THE UNIVERSITY
of ADELAIDE

A STATISTICAL METHODOLOGY TO OPTIMISE SOLUBLE PRODUCTION OF VIRUS-LIKE PARTICLE PROTEIN BY FERMENTATION OF *ESCHERICHIA COLI*

by

Nhat Hoang HUYNH, BE(Hons)(ChemPharma)

School of Chemical Engineering
The University of Adelaide

A thesis submitted for the examination for the degree of
Master of Philosophy
April 2021

STATEMENT OF DECLARATION

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, Nhat Hoang Huynh, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I give permission for the digital version of my thesis to be made available on the web, *via* the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Signature:

Date: 02-04-2021

EXECUTIVE SUMMARY

Hepatitis B core protein (HBc) virus-like particle (VLP) is a self-assembled nanoparticle that resembles the native conformation of a virus without viral genome. HBc has been widely applied as a vaccine production platform to present foreign antigens.

Escherichia coli is recognised as an industrial microbial factory that offers high yield of products and economics of process. However, poor protein folding attributes to the accumulation of newly synthesised recombinant protein in inclusion bodies (IB). Complicated purification process, low yield and high cost are the consequence of IB expression from fermentation. Therefore, the soluble intracellular protein expression with correct conformation of HBc-VLP is desirable in fermentation process.

As biological processes involve many factors and reactions concomitantly, a conventional approach of one-factor-at-a-time is not suitable. However a statistical approach, also known as design-of-experiment, offers an improved means to determine the effects of multiple factors to achieve the most advantageous setting.

A research program was therefore undertaken with the aim to apply statistical approach to investigate the fermentation process factors impact soluble expression of chimeric HBc VLP when they carry foreign epitope in *E. coli* fermentation. Additionally, findings extrapolated from a careful statistical approach are necessary for optimised cultivation at fermenter scale.

A logical and stepwise approach was implemented as a research strategy. This is the first time that design of experiment is used to boost the soluble expression of HBc VLP.

Two (2) VLP protein models includes HBc carrying Hepatitis C virus and HBc carrying Epstein–Barr virus nuclear antigen which are listed as EBNA1-HBc and HCV-HBc.

Fractional factorial design (FFD) was applied to study the effects process factors on the soluble expression of chimeric HCV-HBc in shake-flask fermentation. The greatest yields achieved were 89.7 mg g^{-1} dry cell weight (DCW) and 84.4 mg from one (1) L of culture media. The important process factors were ranked as 1) cell density at induction, 2) post-induction rotation speed of shaker, and 3) post-induction temperature.

Similar to HCV-HBc case, the significant factors for the soluble expression of EBNA1-HBc were determined from FFD. These process factors, particularly cell density at induction, post-induction rotation speed of shaker, and post-induction temperature and were optimised by response surface methodology (RSM). The highest volumetric yield and the

cellular yield achieved were 272.0 mg L^{-1} of culture media and 210.5 mg g^{-1} DCW, respectively.

It was found that using Terrific Broth (TB) as culture media provided optimal conditions for EBNA1-HBc production at an induction time of six (6) h. The production of EBNA1-HBc was scaled up to 5 L fermenter cultivation using the optimal setting achieved from RSM and TB media. Results in fermenter-scale agreed with findings in shake-flask, showing that the soluble yields decreased as the culture was induced at greater cell density.

It was hypothesised that the nutrient in the media is depleted when culture reaches to higher cell density, limiting the protein synthesis. Two (2) feeding strategies were applied in fed-batch mode, namely, 1) constant feeding and 2) dissolved-oxygen (DO) stat feeding. Results showed that both feeding strategies improved productivity of EBNA1-HBc in case of induction of OD_{600} of 20. Constant feeding is significantly improved over DO-stat-based feeding. The greatest yield achieved from constant feeding was 1800 mg L^{-1}, which is greater than from DO-stat-based feeding of 2.57 x times. By using constant feeding strategy, EBNA1-HBc cost per media therefore decreases by a significant 10 %, compared with batch-mode.

Additionally, it was found that the thermal stability of ENBA1-HBc is beneficial to couple with SEC to quantify the EBNA1-HBc capsid in clarified crude lysate.

Results from this research confirm that the newly applied statistical approach is a practical tool to determine the impact of process factors on fermentation and to provide robust data optimise conditions. Process factors, media composition, induction time and feeding strategy contribute to the soluble expression of chimeric HBc-VLP and need to be optimised to boost productivity. However there are present limitations of this research that need to be determined in future work, namely 1) quantitative variation of complex media between batches, 2) an explanation of why this is the optimal combination of factors, and; 3) the reason why carbon limitation from DO-stat-feeding limits the production.

The justification for this is that these findings will aid a detailed understanding of process factors that contribute to soluble expression of chimeric HBc VLP. Not only navigate the design space, but this application of statistical approach is also a general framework for process development and scaling up.

This research is original and not incremental work. Findings will be of direct interest and benefit to recombinant protein production in *E. coli* in general and HBc platform in particular, and the wider biopharmaceutical industries.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Associate Professor Jingxiu Bi, my principal supervisor from the School of Chemical Engineering and Advanced Materials, The University of Adelaide, for her dedicated support, patience, and time. I am grateful to her constructive guidance and help throughout my candidature. The good advice, support, and friendship of my second supervisor, Dr. K R (Ken) Davey, FIChemE, CEng, CSci, are precious to me. Thank you for supporting me in writing. I would also like to thank Prof. Bo Jin for providing the facility at the early stage of my project. It is a pleasure to work with you.

I am grateful to the all the staff members in School of Chemical Engineering and Advanced Materials for their support, both in administration and laboratory services.

I acknowledged the significant support from the Australian Government Research Training Program, for providing me the opportunity to undertake my Master of Philosophy.

My gratitude to my colleagues, Bingyang Zhang, Shuang Yin, Afshin Karami, Lukas Gerstweiler, Thai Thao Ly, Yiran Qu and Yechuan Zhang. It was great to share the journey with you. It was challenging but at the same time, the most interesting and rewarding experience.

I am greatly indebted to my wife, Tien Pham, my sister, Thanh Huynh, my parents and my friends, who have given me unconditional love.

TABLE OF CONTENTS **PAGE**

LIST OF FIGURES **PAGE**

LIST OF TABLES **PAGE**

ABBREVIATION

HBc	Hepatitis B core protein
VLP	Virus-like particle
EBV	Epstein–Barr virus
EBNA1	Epstein–Barr virus nuclear antigen
HCV	Hepatitis C virus
MIR	Major immunogenic region of HBc protein
OFAT	One-factor-at-a-time
FFD	Fractional factorial design
RSM	Response surface methodology
CCD	Central composite design
BBD	Box-Behnken design
DoE	Design of experiment
IPTG	Isopropyl β-D-1-thiogalactopyranoside
OD_{600}	Optical cell density measured by spectrometer at wavelength of 600 nm
IBs	Inclusion bodies (insoluble aggregates)
SDS-PAGE	Sodium dodecyl sulphate - polyacrylamide gel electrophoresis
NAGE	Native agarose gel electrophoresis
HPSEC	High-performance size exclusion chromatography
TEM	Transmission electron microscopy
ANOVA	Analysis of variance
df	Degree of freedom
LB	Luria-Bertani media
TB	Terrific Broth media
SB	Super Broth media
2YT	2 times yeast extract & tryptone media
DO	Dissolved oxygen level in the culture
DO-stat	Dissolved oxygen static-based feeding strategy

CHAPTER 1

INTRODUCTION

1.1 Background

Hepatitis B virus (HBc) is reckoned one of the most versatile platforms to present heterologous antigen. It can be expressed in different expression hosts, both prokaryotic (Sominskaya et al., 2010) and eukaryotic (Freivalds et al., 2011).

Escherichia coli provides a beneficial means because of inexpensive production and high yield (Middelberg et al., 2011). However, the lack of post-modification machinery in *E. coli* is a major drawback. Expressing significant amounts of recombinant protein can overwhelm the cellular metabolism of *E. coli*, impacting protein synthesis and protein folding (Rosano and Ceccarelli, 2014). Formation of misfolded protein often occurs, and results in low levels of soluble production. Recombinant proteins need to be expressed as native form to maintain high bioactivity.

This acknowledged accumulation of insoluble aggregates is a controversial subject of recombinant protein expression in *E. coli* (Markossian and Kurganov, 2004). It is feasible to recover the target protein from inclusion bodies *via* solubilisation and refolding, but these extra stages require substantial cost and, place financial pressure on process economics. Moreover, it is not guaranteed that purification from inclusion bodies is able to maintain the biological activity of the recombinant protein.

Inclusion bodies accumulation is common in chimeric HBc-VLP expressed in *E. coli*. To boost soluble production of HBc-based VLP in *E. coli* process parameters need to be optimised to minimise the physiological stress to the cell and for conditions to target protein synthesis (Kaur et al., 2018). In conventional experiment, process factors are optimised sequentially, one-factor-at-a-time. This approach limits the possibility to determine the most advantageous combination of factors because the interactions between factors are lost. Therefore, to understand quantitatively the effects of fermentation process variables on soluble production of HBc-based VLP, statistical approach such as factorial design and response surface methodology should be employed.

Additionally, findings extrapolated from a careful statistical approach are useful to optimise the production at both shake-flask and fermenter scale to improve the process efficiency and process economics.

A research program was therefore undertaken to apply statistical approaches, including fractional factorial design (FFD) and response surface methodology (RSM) to determine the effects of fermentation process factors on soluble production of two (2) chimeric

HBc-VLPs, EBNA1-HBc and HCV-HBc in microbial system in small-scale. The optimal combination of factors was then applied for larger, fermenter scale cultivation.

1.2 Research aim and objectives

Despite HBc-based VLPs having been intensively studied as a practical vaccine platform, the effects of process factors and their interaction on high yield of soluble expression in *E. coli* remain largely illusive.

The overarching aim is to apply the statistical approach to determine the effects of process factors in order to boost the soluble production of chimeric HBc-VLP in the microbial system. To gain technical insights for chimeric HBc-VLP soluble production, the overarching objectives of this research are to:

1. Use fractional factorial design to quantify the impact of process parameters on production of soluble of HBc-VLP carrying epitope HCV
2. Determine of the impact of process parameters and enhance the soluble production of HBc-VLP carrying epitope EBNA1, using fractional factorial design and response surface methodology
3. Optimise the batch cultivation and fed-batch cultivation based on the findings in small-scale to boost soluble production of HBc-based VLP in fermenter-scale.

A justification for this research is that it will aid greater understanding of factors that contribute to the production of HBc VLP in *E. coli* in terms of vaccine carrier. A better understanding will lead to reliable process development for recombinant protein production and increase chances of commercialisation.

A stepwise and logical approach is adopted, as is presented in Fig 1.1. The research starts with screening experiment of two (2) chimeric HBc-VLPs to navigate the design space in Task 1 and Task 2A. Optimisation experiment is used to obtain the optimal process performance in small-scale cultivation in Task 2B. Scaling up fermentation experiment is followed in Task 3 by both batch and fed-batch cultivation.

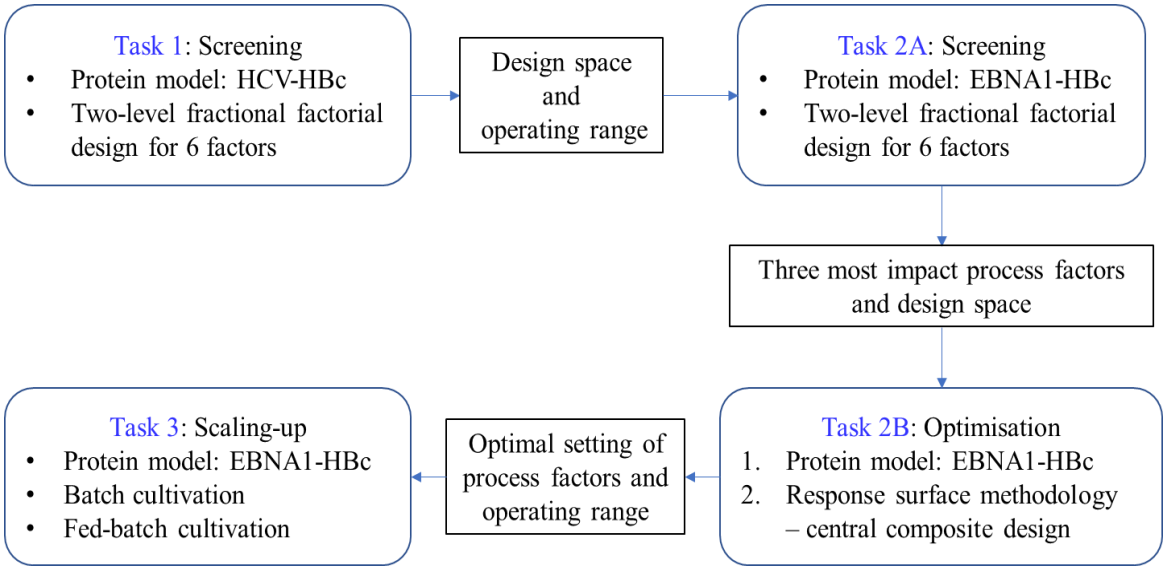

Fig. 1-1: Outline of research.

1.3 Thesis outline

Chapter 1 presents the background, aim and objectives, and thesis outline.

Chapter 2 provides a critical review of HBc-VLP and relevant literature for soluble intracellular protein production in microbial systems, plus potential problems and application of statistical approaches. It is proposed here that statistical approach is able to boost the expression level of recombinant protein expressed by *E. coli* in shake-flask fermentation by optimisation of cultivation process factors.

In Chapter 3 the impact of fermentation process factors on soluble production of HBc-VLP carrying epitope HCV using fractional factorial design is presented. Although key post-induction process factors are identified, namely, 1) the cell density of induction, 2) temperature, and 3) rotation speed of shake, a drawback soon acknowledged is that FFD only investigates two-level of each factor so further optimisation experiment is necessary.

In Chapter 4 the optimisation of soluble production of HBc-VLP carrying epitope EBNA1 in *E. coli* using two steps of statistical approach is presented. The key fermentation process factors were identified from FFD, including the cell density of induction, expression temperature, and shake-flask rotation speed. These factors were further optimised by response surface methodology (RSM) to improve the fermentation process performance.

In Chapter 5 based on the significant fermentation process factors screened by shake-flask fermentation in Chapter 4. The production of soluble EBNA-HBc was further

implemented and optimised in 5 L working volume of 14 L laboratory-bench scale fermenter with automatic control system. This is undertaken because of the differences in cell density at induction between large and small scale. The impact of cell density at induction point and the importance of feeding strategy were further investigated. It is found that fed-batch with constant feeding achieved gram-per-litre scale of volumetric productivity. Furthermore, a brief media cost analysis is made to determine process economics.

The findings and conclusions from this research in promoting soluble production of HBc-based VLP are given in Chapter 6.

Results from this original research will provide technical insights for HBc VLP production in *E. coli*. Navigating the design space of the fermentation process factors in a systematic way is crucial to develop a practical production strategy from shake-flask to fermenter-scale, and; more widely, to researchers interested in improving knowledge of recombinant protein bioprocessing.

A definition of some important terms used in this research is presented in Appendix A. Details of related buffer composition are presented in Appendix B. A set of supporting data obtaining from this research is given in Appendix C.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Globalization has increased the speed of transportation. Despite economic advantages, there are accompanying risks of infectious disease spreading, especially human-to-human. Current COVID-19 global pandemic is urging world-leading research activities, especially in safe and efficient vaccine development.

Vaccination is an effective means to prevent infectious disease. Since it was introduced, vaccination has improved the quality of public health highly significantly. However conventional vaccines, both live-attenuated and inactivated, have drawbacks (Greenwood, 2014). These include the efficiency in controlling multiple strains of infectious pathogen, risk of infection and side effects, plus stability during transportation, and manufacturing cost. Therefore, there is a need to create a vaccine that is rapid to produce, safe, stable, highly immunogenic, cost-effective, and with life-long immunity from a single dose.

Virus-like particle (VLP) is a new generation of vaccine that is potentially safer than conventional vaccines. This is because of the lack of viral genetic materials (Roldao et al., 2010). Amongst various VLP platform, Hepatitis B core (HBc) protein is one of the most promising due to its flexibility to present heterologous antigen on the capsid surface (Roose et al., 2013). A significant feature of VLPs is that they can be expressed in many expression systems, both eukaryotic and prokaryotic.

Expression of a large quantity, soluble, functional, and properly folded recombinant protein is essential because of the requirements in biochemical characterisation and production process. Even with good pre-clinical results, a vaccine candidate will face significant difficulties to reach a commercial market if it is unable to be produced in large quantity.

Escherichia coli has been used intensively to produce recombinant protein. It possesses a high-level of expression, low cost and scale-up ability (Chuan et al., 2014). High level of expression introduces physiological stress on *E. coli,* however affecting the protein synthesis and protein folding. This eventually leads to expression as an insoluble form (Fahnert et al., 2004). Because of lacking post-translation modification mechanism, expression of recombinant protein in *E. coli* as soluble with native conformation is highly regarded. Not only maintaining the specific immunogenicity, but soluble expression also simplifies the recovery and purification process, ultimately reducing production cost.

In this chapter, a detailed review of HBc VLP as vaccine carrier advantages, recent developments, applications, benefits and limitations, in comparison with conventional vaccines is presented. Additionally, two (2) epitopes HCV core and EBNA1 are critically reviewed.

E. coli is reviewed as microbial factory for protein production. This includes metabolic aspects, protein production processes, potential problems, and solutions. The significance of target protein folding state is highlighted and the process factors that influence the soluble intracellular protein are presented. As biological processes involve interconnection between factors and reactions, statistical approaches are reviewed.

To conclude this chapter, a statistical approach consisting of screening experiment and optimisation experiment is selected for soluble production of chimeric. The effects of process factors are investigated for each chimeric HBc VLP in shake-flask cultivation before cultivating in fermenter scale.

2.2 Conventional vaccines versus virus-like particle vaccine

The first successful conventional vaccine was the smallpox vaccine, introduced by Edward Jenner in 1796 (Riedel, 2005). It is noteworthy that conventional vaccines provide protection against rubella, polio, tetanus, etc., and resulted in the complete eradication of smallpox in 1980 (Fenner et al., 1988) and, rinderpest in 2011 (Horzinek, 2011).

There are two types of conventional vaccine, live-attenuated and inactivated vaccine. Live-attenuated vaccines consist of viable bacterial cells or viral particles, which have been reduced in virulence but still provoke an immune response. For example, pathogens can be attenuated by culturing the pathogen in less optimal conditions, forcing them to replicate repeatedly in tissue culture. Another approach is introducing them into different hosts, where the capacity of replication is limited (Mak and Saunder, 2006).

Inactivated vaccines are non-infectious because the infectivity of the pathogen has been destroyed by chemicals or heat. Immunogenicity is retained, but a significant amount needs to be injected in comparison to live-attenuated vaccines (Hoft et al., 2017).

Conventional vaccines have drawbacks, including: 1) limited protection of a certain disease that has various strains or subtype of pathogens (Wong and Webby, 2013), 2) slow-production (Tree et al., 2001), 3) limited shelf-life and storage conditions (Amorij et al., 2008), 4) need for multiple doses and addition of adjuvants to enhance immune response for inactivated vaccine (Handley et al., 2007), and; 5) a risk of side effects.

Production of conventional vaccines is usually based on the propagation of embryonated chicken eggs, yielding one embryonated egg per dose (Tree et al., 2001). Therefore large-scale production is limited. The transportation for conventional vaccines requires a cold-chain refrigerator to maintain the quality of vaccine (Amorij et al., 2008). This raises the barrier for logistics. The side-effects with conventional vaccines have been reported, particularly, Crohn's disease (Afzal and Minor, 2002) and people with a compromised immune system (Nuttall and Eley, 2011). Moreover, it is reported that conventional smallpox vaccine may induce serious side-effects, not only skin rashes at the vaccination site, but also a chance of getting lymphadenopathy, life-threatening illnesses, cardiovascular disease, or death (Lederman et al., 2012).

Amongst various non-conventional vaccine platforms, Virus-like particles (VLPs) are designed to resemble closely the native virus conformation without viral genetic material (Buonaguro et al., 2011; Fuenmayor et al., 2017). VLPs can be derived from a number of sources including, Hepatitis B virus (HBV) (Roose et al., 2013), Human papilloma virus (HPV) (Tumban et al., 2013), Tobacco mosaic virus (TMV) (Smith et al., 2006), and; Cowpea mosaic virus (CPMV) (Montague et al., 2011). Notably, there are VLPs that have been commercially produced, particularly against Hepatitis B virus and Human Papillomavirus. They are VLP-based human vaccines, Cervarix™ and Gardasil™ for cervical cancer; Engerix™ and Recombivax HB™ for hepatitis B viral infection.

VLP are non-pathogenic and unable to replicate, which makes VLP safer compared with live-attenuated vaccine (Wheeler et al., 2008). Classification of VLPs can be based on 1) origin viral taxonomy (*hepadnavirus*, *papillomavirus*, *polyomavirus*, etc.), 2) particulate structures, and; 3) expression systems. VLP has been produced in a variety of enveloped and non-enveloped viruses. Fig. 2-1 presents example structures of VLPs, including a single, double, and triple, capsid proteins. They are categorized as VLP with single or multiple capsid proteins, with or without the lipid membrane.

The structure of VLP confirms either the conformation, or the size of either parental antigenic epitopes or foreign antigens. Naked VLPs are suitable to present simple structure and short peptide. These can be genetically fused into viral capsid protein to form chimeric VLP, for example, Hepatitis B core protein as a carrier for B cell/ T cell epitope (Pumpens and Grens, 2001). Non-enveloped VLP are easier to produce and purify. Alternatively, enveloped VLPs are able to carry full-length monomeric or multimeric conformational proteins, for example, influenza VLP producing from plant-based expression (D'Aoust et al., 2008). The structure of VLP depends on the expression system. For example, it is

reported that prokaryotic host, particular microbial system, can only express naked VLPs, and that eukaryotic systems are preferred for enveloped VLPs (Pushko et al., 2013).

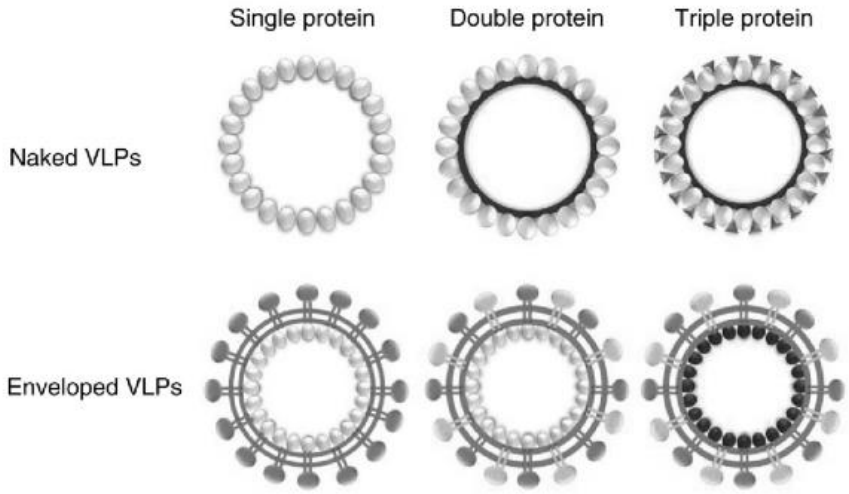

Fig. 2-1: Structure of naked and enveloped VLPs
(adapted from Tagliamonte et al., 2017).

The immunogenicity of VLPs is notable. VLPs present a high-density display of antigen in a highly organized and repetitive structure. Their immunogenicity can trigger the innate and adaptive immune system because of the similarity of pathogen associated molecular patterns (PAMPs) (Bachmann and Jennings, 2010). VLPs can be processed either as 1) exogenous antigens, or 2) endogenous antigens (Grgacic and Anderson, 2006). Buonaguro et al. (2011) reported that B cell activation and high production of antibody are impacted by the quantities of epitope and conformation of VLPs. With these beneficial immunogenic properties, VLPs induce the immune response effectively.

Not only are they safer and more immunogenically effective, VLPs have a shorter production time, compared with embryonated chicken eggs as is presented in Fig. 2-2. As an attractive alternative influenza vaccine can be produced in VLP platform using insect cells, Flublok© has been developed to prevent seasonal influenza and is approved in US in 2013 (Cox et al., 2015).

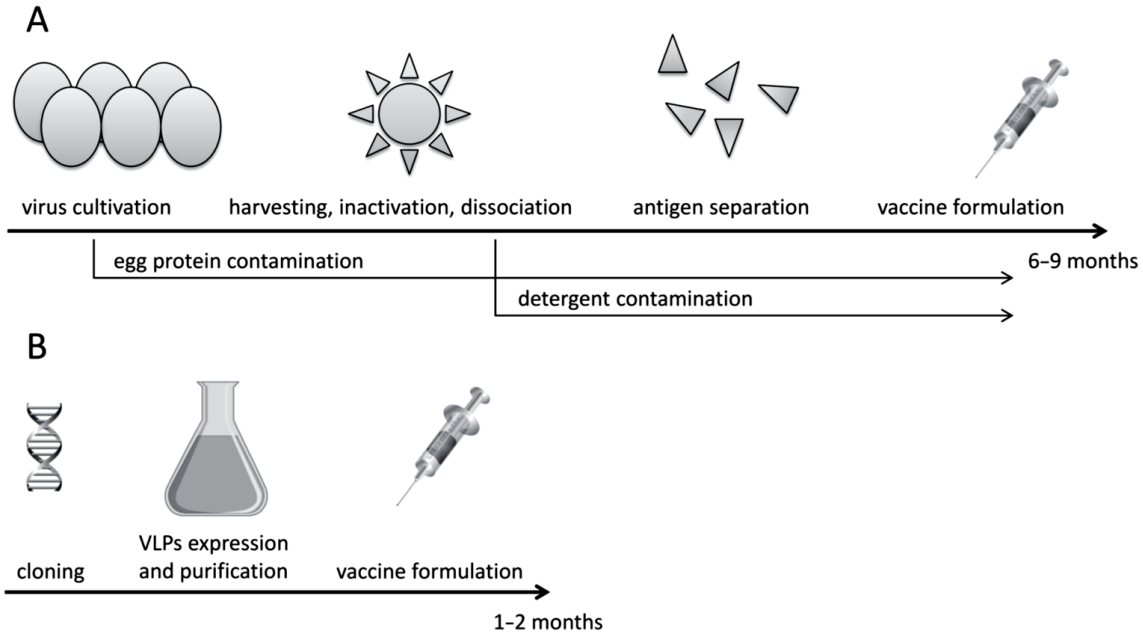

Fig. 2-2: Comparison of conventional vaccine (A) and VLPs (B) manufacturing process (adapted from Naskalska and Pyre, 2015).

Amongst VLP platforms Hepatitis B core protein has been intensively studied since its discovery by Clarke et al. (1987). Since then Hepatitis B core protein has been used to develop vaccine candidates for HCV (Sominskaya et al., 2010), human papilloma virus (HPV) (Chu et al., 2016), meningococcal meningitis and sepsis (Aston-Deaville et al., 2020), and influenza (Ravin et al., 2015). The molecular characteristics, structural feature and application of HBc is reviewed in the next section.

2.3 Chimeric Hepatitis B core virus-like particle as potential vaccine candidate

Hepatitis B core virus-like particle is one of the best known VLPs vaccine carrier. HBc monomeric protein consists of 183 – 185 amino acid residues in length for the *ayw* and *adw* subtypes (Pumpens et al., 2008).

2.3.1 Molecular information of HBc-VLP

Fig. 2-3 presents the domain organisation of HBV core protein. The viral self-assembly (SA) domain is from aa 1 to aa 140 and arginine-rich at C-terminal domain (CTD) for nucleic acid binding is from aa 150 to aa 183 (Birnbaum and Nassal, 1990). These two domains are linked by protease-sensitive hinge peptide 141-STLPETTVV-149 involved in morphogenic functions (Seifer and Standring, 1994).

