COMPARATIVE ANALYSIS OF NUCLEIC ACID RESIDUES IN CORE STREPTAVIDIN USING THREE ISOLATION TECHNIQUES

NURUL NADIA BINTI MOHAMAD ALIAS

UNIVERSITI SAINS MALAYSIA

2024

COMPARATIVE ANALYSIS OF NUCLEIC ACID RESIDUES IN CORE STREPTAVIDIN USING THREE ISOLATION TECHNIQUES

by

NURUL NADIA BINTI MOHAMAD ALIAS

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

August 2024

ACKNOWLEDGEMENT

First and foremost, I would like to express my gratitude to The Almighty for allowing me to continue this journey of my life. To my parents, Mohamad Alias Mohamad Isa and Rosidawati Darus, and my siblings, Nur Diyana, Nurul Aina, and Nur Farah Wahida, thank you for the endless love and immeasurable support, they are always my pillar of strength throughout this journey. To my supervisor, Dr. Eugene Ong Boon Beng, I would like to extend my sincere gratitude for taking me as Eugene's lab member. I appreciate all his valuable knowledge, continuous support, and encouragement which have meant a lot to me for these past two years. To my former supervisor, Dr. Mervyn Liew Wing On, thank you for acknowledging my potential and creating such a great path for me to pursue things I have always wanted to do in this research field. I am highly indebted to all his ideas, life-sharing, and support since 2019.

I would also like to thank these precious people: My *bestie* (Khairina), my GPKs, my seniors (Ping Ying, Ai Ai, Ya Fen, Chua), my *dà jiě* (Helen), my lab members (Ru Shen, Jee Whu, Zhen Pei, Puteri, Fong), my hostel buddies (Syamimi, Izzah, Aziera, Siti), my INFORMM *chingus* (Deeza, Charlie, Huda, Zam, Boon Hui, Melvin, Daniel, Azrin), and my badminton teammates. They have filled up my days and nights with joy and laughter and thank you for this precious friendship. I would also like to thank INFORMM's staff members, especially Mr. Abdul Hakim Jamaluddin, Ms. Suhada, Mr. Qais, and Ms. Syahirah for their kind cooperation and technical assistance. Lastly, to Hakim, thank you for being the rainbow to my rain. I will never forget all the memories, experiences, and meaningful advice he has given to me outside of my research journey. Alhamdulillah for everything.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS

%	percentage/percent
^{o}C	degree Celsius
pH	potential of Hydrogen
K_d	dissociation constant
kDa	kilo Dalton
bp	base pairs
pI	isoelectric point
× *g*	relative centrifugal force
V	voltage
rpm	revolutions per minute
OD	optical density
v/v	volume to volume ratio
t	time taken
p	probability statistical value
*	asterisk
<	less than
N-	amino terminus
C-	carboxyl terminus
s	seconds
min	minutes
h/hr	hours
g	gram
mg	milligram
µg	microgram
ng	nanogram

pg	picogram
M	Molar
mM	millimolar
µM	micromolar
L	liter
mL	milliliter
µL	microliter
mm	millimeter
nm	nanometer

LIST OF ABBREVIATIONS

SAV	streptavidin
cSAV	core streptavidin
IBs	inclusion bodies
DNA	deoxyribonucleic acid
RNA	ribonucleic acid
rRNA	ribosomal RNA
dNTPs	deoxynucleotide triphosphate
siRNA	small interfering RNA
CPPs	cell penetrating peptides
mAbs	monoclonal antibodies
Glu	glutamic acid
Ala	alanine
USD	United States dollar
SDS	sodium dodecyl sulphate
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
qPCR	quantitative PCR/real-time PCR
B-4-F	biotin-4-fluorescein
SEM	standard error of mean
BME	β-mercaptoethanol
DTT	dithiothreitol
M.W	molecular weight
MWCO	molecular weight cut off
CT	cycle threshold
UV	ultraviolet light
IBFQ	Iowa Black FQ
REs	restriction enzymes
PTM	post-translational modification
BSA	bovine serum albumin
HCPs	host cell proteins
CMC	critical micelle concentration

AEX	anion exchange chromatography
HPH	high pressure homogenizer
WHO	World Health Organization
FDA	Food and Drug Administration
IPTG	Isopropyl β-d-1 thiogalactopyranoside
PMSF	phenylmethylsulfonyl fluoride
PBS	phosphate-buffered saline
EDTA	ethylenediaminetetraacetic acid
GdnHCl	guanidine hydrochloride
NaCl	sodium chloride
NaH_2PO_4	sodium dihydrogen phosphate
TE	Tris-EDTA
E. coli	*Escherichia coli*
S. avidinii	*Streptomyces avidinii*
S. lividans	*Streptomyces lividans*
B. subtilis	*Bacillus subtilis*
H. polymorpha	*Hansenula polymorpha*
P. pastoris	*Pichia pastoris*
n. sp	not specified
LB	Pre-lysis
LA	Post-lysis
WTB	Pre-wash Triton X-100
WTA	Post wash Triton X-100
WB	Pre-wash
WA	Post-wash

LIST OF APPENDICES

ANALISIS PERBANDINGAN SISA ASID NUKLEIK DALAM STREPTAVIDIN TERAS MENGGUNAKAN TIGA TEKNIK PENGASINGAN

ABSTRAK

Rekombinan streptavidin teras (cSAV) ialah keratan SAV tetramerik tidak terglikosilasi yang telah digunakan di dalam aplikasi bioteknologi secara meluas. Disebabkan kepentingan komersial cSAV, rekombinan cSAV telah dihasilkan secara menyeluruh dengan menggunakan sistem ekspresi perumah *Escherichia coli* kepada tahap ekspresi yang tinggi. Namun begitu, pengumpulan cSAV di tahap yang tinggi mengakibatkan pembentukan jasad rangkuman (IBs) tidak larut yang memerlukan kaedah pemprosesan hiliran yang berkesan untuk mengatasinya. Proses ini melibatkan pengasingan IBs daripada lisat sel menerusi kaedah gabungan pemusnahan sel, langkah pengemparan berganda, perlarutan IBs dan penstrukturan semula untuk mendapatkan cSAV yang telah distrukturkan semula dengan betul. Namun begitu, ketika penyediaan IBs, kehadiran bahan cemar selular seperti sisa asid nukleik tidak dapat dibendung dan boleh mempengaruhi proses penstrukturan semula berikutnya. Oleh itu, dalam kajian ini, proses pengasingan IBs daripada kumpulan kami sebelum ini (Proses A) telah ditambah baik untuk mendapatkan cSAV IBs yang berkualiti tinggi sebagai langkah untuk mendapatkan cSAV yang telah distrukturkan semula dengan bahan cemar minimal. Penambahbaikan ini telah dipertingkatkan dengan penggabungan tindak balas berantai polimeras kuantitatif (qPCR) untuk memantau sisa DNA. Dengan menggunakan kaedah pemusnahan sel yang telah digabungkan seperti sonikasi menyeluruh (Proses B) dan penggunaan nukleas benzonase (Proses C), cSAV IBs dengan 99% penyingkiran sisa DNA telah diperolehi. Peningkatan 10%

