encodes five homologues of carboxysome shell proteins. *J. Bacteriol.* **181**, 5317–5329 (1999).
- Herring, T. I. et al. Bacterial microcompartment is used for choline fermentation by *Escherichia coli* 536. *J. Bacteriol.* **24**, 200 (2018).
- Petit, E. et al. Involvement of a bacterial microcompartment in the metabolism of fucose and rhamnose by *Clostridium phytofermentans*. *PLoS ONE* **8**, e54337 (2013).
- Erbilgin, O., McDonald, K. L. & Kerfeld, C. A. Characterization of a plantomycetal organelle: a novel bacterial microcompartment for the aerobic degradation of plant saccharides. *Appl. Environ. Microbiol.* **80**, 2193–2205 (2014).
- Penrod, J. T. & Roth, J. R. Conserving a volatile metabolite: a role for carboxysome-like organelles in *Salmonella enterica*. *J. Bacteriol.* **188**, 2865–2874 (2006).
- Chowdhury, C. et al. Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proc. Natl. Acad. Sci. USA* **112**, 2990–2995 (2015).
- Zarzycki, J., Erbilgin, O. & Kerfeld, C. A. Bioinformatic characterization of glycol radical enzyme-associated bacterial microcompartments. *Appl. Environ. Microbiol.* **81**, 8315–8329 (2015).
- Sinha, S. et al. Alanine scanning mutagenesis identifies an asparagine-arginine-lysine triad essential to assembly of the shell of the Pdu microcompartment. *J. Mol. Biol.* **426**, 2328–2345 (2014).
- Tanaka, S., Sawaya, M. R., Phillips, M. & Yeates, T. O. Insights from multiple structures of the shell proteins from the beta-carboxysome. *Protein Sci.* **18**, 108–120 (2009).
- Sutter, M. et al. Visualization of bacterial microcompartment facet assembly using high-speed atomic force microscopy. *Nano Lett.* **16**, 1590–1595 (2016).
- Dryden, K. A., Crowley, C. S., Tanaka, S., Yeates, T. O. & Yeager, M. Two-dimensional crystals of carboxysome shell proteins recapitulate the hexagonal packing of three-dimensional crystals. *Protein Sci.* **18**, 2629–2635 (2009).
- Heldt, D. et al. Structure of a trimeric bacterial microcompartment shell protein, EtuB, associated with ethanol utilization in *Clostridium kluyveri*. *Biochem. J.* **423**, 199–207 (2009).
- Klein, M. G. et al. Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. *J. Mol. Biol.* **392**, 319–333 (2009).
- Cai, F. et al. The structure of CcmP, a tandem bacterial microcompartment domain protein from the  $\beta$ -carboxysome, forms a subcompartment within a microcompartment. *J. Biol. Chem.* **288**, 16055–16063 (2013).
- Sutter, M., Greber, B., Aussignargues, C. & Kerfeld, C. A. Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science* **356**, 1293–1297 (2017).
- Takenoya, M., Nikolakakis, K. & Sagermann, M. Crystallographic insights into the pore structures and mechanisms of the EutL and EutM shell proteins of



- the ethanolamine-utilizing microcompartment of *Escherichia coli*. *J. Bacteriol.* **192**, 6056–6063 (2010).
30. Thompson, M. C., Cascio, D., Leibly, D. J. & Yeates, T. O. An allosteric model for control of pore opening by substrate binding in the EutL microcompartment shell protein. *Protein Sci.* **24**, 956–975 (2015).
  31. Tanaka, S. et al. Atomic-level models of the bacterial carboxysome shell. *Science* **319**, 1083–1086 (2008).
  32. Wheatley, N. M., Gidaniyan, S. D., Liu, Y., Cascio, D. & Yeates, T. O. Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins. *Protein Sci.* **22**, 660–665 (2013).
  33. Kinney, J. N., Salmeen, A., Cai, P. & Kerfeld, C. A. Elucidating essential role of conserved carboxysomal protein CcmN reveals common feature of bacterial microcompartment assembly. *J. Biol. Chem.* **287**, 17729–17736 (2012).
  34. Pan, C. et al. Short N-terminal sequences package proteins into bacterial microcompartments. *Proc. Natl. Acad. Sci. USA* **107**, 7509–7514 (2010).
  35. Lawrence, A. D. et al. Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. *ACS Synth. Biol.* **3**, 454–465 (2014).
  36. Zarzycki, J., Sutter, M., Cortina, N. S., Erb, T. J. & Kerfeld, C. A. In vitro characterization and concerted function of three core enzymes of a glycol radical enzyme-associated bacterial microcompartment. *Sci. Rep.* **7**, 42757 (2017).
  37. Jakobson, C. M., Slininger Lee, M. F. & Tullman-Ereck, D. De novo design of signal sequences to localize cargo to the 1,2-propanediol utilization microcompartment. *Protein Sci.* **26**, 1086–1092 (2017).
  38. Quin, M. B., Perdue, S. A., Hsu, S. Y. & Schmidt-Dannert, C. Encapsulation of multiple cargo proteins within recombinant Eut nanocompartments. *Appl. Microbiol. Biotechnol.* **100**, 9187–9200 (2016).
  39. Cai, F., Bernstein, S. L., Wilson, S. C. & Kerfeld, C. A. Production and characterization of synthetic carboxysome shells with incorporated luminal proteins. *Plant Physiol.* **170**, 1868–1877 (2016).
  40. Lehman, B. P., Chowdhury, C. & Bobi, T. A. The N terminus of the PduB protein binds the protein shell of the Pdu microcompartment to its enzymatic core. *J. Bacteriol.* **199**, e00785-16 (2017).
  41. Fan, C., Cheng, S., Sinha, S. & Bobik, T. A. Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. *Proc. Natl. Acad. Sci. USA* **109**, 14995–15000 (2012).
  42. Cot, S. S., So, A. K. & Espie, G. S. A multiprotein bicarbonate dehydration complex essential to carboxysome function in cyanobacteria. *J. Bacteriol.* **190**, 936–945 (2008).
  43. Chaijarasphong, T. et al. Programmed ribosomal frameshifting mediates expression of the  $\alpha$ -carboxysome. *J. Mol. Biol.* **428**, 153–164 (2015).
  44. Liu, Y. et al. Deciphering molecular details in the assembly of alpha-type carboxysome. *Sci. Rep.* **8**, 15062 (2018).
  45. Craciun, S., Marks, J. A. & Balskus, E. P. Characterization of choline trimethylamine-lyase expands the chemistry of glycol radical enzymes. *ACS Chem. Biol.* **9**, 1408–1413 (2014).
  46. Kalnins, G. et al. Structure and function of CutC choline lyase from human microbiota bacterium *Klebsiella pneumoniae*. *J. Biol. Chem.* **290**, 21732–21740 (2015).
  47. Wang, Z. et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011).
  48. Schiattarella, G. G. et al. Gut microbe-generated metabolite trimethylamine-N-oxide as cardiovascular risk biomarker: a systematic review and dose-response meta-analysis. *Eur. Heart J.* **38**, 2948–2956 (2017).
  49. Lee, M. J. et al. Engineered synthetic scaffolds for organizing proteins within the bacterial cytoplasm. *Nat. Chem. Biol.* **14**, 142–147 (2018).
  50. Liang, M., Frank, S., Lunsdorf, H., Warren, M. J. & Prentice, M. B. Bacterial microcompartment-directed polyphosphate kinase promotes stable polyphosphate accumulation in *E. coli*. *Biotechnol. J.* **12**, 1600415 (2017).
  51. Choudhary, S., Quin, M. B., Sanders, M. A., Johnson, E. T. & Schmidt-Dannert, C. Engineered protein nano-compartments for targeted enzyme localization. *PLoS ONE* **7**, e33342 (2012).
  52. Wagner, H. J., Capitain, C. C., Richter, K., Nessling, M. & Mampel, J. Engineering bacterial microcompartments with heterologous enzyme cargos. *Eng. Life Sci.* **17**, 36–46 (2017).
  53. Hagen, A., Sutter, M., Sloan, N. & Kerfeld, C. A. Programmed loading and rapid purification of engineered bacterial microcompartment shells. *Nat. Commun.* **9**, 2881 (2018).
  54. Lee, M. J. et al. De novo targeting to the cytoplasmic and luminal side of bacterial microcompartments. *Nat. Commun.* **9**, 3413 (2018).
  55. Tissot, A. C. et al. Versatile virus-like particle carrier for epitope based vaccines. *PLoS ONE* **5**, e9809 (2010).
  56. Aussenargues, C., Paasch, B. C., Gonzalez-Esquer, R., Erbilgin, O. & Kerfeld, C. A. Bacterial microcompartment assembly: the key role of encapsulation peptides. *Commun. Integr. Biol.* **8**, e1039755 (2015).
  57. Sutter, M., McGuire, S., Ferlez, B. & Kerfeld, C. A. Structural characterization of a synthetic tandem-domain bacterial microcompartment shell protein capable of forming icosahedral shell assemblies. *ACS Synth. Biol.* **8**, 668–674 (2019).
  58. Sutter, M. et al. Structure of a synthetic beta-carboxysome shell. *Plant Physiol.* **181**, 1050–1058 (2019).
  59. Jorda, J., Leibly, D. J., Thompson, M. C. & Yeates, T. O. Structure of a novel 13 nm dodecahedral nanocage assembled from a redesigned bacterial microcompartment shell protein. *Chem. Commun. (Camb.)* **52**, 5041–5044 (2016).
  60. Mohajerani, P. & Hagan, M. P. The role of the encapsulated cargo in microcompartment assembly. *PLoS Comput. Biol.* **14**, e1006351 (2018).
  61. Mayer, M. J. et al. Effect of bio-engineering on size, shape, composition and rigidity of bacterial microcompartments. *Sci. Rep.* **6**, 36899 (2016).
  62. Parsons, J. B. et al. Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. *J. Mol. Cell* **38**, 305–315 (2010).
  63. Pang, A., Frank, S., Brown, I., Warren, M. J. & Pickersgill, R. W. Structural insights into higher order assembly and function of the bacterial microcompartment protein PduA. *J. Biol. Chem.* **289**, 22377–22384 (2014).
  64. Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, P. K. & Bricogne, G. Tomato bushy stunt virus at 2.9 Å resolution. *Nature* **276**, 368–373 (1978).
  65. Parsons, J. B. et al. Characterisation of PduS, the pdu metabolosome corrin reductase, and evidence of substructural organisation within the bacterial microcompartment. *PLoS ONE* **5**, e14009 (2010).
  66. Erbilgin, O., Sutter, M. & Kerfeld, C. A. The structural basis of coenzyme A recycling in a bacterial organelle. *PLoS Biol.* **14**, e1002399 (2016).
  67. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
  68. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
  69. Zivano, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, e42166 (2018).
  70. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
  71. Pettersen, E. F. et al. UCSF chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
  72. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
  73. Burnley, T., Palmer, C. M. & Winn, M. Recent developments in the CCP-EM software suite. *Acta Crystallogr. D Struct. Biol.* **73**, 469–477 (2017).
  74. Adams, P. D. et al. The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).

### Acknowledgements

This study was supported by University of Latvia Foundation grant “Bacterial microcompartments as synthetic nanoreactors” funded by SIA MikroTik and the State Research Program BIOMEDICINE. CIISB research infrastructure project LM2015043 funded by MEYS CR is gratefully acknowledged for the financial support of the measurements at the CF Cryo-electron Microscopy and Tomography CEITEC MU. We thank Dr. Janis Rumnieks for technical IT support and Dr. Reinis Rutkis for valuable discussions.

### Author contributions

G.K. designed and produced constructs, expressed and purified proteins, solved the cryo-EM structure, and wrote the paper. E.-E.C. designed and produced constructs, and expressed and purified proteins. J.L. wrote the paper. J.J. prepared negative staining TEM samples and performed TEM analysis. A.F. prepared cryo-EM samples and collected cryo-EM data. K.T. coordinated the study and wrote the paper.

### Competing interests

The authors declare no competing interests.

### Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-019-14205-y>.

Correspondence and requests for materials should be addressed to G.K.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.






