EXPRESSION, PURIFICATION, AND CHARACTERIZATION OF PUTATIVE CHOLINE KINASES FROM MICROORGANISMS

MOAD MAHMUD ALARABI KHALIFA

UNIVERSITI SAINS MALAYSIA

2020

EXPRESSION, PURIFICATION, AND CHARACTERIZATION OF PUTATIVE CHOLINE KINASES FROM MICROORGANISMS

by

MOAD MAHMUD ALARABI KHALIFA

Dissertation submitted in partial fulfillment of
the requirements of the degree of
Master of Science (Biomedicine) Mixed Mode

September 2020

ACKNOWLEDGEMENT

To all who have supported me on my journey toward my goals. To all who have been like a pillar in the construction of the foundation that got me here and will keep me going upward and forward towards my dreams. To my family, to my supervisors, to my seniors, to my friends. Thank you. Special thanks are extended to the pioneers, my supervisor Assoc. Prof. Dr. See Too Wei Cun and my co-supervisor Assoc. Prof. Dr. Few Ling Ling for allowing me to be at the frontiers with this research that was supported by the USM RUI grant (1001/PPSK/8012239) and FRGS grant (203/PPSK/6171222).

My heartfelt thanks to everyone

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF PLATES

Page

LIST OF ABBREVIATIONS

ChoK	Choline kinase
ChoKIs	Choline kinase inhibitors
*Sp*ChoK	*Streptococcus pneumoniae* choline kinase
*Sa*ChoK	*Staphylococcus aureus* choline kinase
*Hi*ChoK	*Haemophilus influenzae* choline kinase
*Nm*ChoK	*Neisseria meningitidis* choline kinase
*Na*ChoK	*Neisseria lactamica* choline kinase
*Pf*ChoK	*Plasmodium falciparum* choline kinase
hChoK	Human choline kinase
AMR	Antimicrobial resistance
CDC	Centers for Disease Control and Prevention
IDSA	Infectious Diseases Society of America
MRSA	Methicillin-resistant *Staphylococcus aureus*
VAN	Vancomycin
DAP	Daptomycin
VISA	Vancomycin-intermediate *Staphylococcus aureus*
VRSA	Vancomycin-resistant *Staphylococcus aureus*
XDR	Extensively drug-resistant
PDR	Pandrug-resistant
MDR	Multidrug-resistant
WHO	World Health Organization
NPs	Nanoparticles
PC	Phosphatidylcholine
PE	Phosphatidylethanolamine
PS	Phosphatidylserine
PA	Phosphatidic acid
PG	Phosphatidylglycerol
PI	Phosphatidylinositol
CL	Cardiolipin
LPG	Lysyl-phosphatidylglycerol
GLs	Glycolipids

LPS	Lipopolysaccharide
Cho	Choline
ChoP	Phosphorylcholine
TA	Teichoic acids
CTA	Cell wall teichoic acid
LTA	Lipoteichoic acid
CRP	C-reactive protein
CDP-ethanolamine	Diphosphate-ethanolamine
CDP-choline	Cytidine diphosphate-choline
LPLs	Lysophospholipids
EK	Ethanolamine kinase
ECT	Ethanolamine-phosphate cytidylyltransferase
EPT	Ethanolamine phosphotransferase
CPT	Choline phosphotransferase
CEPT	Choline/ethanolamine phosphotransferase
DAG	Diacylglycerol
LicB	Choline transporter
LicA	Choline kinase
LicC	Phosphorylcholine cytidylyltransferase
LicD	Phosphorylcholine transferases
CCT	Phosphorylcholine cytidylyltransferase
CTP	Cytidine triphosphate
TarI	Cytidylyl transferase
TarJ	Alcohol dehydrogenase
TacF	Teichoic acid flippase
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
PLMT	Phospholipid *N*-methyltransferase
SAM	*S*-adenosylmethionine
MMPE	Monomethylphosphatidylethanolamine
DMPE	Dimethylphosphatidylethanolamine
PSS	Phosphatidylserine synthase
PSD	Phosphatidylserine decarboxylase
Pcs	Phosphatidylcholine synthase

GPC	Glycerophosphocholine
LPC	Lysophosphatidylcholine
GPCAT	GPC acyltransferase
SAH	*S*-adenosylhomocysteine
CDP-DAG	Cytidine diphosphate-diacylglycerol
CDS	Phosphatidate cytidylyltransferase
CDS	CDP-DAG synthase
ChoP-LPS	Phosphorylcholine-lipopolysaccharides
ChoP-LTA	Phosphorylcholine-lipoteichoic acid
*Tg*ChoK	*Toxoplasma gondii* choline kinase
*Lm*ChoK	*Leishmania major* choline kinase
*Pa*ChoK	*Pseudomonas aeruginosa* choline kinase
*Ps*ChoK	*Pseudomonas syringae* choline kinase
nChoK	Nematode ChoK
Asp	Aspartate residue
Asn	Asparagine residue
HC-3	Hemicholinium-3
ΔG	Binding free energy
SC5A7	Human high-affinity Cho transporter
OpuC	Low-affinity Cho transporter
BetT	Low-affinity Cho transporter
PDB	Protein Data Bank
RCSB	Research Collaboratory for Structural Bioinformatics
3D	Three dimensions
CARG	CARGGLKSC
TBI	Traumatic brain injury
PLGA-PLH-PEG	Poly(D,L-lactic-co-glycolic acid)-b-poly(L-histidine)-b-poly(ethylene glycol)
EPS	Extracellular polymeric substances
AD	Alzheimer's disease
mRNA	Messenger RNA
DNA	Deoxyribonucleic acid
MCS	Multiple cloning site
GST	Glutathione S-transferase
MBP	Maltose-binding protein

NusA	N-utilization substance protein A
C-terminus	Carboxyl terminus
N-terminus	Nitrogen terminus
tRNA	Transfer RNA
TAE	Tris-acetate-EDTA
dH2O	Distilled water
LB	Luria Bertani
PEG	Polyethylene glycol
TE	Transformation efficiency
CFU	Colony-forming units
PCR	Polymerase chain reaction
FASTA	FAST-All
PPI	Protein-protein interactions
fld	Folding propensity
dis	Disorder propensity
bet	Beta-strand propensities
pI	Isoelectric point
ent	Sequence entropy
IPTG	Isopropyl β- d-1-thiogalactopyranoside
LD_{50}	Median lethal dose
IC_{50}	Half maximal inhibitory concentration