hasil penstrukturan semula cSAV (72%) dan 83% penurunan sisa DNA daripada penstrukturan semula 1 mg cSAV IBs telah diperiksa menerusi sonikasi menyeluruh. Walaupun penurunan sisa DNA yang amat rendah telah diamati, penstrukturan semula cSAV tidak terjejas dan aktiviti cSAV yang telah distrukturkan semula tidak terkesan. Peningkatan hasil penstrukturan semula cSAV yang dilaporkan dipercayai berkait rapat dengan penyingkiran bahan cemar selular, tidak dispesifikasikan kepada sisa DNA, tetapi selular protein juga. Proses yang telah dioptimumkan yang telah dilaporkan di sini membentangkan kepentingan mendapatkan cSAV IBs dengan kandungan bahan cemar minimal sebelum penstrukturan semula bagi meningkatkan hasil produk dan kebolehlaksanaan kaedah qPCR yang telah dibangunkan dalam memantau tahap bahan cemar asid nukleik sepanjang proses pengasingan IBs.

COMPARATIVE ANALYSIS OF NUCLEIC ACID RESIDUES IN CORE STREPTAVIDIN USING THREE ISOLATION TECHNIQUES

ABSTRACT

Recombinant core streptavidin (cSAV) is a truncated non-glycosylated tetrameric SAV which has been widely utilised in a wide range of biotechnological applications. Due to cSAV commercial importance, recombinant cSAV has been extensively produced using expression host system such as *Escherichia coli* to high expression level. Nevertheless, accumulation of cSAV in high levels results in the formation of insoluble non-functional inclusion bodies (IBs) which requires efficient downstream processing steps to overcome. This process involves isolation of IBs from cell lysates through combination of cell disruption techniques, multiple centrifugation steps, IBs solubilization and refolding to acquire correctly refolded cSAV. However, during IBs preparation, presence of cellular contaminants such as residual nucleic acids is inevitable and can affect subsequent refolding process. Hence, in this study, the IBs isolation process from our previous group (Process A) was improved to obtain high quality cSAV IBs in an effort to attain refolded cSAV with minimal contaminants. The improvements were enhanced with the incorporation of quantitative polymerase chain reactions (qPCR) for residual DNAs monitoring. By employing combined cell disruption approaches such as extensive sonication (Process B) and addition of benzonase nuclease (Process C), cSAV IBs with 99% removal of residual DNAs were achieved. A 10% increment of cSAV refolding yield (72%) and 83% reduction of residual DNAs from refolding of 1 mg cSAV IBs were observed under extensive sonication. Despite perceiving the least residual DNAs, refolding of cSAV was found not affected and the activity of refolded cSAV was not compromised. The increment

of refolding yield of cSAV reported here was believed to be caused by the removal of cellular contaminants not specifically to residual DNAs, but also cellular proteins. The optimized process reported here presented the importance of obtaining cSAV IBs with minimal contaminants prior to refolding to increase product yield and the feasibility of the developed qPCR method in monitoring levels of nucleic acid contaminants throughout IBs isolation process.

CHAPTER 1

INTRODUCTION

1.1 General Introduction

Streptavidin (SAV) is a non-glycosylated tetrameric protein that has high specificity and affinity (Kd $\approx$ 4 x 10-14 M) towards four molecules of biotin (Green, 1990). Due to its strong biotin-binding properties, the propensity of SAV to form nonspecific interactions with other molecules has therefore been greatly reduced. The extraordinarily interaction of SAV-biotin complex has then been widely exploited in various biomedical and diagnostics applications for the detection of biotinylated macromolecules such as protein and DNA (Dundas et al., 2013). The enormous popularity of SAV in molecular biology application is also due to its stability towards extreme pH (Sano & Cantor, 1990a) , temperature (Gonzalez-Montalban et al., 2008), and denaturing conditions (Sano & Cantor, 1990b; Kurzban et al., 1991).

Originally, native SAV was isolated from the culture broth of *Streptomyces avidinii* (*S. avidinii*) (Bayer et al., 1990). However, the production of SAV using its natural producers was hampered by long cultivation time which resulted in lowered productivity (Suter et al., 1988). Therefore, different heterologous expression systems ranging from prokaryotic (bacteria) to eukaryotic (fungi) expression system has been utilized to maximize the production of SAV in much shorter time (Casteluber et al., 2012; Wetzel et al., 2016). Amongst them, *Escherichia coli* (*E. coli*) is the most convenient expression system to express soluble recombinant SAV due to its short generation time, low cultivation costs, high expression level, and can be cultivated to high cell density (Sørensen & Mortensen, 2005a, 2005b). Despite offering higher productivity in comparison to *S. avidinii*, the *E. coli*-based production route is hampered

by (i) cellular toxicity that results from the natural biotin-binding properties of expressed SAV and (ii) accumulation of SAV to high levels in *E. coli* as inactive and insoluble inclusion bodies (IBs) which require an efficient downstream refolding step to overcome.

Nevertheless, due to the ineffectual processing of IBs, cellular contaminants such as host-cell proteins (HCPs) and residual nucleic acids were inevitably carried over into the final recombinant product. This has raised concern for subsequent protein refolding and downstream processing steps as cellular contaminants were previously reported to reduce refolding yield by promoting aggregation (Rudolph & Lilie, 1996). This aggregation effect can be affected by the presence of residual nucleic acids which can form strong electrostatic interaction with partially folded or unfolded hydrophobic surfaces of polypeptides (Wang et al., 2014). In addition to possible interference with protein refolding outcomes, recombinant protein with residual nucleic acids intended for *in vivo* biotechnology applications can maximize oncogenicity and immunogenicity risks (Knezevic et al., 2010). Besides, if these associated contaminants were not removed properly, they can also compromise the assay results in diagnostic applications. These have led us to reevaluate and optimize the IBs isolation steps to acquire highly purified SAV IBs with minimal nucleic acid contaminants, which in return will increase the SAV refolding yield.