**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

### **3.3 CntA oxygenase substrate profile comparison and oxygen dependency of TMA production in *Providencia rettgeri***

- **Isolation, recombinant expression and purification of four CntA oxygenases from different hosts**
- **Substrate profile comparison of CntA from different hosts**
- **Oxygen dependency tests of TMA production of *Providencia rettgeri* culture and purified CntA**

# CntA oxygenase substrate profile comparison and oxygen dependency of TMA production in *Providencia rettgeri*

Gints Kalnins<sup>1</sup>  | Eduards Sevostjanovs<sup>2</sup> | Dace Hartmane<sup>2</sup> |  
Solveiga Grinberga<sup>2</sup> | Kaspars Tars<sup>1,3</sup>

<sup>1</sup>Latvian Biomedical Research and Study Centre, Riga, Latvia

<sup>2</sup>Latvian Institute of Organic Synthesis, Riga, Latvia

<sup>3</sup>Faculty of Biology, University of Latvia, Riga, Latvia

## Correspondence

Gints Kalnins, Latvian Biomedical Research and Study Center, Ratsupites 1, LV-1067 Riga, Latvia.  
Email: gints@biomed.lu.lv

## Funding information

State Research Program Biomedicine

CntA oxygenase is a Rieske 2S-2Fe cluster-containing protein that has been previously described as able to produce trimethylamine (TMA) from carnitine, gamma-butyrobetaine, glycine betaine, and in one case, choline. TMA found in humans is exclusively of bacterial origin, and its metabolite, trimethylamine oxide (TMAO), has been associated with atherosclerosis and heart and renal failure. We isolated four different Rieske oxygenases and determined that there are no significant differences in their substrate panels. All three had high activity toward carnitine/gamma-butyrobetaine, medium activity toward glycine betaine, and very low activity toward choline. We tested the influence of low oxygen concentrations on TMA production in CntA-containing *Providencia rettgeri* cell cultures and discovered that this process, although dependent on the amount of oxygen, is still feasible in environments with 1 and 0.2% oxygen, which is comparable to oxygen levels in some parts of the digestive system.

## KEYWORDS

carnitine, CntA, Rieske oxygenase, trimethylamine

## 1 | INTRODUCTION

Iron-containing metalloenzymes are a diverse class of proteins that perform various oxidative chemical conversions. Iron-containing metalloenzymes can be divided into two groups – heme-containing enzymes and non-heme iron-dependent enzymes. Non-heme iron-dependent enzymes include Rieske oxygenases, a diverse class of two- and three-component systems that characteristically contain an iron atom in the active site and Rieske-type 2S-2Fe clusters [1]. Rieske oxygenase systems always contain oxygenase and reductase components and sometimes contain an additional ferredoxin component [2].

**Abbreviations:** FMO3, flavin-containing monooxygenase 3; TMA, trimethylamine; TMAO, trimethylamine oxide.

The first Rieske oxygenases discovered were *Pseudomonas putida* toluene dioxygenase and naphthalene dioxygenase [3,4]. Since then, numerous Rieske oxygenases that oxidize other aromatic compounds have been identified [5–8]. A common property of these enzymes is the ability to break aromatic rings, making them interesting for biotechnological applications [9,10]. The functions of Rieske oxygenases are not limited to oxidizing aromatic rings. Recently, other Rieske oxygenase systems have been identified that oxidize compounds such as cholesterol [11], amine groups of 4-aminobenzylamine [12], stacchyrine [13] and caffeine [14]. Furthermore, Rieske-type oxygenase systems are not limited to microorganisms. In nematodes and insects, DAF-36/Nvd Rieske oxygenases have been found to oxidize cholesterol [15], in plants, choline monooxygenase has been found to oxidize choline to form betaine aldehyde [16], and some mammals express



CMP-N-acetylneuraminic acid hydroxylase, also a Rieske-type oxygenase [17].

Recently, a two-component CntA/CntB oxygenase system from *Acinetobacter baumannii* has been discovered, capable of cleaving carnitine into trimethylamine (TMA) and malic semialdehyde [18]. CntA is almost exclusively found in *Gammaproteobacteria* [19]. It is related to bacterial aromatic dioxygenases and, curiously, also to plant choline monooxygenases [18]. A closely related iron-sulfur protein, YeaW from *Escherichia coli*, has been described previously, but its enzymatic activity was not determined [20]. The same protein was later reported to have TMA-producing capability, with substrates such as carnitine, gamma-butyrobetaine, glycine betaine, and choline [21]. A significant difference between CntA from *A. baumannii* and YeaW from *E. coli* is the ability to cleave choline, the latter having this property and the former lacking it. The reason for this difference is not clear and requires further investigation.

The CntA-produced metabolite TMA is of particular interest because of its possible involvement in the pathogenesis of cardiovascular diseases [22–24], insulin resistance [25], heart failure [26], and renal failure [27]. TMA in humans is exclusively of bacterial origin, since humans lack the genes to produce TMA. However, TMA can be further metabolized by hepatic monooxygenase FMO3 (flavin-containing monooxygenase 3) into odorless trimethylamine oxide (TMAO). Defective or downregulated FMO3 can cause trimethylaminouria or “fish odor syndrome” [28]. The exact mechanism by which TMA/TMAO exercises its influence is not exactly known, although there have been several suggestions: negative influence of TMAO on cholesterol efflux from foam cells [29]; influence of TMA/TMAO on FMO3, a key known regulator of cholesterol metabolism activity [30]; possible hypersensitization of platelets by TMAO [31]; and impairment of pyruvate and fatty acid oxidation in cardiac mitochondria [32]. There is also some controversy about whether TMAO is actually a cause of atherosclerosis or merely a marker of it [33,34].

Apart from CntA, there are also other enzymes that could be responsible for production of TMA in microbiota – trimethylamine reductase [35], CutC choline lyase [36–38], and ergothionase [39]. Unlike CntA, these enzymes do not need oxygen and therefore could function in anaerobic environments, including the digestive system. However, whether oxygen-consuming CntA can actually contribute to TMA production in humans is unclear.

In this study, we investigated and compared the substrate specificity of CntA oxygenases from different organisms and determined oxygen requirements for aerobic carnitine degradation by TMA-producing *Providencia rettgeri* [40].

## 2 | MATERIALS AND METHODS

### 2.1 | Plasmid construction and protein expression

*E. coli* DE3 cells were purchased from Sigma. *Klebsiella pneumoniae* MSCL535, *Serratia marcescens* MSCL1476, and *Providencia rettgeri* MSCL730 cultures were obtained from The Microbial Strain Collection of Latvia. Genomic DNA was extracted from cells by proteinase K treatment and ethanol precipitation. For primer design, CntA gene sequences were obtained by searching for *Acinetobacter baumannii* CntA [18] homologs in respective organisms (Supporting Information Table S1). DNA fragments were amplified with PCR and subcloned into the pRSF-TEV Duet vector, containing an N-terminal hexahistidine tag and a tobacco etch virus protease cleavage site. Constructs were sequenced for confirmation.

Proteins were expressed in *E. coli* BL21-DE3 cells. Cells were grown at 37 °C with shaking until OD<sub>590</sub> reached 0.5, and then the flask was chilled at 20 °C for 30 min. Protein expression was induced with 1 mM IPTG, and cells were grown overnight at 20 °C. Cells were collected with centrifugation at 6000×g for 5 min and stored at –20 °C.

### 2.2 | Protein purification

Frozen cell paste was suspended in lysis buffer (40 mM Tris-HCl (pH 8.0), 200 mM NaCl, 1% Triton X-100, 20 mM MgSO<sub>4</sub>, 1 mM PMSF, and 2 mM DTT). Cells were lysed by ultrasound for 5 min using an S200 ultrasound generator on the 0.5 s pulsation setting. The lysate was cleared by centrifugation at 14,000×g for 40 min. CntA and CntB proteins were further purified by nickel affinity chromatography using 1-ml or 5-ml HisTrap columns (GE Healthcare). For this step, 20 mM imidazole in 40 mM Tris-HCl (pH 8.0) and 300 mM NaCl was used as a washing buffer, and 300 mM imidazole in 40 mM Tris-HCl (pH 8.0) and 300 mM NaCl was used as an elution buffer. Eluted proteins were further purified by gel filtration on a Superdex 200 16/600 column in 20 mM Tris-HCl (pH 8.0). Brown-colored fractions containing CntA or CntB proteins were pooled, concentrated to 10 mg ml<sup>-1</sup>, flash-frozen in liquid nitrogen and stored at –20 °C.

### 2.3 | *In vitro* CntA oxygenase activity assay

Activity assays were performed in 20 mM Tris-HCl and 50 mM NaCl in a volume of 400 µl, with CntA and CntB protein concentrations of 0.1 mg ml<sup>-1</sup>. Potential substrates – L-carnitine (Acros), choline (Sigma-Aldrich), gamma butyrobetaine (Sigma-Aldrich), and glycine betaine (Sigma-Aldrich) – were added to concentrations of 2 mM.



Two approaches were used for reducing the CntA Rieske cluster. In the first approach, *P. rettgeri* CntB oxidoreductase and 0.5 mM NADH were added to the reaction mixture. In the second approach, the CntA Rieske cluster was reduced directly with 0.5 mM sodium dithionite. After adding all components, reaction mixtures were sealed in 0.5-ml microtubes and incubated for 18 h at room temperature. Reactions were stopped by adding formic acid to a final concentration of 5%. Samples were stored at  $-20^{\circ}\text{C}$ .

For experiments with different oxygen amounts, mixtures were made as described above, with the exception of the *P. rettgeri* CntA protein concentration, which was lowered to  $0.01\text{ mg ml}^{-1}$ . All procedures were performed in a sealable glove bag (Sigma), and all flasks, tubes and solutions were equilibrated in 99.99% argon gas (AGA, Estonia). Solution containing CntA oxygenase was equilibrated in an argon stream for 15 min, and 0.5 M sodium dithionite and 1 M L-carnitine solutions in water were added to final concentrations of 0.5 and 2 mM, respectively. Five milliliters of the mixture was put into open tubes that were placed in 0.5 L flasks. Outside air was added via a tube connected to the flask until concentrations reached 0.2, 1, 5, or 21%. For experiments with pure argon, gas was added to the flasks until oxygen values dropped to undetectable amounts. A Vernier  $\text{O}_2$  Gas Sensor was used for monitoring oxygen levels. Flasks were then sealed and incubated for 18 h at room temperature. Reactions were stopped by adding formic acid to a concentration of 5%. Samples were stored at  $-20^{\circ}\text{C}$ . TMA was quantified as previously described [37].

#### 2.4 | Growth of *P. rettgeri* and *E. coli* at various oxygen concentrations

*P. rettgeri* and *E. coli* BL21-DE3 were grown in 2xTY medium, with 0.1% carnitine, betaine, choline, or gamma-butyrobetaine as additives. For negative controls, bacteria were grown in 2xTY medium without additives. For experiments with *E. coli* BL21-DE3 transformed with pRSF-CntA coli plasmid, cells were grown in 2xTY medium containing  $30\text{ }\mu\text{g ml}^{-1}$  kanamycin and with or without 1 mM IPTG. Cells were grown in 1 ml volumes in 3.5-ml sterile tubes that were placed in 0.5-L flasks. For experiments under hypoxic conditions, flasks were equilibrated with 99.99% argon 4.0 (AGA) in a sealable glove bag (Sigma). Oxygen concentrations were monitored with a Vernier  $\text{O}_2$  Gas Sensor. After reaching undetectable oxygen amounts, 20  $\mu\text{l}$  of seeding stock was added to 2xTY medium containing either 0.1% carnitine or no additives. Outside air was then slowly added to flasks via a tube until desired oxygen concentrations of 0.2, 1, or 5% were reached, except for samples that were left in pure argon. Flasks were then immediately sealed and incubated for 24 h at  $30^{\circ}\text{C}$ . After incubation, formic acid was immediately added to 5% final concentration to stop TMA production, and

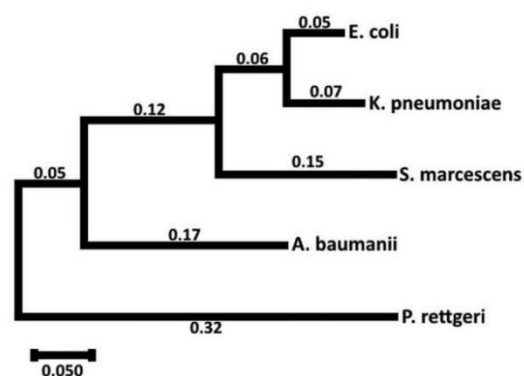
cultures were centrifuged for 1 min at  $15,000\times g$  to remove cells. Supernatant was then collected, transferred to fresh tubes, sealed and stored at  $-20^{\circ}\text{C}$ . TMA was quantified as previously described [37].