EKSPRESI, PENULENAN DAN PENCIRIAN KOLINA KINASE PUTATIF DARIPADA MIKROORGANISMA

ABSTRAK

Kerintangan antimikrob (AMR) telah menjadi suatu ancaman kepada komuniti seluruh dunia. AMR dijangka akan mengakibatkan sepuluh juta kematian setahun menjelang 2050. Dalam bakteria, kolina kinase (ChoK) bertanggungjawab untuk mensintesis fosforilkolina, iaitu prekursor kepada asid lipoteikoik dan asid teikoik dinding sel pada bakteria Gram positif. Dalam bakteria Gram negatif, fosforilkolina digabungkan ke dalam lipopolisakarida membran yang mengawalatur interaksi patogen dan sel hos. Perencat kolina kinase (ChoKIs) yang merosakkan dinding sel telah diuji pada *Streptococcus pneumoniae* dan menunjukkan keputusan yang memberangsangkan. Aktiviti ChoKIs juga boleh ditingkatkan dengan menggunakan partikel nano yang berfungsi sebagai sistem penghantaran ubat. Penghasilan sasaran ubat (iaitu ChoKs daripada bakteria) dalam bentuk protein rekombinan adalah penting untuk menguji keberkesanan ChoKIs. Kajian ini cuba menangani isu AMR dengan meneroka keadaan terbaik untuk penghasilan ChoKs rekombinan daripada *Staphylococcus aureus* (*Sa*ChoK), *Neisseria meningitidis* (*Ni*ChoK) dan *Haemophilus influenzae* (*Hi*ChoK) diikuti dengan penilaian *in silico* ketiga-tiga ChoKs ini sebagai sasaran berpotensi untuk ChoKIs menggunakan kaedah pemodelan struktur dan pelabuhan molekul. Ketiga-tiga ChoKs bakteria pada asalnya diklonkan dalam vektor pET14b untuk ekspresi sebagai protein bertanda His. Walau bagaimana pun, ramalan bioinformatik kelarutan protein menunjukkan bahawa kelarutan ChoKs bertanda His adalah kurang daripada kelarutan purata protein *E. coli* terlarut. Ekspresi protein pada periplasma menggunakan plasmid pGEX telah ditunjukkan dapat meningkatkan kelarutan protein. Oleh itu, gen *Sa*ChoK telah disubklonkan daripada pET14b-

*Sa*ChoK ke dalam vektor pGEX. Berdasarkan ramalan *in silico*, *Sa*ChoK, *Nm*ChoK dan *Hi*ChoK bertanda GST akan lebih terlarut dan dapat dihasilkan dalam kuantiti yang lebih tinggi berbanding protein bertanda His yang dihasilkan daripada vektor pET14b. Pelabuhan molekul untuk struktur model *Sa*ChoK, *Nm*ChoK dan *Hi*ChoK dengan Hemicholinium-3 (HC-3), suatu ChoKI molekul kecil yang terbukti kesannya, telah menunjukkan mod perlekatan sempurna pada poket perlekatan kolina dan menyokong kemungkinan perencatan kompetitif oleh HC-3. Pertindanan tiga struktur model ChoK bakteria berkenaan dengan ChoK manusia menyerlahkan homologi yang ketara dan seterusnya menyokong penggunaan ChoKIs yang telah digunakan ke atas ChoK manusia untuk merencat bakteria AMR. Penghasilan pGEX-*Sa*ChoK dan ramalan bioinformatik telah membuka jalan kepada ekspresi secara optimum *Sa*ChoK, *Nm*ChoK dan *Hi*ChoK dalam sistem *E. coli*. Keputusan pelabuhan molekul menunjukkan potensi aplikasi ChoKIs untuk menentang AMR. Oleh itu, kajian ini telah merintis haluan ke arah ekspresi ChoKs bakteria terlarut untuk diuji dengan ChoKIs sedia ada dan meyerlahkan potensi perencat ini sebagai agen antimikrob.

EXPRESSION, PURIFICATION, AND CHARACTERIZATION OF PUTATIVE CHOLINE KINASES FROM MICROORGANISMS

ABSTRACT

Antimicrobial resistance (AMR) has been a menace to communities worldwide. AMR is estimated to cause ten million deaths a year by 2050. In bacteria, choline kinase (ChoK) is responsible for the synthesis of phosphorylcholine, which is a precursor for lipoteichoic acid and cell wall teichoic acid in Gram-positive bacteria. In Gram-negative bacteria, phosphorylcholine is incorporated into membrane lipopolysaccharides that modulate pathogen-host cell interactions. Choline kinase inhibitors (ChoKIs) that deteriorate the bacterial cell wall, have already been tested on *Streptococcus pneumoniae* with great results. ChoKIs activity can be also be enhanced by nanoparticles that act as a drug delivery system. The generation of the drug targets (bacterial ChoKs) in the form of recombinant proteins, is vital for testing the efficacy of ChoKIs. This study will attempt to address the issue of AMR by searching for the best conditions for the productions of recombinant ChoKs from *Staphylococcus aureus* (*Sa*ChoK), *Neisseria meningitidis* (*Nm*ChoK) and *Haemophilus influezae* (*Hi*ChoK) followed by *in silico* evaluation of these ChoKs as potential targets for ChoKIs by structural modeling and molecular docking approach. All three bacterial ChoKs were originally cloned in the pET14b vector for overexpression as His-tagged proteins. However, bioinformatic protein solubility prediction revealed that the solubility propensity of the His-tagged ChoKs tends to be less than the average soluble *E. coli* proteins. The protein overexpression in the periplasm using pGEX plasmid has been shown to increase protein solubility. Therefore, the *Sa*ChoK gene was subcloned from pET14b-*Sa*ChoK into a pGEX vector. Based on *in silico* prediction, the GST-tagged *Sa*ChoK, *Nm*ChoK, and *Hi*ChoK would be more soluble and produced at higher

yields compared to His-tagged proteins produced from the pET14b vector. Molecular docking of *Sa*ChoK, *Nm*ChoK, and *Hi*ChoK model structures with Hemicholinium-3 (HC-3), an established small-molecule ChoKI, exhibited a fit binding mode inside the choline-binding pocket, indicating promising competitive inhibition by HC-3. Superimpositions of the three bacterial ChoK model structures with human ChoK revealed an ample homology, further supporting the use of ChoKIs previously used to inhibit human ChoK on AMR bacteria. The production of pGEX-*Sa*ChoK and the bioinformatic predictions have laid the groundwork for optimal overexpression of *Sa*ChoK, *Nm*ChoK, and *Hi*ChoK in *E. coli* system. The molecular docking results demonstrate the promising application of ChoKIs to combat AMR. Therefore, this study has paved the way towards successful overexpression of soluble recombinant bacterial ChoKs to be tested with currently available ChoKIs and reveal the potential of these compounds as antimicrobial agents.