Removal of residual nucleic acids in IBs can be achieved through (i) IBs isolation methods that involve cell disruption techniques comprised of mechanical and non-mechanical methods followed by multiple washing and centrifugation steps; and (ii) a refolding strategy that sometimes incorporates chromatography techniques (Middelberg, 2002). Often, most studies focus on removing associated contaminants

post-refolding to attain pure recombinant protein with increased yields. As a result, improvement in refolding strategy is frequently highlighted. Our previous group, for instance had also developed a new refolding strategy to attain higher refolding yield of core SAV (52%) by 8-16% in comparison to conventional refolding methods achieved via dialysis (44%) and dilution (36%) (Chua et al., 2018). However, instead of emphasizing on refolding strategy to obtain high yield of refolded core SAV, our focus of study was shifted towards optimizing core SAV IBs isolation process. As much as nucleic acid-free IBs are desired to result in increased yield of recombinant protein, it is also important to note that the IBs isolation process should not compromise neither the IBs structure nor the quality of recombinant protein embedded inside the IBs.

To complement the efforts of removing residual nucleic acids during preparation of IBs, a highly sensitive detection method such as quantitative real-time PCR (qPCR) is utilized to monitor the level of nucleic acid contaminants (Artika et al., 2022). Previous studies had developed qPCR assay to quantify residual nucleic acids using extracted genomic DNA from *E. coli* cultures, which was further sonicated and fragmented into a short-length DNA using restriction enzymes for assay development purposes (Wang et al., 2013; Fazelahi et al., 2017; Li et al., 2021). Nevertheless, the feasibility to assess the presence of residual nucleic acids throughout IBs preparation and protein refolding using the already established qPCR method, is to the best of our knowledge, has not been previously reported and studied in detail.

In this study, a truncated form of SAV, known as core SAV (cSAV) which offers enhanced proteolytic stability and improved structural stability was expressed in *E. coli* as IBs. The IBs isolation method reported previously from our group was optimized by applying extensive sonication that involves increased washing steps and the addition of

benzonase nuclease in lysis steps. The isolated IBs from each optimized process were solubilized and subjected to refolding by dialysis. Quantitative evaluation of total protein, cSAV protein, and nucleic acid contaminants throughout IBs processing steps were monitored and analysed using bicinchoninic acid (BCA) assay, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and real-time polymerase chain reaction (qPCR) respectively. The activity of refolded cSAV from each process was then assessed using a biotin-4-fluorescein (B-4-F) assay. While yield improvements were observed with the increased mechanical steps, the quality of refolded cSAV was not compromised by the different IBs isolation methods imposed. This improvement is expected to provide a better platform for the recovery of high-yield and high-quality recombinant protein.

1.2 Research Objectives

The main objective of this study was to remove nucleic acid contaminants during cSAV IBs preparation which will affect refolding yield of cSAV.

Specific objectives:

1. To determine the carried-over nucleic acid contaminants in refolded cSAV using previously reported IBs isolation method in reference study.
2. To determine the optimal IBs isolation method for efficient removal of nucleic acid contaminants from cSAV IBs.
3. To examine the effect of nucleic acid removal on the refolding yield of cSAV.
4. To assess the activity of refolded cSAV after nucleic acids removal.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview of recombinant protein production

Recombinant proteins are heterologous proteins produced in a host organism of different species to their origin using genetic engineering techniques and are mainly used for medical, scientific, and industrial purposes (Khan et al., 2016). The production of recombinant proteins requires cloning a gene of interest into an expression vector by transformation means under the regulation of an inducible promoter. Once expressed, the target protein is further purified and characterized to ensure its purity, structure, and functionality where ideally, a purified recombinant protein with soluble and functional bioactivity properties is desired (**Figure 2.1**) (Fakruddin et al., 2013). The increased demand for high-value recombinant proteins has paved the way for the development of various expression systems using prokaryotic cells such as bacteria, and other eukaryotic systems such as yeast, baculovirus, and mammalian cell lines (Overton, 2014). Nevertheless, efficient production of high-level recombinant genes using these expression systems is influenced by a myriad of factors such as cell growth characteristics, production time and costs, upstream and downstream production strategy, optimal expression levels at both transcriptional and translational levels, post-translational modification for correct folding, protein bioactivity, and regulatory approval (Terpe, 2006; World Health Organization, 2014).