### 3 | RESULTS

#### 3.1 | Substrate profiles of different CntA oxygenases

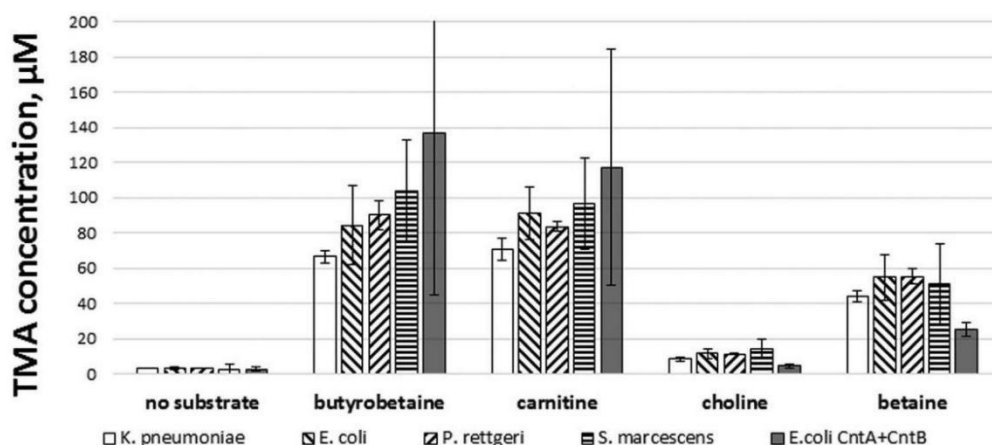
We isolated, expressed, and purified four CntA oxygenases from different microorganisms – *E. coli* BL21-DE3, *S. marcescens*, *P. rettgeri*, and *K. pneumoniae*. The latter three have been reported to be able to produce TMA from carnitine [41–42]. In the case of *E. coli*, the SE11 strain has been reported to be able to produce TMA from carnitine [18], but the BL21-DE3 strain has not yet been characterized. Nevertheless, we were able to isolate CntA and CntB from *E. coli* BL21-DE3 as well.

When compared using the maximum likelihood method, CntA oxygenases from *K. pneumoniae* and *E. coli* are very similar (88%), but *A. baumannii* and *S. marcescens* CntA are no more than 75% similar to one another, and *P. rettgeri* CntA is only 46% similar to its closest relative (Fig. 1). Nevertheless, when we compared the activities of all four oxygenases with four different substrates, there were no significant differences between the substrate profiles of different oxygenases (Fig. 2). The highest TMA production was observed from carnitine and gamma-butyrobetaine (with insignificant differences between them), followed by glycine betaine with 40–60% of the production levels compared to carnitine and gamma-butyrobetaine. TMA production from choline was very low, at least 10-fold lower than that of carnitine and gamma-butyrobetaine and at least fivefold



**FIGURE 1** Maximum likelihood tree of CntA oxygenase amino acid sequences from *E. coli*, *K. pneumoniae*, *S. marcescens*, *A. baumannii*, and *P. rettgeri*. The tree was generated with the MEGA7 program





**FIGURE 2** Substrate profile comparison of different CntA oxygenases. Experiments were performed in aerobic conditions (21% O<sub>2</sub>). Each result is average value of three independent experiments. Error bars indicate standard deviations

lower than that of betaine. Therefore, these results were more in accordance with the study from Zhu et al. [18] than with the study from Koeth et al. [21], the latter showing much higher TMA-production from choline. To exclude the influence of different activity assay methods, we isolated, expressed and purified CntB component from *E. coli* and tested its TMA producing activity together with CntA (Fig. 2). The detected TMA-production activities were not different when compared with results using sodium dithionite method. We found equally high activities for carnitine and gamma-butyrobetaine, lower activity for betaine and very little activity for choline. These results indicate that the presence of the reductase CntB component does not influence CntA substrate specificity.

### 3.2 | Production of TMA from *E. coli* and *P. rettgeri* cultures

To determine if the CntA gene-containing bacteria used in our study are capable of producing TMA, we grew *E. coli* BL21-DE3 and *P. rettgeri* cultures in 2xTY medium containing 0.1% carnitine, gamma-butyrobetaine, glycine betaine, and choline (Table 1). As expected, *P. rettgeri* was capable of producing TMA, but only from carnitine and

gamma-butyrobetaine. Unexpectedly, *E. coli* BL21-DE3 was unable to produce TMA in detectable amounts, despite the fact that it contains a fully functional CntA oxygenase system. When *E. coli* BL21-DE3 cells were transformed with pRSF plasmid containing CntA gene from *E. coli*, the ability to produce TMA from carnitine was observed and it was enhanced by addition of IPTG. Low expression of CntA and/or CntB genes in *E. coli* BL21-DE3 cells is probably the reason for this observation. These results indicate that the ability to produce TMA may be strain-specific and it is possible for certain bacterial strains to have a functional CntA gene with very different expression levels.

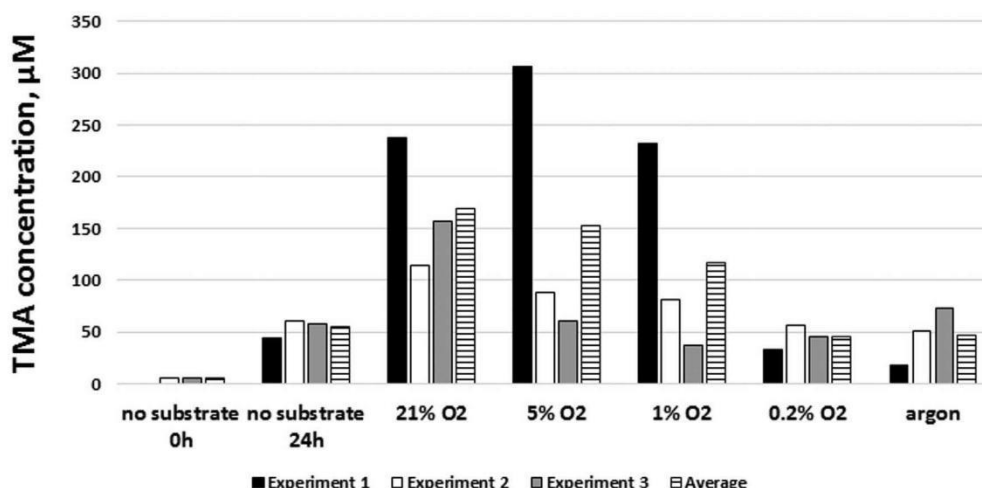
### 3.3 | Production of TMA *in vivo* and *in vitro* at low oxygen levels

To determine the exact oxygen requirements for TMA production in *P. rettgeri* cultures in hypoxic conditions, we grew *P. rettgeri* in rich 2xTY medium containing additional carnitine for 24 h in different atmospheres (Fig. 3). We considered rich medium to be more appropriate since the human gut is also rich in nutrients. Despite the large scattering of our data, a general trend of decreasing amounts of TMA produced in decreasing oxygen concentrations can be noticed.

**TABLE 1** TMA production in *P. rettgeri* and *E. coli* BL21-DE3 cell cultures in aerobic conditions

	Carnitine	Butyrobetaine	Choline	Betaine
<i>P. rettgeri</i>	158.15	212.35	ND	ND
<i>E. coli</i> BL21-DE3	ND	ND	ND	ND
<i>E. coli</i> BL21-DE3 + pRSF_coliCntA	13.49			
<i>E. coli</i> BL21-DE3 + pRSF_coliCntA + IPTG	28.99			

Values indicate detected TMA amounts (µM) in medium after 24 h incubation. ND, not detected.



**FIGURE 3** TMA production from carnitine in *P. rettgeri* cultures in atmospheres with different oxygen concentrations. Controls without substrate were incubated in 21% oxygen environment

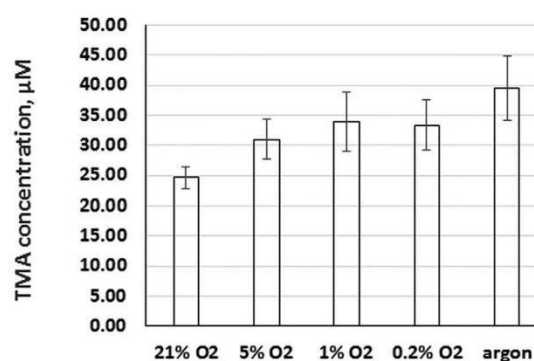
Our data also show that TMA can be produced at even 0.2% oxygen concentration and 99.99% argon. Interestingly, we detected some TMA production in negative controls incubated for 24 h with no carnitine substrate at all. This is most likely due to small amounts of carnitine in yeast extract. In general, our data show that TMA production from carnitine in *P. rettgeri* is dependent on oxygen to a certain level, but even under very anoxic conditions, it is still possible to detect TMA.

To test the influence of low levels of oxygen on *P. rettgeri* CntA oxygenase activity, we performed *in vitro* activity assays in different atmospheres as before with cell cultures (Fig. 4). Our findings unexpectedly showed that CntA is not only active in 99.99% argon, but the catalysis trends towards higher efficiency in more anaerobic environments. This is probably due to the fact that sodium dithionite in water solution slowly combusts and is more stable in environments with lower amounts of oxygen. Therefore, CntA Rieske 2Fe-2S cluster is reduced more efficiently and the observed enzymatic activity is higher. Nevertheless, we can conclude from these data that in principle low oxygen concentrations are not a major obstacle to CntA-mediated catalysis.

## 4 | DISCUSSION

Rieske oxygenases typically have a wide range of substrates, for example, more than 100 substrates have been discovered for naphthalene 1,2-dioxygenase [43]. For CntA oxygenase, four possible substrates have been described – carnitine, gamma-butyrobetaine, glycine betaine, and choline. In a study performed by Zhu et al. [18], it was shown that the CntA

oxygenase system from *A. baumannii* can produce TMA from carnitine, gamma-butyrobetaine and, to a lesser extent, glycine betaine. These results were in agreement with a later study on YeaW oxygenase from *E. coli* [21]. The crucial difference between these two studies is that the former study showed non-detectable activity using choline as a substrate, while the latter study showed the largest activity of all tested substrates for choline. The simplest explanation for such differences would be both enzymes having different substrate profiles – a potential option, considering the diversity of Rieske oxygenases and close CntA relation to plant choline monooxygenases [18]. However, substrate profiles for all



**FIGURE 4** *P. rettgeri* CntA *in vitro* activity in atmospheres with different oxygen concentrations. Each result is medium of three independent experiments. Error bars indicate standard deviation. Carnitine was used as substrate



tested CntA oxygenases, including CntA from *E. coli* (identical to YeaW), were similar among themselves and more consistent with data from experiments with *A. baumannii* CntA oxygenase: high activity for carnitine/butyrobetaine, medium for glycine betaine, very small or non-detectable for choline. Structural data from another C-N bond-cleaving Rieske-type oxygenase, stacchydriine demethylase, shows that the substrate is bound to active site iron atom via a carboxyl group [13]. Since choline has no carboxyl group, this suggests that it could bind to active sites much less efficiently.

The human digestive system itself is a diverse environment, containing gradients of different pH, nutrient concentrations, and oxygen levels. There is a general understanding about the presence of anaerobic conditions in digestive tract lumen, but direct measurements have been performed in only a few studies. Measurements with electrodes and phosphorescence quenching methods have suggested that the amount of oxygen in intestinal and rectal lumens is less than 0.5–1 Torr [44–46]. Methods based on electron paramagnetic resonance have shown larger values – from 32 Torr in the duodenum to 3 Torr in the sigmoidal-rectal junction [47]. This oxygen in digestive tract comes from both diffusion across serosa [46,47] and swallowed air [48]. High affinity cytochrome bd oxygenases are essential to *E. coli* for intestine colonization, indicating the importance of aerobic respiration for gut microflora [49]. *E. coli* can also consume oxygen even at 3 nM concentrations, which is comparable to or even lower than concentrations detected in digestive systems [50]. TMA production from carnitine by bacteria in digestive systems has been clearly demonstrated several times [22,42,51], but the only known enzyme to produce TMA from carnitine is aerobic CntA oxygenase. This raises the question of whether there is enough oxygen in the gut for this process and if not, whether there is another undiscovered TMA-producing enzyme. To determine whether oxygen availability is indeed a limiting factor for TMA production from carnitine, we tested the capability of *P. rettgeri* to produce TMA in various oxygen environments – starting with air (21%), normal human tissue concentration (5%), higher oxygen concentrations from gut environment (1%), lower oxygen concentrations from gut environment (0.2%), and argon gas (less than 0.01%). Our data showed that TMA production is dependent on oxygen concentration, indicating that it is performed by an aerobic enzyme. However, this process is still possible in oxygen levels of 1% (7.6 Torr) and 0.2% (1.5 Torr) that are comparable to the digestive tract environment. Moreover, we found that the CntA oxygenase itself *in vitro* can be fully active in the same conditions (Fig. 4). These data certainly imply that CntA oxygenase, despite being an aerobic enzyme, could still be responsible for carnitine degradation in the human digestive system. More detailed studies of CntA kinetics are required to confirm this hypothesis.

## ACKNOWLEDGMENT

This study was supported by the State Research Program Biomedicine.

## CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interests.

## ORCID

Gints Kalnins  <http://orcid.org/0000-0002-9593-6308>

## REFERENCES

- [1] Barry SM, Challis GL. Mechanism and catalytic diversity of Rieske non-heme iron-dependent oxygenases. *ACS Catal* 2013;3: 2362–70.
- [2] Nam JW, Nojiri H, Yoshida T, Habe H. New classification system for oxygenase components involved in ring-hydroxylating oxygenations. *Biosci Biotechnol Biochem* 2001;65:254–63.
- [3] Yeh WK, Gibson DT, Liu TN. Toluene dioxygenase: a multicomponent enzyme system. *Biochem Biophys Res Commun* 1977;78:401–10.
- [4] Ensley BD, Gibson DT, Laborde AL. Oxidation of naphthalene by a multicomponent enzyme system from *Pseudomonas* sp. strain NCIB 9816. *J Bacteriol* 1982;149:948–54.
- [5] Btinz PV, Cook AM. Dibenzofuran 4,4a-dioxygenase from *Sphingomonas* sp. strain RW1: angular dioxygenation by a three-component enzyme system *J Bacteriol* 1993;175:6467–75.
- [6] Bundy BM, Campbell AL, Neidle EL. Similarities between the antABC-encoded anthranilate dioxygenase and the benABC-encoded benzoate dioxygenase of *Acinetobacter* sp. strain ADP1. *J Bacteriol* 1998;180:4466–74.
- [7] Shepherd JM, Lloyd-Jones G. Novel carbazole degradation genes of *Sphingomonas* CB3: sequence analysis, transcription, and molecular ecology. *Biochem Biophys Res Commun* 1998;247:129–35.
- [8] Dehmel U, Engesser KH, Timmis KN, Dwyer DF. Cloning, nucleotide sequence, and expression of the gene encoding a novel dioxygenase involved in metabolism of carboxydiphenyl ethers in *Pseudomonas pseudoalcaligenes* POB310. *Arch Microbiol* 1995;163:35–41.
- [9] Peng RH, Xiong AS, Xue Y, Fu XY. A profile of ring-hydroxylating oxygenases that degrade aromatic pollutants. *Rev Environ Contam Toxicol* 2010;206:65–94.
- [10] Gally C, Nestl BM, Hauer B. Engineering Rieske non-heme iron oxygenases for the asymmetric dihydroxylation of alkenes. *Angew Chem Int Ed Engl* 2015;54:12952–6.
- [11] Van der Geize R, Yam K, Heuser T, Wilbrink MH. A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages. *Proc Natl Acad Sci USA* 2007;104:1947–52.
- [12] Lee J, Simurdiak M, Zhao H. Reconstitution and characterization of aminopyrrolnitrin oxygenase, a Rieske N-oxygenase that catalyzes unusual arylamine oxidation. *J Biol Chem* 2005;280: 36719–27.



- [13] Daughtry KD, Xiao Y, Stoner-Ma D, Cho E. Quaternary ammonium oxidative demethylation: X-ray crystallographic, resonance Raman, and UV-visible spectroscopic analysis of a Rieske-type demethylase. *J Am Chem Soc* 2012;134:2823–34.
- [14] Summers RM, Louie TM, Yu CL, Subramanian M. Characterization of a broad specificity non-haem iron N-demethylase from *Pseudomonas putida* CBB5 capable of utilizing several purine alkaloids as sole carbon and nitrogen source. *Microbiology* 2011;157:583–92.
- [15] Yoshiyama-Yanagawa T, Enya S, Shimada-Niwa Y, Yaguchi S. The conserved Rieske oxygenase DAF-36/Neverland is a novel cholesterol-metabolizing enzyme. *J Biol Chem* 2011;286:25756–62.
- [16] Rathinasabapathi B, Burnet M, Russell BL, Gage DA. Choline monooxygenase, an unusual iron-sulfur enzyme catalyzing the first step of glycine betaine synthesis in plants: prosthetic group characterization and cDNA cloning. *Proc Natl Acad Sci USA* 1997;94:3454–8.
- [17] Schlenzka W, Shaw L, Kelm S, Schmidt CL. CMP-N-acetylneuraminic acid hydroxylase: the first cytosolic Rieske iron-sulphur protein to be described in Eukarya. *FEBS Lett* 1996;385:197–200.
- [18] Zhu Y, Jameson E, Crosatti M, Schäfer H. Carnitine metabolism to trimethylamine by an unusual Rieske-type oxygenase from human microbiota. *Proc Natl Acad Sci USA* 2014;111:4268–73.
- [19] Rath S, Heidrich B, Pieper DH, Vital M. Uncovering the trimethylamine-producing bacteria of the human gut microbiota. *Microbiome* 2017;5:54.
- [20] Boxhammer S, Glaser S, Kithl A, Wagner AK. Characterization of the recombinant Rieske [2Fe-2S] proteins HcaC and YeaW from *E. coli*. *Biomaterials* 2008;21:459–67.
- [21] Koeth RA, Levison BS, Culley MK, Buffa JA.  $\gamma$ -Butyrobetaine is a proatherogenic intermediate in gut microbial metabolism of L-carnitine to TMAO. *Cell Metab* 2014;20:799–812.
- [22] Wang Z, Klipfell E, Bennett BJ, Koeth R. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 2011;472:57–63.
- [23] Koeth RA, Wang Z, Levison BS, Buffa JA. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat Med* 2013;19:576–85.
- [24] Tang WH, Wang Z, Levison BS, Koeth RA. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med* 2013;368:1575–84.
- [25] Gao X, Liu X, Xu J, Xue C. Dietary trimethylamine N-oxide exacerbates impaired glucose tolerance in mice fed a high fat diet. *Biosci Bioeng* 2014;118:476–81.
- [26] Trøseid M, Ueland T, Hov JR, Svardal A. Microbiota-dependent metabolite trimethylamine-N-oxide is associated with disease severity and survival of patients with chronic heart failure. *J Intern Med* 2015;277:717–26.
- [27] Tang WH, Wang Z, Kennedy DJ, Wu Y. Gut microbiota-dependent trimethylamine N-oxide (TMAO) pathway contributes to both development of renal insufficiency and mortality risk in chronic kidney disease. *Circ Res* 2015;116:448–55.
- [28] Treacy EP, Akerman BR, Chow LM, Youil R. Mutations of the flavin-containing monooxygenase gene (FMO3) cause trimethylaminuria, a defect in detoxication. *Hum Mol Genet* 1998;7:839–45.
- [29] Koeth RA, Wang Z, Levison BS, Buffa JA. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat Med* 2013;19:576–85.
- [30] Warriar M, Shih DM, Burrows AC, Ferguson D. The TMAO-generating enzyme flavin monooxygenase 3 is a central regulator of cholesterol balance. *Cell Rep* 2015;10:326–38.
- [31] Zhu W, Gregory JC, Org E, Buffa JA. Gut microbial metabolite TMAO enhances platelet hyperreactivity and thrombosis risk. *Cell* 2016;165:111–24.
- [32] Makrecka-Kuka M, Volska K, Antone U, Vilskersts R. Trimethylamine N-oxide impairs pyruvate and fatty acid oxidation in cardiac mitochondria. *Toxicol Lett* 2017;267:32–38.
- [33] Cho CE, Caudill MA. Trimethylamine-N-oxide: friend, foe, or simply caught in the cross-fire? *Trends Endocrinol Metab* 2017;28:121–30.
- [34] Fennema D, Phillips IR, Shephard EA. Trimethylamine and trimethylamine N-oxide, a flavin-containing monooxygenase 3 (FMO3)-mediated host-microbiome metabolic axis implicated in health and disease. *Drug Metab Dispos* 2016;44:1839–50.
- [35] Kwan HS, Barrett EL. Purification and properties of trimethylamine oxide reductase from *Salmonella typhimurium*. *J Bacteriol* 1983;155:1455–8.
- [36] Craciun S, Marks JA, Balskus EP. Characterization of choline trimethylamine-lyase expands the chemistry of glyceryl radical enzymes. *ACS Chem Biol* 2014;9:1408–13.
- [37] Kalnins G, Kuka J, Grinberga S, Makrecka-Kuka M. Structure and function of CutC choline lyase from human microbiota bacterium *Klebsiella pneumoniae*. *J Biol Chem* 2015;290:21732–40.
- [38] Bodea S, Funk MA, Balskus EP, Drennan CL. Molecular basis of C-N bond cleavage by the glyceryl radical enzyme choline trimethylamine-lyase. *Cell Chem Biol* 2016;23:1206–16.
- [39] Muramatsu H, Matsuo H, Okada N, Ueda M. Characterization of ergothioneine from *Burkholderia* sp. HME13 and its application to enzymatic quantification of ergothioneine. *Appl Microbiol Biotechnol* 2013;97:5389–400.
- [40] Ellsner T, Preusser A, Wagner U, Kleber HP. Metabolism of L(-)-carnitine by *Enterobacteriaceae* under aerobic conditions. *FEMS Microbiol Lett* 1999;174:295–301.
- [41] Unemoto T, Hayashi M, Miyaki K, Hayashi M. Formation of trimethylamine from DL-carnitine by *Serratia marcescens*. *Biochim Biophys Acta* 1966;121:220–2.
- [42] Kuka J, Liepinsh E, Makrecka-Kuka M, Liepins J. Suppression of intestinal microbiota-dependent production of pro-atherogenic trimethylamine N-oxide by shifting L-carnitine microbial degradation. *Life Sci* 2014;117:84–92.
- [43] Escalante DE, Aukema KG, Wackett LP, Aksan A. Simulation of the bottleneck controlling access into a Rieske active site: predicting substrates of naphthalene 1,2-dioxygenase. *J Chem Inf Model* 2017;57:550–61.
- [44] Crompton DW, Shrimpton DH, Silver IA. Measurements of the oxygen tension in the lumen of the small intestine of the domestic duck. *J Exp Biol* 1965;43:473–8.
- [45] Lind Due V, Bonde J, Kann T, Perner A. Extremely low oxygen tension in the rectal lumen of human subjects. *Acta Anaesthesiol Scand* 2003;47:372.
- [46] Albenberg L, Esipova TV, Judge CP, Bittinger K. Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota. *Gastroenterology* 2014;147:1055–63.

- [47] He G, Shankar RA, Chzhan M, Samouilov A. Noninvasive measurement of anatomic structure and intraluminal oxygenation in the gastrointestinal tract of living mice with spatial and spectral EPR imaging. *Proc Natl Acad Sci USA* 1999;96:4586–91.
- [48] Levitt MD, Bond JH. Flatulence. *Annu Rev Med* 1980;31:127–37.
- [49] Jones SA, Chowdhury FZ, Fabich AJ, Anderson A. Respiration of *Escherichia coli* in the mouse intestine. *Infect Immun* 2007;75:4891–9.
- [50] Stolper DA, Revsbech NP, Canfield DE. Aerobic growth at nanomolar oxygen concentrations. *Proc Natl Acad Sci USA* 2010;107:18755–60.
- [51] Weinert CH, Empl MT, Krüger R, Frommherz L. The influence of a chronic L-carnitine administration on the plasma metabolome of male Fischer 344 rats. *Mol Nutr Food Res* 2017;61:1600651–62.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Kalnins G, Sevostjanovs E, Hartmane D, Grinberga S, Tars K. CntA oxygenase substrate profile comparison and oxygen dependency of TMA production in *Providencia rettgeri*. *J Basic Microbiol.* 2017;1–8.  
<https://doi.org/10.1002/jobm.201700428>

## 4. DISCUSSION

### 4.1 Structure of CutC choline lyase

Glycyl radical enzyme CutC choline lyase is a significant contributor to the TMA pool in mammal organism (Roberts et. al., 2018). Since inhibition of this enzyme has potential therapeutic and preventive effect, structural characterization of this enzyme could be useful for rational drug design. As both our results and later structural studies of GRM1 type CutC from *Desulfovibrio alaskensis* (Bodea et. al., 2016) has demonstrated, the choline molecule is located in a very tightly fitting pocket. A rational drug design benefits from large-sized active sites where a large variety of chemical compounds can easily be localized; however, for CutC such opportunities are therefore few. Still, some promising inhibitors has nevertheless been identified – 3,3-dimethylbutanol (Wang et. al., 2015), halomethylcholine and halomethylbetaine analogues (Roberts, et. al. 2018) and betaine aldehyde (Orman et. al., 2019). These inhibitors are all highly similar to choline, having only small additional modifications. Iodomethylcholine even could be an irreversible inhibitor of CutC, binding covalently to a nucleophilic residue in the active site (Roberts, et. al. 2018). This illustrates that there could be some future prospects designing synthetic ligands targeted to the CutC active site.

The conformational shift driven by the substrate binding discovered in *Klebsiella pneumoniae* CutC is unique among GRE. A very similar shift was noticed for benzylsuccinate synthase, but in this particular case the switching from open to closed form was driven by the binding of accessory subunits (Funk et. al., 2014). Curiously, this was not the case for GRM1 type CutC (Bodea et. al., 2016) – both choline-bound and choline-free mutant forms are matching the GRM2 type choline bound closed structure. In that case it was actually impossible to purify a native choline-free GRM1 type CutC since the substrate was collected by the enzyme from the cell environment. The role of the GRM2 type CutC conformational shift is not clear – it could be associated with making the GRE active site accessible to GRE-AE or it could be also a mechanism for increasing the affinity for substrate. It could be possible that advantage could be taken of this conformational shift – by designing an appropriate compound binding to the open, choline free CutC form it could be possible to lock the whole enzyme in a permanent open state.