CHAPTER 1

INTRODUCTION

1.1 Introduction:

The swift adaptation of the bacteria in the revolutionary development of antibiotics made it easy for it to develop resistance to the antibiotics. With the halt of antibiotic discovery nowadays, there are no guarantees that humanity can manage to combat these adapted bacteria. This signals the looming perspective of the return of the "pre-antibiotic era" (Jayachandran, 2018). Therefore, a novel solution is needed (Draenert *et al.*, 2015; Duval *et al.*, 2019) to balance the scales and combat these bacteria, especially in this era, that is designated by the Centers for Disease Control and Prevention (CDC) as a ''post-antibiotic era'' (Gupta et al., 2019). In which the discovery of antibiotics has come to a standstill marking the "golden era" of antibiotic development as extinct (Davies, 2006; Aminov, 2010). More so, where every known antibiotic has been outmaneuvered by the bacteria via developing resistance to it (Payne et al., 2007). The stalemate in antibiotic development and the continuous antimicrobial resistance (AMR) escalation has led to the emergence of this predicament (Coates et al., 2011; Nathan and Cars, 2014). Along with the lack of research, the continuous use of the existing antibiotics has greatly stimulated AMR nowadays more than ever (Chambers, 2001; Davies and Davies, 2010; Llor and Bjerrum, 2014; Wojkowska-Mach et al., 2018). It has even been estimated that AMR will cause ten million deaths each year by 2050 (de Kraker *et al.*, 2016; Sugden *et al.*, 2016; World Health Organization, 2019). Therefore, AMR must be dealt with immediately as the Infectious Diseases Society of America (IDSA) has stated concerning AMR "one of the greatest threats to human health worldwide" (Infectious Diseases Society of America *et al.*, 2011).

No bacteria have ever shown a more remarkable adaptive example of AMR than *Staphylococcus aureus* (Figure 1.1). Along with the revolution of the "antibiotic era", came the inauguration of the AMR of *S. aureus* by first developing it towards penicillin after being introduced in the 1940s (Chambers and Deleo, 2009; Hiramatsu et al., 2014; Foster, 2017). Particularly from 1942 (Rammelkamp and Maxon, 1942), which has proven *S. aureus* to be a formidable adversary. It has given rise to a pandemic by the penicillin-resistant strain known as the phage-type 80/81 (Brodie et al., 1956). The strain and the pandemic vanished soon after the inception of methicillin in 1960 in uncertain circumstances (DeLeo et al., 2011). Soon after that, the genesis of methicillin-resistant *S. aureus* type I (MRSA-I) has been established until the 1970s (Jevons, 1961; Chambers and Deleo, 2009; Foster, 2017). Eventually, MRSA-II and III have emerged in the mid to late 1970s, marking the MRSA pandemic, then the smaller more mobile MRSA-IV surfaced in the 1990s (Chambers and Deleo, 2009).

Vancomycin (VAN) has been usually used to treat MRSA, but only as a last-line treatment and daptomycin (DAP) as well (Barros et al., 2019). Linezolid is now applied for MRSA as a clinical alternative for vancomycin (Endimiani et al., 2011). *S. aureus* has stood as the victor yet again by developing AMR towards them. Beginning with vancomycin known as vancomycin-intermediate *S. aureus* (VISA) in the 1990s, vancomycin-resistant *S. aureus* (VRSA) in 2002 (Chambers and Deleo, 2009), and daptomycin in 2003 (DAP-R) (Marty et al., 2006). VISA and VRSA are treated with daptomycin (Tran et al., 2015) and linezolid (Safa et al., 2016), however, AMR to linezolid has also emerged in 2004 (Endimiani et al., 2011). AMR in *S. aureus* has even led to elevated hemolysis with alcohol (Korem *et al.*, 2007) and biofilms to thrive on alcohol treatment (Redelman *et al.*, 2012; Luther *et al.*, 2015; Liu *et al.*, 2020).

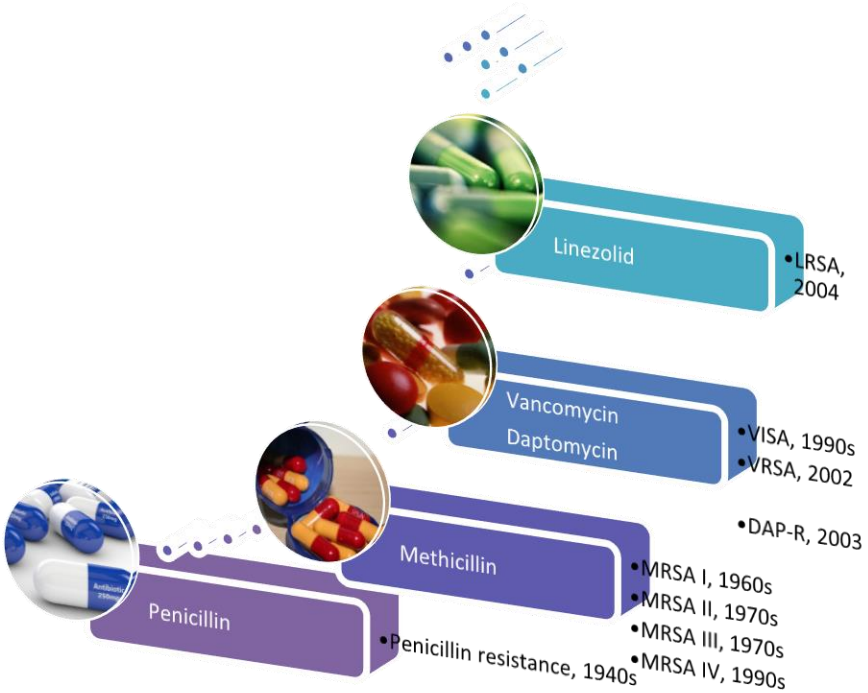

Figure 1.1 Trends of treatment and subsequent AMR in *S. aureus.*

Many bacteria other than *S. aureus* have shown remarkable resistance to antibiotics such as *Haemophilus influenzae* (Campos, 2001; Nag et al., 2001; Campos et al., 2003; Tristram et al., 2007; Yokota et al., 2009; Bae et al., 2010; Kostyanev and Sechanova, 2012), *Neisseria meningitidis* (Oppenheim, 1997; Gorla et al., 2018; Vacca et al., 2018; Zouheir et al., 2019), *Neisseria gonorrhoeae* (Unemo and Shafer, 2011; Wi et al., 2017; MacFadden et al., 2018; Cristillo et al., 2019; Rubin et al., 2020), *Streptococcus pneumoniae* (Appelbaum et al., 1977; Jacobs, 1999; Kim et al., 2016; Cherazard et al., 2017), *Streptococcus mitis* and *Streptococcus oralis* (Ono et al., 2000; Humphries et al., 2013; Adams et al., 2017; Mishra et al., 2017; van Prehn et al., 2019), and *Pseudomonas aeruginosa* and *Klebsiella pneumoniae* that are known as extensively drug-resistant (XDR) and pandrug-resistant (PDR) (Pontikis et al., 2014; Zhi-Wen et al., 2015). AMR is even considered a virulence factor for the bacteria (Aslam et al., 2018), where the outcome of infection with multidrug-resistant (MDR) bacteria is worse than the susceptible counterpart (Bodi *et al.*, 2001; Vardakas *et al.*, 2013; van Duin and Paterson, 2016; Perdikouri *et al.*, 2019). Many strategies explain that including i) the destruction or inactivation of the antibiotic molecule, ii) reducing the antibiotic within the cell via decreased penetration or extracellular expulsion of the antibiotic via the manifestation of efflux pumps, iii) structural change of the target site via a gene mutation, enzymatic reaction, or swiping the target site with a new antibiotic-invincible site (Munita and Arias, 2016). *Salmonella enterica* can even develop AMR upon acquiring mobile genetic elements (Hoffmann et al., 2017; Kudirkiene et al., 2018; Sultan et al., 2018).