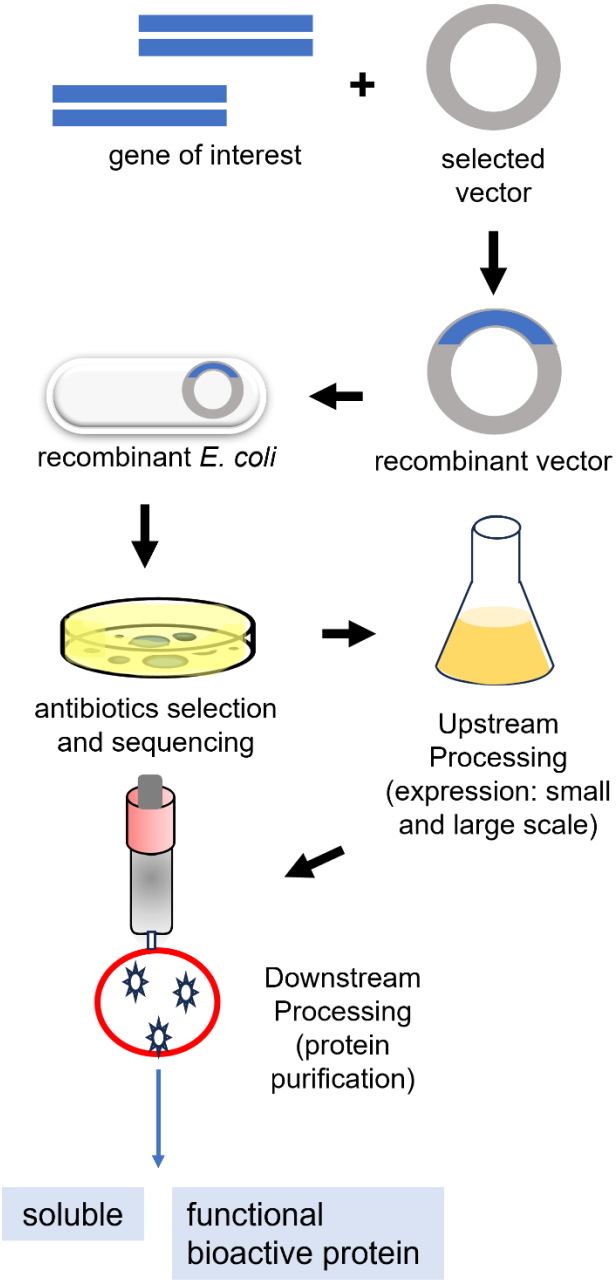

Figure 2.1 The common flow process of recombinant protein production

2.2 Recombinant streptavidin (SAV) production

Among the high-value recombinant proteins, streptavidin (SAV) is one of the commercially important proteins due to its extraordinary function and properties, which will be further reviewed in **Section 2.2.1**. This protein has been extensively produced up to industrial scale using its original host (Cat. No. S888, 434302, and 434301, Thermo Scientific) and recombinant hosts (Cat. No. 11721674001, Sigma), and commercialized in the market at an exorbitant price (≈ 220 USD per mg as in July 2024) (Merck, 2024). The exorbitant cost of SAV can be attributed to its intrinsic toxicity property and inefficient existing manufacturing processes that result in low productivity. Despite its pricey cost, SAV remains popular, in fact, from the Google Scholar literature database up to this date, about 552, 000 articles contain the term "streptavidin" in their respective research fields. The enormous popularity of SAV is due to its robust interaction with its binding molecules, biotin which greatly reduces non-specific interaction, thereby allowing this protein complex to be conveniently utilized in a myriad of biotechnology applications (**Figure 2.2**).

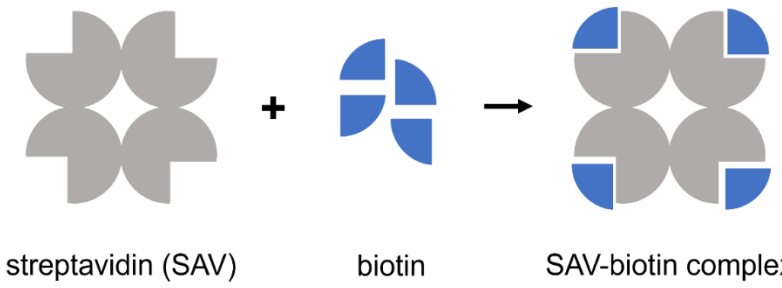

Figure 2.2 An illustration of SAV, biotin, and SAV-biotin complex

2.2.1 Properties of SAV

SAV or native full-length SAV is a non-glycosylated homotetrameric protein with a molecular mass of ~60 kDa (Argaraña et al., 1986; Bayer et al., 1990). It consists of six tryptophan residues per monomer with a conserved Trp-120 responsible for the linkage of monomers to its neighbouring subunits (Sano & Cantor, 1995). Crystallography data of SAV revealed that the tertiary structure of SAV monomer is an antiparallel β-barrel geometry composed of eight-stranded antiparallel β-sheets. The association of four identical monomeric SAV results in the formation of tetrameric SAV with four biotin-binding pockets, located at each of SAV β-barrel (Weber et al., 1987; Hendrickson et al., 1989). Since SAV contains no cysteine residues hence no disulphide bridges, its dimer structural stability and rigidity are derived from the extensive hydrogen bonds and salt bridges interactions between its monomeric units, while its dimer-dimer interaction is held together more weakly by hydrophobic interactions (Waner et al., 2004). SAV also shows a lower level of nonspecific binding to untargeted molecules due to the absence of glycosylation, resulting in its widespread utilization in biotechnological applications.

SAV exhibits high specificity and affinity towards biotin through strong non-covalent interaction (Kd = 4 × 10-4). Biotin, which is also known as vitamin H or B7, is an important micronutrient synthesized naturally inside the microbial system through *de novo* synthesis, critical for microbial survival and cell viability (Sirithanakorn & Cronan, 2021). Upon the binding of SAV to biotin, Trp-120 from the neighbouring dimer interacts with biotin, leading to the enhancement of its tetrameric stability (González et al., 1999). SAV has been found to be extremely stable at high concentrations of denaturant and across a wide pH range (Chaiet & Wolf, 1964; Waner et al., 2004). SAV is also stable at high temperatures, whereby dissociation of tetrameric

structure into monomeric subunits requires boiling at 100ºC in the presence of sodium dodecyl sulphate (SDS) detergent (Bayer et al., 1986; González et al., 1997). As a result of high stability and strong affinity to biotin, SAV has been exploited in a myriad of biotechnology applications ranging from diagnostic system to *in vivo* nanoscale applications such as drug delivery, which will be reviewed in **Section 2.2.3**.

Native full-length SAV is susceptible to proteolytic degradation by proteases that inconsistently cleave SAV monomers at both amino- and carboxyl- terminals (Bayer et al., 1989). To attain product consistency, digestion at N- and C- terminals of native full-length SAV with proteases such as proteinase K is often performed. The truncated form of SAV known as core SAV (cSAV) contains amino acid residues from Glu-14 to Ala-138 with its tetrameric molecular mass falls within 53 to 55 kDa and has a near-neutral isoelectric point, pI (5 to 6) (Bayer et al., 1989). Although in a truncated form, cSAV retains biotin binding affinity similar to native full-length SAV and offers improved stability and solubility, which were the reasons for its utilization in this study.

2.2.2 Production of SAV

Native full length SAV can be isolated from its natural producer, *Streptomyces avidinii* (*S. avidinii*) in the culture supernatants (Bayer et al., 1990). However, production of native SAV using *S. avidinii* is hampered by several factors. Cultivation of S. *avidinii* for SAV production requires five to eight days, which translates to low productivity (Suter et al., 1988). The propensity of SAV to bind to biotin molecules that are naturally synthesized and secreted by host bacteria into the culture supernatant, which can be detrimental to the cellular growth and lower the biotin-binding capability prior to application is another challenge (Wu et al., 2002). The need for proteases to remove N- and C- terminals of native SAV for product homogeneity is also cost-ineffective (Bayer et al., 1989, 1990). To overcome the challenges above, the production

of recombinant SAV using various expression system such as *Streptomyces lividans* (*S. lividans*), *Bacillus subtilis* (*B. subtilis*), *Escherichia coli* (*E. coli*), *Hansenula polymorpha (H. polymorpha*), and *Pichia pastoris* (*P. pastoris*) have been investigated and presented in detailed in **Table 2.1** (Chen et al., 2011; Casteluber et al., 2012; Noda et al., 2015; Wetzel et al., 2016).