## 4.2 CutC-associated GRM2 type bacterial microcompartment structure and encapsulation mechanisms

Bacterial microcompartments are proteinaceous self-assembling organelles that are widespread among prokaryotes. These organelles are built with a specific purpose of enclosing enzymatic pathways and increasing their efficiency and protecting the cell from toxic intermediates. Since CutC is a BMC-associated enzyme, this supramolecular assembly itself is a critical factor for functioning of this enzyme in bacteria.

There are two general ways the assembly process is thought to occur: core-first or concomitant assembly (Kerfeld et. al., 2018). In a core-first mechanism the enzymatic core is built first and the shell is then formed around it. In concomitant assembly, the shell and the core are formed simultaneously. A core-first mechanism seems to be characteristic to  $\beta$ -carboxysomes and concomitant assembly to  $\alpha$ -carboxysomes. A computational study has suggested that these building modes are dependent on the binding efficiency of components – weak core interactions result in concomitant assembly and strong core interactions result in a core-first assembly (Perlmutter, et. al., 2016). The properties of shell also have to be noted – different spontaneous curvature of the shell components could result in different BMC particle sizes and different building modes (Mohajerani & Hagan, 2018).

Our data shows that even highly similar BMC-H proteins can have different shell-forming properties. *cmcA* and *cmcB*, having properties matching low spontaneous curvature components are unable to form any particles as such or together with pentameric subunits. *cmcC*<sup>Δ</sup>, *cmcC*<sub>trunc</sub>, *cmcE* and, to a lesser extent, native *cmcC* would correspond to high spontaneous curvature components. It is possible that some synergic effects could also be involved in the determination of curvature – *cmcAB+D* are able to form small type particles in contrast to individual *cmcA+D* and *cmcB+D* unable to do so. We propose that a large type BDP shell, up to 200-300 nm in size, is a balanced combination of components with both high and low spontaneous curvature properties and that their proportion is a critical factor for the particle size.

Our results also show that in the GRM2 core encapsulation process CutC is most likely playing a central adaptor role. It turned out that the two domains of CutC have different functions – while the N-terminal extension of 336 residues were responsible for binding of CutF aldehyde dehydrogenase and increasing the resulting complex size, the C-terminal part, in addition to its catalytic functions, was responsible for encapsulation of it in the BDPs and binding of CutO alcohol dehydrogenase. Interestingly, CutF still contains an EP-like sequence at the C-terminal end and CutH has an analogical sequence at the N-terminal end,

but their functions in our recombinant BDP system, if any, are limited purely to crosslinking. This illustrates that the core is built in a hierarchical manner. For proper recycling of NAD<sup>+</sup>/NADH it is necessary to ensure that the aldehyde intermediate is more or less equally divided between alcohol and aldehyde dehydrogenases, and locating both enzymes in a close proximity to the signature enzyme could be a good way how to achieve this.

Data about the assembly of the BDP shell is consistent with previous studies. These particles are formed similarly to previously reported larger pT=9 pseudosymmetric particles from *Haliangium ochraceum* (Sutter et. al., 2017), pT=4 particles comprised of double-fused *Haliangium ochraceum* BMC-H proteins (Sutter et. al., 2019a), mixed icosahedral and elongated pT=4 and pT=3 particles from *Halotheca sp.* (Sutter et. al., 2019b) and smaller pT=1 particles comprised of circularly permuted BMC-H proteins (Jorda et. al., 2016) – in all cases convex side of BMC-H and BMC-P proteins are directed towards lumen.

### **4.3 CntA oxygenase substrate profile and oxygen dependency**

TMA production from carnitine by the gut microflora has been convincingly demonstrated several times (Wang et. a., 2011; Kuka et. al., 2014; Weinert et. al., 2017), but the identity of the responsible bacterial enzyme is less clear. CntA Rieske oxygenase is a Rieske 2Fe-2S cluster-containing enzyme capable of producing TMA from carnitine, and so far it is the best candidate for this process.

*In vitro* CntA characterization has been scarce so far, with only two studies investigating the substrate panel (Zhu et. al, 2014; Koeth et. al., 2014). While these two studies are in agreement on a high TMA-producing capability for carnitine and gamma-butyrobetaine, and a somewhat lower activity for glycine betaine, the results for choline were markedly different. Study by Zhu (2014) reported no detected activity for choline at all, but study by Koeth (2014) reported a higher TMA producing activity from choline than from carnitine. We investigated CntA oxygenase activities from four different organisms and concluded that our results are in a better agreement with Zhu (2014) study, since our results demonstrated a high activity for carnitine and gamma-butyrobetaine, medium for betaine and very low for choline. The CntA activity in bacteria most likely is targeted towards carnitine and gamma-butyrobetaine, since *Providencia rettgeri* culture was unable to produce any TMA in aerobic conditions from choline or glycine betaine.

Since CntA requires oxygen for proper functioning and the gastrointestinal tract is mostly anaerobic, we tested the capability of *Providencia rettgeri* cultures and purified CntA enzyme to produce TMA in different oxygen concentrations. We discovered that there is a

trend for lower TMA amounts produced in decreasing oxygen concentrations in the presence of carnitine, but even at very low oxygen concentrations comparable to ones in gastrointestinal tract – 1%, 0.2% and <0.01% – it was still possible to detect some TMA production. This suggests that the oxygen availability as such in gastrointestinal system is not a principal obstacle for CntA and it could indeed contribute to a degree to TMA production from carnitine and gamma-butyrobetaine in the intestines. This does not exclude the possibility that this contribution is small or that there are other pathways involved in bacterial carnitine degradation – a recent study, actually, has found no correlation of fecal CntA gene abundance with TMAO production in mice (Wu et. al., 2018).



## 5. CONCLUSIONS

1. GRM2 type CutC has a typical GRE fold of 10- $\beta/\alpha$  barrel.
2. GRM2 type CutC choline lyase exists in two different conformations: more structured, choline bound form and more disordered, choline free form.
3. The conformational CutC differences between choline bound and choline free states exist in solution as well as in crystal.
4. The minimal gene set required for production of recombinant GRM2 type BDPs in *Escherichia coli* expression system is cmcC and cmcD.
5. Adding C-terminally elongated BMC-H proteins cmcC<sup>+</sup> and cmcE and C-terminally truncated cmcC<sub>trunc</sub> to the BDP shell gene set can cause formation of larger BDPs.
6. CutC is the only GRM2 core enzyme capable of encapsulation in BDPs in detectable amounts as such and it mediates the encapsulation of CutO alcohol dehydrogenase and, possibly, also CutF aldehyde dehydrogenase.
7. Small type, 25 nm large cmcABC<sup>+</sup>+D+CutC<sub>336-1128</sub> GRM2 BDPs are pT=4 quasi symmetric.
8. Rieske type CntA oxygenase has high specific TMA-producing activity for carnitine and GBB, medium for betaine and very low for choline.
9. Availability of oxygen has a significant impact on the TMA production by *Providencia rettgeri* in the presence of carnitine, but this process could still be feasible at oxygen concentrations comparable to ones in gastrointestinal tract.

## 6. MAIN THESIS OF DEFENSE

1. CutC choline lyase have a typical GRE fold of 10- $\beta/\alpha$  barrel and binding of choline to its active site triggers conformational shift from open to closed state.
2. GRM2 signature enzyme CutC plays a central adaptor role in the encapsulation of GRM2 type enzymatic BMC core and the BMC-H proteins in the GRM2 locus have specialized roles in the formation of BMC shell.
3. Rieske type non-heme CntA oxygenase can produce TMA with high efficiency from carnitine and gamma-butyrobetaine, with medium efficiency from betaine and in with very low efficiency from choline and, despite low oxygen concentrations in gastrointestinal system, CntA could nevertheless contribute to a certain amount to TMA production in mammalian intestines.

## 7. PUBLICATIONS

I. 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*. *J Biol Chem.* **290**: 21732-40.

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

III. Kalnins G, Sevostjanovs E, Hartmane D, Grinberga S, Tars K. (2018). CntA oxygenase substrate profile comparison and oxygen dependency of TMA production in *Providencia rettgeri*. *J Basic Microbiol* **58**: 52-59.



## 8. APPROBATION OF RESEARCH

1. Kalnins G. Encapsulation mechanisms and structural studies of GRM2 bacterial microcompartment particles. FEBS3+ conference. Oral presentation. Riga, Latvia, 2019.

2. Kalnins G., E.E. Cesle, Liepins J., Rutkis R., Filimonenko A. and Tars K. Encapsulation mechanisms and structural studies of GRM2 bacterial microcompartment particles. Instruct Biennial Structural Biology Conference “Structural Biology: Deeper into the Cell”. Poster presentation. Madrid, Spain, 2019.

3. Česle EE, Kalniņš G. *Klebsiella pneumoniae* bakteriālo mikrokompartmentu pētījumi. Molekulārās bioloģijas sekcija LU 76. konferences molekulārās bioloģijas sekcija un Latvijas Bioķīmijas biedrības 3. konference. Mutisks referāts. Rīga, Latvija, 2018.

4. Kalniņš G., Tārs K. Trimetilamīnu producējošo bakteriālo enzīmu strukturālie un funkcionālie pētījumi. LU 75. konferences molekulārās bioloģijas sekcija un Latvijas Bioķīmijas biedrības 2. konference. Mutisks referāts. Rīga, Latvija, 2017.

5. Kalniņš G. CutC holīna liāzes strukturālie un funkcionālie pētījumi. LU 74. konferences molekulārās bioloģijas sekcija un Latvijas Bioķīmijas biedrības 1. konference. Mutisks referāts. Rīga, Latvija, 2016.

6. Kalnins G. Structure and function of bacterial trimethylamine producing enzymes. Drug Discovery Conference. Oral presentation. Riga, Latvia, 2015.

7. Kalniņš G., Kūka J, Makrecka M, Tārs K. Trimetilamīnu producējošo bakteriālo enzīmu strukturāls un funkcionāls raksturojums. LU 73. konferences molekulārās bioloģijas sekcija. Mutisks referāts. Rīga, Latvija, 2015.

## 9. ACKNOWLEDGEMENTS

This study has been supported by the state research program „Biomedicīna sabiedrības veselībai (BIOMEDICINE) and University of Latvia Foundation grant “Bacterial microcompartments as synthetic nanoreactors” funded by SIA Mikrotik. CIISB research infrastructure project LM2015043 funded by MEYS CR is gratefully acknowledged for the financial support of the measurements at the CF Cryo-electron Microscopy and Tomography CEITEC MU.

I would like to extend my deepest gratitude to my supervisor Kaspars Tārs for the almost endless support of my ideas, allowing me an extraordinary freedom in pursuing my research and simply for the opportunity to work in his group. My gratitude also goes to all my colleagues from the structural biology group of Latvian Biomedical Research and Study Centre, and especially my student Eva Emīlija Česle for supporting my research with her work on her bachelor thesis and also Ināra Akopjana for support and advice with protein expression and production. I also thank Juris Jansons for support with electronmicroscopy.

I am also grateful to Solveiga Grīnberga and her colleagues from the mass spectrometry group of Latvian Institute of Organic Synthesis for the quantitative TMA analysis. Likewise, I am thankful to Jānis Liepiņš and Reinis Rutkis from Institute of Microbiology and Biotechnology for valuable discussions and advices. Staff of Microbial Strain Collection of Latvia is acknowledged for providing microbial strains.

I am grateful to the head of Core Facility Cryo-Electron Microscopy and Tomography Jiří Nováček and Anatolij Filimoněnko from CEITEC in Brno, Czech Republic for their advices and support with cryoEM data collection and procession. MAX-lab staff in Lund, Sweden is acknowledged for their support during the diffraction data collection at MAX II synchrotron.

And finally, I give thanks to my family for support and encouragement and my former biology/chemistry teacher Laimdota Peļše for setting me on this path.