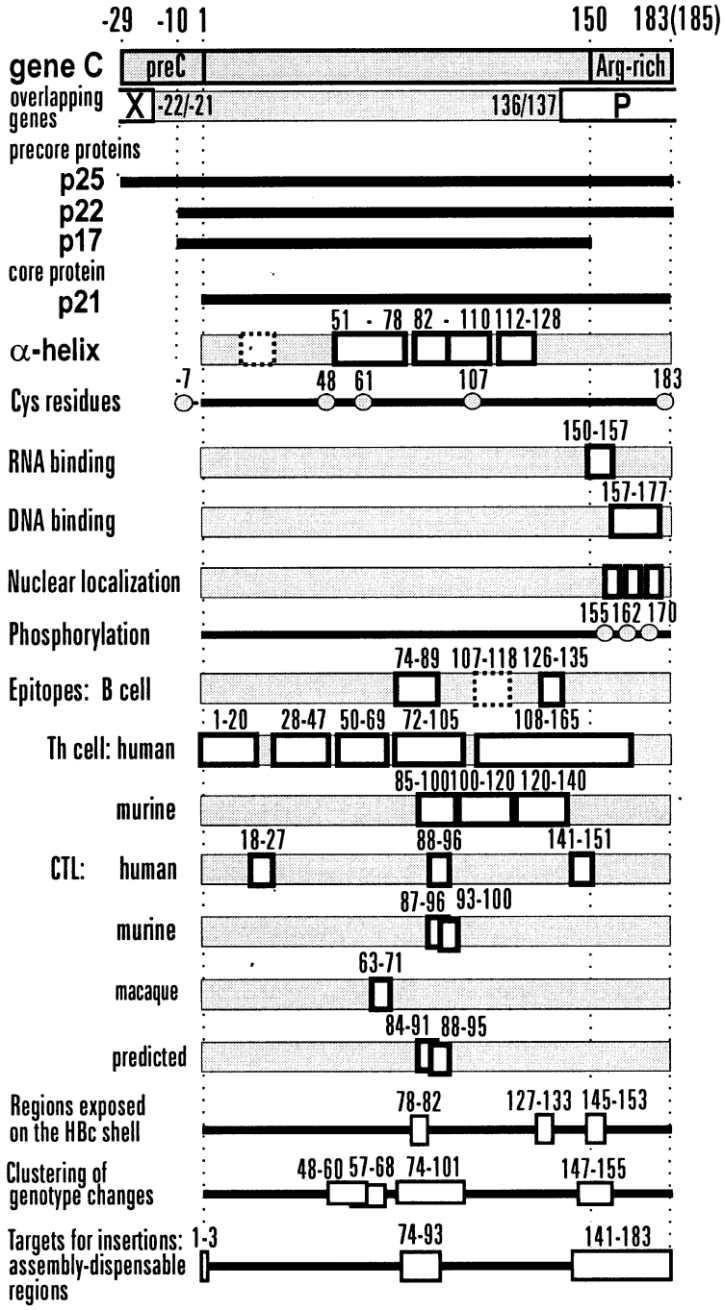

Fig. 2-3: Portrait of Hepatitis B core protein (adapted from Pumpens and Grens, 1999).

In addition, this peptide linker manipulates the ratio between the two HBc isomorphs upon recombinant expression (Zlotnick et al., 1996). The SA domain is responsible for immunological B-cell epitopes. The linking peptide and CTD do not appear to possess any immunological importance (Pumpens and Grens, 1999; Chain and Myers, 2005).

The HBcΔ variants and truncated HBc protein are HBV core protein without CTD, they still however possess self-assembly characteristic but are unable to encapsulate RNA from host cell (Borisova et al., 1988). Although the CTD enhances the encapsulation of

nucleic acid (between 100 to 3,000 nucleotides depended on the length of CTD), the interactions between protein-nucleic acid contribute to capsid stability (Birnbaum and Nassal, 1990).

HBc monomer protein is formed by five α-helices. In order to form a central helical bundle (tip of the spike), the helix-3 and helix-4 of HBc monomer fold themselves into "hairpin" structure and dimerize with a second monomer (Wynne et al., 1999). This ultimately creates the building blocks for capsid formation. At position of aa 61, cysteine residues from each monomer create a disulfide bridge.

Protein dimer is the basic structure of HBc protein. Fig. 2-4 presents the structure of HBc protein. The icosahedral HBc particles are assembled into two isomorphs, T = 3 (90 dimeric units) and T = 4 (120 dimeric units) according to the triangulation number. Mature HBc viral capsid has outer diameter of 32 nm and 35 nm, respectively (Crowther et al., 1994).

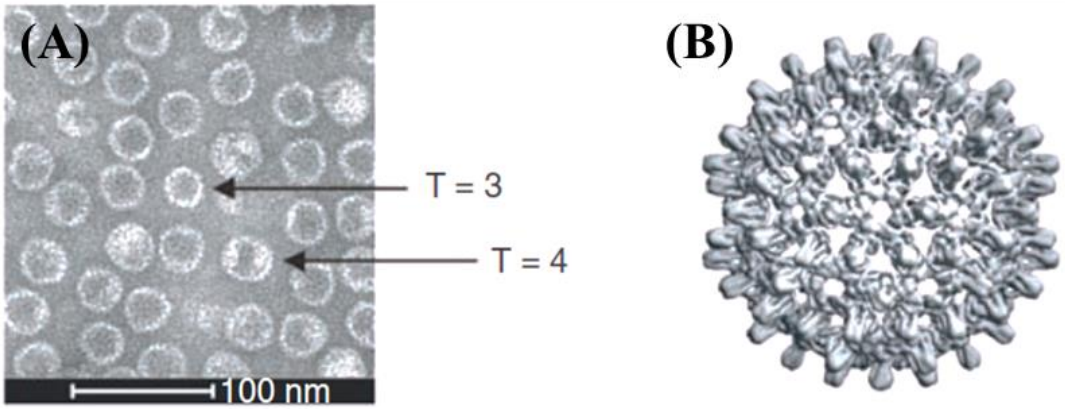

Fig. 2-4: Structure of Hepatitis B core VLP; TEM image presenting T=3 and T=4 symmetry (A) and Cryo-electron microscopy picture of HBc particles (B) (adapted from Roose et al., 2013).

The morphology and antigenicity of Hepatitis B core VLP expressed in *E. coli* are similar to the viral cores that are seen in HBV-infected human or infected chimpanzee (Cohen and Richmond, 1982). Due to this unique feature *E. coli*-derived HBc VLP has been exploited as an experimental vaccine carrier. Table 2-1 summarises the current HBc-based VLP vaccine candidates. These vary by the degree of epitope size and insertion location, including, 1) major immunogenic region (MIR) (aa residues 74-93 of HBc protein), 2) N-terminus and 3) C-terminus.

Table 2-1: *Escherichia coli*-derived HBc VLP vaccine candidates.

Pathogens /Targets	Antigen/ Epitopes	Insertion Site	Expression level/ Promoter	Reference
Alzheimer's disease and frontotemporal dementia	*tau* protein (aa residues 294-305)	MIR (between aa 78 and aa 79) of the truncated HBc gene (aa 1-149)	Tryptophan operon Soluble expression Yield: not available	Ji et al. (2018)
Dengue virus (DENV-2)	Envelope domain III (EDIII)	MIR (replacing aa residues 76–80) of the truncated HBc gene (lacking aa 166 to 183)	His-tag N-Terminal Lac Operon Insoluble yield: 7 mg L^{-1}	Arora et al. (2012)
Enterovirus (Hand-Foot-Mouth disease)	SP70 (sequence of 15 aa)	MIR (aa 78 and aa 79) of the truncated HBc gene (aa 1-144)	Lac operon Soluble expression Yield: not available	Pu et al. (2018)
Foot and mouth disease virus (FMDV)	Amino acids 142-160 of virus protein VPl,	N-terminus of HBc	Expression yield and promoter are not available	Clarke et al. (1987)
Quaternary Structure of Epitope	Enhanced green fluorescent protein (eGFP) (238 aa)	MIR (between aa 78 and 80) of the truncated HBc gene (aa 1–149)	His-tag C-Terminal Lac operon Soluble & Insoluble Soluble yield: 1 to 32.5 mg L^{-1}	Vogel et al. (2005)
Hantavirus	120 amino-terminal aa of N protein from DOBV, HTNV & PUUV	MIR at aa 78 of the truncated HBc gene (aa 1-144)	Tryptophan operon Soluble expression Yield: not available	Geldmacher et al. (2004)
Hepatitis B virus (HBV)	HBV Pre-S1; pure (aa 31-35); extended (aa 20-47); PreS1-phil (aa 16-60 and aa 89-119)	C-terminal and modification of aa 144-183 of full-length HBc gene (arginine codon to glycine codon)	Tryptophan operon Soluble expression Yield: not available	Dishlers et al. (2015)
Hepatitis B virus (HBV)	HBV Pre-S1-hydrophilic (aa 12-60 and aa 89-119)	MIR (between aa 78 and aa 79) of the truncated HBc gene (aa 1-144) and full-length HBc gene (183 aa)	Tryptophan operon Soluble expression Yield: not available	Skrastina et al. (2008)
Hepatitis C Virus (HCV)	HCV core (aa 1-98, aa 1-173, aa 1-151); HCV NS3 (aa 327-482, aa 202-326)	MIR at aa 78 or at aa 144 of the truncated HBc gene (aa 1-144)	Tryptophan operon Soluble expression Yield: not available	Mihailova et al. (2006)
Hepatitis C virus (HCV)	HCV core (aa 1-60); HBV Pre-S1 (aa 20 - 47)	MIR at aa 78 or C-terminal at aa 144 of the truncated HBc gene (aa 1-144)	Tryptophan operon Soluble expression Yield: not available	Sominskaya et al. (2010)

Table 2-1: *Escherichia coli*-derived HBc VLP vaccine candidates (*cont.*)

Pathogens /Target	Antigen/ Epitopes	Insertion Site	Expression level/ Promoter	Reference
Hepatocellular carcinomas (HCC)	MAGE-1 (aa 278–286); MAGE-3 (aa271–279); AFP1 (aa 158–166) and AFP2 (aa 542–550)	C-terminal (aa 149) with GA linker of the truncated HBc gene (aa 1-149)	Lac operon Insoluble expression Yield: not available	Zhang et al. (2007)
Human Papillomavirus (HPV)	Cytotoxic T lymphocytes epitope E7 (aa 49-57)	MIR (between aa 78 and aa 79) of the truncated HBc gene (aa 1-149)	Lac operon Soluble expression Yield: not available	Chu et al. (2016)
Infectious bursal disease virus (IBDV)	5EPIS epitope (78 aa)	MIR (between aa 79 and aa 80) of the full-length HBc gene (183 aa)	Lac operon Insoluble expression Yield: not available	Wang et al. (2012)
Influenza virus	1, 2 or 4 copies of trans membrane protein M2e (aa 2-24) of human and avian strain	MIR (replacing 78-81 aa) of the truncated HBc gene (aa 4-149+ Cys)	Lac operon Soluble/ Insoluble expression Yield: not available	Ravin et al. (2015)
Influenza virus	Long alpha helix (LAH) from HA2 (aa 76–130)	MIR (replacing 75–85 aa) of the HBc gene (aa 1-149)	His-tag Lac operon Insoluble expression Yield: not available	Chen et al. (2015)
Influenza virus	Multiple M2e epitopes	N-terminal and MIR (78-81 aa) of the truncated HBc (aa 1-149) and N-terminal of the truncated HBc (aa 1−149+Cys).	His-tag C-Terminal Lac operon Soluble /Insoluble expression Yield: not available	Sun et al. (2015)
Melanoma	MAGE-3 (aa 168–176); MAGE-3 (aa 163–181)	MIR (replacing 76-84aa) and C terminal of aa 161, 168, 176 of variant truncated HBc gene	Tryptophan operon Soluble expression Yield: not available	Kazaks et al. (2008)
Meningococcal meningitis and sepsis	Linker of glycine and serine; FHbp ORF (aa 157-274); NadA ORF (aa 26-309)	MIR (replacing aa 78-81) and aa 148 of the HBc gene (aa 1-148)	Strep-tag C-Terminal Lac operon Soluble expression Yield: 8 mg L^{-1}	Aston-Deaville et al. (2020)
Improving solubility of HBc VLP	β-galactosidase (8 aa); Xpress™ epitope tag (8 aa)	N-terminal of the truncated HBc gene (aa 1-146)	His-tag N-terminal Lac operon Soluble expression Yield: 75 mg L^{-1}	Yap et al. (2009)

Table 2-1: *Escherichia coli*-derived HBc VLP vaccine candidates (*cont.*)

Pathogens /Target	Antigen/ Epitopes	Insertion Site	Expression/ Promoter	Reference
Porcine epidemic diarrhea virus (PEDV)	PEDV linear B-cell epitopes (aa 748-755, aa 764-771, aa 1368-1374, aa 195-200)	MIR (replacing aa 78-82), MIR (aa 79), and C-terminal (aa 149) of the HBc gene (aa 1-149)	His-tag N-Terminal Lac operon Insoluble expression Yield: not available	Gillam and Zhang (2018)
Puumala (PUU) Hantavirus	PUU N Vranica epitope (aa 1-45); PUU N CG18-20 (aa 1-45; aa 75-119)	N-terminal, MIR (aa 78) and C-terminal of the truncated HBc gene (aa 1-149)	Tryptophan operon Soluble expression Yield: not available	Koletzki et al. (2000)
Respiratory Syncytial Virus (RSV)	G protein (aa residues 144-204); M2 protein (aa residues 82-90)	MIR (between aa78 and aa79) and C-terminal of the truncated HBc gene (aa 1-149)	His-tag N-Terminal Lac operon Insoluble expression Yield: not available	Qiao et al. (2016)
Rubella virus	E1 glycoprotein (aa 214-285, aa 214-240, aa 245-285)	MIR (aa 78 and aa 79) of the HBc gene (aa 1-144)	Tryptophan operon Insoluble expression Yield: not available	Skarastina et al. (2013)

A comparison of potential insertion sites suggests MIR is more effective in inducing antigen-specific antibodies. This can be explained by the structure of HBc dimer because MIR allows maximum exposure of the inserted protein on the tip of the spike (Schodel et al., 1992). Although the insertion capacity at MIR is not fully studied, Geldmacher et al. (2004) successfully expressed HBc-VLP bearing two copies of 120 aa hantavirus nucleocapsid protein, and Vogel et al. (2005) employed HBc-VLP to present full-length 238 aa of green fluorescent protein.

Insertion at N-terminal guarantees the outer exposure of epitopes on the capsid surface, which was demonstrated in the first HBc-VLP vaccine candidate carrying FMDV epitope by Clarke et al. (1987). To support the argument, an analysis in crystallization and X-ray of T=4 particle of hepatitis B capsid protein, conducted by Tan et al. (2007), confirmed that fused peptides at N-terminal insertion was exposed on the exterior of HBc capsid. However, it is reported that the assembly of the capsid is influenced by the peptide fused to the N-terminus, unless there is a flexible linker included prior to the N-terminus of HBc (Murray and Shiau, 1999).

However to the contrary, insertion at C-terminal of HBc protein does not guarantee outside exposure of foreign epitope since C-terminus of HBc is located interiorly of the capsid (Zlotnick et al., 1997). Usually, the attached protein of interest is buried within the

HBc capsid, leading to poor immunogenicity. Therefore, the use of N-terminal fusion HBc in the development of other HBc vaccine candidates is limited. To overcome this disadvantage a modified HBc protein was developed by fully or partially replacing the arginine residues of the CTD with glycine residue, namely HBcG vector (Dishlers et al., 2015). This modification allowed the outer surface exposure of C-terminally inserted epitope and improved immunogenicity.

Although full-length HBc protein has the advantage of capsid stability, the C-terminally truncated HBcΔ protein, consisting of HBc aa 1-144 or HBc aa 1-149, are adequate to be used as vaccine scaffolds (Sominskaya et al., 2013; Roseman et al., 2012).

As a versatile vaccine-carrier platform, the potential of using HBc-derived VLPs in vaccine production is great. HBc-derived VLPs can be modified in various means, including alteration of the length of antigen epitopes, the choice of insertion site in the protein, and the adjustment in C-terminal sequence or MIR region. Although these means sound promising in creating an effective recombinant HBc-based VLPs, there is no guarantee that the designated recombinant protein will be expressed as soluble, native-folded, and particulate structure. It is necessary to explore how molecular design, regarding the length of inserting epitope at either N or C-terminal insertion site, and environmental conditions influence the expression of soluble chimeric HBc-VLP in order to establish a cost-effective production.

It is stated that there is no difference in structure between HBc particles from liver of infected patients and recombinant HBc particles that derived from *E. coli* (Kenny et al., 1995). Recombinant HBc protein forms a dimer then self-assemble into two (2) forms of icosahedral particles (Roseman et al., 2005), it is in agreement with the vaccine candidates listed in Table 2-1, recombinant HBc protein were expressed as HBc monomer then self-assembled into VLP in *E. coli*. Based on the product folding state and nature of the recombinant protein, different purification pathways have been developed, which will be reviewed in the next section.

An additional advantage of full-length Hepatitis B core protein is high thermal stability. For example in the study by Naito et al. (1997) Hepatitis B core antigen (HBcAg) was heated to temperatures of 60, 65 and 70, ^{o}C for 60 min. Only the sample heated at 70 ^{o}C decreased in quantity. This finding suggests HBcAg in the crude lysate had thermal resistance up to 65 ^{o}C. This feature was confirmed by Ng et al. (2006) where HBcAg retained its structure and antigenicity when heated at 80 ^{o}C for 45 min.

The thermal stability of chimeric HBc VLP has also been utilised to remove host cell protein in order to obtain a less complicated feed for chromatography (Li et al., 2018). In addition, the size of assembled VLP is considered large, compared with other host protein. Size exclusion chromatography has been used to monitor the expression of HBc VLP (Yang et al., 2015). These two features can be use in sample treatment before the determination of assembly expression level. It is expected that insertion of foreign epitope may influence the stability and assembly of HBcAg (Karpenko et al., 2000). Therefore, the condition in sample treatment should be considered.

2.3.2 Potential epitopes for new vaccine development

2.3.2.1 VLPs based vaccine against Epstein-Barr virus (EBV) infection

Epstein-Barr virus (EBV) is one of the most widely distributed pathogens since the primary spreading pathway is *via* saliva. EBV-infection impacts significantly on public health and world economics since EBV infection often develops to infectious mononucleosis (IM) and patients with IM exhibit symptoms such as fever, sore throat, swollen lymph node, severe fatigue, immune dysfunction and these patients have higher chance to EBV-associated malignancies (Cohen et al., 2011). These malignancies include lymphoma (Burkitt's lymphoma, Hodgkin's disease, natural killer cell and T cell lymphomas) and carcinoma (nasopharyngeal carcinoma and gastric carcinoma), which contributes to about 1.5 % of all cases of human cancer worldwide (Farrell, 2019).

It is reported that although chemotherapy is an effective approach to battle against EBV-associated malignancy, chemotherapy-associated complications, disease relapse or secondary malignancies have been putting an end to 50 % of survivors of all forms Hodgkin lymphoma (Cohen et al., 2011). EBV-associated lymphoma is the second most common malignancy, which occur in bone marrow and organ transplant recipients (Cohen et al., 2011).

Since the first vaccine proposal was developed by Epstein (1976), most of the EBV prophylactic vaccine trials was focused on EBV membrane glycoprotein *gp350* as the monomeric EBV *gp350* reduced the incidence of infectious mononucleosis (Cohen, 2018; Sokal et al., 2007). Besides of glycoprotein *gp350*, EBNA1, a crucial protein that meditates EBV genome synthesis and partitioning to daughter cells, and regulates viral gene transcription (Altmann et al., 2006; Smith and Sugden, 2013), has been utilised to boost humoral response and trigger cell-mediated immunity (Perez et al., 2017). A full-length

EBNA1 consists of 641 amino-acids (EBV strain B95-8), and two of this polypeptide compose together to form dimeric protein (Lindner and Sugden, 2007). Although EBV nuclear antigen (EBNA1) has been successfully expressed in *E. coli*, because of the composition of the rare codon, the yield of EBNA1 in *E. coli* was low, approximately 2 mg L^{-1} (Mayer et al., 2012).

For development of effective EBV vaccine against virus-associated cancers, the vaccine candidate must be able to induce a strong T cell-mediated immune response and fewer side effects. It is challenging to fuse the full-length EBNA1 to any vaccine carrier, therefore, using EBNA1-derived epitope HPVGEADYFEY as an alternative approach has high potential. HPVGEADYFEY is a linear epitope from the EBNA1 primary sequence (aa 407–417). Effectors that recognize and kill infected target cells are generated by the contribution form $CD8^+$ cytotoxic T lymphocyte (CTL) response. It is reported that EBNA1 epitopes (aa 407–417) can induce $CD8^+$ T cell responses *in vivo* through uptake of the exogenous protein (Blake et al., 1997; 2000). Another approach is using VLP platform to generate EBV VLP vaccine candidate, this EBV-derived VLP was expressed in mammalian expression system and demonstrated highly immunogenic in terms of humoral and strong $CD8^+$ and $CD4^+$ T cell responses *in vitro* and *in vivo* (Ruiss et al., 2011), however, expression in mammalian system is not an economical process.

Therefore, construction of *E coli*-derived HBc VLP carrying EBNA1 epitope is promising for developing an effective vaccine candidate. Even though there are progress in vaccine development against EBV-infection, there is no therapeutic or prophylactic vaccination available at present.

2.3.2.2 VLPs based vaccine against Hepatitis C virus (HCV) infection

Hepatitis C virus is a single-stranded RNA (ssRNA) enveloped virus, classified in the genus *Hepacivirus* in the *Flaviviridae* family (Choo et al., 1989). HCV is a blood-borne virus, including blood exposure, drug use, sexual transmission, mother-to-baby and sharing hygiene items if infected blood present. HCV infects approximately 3 % of the world population with a breakdown of an estimated 100 million people are serologically positive and at least 71 million people are diagnosed as persistent infection (Soriano et al., 2018). It is reported that HCV-associated end-stage liver disease and its complication cause nearly 476,000 deaths per year globally (Calvaruso et al., 2018). In Australia, HCV infection occurs at a rate of 10,000 new case annually (Holmes et al., 2013).

In past decade treatment of HCV primarily is focused on the combination of PEGinterferon alfa-2a and ribavirin. This treatment is administered for 24 or 48 weeks. However the high percentage (up to 60 %) of treated subjects cannot eliminate the virus completely (Fried et al., 2002), eventually making this treatment less ideal. Therefore, the preventative vaccine is a priority to fight against HCV infection.

It is a challenging and rewarding progress to develop an effective HCV vaccine candidate because there are multiple genotypes of HCV with complex nature of immunological response to HCV and limited availability of animal model (Stoll-Keller et al., 2009). To provide a proper virus clearance, it is requires that vaccine candidates possess neutralizing antibody (NAb) responses with strong and broadly cross-reactive $CD4^+$, $CD8^+$ (Alvarez-Lajonchere and Duenas-Carrera, 2012). For this reason, HCV vaccine often combines HCV NS3, core, NS4, NS5 protein or their epitope to induce strong cellular response meanwhile E1 and E2 focus on Nab (Bellier and Klatzmann, 2013). Particularly, HCV core, E1 and E2 are structural protein, which demonstrate significant protective immune responses compared with BS3, NS4, NS5 as they are non-structural protein (Dahari et al., 2010). Moreover, a study of Boulant et al. (2005) stated that HCV core protein structure is a dimeric α-helical. To induce antigen-specific board NAbs, conformational epitopes need to be presented as their correct 3D structure.

Virus-like particle is a promising platform to fulfil all the criteria for HCV vaccine, especially HBV core-based VLP. At first, for serological diagnosis purpose, HCV core (aa 1-98) was fused with HBc in compliance with preS1 and preS2 to form HBcPreS1PreS2HCc (1-98). In terms of anti-HCc induction, this synthesised protein possessed higher immunogenic property than HCc (1-98) (Wu et al., 1999). Later Sominskaya et al. (2010) has constructed multivalent HBc VLP carrying HBC (PreS) and HCV core epitope as vaccine candidate. Moreover, chimeric HBc VLP carrying HCV epitopes (HCV core 98 aa and HCV NS3 155 aa) was also produced by Mihailova et al. (2006). Although both chimeric HBc VLP were able to be constructed and expressed in *E. coli*, the antibody and T-cell proliferative responses to HCV epitopes were low. One important factor about these chimeric HBc VLP is the insertion site of the epitopes, C-terminally fusion directes the epitopes to be buried inside the capsids, leading to poor antibody responses (Schodel et al., 1992).

Since there is no therapeutic or preventative vaccination available for HCV infection yet, it is a promising approach to fuse HCV core (aa 10-53), also known as *c22p*, to HBc VLP at N-terminus to develop new vaccine candidates. HCV core consists of aa 10-53 was

used as novel Hepatitis C virus major epitope for diagnosis of HCV infection (Chien et al., 1999; Sominskaya et al., 2010).

To summary the review of HBc VLP, six (6) key aspects are:

1. Hepatitis B core VLP is a promising platform for vaccine development and it is expressed in both prokaryotic and eukaryotic expression system
2. The immunogenicity of chimeric HBc-VLP depends on the location of epitope. There are three (3) insertion sites: 1) N-terminus, 2) C-terminus, 3) MIR. They determine the location of epitope, either displayed on exterior surface or interior of the viral capsid
3. Full-length HBc protein has the advantage in capsid stability
4. Although many molecular design approaches have been conducted, including insertion site, with tag or without tag, choice of promoter, etc., it is not guarantee that chimeric HBc VLP is expressed as soluble intracellular protein
5. The thermal stability and size of HBc is beneficial for purification process. However, the inserted epitope may alter these properties
6. Although HBc VLPs tolerate the structural modification, the influence of structure and size of the inserted epitope on the formation of VLP is not well-studied. It is expected that the soluble expression of HBc protein carrying complex structure epitope is less than the HBc protein and HBc protein carrying simple structure epitope.
7. HCV core protein has α-helical structure, in comparison with EBNA1 protein, which is a linear epitope, HCV vaccine and EBV vaccine are potential innovative vaccines against HCV and EBV infectious disease.

The nature and flexibility of HBc platform has been reviewed and summarised. The next section presents the microbial expression system, particularly process factors that influence the soluble recombinant protein expression.

2.4 *Escherichia coli* as microbial expression system

Escherichia coli has been used intensively as microbial expression system for producing biopharmaceutical products from both prokaryotic and eukaryotic origins, i.e. the human insulin (Goeddel et al., 1979) and human protein somatostatin (Itakura et al., 1977). The benefits of short doubling time (~20 min) and ability to culture in expensive media makes *E. coli* as one of the most popular means to produce recombinant pharmaceutical, particularly virus-like particle. The followings are some examples of commercialised vaccine or vaccine candidates in the pipeline produced by *Escherichia coli*.

Hepatitis E Virus (HEV 239) vaccine has been approved in China, under the name of Hecolin©. It was reported as the first *E. coli*-derived VLP-based vaccine (Proffitt, 2012). Moreover, expression of murine polyomavirus VLP, carrying M2e epitope from influenza, in *E. coli* has potential to produce a safe and effective vaccine (Middelberg et al., 2011). There are several vaccine candidates, in which epitope is fused into Hepatitis B core (HBcAg) or Hepatitis B surface antigen (HBsAg), have been in clinical trials, including:

1. ACAM-FLU-A© (Phase I), Hepatitis B core antigen (HBcAg) bearing M2e epitope (Ibanez et al., 2013)
2. MalariVax© (ICC-1132) (Phase I) - HBcAg bearing *Plasmodium falciparum* circumsporozoite protein (Gregson et al., 2008)
3. HBV ABX203 (HeberNasvac©) (Licensed) - HBsAg/HBcAg as vaccine antigens (Lobaina et al., 2015).

These are strong evidence that *E. coli* is feasible to generate new HBc-based vaccine candidate. In the following sections, the recombinant protein production process, production folding state associated with purification process are discussed. Moreover, the cultivation factors affecting the soluble expression of recombinant protein are also summarised.

2.4.1 Recombinant protein production process by *E. coli*

Because of advances in genetic engineering, *E. coli* has become the first choice for laboratory study and initial development for large-scale production. Bacterial system offers easily scale-up and economical production process (Chuan et al. 2014).

The key components of recombinant protein production process are presented in Fig. 2-5. They include the 1) design of expression vector, 2) selection of host strain and transformation, 3) growth of bacteria and protein production, and; 4) protein recovery and purification.

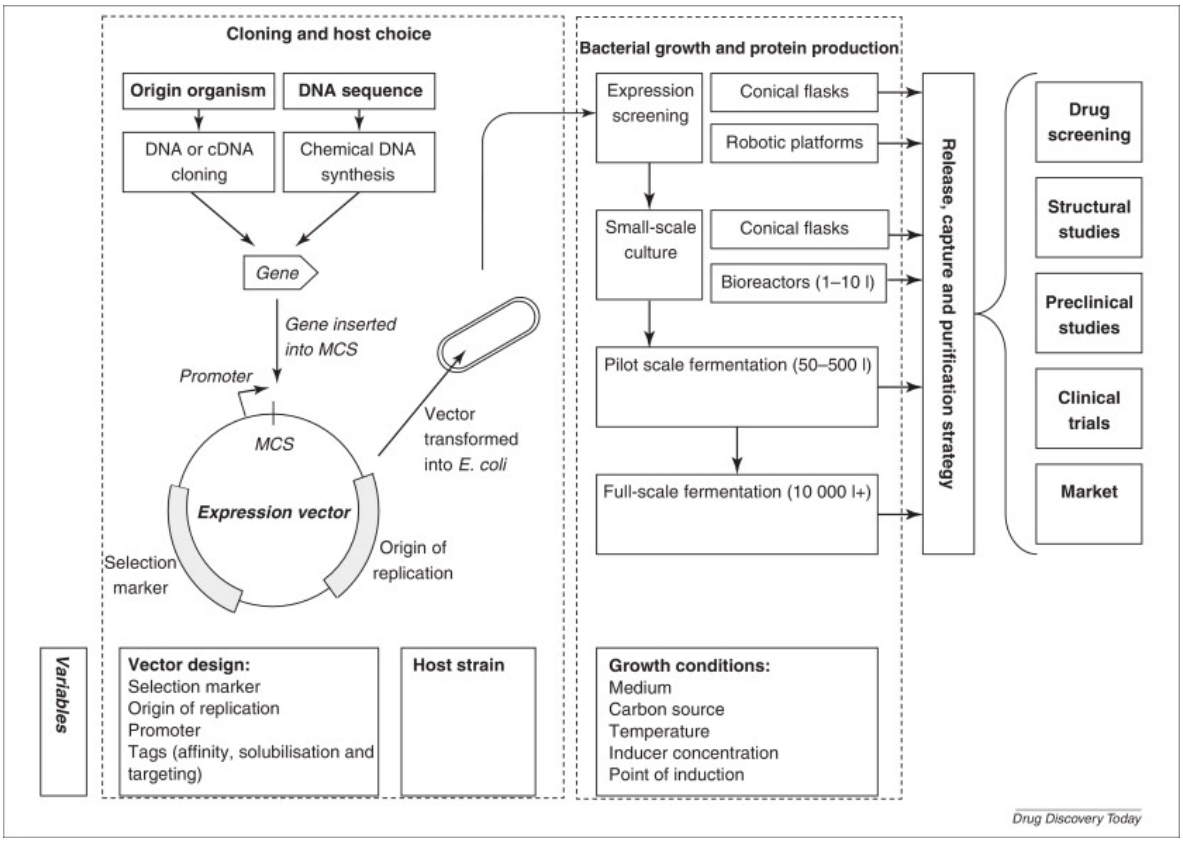

Fig. 2-5: Outline of recombinant protein production in *Escherichia coli*
Abbreviation: MCS, multiple cloning site (adapted from Overton, 2014).