These mechanisms of AMR are the reason behind the incredible speed of bacterial adaptation towards antibiotics. Even the World Health Organization (WHO) has declared AMR as the ‘growing threat’ (World Health Organization, 2014;

Lyddiard *et al.*, 2016). The community, however, have played a big part in it as well, most notably, self-medication from previously prescribed drugs (Grigoryan et al., 2007), the administration of antibiotics over the counter in developing countries (Zaman et al., 2017), and the inappropriate use of antibiotics by the community and health workers alike (Ena *et al.*, 1993; Mehrad *et al.*, 2015; Adedeji, 2016; van Duin and Paterson, 2016). To curb this predicament, public awareness has to be increased and a novel method needs to be developed. The key perhaps would be the application of novel drugs designated for eukaryotic cells with the ability to debilitate prokaryotic cells. Nanoparticles (NPs) are yet another key to the puzzle of AMR. Better yet, a combination between those two solutions would probably produce highly effective antimicrobic, impervious to AMR, even in the future.

One promising novel antimicrobic is choline kinase inhibitors (ChoKIs). Moreover, choline kinase (ChoK) is an acknowledged drug target in eukaryotes (Lacal, 2015) that exists in many species (Peisach et al., 2003). Several ChoKIs amounted to a surmountable inhibitory effect on the human choline kinase (hChoK) (Zimmerman and Ibrahim, 2017). It even demonstrated an effect on cancer (Janardhan et al., 2006; Lacal, 2008; Arlauckas et al., 2016) where hChoK is overexpressed (Glunde et al., 2005; Krishnamachary et al., 2009; Granata et al., 2014), parasites such as *Plasmodium falciparum* (Zimmerman *et al.*, 2013; Serrán-Aguilera *et al.*, 2016), and autoimmune diseases (Guma et al., 2015). In prokaryotes, many pathogens possess a putative ChoK gene, such as *S. aureus*, *Bacillus subtilis*, *Clostridium perfringens*, and *Clostridium botulinum* (Zimmerman and Ibrahim, 2017). *S. pneumoniae* also has demonstrated a confirmed ChoK activity (Wang et al., 2015) and inhibition by ChoKIs (Zimmerman and Ibrahim, 2017). That signifies the high probability of applying eukaryotic ChoKIs on prokaryotes.

1.2 Objectives:

1.2.1 General objective:

This study aims to produce pGEX plasmid constructs of *S. aureus*, *H. influenzae*, and *N. meningitidis* putative choline kinases for overexpression as GST-tag recombinant proteins. This study also aims to use *in silico* approach to predict the heterologous protein expression, protein solubility, protein structure, and protein-protein interactions to guide the overexpression of these proteins in *E. coli* and to model the structure of one of the bacterial choline kinases for molecular docking with a potential choline kinase inhibitor.

1.2.2 Special objectives:

- To produce pGEX plasmid constructs of *S. aureus*, *H. influenzae*, and *N. meningitidis* putative choline kinases
- To predict the heterologous periplasmic protein expression, protein solubility, protein-protein interactions, and protein structure of the respective bacterial putative choline kinases.
- To perform molecular docking of the putative choline kinase model structures of *S. aureus*, *H. influenzae*, and *N. meningitidis* with HC-3 and subsequently perform tertiary structure protein alignment i.e. protein superimposition with the human choline kinase.

1.3 The rationale of the study:

To test out different ChoKIs to see their effect on the ChoK of different species of antibiotic-resistant bacteria, we first need to produce the ChoK recombinant protein of these bacteria *in vitro*. Specifically, *S. aureus*, *H. influenzae*, and *N. meningitidis*.

This way *S. aureus* ChoK (*Sa*ChoK), *H. influenzae* ChoK (*Hi*ChoK), and *N. meningitidis* ChoK (*Nm*ChoK) would serve as a readymade drug target for the investigation of ChoKIs effectiveness. Therefore, expression, purification, and characterization of putative choline kinases from these bacteria are paramount to further demonstrate the promising growth inhibition activity of the ChoKIs. The first step towards ending the menace of the AMR is producing a novel drug target i.e. ChoK, in the form of a recombinant protein standing by for testing.

The discovery of new protocols for the production of an active and stable recombinant protein is paramount (Penning and Jez, 2001; Brannigan and Wilkinson, 2002) for a better understanding of selectivity and function of the protein to inhibitors (Antikainen and Martin, 2005). To achieve that, high yields of the protein must be obtained in native form, which is difficult to do from natural sources, unlike the heterologous systems such as the commonly used *E. coli* system (Ferrer-Miralles et al., 2015). As such, this study attempted to clarify and predict the correct system and protocols to use to obtain satisfactory amounts of the ChoK from these microorganisms. Therefore, several bioinformatic predictions were performed to ascertain the most probable outcome and the likely parameters to achieve it.

After the purification and verification, comes the characterization of the protein. This is a vital process where not only the produced protein is confirmed by the enzymatic activity, but data is produced that can act as a reference when testing modulators on the protein. In this case, the inhibitors are supposed to lower the enzymatic activity below the normal control protein. This is where the enzymatic assay plays an important role in the validation of the inhibitor activity on the protein.

Molecular docking has been the centerpiece for drug discovery for decades (Pinzi and Rastelli, 2019; Torres *et al.*, 2019). This study has conducted molecular docking to demonstrate the viability and vitality of the application of ChoKIs to suppress the growth of the bacteria in question. Therefore, the most vital bioinformatics tool in this study is molecular docking to realize the potential of the end goal of this research, which is the eradication of AMR. Model structures of ChoK of the respective bacteria have been produced to facilitate the procedure of molecular docking.

CHAPTER 2

LITERATURE REVIEW

2.1 Recombinant protein as "The backbone":

The expression of the protein in the heterologous systems in the form of the recombinant protein has opened wide a lot of avenues and applications for the researchers to exploit. This has made it easier to obtain an in-depth understanding of the molecular processes of living beings. It would not be an exaggeration to claim that recombinant proteins have become the backbone of the world of proteomics.