## 10. REFERENCES

- Adams P.D., Afonine P.V., Bunkóczi G., Chen V.B., Echols N., Headd J.J., Hung L.W., Jain S., Kapral G.J., Grosse Kunstleve R.W., McCoy A.J., Moriarty N.W., Oeffner R.D., Read R.J., Richardson D.C., Richardson J.S., Terwilliger T.C., Zwart P.H. (2011). The Phenix software for automated determination of macromolecular structures. *Methods* **55**:94-106
- Al-Waiz M., Ayesh R., Mitchell S.C., Idle J.R., Smith R.L. (1987a). Disclosure of the metabolic retroversion of trimethylamine N-oxide in humans: a pharmacogenetic approach. *Clin Pharmacol Ther* **42**:608-12.
- Al-Waiz M., Ayesh R., Mitchell S.C., Idle J.R., Smith R.L. (1987b). A genetic polymorphism of the N-oxidation of trimethylamine in humans. *Clin Pharmacol Ther* **42**:588-94.
- Amoore, J.E., Hautala, E. (1983). Odor as an aid to chemical safety: Odor thresholds compared with threshold limit values and volatilities for 214 industrial chemicals in air and water dilution. *Journal of Applied Toxicology* **3**:272-290.
- Axen S.D., Erbilgin O., Kerfeld C.A. (2014). A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Comput Biol* **10**:e1003898.
- Bae S., Ulrich C.M., Neuhauser M.L., Malysheva O., Bailey L.B., Xiao L., Brown E.C., Cushing-Haugen K.L., Zheng Y., Cheng T.Y., Miller J.W., Green R., Lane D.S., Beresford S.A., Caudill MA. (2014). Plasma choline metabolites and colorectal cancer risk in the Women's Health Initiative Observational Study. *Cancer Res* **74**:7442-52.
- Barry S.M., Challis G.L. (2013). Mechanism and Catalytic Diversity of Rieske Non-Heme Iron-Dependent Oxygenases. *ACS Catal* **3**.
- Becker A., Fritz-Wolf K., Kabsch W., Knappe J., Schultz S., Volker Wagner A.F. (1999). Structure and mechanism of the glycyl radical enzyme pyruvate formate-lyase. *Nat Struct Biol* **6**:969-75.
- Bidulescu A., Chambless L.E., Siega-Riz A.M., Zeisel S.H., Heiss G. (2007). Usual choline and betaine dietary intake and incident coronary heart disease: the Atherosclerosis Risk in Communities (ARIC) study. *BMC Cardiovasc Disord* **7**:20.
- Bodea S., Funk M.A., Balskus E.P., Drennan C.L. (2016). Molecular Basis of C-N Bond Cleavage by the Glycyl Radical Enzyme Choline Trimethylamine-Lyase. *Cell Chem Biol* **23**:1206-1216.



- Boxhammer S., Glaser S., Kühl A., Wagner A.K. (2008). Characterization of the recombinant Rieske [2Fe-2S] proteins HcaC and YeaW from *E. coli*. *Biometals* **21**:459–67.
- Burnley T., Palmer C.M., Winn M. (2017). Recent developments in the CCP-EM software suite. *Acta Crystallogr D Struct Biol* **73**, 469-477.
- Busby M.G., Fischer L., da Costa K.A., Thompson D., Mar M.H., Zeisel S.H. (2004). Choline- and betaine-defined diets for use in clinical research and for the management of trimethylaminuria. *J Am Diet Assoc* **104**:1836-45.
- Cai F., Menon B.B., Cannon G.C., Curry K.J., Shively J.M., Heinhorst S. (2009). The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a CO<sub>2</sub> leakage barrier. *PLoS One* **4**:e7521.
- Cai F., Bernstein S.L., Wilson S.C., Kerfeld C.A. (2016). Production and Characterization of Synthetic Carboxysome Shells with Incorporated Luminal Proteins. *Plant Physiol* **170**:1868-77.
- Chalmers RA, Bain MD, Michelakakis H, Zschocke J, Iles RA. (2006). Diagnosis and management of trimethylaminuria (FMO3 deficiency) in children. *J Inherit Metab Dis* **29**:162-72.
- Chen M.L., Yi L., Zhang Y., Zhou X., Ran L., Yang J., Zhu J.D., Zhang Q.Y., Mi M.T. (2016). Resveratrol Attenuates Trimethylamine-N-Oxide (TMAO)-Induced Atherosclerosis by Regulating TMAO Synthesis and Bile Acid Metabolism via Remodeling of the Gut Microbiota. *MBio* **7**:e02210-15.
- Chen M.L., Zhu X.H., Ran L., Lang H.D, Yi L., Mi M.T. (2017). Trimethylamine-N-Oxide Induces Vascular Inflammation by Activating the NLRP3 Inflammasome Through the SIRT3-SOD2-mtROS Signaling Pathway. *J Am Heart Assoc* **6**: pii: e006347.
- Cheng S., Liu Y., Crowley C.S., Yeates T.O., Bobik T.A. (2008). Bacterial microcompartments: their properties and paradoxes. *Bioessays* **30**:1084-95.
- Christodoulou J. (2012). Trimethylaminuria: an under-recognised and socially debilitating metabolic disorder. *J Paediatr Child Health* **48**:E153-5
- Cho C.E, Caudill M.A. (2017). Trimethylamine-N-Oxide: Friend, Foe, or Simply Caught in the Cross-Fire? *Trends Endocrinol Metab* **28**:121-130.
- Chowdhury C., Chun S., Pang A., Sawaya M.R., Sinha S., Yeates T.O., Bobik T.A. (2015). Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proc Natl Acad Sci USA* **112**:2990-5.
- Cohen J.J., Krupp M.A., Chidsey C.A. (1958). Renal conservation of trimethylamine oxide by the spiny dogfish, *Squalus acanthias*. *Am J Physiol* **194**:229-35.

- Collins H.L., Drazul-Schrader D., Sulpizio A.C., Koster P.D., Williamson Y., Adelman S.J., Owen K., Sanli T., Bellamine A. L-Carnitine intake and high trimethylamine N-oxide plasma levels correlate with low aortic lesions in ApoE(-/-) transgenic mice expressing CETP. (2016). *Atherosclerosis* **244**:29-37.
- Craciun S., Balskus E.P. (2012). Microbial conversion of choline to trimethylamine requires a glyceryl radical enzyme. *Proc Natl Acad Sci USA* **109**:21307-12.
- Craciun S., Marks J.A., Balskus E.P. (2014). Characterization of choline trimethylamine-lyase expands the chemistry of glyceryl radical enzymes. *ACS Chem Biol* **9**:1408-13.
- Cromwell B.T., Richardson M. (1966). Studies on the biogenesis of some simple amines and quaternary ammonium compounds in higher plants: Trimethylamine in *Chenopodium vulvaria* L. *Phytochemistry* **5**:735–746.
- Dalmeijer G.W., Olthof M.R., Verhoef P., Bots M.L., van der Schouw Y.T. (2008). Prospective study on dietary intakes of folate, betaine, and choline and cardiovascular disease risk in women. *Eur J Clin Nutr* **62**:386-94.
- Dambrova M., Skapare-Makarova E., Konrade I., Pugovics O., Grinberga S., Tirzite D., Petrovska R., Kalvins I., Liepins E. (2013). Meldonium decreases the diet-increased plasma levels of trimethylamine N-oxide, a metabolite associated with atherosclerosis. *J Clin Pharmacol* **53**:1095-8.
- Dambrova M., Latkovskis G., Kuka J., Strele I., Konrade I., Grinberga S., Hartmane D., Pugovics O., Erglis A., Liepinsh E. (2016). Diabetes is Associated with Higher Trimethylamine N-oxide Plasma Levels. *Exp Clin Endocrinol Diabetes* **124**:251-6.
- Daughtry K.D., Xiao Y., Stoner-Ma D., Cho E., Orville A.M., Liu P., Allen K.N. (2012). Quaternary ammonium oxidative demethylation: X-ray crystallographic, resonance Raman, and UV-visible spectroscopic analysis of a Rieske-type demethylase. *J Am Chem Soc* **134**:2823-34.
- Dessaignes M. (1856). Trimethylamin aus Menschenharn. *Justus Liebigs Ann Chem* **100**:218.
- Dolphin C.T., Janmohamed A., Smith R.L., Shephard E.A., Phillips I.R. (1997). Missense mutation in flavin-containing mono-oxygenase 3 gene, FMO3, underlies fish-odour syndrome. *Nat Genet* **17**:491-4.
- Dorée C., Golla F. (1911). Trimethylamine as a normal constituent of human blood, urine and cerebrospinal fluid. *Biochem J* **5**:306–324.
- Dos Santos J.P., Iobbi-Nivol C., Couillault C., Giordano G., Méjean V. (1998). Molecular analysis of the trimethylamine N-oxide (TMAO) reductase respiratory system from a *Shewanella* species. *J Mol Biol* **284**:421-33.

- Emsley P, Cowtan K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**:2126-32.
- Evans P. R. (1997). SCALA. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, **No. 33**.
- Falls J.G., Ryu D.Y., Cao Y., Levi P.E., Hodgson E. (1997). Regulation of mouse liver flavin-containing monooxygenases 1 and 3 by sex steroids. *Arch Biochem Biophys* **342**:212-23.
- Fan C., Cheng S., Liu Y., Escobar C.M., Crowley C.S., Jefferson R.E., Yeates T.O., Bobik T.A. (2010). Short N-terminal sequences package proteins into bacterial microcompartments. *Proc Natl Acad Sci USA* **107**:7509-14.
- Fan C., Cheng S., Sinha S., Bobik T.A. (2012). Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. *Proc Natl Acad Sci USA* **109**:14995-5000.
- Fennema D., Phillips I.R., Shephard E.A. (2016). Trimethylamine and Trimethylamine N-Oxide, a Flavin-Containing Monooxygenase 3 (FMO3)-Mediated Host-Microbiome Metabolic Axis Implicated in Health and Disease. *Drug Metab Dispos* **44**:1839-1850.
- Funk M.A., Judd E.T., Marsh E.N., Elliott S.J., Drennan C.L. (2014). Structures of benzylsuccinate synthase elucidate roles of accessory subunits in glycyl radical enzyme activation and activity. *Proc Natl Acad Sci USA* **111**:10161-6.
- Gao X., Liu X., Xu J., Xue C., Xue Y., Wang Y. (2014). Dietary trimethylamine N-oxide exacerbates impaired glucose tolerance in mice fed a high fat diet. *J Biosci Bioeng* **118**:476-81.
- Geng J., Yang C., Wang B., Zhang X., Hu T., Gu Y., Li J. (2018). Trimethylamine N-oxide promotes atherosclerosis via CD36-dependent MAPK/JNK pathway. *Biomed Pharmacother* **97**:941-947.
- Goldstein L., Funkhouser D. (1972). Biosynthesis of trimethylamine oxide in the nurse shark, *Ginglymostoma cirratum*. *Comp Biochem Physiol A Comp Physiol* **42**:51-7.
- Goldstein L., Palatt P.J. (1974). Trimethylamine oxide excretion rates in elasmobranchs. *Am J Physiol* **227**:1268-72.
- Hanna W.F., Vickery H.B., Pucher G.W. (1932). The isolation of trimethylamine from spores of *Tilletia levis*, the stinking smut of wheat. *J Biol Chem* **97**:351-358.
- Havemann GD, Bobik TA. (2003). Protein content of polyhedral organelles involved in coenzyme B12-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar Typhimurium LT2. *J Bacteriol* **185**:5086-95.

- Hayward HR, Stadtman TC. (1960). Anaerobic degradation of choline. II. Preparation and properties of cell-free extracts of *Vibrio cholonicus*. *J Biol Chem* **235**:538-43.
- Heianza Y., Ma W., Manson J.E., Rexrode K.M., Qi L. (2017). Gut Microbiota Metabolites and Risk of Major Adverse Cardiovascular Disease Events and Death: A Systematic Review and Meta-Analysis of Prospective Studies. *J Am Heart Assoc* **6**: pii: e004947.
- Heldt D., Frank S., Seyedarabi A., Ladikis D., Parsons J.B., Warren M.J., Pickersgill R.W. (2009). Structure of a trimeric bacterial microcompartment shell protein, EtuB, associated with ethanol utilization in *Clostridium kluyveri*. *Biochem J* **423**:199-207.
- Herring T.I., Harris T.N., Chowdhury C., Mohanty S.K., Bobik T.A. (2018). A Bacterial Microcompartment Is Used for Choline Fermentation by *Escherichia coli* 536. *J Bacteriol* **200**: e00764-17.
- Hofmann A.W. (1851). Researches into the molecular constitution of the organic bases. *Phil Trans Roy Soc London* **141**:357–398.
- Hofmann A.W. (1853). On the occurrence of trimethylamine in herring-brine. *Quart J Chem Soc* **5**:288.
- Jakobson C.M., Slininger Lee M.F., Tullman-Ercek D. (2017). De novo design of signal sequences to localize cargo to the 1,2-propanediol utilization microcompartment. *Protein Sci* **26**:1086-1092.
- Jameson E., Fu T., Brown I.R., Paszkiewicz K., Purdy K.J., Frank S., Chen Y. (2016). Anaerobic choline metabolism in microcompartments promotes growth and swarming of *Proteus mirabilis*. *Environ Microbiol* **18**:2886-98.
- Jorda J., Lopez D., Wheatley N.M., Yeates T.O. (2013). Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci* **22**:179-95.
- Jorda J., Leibly D.J., Thompson M.C., Yeates T.O. Structure of a novel 13 nm dodecahedral nanocage assembled from a redesigned bacterial microcompartment shell protein. *Chem Commun (Camb)*. **52**, 5041-4 (2016).
- Kenyon S., Carmichael P.L., Khalaque S., Panchal S., Waring R., Harris R., Smith R.L., Mitchell S.C. (2004). The passage of trimethylamine across rat and human skin. *Food Chem Toxicol* **42**:1619-28.
- Kerfeld C.A., Sawaya M.R., Tanaka S., Nguyen C.V., Phillips M, Beeby M., Yeates T.O. (2005). Protein structures forming the shell of primitive bacterial organelles. *Science*, **309**, 936–938.