It is common that the during protein production in bacteria, potential problems that come from cell viability, plasmid loss or protein folding occurred (Overton, 2014). Cell viability problems are described as extensive cell death during growth phase or at the time of inducer introduction. Plasmid loss is another common problem, since production of recombinant protein put cells in metabolic pressure, the energy demand for maintenance

and replication of plasmid vectors is significant (Silva et al., 2012). The next issue is protein folding, which is also the main focus of this research project.

In general, the overexpression of heterologous recombinant protein leads to metabolic burden in host cell (Glick, 1995). Cells start to overconsume the metabolic precursors such as adenine triphosphate (ATP), ribosomal ribonucleic acid (rRNA) and amino acids to support the protein synthesis (Ramchuran et al., 2005). Expression of recombinant protein in *E. coli* remains a major challenge. Because of lacking post-translation modification, the protein folding which is crucial for protein solubility, stability, biological activity, immunogenicity and purification strategy (Walsh and Jefferis, 2006). It is proposed by Peternel and Komel (2011) that various transitional folding states of target protein are presented during the expression, not just only the correct folded protein. Fig. 2-6 demonstrates the mechanism of protein folding and aggregation. The aim of protein folding is to fold the newly synthesized protein into correct folding, however, if the folding intermediates fails to achieve the native folding, they will be degraded or form into small proto-aggregates (soluble aggregates) (Schrodel, and de Marco, 2005). A significant note is protein folding is assisted with cell quality control machinery, including chaperone, heat-shock protein and proteases. This network establishes the kinetic equilibrium between soluble and aggregates form of protein (reviewed by Peternel and Komel, 2011). Therefore, soluble fraction of crude cell lysates includes both protein with native conformation and proto-aggregates. In addition, proto-aggregations accumulate themselves, promoting the accumulation of insoluble aggregates, also known as inclusion bodies (Fahnert et al., 2004).

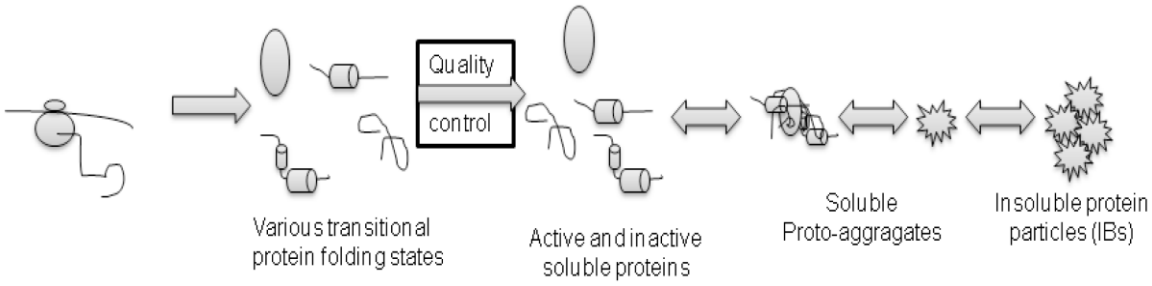

Fig. 2-6: A potential mechanism of protein folding and aggregation in *E. coli* (adapted from Peternel and Komel, 2011).

In some cases, protein from inclusion bodies can be recovered by solubilisation and refolding, however, there is uncertainty that protein will gain back their native function fully (Ventura and Villaverde, 2006). Proper folding not only enhances the solubility of the

protein but also maintain the specific structure for clinical assessment, particularly for HBc-based VLP, where the epitope of interest was displayed on the surface of the capsid in repetitive manner.

Table 2-2 illustrates the typical downstream process for each upstream product folding state. A generic purification process of soluble recombinant protein includes protein extraction, clarification, bioburden reduction, purification chromatography and polishing chromatography (Kazaks et al., 2008; Vicente et al., 2011), meanwhile, the downstream processing of inclusion bodies includes the solubilisation of inclusion bodies, refolding, purification and assembly (Suffian et al., 2017; Zhang et al., 2020; Jevsevar et al., 2005). Moreover, the chance that host cell proteins or DNA are encapsulated in recombinant inclusion bodies is high, eventually affecting the refolding yield negatively (Maachupalli-Reddyet al., 1997).

In comparison, the purification of soluble protein is less complicated and more productive, both in time and cost, than the recovery of bioactive material from inclusion bodies. The complexity of downstream processes increases significantly when recombinant protein is purified from inclusion body protein. Therefore, maximizing the expression of HBc-based VLP as soluble form is highly regarded. It is found that adjusting the process factors during induction phase increases the soluble expression of recombinant protein in *E. coli* because they influence both folding state and yield of the target product (Tripathi et al., 2009; Papaneophytou and Kontopidis, 2014). These are discussed in the following section.

Table 2-2: Product folding state determines the purification strategy.

	Product folding state	
	Soluble intracellular protein[a]	**Inclusion body protein**[b]
Steps	Cell harvest	Cell harvest
	Cell disruption	Cell disruption
	Clarification	Clarification
	Heat and salt precipitation	Denaturing/Solubilisation
	Solubilisation of precipitate	Purification in monomer form
	Diafiltration	Assembly in step-wise reduction
	Size-exclusion chromatography	Diafiltration

(a) Kazaks et al., 2008; (b) Suffian et al., 2017

2.4.2 Genetic factors and host strain

In expression vector design, there are some variables to be taken into consideration, including the selection of promoters, the insertion of target gene into multiple cloning site, the origin of replication and optional tags. The selection and development of promoter, based on the characteristic of each protein, are important because it should be tuneable to achieve high yield of the target protein able to and minimise the leaky expression (Rosano and Ceccarelli, 2014).

There are several promoters that are widely used in protein production, such as Arabinose, hybrid (*trc*, *tac)* and T7. Although they are considered as strong promoters, they also exhibit their own drawbacks (Jia and Jeon, 2016). For example, arabinose promoters possess the lowest basal transcriptional activity among promoters because they rely on positive control (Siegele and Hu, 1997). However they depend on gene-regulation of the recombinant protein, and the minimal basal transcriptional activity does not always guarantee to reach zero (Guzman et al., 1995). Moreover, hybrid promoters demonstrate leaky expression, affecting plasmid stability and eventually decreasing the yield of the target protein (De Boer et al., 1983; Tegel et al., 2011).

It is undeniable that *lac* promoter, key opponent of *lac* operon, is extremely popular for protein expression. The *lac* operon is responsible for the transport and metabolism of lactose in *E. coli*. Lactose or its non-hydrolysable analog Isopropyl β-D-1-thiogalactopyranoside (IPTG) are two common inducer that can induce *Lac* operon. However, *lac* operon is only a moderate promoter, in the presence of glucose, it is difficult for *lac* operon to be completely activated. For a complete activation of lac operon, the cyclic adenosine monophosphate (cAMP) is necessary, which is produced when glucose has been utilized (Kimata et al., 1997). It is suggested that *lac* promoter and its derivatives are rather weak (Deuschle et al., 1986).

The T7 RNA polymerase (RNAP) system, also known as pET system, which is developed based on the T7 bacteriophage (Studier and Moffatt, 1986; Studier, 1991) to overcome the dependence on the cAMP and increase the strength of the promoter. The gene of the T7 RNA polymerase is chromosomally integrated under the influence of a *lac* promoter derivate, the *lac*-UV5, less sensitive to the presence of glucose (Grossman et al., 1998). With this hybrid fusion, the T7*lac* system allows strong induction and high level of target gene expression. To produce the recombinant protein, IPTG or lactose is required for pET system. The connection between T7 promoter, lac operon and IPTG are well-studied

by Studier and Moffatt (1986). In the absence of IPTG, the transcription of T7 RNAP is prevented due to the binding between *lac* promoter and Lac represseor LacI. When the IPTG is added, LacI binds to the IPTG molecules and releases *lac*-UV5, therefore, *E. coli* RNAP is able to transcribe T7 RNAP, which activates the promoter on expression vector to produce the target gene. Another crucial point is the native *E. coli* RNAP cannot activate T7 promoter on the pET expression plasmid. This activation is significantly dependent on the presence of inducer (Overton, 2014). Regarding metabolic pathway, inducers should be able to enter the cell *via* active transport system or passive diffusion through cell membrane. For example, the uptake of natural inducer (lactose) in *E. coli* depends on the lactose permease LacY (Guan and Kaback. 2006) and the uptake of synthetic inducer (IPTG) is mediated by both passive diffusion and lactose permease LacY active transport (Fernandez-Castane et al., 2012).

The choice of host strain is also a major factor to determine the efficiency of the protein production process. In principle, the requirements of the plasmid expression system determines which host strain should be used. Regarding pET expression system, the chromosomally encoded T7 RNA polymerase is required. Because of this reason, the choice of host strain limits into two (2) strain: K-12 and BL21 (Overton, 2014). The *E. coli* BL21 (DE3) strain is preferred for the expression of protein that are cloned under the control of T7 promoter (Studier and Moffatt, 1986) Moreover, in comparison to *E. coli* K-12, *E. coli* BL21 demonstrates a higher biomass yield and lower acetate production (Waegeman et al., 2013).

2.4.3 Cultivation factors effecting soluble expression of recombinant protein

As is presented in Fig. 2-5 the scale of bacteria culture varies from screening to commercial production, particularly from 10 mL to larger than 10000L. The expression screening studies can be done using conventional shake-flask cultivation (Li et al., 2015) or microtiter plate-based cultivation system (Rohe et al., 2012). After determined the growth condition from the initial batch culture, the production is often tested in stirred-tank reactor (STR) on batch and fed-batch regime, in which the condition of culture are well-controlled and monitored, in comparison to shake-flask cultivation. At this point, the growth and expression condition decide the success of the production process.

In ideal recombinant protein production, a strong promoter supports a high soluble protein-production with native conformation. A strong promoter often has these properties:

1) it must be tightly regulated to ensure the growth and survival of cell by preventing the product toxicity when the recombinant gene is overexpressed, and 2) it allows the accumulation of protein-based products up to 50 % of the total cell protein (Studier and Moffatt, 1986).

As significant transcription rate occurs, the destruction of ribosome and cell death remain as disadvantages (Studier, 1991). Expression of a foreign recombinant gene creates stress on the expression host. The metabolic burden, either energy or nutrients, will interfere the cell metabolism and cell growth. The resources will be utilized for cell maintenance and expression of target gene. The foreign recombinant protein may have different properties from the microenvironment of *E. coli*. The differences in pH, redox potential, osmolarity and folding mechanism as well as high concentration of hydrophobic stretches in the newly synthesized protein lead to the formation of inclusion bodies (Carrio and Villaverde, 2002).

As is presented in Table 2-1, most of HBc-based VLPs is expressed as inclusion bodies under *lac* promoter, including vaccine candidate of porcine epidemic diarrhea virus (Gillam and Zhang, 2018), influenza virus (Chen et al., 2015; Sun et al., 2015), hepatocellular carcinomas (Zhang et al., 2007), infectious bursal disease virus (Wang et al., 2012). To obtain maximum soluble yield of target products with T7*lac*, an optimal induction strategy is required.

The environmental expression conditions contribute to the yield and ratio soluble/insoluble protein (Papaneophytou and Kontopidis, 2014). The influence from protein synthesis rate and protein folding rate are important for the expression of chimeric HBc in *E. coli* in order to maintain its structure and immunogenicity. In general, expression conditions are optimised in shake-flask or microlitre cultivation prior to fermenter because of process economics. In terms of induction in batch mode, no feeding is required. However, regarding fed-batch mode, feeding strategy and feeding rate have strong influence on protein production (Wechselberger et al., 2012).

Therefore, optimisation of process parameters during induction phase is necessary to ensure the soluble production of chimeric HBc VLP in fermenter scale.

2.4.3.1 Media composition

Theoretically, complex media contains yeast hydrolysates, it offers a rapid growth for bacterial culture and a great reservoir of amino acids for recombinant protein production. Organic nitrogenous source, such as yeast extract, provides precursors for cellular building blocks synthesis and reduces the inhibitory effects of acetate in the culture. These benefits improve the specific cellular yield of the target protein (Tripathi et al., 2008; Lim et al., 2000). For example, Luria-Bertani (LB) media is a classic media for production of recombinant protein in *E. coli*, which provides basic nutrient content and suitable osmolarity for *E. coli* growth at early log phase (Sezonov et al., 2007). However limited quantity of divalent cations and carbon source in LB cannot support the culture to reach high cell density (Sezonov et al., 2007).

The supplement of divalent cations, i.e. $MgSO_4$, enhances the cell growth by extending the exponential phase (Studier, 2005). Magnesium has also been reported to enhance the cell growth in multiple strains of *E. coli* and decrease acetylation based on the ratio between carbon and magnesium (Christensen et al., 2017). Moreover, lack of Mg^{2+} in peptide media may lead to ribosome degradation (Nierhaus, 2014; McCarthy, 1962). It is reported that in the presence of magnesium (0.1 mM), gloshedobin, snake venom enzyme, had 50 % increase in solubility when expressed at 25 ^{o}C (Yang et al., 2003). With these arguments, addition of Mg^{2+} is predicted to boost the production rate.

Because of the lack of carbon source in LB, other advanced complex media such as Terrific Broth has been used. The carbon source in these complex media comes from the addition of glucose or glycerol. *E. coli* can be cultivated aerobically or anaerobically. In case of aerobic cultivation, *E. coli* consumes glucose for respiration. Carbon source enters the central carbon metabolism and tricarboxylic acid (TCA) cycle to make adenine triphosphate (ATP) in the presence of oxygen (Martinez-Gomez et al., 2012; Hempfling and Mainzer, 1975). In case of anaerobic cultivation, *E. coli* switches the metabolism pathway of glucose from respiratory pathway to fermentative pathways to create mixed acids, also known as by-products (Clark, 1989). The fermentative pathway is undesirable because it promotes acetate accumulation, leading to both lower biomass and energy yields.

Limitation in carbon source may reduce the yield of biomass and protein production but excess nutrient is also a drawback. The excess of glucose in culture media can direct the culture to overflow metabolism despite the presence of oxygen. When the glucose concentration in the medium exceeds a certain concentration, instead of utilizing glucose

as energy source *via* respiratory pathway to produce ATP, the overflow of glucose flux through the TCA cycle leads to acetate generation (Contiero et al., 2000; Eiteman and Altman, 2006). Therefore, the initial concentration of carbon source or feeding strategy in fed-batch regimens should be considered carefully.

Another choice of carbon source is glycerol, which is recognized as renewable feedstock. The growth rate of *E. coli* on glycerol is much lower than growth rate on glucose, particularly, the growth rate of *E. coli* JM101 on glucose and glycerol is 0.69 h^{-1} and 0.49 h^{-1}, respectively (Martinez-Gomez et al., 2012). Glycerol cannot be readily consumed so it does not generate acid as much as glucose. Therefore, substitution glucose with glycerol is an attractive approach to reduce the acetate formation as well as overflow metabolism (Holms, 1986). Glycerol has been used in auto-induction media (Stuider, 2005), in which glycerol, glucose and lactose are compromised. It is reported that the yield of target protein obtained *via* auto-induction media is greater than those obtained by conventional IPTG induction.

2.4.3.2 Cell density at induction

Cultivation of *E. coli* is an aerobic process. This operation is conducted in stirred and aerated fermenter in various scales. It is reported that productivity of *E. coli* can reach higher than 15 g L^{-1} for protein as inclusion bodies, 5 g L^{-1} for simple monomeric protein and 0.5 g L^{-1} for complex protein (Sagmeister et al., 2014). To achieve these productivity scale, one of the most common approach is expression of the target product in high cell density culture by employing two-step cultivation, including accumulation of biomass step and induction step (Faulkner et al., 2006; Lee, 1996; Choi et al., 2006)

The cell density at induction is a crucial factor for soluble protein synthesis, which is monitored by optical density at wavelength 600 nm (OD_{600}). In the study of Liew et al. (2010) the effects of cell density at induction for murine polyomavirus (MPV) VLP production was determined. The volumetric yield was recorded at 4.38 g L^{-1} (culture media) of MPV VLP with solubility of 60.4 % when the induced at OD_{600} of 60 using fed-batch culture. Induction at OD_{600} of 100 decreased the solubility and volumetric yield. Chua et al. (2008) also investigated the effects of cell density at induction on the production of core streptavidin to find out that induction at OD_{600} of 100 achieved the highest volumetric yield.

Other examples of the effects of cell density at induction on the expression are studies of Chuan et al. (2008) and Ladd Effio et al. (2016). These studies evaluated the significance of induction at early exponential phase and late exponential phase (OD_{600} of 0.5 and 4.0) in small-scale cultivation to point out the negative impact when culture was induced at late exponential phase in the production of MPV VLP.

As is mentioned in Table 2-1, most of the HBc VLP expressed under the control of *lac* operon was induced with IPTG at early exponential phase (OD_{600} of 0.6 to 0.8) in shake-flask fermentation. In fermenter-scale, cell density at induction is varied from 20 to 100 of OD_{600}. However, only a few of expressions are expressed in soluble fraction. In terms of batch cultivation, at higher cell density, the activity of cells decreases due to the decline in nutrient concentration, the availability of dissolved oxygen, acetate accumulation and the rise of carbon dioxide in the culture. Eventually, these factors negatively impact the gene expression (Miao and Kompala, 1992).

Not only affects the protein production, the cell density at induction also has impacts on the activity of the protein of interest. Study of Olaofe et al. (2010) for amidase production demonstrates the volumetric amidase activity enhanced 70-fold when culture was induced at early exponential phase.

These studies suggest that optimisation of cell density at induction is essential for soluble intracellular protein expression.

2.4.3.3 Inducer concentration

Lactose and IPTG are two common inducers that are used to induce protein expression in *lac* promoter. The inducer concentration determines the rate of protein synthesis. The excess of inducer has negative impacts on both cell growth and protein expression, meanwhile, low concentration of inducer may lead to insufficient induction, especially in high cell density culture (Baneyx and Mujacic, 2004). It is reported that IPTG concentration will not affect the specific growth rate and biomass when the culture is induced with the concentration between 0-1 mM (Ramirez et al., 1994; Olaofe et al., 2010).

High concentration of inducer in the culture media may lead to aggressive expression of target protein *in vivo*, which increases the intracellular protein concentration and overwhelms the folding machinery of *E. coli*. This results in the formation of inclusion bodies (Fahnert et al., 2004). IPTG concentration in range of 0.5-1.0 mM has been used in expression of HBc-based VLP, listed in Table 2-1. Another approach for minimizing the

toxic effects of IPTG on culture is multiple times of induction (Hu et al., 2015). Not only increases the Pfu DNA polymerase activity, but this strategy also avoids overaccumulation of IPTG and reduces the cost of production. The nature of each recombinant protein is different, it is hard to use one IPTG concentration as standard. It must be characterized for specific strain/ vector/ protein to achieve soluble expression.

2.4.2.4 Post-induction temperature

The optimal temperature for *E. coli* growth is 37 ^{o}C. This temperature may cause plasmid loss and stimulate the hydrophobic interactions that are responsible for aggregation reaction (Hunke and Betton, 2003; Strandberg and Enfors, 1991). It is possible to interfere the accumulation of protein aggregates by partially controlling the factors that influenced the aggregation reaction. One of the common practices is to reduce the protein synthesis rate. The post-induction temperature is adjusted to lower than 37 ^{o}C.

This approach offers the newly expressed protein time to fold properly and overcome the overcrowded intracellular protein in cytoplasm (Schein, 1989; Vera et al., 2007). Furthermore, in the scenario in which the target protein is expressed as insoluble form, the protein quality in the inclusion bodies is influenced by temperature. For instance, the activity Aβ42 Alzheimer peptide in the aggregates inversely correlates with the temperature (De Groot and Ventura, 2006).

Regarding HBc-based VLP expression in *E. coli*, the study of Ravin et al., 2015 demonstrates the success of soluble expression of M2e-HBc VLP, vaccine candidate against influenza. The culture was induced with 0.5 mM of IPTG and was cultivated at 20-30 ^{o}C for 16-18 h. Meanwhile, the vaccine candidate for PEDV was expressed as inclusion bodies when the culture was induced at 28 ^{o}C with 1 mM of IPTG (Gilliam and Zhang, 2018).

The effect of temperature on the solubility of His-tag HBcAg has been investigated by Yap et al. (2009). The culture was induced with IPTG of 0.5 mM and incubated for 16-18 h at temperature of 27, 30, and 37 ^{o}C. The greatest soluble yield was recorded at 75 mg L^{-1} (culture media) with the solubility of 95 % when the culture was incubated at 30 ^{o}C. It is reported that low temperature incubation (30 ^{o}C) also enhanced the activity and expression of *E. coli* chaperones (Mogk et al., 2002). Moreover, low expression temperature (28 ^{o}C) and co-expression of Trigger Factors (TF) and *E. coli* chaperones GroEL–GroES improved

the soluble production of GST–VAS (vasostatin 120–180) by 6-fold in comparison to the yield at 37 ºC (Sun et al., 2005).

Although expression at low temperature enhances the soluble yield, low temperature may inhibit the cell growth, resulting in a decline of biomass. These arguments should be taken into consideration when evaluating the operating range.

2.4.3.5 Dissolved oxygen level

Bacterial metabolism depends on the availability of oxygen level in the culture, particularly the demand of oxygen for cellular respiratory when the culture reaches high cell density (Whiffin et al., 2004). It is essential to maintain the aerobic condition for supporting high cell growth and providing oxidising environment for protein synthesis (Phue and Shiloach 2005).

Oxygen limitation is classified as a stress for culture. *E coli* will adjust their own metabolic capacity to fit their energy generation and availability of oxygen (Bettenbrock et al., 2014). In aerobic metabolism, in electron transport chain, oxygen is served as terminal electron acceptor (Anraku, 1988). Molecular oxygen directly influences the generation of adenosine triphosphate (ATP), an energy currency that involves in various biological activities in cell, including metabolic changes, protein oxidation, DNA oxidation, protein synthesis and plasmid replication (Konz et al., 1998; Henkel et al., 2014). Therefore, a fluctuation in oxygen supply can cause intracellular oxidative stress, affecting the quality of the final product.

Kim et al. (2012) demonstrates that high intracellular levels of ATP enhanced the production of green fluorescence protein in *E. coli*. In agreement with the findings, Kim et al. (2013) discovered that ATP hydrolysis is required for folding of non-native protein, for example, the GroEL-GroES and DnaK systems requires ATP for functioning. ATP has been proved to be involved in Human Immunodeficiency virus (HIV) VLP assembly (Tritel and Resh, 2001).

It is reported that even a short exposure to anaerobic condition, pre-proinsulin productivity can be substantially reduced by intermittent switching between aerobic and mixed acid fermentation (Sandoval-Basurto et al., 2005). The formation of by-products, such as acetate, lactate, and succinate causes the loss in the biomass and protein yields.

There are several approaches to maintain dissolved oxygen level in a bioprocess culture. They include 1) changing the agitation speed of the impeller (Shin et al., 1996),

2) increasing fermenter pressure (Ma et al., 2010), 3) supplying oxygen-enriched air (Ma et al., 2010); and 4) using oxygen vector to enhance the solubility of oxygen in culture medium (Pilarek et al., 2011).

Significant increase in agitation speed or aeration may lead to foam formation and enormous power consumption, which has negative impact on the cost of production. Addition of hydrophobic liquids organic solvents with higher affinity for oxygen, such as, immiscible oxygenated oils (Sklodowska and Jakiela 2017), alkanes (Zhang et al., 2018; Li et al., 2012), paraffin oil (Narta et al., 2011), oleic acid (Liu et al., 2016), and perfluoro-chemicals (Westbrook et al., 2018), is able to reduce the surface tension of water and retain oxygen molecules in the two-phase system (Rols et al., 1990; Nielsen et al., 2003), eventually increase the oxygen transfer rate in the reactor.

2.4.3.6 Batch and fed-batch in fermenter scale cultivation

All the mentioned fermentation process factors in previous section apply to both shake-flask and fermenter-scale. In fermenter-scale cultivation, as the high cell density and large culture volume, the consumption of nutrient is enormous. Consequently, guarantee of an adequate supply of nutrients is of great importance.

Initially, in batch cultivation, the availability of nutrients is defined, the cultivation ends when culture medium is depleted. This provides an overflow of nutrient for non-limiting growth for *E. coli* at the start of cultivation and possibility of nutrient shortage at the end. These could result in the formation of unwanted metabolites, such as acetate, formate (Szenk et al., 2017), leading to the inhibition of cell growth and decrease of the production yield of target recombinant protein (Lee, 1996). Despite of the availability of nutrient, the maximum growth rate is also temperature-dependant (Kovarova et al., 1996). Therefore, the initial carbon source concentration and temperature should be determined with caution regarding the technical or physical constraints.

Following the batch phase, fed-batch is initiated by addition of the supplementary media when the culture shows the sign of the nutrient limitation. A sudden rise in pH or dissolved oxygen level indicates the growth limiting substrate (carbon source) is depleted.

There are several types of feeding regimes to control the growth rate of the culture and avoid the overflow of nutrient (Lee, 1996). They are classified into two main categories: 1) feeding strategies without feedback control (linear or exponential) and 2) with feedback control (DO-stat or pH-stat) feeding strategies (Mears et al., 2017). Since there is no

“one size fits all” in protein production, each recombinant protein has favoured expression condition. The efficiency of each feeding strategy should be investigated.

In fed-batch cultivations, feeding strategies without feedback control (feed forward) are mainly based on the cell growth rate which is strongly influenced by feeding rate. This feeding rate is determined by the initial biomass concentration of the fed-batch phase, the current working volume of the culture, the biomass yield coefficient, the concentration of feed solution, and the desired specific growth rate (Kim et al., 2004). Meanwhile, in feedback control feeding strategies, the feeding solution is supplied under the control of DO or pH level of the cell culture.

DO-stat feeding based on the concept of DO rises because of the reduction of respiration (oxygen consumption by cell) as the nutrient is limited. During the cultivation, the configuration of fermenter such as agitation speed and aeration rate are fixed, the DO-stat maintains the culture at constant DO set point by controlling the nutrient feed rate in relation to DO level. Nutrient is supplied when DO rises higher than the set point and nutrient feeding rate is stopped or reduced when DO drops below the setpoint.

In pH-stat feeding, the pH of the culture increases as the carbon source is depleted. Similar to DO-stat mechanism, pH of the culture is maintained at constant level by nutrient feeding. Nutrient is supplied when pH rises higher than the set point and nutrient feeding rate is stopped or reduced when pH drops below the set point. During the cultivation, the DO level of the culture is usually controlled in cascade mode. Auto-adjustment of agitation speed and aeration is performed by proportional–integral–derivative controller (PID) controller of the fermenter.

Therefore, not only nutrient is important, but the availability of oxygen throughout the cultivation is also a growth limiting factor (Farrell et al. 2015). In general, the oxygen uptake rate (OUR) of the cell has to be smaller than oxygen transfer rate (OTR) to ensure the physiological condition of the working cell. In fed-batch cultivation, culture will reach to a cell density where OUR larger than OTR in which the reactor is unable to adjust its impeller agitation and gas aeration velocities to cover the demand of oxygen in cascade control, leading to oxygen limitation. This limitation switches the metabolism pathway of *E. coli*, from respiratory to fermentative metabolism (Spiro and Guest, 1991).

Farrell et al. (2012) conducted pH-stat fed-batch cultivation to produce 41 kDa recombinant protein vaccine candidate. The study demonstrates the constraint of maintaining the DO level in 200-L bio-reactor when using pH-stat. Large bio-reactor has

its agitation speed limit (approximately 500 rpm), which narrows down the maximum volumetric oxygen mass transfer coefficient (k_La) to 0.06 (s^{-1}).

For a consistent process performance between reactor scales, DO-stat has more advantages since the method use the feeding rate to maintain the dissolved oxygen level at a designated set-point to avoid the oxygen limitation (Farrell et al., 2015).

As a result of using *E. coli* as expression host, following six (6) key aspects were discussed. They are:

1. General recombinant protein production process includes 1) design of expression vector, 2) selection of host strain and transformation, 3) growth of bacteria and protein production, 4) protein recovery and purification
2. Potential problems in fermentation process include cell viability, plasmid loss, and protein folding. High expression rate of recombinant protein with poor protein folding rate leads to accumulation of inclusion bodies
3. Protein folding state determines the bioactivity of the protein. The bioactivity of the target protein recovered from inclusion bodies is not guaranteed. It is feasible to control the protein synthesis rate *via* process factors to improve solubility of the expression
4. Expression vectors and choice of promoter determine the host strain
5. Process factors that influence the soluble production of recombinant protein include post-induction temperature, the cell density and pH level at induction, inducer concentration, metal ions, and dissolved oxygen level in both shake-flask and fermenter-scale
6. Feeding strategy in fed-batch cultivation is a crucial factor.