That being said, many challenges faced by expression systems, especially the most preferred one, *E. coli*, serve as roadblocks that hinder or even halt the expression (Fakruddin et al., 2013; Kaur et al., 2018). The roadblocks that require optimization to overcome include, codon biasing, vector and organism choice, the temperature of culture, induction time, formation of the insoluble inclusion bodies, inducer concentration, recombinant protein size, media additives (Kaur et al., 2018), and size of the DNA insert (Rai and Padh, 2000).

With all these limitations, *E. coli* is still the most used expression system. A vast range of advantages has made *E. coli* the first choice for expression e.g. *E. coli* permits a manipulation with simple equipment, has the potential to produce unlimited amounts of the recombinant protein, economic choice (Fakruddin *et al.*, 2013), the unparalleled density of cell culture (Shiloach and Fass, 2005), easy and rapid growth (Sezonov *et al.*, 2007; Kaur *et al.*, 2018), and simple media making in a readily available form (Pope and Kent, 1996). That is just the tip of the iceberg. All these factors have made *E. coli* being studied extensively, which resulted in the emergence of many strains fit for each specific function. The best example would be the *E. coli*

BL21 strain for protein expression that lacks proteases such as *Lon* (Gottesman, 1996) and *OmpT* (Grodberg and Dunn, 1988). Nowadays, the huge selection of the available strains has made it easier to tackle any hurdles the researchers might face to obtain an overexpression of the recombinant protein of interest and subsequent desired testing.

2.2 Pros and Cons of *E. coli* as a host:

Ever since the production of the heterologous insulin produced in *E. coli* that was first clinically deployed in 1982 (Crowl *et al.*, 1985), *E. coli* has been the go-to host for recombinant protein expression. That would be due to the many advantages *E. coli* possess. For better or worse, however, many factors contribute to this choice may help or hinder obtaining the protein in satisfactory quantities including:

2.2.1 The unique sequence of the gene:

DNA is involved in the transcriptions and translation stages of the recombinant protein synthesis inside the *E. coli* host (Fakruddin *et al.*, 2013).

2.2.1(a) Transcription stage vital sequences:

Those vital sequences involved in transcription include:

2.2.1(a)(i) The promoter:

The promoter consists of the −35 and the −10 box regions and the spacer region between them (Glick and Whitney, 1987; Fakruddin *et al.*, 2013). The optimal promoter is composed of the consensus sequence produced by the alignment of several promoters and it contains a spacer of 17 nucleotides (Glick and Whitney, 1987). For higher yields, the sequence of interest should be immediately downstream of a strong promoter (Carrier *et al.*, 1983). Regulation of this strong promoter is important to prevent the loss of the plasmid (Ringquist *et al.*, 1992). One of the most popular

promoters is T7 promoter systems present in the pET vectors, most notably, lacUV5 promoter (Rosano and Ceccarelli, 2014).

2.2.1(a)(ii) The terminator sequence:

This sequence is essential to allow the stoppage of the transcription and expression at a certain point and exists in two classes, factor-independent and factor-dependent terminators (Stormo *et al.*, 1982).

2.2.1(a)(iii) The regulatory sequence:

The regulatory sequence regulates the transcription via activation through the transcriptional activators class or inhibition via transcriptional repressors class that binds to the promoter or immediately downstream from it preventing the RNA polymerase binding to the promoter (Schumann and Ferreira, 2004).

2.2.1(a)(iv) The RNA polymerase:

The RNA polymerase is the engine driving the transcription. It consists of α, β, β′, ω, and σ components. σ being the crucial component that recognizes the promoter (Gruber and Gross, 2003).

2.2.1(b) Translation stage vital sequences:

Those vital sequences involved in translation that might hinder it is the sequence of the 5′ end of each mRNA (Fakruddin *et al.*, 2013). The initiation region of the translation comprises of four sequences: i) the start codon, ii) the Shine-Dalgarno sequence, (iii) the spacer region between those two sequences with a length of 4 to 8 nucleotides optimally, and iv) translational enhancers (Ringquist *et al.*, 1992; Schumann and Ferreira, 2004). The translation initiation region is paramount for an efficient gene expression e.g. the omittance of the Shine-Dalgarno sequence or the start

codon, blocks access to the 30S ribosomal unit, which inhibits the translation (Ramesh *et al.*, 1994).

2.2.2 Vector choice and stability:

The vector that inserts of choice has been introduced to it, is transformed into competent *E. coli* cells, and becomes the source of the foreign recombinant protein of interest (Fakruddin et al., 2013). Vectors are commonly composed of many components that define the stability such as (Rosano and Ceccarelli, 2014), replicons that contain the copy number that must be adequate and too high that does not burden the metabolic machinery (del Solar and Espinosa, 2000), promoters (Rosano and Ceccarelli, 2014), selection markers such as antibiotic resistance genes that are used to retard the growth of cells without the vector (Korpimaki *et al.*, 2003), multiple cloning sites, and fusion protein/fusion protein removal strategies (Rosano and Ceccarelli, 2014). The stability of the vector is also influenced by many other factors including the genotype of the vector and host, the size and origin of the DNA of interest (Rai and Padh, 2000), Plasmid loss in the defective segregation at cell division (Ashby and Stacey, 1984) that is commonly solved with the addition of the appropriate antibodies (Pierce and Gutteridge, 1985), elevated metabolic energy requirement for vector maintenance (Aiba *et al.*, 1982), and physiological parameters such as pH, medium constituents, and temperature (Rai and Padh, 2000). Therefore, the structure of vectors (Figure 2.1) plays an important role in the successful overexpression of the protein of interest.

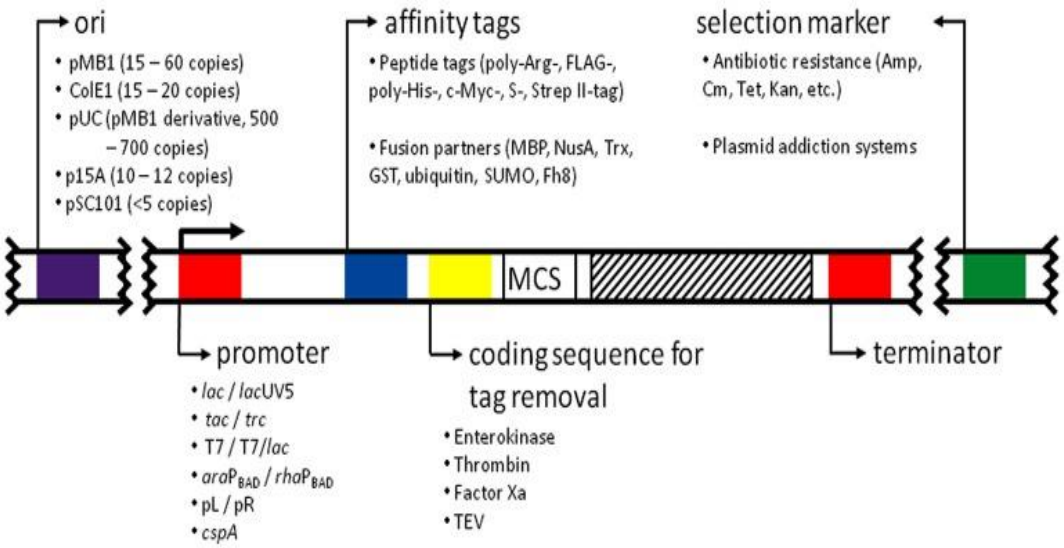

Figure 2.1 Structure of the most common expression vectors detailing the most prominent features. The structure was obtained from Rosano and Ceccarelli (2014).