Amongst these expression systems, *E. coli* is mostly preferred to express SAV protein due to its rapid growth (short generation time of 20 minutes), capability to be scaled-up to high-cell density and inexpensive culture medium requirements (Makrides, 1996; Baneyx & Mujacic, 2004; Terpe, 2006). As high levels of soluble SAV production is anticipated in each of expression systems above, improvement efforts were attempted using an array of techniques such as (i) modifying media composition (by adding carbon sources and nutrients), (ii) engineering host strains and plasmid construct (iii) optimizing cultivation parameters such as pH, oxygen levels and temperature, (iv) change of cultivation mode (from batch (shake-flask) to fed-batch cultivation (fermenter)), (v) use of signal or fusion protein to improve solubility, and (vi) utilizing suitable purification techniques as described in **Table 2.1**.

Table 2.1 Soluble expression of SAV in different heterologous expression system ranging from prokaryotic (bacteria) to eukaryotic (fungi)

Expression host	Type of SAV	Post-expression location	Strain/plasmid/construct design/promoter/fusion protein	Cultivation mode	Purification of SAV	References
S. avidinii	Full-length	Extracellular (secretion)	ATCC 27419	Batch (shake flask and fermenter)	• Concentration using 10, 000 M.W cutoff filter. • Iminobiotin column	(Aldwin et al., 1990)
	Full-length	Extracellular (secretion)	ATCC 27419	Batch (shake flask)	• Ammonium sulphate precipitation • Dialysis	(Cazin et al., 1988)
	Full-length	Extracellular (secretion)	DSMZ 40526; CBS 730.72	Batch (shake flask) Fed-batch (fermenter)	*n.sp*	(Müller et al., 2013)
S. lividans	Full-length	Extracellular (secretion)	Plasmid pTONA4 harbouring native full-length SAV	Batch (shake flask)	• Ammonium sulphate precipitation	(Noda et al., 2015)

B. subtilis	Full-length	Extracellular (secretion)	BE1500; BE1510 containing plasmid pBE83 secrete Lvs	Batch (shake flask)	• Ammonium sulphate precipitation • Iminobiotin column • Dialysis	(Nagarajan et al., 1993)
	Full-length	Extracellular (secretion)	WB700 PUB110-based expression vector containing native SAV, designed as pNSAV plasmid	Batch (shake flask)	• Cation-exchange column	(Wu et al., 2002)
	Full-length	Extracellular (secretion)	Engineered WB800BIO consisting of groE promoter and gluconate operator, harbouring pSSAV-T*cry* plasmid	Batch (shake flask)	• Cation-exchange column	(Wu & Wong, 2002)
E. coli	Full-length	Extracellular (secretion)	BL21(DE3) harbouring pUC8SZ plasmids	Batch (fermenter)	• Iminobiotin column	(Gallizia et al., 1998)
	n.sp	Extracellular (secretion)	BL21(DE)pLysS harbouring pET-15b-SAV construct, with domain 1 of IF2 as solubility partner	Batch (shake flask)	*n.sp*	(Sørensen et al., 2003a)

	Full-length	Intracellular	BL21(DE3) harbouring pET-SAVM4 plasmid consisting of monomeric SAV mutein M4 using T7 promoter	Batch (shake flask)	• M4-agarose affinity matrix	(Wu & Wong, 2006)
	Full-length	Extracellular (supernatant)	BL21(DE3) using leader sequence of *pho*A gene, and p616 promoter, medium supplemented with glycine	Batch (shake flask)	• Centrifugation and sonication	(Miksch et al., 2008)
	Full-length	Extracellular (supernatant)	JW1667-5 with leader sequence *bgl*A, harbouring pEL plasmid	Batch (shake flask) Fed-batch (fermenter)	• Cell disruption by centrifugation, osmotic shock, and sonication	(Müller et al., 2016)
H. polymorpha	Full-length	Extracellular (supernatant)	Methanol-inducible promoter P_{FMD}	Fed-batch (fermenter)	*n.sp*	(Wetzel et al., 2016)
P. pastoris	Full-length	Extracellular (supernatant)	Plasmid pPICZαA in frame with the *S. cerevisiae* α-mating factor secretion signal	Fed-batch (fermenter)	• Tangential flow filtration system • Iminobiotin column	(Nogueira et al., 2014)
	Core	Extracellular (supernatant)	Methanol-induced cStp gene cloned into pP1C9 plasmid	Fed-batch (fermenter)	• Immobilization in calcium alginate beads	(Casteluber et al., 2012)
	Core	Extracellular (supernatant)	X-33 strain with GAP promoter	Fed-batch (fermenter)	*n.sp*	(Müller, et al., 2016)

n.sp - not specified

While success of each techniques varies, there are several general limitations to attain high levels of soluble SAV that may result from the optimizations above. As an example, while cultivation temperature is easily optimized at small scale (shake-flask), heat transfer can be significantly complicated to be conducted with large scale fermenter. In another example, expression of target protein with a fused-signal protein may lead to inhibition of structure-activity related functions. As a result, expensive enzyme is necessitated for protein cleavage followed by subsequent purification steps to remove the solubility-enhanced signal protein. Besides, cultivation of soluble SAV always results in low productivity although improvements such as engineering its plasmid and host constructs were made to enhance its solubility. Therefore, an alternative to expressing recombinant SAV as soluble product is to allow the formation of inclusion bodies which will be further explained in **Section 2.3**.

2.2.3 Application of SAV

Since streptavidin is a tetrameric protein composed of four functional biotin-binding sites, it works as a protein-based linker to interact and bind with biotinylated molecules such as monoclonal antibodies and radioisotopes, forming a robust-combined complex (Clark et al., 2004). Its symmetrical tetravalency offers many advantages such as delivering signals resulted from surface or protein tethering as well as becoming a tool for molecular orientation (Sedlak et al., 2020).

The high affinity binding of SAV to biotin is also a reason for this protein complex to be exploited in various biotechnological applications. Amongst them are immunoassay applications that involve conjugation of SAV with reporter enzymes or tag to isolate or detect biotinylated molecules such as protein and DNA (Balzer & Whitehurst, 2023). For instance, HRP conjugated SAV was used in immunodetection systems such as ELISA, western blot analysis, and in situ hybridization imaging

applications to capture the interaction between biotinylated antibody and targeted antigen (Berg & Fishman, 2019). SAV-biotin complex system has also been incorporated in biosensors development where the sensor probe surface was coated with SAV to immobilize biotinylated capture DNA for the detection of pathogen contamination (Tombelli et al., 2005; Chua et al., 2011; Poltronieri et al., 2014).