The recombinant protein production process, potential problems and process factors contributed to soluble production of target protein in *E. coli* have been reviewed and summarised. Because of the interactions between process factors, to extrapolate the insights from these process factors, statistical approaches are highly regarded.

2.5 Statistical approach – Design of Experiment (DoE)

As is discussed in section 2.4, there are many variables that influence the soluble production of HBc-based VLP in *E. coli*. Besides of genetic modification, fermentation process factors are also crucial to soluble production of chimeric HBc-based VLP.

One-factor-at-a-time (OFAT) approach is described as the method of changing one independent variable and while the other factors remains unchanged (Frey et al., 2003). Experiment continues with the determination of the next factor based on the optimal condition of previous factor. It has been applied for several decades to improve the performance of biopharmaceutical fermentation process.

The statistical approach, also known as design of experiment (DoE), offers a controlled model to screen and optimise the process factors to achieve single or multiple optimal responses by exploring the relationship between variables and experimental responses (Myers et al., 2016).

Table 2-3 demonstrates the comparison between conventional and statistical approaches, i.e. factorial design and response surface design. In general, although OFAT is simple, it requires intensive labour work and is time-consuming. In addition, the interactive effects among the variables studied are lost, leading to an incomplete understanding of system behavior and the approach. Ultimately, OFAT fails to identify the true optimum of process settings (Abou-Taleb and Galal, 2018). Not only beneficial in time and cost of experiment, the amount of information extrapolated from the statistical design is significantly greater than those from OFAT with fewer experiment runs. This approach provides a mean to navigate the design space. In recombinant protein expression, from the review by Papaneophytou and Kontopidis (2014), all of the process factors during the expression are interconnected, it is feasible that the interaction between factors has higher influence than the effect of single factor (Noguere et al. 2012).

Another key advantage of DoE is the sequential experiment. A typical DoE workflow includes the planning, screening and optimisation. Screening experiment is completed in factorial design (first-order models) and optimisation is achieved using response surface methodology (polynomial equation) (Bezerra et al., 2008; Uhoraningoga et al., 2018; Snedecor and Cochran, 1989). Because of these arguments, statistical approaches offer ideal means for optimisation of soluble production of HBc-based VLP.

Table 2-3: Comparison between conventional and statistical approaches.

	Conventional approach	**Statistical approaches**	
Design	**-**	**Factorial**	**Response surface**
Popular design model	One-factor-at-a-time	i. Two-level full factorial ii. Fractional factorial iii. Placket -Burman	i. Central composite ii. Box-Behnken
Aim	Optimisation	Screening/Optimisation	Optimisation
Assessment	Linear equation with one variable	Linear equation of multiple variables and their two-factor interaction	Quadratic equation with multiple variables and their two-factor interaction
Advantages	i. Both continuous and categorical factors ii. Simple and straightforward	i. Both continuous and categorical variables ii. Knowledge of main effects and interaction between factors iii. The number of experiment runs can be reduced using fractional factorial design iv. Prediction in design range v. Non-bias vi. Computer-aid	i. Graphical or numerical option for point optimisation ii. Extensive information on experimental factor effects and interaction between factors iii. Less number of factors but greater number of levels for each factor. iv. Prediction in design space v. Non-bias vi. Computer-aid
Disadvantages	i. Large number of experiments runs ii. Time consumption iii. Expensive iv. Loss of interaction between factors v. Lack of predictive ability.	i. Only investigated the high (+) and low (-) level of each factor ii. Approach become expensive in relation to the numbers of factors iii. The number of experiments determines the amount of information extrapolate from the design iv. Possibility of sequential experimentation	i. Continuous variables only ii. Requirement of screening experiment or knowledge to design the experimental matrix iii. The investigated range of design factors should be carefully determined to achieve curvature iv. Sometimes optimal combination of factors does not work for multiple responses

2.5.1 Screening experiments by full/fractional factorial design

Screening experiment focuses mainly on media composition and culture conditions (Swalley et al., 2006; Marini et al., 2014). To evaluate all the variables that influence the soluble production of HBc-based VLP in microbial expression system, a wide range of factors should be included in experimental screening to narrow down the significant factors (Mandenius and Brundin, 2008; Box et al., 2005).

The classical designs for screening purpose include full factorial design, fractional factorial design and Plackett-Burman designs. The choice of design depends on the screening purpose. In terms of effects achieved by the model, full factorial design provides information regarding the main effect and two-factor interactions in exactly the same way; fractional factorial design offers either only the main effect or both main effects and two-factor interactions based on the resolution of the design; and Plackett-Burman design offers main effect only (Montgomery, 2012).

Although the protein expression condition has been intensively studied, the selection of variables and their levels should be appropriate, following the pre-knowledge to avoid experimental bias and bad judgement (Mandenius and Brundin, 2008).

In case of lacking pre-knowledge of expression of protein of interest, it is recommended to conduct a screening experiment to examine as many factors as possible. A full factorial design, 2^k, is an ideal experimental design. The controlled model includes two coded levels for each factor, lower (-) and (+) higher value, and k is number of factors, which is larger than two (2). The drawback of full factorial design is the number of experiments increases in exponential manner if the number of factors is large.

Full factorial design of experiment has been used to optimise the soluble expression of VP1 of murine polyomavirus VLP in the study of Chuan et al. (2008) and Ladd-Effio et al. (2016). These studies suggest that strategical modification of host cell and optimal induction strategy are better alternatives to boost the soluble expression of recombinant protein rather than codon optimisation.

In comparison, fractional factorial design is recommended when factors number is larger than five (5), this design offers less experimentation but come with the cost of losing some information about potential interactions (Papaneophytou and Kontopidis, 2014). The design is described as $2^{(k-p)}_{R}$, in which 2 is the number of levels, k is the number of factors, p extra columns required, and R is the resolution of the design. Resolution is determined as how much the effects in a fractional factorial design are aliased with other effects.

A significant feature of full or fractional factorial design is the evaluation for both continuous factors (temperature, induction time, IPTG conc., etc.) and categorical factors (cell line, expression vector, fusion tag, culture media, etc.) (Papaneophytou and Kontopidis, 2014). For example, to boost the soluble expression of pneumolysin (250 mg L^{-1}), Marini et al. (2014) studied the effects of induction time, IPTG conc., expression temperature and media composition using fractional factorial design.

To rank the effects from design factors, the half-normal probability plot coupled with Lenth's method (Lenth, 1989) is an established statistical technique (Taylor, 1994). The absolute value of the effect estimates is plotted against their cumulative normal probabilities. Significant effects of design factor will be shown as outliers from the straight line that passing through the origin and close to the fiftieth percentile data value. Analysis of variance and residual plot are integral part of regression model building.

After completing the observation from experimental design, a quantitative relationship (equation) between the response and important design factors is established. In general, a low-order polynomial model is appropriate for screening or characterisation purpose. Equation 2-1 presents a common model includes main effects and interaction between two-factors:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \ldots + \varepsilon \qquad \textbf{(Eq. 2-1)}$$

where y is the response, β is the unknown parameter that is estimated from the data in the experiment, the X is the design factors, the cross-product term X_1X_2 is the two-factor interaction between the design factors and ε is the experimental error in the system (Montgomery, 2012)

The limitation of these designs is they can only generate first-order equation, which is inadequate to demonstrate the curvature of the process. The importance of each factor influencing protein production is estimated using statistical software such as Minitab™ (Minitab Inc., State Collage, PA), and Design-Expert™ (Stat-Ease©, Minnesota, USA).

2.5.2 Optimisation experiment by response surface methodology

After identifying the significant process factors, simultaneous optimisation of the greater number of levels of the investigated variables is necessary to achieve the best system performance, which can be done by response surface methodology (RSM). The optimal objective can be a single or a set of responses, for example, biomass, volumetric yield and solubility of expressed protein (Papaneophytou and Kontopidis, 2012; Muntari et al., 2012; Puente-Massaguer et al., 2020).

The RSM uses a collection of statistical and mathematical techniques to achieve the fitting of the experimental data to the quadratic equation (second-order equation) (Montgomery, 2012), increasing the precision of the model, compared to fractional factorial design. Equation 2-2 demonstrates the relationship between design factors and response is similar to Eq. 2-1 with the extension of higher order of interaction ($\beta_{ii}(X_{ii})^2$):

$$y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_{12}X_1X_2 + \beta_{11}(X_{11})^2 + \beta_{22}(X_{22})^2 + \ldots + \varepsilon \qquad \textbf{(Eq. 2-2)}$$

Usually, before applying the RSM methodology, it is essential to have pre-knowledge about the expression of the protein of interest or perform screening experiment to choose the optimal plasmid vector and strain (Papaneophytou and Kontopidis 2014). Experimental designs for response surface methodology include Box–Behnken (BBD), central composite designs (CCD), and full three-level factorial designs (Bezerra et al., 2008). Both CCD and BBD are widely used in optimisation of protein expression. For example, Balderas et al. (2008) utilised the BBD to enhance the production of human interferon gamma by 13-fold. The expression of extracellular Zinc-Metalloprotease (SVP2) protein was enhanced by 15-fold using CCD (Beigi et al., 2012).

Central composite design, first presented by Box and Wilson (1951), consists of three components: 1) a full or fractional factorial, 2) an axial point (star design) and 3) central points. An optimisation study involves four (4) variables will generates a matrix of 30 experiments, covering five (5) levels of each factors (-2, -1, 0, 1, 2). Replication of centre point provides the measurement of reproducibility and model lack of fit (Ferreira et al., 2007). Based on three-level incomplete factorial design, BBD is composed of the middle points of the edges and centre point from a cube, leading to a spherical, rotatable second-order design (Box and Behnken, 1960). With 4 variables, BBD only need twenty-seven (27) experiments. Although the smaller number of run benefits the time and cost, the lower number of experiments leads to smaller degrees of freedom.

Fig. 2-7 demonstrates the examples of a 3-factor CCD and a 3-factor BBD. The number of experiments in 3-factor BBD is less than the number of experiments in 3-factor CCD. However, because of the difference in localisation of experimental points, the predictive ability of the model to calculate the response of the new experimental point, particularly the extreme combinations, is affected. CCD contains the extreme factor combinations, whereas BBD does not examine borderline region. Therefore, CCD has better prediction than BBD (Zolgharnein et al., 2013; Rakic et al., 2014).

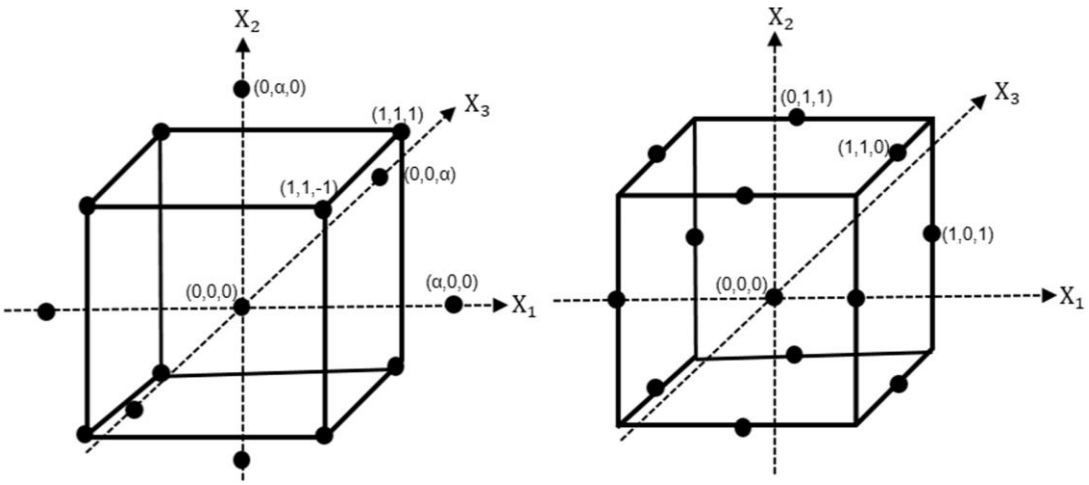

Fig. 2-7: Example of a 3-factor central composite design (left) and a 3-factor Box-Behnken design (right) (adapted from Ockuly et al., 2017)

In case of optimising of serval responses, changing the level of a factor can have positive effect on a specific response and negative effects on the others. It is common that all surfaces found do not provide the optimal condition for the data set. There are two main approaches to find the optimum regions, visual inspection and Derringer function, also known as desirability function (Murphy et al., 2005; Derringer and Suich, 1980).

The design of experiment with their special properties compared with conventional approach (OFAT) has shown their advances in boosting the soluble expression of recombinant protein in *E. coli*, particularly VLP vaccine candidate. Therefore, it is wise to employ statistical approach to optimise the expression of HBc VLP as soluble form and navigate the operational process parameter range before scaling-up the production in industrial scale.

2.6 Chapter summary and conclusions

Based on a critical review of the literature, the following important factors emerge:

1. Hepatitis B core virus-like particle is a promising platform for vaccine development. Compared with conventional vaccines, it is more advanced because it is potentially safer and has the flexibility to carry epitope of interest at three different insertion sites to induce the immune response. Two epitope models, EBNA1 and HCV core have been reviewed in terms of structure and immunogenicity aspects. HBc VLP is a thermal stable protein that is beneficial for protein purification.

2. *E. coli* has advantages in recombinant protein expression. These include, high yield of product, the availability of well-studied hosts and vectors, short cultivation time and, process economics. However, solubility of the expression is a potential challenge. This is because mis-folding protein accumulates insoluble protein into inclusion bodies. The expression of chimeric HBc-based VLP tends to be in inclusion bodies. The soluble expression of chimeric HBc-based VLP is highly preferred so the fermentation process factors, including media composition, the concentration of inducer, cell density at induction, post-induction temperature, dissolved oxygen level during the expression and feeding strategy, need to be optimised.

3. One-variable-at-a time is not a good approach because multiple factors and biological reactions occur at the same time in biological process. Statistical approaches provide a platform to identify the most influence factors and interaction for a specific response. In general, statistical approach consists of 1) fractional factorial design (FFD) and 2) response surface methodology (RSM). FFD can be used to rank the effects of design factors and RSM provides the optimal setting and prediction with higher precision.

In the next chapter, Chapter 3, the expression of HBc VLP carrying HCV core peptide is investigated in shake-flask fermentation. Fractional factorial design is employed with six (6) factors and two (2) levels. The protein is characterized with size-exclusion chromatography and transmission electron microscopy. The impacts of factors on experimental outcome are statistically demonstrated and critically discussed.

CHAPTER 3

ENHANCEMENT OF SOLUBLE EXPRESSION OF *ESCHERICHIA COLI*-DERIVED VIRUS-LIKE PARTICLES CARRYING STRUCTURAL EPITOPE BY FACTIONAL FACTORIAL DESIGN

3.1 Introduction

A review of the literature (Chapter 2) showed that the microbial expression system is able to express HBc VLP, particularly inducible expression system.

E. coli has been used widely as an exceptional host because it is well-characterized and is readily available in a range of commercial strains. As is mentioned in Chapter 2, chimeric Hepatitis B core (HBc) virus-like particle (VLP) features intracellular self-assembly properties. Although it is reported that *E. coli*-derived VLP can only be expressed as naked VLP (Pushko et al., 2013), the size of naked VLP is considerably large to *E. coli* folding machinery. Moreover, T7 expression system triggers strong and rapid over-expression of recombinant protein when exposures to IPTG, putting the host cell into metabolic burden (Glick, 1995). Gasser et al. (2008) reviewed the protein folding and conformational stress in microbial cell, suggesting that when the rates of protein synthesis overwhelm post-translational modifications such as folding machinery, the incorrectly folded intermediates often accumulate.

Therefore, soluble expression of recombinant protein from *E. coli* is limited and it is not beneficial if the product is expressed as inclusion bodies. Inclusion bodies complicate the purification process and refolding solubilized inclusion bodies to a bioactive form is a challenging task and protein dependent (Burgess et al., 2009). Besides of strain and plasmid, many fermentation process factors are being manipulated to reduce inclusion bodies formation and enhance soluble expression of recombinant protein, typically post-induction temperature, IPTG concentration, pH level at induction, the cell density at induction time (Chen, 2012; Papaneophytou and Kontopidis, 2014).

It is not known how the inserted epitope and process factors influence the soluble expression in a systematic way. Evidence from the literature is strong nevertheless that most of study is focused on synthesis new vaccine candidate and their immunogenicity, not many studies are focused on improving the soluble production of HBc-based VLP. Because of significant responsive regulatory systems in *E. coli*, cells rapidly adapt to environmental change to secure survival. Therefore utilization of conventional experimental design, and trial and error, will be time-consuming and require laborious effort. This is unlikely therefore a good method for study (Uhoraningoga et al., 2018).

A statistical method using the design of experiments can reduce the number of the experiments but still extract significant information, including any interactions between factors. Fractional factorial methods can include both categorical and continuous variables

(Wu and Hamada, 2009). It has been applied to improve the production of VP1 of murine polyomavirus VLP in the study of Chuan et al. (2008) and Ladd-Effio et al. (2016). This is therefore suitable to this research into enhancing soluble production of recombinant protein in *E. coli*.

Hepatitis C virus infection is a cause of liver cirrhosis and hepatocellular carcinoma (Caselmann and Alt, 1996). The potential of using HBc as carrier for HCV vaccine candidate was reported by Mihailova et al. (2006). HBc VLP bearing virus-neutralizing HBV pre-S1 epitope and HCV core epitope was also reported by Sominskaya et al. (2010). Both demonstrated the flexibility of the HBc protein as a vaccine carrier.

Aim of this chapter was to boost the soluble production of HBc VLP carrying structural epitope in *E. coli*. A set of 32 experiment was conducted to identify the effect of the main fermentation process factors on the soluble expression of HBc-based VLP using fractional factorial design in shake-flask fermentation. The structural epitope, HCV, has 44 amino acids and is able to form into α-helices structure. Because of this, it is used as model to investigate the impact of parameters on soluble production of HCV-HBc VLP.

It is hoped that insight gained from this chapter could be used to provide needed technical insight to navigate the operating range of process parameters in laboratory fermenter and methods of up-scaling.

3.2 Material and methods

3.2.1 Expression strain

The epitope of Hepatitis C Virus core protein was inserted at N-terminus of HBc protein to form chimeric HCV-HBc (HBc: 183 amino acid). The amino acid sequence of HCV core is KTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTS). The plasmid pET-30a (+) (Invitrogen ®, USA) was used to form pET-30a (+)/HCV-HBc, carrying the gene of HCV-HBc protein under control of T7 promoter with *lac* operon.

Via heat-shock transformation the plasmid was transformed into *E. coli* by using One Shot™ BL21 (DE3) chemically competent *E. coli* (Invitrogen©, USA). Cell were cultured on Luria Bertani (LB) agar plate. Cells were stored at minus 20 ^{o}C as the working cell.

HBc protein (the carrier – without epitope) was also transformed into *E. coli* as control.

All reagents were analytical grade (AR) unless otherwise stated.

3.2.2 Fractional factorial experimental design

Table 3-1 shows a fractional factorial design that was used to boost the soluble expression of HCV-HBc. Design included 32 expression condition, made up from six (6) design factors with two (2) designated levels as low (-1) and high (+1) level. Process factors were presented as coded value. They included (**A**) post-induction temperature (30 ^{o}C and 40 ^{o}C), (**B**) post-induction rotation speed of the shaker (120 rpm and 250 rpm), (**C**) pH level at induction (6.2 and 7.2), (**D**) isopropyl β-D-1-thiogalactopyranoside (IPTG) concentration (0.2 mM and 0.8 mM), (**E**) optical cell density measured at wavelength 600 nm at induction (OD_{600} of 0.8 and OD_{600} of 2.0), and; the (**F**) presence of Mg^{2+} (0 mM and 10 mM).

Three (3) responses were recorded, they were: 1) biomass per litre of culture media (g L^{-1}), 2) the volumetric yield of target protein per litre of culture media (mg L^{-1}), and 3) the yield of soluble VLP per dry cell mass (mg g^{-1}), which was set as main response for the experimental design. To clearly determine the main effects of process factors and two-factor interaction, a 2^{6-1} fractional factorial design consists of 32 experiments was used. Data analyses were performed by Design-Expert™ (version 11, Stat-Ease©, Minnesota, USA) where the response was kept as normal (no transformation required by the Box-Cox method) (Box and Cox, 1964). As is included in Design-Expert, half-normal probability plot and analysis of variance (ANOVA) were computed to determine the significance of each factor.

Table 3-1: Factors and coded levels of the fractional factorial design for HCV-HBc VLP.

Exp. no.	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
	A: Temperature	B: Rotation speed	C: pH	D: IPTG conc.	E: Cell density	F: Mg^{2+}
	(-1): 30	(-1): 120	(-1): 6.2	(-1): 0.2	(-1): 0.8	(-1): 0
	(+1): 40	(+1): 250	(+1): 7.2	(+1): 0.8	(+1): 2.0	(+1): 10
	ºC	rpm		mM	OD_{600}	mM
1	-1	-1	+1	-1	-1	+1
2	-1	-1	+1	+1	-1	-1
3	+1	+1	-1	-1	-1	-1
4	-1	-1	-1	-1	+1	+1
5	+1	+1	+1	-1	+1	-1
6	+1	+1	+1	+1	-1	-1
7	+1	+1	+1	-1	-1	+1
8	-1	+1	+1	+1	-1	+1
9	-1	+1	+1	+1	+1	-1
10	+1	-1	+1	+1	-1	+1
11	-1	-1	+1	-1	+1	-1
12	+1	+1	-1	+1	+1	-1
13	+1	+1	-1	+1	-1	+1
14	-1	+1	-1	+1	+1	+1
15	-1	+1	-1	-1	-1	+1
16	+1	-1	-1	-1	+1	-1
17	-1	+1	+1	-1	-1	-1
18	+1	-1	-1	-1	-1	+1
19	-1	-1	+1	+1	+1	+1
20	+1	+1	-1	-1	+1	+1
21	-1	+1	-1	+1	-1	-1
22	+1	-1	-1	+1	+1	+1
23	+1	-1	+1	+1	+1	-1
24	-1	-1	-1	+1	-1	+1
25	+1	-1	+1	-1	-1	-1
26	-1	-1	-1	+1	-1	-1
27	-1	+1	-1	-1	-1	-1
28	-1	+1	+1	-1	+1	+1
29	+1	+1	+1	+1	+1	+1
30	+1	-1	+1	-1	+1	+1
31	-1	-1	-1	-1	-1	-1
32	+1	-1	-1	+1	-1	-1

3.2.3 Expression of HCV-HBc VLP

The *E. coli* transformed with the pET-30a/HCV-HBc were cultured in 50 mL of sterilised Luria-Bertani media (10 g L^{-1} tryptone (Oxoid, USA), 5 g L^{-1} yeast extract (Oxoid, USA), 10g L^{-1} NaCl (Chem-supply, Australia) in 250-mL Erlenmeyer flask, supplemented with 50 µg mL^{-1} kanamycin sulfate (Invitrogen©, USA), at 37 °C and 200 rpm (round per minute) for 16 h.

The overnight culture was added into 200 mL of fresh sterilized LB media volume with 50 µg mL^{-1} kanamycin sulfate, in 500-mL Erlenmeyer flask at a ratio of 1 v/v % (seeding/culture volume), and cultivated at 37 °C and 200 rpm. $MgCl_2$ (Chem-supply, Australia) was added to the culture following the experiment design, as is listed in Table 3-1. The optical density (OD_{600}) was measured by UV-1600PC Spectrometer (VWR International©, USA).

The expression of HCV-HBc was conducted followed fractional factorial design matrix, as is presented in Table 3-1. Experiments were conducted in randomised sequence to avoid bias. When culture reached to designated value of cell density, the fermentation process factors were adjusted according to FFD matrix. K_2HPO_4 2 M (Chem-supply, Australia) and KH_2PO_4 0.5 M (Chem-supply, Australia) were used to adjust the pH of the culture at induction. IPTG (Invitrogen©, USA) was added to the culture to express the target protein. Temperature and rotation speed of the shaker were also adjusted. The expression was continued for the next six (6) h.

Before cells were harvested by centrifugation (3700 rpm, 20 °C, 30 min), the final cell density was measured to obtain the biomass concentration. The cell pellet was washed with deionized water, then stored at minus 20 °C for further analyses.

Cells were lysed at a ratio of 1 g wet cell weight (WCW) per 10 mL of lysis buffer (20 mM of Tris-HCl (Chem-supply, Australia), 0.1 v/v % Triton X-100 (Glentham, United Kingdom), 3 mM EDTA (Invitrogen©, USA), pH 8.0) by ultra-sonication (350 W, 10 min) on ice-bath (0 °C). Each cycle has a pulse of four (4) s on with six (6) s off. Cell crude lysate was clarified by centrifugation (12000 rpm, 4 °C, 30 min) to separate supernatant fraction (soluble protein) and insoluble precipitate (inclusion bodies). These fractions were further analysed by 12 % reducing SDS-PAGE.

The supernatant fraction was identified as clarified crude lysate. Protein concentration of each fractions was determined by Bradford assay (Bradford, 1976). Bovine serum albumin (Sigma-Aldrich, USA) was used as standard.

To determine the constant number between dry cell weight (DCW) and optical density (OD_{600}), 10 mL of culture media at different optical density was centrifuged. The cell pellet was collected and washed with deionized water then dried in oven at 60 ^{o}C for 24 h to remove water. In one (1) L of culture media at OD_{600} of 1 has 0.4 g DCW.

It is hypothesised that in the clarified crude lysate, there were: 1) HCV-HBc soluble monomer protein, 2) soluble aggregation, and 3) soluble HCV-HBc VLPs. Reducing SDS-PAGE 12 % (w/v) was used to break down the soluble aggregation and soluble VLPs into HCV-HBc protein monomer. The percentage (%) of HCV-HBc protein expressed in clarified crude lysate was estimated from the gel densitometry. High-performance size-exclusion chromatography (HPSEC) was used to quantify the soluble aggregates.

The soluble HCV-HBc VLP mass (mg) is calculated, following Equation (3-1), where **X** is the percentage of protein monomer obtained from SDS-PAGE gel densitometry, **Y** is the percentage of soluble aggregates obtained from HPSEC and **Z** is the total soluble protein mass of clarified crude lysate.

$$\text{Soluble}_{\text{HCV-HBC VLP}}\ \text{mass (mg)} = [X\ (\%) - Y\ (\%)] \times Z\ (\text{mg}) \qquad \textbf{(Eq. 3-1)}$$

The calculated soluble HCV-HBc VLP mass (mg) is converted to volumetric yield (mg L^{-1}) of culture media and cellular yield of protein (mg g^{-1} DCW) over biomass (g).

3.2.4 Gel densitometry of SDS-PAGE

For evaluation of monomer HBc-based VLP expression clarified crude lysate from each designed expression condition was separated on SDS-PAGE gel (Polyacrylamide 12 w/v %) using Mini-Protean Handcast System (Bio-rad©, USA). Sample was mixed with 5X SDS sample-loading at ratio 4:1 (v/v). Samples were heated on heat block at 100 ^{o}C for 20 min to denature the protein sample. Sample was deposited into the gel and was resolved at 100 V for 1.5 h at 25 ^{o}C in 1X SDS-PAGE running buffer. The gel then was staining with Coomassie Blue (Bio-rad©, USA). After de-staining, gel was captured by Syngene G-box (Syngene©, USA). The gel densitometry was analysed by *Image J* software (Schneider et al., 2012). VLPs collected from HPSEC peak was detected by reducing SDS-PAGE with silver staining using SilverQuest™ Silver Staining Kit (Invitrogen©, USA). The staining procedure was followed SilverQuest™ Silver Staining Kit manual.

The component of sample loading buffer, running buffer, staining solution and de-staining solution was summarized in Appendix B.

3.2.5 High-performance size-exclusion chromatography (HPSEC)

HPSEC has been used to analyse the ratio between soluble-aggregates and soluble VLPs capsid (Yang et al., 2015). The measurement was performed on the Prominence HPLC system (Shimadzu, Japan) equipped with TSK-G5000 GW_{XL} (300 × 7.8 mm, I.D., pore size 100 nm) (Tosoh Bioscience, Germany) with UV monitoring at wavelength 260 nm and 280, nm. Prior injection, all clarified crude lysate samples were filtered with 0.2 µm membrane and 50 µL of the filtered sample was resolved at flowrate 0.3 mL min^{-1}. The mobile phase was phosphate buffer (50 mM, pH 7.4) (Chem-supply, Australia). Peak integration was analysed by Lab Solution (Shimadzu, Japan).

3.2.6 Transmission electron microscopy

To confirm VLP structure, a 5 µL semi-purified HCV-HBc VLP from HPSEC fraction, was applied on the carbon-coated grid and stained with 5 µL 1 w/v % aqueous uranyl acetate. Micrographs were recorded by FEI Tecnai G2 Spirit TEM at 100 kV (FEI Company, Japan). Images were captured at 100 nm scale.

3.2.7 Expression of HBc carrier protein

To determine whether HCV epitope has influence on the soluble expression of HBc VLP, the *E. coli* transformed with the pET-30a/HBc were cultured in LB media under the optimal expression condition of HBc-VLP. The cultivation, expression and analysis procedure are followed as is described in section 3.2.3. The volumetric yield and cellular yield of HBc VLP are estimated.