The pET vectors being the most used are an appropriate choice for the expression of the recombinant protein due to the high selectivity of T7 RNA polymerase to the T7 promoter sequence, highly active RNA polymerase, and the high translation efficiency of the gene 10 5´ leader translation initiation region signals (Agilent, 2017). pET vectors express His-tagged protein that makes it easily purified with Ni-NTA affinity chromatography (Xue *et al.*, 2012; Trigoso *et al.*, 2016). pGEX vectors on the other hand, despite not being very common, are very proficient and can express proteins otherwise difficult; where it has many advantages including the possible expression of soluble proteins with easy and proper folding that are otherwise expressed as inclusion bodies, the ability to purify it with ease with anti-GST, and the simplicity of analyzing GST-tagged proteins with western blot (Rukmana and Yasmon, 2018).

2.2.3 Affinity tags vs inclusion bodies:

Affinity tags are also an important component of the vectors that contribute vastly to the stability of the vector and are vital for the purification of the recombinant protein in a soluble active form (Fakruddin *et al.*, 2013; Rosano and Ceccarelli, 2014). The goal is to form a fused or chimeric protein with a small sequence of amino acids (Nilsson *et al.*, 1997). The downside is that there is a small possibility, it might have a negative impact on the tertiary structure or biological activity (Klose *et al.*, 2004; Chant *et al.*, 2005; Khan *et al.*, 2012). The tag is preferably placed at the C-terminal to ensure the expression of the protein and in the solvent-accessible end (Rosano and Ceccarelli, 2014). The common peptide affinity tags used are the FLAG-, S-, poly-His-, poly-Arg-, c-Myc-, and Strep II-tags (Terpe, 2003). In contrast, the common non-peptide affinity tags that enhance solubility (Hammarström *et al.*, 2002) with an unknown mechanism (Raran-Kurussi and Waugh, 2012) include glutathione S-

transferase (GST) (Smith and Johnson, 1988), the maltose-binding protein (MBP) (Kapust and Waugh, 1999), N-utilization substance protein A (NusA) (Davis *et al.*, 1999), and ubiquitin (Baker, 1996). These solubility enhancers are highly efficient to the point that removing them will render the solubility of the protein of interest as unpredictable (Esposito and Chatterjee, 2006). On the other hand, if these tags are removed after expression, then the protein will attain the new higher solubility (Costa *et al.*, 2013).

2.2.4 The formation of the inclusion bodies:

The formation of insoluble aggregates due to the rapid expression is known as inclusion bodies (Betts and King, 1999). These are large and spherical bodies that result from failure to remove the misfolded proteins (Blackwell and Horgan, 1991). The common practice, in this case, is to decrease the rate of expression by lowering the temperature, pH, copy number, and fusion with a solubilizing partner (Schumann and Ferreira, 2004). The fusion partners can also be used to increase the yield of expression by fusion with highly expressed fusion partner at the C-terminus (Sørensen and Mortensen, 2005) or to increase the detection by western blot via antibody-recognizable peptide (Makrides, 1996).

2.2.5 Manipulation of the location of the recombinant protein:

The possibility to direct the recombinant protein to a specific location such as the cytoplasm, the inner or outer membrane, or the periplasmic space, has made it more advantageous to the manipulation of the expression of the recombinant protein (Baneyx, 1999). The expression of the protein in the periplasmic space allows for a higher yield due to the absence of proteases that might degrade the recombinant protein of interest (Hoffman and Wright, 1985).

2.2.6 Codon biasing:

Imagine that 61 codons encode for 20 amino acids and with variable preference i.e. frequency that reflects the pool of tRNAs in the organism (Berg and Kurland, 1997). The codons that have a large pool of tRNAs are used by highly expressed genes, while codons that have a small pool of tRNAs are used by the regulatory genes, however, the foreign gene expression is hinged on the availability of a specific tRNA (Stoletzki and Eyre-Walker, 2007). If the rare codons (least frequent) are overexpressed, it leads to the synthesis of a defective protein, even the location and amount of these rare codons can influence the translation and overall expression greatly e.g. rare codons near the promoter can stall the ribosome and halt the translations; however, this can be overcome by increasing the corresponding tRNAs or changing the rare codons with codons of high usage frequency (Chen and Inouye, 1994). The frequency even differs between species for the same codon (Table 2.1).

Table 2.1 Arginine codon usage frequency in four species. Adapted from Fakruddin et al. (2013).

Codon	***E. coli***	***B. subtilis***	***S. cerevisiae***	***Homo sapiens***
CGU	38	18	14	8
CGC	40	21	6	19
CGA	6	10	7	11
CGG	10	16	4	22
AGA	4	26	48	20
AGG	2	9	21	20

2.3 Phosphatidylcholine vs Phosphorylcholine:

The bacterial membranes are composed of many complex lipids, including phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylserine (PS), phosphatidic acid (PA), phosphatidylglycerol (PG), phosphatidylinositol (PI), cardiolipin (CL), lysyl-phosphatidylglycerol (LPG), and glycolipids (GLs), (Sohlenkamp and Geiger, 2015). Gram-positive bacteria have a unique thick murein cell wall with an underlying cytoplasmatic membrane; Gram-negative bacteria, on the other hand, owns a sandwich-like structure of an outer, inner membrane, and a thin murein cell wall in between (Raetz and Whitfield, 2002; Silhavy et al., 2010; Reichmann and Grundling, 2011). Lipopolysaccharide (LPS) constitutes most of the outer membrane (Kamio and Nikaido, 1976; Raetz and Dowhan, 1990), especially, lipid A that constitutes the framework of the outer leaflet (Raetz et al., 2007) and mediates virulence (Rietschel et al., 1982). Besides being an essential component, lipid A is an established drug target in Gram-negative bacteria (Barb and Zhou, 2008; Zhou and Zhao, 2017; Wang et al., 2018).