- Kerfeld C.A., Melnicki M.R. (2016). Assembly, function and evolution of cyanobacterial carboxysomes. *Curr Opin Plant Biol* **31**:66-75.
- Kerfeld C.A., Aussignargues C., Zarzycki J., Cai F., Sutter M. (2018). Bacterial microcompartments. *Nat Rev Microbiol* **16**:277-290.
- Kim R.B., Morse B.L., Djurdjev O., Tang M., Muirhead N., Barrett B., Holmes D.T., Madore F., Clase C.M., Rigatto C., Levin A; CanPREDDICT Investigators. (2016). Advanced chronic kidney disease populations have elevated trimethylamine N-oxide levels associated with increased cardiovascular events. *Kidney Int* **89**:1144-1152.
- Klein M.G., Zwart P., Bagby S.C., Cai F., Chisholm S.W., Heinhorst S., Cannon G.C., Kerfeld C.A. (2009). Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. *J Mol Biol* **392**:319-33.
- Knappe J, Wagner AF. (1995). Glycyl free radical in pyruvate formate-lyase: synthesis, structure characteristics, and involvement in catalysis. *Methods Enzymol* **258**:343-62.
- Koeth R.A., Wang Z., Levison B.S., Buffa J.A., Org E., Sheehy B.T., Britt E.B., Fu X., Wu Y., Li L., Smith J.D., DiDonato J.A., Chen J., Li H., Wu G.D., Lewis J.D., Warrier M., Brown J.M., Krauss R.M., Tang W.H., Bushman F.D., Lusis A.J., Hazen S.L. (2013). Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat Med* **19**:576-85.
- Koeth R.A., Levison B.S., Culley M.K., Buffa J.A., Wang Z., Gregory J.C., Org E., Wu Y., Li L., Smith J.D., Tang W.H.W., DiDonato J.A., Lusis A.J., Hazen S.L. (2014).  $\gamma$ -Butyrobetaine is a proatherogenic intermediate in gut microbial metabolism of L-carnitine to TMAO. *Cell Metab* **20**:799-812.
- Kuka J., Liepinsh E., Makrecka-Kuka M., Liepins J., Cirule H., Gustina D., Loza E., Zharkova-Malkova O., Grinberga S., Pugovics O., Dambrova M. (2014). Suppression of intestinal microbiota-dependent production of pro-atherogenic trimethylamine N-oxide by shifting L-carnitine microbial degradation. *Life Sci* **117**:84-92.
- LaMattina J.W., Keul N.D., Reitzer P., Kapoor S., Galzerani F., Koch D.J., Gouvea I.E., Lanzilotta W.N. (2016). 1,2-Propanediol Dehydration in *Roseburia inulinivorans*: STRUCTURAL BASIS FOR SUBSTRATE AND ENANTIOMER SELECTIVITY. *J Biol Chem* **291**:15515-26.
- Lang D.H., Yeung C.K., Peter R.M., Ibarra C., Gasser R., Itagaki K., Philpot R.M., Rettie A.E. (1998). Isoform specificity of trimethylamine N-oxygenation by human flavin-containing monooxygenase (FMO) and P450 enzymes: selective catalysis by FMO3. *Biochem Pharmacol* **56**:1005-12.

- Lawrence A.D., Frank S., Newnham S., Lee M.J., Brown I.R., Xue W.F., Rowe M.L., Mulvihill D.P., Prentice M.B., Howard M.J., Warren M.J. (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.
- Leonardos, G., Kendall, D., Barnard, N. J. (1969). Odor threshold determinations of 53 odorant chemicals. *Air Pollu Control Assoc* **19**:2, 91-95.
- Leslie A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, No. 26
- Leuthner B., Leutwein C., Schulz H., Hörth P., Haehnel W., Schiltz E., Schägger H., Heider J. (1998). Biochemical and genetic characterization of benzylsuccinate synthase from *Thauera aromatica*: a new glycyl radical enzyme catalysing the first step in anaerobic toluene metabolism. *Mol Microbiol* **28**:615-28.
- Levin B.J., Huang Y.Y., Peck S.C., Wei Y., Martínez-Del Campo A., Marks J.A., Franzosa E.A., Huttenhower C., Balskus E.P. (2017). A prominent glycyl radical enzyme in human gut microbiomes metabolizes trans-4-hydroxy-l-proline. *Science* **355** pii: eaai8386.
- Li Q., Korzan W.J., Ferrero D.M., Chang R.B., Roy D.S., Buchi M., Lemon J.K., Kaur A.W., Stowers L., Fendt M., Liberles S.D. (2013). Synchronous evolution of an odor biosynthesis pathway and behavioral response. *Curr Biol* **23**:11-20.
- Li T., Chen Y., Gua C., Li X. (2017). Elevated Circulating Trimethylamine N-Oxide Levels Contribute to Endothelial Dysfunction in Aged Rats through Vascular Inflammation and Oxidative Stress. *Front Physiol* **8**:350.
- Liberles, S.D., Buck, L.B. (2006). A second class of chemosensory receptors in the olfactory epithelium. *Nature* **442**:645-650.
- Logan D.T., Andersson J., Sjöberg B.M., Nordlund P. (1999). A glycyl radical site in the crystal structure of a class III ribonucleotide reductase. *Science* **283**:1499-504.
- Ma G., Pan B., Chen Y., Guo C., Zhao M., Zheng L., Chen B. (2017). Trimethylamine N-oxide in atherogenesis: impairing endothelial self-repair capacity and enhancing monocyte adhesion. *Biosci Rep* **37**: pii: BSR20160244.
- Makrecka-Kuka M., Volska K., Antone U., Vilskersts R., Grinberga S., Bandere D., Liepinsh E., Dambrova M. (2017). Trimethylamine N-oxide impairs pyruvate and fatty acid oxidation in cardiac mitochondria. *Toxicol Lett* **267**:32-38.

- Méjean V., Iobbi-Nivol C., Lepelletier M., Giordano G., Chippaux M., Pascal M.C. (1994). TMAO anaerobic respiration in *Escherichia coli*: involvement of the tor operon. *Mol Microbiol* **11**:1169-79.
- Miller C.A., Corbin K.D., da Costa K.A., Zhang S., Zhao X., Galanko J.A., Blevins T., Bennett B.J., O'Connor A., Zeisel S.H. (2014). Effect of egg ingestion on trimethylamine-N-oxide production in humans: a randomized, controlled, dose-response study. *Am J Clin Nutr* **100**:778-86.
- Miao J, Ling AV, Manthena PV, Gearing ME, Graham MJ, Crooke RM, Croce KJ, Esquejo RM, Clish CB; Morbid Obesity Study Group, Vicent D, Biddinger SB. (2015). Flavin-containing monooxygenase 3 as a potential player in diabetes-associated atherosclerosis. *Nat Commun* **6**:6498.
- Missailidis C., Hällqvist J., Qureshi A.R., Barany P., Heimbürger O., Lindholm B., Stenvinkel P., Bergman P. (2016). Serum Trimethylamine-N-Oxide Is Strongly Related to Renal Function and Predicts Outcome in Chronic Kidney Disease. *PLoS One* **11**:e0141738.
- Mohammadi A., Najar A.G., Yaghoobi M.M., Jahani Y., Vahabzadeh Z. (2016). Trimethylamine-N-Oxide Treatment Induces Changes in the ATP-Binding Cassette Transporter A1 and Scavenger Receptor A1 in Murine Macrophage J774A.1 cells. *Inflammation* **39**:393-404.
- Mohajerani F., Hagan M.F. (2018). The role of the encapsulated cargo in microcompartment assembly. *PLoS Comput Biol* **14**, e1006351.
- Muramatsu H., Matsuo H., Okada N., Ueda M., Yamamoto H., Kato S., Nagata S. (2013). Characterization of ergothionase from *Burkholderia* sp. HME13 and its application to enzymatic quantification of ergothioneine. *Appl Microbiol Biotechnol* **97**:5389-400.
- Murshudov G.N., Vagin A.A., Dodson E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**:240-55.
- Nagata C, Wada K, Tamura T, Konishi K, Kawachi T, Tsuji M, Nakamura K. (2015). Choline and Betaine Intakes Are Not Associated with Cardiovascular Disease Mortality Risk in Japanese Men and Women. *J Nutr* **145**:1787-92.
- O'Brien J.R., Raynaud C., Croux C., Girbal L., Soucaille P., Lanzilotta W.N. (2004). Insight into the mechanism of the B12-independent glycerol dehydratase from *Clostridium butyricum*: preliminary biochemical and structural characterization. *Biochemistry* **43**:4635-45.

- Organ C.L., Otsuka H., Bhushan S., Wang Z., Bradley J., Trivedi R., Polhemus D.J., Tang W.H., Wu Y., Hazen S.L., Lefer D.J. (2016). Choline Diet and Its Gut Microbe-Derived Metabolite, Trimethylamine N-Oxide, Exacerbate Pressure Overload-Induced Heart Failure. *Circ Heart Fail* **9**:e002314.
- Orman M., Bodea S., Funk M.A., Campo A.M., Bollenbach M., Drennan C.L., Balskus E.P. (2019). Structure-Guided Identification of a Small Molecule That Inhibits Anaerobic Choline Metabolism by Human Gut Bacteria. *J Am Chem Soc* **141**:33-37.
- Peck S.C., Denger K., Burrichter A., Irwin S.M., Balskus E.P., Schleheck D. (2019). A glyceryl radical enzyme enables hydrogen sulfide production by the human intestinal bacterium *Bilophila wadsworthia*. *Proc Natl Acad Sci USA* **116**:3171-3176.
- Penrod J.T., Roth J.R. (2006). Conserving a volatile metabolite: a role for carboxysome-like organelles in *Salmonella enterica*. *J Bacteriol* **188**:2865-74.
- Perlmutter J.D., Mohajerani F., Hagan M.F. (2016). Many-molecule encapsulation by an icosahedral shell. *Elife* **5** pii: e14078.
- Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C., Ferrin T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**:1605-12.
- Randrianarisoa E., Lehn-Stefan A., Wang X., Hoene M., Peter A., Heinzmann S.S., Zhao X., Königsrainer I., Königsrainer A., Balletshofer B., Machann J., Schick F., Fritsche A., Häring H.U., Xu G., Lehmann R., Stefan N. (2016). Relationship of Serum Trimethylamine N-Oxide (TMAO) Levels with early Atherosclerosis in Humans. *Sci Rep* **6**:26745.
- Rath S., Heidrich B., Pieper D.H, Vital M. (2017). Uncovering the trimethylamine-producing bacteria of the human gut microbiota. *Microbiome* **5**:54.
- Rathinasabapathi B., Burnet M., Russell B.L., Gage D.A., Liao P.C., Nye G.J., Scott P., Golbeck J.H., Hanson A.D. (1997). Choline monooxygenase, an unusual iron-sulfur enzyme catalyzing the first step of glycine betaine synthesis in plants: prosthetic group characterization and cDNA cloning. *Proc Natl Acad Sci USA* **94**:3454-8.
- Roberts A.B., Gu X., Buffa J.A., Hurd A.G., Wang Z., Zhu W., Gupta N., Skye S.M., Cody D.B., Levison B.S., Barrington W.T., Russell M.W., Reed J.M., Duzan A., Lang J.M., Fu X., Li L., Myers A.J., Rachakonda S., DiDonato J.A., Brown J.M., Gogonea V., Lulis A.J., Garcia-Garcia J.C., Hazen S.L. (2018). Development of a gut microbe-targeted nonlethal therapeutic to inhibit thrombosis potential. *Nat Med* **24**:1407-1417.