3.3 Results

3.3.1 Characteristic of HCV-HBc VLP in *E. coli*

In the preliminary experiment, it was found that HCV-HBc was expressed in the insoluble form of inclusion bodies when culture was incubated at conventional condition (37 ^{o}C, OD_{600} of 0.8, IPTG of 1 mM and 200 rpm). Therefore, it is essential to optimise the fermentation process factors to achieve soluble expression of HCV-HBc. These factors, including: 1) post-induction temperature, 2) post-induction rotation speed of the shaker, 3) pH level at induction, 4) IPTG concentration, 5) cell density at induction, and; the 6) presence of Mg^{2+}, were investigated to determine their effects on the response, cellular yield (mg g^{-1} DCW). The detail of expression condition is listed in Table 3-2.

To investigate the expression form (protein folding state) either of soluble or insoluble protein, SDS-PAGE is applied. Fig. 3-1 presents the HCV-HBc folding state in two different expression conditions, namely, experiment (exp.) no. 28 (30 ^{o}C, OD_{600} of 2.0, 0.2 mM of IPTG and 250 rpm) as Fig. 3-1A and exp. no. 21 (30 ^{o}C, OD_{600} of 0.8, 0.8 mM of IPTG, and 250 rpm) as Fig. 3-1B. Cell pellet was lysed and cloudy cell lysate (Lane 1) was further centrifuged to obtain clarified crude lysate (Lane 2, soluble protein) and crude lysate pellet (Lane 3, insoluble).

In both of the experimental conditions, lane 1 of both of Fig. 3-1A and Fig. 3-1B shows an obvious band at 23 kDa, showing that HCV-HBc was expressed successfully by *E. coli* using IPTG inducible expression system. However, the folding state was different. Most of the target protein from exp. no. 28 formed inclusion bodies (Lane 3, Fig. 3-1A) comparing with the exp. no. 21 where most of the target protein was expressed in the soluble form (Lane 2, Fig. 3-1B). This phenomenon suggests that the expression protein folding state is significantly influenced by the process factors, particularly cell induction conditions.

The compositions of clarified crude lysate from experiment no. 21 was further analysed by HPSEC using purified HBc-VLP as reference with the retention time of 23.8 min. Fig. 3-2 demonstrates the chromatogram of clarified crude lysate when resolved by HPSEC. The soluble HBc was separated at retention time of 20 min (peak 20), 23.8 min (peak 23.8) and 25.5 min (peak 25.5).

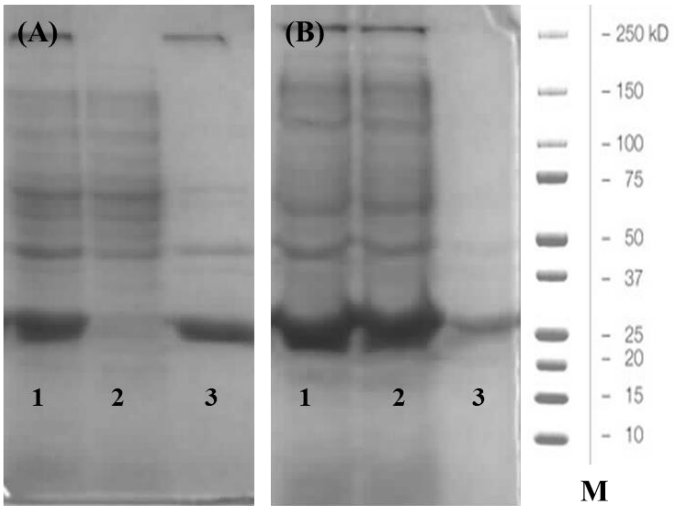

Fig. 3-1: Expression of HCV-HBc in different conditions. (A) Experiment no. 28 and (B) Experiment no. 21. Both (A) and (B) Lane 1: crude lysate, Lane 2: clarified of crude lysate and pellet of crude lysate Lane 3: inclusion bodies. M: marker.

Two (2) peak fractions were analysed by SDS-PAGE and silver staining to confirm the size of monomer, which is 23 kDa (Fig. 3-2B). Then TEM was used to confirm the spherical structure of HBc-VLP. From TEM images the soluble HBc-VLP aggregates were seen in peak 20 (Fig. 3-2C), a mixture of T=4 and T=3 of fully assembled HCV-HBc VLP were seen in peak 23.8 (Fig. 3-2D) and T=3 of HCV-HBc VLP were seen in peak 25.5 (Fig. 3-2E). It is noticed that there was only a few of particles in shoulder peak 25.5, suggesting this is a tailing effects from peak 23.8. Moreover, two (2) peaks at retention time of 33 and 38 min were mostly impurities from host cells and no HCV-HBc protein monomer were detected (data not shown).

The percentage of soluble aggregates detected by HPSEC was about 0.5 to 3 % of soluble HCV-HBc expression. These estimations were achieved from peak integration using LabSolution©.

Noticeably, the complicated composition of clarified crude lysate makes the chromatogram (Fig. 3-2A) become poorly resolved. The VLP percentage is not a precise value due to peak overlapping. Therefore, it is unable to directly measure the fully assembled HCV-HBc VLP from HPSEC. Because there was no HCV-HBc protein monomer were detected in other peaks from chromatogram, it is assumed that all the soluble monomer can self-assemble into soluble VLPs or soluble aggregates. To consider this, the yield of HCV-HBc VLP was determined by SDS-PAGE densitometry minus

percentage of soluble aggregates from HPSEC. This method is more accurate than directly using the peak area ratio of VLP from HPSEC chromatogram in the further study.

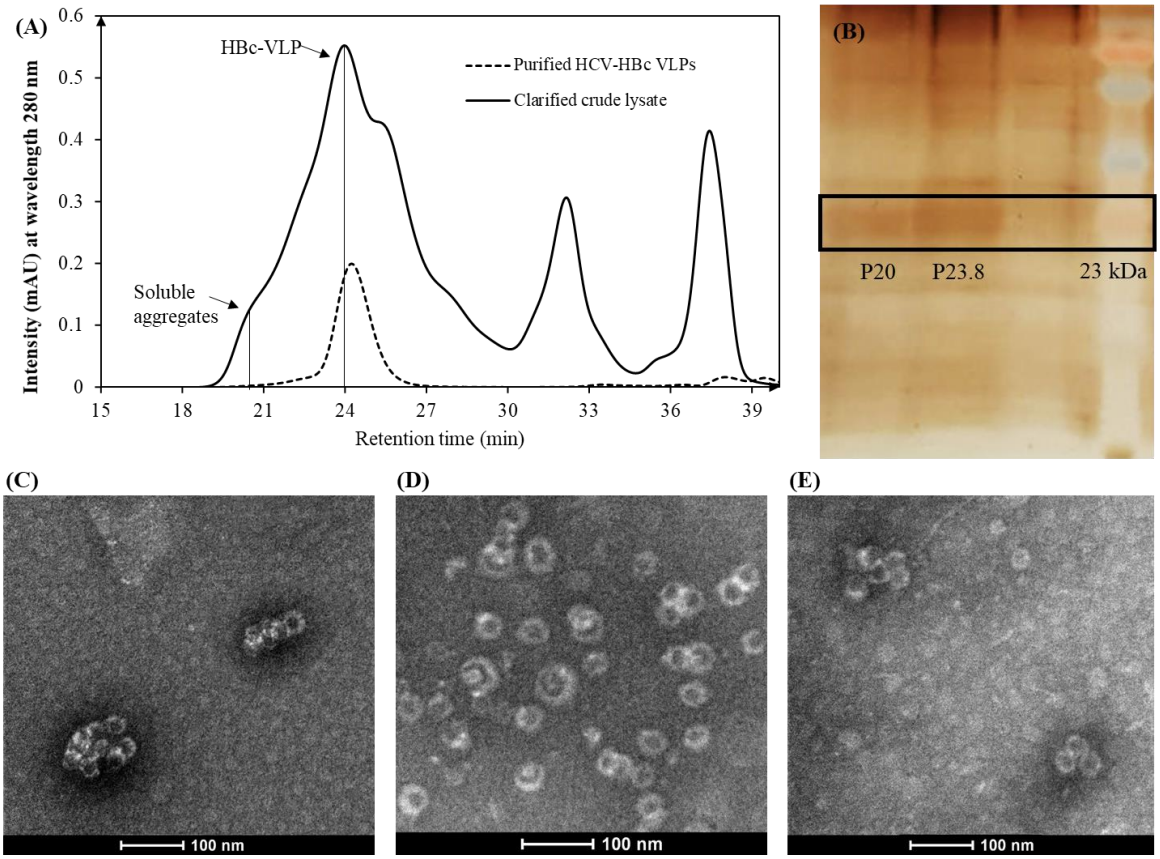

Fig. 3-2: Characteristic of HCV-HBc VLP using HPSEC; (A) HPSEC chromatogram of clarified crude lysate, resolving in TSK-G5000 GW_{XL} at 0.3 mL min^{-1}; (B) Peak fractions of minute 20, and 23.8 were analysed by SDS-PAGE with silver staining, (C) TEM image aggregated soluble VLPs in Peak 20; (D) TEM image of a mixture of VLP in Peak 23.8; (E) TEM image of a few VLP in shoulder peak 25.5.

3.3.2 Model analysis and diagnostics for soluble HCV-HBc VLP expression

Table 3-2 lists the process factor, their two levels and responses in 32 different expression conditions. The response was set as the soluble yield of HCV-HBc VLP per unit biomass, also known as cellular yield (mg g^{-1} DCW) The biomass yield range of soluble expression of HCV-HBc VLP is from 0.7 to 1.2 g L^{-1} DCW, the volumetric yield range of soluble HCV-HBc VLP is 33.4 to 91.1 mg L^{-1}, and the cellular yield range of soluble HCV-HBc VLP is 32 to 90 mg g^{-1} DCW.

Table 3-2: Fractional factorial design data of HCV-HBc VLP soluble production.

	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**	**Factor 6**	**Response 1**	**Response 2**	**Response 3**
Exp. no.	A: Temp.	B: Rotation speed	C: pH	D: IPTG conc.	E: Cell density	F: Mg^{2+}	Biomass DCW	Volumetric yield	Cellular yield
	°C	rpm		mM	OD_{600}	mM	g L^{-1}	mg L^{-1}	mg g^{-1}
1	30	110	7.2	0.2	0.8	10	0.8	40.2	48.2
2	30	110	7.2	0.8	0.8	0	0.8	42.3	52.0
3	40	250	6.2	0.2	0.8	0	1.0	33.4	33.9
4	30	110	6.2	0.2	2.0	10	1.2	68.4	41.5
5	40	250	7.2	0.2	2.0	0	1.2	56.1	49.6
6	40	250	7.2	0.8	0.8	0	1.0	38.4	37.8
7	40	250	7.2	0.2	0.8	10	1.0	42.7	42.1
8	30	250	7.2	0.8	0.8	10	1.0	91.1	87.5
9	30	250	7.2	0.8	2.0	0	1.1	78.1	54.4
10	40	110	7.2	0.8	0.8	10	0.8	43.9	58.5
11	30	110	7.2	0.2	2.0	0	1.2	56.3	38.7
12	40	250	6.2	0.8	2.0	0	1.1	76.7	56.9
13	40	250	6.2	0.8	0.8	10	1.0	49.8	50.0
14	30	250	6.2	0.8	2.0	10	1.1	72.2	52.8
15	30	250	6.2	0.2	0.8	10	1.0	77.1	76.7
16	40	110	6.2	0.2	2.0	0	1.1	34.4	27.1
17	30	250	7.2	0.2	0.8	0	1.0	74.4	72.3
18	40	110	6.2	0.2	0.8	10	0.7	41.6	55.8
19	30	110	7.2	0.8	2.0	10	1.1	55.7	42.7
20	40	250	6.2	0.2	2.0	10	1.2	63.1	41.6
21	30	250	6.2	0.8	0.8	0	0.9	84.4	89.7
22	40	110	6.2	0.8	2.0	10	1.1	41.2	38.8
23	40	110	7.2	0.8	2.0	0	1.1	34.0	32.4
24	30	110	6.2	0.8	0.8	10	0.8	34.5	41.8
25	40	110	7.2	0.2	0.8	0	0.7	42.7	60.5
26	30	110	6.2	0.8	2.0	0	1.1	56.0	43.6
27	30	250	6.2	0.2	2.0	0	1.1	76.4	49.2
28	30	250	7.2	0.2	2.0	10	1.1	77.5	53.9
29	40	250	7.2	0.8	2.0	10	1.1	75.9	53.6
30	40	110	7.2	0.2	2.0	10	1.0	42.8	42.8
31	30	110	6.2	0.2	0.8	0	0.9	47.0	54.2
32	40	110	6.2	0.8	0.8	0	0.7	40.8	55.9

Fig. 3-3A shows the half-normal plot, obtained from Design-Expert Software. Significant factors and interactions were identified, showing that **A** (post-induction temperature), **B** (post-induction rotation speed of the shaker) , **D** (IPTG concentration), **E** (optical cell density at induction OD_{600}), **AB** (post-induction of temperature - rotation speed of the shaker), **AE** (post-induction of temperature – cell density at induction), **ABE** (post-induction of temperature - rotation speed of the shaker - cell density at induction), and **ACD** (post-induction temperature-pH-cell density at induction) are significant model terms as their effects were shown as outliers from fiftieth percentile line (red line). The terms, C (pH), AC (post-induction temperature – pH), AD (post-induction temperature – IPTG), and CD (pH-IPTG) were added to maintain the hierarchy. Fig. 3-3B presents the normal probability plot of the external studentised residuals to support the reliability of the model, demonstrating the errors are normally distributed and not significant.

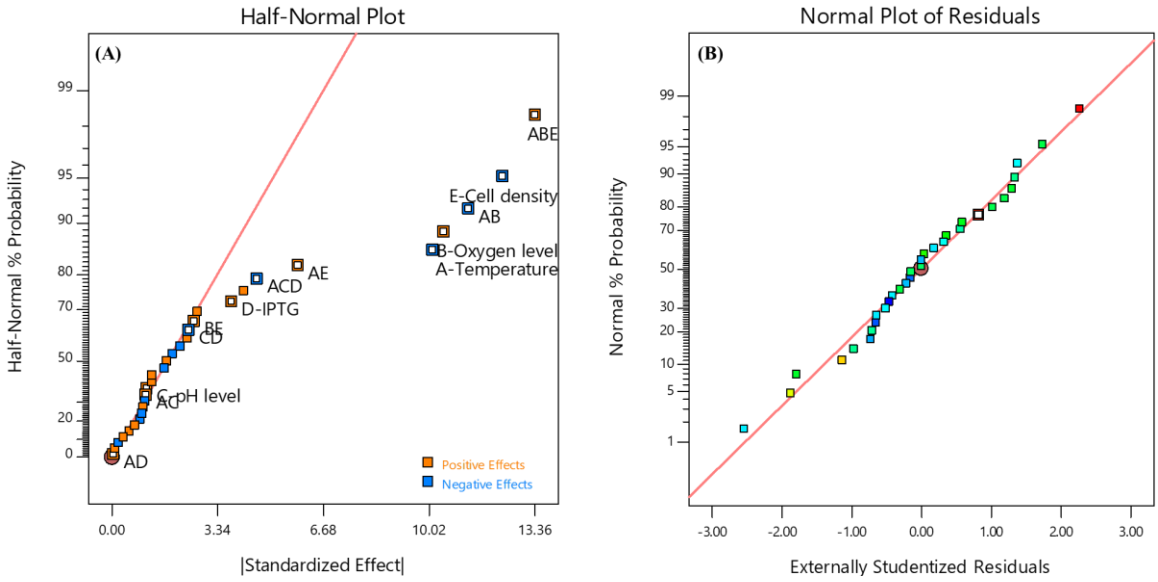

Fig. 3-3: Design analysis and model diagnosis of HCV-HBc VLP; (A) Half-normal plot of standardized effect and (B) the normal probability plot of externally studentised residual for the model.

Table 3-3 presents the ANOVA results for cellular yield of soluble HCV-HBc VLP production. To examine the reliability of the model, ANOVA was computed. The main effects and interaction effects of each design factor having P value less than 0.05 are considered as potentially significant, and design factor that has value larger than 0.1 is not significant. The F value of 28.93 shown the model is significant, with $R^2 = 0.94$, and importantly reflects the ability of the model to predict experimental outcome in the

investigated range. Also, R^2 = 0.94 implies that 94 % of the variations in the soluble production of HCV-HBc can be explained by the model. Moreover, adjusted R^2 = 0.89 agrees reasonably with predicted R^2 = 0.80 and Adequate Precision = 17.51. This finding indicates the model is applicable for navigating the design space and prediction of soluble expression of HCV-HBc VLP.

Table 3-3: ANOVA results of FFD for soluble cellular yield of HCV-HBc VLP.

Source	**Sum of squares**	**df**	**Mean square**	**F-value**	***p*-value**	
Model	6028.07	13	463.70	20.05	< 0.0001	Significant
A-Temperature	819.11	1	819.11	35.41	< 0.0001	Significant
B-Rotation speed	876.76	1	876.76	37.91	< 0.0001	Significant
C-pH level	9.57	1	9.57	0.4138	0.5282	Not significant
D-IPTG	113.63	1	113.63	4.91	0.0398	Significant
E-Cell density	1216.48	1	1216.48	52.59	< 0.0001	Significant
AB	1013.63	1	1013.63	43.82	< 0.0001	Significant
AC	9.14	1	9.14	0.3951	0.5375	Not significant
AD	0.0153	1	0.0153	0.0007	0.9798	Not significant
AE	275.54	1	275.54	11.91	0.0028	Significant
CD	46.80	1	46.80	2.02	0.1720	Not significant
ABE	1427.12	1	1427.12	61.70	< 0.0001	Significant
ACD	166.99	1	166.99	7.22	0.0151	Significant
Residual	416.33	18	23.13			
Cor Total	6444.40	31				

R^2 = 0.94; adjusted R^2 = 0.89; predicted R^2 = 0.80; Adequate precision = 17.51, df: degree of freedom

Equation (3-2) predicts the yield of HCV-HBc VLP (mg g^{-1}) by fitting the eperimental data to first order empirical model (Eq. 2-1), where β is estimated from the experiment data, and X is the design factor:

$$\text{Yield}_{\text{HCV-HBc VLP}}\ (\text{mg g}^{-1}) = +51.14 - 5.06*\mathbf{A} + 5.23*\mathbf{B} + 0.5469*\mathbf{C} + 1.88*\mathbf{D} - 6.17*\mathbf{E} - 5.63*\mathbf{AB} + 0.5344*\mathbf{AC} + 0.0219*\mathbf{AD} + 2.93*\mathbf{AE} + 1.29*\mathbf{BE} - 1.21*\mathbf{CD} + 6.68*\mathbf{ABE} - 2.28\ \mathbf{ACD} \quad (\textbf{Eq. 3-2})$$

Regarding the HBc VLP, the expression was conducted followed the condition of experiment no. 21 in Table 3-2. The volumetric yield and cellular yield of HBc VLP were estimated as 184.3 mg L^{-1} and 137.5 mg g^{-1}.

3.3.3 Influence of process factors on HCV-HBc soluble yield

The greatest expression levels were seen at 30 ^{o}C, 250 rpm, 0.8 mM of IPTG, induced at OD_{600} of 0.8, regardless of pH level and the presence of Mg^{2+} (cellular yield 87.5 mg g^{-1} in experiments no.8 and 89.1 mg g^{-1} in experiment no.21). In general, low temperature (30 ^{o}C) and high shaker speed (250 rpm) and induction at early exponential phase were beneficial for the soluble production HCV-HBc.

Fig. 3-4 and Fig. 3-5 present the percentage of soluble HCV-HBc over total HCV-HBc expressed and volumetric yield of soluble HCV-HBc VLP in one (1) litre of culture media. Overall, the solubility of HCV-HBc VLP was 25 to 70 % in case of early induction (OD_{600} of 0.8) (Fig. 3-4A) and 20 to 35 % when culture induced at higher cell density (OD_{600} of 2.0) (Fig. 3-4B). Significantly under aerobic conditions low expression temperature and early induction, Fig. 3-4A presents the solubility of HCV-HBC VLP vary from 62 to 75 % and volumetric yield increased by 2.5-fold in comparison to the induction at higher cell density as is presented in Fig. 3-5.

Fig. 3-6 demonstrates the effect of interactions between AB, CD and AE on soluble production of HCV-HBc VLP. At early induction and high rotation speed (250 rpm), the influence of temperature on soluble yield was significant, the yield when protein expressed at 30 ^{o}C was 80 mg g^{-1}, 2 - fold in comparison to protein expressed at 40 ^{o}C (40 mg g^{-1}).

Additionally, under early induction with low rotation speed (120 rpm), the effect of temperature did not influence the expression much, the yield was recorded at 50 to 60 mg g^{-1} (Fig. 3-6A). Further regarding the expression at mid-exponential phase (OD_{600} of 2.0) the temperature effect was similar in both cases of rotation speed of shaker (Fig. 3-6B). Although expression at 40 ^{o}C decreased the soluble yield. Noticeably, high post-induction rotation speed of shaker gave higher soluble yield.

Fig. 3-6C and Fig. 3-6D present the effects of interaction between IPTG concentration and pH level at induction at 30 ^{O}C and 40 ^{O}. Both IPTG concentration and pH level did not significantly influence soluble expression when protein was expressed at 30 ^{o}C (55 mg g^{-1}). However when the expression was conducted at 40 ^{o}C and pH of 6.2 the addition of 0.8 mM IPTG gave greater yield than addition of 0.2 mM IPTG (the difference was approximately 10 mg g^{-1}).

Fig. 3-6E and Fig. 3-6F present the effects of interaction between cell density at induction and post-induction temperature when induced with IPTG of 0.8 mM and 0.2 mM. The yield had a meaningfully small increase when cell culture was induced with IPTG of

0.8 mM. The result highlights the importance of post-induction temperatures, post-induction rotation speed of shaker, IPTG concentration and cell density at induction time. The soluble production can be increased if the cell culture was induced at early exponential phase (OD_{600} of 0.8), low expression temperature (30 ^{o}C), IPTG of 0.8 mM and high rotation speed of the shaker (250 rpm).

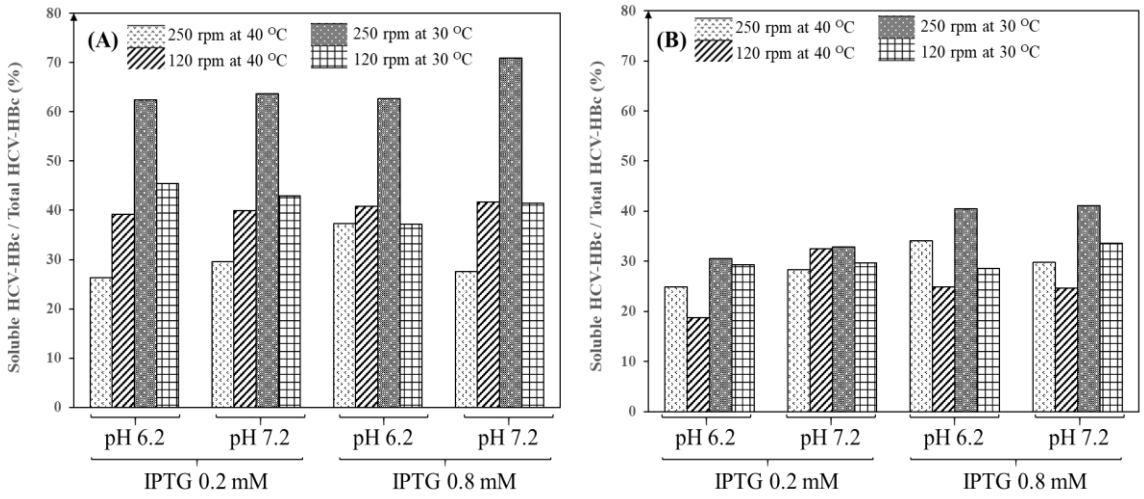

Fig. 3-4: Percentage of soluble HCV-HBc over the total HCV-HBc protein expressed. (A) Culture was induced at OD_{600} of 0.8 and (B) culture was induced at OD_{600} of 2.0.

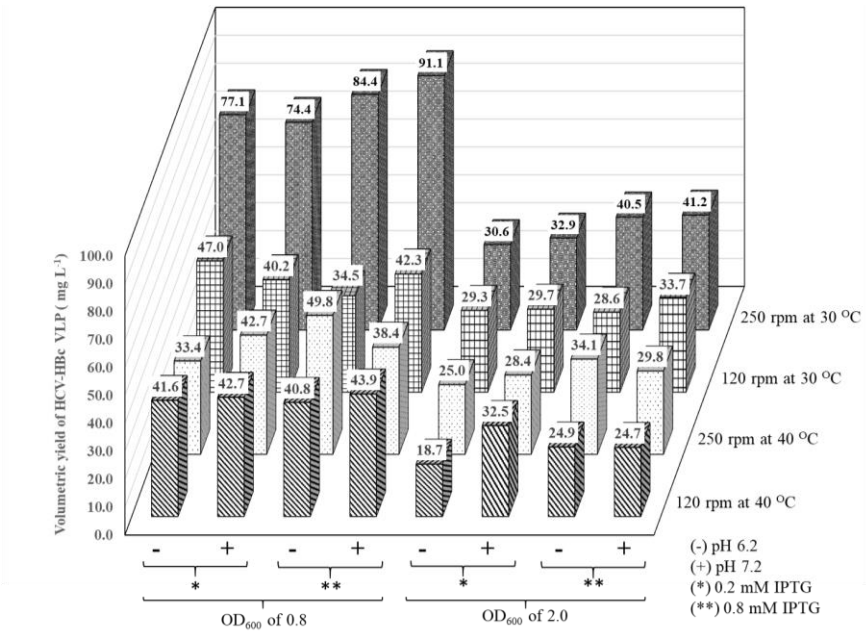

Fig. 3-5: Volumetric yield of soluble HCV-HBc in 32 different conditions shown in Table 3-2.

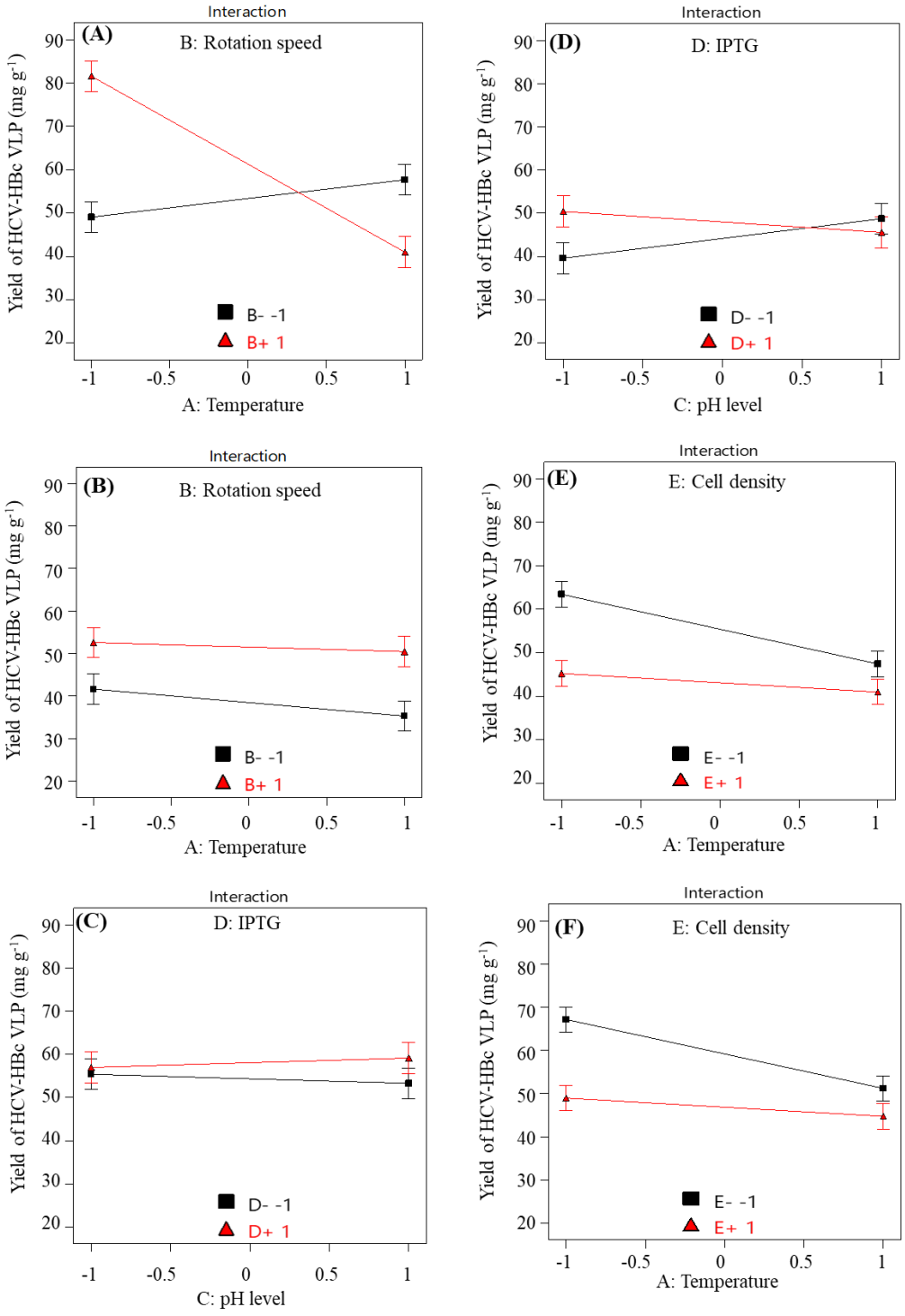

Fig. 3-6: Interactions between AB, CD, AE on the soluble expression on HCV-HBc VLP; (A) and (B) is the effects of interaction between A and B at induction cell density OD_{600} of 0.8 and 2.0; (C) and (D) is the effects of C and B at induction temperature of 30 ^{o}C and 40 ^{o}C; (E) and (F) is the effects of interaction between A and E level with IPTG of 0.2 mM and 0.8 mM; (+1) means high level and (-1) means low level in coded value.