The infection magnitude and survival of *S. pneumoniae* hinge on the cell wall (Wang et al., 2015). Choline (Cho) is vital for cell wall integrity (Tomasz, 1967). The synthesis of phosphorylcholine (ChoP) from Cho is performed by ChoK (Whiting and Gillespie, 1996; Elswaifi et al., 2009). ChoP is necessary to produce teichoic acids (TA) in the form of cell wall teichoic acid (CTA) and lipoteichoic acid (LTA) adhered to the membrane, stipulating the essentiality of ChoP (Rane and Subbarow, 1940; Brundish and Baddiley, 1968; Mosser and Tomasz, 1970; Fischer and Tomasz, 1985; Skov Sorensen et al., 1988; Fischer, 1994; Whiting and Gillespie, 1996; Grundling and Schneewind, 2007; Seo et al., 2008; Denapaite et al., 2012; Gisch et al., 2013; Young et al., 2013). Remarkably, LTA and CTA have similar structures (Fischer et al., 1993).

Not only that, but ChoP also plays a role in the synthesis of type IV LTA of *S. mitis* and *S. oralis* (Fischer, 1997; Bergstrom et al., 2000; Seo et al., 2008; Denapaite et al., 2012; Gisch et al., 2015). LTA mediates the development of AMR to Beta-lactams, which makes it a vital virulence factor and a potential drug target (Zhang et al., 1999; Ginsburg, 2002).

LPS is essential for virulence (Zhang et al., 2013). LPS is related to the initiation of endotoxic shock, molding the protective barrier of the outer membrane (Raetz and Whitfield, 2002), and the host immune system sensitization (Silhavy et al., 2010). The virulence is capitalized by the critical addition of ChoP to LPS and TA (Galán-Bartual et al., 2015), regardless of the rarity of this procedure (Young et al., 2013). This modification of ChoP helps the survival of bacteria inside the host (Lysenko et al., 2000b; Mukherjee et al., 2011; Clark et al., 2012) and even nematodes (Harnett et al., 2010), aids in the host immunity recognition (Clark and Weiser, 2013) through C-reactive protein (CRP) (Volanakis and Kaplan, 1971), adhesion facilitation (Cundell et al., 1995; Weiser et al., 1997; Swords et al., 2000; Clark and Weiser, 2013), colonization (Lysenko et al., 2000b; Kharat and Tomasz, 2006; Clark et al., 2012), act as an adherence liaison to bacteriophage anchor and surface proteins ligands (Lopez and Garcia, 2004), and decrease genetic alteration and bacterial autolysis (Briese and Hakenbeck, 1985; Fischer, 2000).

PC also plays an important role in many processes including the formation of the bilayer structure (Sohlenkamp et al., 2003), appropriate folding of the membrane protein (Bogdanov et al., 1996; Bogdanov et al., 1999), withstanding harsh environmental changes (Sohlenkamp et al., 2003), reducing susceptibility to antibiotics that impede with the bacterial membranes (Lysenko et al., 2000b), and is

paramount for the microbe-host interactions (Aktas et al., 2010; Joyce et al., 2019). As an example, when PC is diminished in several bacteria, it demonstrated reduced virulence such as *Legionella* (Conover et al., 2008), *Brucella abortus* (Comerci et al., 2006), and *Agrobacterium tumefaciens* (Aktas et al., 2010; Aktas et al., 2014). In contrast, the virulence in *P. aeruginosa* did not undergo any change in virulence in the absence of PC (Malek et al., 2012). The selection of a novel drug target includes the consideration of the virulence factors (Maestro and Sanz, 2016). Therefore, ChoK would easily meet this requirement considering that ChoP and PC are like the building blocks of virulence.

PC and PE are the most copious major lipids in eukaryotes (Nelson et al., 2008; Gibellini and Smith, 2010; Fagone and Jackowski, 2013; Vance and Tasseva, 2013). However, PC is estimated to be present in only 15% of bacteria (Geiger et al., 2013). In eukaryotes, both PC and PE are an essential component of the cell membrane and play an important role in various cellular functions (Farine et al., 2015). For example, PC degradation is essential for the production of secondary messengers i.e. the regulation of many cellular functions (Billah and Anthes, 1990). Therefore, any disruption in the PC pathway would hinder membrane movement (Fagone and Jackowski, 2013) and cell cycle (Northwood et al., 1999).

2.4 Phosphatidylcholine and phosphorylcholine pathways:

PE and PC synthesis in eukaryotes occurs mainly via the Kennedy pathway (Bakovic et al., 2007; Gibellini and Smith, 2010). PE is synthesized via the cytidine diphosphate-ethanolamine (CDP-ethanolamine) branch, while PC is synthesized via the cytidine diphosphate-choline (CDP-choline) branch (Kennedy and Weiss, 1956). However, PE can also be synthesized by decarboxylating PS (Vance, 2008) and base-

exchange reactions with PS that also occur with PC (Sundler et al., 1974; Vance, 2008). PC is also synthesized by PE methylation (Bremer et al., 1960; Bremer and Greenberg, 1961). Additionally, PE and PC are also obtained via acylation of lysophospholipids (LPLs) (Stein and Stein, 1966; Homma and Nojima, 1982).

CDP-ethanolamine branch of the Kennedy pathway is initiated by phosphorylation of ethanolamine to ethanolamine-phosphate by ethanolamine kinase (EK) utilizing ATP, then CDP-ethanolamine is produced by ethanolamine-phosphate cytidylyltransferase (ECT) (Kennedy and Weiss, 1956; Sundler and Akesson, 1975; Gibellini and Smith, 2010; Farine et al., 2015). CDP-choline branch uses choline in similar reactions to produce CDP-choline (Gibellini and Smith, 2010). Finally, PE and PC are synthesized by ethanolamine, choline, and choline/ethanolamine phosphotransferases (EPT, CPT, and CEPT, respectively) (Gibellini and Smith, 2010; Farine et al., 2015). This last step of the CDP-choline pathway that produces PC is specifically catalyzed by 1,2-diacylglycerol choline phosphotransferase i.e. CPT utilizing diacylglycerol (DAG) in eukaryotes and CPT homologous in *Treponema* such as *Treponema denticola*, perhaps, not in any other prokaryotes (Vences-Guzmán et al., 2017). The study is needed here where it could be catalyzed by an enzyme in the bacteria other than *Treponema*.