The strong streptavidin-biotin interaction has also been widely utilized in *in vivo* applications such as drug delivery. For example, SAV has been applied for targeted delivery to increase cellular uptake of small interfering RNA (siRNA) in liver fibrosis treatments (Shukla et al., 2013). In another example, SAV was non-covalently coupled with biotinylated transporters such as cell-penetrating peptides (CPPs) conjugated with siRNA to form siRNA-CPP-biotin streptavidin complexes to promote cellular internalization of siRNA via independent-endocytosis and maximize efficient transport of siRNA into cell lines (Wierzbicki et al., 2014).

SAV has also been applied in the fields of imaging and diagnostics for cancer treatment. Radioimmunotherapy, a type of radiation therapy for cancer treatment utilizes radiolabelled monoclonal antibodies (mAbs) to target tumour-associated antigens (Barbet et al., 2012). However, direct administration of radiolabelled mAbs to tumour cells results in slow and inefficient tumour diffusion as localization of radiolabelled mAbs in targeted cells often takes several days. Additionally, mAbs accumulate at both targeted and non-targeted cells after injection. To overcome the above-mentioned problem, a pretargeted radioimmunotherapy approach was developed where biotinylated monoclonal antibodies (mAbs) against cancer cell-specific antigens were used to recruit radiolabelled SAV to deliver radiation only to cancerous cells (Boerman et al., 2003; Lesch et al., 2010). Generally, biotinylated mAbs is first

administered, this is followed by the administration of avidin to clear unbound antibodies in the systemic circulation. Radiolabelled SAV is then administered to target the cancer cells. The pretargeting strategy using radiolabelled SAV has been found to maximise exposure of radiation towards cancer tissues in comparison to normal tissues (Verhoeven et al., 2019).

Incorporation of SAV-biotin system for *in vivo* applications must undergo a strict regulatory approval. This includes assessment of safety, purity, and potency of the recombinant product. Identification, characterization, and quantification of product-related impurities and process-related impurities must be carried out and compared with multiple reference recombinant product prior to application to minimise risks associated with immunogenicity and oncogenicity (World Health Organization, 2014). Impurities such as nucleic acid contaminants during recombinant SAV preparations must be removed to a limit guideline (10 ng/dose) provided by World Health Organisations (WHO). Removal of these cellular contaminants is also critical to reduce cross contamination and false positive results in diagnostic applications. The various approaches that could be utilized to remove these associated contaminants that are present in recombinant SAV preparations is detailed in **Section 2.3.3(d).**

2.3 Inclusion bodies

2.3.1 Formation of inclusion bodies

E. coli has been widely exploited for the production of recombinant proteins at industrial scale (Baneyx & Mujacic, 2004). Utilisation of *E. coli* as a host system for expression of recombinant proteins offers advantages such as ease of cultivation, growth on inexpensive media, amenability to scale-up to high cell density and high expression levels with a short doubling time which translates to high productivity

(Sørensen & Mortensen, 2005a, 2005b). Numerous recombinant proteins which do not require complex post translation modifications (PTMs) have been successfully expressed as soluble proteins using *E. coli*. Nevertheless, expression of recombinant proteins using *E. coli* can also be hampered by the formation of insoluble aggregates known as inclusion bodies (IBs) (Thomas & Baneyx, 1996).

IBs are very dense, large refractile protein aggregates which have been observed in both prokaryotic and eukaryotic expression systems such as bacteria, yeast, and mammalian systems (Rudolph & Lilie, 1996). The structure of IBs can be classified into two types: i) highly ordered, such as amyloid fibrils and ii) disorderly amorphous structure (Wang et al., 2008). Amyloid fibrils are always associated with debilitating human pathological conditions such as Alzheimer disease and Parkinson disease (Kalia & Lang, 2015; De Strooper & Karran, 2016). On the other hand, amorphous aggregates are normally found in the bacterial cytoplasmic or periplasmic space. These amorphous IBs are usually observed as spherical or cylindrical in shape with lengths ranging from 0.2 to 1.2 μm and are composed of non-functional misfolded polypeptides (Ramón et al., 2014).

Intracellular formation of IBs is often correlated with overexpression of recombinant proteins in *E. coli*. Overloading of the *E. coli* chaperone machinery which facilitates refolding results in agglomeration of target protein folding intermediates as IBs (Jürgen et al., 2010). In addition to overexpression, formation of IBs could also be a result of reduced environment in *E. coli* cytoplasmic space which is unfavourable for the formation of proper disulphide bonds (Fahnert et al., 2004). Although the formation of IBs, maybe in some instances be undesired, there are also several advantages that can be derived from the production of recombinant proteins as IBs. The advantages are: i) accumulation of recombinant protein to high levels, especially proteins that are toxic to

host cells, ii) reduced associated cellular proteolytic degradation, and; iii) ease of isolation from cell lysate (Clark, 2001). However, many IBs need to be processed before desired structure correlated bioactivity of target protein can be attained. This often involves IB isolation, solubilisation, removal of associated contaminants and refolding steps that will be further discussed in **Section 2.3.3, Section 2.3.4,** and **Section 2.3.5.**

2.3.2 Constitution of IBs

E. coli IBs are generally composed of self-aggregated polypeptides (80-95%) of expressed target recombinant protein (Ramón et al., 2014). The self-aggregated polypeptides in IBs may consist of variable amounts of properly folded proteins or partially folded proteins (Gonzalez-Montalban et al., 2008). Host cell proteins (HCPs) such as membrane proteins, molecular chaperones, and non-proteinaceous components such as peptidoglycans, lipopolysaccharides, phospholipids, and nucleic acids have also been detected in IBs (Thatcher, 1990; Valax & Georgiou, 1993). Groups of membrane proteins such as OmpF, OmpC, OmpA and OmpT have been co-isolated with expressed target protein from IBs (Jürgen et al., 2010). Molecular chaperones that are responsible for assisting correct protein folding such as IbpA and IbpB as well as DnaK and GroEl have also been found but in minimal concentrations (Carrió & Villaverde, 2002). Amongst the nucleic acid materials that have been co-purified and precipitated together with IBs are genomic DNA, plasmid DNA, ribosomal RNA and DNA (Krachmarova et al., 2020). Both 16S and 23S rRNA have also been extracted from IBs. The composition of *E. coli* IBs consisted of cellular contaminants and recombinant proteins conformations are further presented in **Figure 2.3**.