3.4 Discussion

HBc is a promising carrier platform because of its flexibility to carry antigen epitope. The size of HBc-based VLP even without epitopes, is highly significant to the capacity of *E. coli*, especially when assembled intracellularly. As an effective promoter for expression of recombinant protein in *E. coli*, T7 promoter can boost the expression level of product to 50 % of total cell protein (Studier and Moffatt, 1986). However, due to overexpression from T7, cell will be overflowed with protein that impacts protein folding velocity and promotes formation of inclusion bodies (Kesik-Brodacka et al., 2012). To overcome this drawback and continue using the T7 promoter, adjusting fermentation parameters is essential to achieve higher soluble production.

In the agreement with the literature review, HBc capsid has two isomorphs with T=3 (180 subunits) and T=4 (240 subunits), where the theoretical hydrodynamic radius of T=4 is 17 nm (Crowther et al., 1994). From the TEM images of Fig. 3-2D, the HCV-HBc VLP had majority of T=4 capsid with diameter of 34 nm. The reasons of switching the assembly structure between particle of 180 subunits and particles of 240 subunits remained unclear. It is mentioned that the assembly of *E. coli* derived HBc VLP is similar to the assembly of the core of the wild-type virus, and the different in size is not caused by the interaction between the protein and the contents inside the capsid (Crowther et al., 1994). The study of Crowther et al., 1994 also hypothesised that the size differences wild-type virus caused by the packaging of the complete RNA pre-genome and the viral polymerase. The limitation of this characterisation experiment is lack of size distribution of HCV-HBc VLP and exact measurement of each capsid types in terms of hydrodynamic diameter and molecular weight. To improve the size distribution of VLP, using gel filtration as a polishing step will help to achieve a uniform HBc VLP and further analyses using dynamic light scattering and multi angle laser light scattering will help to provide a complete characterisation of the protein.

Fractional factorial experimental design approach is a practical tool to 1) gain technical insight to boost the soluble production of HCV-HBc and 2) develop empirical equation for the prediction. It is feasible to manipulate the process factors to improve the solubility of expression in a systematic way as proved in this study. The statistical analysis such as ANOVA and residual plot are useful to validate the reliability of the experimental model. The order of significant process factors was as following: 1) cell density at induction,

2) post-induction rotation speed of shaker, 3) post-induction temperature and 4) their interaction.

Although the statistical approach successfully extracted significant information from the experiments, because there were only two (2) levels, the significance of factors identified in this design are not sufficiently extensive to explain the mechanisms involved (Uhoraningoga et al., 2018). Following the identification of the significant parameters, further experiments need to be conducted with fewer factors and more levels. This resulting response surface model will better aid optimisation. Some potential mechanisms why these process factors are important to soluble expression of HCV-HBc are proposed.

The first major factor is the cell density at induction. Before induction, cells consumed nutrients to produce more host cell protein (HCP) to support the cell proliferation, in which dissolved oxygen rate and temperature mainly controlled the growth rate. However, when IPTG was added, cells shifted from expression of HCP to expression of foreign recombinant protein, which later caused metabolic burden to the host cells, either nutrient or energy. Fig. 3-4A and Fig. 3-4B show that the solubility of HCV-HBc VLP expression was dependent on the cell density at induction, the lower cell density at induction was, the greater solubility of the expression was. Also, the volumetric yield enhanced from 41.2 to 91.1 mg L^{-1} (culture media) when shifting the induction to low cell density. Therefore, cell density at induction and inducer concentration are linked together in terms of dose dependent.

The rotation speed of the shaker is a dominating factor in the expression of soluble HCV-HBc VLP. The cell population at late induction (OD_{600} at 2.0) was greater than those at early induction (OD_{600} at 0.8) so the needs of oxygen was also greater. Low rotation speed (120 rpm) did not provide enough oxygen to the culture. Limited oxygen level had a negative impact, suggesting that oxygen limitation induced stress on the culture. The results and explanation highlight the importance of oxygen level in bacterial growth and soluble protein expression, it also indicates the limitation of shake-flask cultivation (Losen et al., 2004), such as low oxygen transfer rate and poor heat transfer. It is found that *E. coli* shifts its own metabolic pathway, switching between aerobic growth and anaerobic growth, depending on the level of oxygen in the culture. This behaviour consequently affects the tricarboxylic acid cycle (TCA) and expresses multiple genes to adapt to the availability of oxygen (von Wulffen et al., 2016). In aerobic conditions *E. coli* uses oxygen as the terminal electron acceptor, oxidizing pyruvate *via* the TCA cycle to produce adenosine triphosphate (ATP). ATP not only provides energy for protein translation (Konz et al., 1998), it is also

reported to be involved in virus-like particle assembly in case of Human immunodeficiency virus type 1 (HIV-1) VLP (Tritel and Resh, 2001). It is used by heat shock protein, Hsp70, in the presence of co-chaperones to stabilize native protein folding (Xu, 2018). Because oxygen is limited *E. coli* uses inorganic compounds as electron receptors instead of using oxygen. This altered electron transport chain leads to a decrease in ATP production in cells (Shalel-Levanon et al., 2005).

In this research even though induction at 30 ^{o}C favoured the expression of the soluble virus-like particle, the biomass achieved was less than those induction at 37 ^{o}C. Higher temperatures can achieve higher biomass but potentially increasing the inclusion body formation. It is reported aggregation reaction is temperature-dependent because it is related to hydrophobic interaction (van Dijk et al., 2015). The solubility of HCV-HBc expression enhanced approximately 2.5-fold when culture was cultivated at 30 ^{o}C instead of 40 ^{o}C. In particular, when induced at pH 6.2 and 0.2 mM of IPTG, volumetric yield increased from 33.4 to 77.1 mg L^{-1}, and volumetric yield increased from 38.4 to 91.1 mg L^{-1} when induced at pH 7.2 and 0.8 mM of IPTG (Fig. 3-4A). These observations align with the argument, highlighting the importance of temperature during expression.

In conventional induction, 1 mM of IPTG is required for induction. In this study 0.2 mM of IPTG was sufficient to express HCV-HBc VLP. It agrees with Muhlmann et al. (2017) that low IPTG concentrations (0.05 – 0.1 mM) are beneficial for the soluble expression. The pH level was adjusted with K_2HPO_4 2M and KH_2PO_4 0.5M. Media supplemented with $MgCl_2$ is reported to prevent ribosome degradation (Nierhaus, 2014). However, DoE did not show that Mg^{2+} concentration was a significant factor. To adjust to desired pH, the amount of K_2HPO_4 used at the time induction was estimated at about 80 - 100 mM, and KH_2PO_4 was about 5 - 10 mM. Fig 3-6D indicates there was an interaction between IPTG concentration and K^+ ion on the soluble expression of HCV-HBc in shake-flask fermentation but the effect of this interaction was not significant as is listed in Table 3-3.

Because the theoretical structure of HCV epitope in use is α-helices, protein synthesis and protein folding are more complicated. This necessitates the slow folding rate to ensure the native folding. The requirement of a high rotation speed (250 rpm) and low temperature (30 ^{o}C) following induction for soluble expression of HBc-based VLP is partially explained by these arguments. Protein expression rate, protein folding, and biomass are interconnected by post-induction process factors.

In comparison to HBc VLP, both volumetric yield and cellular yield of HBc-VLP were greater than HCV-HBc VLP under same optimal condition (experiment no. 21 in Table 3-2), indicating that the insertion of epitope has significant influence on the soluble expression of HBc in *E. coli*. However, how a particular structure, i.e., α-helix, influences on the formation of HBc VLP remains unknown. It is reported in the study of Vogel et al., 2005 that the interaction between the insert epitope and core protein is potentially lead to misfolding of the protein monomer. Although the study focused on the insertion at MIR, it had suggested some possible reasons behind this phenomenon. A complicated and bulky, such as HCV epitope, may prevent the central helices of the HBc protein fold into juxtaposition. Moreover, Vogel et al., 2005 mentioned that the constraints from 1) the spatial proximity of the insertion site, and 2) the display of high-density of foreign epitope on the capsid surface, leading to the change in capsid stability, are also significant factors. Therefore, epitope that has unstable structure and bulky will significantly affect the capsid assembly, eventually decreasing the soluble expression of HBc VLP.

It is clear that statistical approach extrapolates more information from experimental responses, compared with the one-variable-at-time approach. A boost of 4.8-fold in volumetric soluble yield is the evidence of the applicability of statistical approach to recombinant protein expression.

This novel approach should be applied for other recombinant protein bioprocessing because of its predication on underlying interaction between process factors. An advantage of this approach is that the design of experiment is formed based on the aims of the study, allowing researchers to know the number of experiments they need to observe and detailed timeframe.

3.5 Chapter summary and conclusions

Based on experimental results and analyses in this chapter, the following important factors emerge:

1. A HCV vaccine candidate has been produced in *E. coli*. It was found that HCV-HBc VLP tends to be expressed as inclusion bodies. The collected peak fraction in clarified crude lysate from chromatography at a retention time of 23.8 min has been examined by TEM to reveal that HCV-HBc protein folds into correct VLP spherical structure.

2. In comparison to HBc VLP without carrying foreign epitope, HCV-HBc VLP expressed less soluble yield. It is hypothesised that HCV epitope has influence the HBc VLP formation, potentially increasing the protein misfolding.

3. FFD is an effective statistical tool to investigate process factors that impact the expression of recombinant protein. This approach is novel because it has been used for other recombinant protein but not for HBc platform. Analyses show that the greatest yield of HCV-HBc has been recorded at 89.7 mg g^{-1} DCW. The influence of process factors could be ranked as: 1) cell density at induction, 2) post-induction rotation speed of shaker, and 3) post-induction temperature.

4. These new findings provide the design space to improve soluble production of HBc VLP in microbial system.

Despite its advance, the novel statistical approach developed in this chapter has some drawbacks, namely, the expression and analysis heavily rely on equipment performance, especially limitations in shake-flask cultivation. Additionally, design of experiment does not explain the mechanisms involved.

In the next chapter, Chapter 4, a novel statistical approach for boosting the soluble production of HBc VLP carrying simple structure epitope EBNA1 is developed. After recognizing the limitation in the experiment design and analytical methods in Chapter 3, the design of experiment for EBNA1-HBc starts with fractional factorial design for screening of factors, then optimising using response surface methodology in which, instead of evaluating from densitometry of SDS-PAGE, native agarose gel electrophoresis is an alternative approach to estimate expression.

CHAPTER 4

ENHANCEMENT OF SOLUBLE EXPRESSION OF *ESCHERICHIA COLI*-DERIVED VIRUS-LIKE PARTICLES CARRYING NON-STRUCTURAL EPITOPE BY RESPONSE SURFACE METHODOLOGY

4.1 Introduction

A critical review of soluble intracellular *E. coli*-derived recombinant protein and application of statistical approach have been presented in Chapter 2. In addition, fractional factorial design has been applied to boost the soluble expression of HCV-HBc as is presented in Chapter 3, to show that it is a useful tool to optimise process factors. Moreover, it was found that HCV-HBc is easy to aggregate in both soluble and insoluble form.

In this chapter, EBNA1-HBc is developed as vaccine candidate against Epstein-Barr virus infection and statistical approach is applied for the first time to increase the soluble yield of HBc VLP carrying non-structural epitope EBNA1. EBNA1 has a role in suppressing reactivation in latent infection and involve in lytic gene expression and DNA amplification in lytic infection (Sivachandran et al., 2012).

To boost the soluble production of target protein and overcome the limitation in experiment design and analytical method in Chapter 3, alternative analytical methods and design of experiment are developed. After recognising that two (level) for each factor in fractional factorial design (FFD) does not include the curvature for continuous factor, the statistical approach used in this chapter includes FFD and response surface model (RSM). First, the process factors are screened using fractional factorial design (FFD). Three (3) of the process factors that have significant impacts on the soluble expression of EBNA1-HBc are identified and then their levels are optimised using RSM. The optimal combination is achieved from point prediction feature from RSM cultivations at the optimal condition are conducted to validate the optimal point.

In addition, analytical method was further development, taking the advantages from HBc-VLP size, native agarose gel electrophoresis (NAGE) is an effective method to separate HBc-capsid out of cell lysate (Yoon et al., 2012). The fully assembled HBc VLP and soluble aggregates can be quantified using gel densitometry from NAGE. It is also noticed that thermal stability of chimeric HBc-VLP enables the possibility to use heat precipitation to remove host cell protein (HCP) as initial step of protein separation (Ng et al., 2006; Li et al., 2018). This pre-treatment step provides high-quality sample molecular characterisation by for High-performance Size-exclusion chromatography (HPSEC) (Yang et al., 2015).

It is hoped that insight gained from this chapter might demonstrate the advance of statistical approach in sequential order from FFD to RSM to boost the soluble production of EBNA1-HBc in *E. coli*. These findings could thereby benefit the high-throughput

fermentation process factor screening and later on to apply for the scale-up in recombinant protein bioprocessing.

4.2 Material and methods

4.2.1 Expression strain

The epitope used this chapter is different from the epitope used in Chapter 3. The epitope of EBNA1 (EBNA1: HPVGEADYFEY) was inserted at N-terminus of HBc (HBc: 183 amino acid) to form EBNA1-HBc. The plasmid pET-30a (+) (Invitrogen ®, USA) was used to form pET-30a (+)/HCV-HBc, carrying the gene of EBNA1-HBc protein under control of T7 promoter with *lac* operon.

Via heat-shock transformation the plasmid was transformed into *E. coli* by using One Shot™ BL21 (DE3) chemically competent *E. coli* (Invitrogen©, USA). Cells were cultured on Luria Bertani (LB) agar plate. Cells were stored at minus 20 ^{o}C as the working cell.

All reagents were analytical grade (AR) unless otherwise stated.

4.2.2 Fractional factorial design for factor screening

Preliminary experimental design was used to point out the significant variables that affected the soluble production of EBNA1-HBc VLP. To obtain main effects of process factor and two-factor interaction, a resolution V of fractional factorial design (FFD) was prepared by Design-Expert software (version 11, Stat-Ease©, Minnesota, USA) (Wu and Hamada, 2009). The design included 32 expression condition, made up from six (6) design factors with two (2) designated levels as low (-1) and high (+1) level.

Table 4-1 shows a fractional factorial design that was used to boost the soluble expression of EBNA1-HBc. The six (6) factors included (**A**) post-induction temperature (30 ^{o}C and 40 ^{o}C), (**B**) dissolved oxygen rate in terms of post-induction rotation speed of the shaker (110 and 250 rpm), (**C**) pH of the culture at induction point (6.2 and 7.2), (**D**) IPTG concentration (0.05 and 0.5 mM) (Invitrogen©, USA), (**E**) optical cell density at induction point (OD_{600} of 0.8 and 2.0), and (**F**) the presence of Mg^{2+} (0 and 10 mM). The pH of the culture was adjusted with K_2HPO_4 2M (Chem-supply, Australia) and KH_2PO_4 0.5M (Chem-supply, Australia).

The main response was set as the cellular yield of soluble VLP per dry cell weight (mg g^{-1}), in relation to the biomass per litre of culture media (g L^{-1}) and the volumetric yield

of target protein per litre of culture media (mg L^{-1}). Data analysis was performed by Design-Expert. The response was kept as normal (no transformation required by the Box-Cox method) (Box and Cox, 1964). The significance of each factor was determined by the half-normal probability plot and computed with the analysis of variance (ANOVA) as featured in Design-Expert.

4.2.3 Response surface model for factor optimisation

Response surface model (RSM) was used to improve the soluble production of EBNA1-HBc in *E. coli*. To investigate three independent factors with five levels, a 2^3 factorial central composite rotary design (CCD) was used. Table 4-2 presents the experimental matrix, which includes 20 experiments and six (6) of them are centre points. The experimental matrix was created by Design-Expert software.

Variables and their levels were chosen followed the screening experiment results. They are (**A***) post-induction dissolved oxygen rate in relation to rotation speed of the shaker (170, 190, 220, 250, and 270 rpm), (**B***) post-induction temperature (28, 30, 33, 36, and 38 ^{o}C), and (**C***) the optical density of cell population at induction of OD_{600} (0.6, 0.8, 1.1, 1.4, and 1.6). The response of the RSM matrix was set as the cellular yield, which is described as the amount of soluble HBc VLP per dry cell weight (mg g^{-1}).

The IPTG concentration is fixed at 0.5 mM for every experiment. Design-Expert software features point optimisation, which was employed to optimise the level of each factors for maximum response. The combination of optimised parameters was experimentally tested to validate the model accuracy in triplicate (n=3).

Table 4-1: Fractional factorial design matrix for factor screening of soluble expression of EBNA1-HBc VLP.

	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**	**Factor 6**
Exp. no.	A: Temperature	B: Rotation speed	C: pH	D: IPTG conc.	E: Cell density	F: Mg^{2+}
	(-1): 30 (+1): 40	(-1): 120 (+1): 250	(-1): 6.2 (+1): 7.2	(-1): 0.05 (+1): 0.5	(-1): 0.8 (+1): 2.0	(-1): 0 (+1): 10
	°C	rpm		mM	OD_{600}	mM
1	30	110	6.2	0.5	2.0	0
2	30	250	6.2	0.05	2.0	0
3	30	250	6.2	0.05	0.8	10
4	40	110	7.2	0.05	0.8	0
5	30	250	7.2	0.5	0.8	10
6	40	110	7.2	0.5	0.8	10
7	40	250	7.2	0.05	0.8	10
8	40	110	7.2	0.5	2.0	0
9	30	250	7.2	0.5	2.0	0
10	40	250	6.2	0.05	0.8	0
11	30	110	7.2	0.5	2.0	10
12	40	250	7.2	0.05	2.0	0
13	40	110	6.2	0.5	2.0	10
14	40	250	6.2	0.5	2.0	0
15	40	110	6.2	0.5	0.8	0
16	40	250	7.2	0.5	2.0	10
17	40	110	6.2	0.05	2.0	0
18	40	250	6.2	0.5	0.8	10
19	30	250	7.2	0.05	0.8	0
20	30	250	6.2	0.5	2.0	10
21	40	110	7.2	0.05	2.0	10
22	30	110	7.2	0.5	0.8	0
23	30	110	6.2	0.05	2.0	10
24	30	110	6.2	0.5	0.8	10
25	30	110	6.2	0.05	0.8	0
26	30	110	7.2	0.05	2.0	0
27	40	250	6.2	0.05	2.0	10
28	30	250	7.2	0.05	2.0	10
29	30	250	6.2	0.5	0.8	0
30	40	250	7.2	0.5	0.8	0
31	40	110	6.2	0.05	0.8	10
32	30	110	7.2	0.05	0.8	10

Table 4-2: Response surface model – central composite design for factor optimisation of soluble expression of EBNA1-HBc VLP.

	Factor 1	**Factor 2**	**Factor 3**
Exp. no.	A*: Rotation speed	B*: Temperature	C*: Cell density at induction
	(-1.68): 170 (-1): 190 (0): 220 (+1): 250 (+1.68): 270	(-1.68): 28 (-1): 30 (0): 33 (+1): 36 (+1.68): 38	(-1.68): 0.6 (-1): 0.8 (0): 1.1 (+1): 1.4 (+1.68): 1.6
	rpm	^{o}C	OD_{600}
1	-1	-1	1
2	0	0	0
3	1	-1	-1
4	-1	-1	-1
5	1	1	-1
6	0	0	-1.68
7	-1	1	-1
8	0	0	0
9	0	0	0
10	0	1.68	0
11	1	-1	1
12	0	0	0
13	0	0	0
14	0	0	1.68
15	-1.68	0	0
16	1	1	1
17	-1	1	1
18	1.68	0	0
19	0	-1.68	0
20	0	0	0

4.2.4 Expression of EBNA1-HBc VLP

The *E. coli* transformed with the recombinant plasmids were cultured in 50 mL of sterilized LB media (10 g L^{-1} tryptone (Oxoid, USA), 5 g L^{-1} yeast extract (Oxoid, USA), 10g L^{-1} NaCl (Chem-supply, Australia), supplemented with 50 µg mL^{-1} kanamycin sulfate (Invitrogen©, USA), at 37 ºC and 200 rpm (round per minute) for 16 h.

The overnight culture was added into 200 mL of fresh sterilized LB media volume with 50 µg mL^{-1} kanamycin sulfate, in the 500-mL flask, at a ratio of 1 v/v % (seeding/culture volume), then was cultivated at 37 ºC and 200 rpm. Optical cell density (OD_{600}) was measured by UV-1600PC Spectrometer (VWR International©, USA) at wavelength 600 nm.

In terms of factor screening, the expression of EBNA1-HBc was conducted following FFD matrix, as is shown in Table 4-1. Experiments were conducted in randomised sequence to avoid bias. The $MgCl_2$ (Chem-supply, Australia) was supplied to the culture at the beginning of the cultivation according to the FFD. When the cell density reached the designated value of OD_{600}, the fermentation process factors were adjusted to specific experiment number in Table 4-1. K_2HPO_4 2 M (Chem-supply, Australia) and KH_2PO_4 0.5 M (Chem-supply, Australia) were used to adjust the pH of the culture at induction. IPTG (Invitrogen©, USA) was added to the culture to express the EBNA1-HBc. Temperature and rotation speed of the shaker were also adjusted. The expression was continued for the next six (6) h.

In terms of factor optimisation, the expression of EBNA1-HBc was followed RSM matrix. Experiments were performed in randomised sequence to avoid bias. When the cell density reached the designated value of OD_{600}, the temperature and rotation speed of shaker were adjusted to specific experiment number in Table 4-2. IPTG of 0.5 mM was supplied to the culture to begin expression.

The final cell density (OD_{600}) was measured to determine the final biomass concentration. It is established that one (1) litre of culture media at OD_{600} of 1.0 equals to 0.4 g of dry cell weight (DCW). Cells were harvested by centrifugation (3700 rpm, 20 ºC, 30 min) and were washed with deionized water. The cell pellet was stored at minus 20 ºC for further analyses.

Cells were suspended in lysis buffer and were disruption as is described in section 3.2.3. The supernatant fraction was identified as clarified crude lysate.

Protein concentration was determined by Bradford assay (Bradford, 1976), using bovine serum albumin (Sigma-Aldrich, USA) as protein standard.

4.2.5 NAGE and SDS-PAGE

Native agarose gel electrophoresis was used to separate small and large capsid of HBc-based VLP (Yoon et al., 2012), which was used to evaluate the soluble expression level of EBNA1-HBc VLP for both preliminary and optimisation study.

Samples were mixed with NAGE-loading buffer at ratio 4:1 (v/v) and incubated at 25 ^{o}C for 20 min before running. A 1 w/v % agarose (Sigma-Aldrich, USA) was prepared in 1X TAE buffer (Thermo Fisher Scientific, USA). Agarose gel was run with Owl™ EasyCast™ B2 Mini Gel Electrophoresis Systems (Bio-rad©, USA). Wells were loaded with 45 to 50 µg of clarified crude lysate of each condition. Ferritin nano-particle (approximately 500 kDa) with designated mass from 15 to 55 µg, was used as the marker. Protein was resolved at 80 V for 6 h at 4 ^{o}C in 1X TAE buffer. The procedure for staining, de-staining, gel image capture and densitometry analysis was followed the procedure of SDS-PAGE.

SDS-PAGE (Polyacrylamide 12 w/v %) was used to determine the purity of EBNA1-HBc in protein sample. The procedure for sample preparation, electrophoresis, staining, de-staining, gel image capture and densitometry analysis are described in section 3.2.4. In terms of determination of VLP in HPSEC collected peak, 12 w/v % reducing SDS-PAGE with silver staining was conducted using SilverQuest™ Silver Staining Kit (Invitrogen©, USA).

The component of TAE buffer, NAGE-loading buffer, SDS- loading buffer, SDS-running buffer, staining buffer and staining buffer was listed in Appendix B.

4.2.6 Host cell protein removal by heat precipitation

To avoid the complicated nature of clarified crude lysate, removal of impurities, i.e. unstable thermal host cell protein, is essential. Similar to study of Li et al. (2018), the clarified crude lysate was heated on heat block for 30 min at 40, 50, 60, 65, 70, and 75, ^{o}C, respectively. The treated lysate was cool down to the 25 ^{o}C and centrifuged (12000 rpm, 10 ^{o}C, 30 min). The purity (%) and recovery yield (%) of EBNA1-HBc are determined as Equation (4-1) and Equation (4-2), respectively.

$$\text{Purity (\%)} = \frac{\text{The amount of EBNA1}-\text{HBC in sample}}{\text{The amount of total protein in sample}} \times 100\ \% \qquad \textbf{(Eq. 4-1)}$$

$$\text{Recovery yield (\%)} = \frac{\text{The amount of EBNA1}-\text{HBC in supernantant after heat treatment}}{\text{The amount of EBNA1}-\text{HBc in clarified crude lysate}} \times 100\ \% \qquad \textbf{(Eq. 4-2)}$$

4.2.7 High-performance size-exclusion chromatography

HPSEC has been used to quantify HBc-based VLP particles (Yang et al., 2015). HPSEC was performed on the Prominence HPLC system (Shimadzu, Japan), equipped with TSK-G5000 GW_{XL} (300 × 7.8 mm, I.D., pore size 100 nm), (Tosoh Bioscience, Germany) and column-guard, with UV monitoring at wavelength 260 nm and 280 nm.

Purified EBNA HBc-VLPs with known concentration was injected into HPSEC to develop a calibration curve. For each measurement, 50 µL of heat-treated clarified crude lysate was resolved at flowrate 0.5 mL min^{-1}, the mobile phase was phosphate buffer (50 mM, pH 7.4) (Chem-supply, Australia). Peak integration was analysed by LabSolution© (Shimadzu, Japan). Protein fractions were analysed by SDS-PAGE and silver-staining. HBc-based VLP was quantified by comparing the peak area between standard curves and results obtained from peak integration.

4.3 Results

4.3.1 Quantification of EBNA-HBc particles by NAGE

Conventionally, the expression level of target protein is determined by SDS-PAGE. Sodium dodecyl sulfate breaks not only the fully assembled capsid but also soluble aggregates into monomer. Native agarose gel electrophoresis can separate the VLP capsid and soluble aggregates out of the impurities in the clarified crude lysate based on the size and overall charge of the particles without compromising the structure.

Fig. 4-1 demonstrates the quantification of EBNA1-HBc VLP using NAGE. Fig. 4-1A presents the position of ferritin as reference and EBNA1-HBc capsid is fractionated by NAGE, according to the position of VLP standard. In this study, ferritin was used as internal marker (Lane 1, lane 2, and lane 3 is 53 µg, 35 µg and 17 µg of ferritin, respectively).

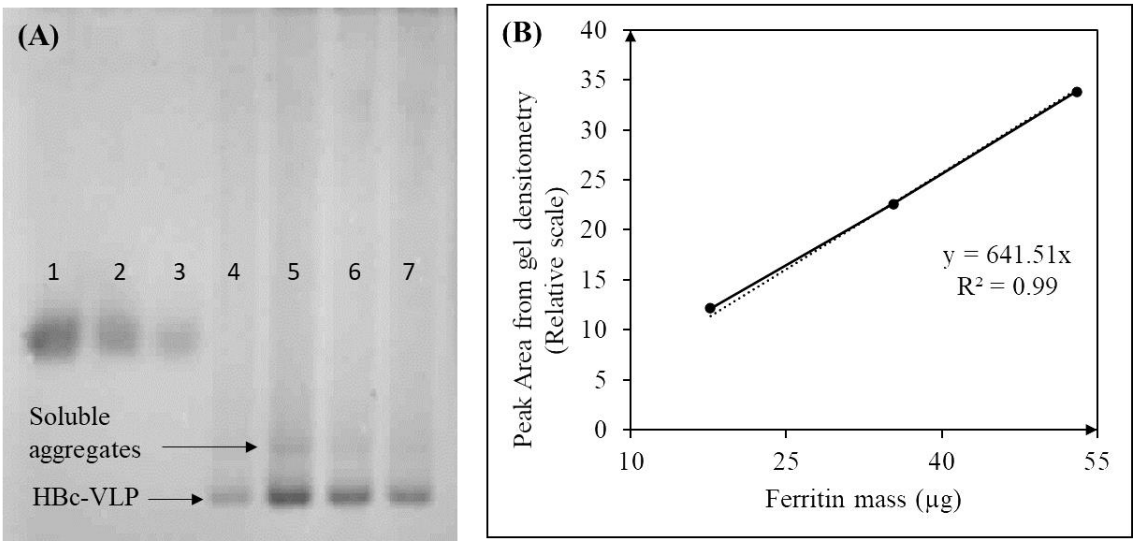

Fig. 4-1: Quantification of HBc-VLP by NAGE; (A) Lane 1, 2, and 3 are ferritin with known mass. Lane 4 is purified HBc-VLP standard. Lane 5 6 7 are clarified crude lysate with known protein mass on 0.8 % (w/v) native agarose gel; (B) The calibration curve between band intensity and ferritin mass.