- Romano K.A., Vivas E.I., Amador-Noguez D., Rey F.E. (2015). Intestinal microbiota composition modulates choline bioavailability from diet and accumulation of the proatherogenic metabolite trimethylamine-N-oxide. *MBio* **6**:e02481.
- Sarma R., Paul S. (2013). Crucial importance of water structure modification on trimethylamine N-oxide counteracting effect at high pressure. *J Phys Chem B* **117**:677-89.
- Schiattarella G.G., Sannino A., Toscano E., Giugliano G., Gargiulo G., Franzone A., Trimarco B., Esposito G., Perrino C. (2017). Gut microbe-generated metabolite trimethylamine-N-oxide as cardiovascular risk biomarker: a systematic review and dose-response meta-analysis. *Eur Heart J* **38**:2948-2956.
- Schneider J., Girreser U., Havemeyer A., Bittner F., Clement B. (2018). Detoxification of Trimethylamine N-Oxide by the Mitochondrial Amidoxime Reducing Component mARC. *Chem Res Toxicol* **31**:447-453.
- Seibel B.A., Walsh P.J. (2002). Trimethylamine oxide accumulation in marine animals: relationship to acylglycerol storage. *J Exp Biol* **205**:297-306.
- 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. *J Mol Biol* **426**:2328-45.
- Shephard E.A., Treacy E.P., Phillips I.R. (2015). Clinical utility gene card for: Trimethylaminuria - update 2014. *Eur J Hum Genet* **23**.
- Shisler K.A., Broderick J.B. (2014). Glycyl radical activating enzymes: structure, mechanism, and substrate interactions. *Arch Biochem Biophys* **546**:64-71.
- Skye S.M., Zhu W., Romano K.A., Guo C.J., Wang Z., Jia X., Kirsop J., Haag B., Lang J.M., DiDonato J.A., Tang W.H.W, Lusic A.J., Rey F.E., Fischbach M.A., Hazen S.L. (2018). Microbial Transplantation With Human Gut Commensals Containing CutC Is Sufficient to Transmit Enhanced Platelet Reactivity and Thrombosis Potential. *Circ Res* **123**:1164-1176.
- Stephens, E.R. (1971). Identification of odors from cattle feed lots. *Calif Agr* **25**:10–11.
- Summers R.M., Louie T.M., Yu C.L., Subramanian M. (2011). Characterization of a broad-specificity non-haem iron N-demethylase from *Pseudomonas putida* CBB5 capable of utilizing several purine alkaloids as sole carbon and nitrogen source. *Microbiology* **157**:583-92.
- Sun X., Ollagnier S., Schmidt P.P., Atta M., Mulliez E., Lepape L., Eliasson R., Gräslund A., Fontecave M., Reichard P., Sjöberg B.M. (1996). The free radical of the anaerobic

- ribonucleotide reductase from *Escherichia coli* is at glycine 681. *J Biol Chem* **271**:6827-31.
- Sutter M., Faulkner M., Aussignargues C., Paasch B.C., Barrett S., Kerfeld C.A., Liu L.N. (2016). Visualization of Bacterial Microcompartment Facet Assembly Using High-Speed Atomic Force Microscopy. *Nano Lett* **16**:1590-5.
- Sutter M., Greber B., Aussignargues C., Kerfeld C.A. (2017). Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science* **356**:1293-1297.
- Sutter M., McGuire S., Ferlez B., Kerfeld C.A. (2019a). Structural Characterization of a Synthetic Tandem-Domain Bacterial Microcompartment Shell Protein Capable of Forming Icosahedral Shell Assemblies. *ACS Synth Biol*, **8**:668-674
- Sutter M., Laughlin T.G., Sloan N.B., Serwas D., Davies K.M., Kerfeld C.A. (2019b). Structure of a synthetic beta-carboxysome shell. *Plant Physiol* pii: pp.00885.2019
- 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*. *J Bacteriol* **192**:6056-63.
- 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 (5866):1083-6.
- Tang W.H., Wang Z., Levison B.S., Koeth R.A., Britt E.B., Fu X., Wu Y., Hazen S.L. (2013). Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med* **368**:1575-84.
- Tang W.H, Wang Z., Kennedy D.J., Wu Y., Buffa J.A., Agatista-Boyle B., Li X.S., Levison B.S., Hazen S.L. (2015). Gut microbiota-dependent trimethylamine N-oxide (TMAO) pathway contributes to both development of renal insufficiency and mortality risk in chronic kidney disease. *Circ Res* **116**:448-55.
- Tang W.H., Wang Z., Li X.S., Fan Y., Li D.S., Wu Y., Hazen S.L. (2017). Increased Trimethylamine N-Oxide Portends High Mortality Risk Independent of Glycemic Control in Patients with Type 2 Diabetes Mellitus. *Clin Chem* **63**:297-306.
- Thibodeaux C.J., van der Donk W.A. (2012). Converging on a mechanism for choline degradation. *Proc Natl Acad Sci USA* **109**:21184-5.
- Thompson M.C., Cascio D., Leibly D.J., Yeates T.O. (2015). An allosteric model for control of pore opening by substrate binding in the EutL microcompartment shell protein. *Protein Sci* **24**:956-75.

- Treacy E., Johnson D., Pitt J.J., Danks D.M. (1995). Trimethylaminuria, fish odour syndrome: a new method of detection and response to treatment with metronidazole. *J Inherit Metab Dis* **18**:306-12.
- Treacy E.P., Akerman B.R., Chow L.M., Youil R., Bibeau C., Lin J., Bruce A.G., Knight M., Danks D.M., Cashman J.R., Forrest S.M. (1998). Mutations of the flavin-containing monooxygenase gene (FMO3) cause trimethylaminuria, a defect in detoxication. *Hum Mol Genet* **7**:839-45.
- Treberg J.R., Driedzic W.R. (2002). Elevated levels of trimethylamine oxide in deep-sea fish: evidence for synthesis and intertissue physiological importance. *J Exp Zool* **293**:39-45.
- Trøseid M., Ueland T., Hov J.R., Svardal A., Gregersen I., Dahl C.P., Aakhus S., Gude E., Bjørndal B., Halvorsen B, Karlsen T.H., Aukrust P., Gullestad L., Berge R.K., Yndestad A. (2015). Microbiota-dependent metabolite trimethylamine-N-oxide is associated with disease severity and survival of patients with chronic heart failure. *J Intern Med* **277**:717-26.
- Van der Geize R., Yam K., Heuser T., Wilbrink M.H., Hara H., Anderton M.C., Sim E., Dijkhuizen L., Davies J.E., Mohn W.W., Eltis L.D. (2007). A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages. *Proc Natl Acad Sci USA* **104**:1947-52.
- Vagin A, Teplyakov A. (1997). MOLREP: an automated program for molecular replacement. *J Appl Cryst.* **30**, 1022–1025.
- Wagner A.F, Frey M., Neugebauer F.A., 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.
- Wang Z., Klipfell E., Bennett B.J., Koeth R., Levison B.S., Dugar B., Feldstein A.E., Britt E.B., Fu X., Chung Y.M., Wu Y., Schauer P., Smith J.D., Allayee H., Tang W.H., DiDonato J.A., Lusis A.J., Hazen S.L. (2011). Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**:57-63.
- Wang Z., Roberts A.B., Buffa J.A., Levison B.S., Zhu W., Org E., Gu X., Huang Y., Zamanian-Daryoush M., Culley M.K., DiDonato A.J., Fu X., Hazen J.E., Krajcik D., DiDonato J.A., Lusis A.J., Hazen S.L. (2015). Non-lethal Inhibition of Gut Microbial Trimethylamine Production for the Treatment of Atherosclerosis. *Cell* **163**:1585-95.
- Warrier M., Shih D.M., Burrows A.C., Ferguson D., Gromovsky A.D., Brown A.L., Marshall S., McDaniel A., Schugar R.C., Wang Z., Sacks J., Rong X., Vallim T.A., Chou J.,

- Ivanova P.T., Myers D.S., Brown H.A., Lee R.G., Crooke R.M., Graham M.J., Liu X., Parini P., Tontonoz P., Lusis A.J., Hazen S.L., Temel R.E., Brown J.M. (2015). The TMAO-Generating Enzyme Flavin Monooxygenase 3 Is a Central Regulator of Cholesterol Balance. *Cell Rep* **10**:326-338.
- Weinert C.H., Empl M.T., Krüger R., Frommherz L., Egert B., Steinberg P., Kulling S.E. (2017). The influence of a chronic L-carnitine administration on the plasma metabolome of male Fischer 344 rats. *Mol Nutr Food Res* **61**.
- Wheatley N.M., Gidaniyan S.D., Liu Y., Cascio D., Yeates T.O. (2013). Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins. *Protein Sci* **22**:660-5.
- Wu W.K., Chen C.C., Liu P.Y., Panyod S., Liao B.Y., Chen P.C., Kao H.L., Kuo H.C., Kuo C.H., Chiu T.H.T., Chen R.A., Chuang H.L., Huang Y.T., Zou H.B., Hsu C.C., Chang T.Y., Lin C.L., Ho C.T., Yu H.T., Sheen L.Y., Wu M.S. (2018). Identification of TMAO-producer phenotype and host-diet-gut dysbiosis by carnitine challenge test in human and germ-free mice. *Gut pii: gutjnl-2018-317155*.
- Yancey P.H., Clark M.E., Hand S.C., Bowlus R.D., Somero G.N. (1982). Living with water stress: evolution of osmolyte systems. *Science* **217**:1214-22.
- Yancey P.H., Siebenaller J.F. (1999). Trimethylamine oxide stabilizes teleost and mammalian lactate dehydrogenases against inactivation by hydrostatic pressure and trypsinolysis. *J Exp Biol* **202**:3597-603
- Yancey P.H., Fyfe-Johnson A.L., Kelly R.H., Walker V.P., Auñón M.T. (2001). Trimethylamine oxide counteracts effects of hydrostatic pressure on proteins of deep-sea teleosts. *J Exp Zool* **289**:172-6.
- Yancey P.H., Gerringer M.E., Drazen J.C., Rowden A.A., Jamieson A. (2014). Marine fish may be biochemically constrained from inhabiting the deepest ocean depths. *Proc Natl Acad Sci USA* **111**:4461-5.
- Yang J., Naik S.G., Ortillo D.O., García-Serres R., Li M., Broderick W.E., Huynh B.H., Broderick J.B. (2009). The iron-sulfur cluster of pyruvate formate-lyase activating enzyme in whole cells: cluster interconversion and a valence-localized [4Fe-4S]<sup>2+</sup> state. *Biochemistry* **48**:9234-41.
- Yoshiyama-Yanagawa T., Enya S., Shimada-Niwa Y., Yaguchi S., Haramoto Y., Matsuya T., Shiomi K., Sasakura Y., Takahashi S., Asashima M., Kataoka H., Niwa R. (2011). The conserved Rieske oxygenase DAF-36/Neverland is a novel cholesterol-metabolizing enzyme. *J Biol Chem* **286**:25756-62.



- Zarzycki J, Erbilgin O, Kerfeld CA. (2015). Bioinformatic characterization of glyceryl radical enzyme-associated bacterial microcompartments. *Appl Environ Microbiol*, **81**:8315-29.
- Zarzycki J., Sutter M., Cortina N.S., Erb T.J., Kerfeld C.A. (2017). In Vitro Characterization and Concerted Function of Three Core Enzymes of a Glyceryl Radical Enzyme - Associated Bacterial Microcompartment. *Sci Rep* **7**:42757.
- von Zeynek R. (1899). Um das Wissen der Menschlichen Leber Galle. Wien kiln Wochen **12**:568–569.
- Zhang A.Q., Mitchell S.C., Smith R.L. (1996). Discontinuous distribution of N-oxidation of dietary-derived trimethylamine in a British population. *Xenobiotica* **26**:957-61.
- Zhang K. (2016). Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**:1-12.
- Zheng S.Q., Palovcak E., Armache J.P., Verba K.A., Cheng Y., Agard D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332.
- Zhu Y., Jameson E., Crosatti M., Schäfer H., Rajakumar K., Bugg T.D., Chen Y. (2014). Carnitine metabolism to trimethylamine by an unusual Rieske-type oxygenase from human microbiota. *Proc Natl Acad Sci USA* **111**:4268-73.
- Zhu W., Gregory J.C., Org E., Buffa J.A., Gupta N., Wang Z., Li L., Fu X., Wu Y., Mehrabian M., Sartor R.B., McIntyre T.M., Silverstein R.L., Tang W.H.W., DiDonato J.A., Brown J.M., Lusic A.J., Hazen S.L. (2016). Gut Microbial Metabolite TMAO Enhances Platelet Hyperreactivity and Thrombosis Risk. *Cell* **165**:111-124.
- Zhu W., Buffa J.A., Wang Z., Warriar M., Schugar R., Shih D.M., Gupta N., Gregory J.C., Org E., Fu X., Li L., DiDonato J.A., Lusic A.J., Brown J.M., Hazen S.L. (2018). Flavin monooxygenase 3, the host hepatic enzyme in the metaorganismal trimethylamine N-oxide-generating pathway, modulates platelet responsiveness and thrombosis risk. *J Thromb Haemost* **16**:1857-1872.
- Zivanov J., Nakane T., Forsberg B.O., Kimanius D., Hagen W.J., Lindahl E., Scheres S.H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, e42166.
- Zou, Q., Bennion, B.J., Daggett, V., Murphy, K.P. (2002). The molecular mechanism of stabilization of proteins by TMAO and its ability to counteract the effects of urea. *J Am Chem Soc* **124**: 1192–1202.