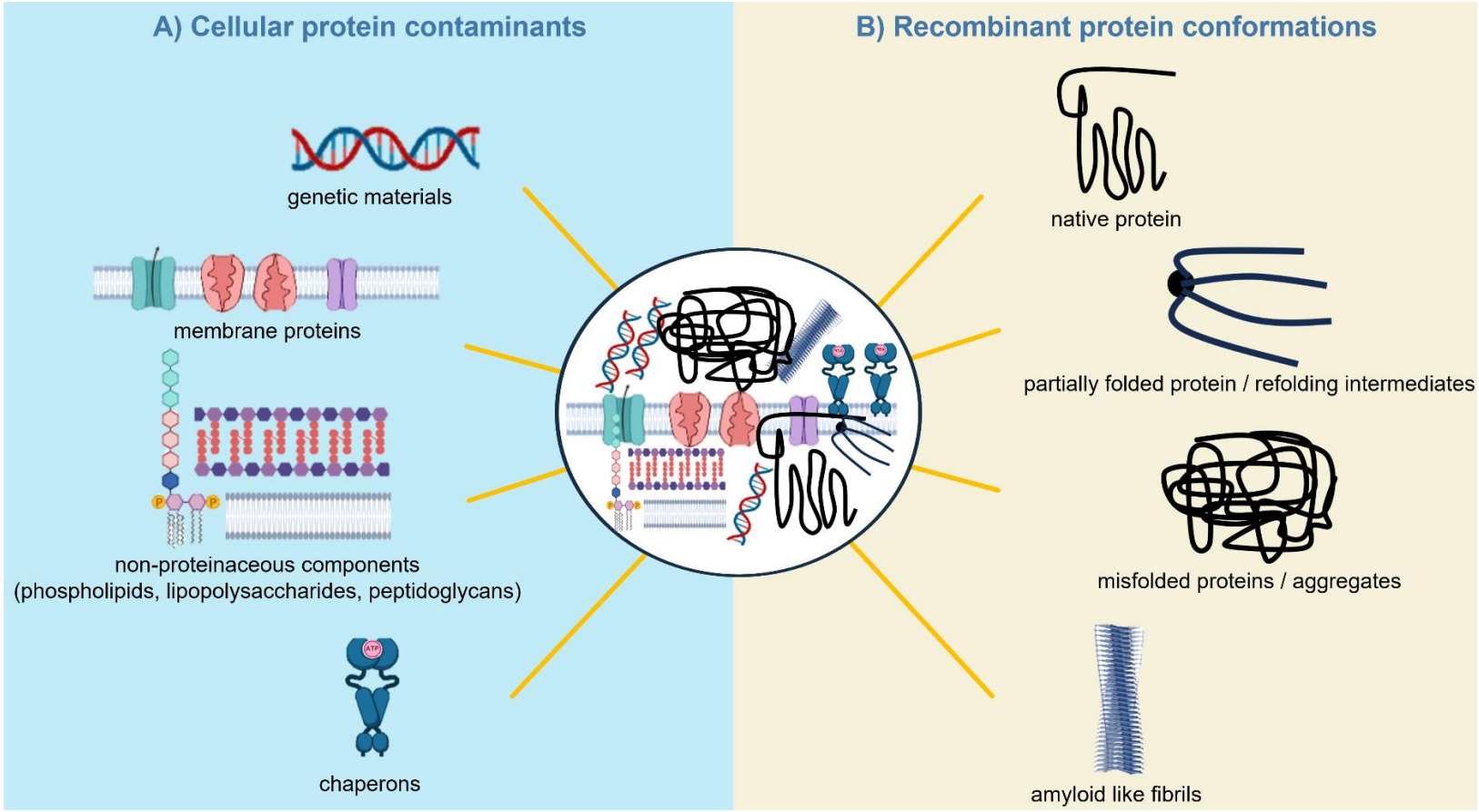

Figure 2.3 IBs composition in the cytoplasmic of *E. coli* (A) Cellular protein contaminants are composed of bacterial genetic material, membrane proteins, chaperones and non-proteinaceous components. (B) Recombinant protein conformations are composed of native, misfolded, partially folded proteins, and amyloid like fibrils.

2.3.3 Isolation of IBs

The presence of various cellular components as described in **Section 2.3.2** during IBs isolation is the result of cell disruption process (Rathore et al., 2003). Since SAV is expressed intracellularly which is inside the cell cytoplasm, disruption of cells is necessitated to break open the cell and release cellular components specifically the desired expressed protein. Disruption of cells can be performed by using two different approaches: (i) non-mechanical and (ii) mechanical approaches. Once cells are destroyed and ruptured, IBs can be further isolated through centrifugation step to remove undesired cellular contaminants which have been released and solubilized into the supernatant of IBs suspension.

2.3.3(a) Non-mechanical approaches

In some IBs preparation especially involving recombinant proteins that are sensitive to environmental changes, incorporation of mild and inexpensive non-mechanical approaches such as physical and chemical methods are sufficient to isolate IBs. Physical method includes thermal lysis that involves repeated freezing and thawing of the cells at room temperature (Johnson & Hecht, 1994). This repeated action can cause swollen of microbial cells due to the formation and contraction of ice crystals during freezing and thawing process, which will eventually break the cell wall and disrupt cell components. Meanwhile, chemical method uses additives such as buffering salts, detergents, glucose, metal chelators, metal ions, chaotropic agents (urea, guanidine, and EDTA), reducing agents, and protease inhibitors (PMSF) in the lysis and wash buffer (Seddon et al., 2004; Leibly et al., 2012; Peach et al., 2012). These additives are equally important to regulate the pH and osmolarity of cell lysate, enhance protein solubility and stability, and facilitate microbial cells disintegration. Among the above additives, detergents are the main ingredient that determine the lysis strength of a

prepared buffer (Massiah et al., 2016). Since the outer membrane of *E. coli* is composed of both lipid bilayers and peptidoglycans, detergents are often used in combination with lysozymes to disintegrate the hydrophobic-hydrophilic interactions between lipids and proteins of the cell wall and membranes. Detergents, also known as surfactants, can be divided into three types which are cationic, anionic, and non-ionic detergents. Amongst all, anionic detergent such as sodium dodecyl sulphate (SDS) and sarcosyl, as well as non-ionic detergents comprised of zwitterionic detergent, Triton X-100 and Tween series are commonly utilized detergents during cell lysis (Hadj Sassi et al., 2017; Ko et al., 2021). SDS is a strong lysis agent and can rapidly denature protein structure. However, due to its strong properties, SDS can disrupt protein functionality. As a result, a non-ionic detergent such as Triton X-100 which offers non-denaturing protein condition is preferred over SDS during cell disruption process.