It is feasible to plot the intensity of ferritin versus their loading mass as a standard curve, as is presented in Fig. 4-1B. The mass of VLP of each condition was quantified to determine the expression level based on their densitometry. For example, the EBNA1-HBc VLP band's intensity in lane 5, lane 6 and lane 7 correspond with 21.3 µg, 17.7 µg and 13.8

µg of EBNA1-HBc VLP in clarified crude lysate, respectively. All the quantification of EBNA1-HBc VLP in clarified crude lysate for screening experiment and optimisation experiment is determined using NAGE.

4.3.2 Factor screening by fractional factorial design

The focus of the preliminary experiment is to narrow down the number of significant process factors that influence the soluble production of EBNA1-HBc. Table 4-3 summarizes the fractional factorial design matrix, including six (6) factors, to determine the the yield of product per unit biomass, also known as, cellular yield (mg g^{-1} DCW), which was set as response 3.

The response was experimentally estimated based on the densitometry from NAGE, as is summarised in Table 4-3, the range of cellular yield of soluble EBNA1-HBc VLP varies from 42 to 163 mg g^{-1} DCW. The greatest yield is 163 mg g^{-1} DCW, which is achieved at 30 ^{o}C, 250 rpm, pH 6.2, induced at OD_{600} of 0.8 with 0.5 mM of IPTG without Mg^{2+} (experiment no. 29).

It is noticed that cultivation induced at OD_{600} of 2.0 produces a less soluble yield of EBNA1-HBc VLP (experiment no. 20), in comparison to culture induced at OD_{600} of 0.8 (experiment no. 29).

Fig. 4-2A presents the half-normal plot, which indicated the significant process factors and two-factor interaction (2FI). They are listed as **A** (Post-induction temperature), **B** (post-induction rotation speed of shaker), **C** (pH), **E** (cell density at induction), **AB** (post-induction temperature – rotation speed of shaker), **BD** (post-induction rotation speed of shaker and cell density at induction), **CD** (pH and IPTG concentration). Their p-values, which were achieved from the analysis of variance (ANOVA), were less than 0.05. Factor **D** (IPTG concentration) was added into the model to support the hierarchy. To validate the reliability of the model, Fig. 4-2B presents the normal probability plot of the external studentised residuals showing the error were normally distributed and insignificant.

The value of R^2 = 0.74 demonstrated how the equation fits the dataset, implying that 74 % of the variations in the soluble production of EBNA1-HBc can be explained by the model. The Adjusted R^2 = 0.65 is in reasonable agreement with Predicted R^2 = 0.50, less than 0.2 difference, and the Adequate Precision = 9.45. This finding indicates the model is applicable for navigating the design space.

Table 4-3: Preliminary experimental matrix for soluble production of EBNA1-HBc VLP.

	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**	**Factor 6**	**Response 1**	**Response 2**	**Response 3**
Exp. no.	A: Temp.	B: Rotation Speed	C: pH	D: IPTG conc.	E: Cell density	F: Mg^{2+}	Biomass DCW	Volumetric yield	Cellular yield
	^{o}C	rpm		mM	OD_{600}	mM	g L^{-1}	mg L^{-1}	mg g^{-1}
1	30	110	6.2	0.5	2.0	0	1.0	97.2	97.7
2	30	250	6.2	0.05	2.0	0	1.3	149.9	111.4
3	30	250	6.2	0.05	0.8	10	1.1	127.3	113.2
4	40	110	7.2	0.05	0.8	0	0.7	42.2	58.9
5	30	250	7.2	0.5	0.8	10	1.1	96.6	85.9
6	40	110	7.2	0.5	0.8	10	0.6	42.1	65.3
7	40	250	7.2	0.05	0.8	10	1.0	136.9	135.6
8	40	110	7.2	0.5	2.0	0	1.0	61.1	60.7
9	30	250	7.2	0.5	2.0	0	1.5	102.5	68.3
10	40	250	6.2	0.05	0.8	0	1.1	175.4	154.3
11	30	110	7.2	0.5	2.0	10	1.0	84.7	79.1
12	40	250	7.2	0.05	2.0	0	1.2	149.1	118.8
13	40	110	6.2	0.5	2.0	10	1.0	62.7	58.1
14	40	250	6.2	0.5	2.0	0	1.2	149.0	124.2
15	40	110	6.2	0.5	0.8	0	0.7	62.3	92.6
16	40	250	7.2	0.5	2.0	10	1.3	112.2	88.6
17	40	110	6.2	0.05	2.0	0	1.2	48.3	42
18	40	250	6.2	0.5	0.8	10	0.9	98.9	111.5
19	30	250	7.2	0.05	0.8	0	1.1	142.9	129.8
20	30	250	6.2	0.5	2.0	10	1.4	194.9	137.5
21	40	110	7.2	0.05	2.0	10	1.0	47.5	47.1
22	30	110	7.2	0.5	0.8	0	0.8	100.8	125.4
23	30	110	6.2	0.05	2.0	10	1.0	52.3	52.3
24	30	110	6.2	0.5	0.8	10	0.7	109.8	149.2
25	30	110	6.2	0.05	0.8	0	0.7	61.1	84.3
26	30	110	7.2	0.05	2.0	0	1.0	104.0	104.0
27	40	250	6.2	0.05	2.0	10	1.2	128.0	104.3
28	30	250	7.2	0.05	2.0	10	1.4	168.0	117.3
29	30	250	6.2	0.5	0.8	0	1.1	179.7	163.4
30	40	250	7.2	0.5	0.8	0	0.9	58.1	61.7
31	40	110	6.2	0.05	0.8	10	0.8	74.4	94.9
32	30	110	7.2	0.05	0.8	10	0.8	93.4	110.7

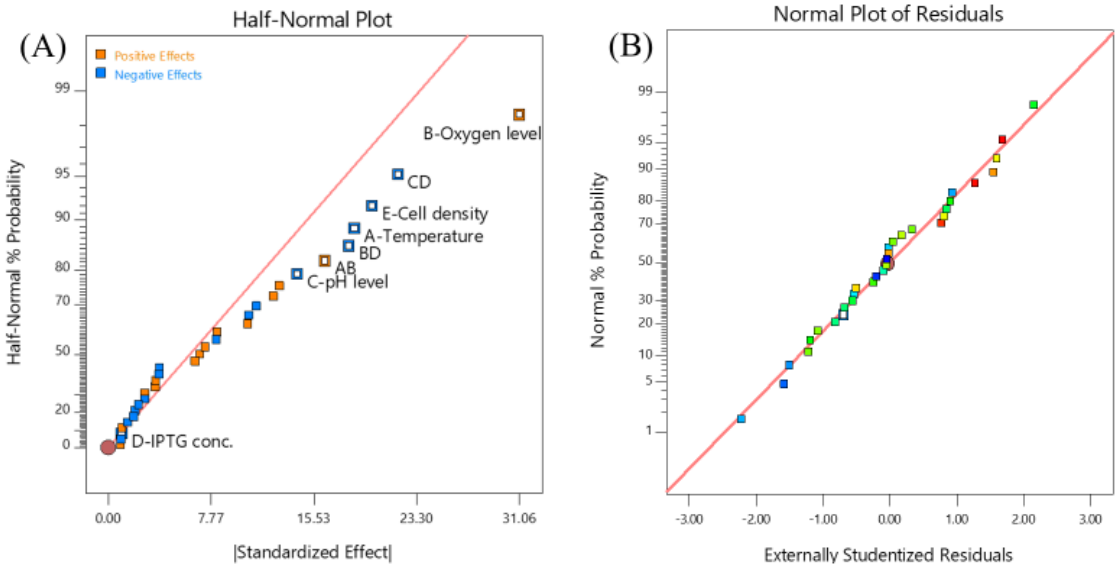

Fig. 4-2: Fractional factorial design analysis and model diagnosis for soluble production of EBNA1-HBc VLP; (A) The half-normal plot of standardized effect and (B) the normal (%) probability plot of externally studentised residual for the model.

Equation (4-3) predicts the yield of EBNA1-HBc VLP (mg g^{-1}) in preliminary study by fitting the experiment data to first-order polynomial equation (Eq. 2-1), where β is estimated from the experiment data, and X is the design factors:

$$\text{Yield}_{\text{EBNA1-HBc}}\ (\text{mg g}^{-1}) = +97.95 - 9.30*\mathbf{A} + 15.53*\mathbf{B} - 7.14*\mathbf{C} - 0.5073*\mathbf{D} - 9.96*\mathbf{E} + 8.18*\mathbf{AB} - 9.09*\mathbf{BD} - 10.94*\mathbf{CD} \quad \textbf{(Eq. 4-3)}$$

Fig. 4-3 summaries the interaction between **AB**, **BD**, and **CD**, even though IPTG by itself is not a significant factor, the interaction between IPTG concentration with pH, and the interaction between IPTG concentration with rotation speed influence the final yield.

With the addition of IPTG 0.05 mM, there was a huge boost in yield (70 mg g^{-1} to 120 mg g^{-1}) when shaker speed was set at 120 and 250 rpm, meanwhile, with IPTG 0.5 mM, the difference in yield between two shaker speed was not significant. Regarding pH level at induction, the difference in yield when the culture was induced with IPTG 0.05 mM was not significant in both pH 6.2 or pH 7.2, however, there was a remarkable negative impact in yield when the induction was conducted at pH 7.2 with IPTG of 0.5 mM, in comparison with culture induced at pH 6.2 (80 mg g^{-1} to 120 mg g^{-1}).

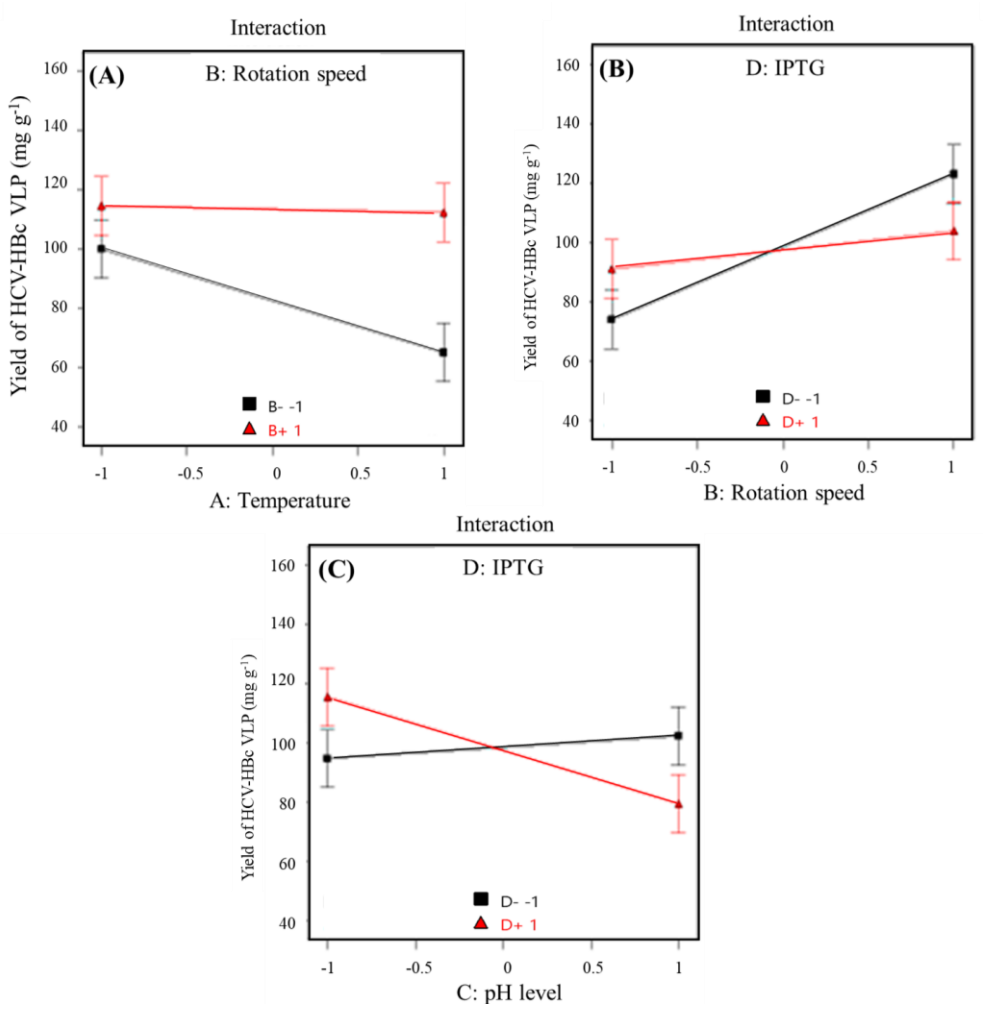

Fig. 4-3: The interaction of process parameter on soluble production of EBNA1-HBc VLP; (A) the effect of shaker speed and temperature; (B) the effect of shaker speed and IPTG concentration; (C) the effect of IPTG concentration and pH level at induction; (+1) means high level and (-1) means low level in coded value.

The influence of post-induction temperature was not significant when the level of post-induction shaker speed was high (250 rpm) but in case of low post-induction shaker speed (120 rpm), high temperature (40 ^{o}C) had a negative impact on the yield. Moreover, the negative sign of term E (cell density at induction) in Eq. 4-3 indicated there was a negative effect when culture was induced at mid-exponential phase (OD_{600} of 2.0).

4.3.3 Optimisation by response surface methodology

The outcome of the preliminary study revealed the significant factors that affecting the soluble production of EBNA1-HBC VLP. They included: 1) post induction shaker speed, 2) post-induction temperature and 3) cell density at induction. Soluble production of recombinant protein includes many biological pathways to support the cell viability and expression of foreign protein. Response surface model – central composite design (RSM-CCD) was applied for this complicated process. The design matrix included 20 experiments that consists of 8 cubic points, 6 axial points and 6 central points.

Table 4-4 presents the design of experiment for optimisation purpose where the response were set as the cellular yield of soluble EBNA1-HBc VLP (mg g^{-1} DCW) and A*, B*, C* were coded terms for the 3 variables, the post induction shaker speed, the post-induction temperature, and cell density at induction, respectively. The response was experimentally estimated based on the densitometry from NAGE.

Table 4-5 presents the ANOVA results of the quadratic model of soluble production of EBNA1-HBc. To evaluate whether a coefficient is significant or not, P value is determined. With a $P < 0.05$, the model terms are evaluated as significant. Particularly, all the linear terms **A*** and **C***, the interaction of **A*B*** (post-induction temperature and shaker speed) and quadratic terms $\mathbf{A^{*2}}$, $\mathbf{B^{*2}}$, and $\mathbf{C^{*2}}$ (higher order of process factors) have significant effects on the soluble production of the VLP. Factor **B*** was not a significant factor but was added to the model to support the hierarchy.

Table 4-4: Response surface model – central composite design matrix for optimisation of soluble expression of EBNA1-HBc VLP.

	Factor 1	**Factor 2**	**Factor 3**	**Responses**			
	Actual value			Actual value			Predicted value
Exp. no.	A* Shaker speed	B* Temp.	C* Cell density	Biomass DCW	Volumetric yield	Cellular yield	Cellular yield
	rpm	°C	OD600	g L^{-1}	mg L^{-1}	mg g^{-1}	mg g^{-1}
1	190	30	1.4	1.06	113.5	106.8	115.7
2	220	33	1.1	0.95	193.5	203.6	207.2
3	250	30	0.8	1.05	168.3	160.8	172.2
4	190	30	0.8	1.03	115.9	112.2	100.3
5	250	36	0.8	1.52	236.4	155.5	151.7
6	220	33	0.6	1.32	203.5	154.2	155.4
7	190	36	0.8	1.24	154.9	125.2	127.0
8	220	33	1.1	0.73	147.7	202.3	207.2
9	220	33	1.1	0.97	205.7	212.0	207.2
10	220	38	1.1	1.70	244.7	143.9	147.1
11	250	30	1.4	1.30	250.5	192.7	187.7
12	220	33	1.1	0.93	185.8	199.8	207.2
13	220	33	1.1	1.13	244.7	216.5	207.2
14	220	33	1.6	1.30	236.0	181.7	181.4
15	170	33	1.1	1.39	121.3	87.5	93.6
16	250	36	1.4	1.75	283.0	161.6	167.2
17	190	36	1.4	1.28	193.7	151.9	142.4
18	270	33	1.1	1.30	234.0	180.0	174.9
19	220	28	1.1	1.53	221.3	144.2	141.9
20	220	33	1.1	0.93	192.7	208.4	207.2

Table 4-5: ANOVA results of quadratic model for soluble EBNA1-HBc VLP production.

Source	Sum of Squares	df	Mean Square	F-value	p-value	
Model	26410.93	7	3772.99	56.1	< 0.0001	Significant
A (Rotation speed, rpm)	7977.2	1	7977.2	118.62	< 0.0001	Significant
B (Temperature, ^{o}C)	32.9	1	32.9	0.4892	0.4976	Not significant
C (Induction point, OD_{600})	815.76	1	815.76	12.13	0.0045	Significant
AB	1116.28	1	1116.28	16.6	0.0015	Significant
A^2	9590.94	1	9590.94	142.62	< 0.0001	Significant
B^2	7074.33	1	7074.33	105.2	< 0.0001	Significant
C^2	2707.25	1	2707.25	40.26	< 0.0001	Significant
Residual	807	12	67.25			
Lack of Fit	596.12	7	85.16	2.02	0.2283	Not significant

R^2 = 0.97; adjusted R^2 = 0.95; predicted R^2 = 0.91; Adequate precision = 21.91; df: degree of freedom

From ANOVA analysis, the fit statistics provided the good fit of the selected model where R^2 = 0.97 and adjusted R^2 = 0.95. With the difference less than 0.2, the predicted R^2 = 0.913 is in reasonable agreement with the adjusted R^2. Adequate precision was calculated to measure the signal to noise ratio, with the value of 21.90, the signal is larger than 4, showing the signal is adequate for predictive purpose (Myers et al., 2016). With R^2 = 0.97, the empirical model can explain 97 % of the variation in the soluble expression of EBNA-HBc in shake-flask fermentation.

ANOVA analysis calculated the F-value of model is 56.10 and F-value of Lack of Fit is 2.02, implied that the model is significant and the lack of fit is not significant relative to the pure error. These calculated statistic figures shown that the model predicts the mean value well.

As a major tool for evaluating the adequacy of the adopted model, Fig. 4-4 illustrates the normal probability plot of the external studentised residuals, demonstrating the errors are normally distributed and insignificant.

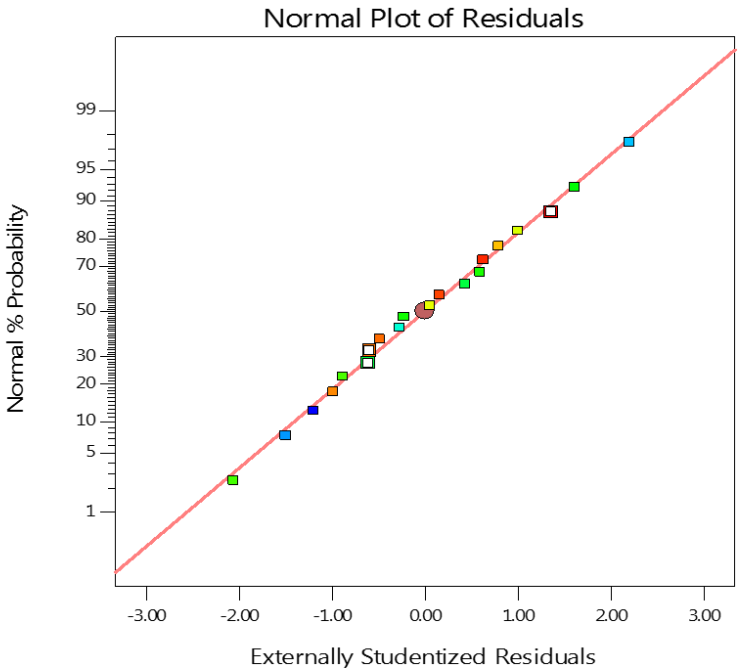

Fig. 4-4: Normal (%) probability plot of externally studentised residuals for the soluble production of EBNA1-HBc VLP in RSM model.

In coded forms, Equation 4-4 presents the empirical model that illustrated the response created the following regression equation (Eq. 2-2), where the positive and negative effects of the variables on the response are presented as plus (+) and (-) symbols. Eq. 4-4 was used to calculate the predicted value as is presented in Table 4-4.

$$\text{Yield}_{\text{EBNA1-HBc}}\ (\text{mg g}^{-1}) = +207.17 + 24.17*(\mathbf{A^*}) + 1.55*(\mathbf{B^*}) + 7.73*(\mathbf{C^*}) - 11.8*(\mathbf{A^*B^*}) - 25.80*(\mathbf{A^{*2}}) - 22.16*(\mathbf{B^{*2}}) - 13.71*(\mathbf{C^{*2}}) \quad \textbf{(Eq. 4-4)}$$

As is presented in Table 4-4, after conducting various combinations of fermentation process parameters, the cellular yield of EBNA1-HBc VLP increases from 87.5 to 217.0 mg g^{-1}. The lowest yield (87.5 mg g^{-1}) was recorded at experiment 15, in this case, post-induction rotation speed was 170 rpm, post-induction temperature was 33 ^{o}C and the cell density at induction point (OD_{600}) was 1.1, suggesting that insufficient dissolved oxygen level after induction has negative impact on the soluble production of EBNA1-HBc VLP.

The highest yield, 217.0 mg g^{-1}, was observed at one of the central points, the expression condition includes post-induction rotation speed at 230 rpm, post-induction temperature at 33 ^{o}C and the cell density at OD_{600} of 1.1. The central point data yielded a good agreement between six replicated, from 199.8.4 to 217.0 mg g^{-1}, showing that in

shake-flask fermentation, 1) EBNA1-HBc VLP favours the soluble production at 33 °C and 2) the increase of oxygen level by increasing the rotation speed after induction is necessary.

Moreover, from the empirical model, there is a good agreement between the predicted and actual values for EBNA1-HBc-VLP yield, showing the model is reliable to predict the soluble yield of EBNA1-HBc in shake-flask fermentation.

There is a negative sign of A*B* model term in Equation 4-4, it is assumed there is a negative effect of the interaction between those process factors. Fig. 4-5 presents the contour plot of EBNA1-HBc VLP soluble production at three (3) different induction points (0.8, 1.1 and 1.4 of OD_{600}). The soluble production of HBc VLP increases as the post-induction temperature rises until reaching their central point. For post-induction rotation speed of shaker, the production reaches its peak at point of 0.5 (235 rpm). Induction at early exponential phase (OD_{600} of 0.8) had lower soluble yield, from 120 to 191 mg g^{-1}. Cultures induced at mid exponential phase (OD_{600} of 1.2 to 1.4) have yield from 140 to 210 mg g^{-1}.

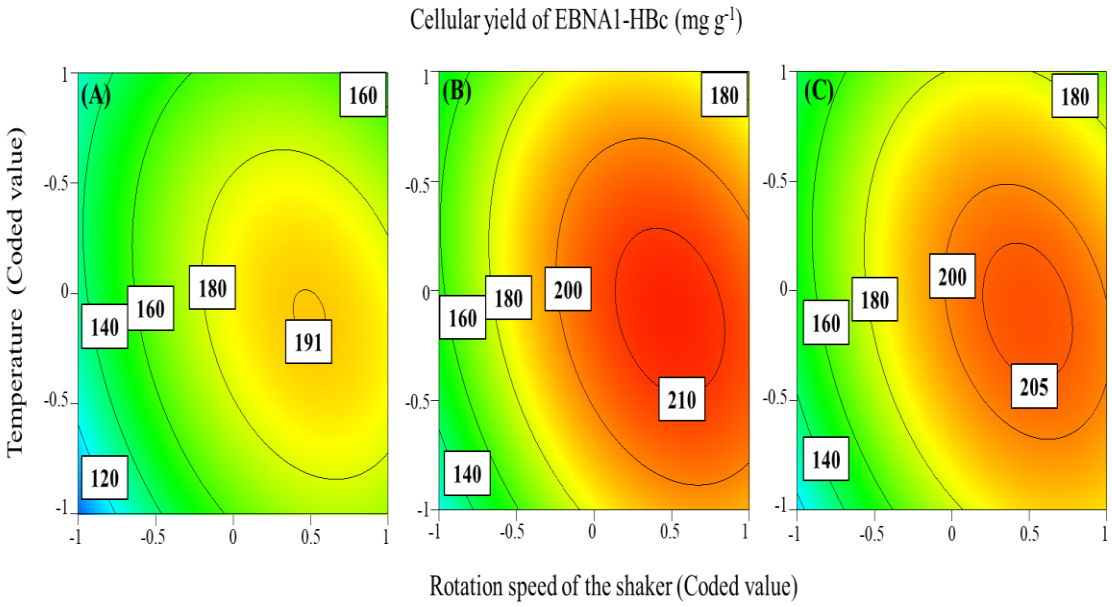

Fig. 4-5: Contour plots of the impact of two-factor interaction on the soluble production of EBNA1-HBc at three different induction points; (A) at coded value of minus 1, OD_{600} of 0.8; (B) at the central point (0), OD_{600} of 1.1; and (C) at coded value of 1, OD_{600} of 1.4.

4.3.4 Point optimisation feature and experimental validation

The potential direction for maximizing soluble production of EBNA1-HBc VLP was evaluated by Design-Expert software package. By using "point optimisation" technique, the optimal expression condition was obtained.

Fig. 4-6 illustrates contour of the desirability and soluble yield of EBNA1-HBc VLP. The mean value is predicted at 214.1 mg g^{-1} when culture is induced at the cell density at OD_{600} of 1.2, post-induction rotation speed of the shaker of 235 rpm, 0.5 mM of IPTG, and post-induction temperature at 33 ^{o}C.

The predicted value is given in the range of 201.5 to 226.6 mg g^{-1}. To validate the accuracy of the model, experiments were conducted in triplicate (n=3) using the optimal combination of factors. The soluble production was recorded at the mean value of 210.5 mg g^{-1}, and volumetric yield of 272 mg L^{-1}, showing there is a good agreement between predicted and experimental value.

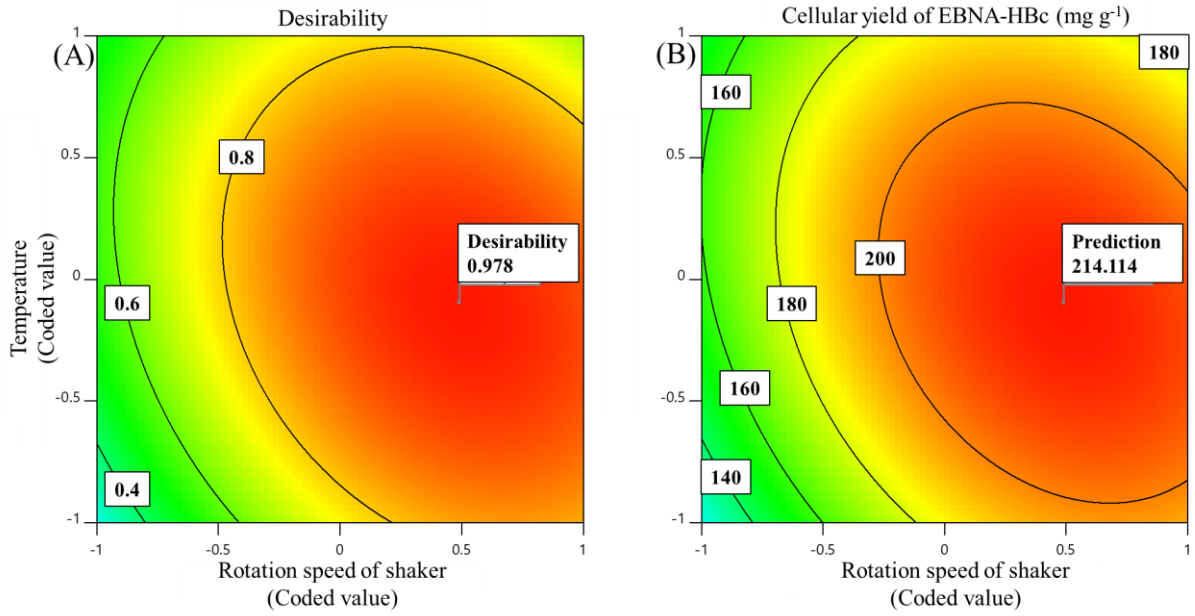

Fig. 4-6: Point optimisation achieved from RSM. The contour plot of (A) the desirability and (B) the soluble yield of EBNA1-HBc at optimal prediction point where A*=0.490 (235 rpm), B*= -0.096 (33^{O}C) and C*=0.281 (OD_{600} of 1.2) and IPTG of 0.5 mM.

4.3.5 Quantification of HBc-based VLP using heat precipitation & HPSEC

Heat precipitation is a relatively cheap and simple method to extract the HBc-capsid out of the clarified crude lysate.

The study of Ng et al. (2006) provides the range of temperature treatment that would not affect the structure of HBc-VLP. Fig. 4-7 presents the results of host cell protein removal by thermal treatment, including the purity of the EBNA1-HBc after treatment (Fig. 4-7A), and the recovery percentage of EBNA1-HBc before and after heat treatment (Fig. 4-7B) from SDS-PAGE result (Fig. 4-7C). The purity and recovery percentage of EBNA1-HBc were calculated based on Eq. 4-1 and Eq. 4-2, respectively.

As is shown in Fig. 4-7A, the purity of VLP after treatment increased from 23 % to 65.5 % as the sample was treated at 40 ^{o}C and 70 ^{o}C. There was no significant difference in the purity of VLP (65.5 % and 64.1 %) when samples were heat-treated at 70 and 75, ^{o}C.