2.4.1 ChoP and PC prokaryotic pathways:

ChoK is involved with the CDP-choline branch of the Kennedy pathway, which is present in bacteria as well (Sohlenkamp et al., 2003; Geiger et al., 2013; Sohlenkamp and Geiger, 2015; Joyce et al., 2019). This ChoP pathway is maintained by the uptake of Cho outside the cell via the choline transporter (*LicB*) (Fan et al., 2001; Fan et al., 2003; Denapaite et al., 2012; Young et al., 2013). Cho is then phosphorylated using adenosine triphosphate (ATP) to ChoP by ChoK (*LicA*) in the

cytoplasm (Whiting and Gillespie, 1996; Eberhardt et al., 2009; Elswaifi et al., 2009; Sohlenkamp and Geiger, 2015; Zimmerman and Ibrahim, 2017; Zimmerman et al., 2019). Therefore, *LicA* is critical for the ChoP pathway (Serino and Virji, 2002). Afterward, in the cytoplasm by phosphorylcholine cytidylyltransferase (CCT or *LicC*) utilizing cytidine triphosphate (CTP), ChoP is transformed into CDP-choline (Kwak et al., 2002; Zimmerman and Ibrahim, 2017; Zimmerman et al., 2019). to synthesize TA, Chop molecule is transferred from CDP-choline by phosphorylcholine transferases (*LicD*), particularly *LicD1* and *LicD2* in *S. pneumoniae* (*LicD* in *H. influenzae* transfers ChoP to LPS) to pre-teichoic acid produced by cytidylyl transferase (*TarI*) and alcohol dehydrogenase (*TarJ*) (Zhang et al., 1999; Lysenko et al., 2000a; Sohlenkamp et al., 2003; Baur et al., 2009; Denapaite et al., 2012; Geiger et al., 2013; Waldow et al., 2018). Finally, teichoic acid flippase (*TacF*) integrates TA into the cell wall and membrane (Damjanovic et al., 2007; Zimmerman and Ibrahim, 2017; Zimmerman et al., 2019). This pathway (Figure 2.2) presents an opportunity to halt ChoP and chip away at many pathogens for good.

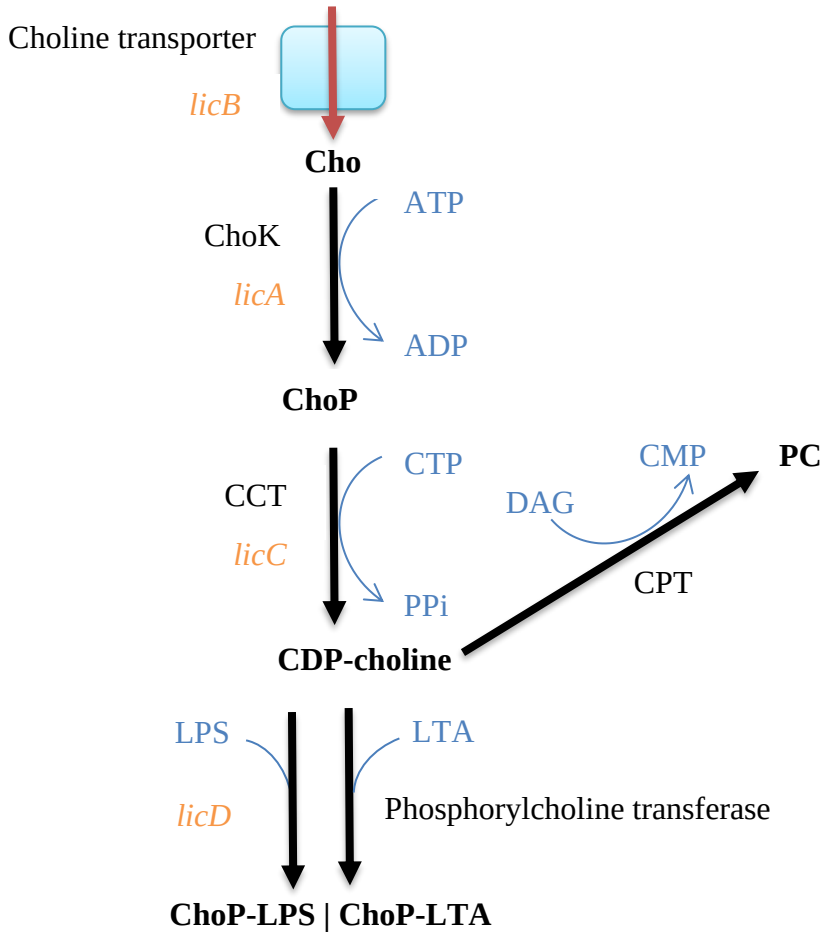

Figure 2.2 Phosphorylcholine pathway in bacteria and CDP-choline pathway of phosphatidylcholine synthesis. The genes encoding for the responsible enzymes are in orange and substrates are in blue. Abbreviations not in the text: ADP, Adenosine diphosphate; PPi, Pyrophosphate.

PC synthesis articulate in three other pathways too (Figure 2.3A-C) (Geiger et al., 2013; Sohlenkamp and Geiger, 2015; Joyce et al., 2019). The PE methylation pathway (Figure 2.3A) (Kaneshiro and Law, 1964; Sohlenkamp et al., 2003; Sohlenkamp and Geiger, 2015), where phospholipid *N*-methyltransferase (PLMT) catalyzing the methyl donor *S*-adenosylmethionine (SAM) to synthesize monomethylphosphatidylethanolamine (MMPE), dimethylphosphatidylethanolamine (DMPE), then PC from PE (Kaneshiro and Law, 1964; Arondel et al., 1993; de Rudder et al., 2000; Keogh et al., 2009; Sohlenkamp and Geiger, 2015). In some bacteria such as *X. campestris*, however, MMPE is the last product (Goldfine and Ellis, 1964; Tornabene, 1973; Moser et al., 2014). PE is synthesized in two steps, first from the collusion of CDP-DAG and serine by phosphatidylserine synthase (PSS) to form PS, which phosphatidylserine decarboxylase (PSD) decarboxylates it to PE (DeChavigny et al., 1991; Sohlenkamp and Geiger, 2015). The PC synthase (Pcs) pathway (Figure 2.3B) (de Rudder et al., 1999; Sohlenkamp et al., 2000), consists of the Pcs enzyme condensation of CDP-DAG with Cho (de Rudder et al., 1999; Sohlenkamp et al., 2000; Sohlenkamp and Geiger, 2015). The Glycerophosphocholine (GPC) pathway (Figure 2.3C) (Moser et al., 2014; Sohlenkamp and Geiger, 2015; Joyce et al., 2019) was reported only in, *S. mitis*, *S. oralis* (Joyce et al., 2019), *Xanthomonas campestris* (Moser et al., 2014), and *Saccharomyces cerevisiae* (Stalberg et al., 2008). Here, extracellular GPC (Fisher et al., 2005) is transported and then transformed to lysophosphatidylcholine (LPC) then to PC via two acyl-CoA-dependent acylations by unknown enzymes except by *X. campestris* acyltransferases *Xc_0188* and *Xc_0238* for PC (Moser et al., 2014). The first acylation enzyme was identified in yeast as GPC acyltransferase (GPCAT) (Stalberg et al., 2008; Głąb et al., 2016), but not resolved in bacteria (Rottig and Steinbuchel, 2013; Moser et al., 2014; Joyce et al., 2019).