Apparently, Triton X-100 has the ability to disrupt lipid-lipid and lipid-protein interactions, sequestered to, and separated recombinant IBs from cellular contaminants by forming spherical structures known as micelles. The formation of micelle consisting of protein and detergent is due to the physiological characteristics of Triton X-100 having a very low critical micelle concentration (CMC) (~0.3 mM, ~0.02% v/v) (Cui et al., 2008). Nevertheless, micelles need to be removed as this protein-detergent complex was found to further inhibit proper refolding due to their strong interactions with refolding intermediates. The removal of detergent is usually done via additional purification steps such as chromatography techniques and dialysis (Opitz et al., 2015). However, since any chromatography approaches were not attempted throughout this study and dialysis being unsuitable approach to remove Triton X-100 due to the high molecular mass of formed micelle, subsequent washing steps without Triton X-100 are often integrated in the IBs isolation steps, where this additional step was carried out

herein. This step is performed by resuspending IBs pellets with non-detergent wash buffer to decrease the detergent concentration below its CMC, allowing dispersion of micelles into its detergent monomers thus removing micelle from contaminating IBs (Yang et al., 2011; Mohseni et al., 2016; Ni et al., 2016; Maksum et al., 2022). In this study, buffering salts, sucrose, Triton X-100, EDTA, and PMSF were utilized as the additives in the lysis and wash buffer to enhance the isolation process of cSAV IBs.

Previous studies have isolated protein IBs using the combination of lysozyme and detergents, however, since lysis of bacteria cells often leads to the release of nucleic acid contaminants which greatly increase the lysates viscosity and can complicate subsequent processing steps, nucleases such as DNase and RNase were frequently added (Bajaj et al., 2015; Piao et al., 2016; Linova et al., 2020). These enzymes have the capacity to hydrolyze the phosphodiester bonds of the DNA and RNA respectively. Addition of another nucleases, known as benzonase nuclease can also eliminate nucleic acids contamination by degrading all forms of DNA and RNA (Singhvi et al., 2020). This nuclease has been widely used in downstream processing of recombinant protein due to the absence of proteolytic activity and its specific activity and capability to destroy any recombinant DNA molecules that might contaminate the target protein (Yang et al., 2013). Nevertheless, the utilization of these biological enzymes during cell lysis is exceptionally expensive and they are often coupled with detergent to attain complete cell lysis. Besides, the addition of enzymes may complicate purification steps as they have to be removed from the recombinant products especially when nucleases-free product is desired. Even so, these nucleases can be removed by subsequent washing and centrifugation steps. Their activity can also be inhibited and subsequently lost when IBs are reconstituted with high concentration of denaturants and salts, which is commonly achieved during IBs solubilization step.

2.3.3(b) Mechanical approaches

Apart from non-mechanical approaches being the convenient method to isolate IBs, mechanical approach has also been extensively used in the recombinant protein production. High-pressure homogenizer (HPH) and sonicator are the two popular mechanical tools that utilize force to homogenize samples without damaging the desired intracellular components. HPH is a powerful mechanical equipment as high pressure is applied when microbial cells in suspension are forced through an orifice valve creating a high shear force that result in cells disruption (Gomes et al., 2020). Some studies have utilized HPH to isolate *E. coli* recombinant IBs, however, since the system units are expensive, heavy and cumbersome, its utilization is frequently adopted for large scale production, not in the small lab-scale setting (Humer et al., 2020; Xi et al., 2020; Reddy Patakottu et al., 2023). Besides, this equipment is inconvenient for applications that require processing of multiple samples as the cleaning process is labour-intensive and time consuming. Therefore, an alternative to this method is by subjecting cell suspension to sonication.

Sonication is a laboratory technique that has been effectively used for the disintegration of microbial, plant and animal cells (Huh & Kim, 2003). Previous study has reported that sonication was the best method to completely lyse various types of microbial cells such as *E. coli* BL21 (DE3) (bacteria), *Leucosporidium muscorum* (yeast), and *Aspergillus niger* (fungi) for the release of L-asparaginase enzyme. Besides, as of today, regardless of new technological advancements, most laboratories have reported using sonicator to isolate various types of recombinant proteins (Chung et al., 2015; Upadhyay et al., 2016; Ashcheulova et al., 2018; Zhou & He, 2020). This technique produces a high frequency sound wave (15-20 kHz/s) which is then converted into mechanical oscillations by a transducer through a sonicator probe that is immersed

into the cell suspension. The high energy sound waves cause the release of complex byproducts and micronization of cell debris which need to be separated and removed during IBs preparation by centrifugation (Costa-Silva et al., 2018). Since dissipation of mechanical energy from sonication causes generation of heat that can induce sample degradation, intensive cooling of samples during and after the operation is required by incubating the sample on ice or performing cell disruption at low temperatures.

2.3.3(c) Combination of cell disruption approach

As much as contaminants-free IBs are desired, it is crucial to note that the utilization of cell disruption methods should not compromise the applicability of IBs neither their structure nor the quality of the final recombinant product (García-Fruitós, 2010). Previous studies have explored the effectiveness of cell disruption methods ranging from mechanical approach such as HPH and sonication to non-mechanical approach involving the use of physical, chemical and biological lysis (Rodríguez-Carmona et al., 2010). The obtained results show that mechanical methods improved the lysis process for most of the cells, but the quality of recombinant proteins that are trapped inside the IBs was compromised. On the other hand, while non-mechanical methods are gentle toward IBs, the methods when incorporated alone were not effective enough to completely disrupt the cells and release the desired intracellular components (Peternel & Komel, 2010). Therefore, they have concluded that combination of both mechanical and non-mechanical approaches could be a suitable approach to maximize the cell lysis process while not compromising the desired recombinant proteins. To achieve maximum cell disruption, repeated steps of lysis and washing, followed by multiple centrifugation steps are frequently integrated during IBs isolation (Zakharova et al., 2015; Lipničanová et al., 2020; Li et al., 2021; Restrepo-Pineda et al., 2022) .