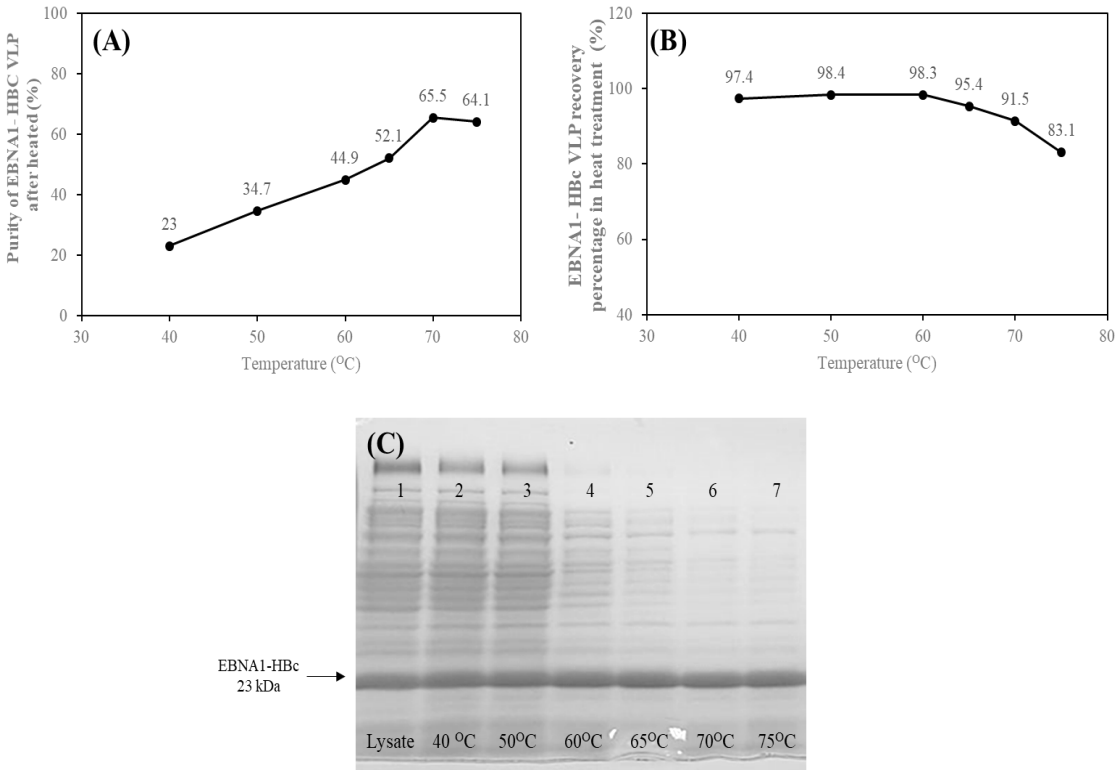

Fig. 4-7: Host cell protein removal by thermal treatment at different temperature, particularly, 40, 50, 60, 65, 70, and 75, ^{o}C; (A) the purity of VLP in the clarified crude lysate after heat treated; (B) recovery percentage of HBc-VLP before and after heat treatment; (C) SDS-PAGE result of heated samples.

The EBNA1-HBc recovery percentage varied from 97.4 % to 83.1 % as the samples were heat-treated from 40 to 75 ^{o}C. At 40, 50, 60, and 65, ^{o}C, the recovery rates were higher than 95 %. When treated at 70 ^{o}C, the recovery rate was 91.5 %, as is presented in Fig. 4-7B. However, the recovery percentage decreased to 83.1 % when sample was treated at 75 ^{o}C. Based on the recovery percentage and purity, results pointed out the most effective temperature was 70 ^{o}C.

Fig. 4-8 presents the quantification of VLP when treated by heat using the NAGE and HPSEC. Fig. 4-8A shows the densitometry of VLP after gel separation, the intensity of lane 4 (heat treated samples at 70 ^{o}C) is approximately equal to the intensity of lane 1 (clarified crude lysate). Moreover, lanes of heated sample have better resolution, these lanes have less noise from protein than lane 1.

Preliminary results from HPSEC shown the retention time of purified soluble EBNA1-HBc by TSK-5000 GW_{XL} at flow rate 0.5 mL min^{-1} is 15 min. Fig. 4-8B illustrates the difference between the chromatograms of before and after heat-treated at 70 ^{o}C at wavelength 280 nm. Both of them were resolved in TSK-5000 GW_{XL} at elution rate of 0.5 mL min^{-1}. The peak at 15 min before and after heat treatment were similar, demonstrating the HBc-VLP was reserved after thermal treatment. The composition of heat-treated sample was further investigated by collecting the peak fractions of different retention time (minute of 13, 15, 21, 23, and 24). The collected fractions were analysed by SDS-PAGE with silver staining, as is presented in Fig. 4-8C. From SDS-PAGE, EBNA1-HBc capsid were located at the retention time of 15 min. This evidence supported the result of HPSEC in Fig. 4-8B.

Fig. 4-8D presents the chromatogram of sample treated at 70 ^{o}C at two (2) wavelengths (260nm and 280nm). At the retention time of 15 min, the ratio of wavelength 260nm intensity over wavelength 280nm intensity (260/280 nm) is 1.7, showing there was a complex formation of protein and nuclear acid (Glasel, 1995). Sominskaya et al., 2013 reported that during self-assembly process of HBc capsid, encapsulation of host cell nucleic acid occurs simultaneously. To confirm the structure of EBNA1-HBc VLP in peak 15, TEM image was captured, as is shown in Fig. 4-8E, there was a mixture of T=4 and T=3 VLP with diameter of 34 and 32 nm.

At the retention time of 21 min of treated lysate, the ratios of 260/280 nm are larger than 2.0, showing the components of this peak mostly was nuclear acid. The shoulder peak at 22.5 of treated lysate suggested that there were small size impurities, in agreement to the result of SDS-PAGE, as is shown in Lane 6, Fig. 4-7C. It is noticed that there was a shift of peak 22.5 of clarified crude lysate to peak 21 after treatment by heat. This shifting is

predicted to be caused by the aggregation of host cell protein when heated. Although clarification by centrifugation has removed most of the insoluble precipitate, there remained some of small soluble aggregates of host cell protein at peak 21 of treated lysate.

The applicability of heat precipitation coupled with HPSEC could provide an alternative approach to quantify HBc VLP in crude lysate. To confirm the accuracy of this approach, peak area of different purified HBc-VLP masses was measured to develop the calibration curve in Fig. 4-8F. The mass of EBNA1-HBc VLP from clarified crude lysate was quantified based on the peak area achieved from peak integration.

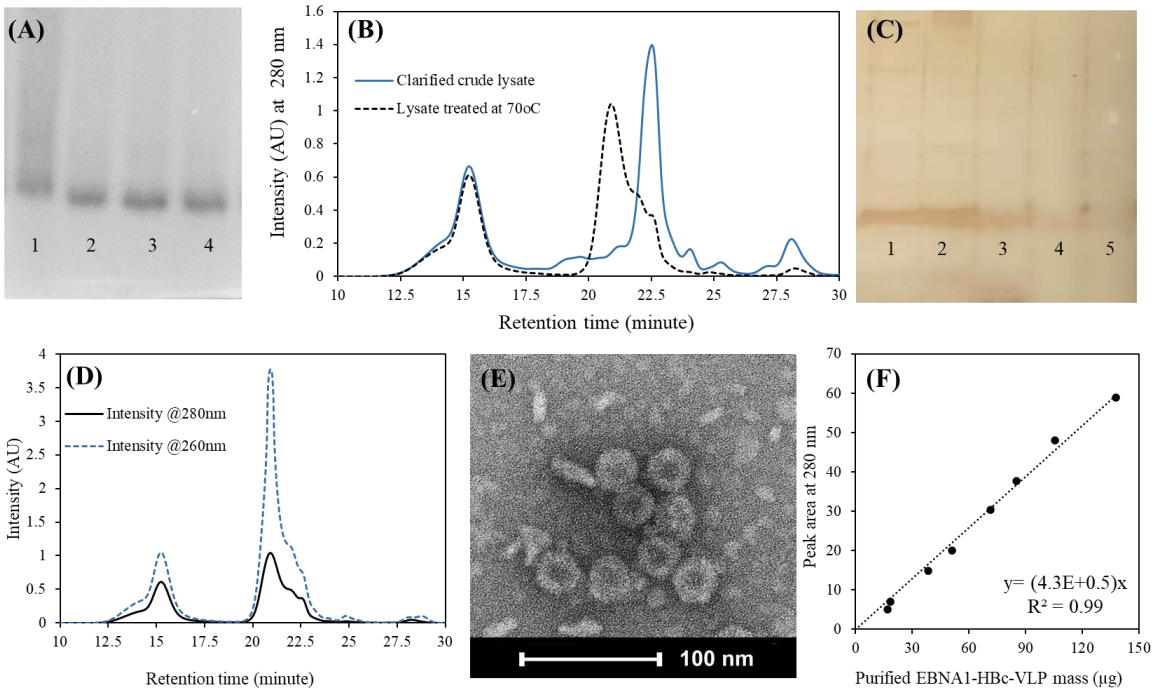

Fig. 4-8: Quantification of EBNA1-HBc VLP using heat precipitation and HPSEC; (A) the quantity of HBc-VLP in lysate before and after treated at 60, 65, and 70, ^{o}C, by NAGE; (B) the chromatograms of clarified crude lysate before and after heat treated at 70 ^{o}C, resolving in TSK-5000 GW_{XL} at elution rate of 0.5 mL min^{-1}; (C) SDS-PAGE with silver staining of the collected fractions at different retention time (minute of 13, 15, 21, 23, and 24) of 70 ^{o}C treated lysate, resolved by SEC; (D) the chromatogram of clarified crude lysate after treated at 70 ^{o}C; (E) the TEM image of soluble EBNA1-HBc VLP in peak 15; (F) the calibration curve of standard HBc-VLP.

As there was interference from nuclear acid, the mass quantified by HPSEC was higher than the mass quantified by NAGE. Fig. 4-9 illustrates the difference in VLP mass

quantified by two (2) methods from the set of 20 experiments in RSM-CCD. The experiment data were fitted in linear trend to generate a linear equation with $R^2 = 0.84$. This correlation could support the development of this analytical method to quantify the amount of VLPs in the clarified crude lysate.

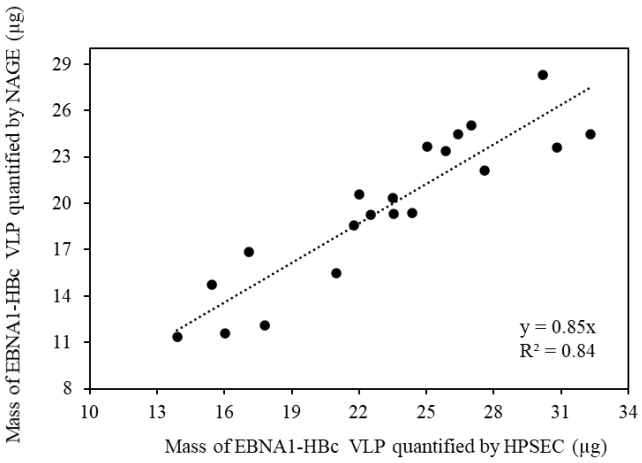

Fig. 4-9: The correlation of VLP mass quantified by NAGE and HPSEC from 20 experiments in RSM-CCD design.

In addition, Table 4-6 demonstrates the advantages and disadvantages of both methods. Importantly, NAGE requires an internal protein marker for every gel. This marker can be HBc-VLP itself or protein in appropriate size, such as ferritin, so they will not be flush out of the agarose gel. The limit of detection of NAGE also a constraint of this methods. The sample collected during fermentation and volume of lysis buffer should be considered to match the limit of detection of NAGE. On the contrary, HPSEC can detect small quantity of VLP in small culture media sample, enabling an approach to monitor the expression and assembly of VLP during cell culture process. If the drawback in nucleic acid contamination is eliminated, HPSEC approach is faster, and more reliable than NAGE.

Table 4-6: NAGE and HPSEC as quantitative methods.

No.	Criteria	Methods	
		NAGE	**HPSEC**
1	Sample preparation	Not required Heat treatment improves resolution	Heat treatment is required to improve the integrity of the column
2	Standard curves	Internal calibration curve is required for every analysis	Standard curve is developed only once.
3	Number of samples per analysis	Standard gel has 10 wells • 4 wells for internal makers • 6 wells for samples	One sample of a time
4	Process types	Manual	Automatic system
5	Process time	Seven (7) h per gel, including: • Running gel • Staining and de-staining	Approximately one (1) h per sample
6	Sample quantification	Manual by densitometry software	Peak integration in built-in software
7	Limit of detection	5 µg to 70 µg per well	Depends on the standard curves
8	Drawbacks	Time and labour-consuming	Nucleic acid interference

4.4 Discussion

The demand of producing large quantity of soluble and native folding of recombinant protein is high in pharmaceutical industry. Since the study of Clark et al. (1987), HBc platform has been intensively studied to be used as epitope carrier, presenting the antigen from different pathogen and diseases (Roose et al., 2013).

Using *E. coli* as expression system brings many benefits to the production process, however, the low-level of soluble expression of target product remains as challenge. Without optimised operation condition, recombinant protein has tendency to express in inclusion bodies. HBc-based VLP carrying EBNA1 had been successfully expressed in *E. coli* in this study. Based on the results, shake-flask fermentation and design of experiment provide the means to determine the effects of the investigated process parameters and their interaction on the soluble production of EBNA1-HBc VLP.

Noticeably, the effects of the interaction between post-induction shaker speed and post-induction temperature in both screening and optimisation experiment is consistent. In case of oxygen limitation (shaker speed of 120 rpm), the influence of temperature shows its significant impact on soluble yield of HBc-VLP (65 mg g^{-1} at 40 ^{o}C and 100 mg g^{-1} at 30 ^{o}C); low temperature expression increased the soluble production (Fig. 4-3A). It is reasonable since soluble expression of recombinant protein in *E. coli* favoured low temperature (less than 37 ^{o}C), taking the advantages from the lower specific growth rate, the increase in chaperone expression and activity (Ferrer et al., 2003), and lower protein synthesis and folding rate. When the oxygen level was increased (shaker speed in range of 190 to 250 rpm), induction at OD_{600} of 0.8, the influence of temperature has minimal effect on soluble yield (160 to 180 mg g^{-1}) as can be seen in Fig. 4-5A.

The result from Fig. 4-3C suggests that IPTG and pH level has strong interaction on the soluble expression. At low dose of inducer, IPTG of 0.05 mM, the yield of soluble expression at pH 7.2 is higher than the expression at pH 6.2. IPTG of 0.05 mM is enough to induce the expression of EBNA-HBc-VLP in *E. coli*, but in terms of high cell density culture in bioreactors, the level of IPTG should be proportional to higher cell level. This is the reason why IPTG of 0.5 mM was chosen for later optimisation. Cells induced with IPTG of 0.5 mM favoured the expression at pH 6.2, suggesting that acidic environment was beneficial to soluble expression of HBc-VLP.

The interaction between post-induction rotation speed of shaker and IPTG concentration highlights the demand of oxygen availability when *E. coli* express the foreign

protein. In *E. coli*, most of the major chaperone systems use the cycles of ATP binding and hydrolysis to stabilize non-native proteins, unfold misfolded proteins and provide the optimal condition for protein folding (Saibil, 2013). The level of ATP production was depended on the intracellular oxygen level. Moreover, oxygen also has influence of the gene expression by affecting the oxidative status of many enzyme (Knoz et al., 2008). Even with small dose of inducer (0.05 mM of IPTG), *E. coli* was able to produce a good yield of soluble EBNA1-HBc VLP when expressed at 250 rpm. It could be explained by the influence of strong promoter, T7 RNA polymerase (Studier and Moffatt, 1986).

Table 4-7 summaries the yield obtained from different approach, namely: 1) before optimisation, 2) fractional factorial design, 3) RSM before point optimisation and 4) RSM after point optimisation. By using "point optimisation" from Design Expert software, the soluble expression yield has increased 2-fold in volumetric yield and 1.7-fold in specific yield. Although fraction factorial design only improved the specific yield (mg g^{-1} DCW) by 1.3 x times, it is suitable to investigate both categorical variables (strain and plasmid) and continuous variables (temperature, inducer concentration, etc.). Meanwhile, RSM is more suitable for continuous variables.

Comparing the soluble yields of WT HBc and EBNA1-HBc under the same cultivation condition suggests that the insertion of non-structural epitope to N-terminus of HBc protein did not affect the soluble production of chimeric HBc-based VLP. However, this hypothesis needs to be confirmed with major immunogenic region and C-terminus.

In comparison to study of Yap et al, 2009, truncated HBc-based VLP carrying His-tag and β-galactosidase at N-terminal, was expressed in different temperature, 27, 30, and 37, ^{o}C. The yield from their study was reported at 70 – 75 mg L^{-1} with the solubility is greater than 90 % when culture was induced at 0.6-0.8 of OD_{600} with 0.5 mM IPTG for 16-18 h at 30 ^{o}C. These process parameters were similar to our operational process factor range, suggesting early induction and low temperature has beneficial impact on the soluble production of HBc-VLP. EBNA1-HBc VLP volumetric yield (bulk production) after point optimisation was approximately 4 x times, compared to Yap et al, 2009.

In addition, the thermal stability of EBNA1-HBc was noticed during analysis. The quantification of EBNA1-HBc VLP in crude lysate using heat precipitation and HPSEC was conducted. The results highlight the efficiency of thermal treatment when clarified crude lysate containing HBc-VLP was heated at 70 ^{o}C to remove host cell protein as is presented in Fig. 4-7. This method prevents the chance of column blocking and provides a clean feed into the HPSEC system.

Table 4-7: The comparison in soluble yields of EBNA1-HBc-VLP in different steps.

Methods	Vol. yield	Yield	Expression condition
	mg L^{-1}	mg g^{-1}	Temp ^{o}C/RPM/OD_{600}/IPTG mM
Before optimisation	154.9	125.2	36/190/0.8/0.5
Fractional factorial	179.7	163.4	30/250/0.8/0.5 – Exp. no. 29 Table 4-3
RSM before P.O[(a)]	147.7 - 244.7	199.8 - 216.5	33/220/1.1/0.5 – Centre points
RSM after P.O	272.0	210.5	33/235/1.2/0.5

[a] P.O: Point optimisation features from RSM in Design-Expert

Within the same sample, the mass of VLP quantified by HPSEC was greater than those quantified by NAGE, which can be explained by the interference of nucleic acid in HPSEC measurement. This interference can be explained by the encapsulation of host cell nucleic acid when full-length HBc protein is used (Sominskaya et al., 2013). In particular, C-terminal of full-length HBc contains arginine-rich domain (Porterfield et al., 2010). Moreover, according to Li et al., 2018 study, when the clarified crude lysate is heated at 70 ^{o}C, there is a chance that the pore of HBc capsid has been expanded, entrapping host cell nucleic acid and protein inside the capsid.

To overcome this drawback, DNAse I was used to minimise nucleic acid contamination (data not shown), but EBNA1-HBc became unstable immediately, suggesting the interactions between protein-nucleic acid contributed to capsid stability (Birnbaum and Nassal, 1990). A further investigation to remove nucleic acid and stabilize DNA-free targeting protein after removal of nucleic acid is critical in future research. Some purification approaches have been suggested by Li et al., 2018 and Zhang et al., 2021, including 1) using two-step heat treatment followed by hydrophobic interaction chromatography and 2) using ammonium sulfate precipitation to recover HBc VLP, followed by the disassembly-purification-reassembly process to obtain nucleic acid free HBc particles.

In addition, it is found that EBNA1-HBc VLP in clarified crude lysate can stand the temperature of 70 ^{o}C for 30 min but for purified EBNA1-HBc in 20 mM Tris-HCl buffer at pH 7.4. This phenomenon shows that component of lysis buffer has impacts on the thermal stability of EBNA1-HBc because it contains Triton X-100. It is reported by Alqueres et al. (2011) that Triton X-100 enhances the thermostability of target protein.

There are some limits of the current study, including the lack of media optimisation, and the disadvantages of RSM. First, LB media does not contain carbon source, growth and protein expression utilise mostly on nitrogen source. A media screening before cultivation in fermenter is essential. Second, even though RSM shows the effect of these investigated process parameters on the soluble production of EBNA1-HBc VLP but it only tells what happened, not the mechanism behind these observations. Knowing what happen is a key to further research to explain the underlying cause.

4.5 Chapter summary and conclusions

Based on experimental results and discussion of this chapter, the following important factors emerge:

1. An EBV vaccine candidate has been produced in *E. coli*. NAGE is an effective means to quantify EBNA1-HBc VLP from clarified crude lysate.

2. Six (6) process factors were investigated in two level fractional factorial design (FFD). The significant process factors screened from FFD, including post-induction temperature, post-induction rotation speed of shaker and cell density at induction, were optimised using response surface methodology – central composite design. After conducting 20 experiments and confirmed "point prediction" optimisation, the greatest yield of EBNA1-HBc was 210.5 mg g^{-1} DCW at 33 ^{o}C, 235 rpm, OD_{600} of 1.2, and IPTG of 0.5 mM.

3. It was found that thermal stability of EBNA1-HBc is suitable for the removal of host cell protein by heat precipitation. With the applicability of thermal treatment, size exclusion chromatography is able to quantify the amount of EBNA1-HBc VLP in clarified crude lysate. These new findings provide research direction to use HPSEC in high throughput screening to meet the demand for industrial production.

It is evident from this chapter that the expression folding state and productivity of EBNA1-HBc can be improved by optimisation of fermentation process factors. The sequential order of using FFD and RSM provides a mean to boost the production of HBc-based VLP in *E. coli*.

In the next chapter, Chapter 5, a media screening is performed and the optimal combination of process factors operated in laboratory-scale fermenter will be presented to further optimise the fermentation process to improve the soluble yield of EBNA1-HBc VLP.

CHAPTER 5

PROCESS OPTIMISATION TO IMPROVE SOLUBLE EXPRESSION OF VIRUS-LIKE PARTICLES IN BATCH AND FED-BATCH FERMENTATION

5.1 Introduction

A review of batch and fed-batch cultivation for soluble intracellular *E. coli*-derived recombinant protein and feeding strategy has been presented in Chapter 2. Regarding EBNA1-HBc, the productivity has been optimised using fractional factorial design and response surface methodology as outlined in Chapter 4. The optimal combination of post-induction process factors for shake-flask cultivation is 235 rpm, 33 ^{o}C, and OD_{600} of 1.2 with 0.5 mM of IPTG.

Luria-Bertani broth (LB) is a common peptide media in which *E. coli* can grow to high cell density of up to OD_{600} of 7 (Sezonov et al., 2007). The presence of yeast extract not only provides a nitrogenous source for building block synthesis in cell, but also reduces acetic acid secretion during bacterial growth (Tripathi et al., 2009). Because of the applicability of fermenter, cells are cultivated under a well-controlled condition, it is feasible to reach to high cell density culture (HCDC) (OD_{600} > 20). Therefore LB media is not suitable for HCDC. Finding an alternative media is crucial for fermenter scale cultivation. Additionally, cultivation time is a vital factor in HCDC because of plasmid loss (Nugent et al., 1983). Consequently, the cultivation loses the capability to produce recombine protein. Moreover, the extended culture period of batch culture leads to nutrient limitation, in which cells inhibit the synthesis of products but promote the production of protease, an enzyme that catalyses proteolysis (Rozkov and Enfors, 2004). It is hypothesized that when the culture reaches to high cell density, the limitation in nutrient or accumulation of by-products disrupts protein production (Eiteman and Altman, 2006).

It was also found that the cell density at induction point has significant impact on soluble production of EBNA1-HBc in shake-flask fermentation in Chapter 4. However, because of the significant differences in cell density between shake-flask and fermenter-scale, the influence of cell density in fermenter should be further investigated. In addition the feeding strategies proved to have significant impact on the production of recombinant protein (de Andrade et al., 2018). It is essential to determine these impacts on the soluble production of EBNA1-HBc.

The aim of Chapter 5 is to gain quantitative insight into batch and fed-batch cultivation. The following investigation are included: 1) the optimal combination of post-induction fermentation process factors achieved from Chapter 4 was applied to different culture media before scaling-up to a 14-L fermenter; 2) in laboratory-bench scale fermenter, batch

cultivations at different induction points were conducted; and 3) two (2) fed-batch cultivations with DO-stat feeding and constant feeding were operated.

A justification is that this provides the technical insights to improve soluble production of *E. coli*-derived HBc VLP. With promising results, it will be practical to boost production of recombinant protein in general and HBc VLP in particular by optimisation of process factors. This approach increases the chance of the vaccine candidate reaching pilot-scale production (100 L) and further commercialisation.

5.2 Material and methods

5.2.1 Expression strain

The expression strain that was used in this study was the same strain used in the optimisation of shake-flask fermentation in Chapter 4. Briefly, the recombinant *E. coli* BL21 (DE3) (Invitrogen©, USA) harbouring an IPTG inducible pET30a (+)/EBNA1-HBc plasmid is used for the expression of a 23 kDa EBNA1-HBc VLP.

All reagents were analytical grade (AR) unless otherwise stated.

5.2.2 Media screening

E. coli transformed with the recombinant plasmids from glycerol stock was cultured in 50 mL of sterilized LB media, containing tryptone (Oxoid, USA), yeast extract (Oxoid, USA), and NaCl (Chem-supply, Australia) with the composition listed in Table 1. The media were supplemented with 50 µg mL^{-1} kanamycin sulfate (Invitrogen©, USA), at 37 ^{o}C and 200 rpm (round per minute) for 16 h. Four (4) different media were prepared according to Table 5-1, including LB, TB, SB and 2xYT.

The overnight culture was added into 200 mL of different fresh sterilized media with 50 µg mL^{-1} kanamycin sulfate, in a 500-mL flask, at a ratio of 1 v/v % (seeding/culture volume), and cultivated at 37 ^{o}C and 200 rpm. Cell density of the culture (OD_{600}) was measured offline using UV-1600PC Spectrometer (VWR International©, USA). When the cell density reached OD_{600} of 1.2, the expression condition was adjusted to 32 ^{o}C and 230 rpm, and IPTG (Invitrogen ®, USA) added to a final concentration of 0.5 mM. The expression was continued for the next six (6) or nine (9) h. The cell culture samples were centrifuged at 12000 rpm, 20 ^{o}C, 10 min, washed with deionized water then stored at minus 20 ^{o}C.

Table 5-1: Media composition for screening experiment.

Media	Ingredient and concentration					
	Tryptone	Yeast extract	NaCl	K_2HPO_4	KH_2PO_4	Glycerol
	g L^{-1}	g L^{-1}	g L^{-1}	mM	mM	g L^{-1}
TB - Terrific Broth	12	24	-	72	17	5
SB - Super Broth	30	20	5	-	-	-
LB - Luria-Bertani	10	5	10	-	-	-
2YT - 2x yeast extract & tryptone	16	10	5	-	-	-

5.2.3 Batch cultivations

To determine the impact of induction point in batch fermentation, five (5) L of Terrific Broth media was used with the composition as is listed in Table 5-1. Yeast extract, tryptone and glycerol (Chem-supply, Australia) were dissolved and sterilised at 121 ^oC for 50 min. K_2HPO_4 and KH_2PO_4 (Chem-supply, Australia) were sterilised separately from the other three (3) component at 121 ^oC for 50 min. Following cooling to 37 ^oC, all of the media component was transferred to the 14-L fermenter and was supplemented with kanamycin sulfate at the final concentration of 50 µg mL^{-1} plus 0.01 % (v/v) of Antifoam 204 (Sigma-Aldrich, USA). The inoculum was prepared as described in section 5.2.2, i.e. seeding from the glycerol stock to 50 mL culture then 200 mL culture. When the optical density of 200 mL culture reached the OD_{600} of 4.5 to 5.0. The culture was used to inoculate five (5) L of sterilized TB media in stainless steel dished-bottom of 14-L water-jacketed vessel (57.4 cm height x 29.3 cm outer diameter) with BioFlo© 320 bioprocess control station (New Brunswick Scientific Co., Inc., NBS). To avoid evaporation on the head plate of the fermenter, an air-outlet condenser was used. For mixing and air dispersion, two six-bladed (2.11 cm height x 8.43 cm outer diameter) Rushton type impellers, stirred by top-mounted motor, were used.

All the cultures were conducted in biphasic of temperature, 37 ^oC for growth phase and 32 ^oC for induction phase. The electrode of pH and Dissolved Oxygen (DO) sensor (Mettler Toledo, USA) were used to monitor the pH and DO level of the culture. To maintain pH of 6.8, liquid ammonia NH_4OH 7 M (Chem-supply, Australia) and phosphoric acid H_3PO_4 4 M (Chem-supply, Australia) were added. Instrument grade compressed air was filtered with 0.2 µm Polytetrafluoroethylene (PTFE) membrane (Pall Corporation, USA), and was

sparged into the fermenter below the impeller using an L-type sparger. The dissolved oxygen level was maintained at 30 % by adjusting the agitation speed (300 to 800 rpm) and aeration rate (0.2 to 1 VVM – the volume of air per volume of liquid per minute) in cascade mode. The initial optical density of the fermenter culture was kept at 0.15 to 0.2 of OD_{600} to ensure consistency between cultivations. During cultivation parameters including optical cell density, temperature, pH, dissolved oxygen level, aeration, and agitation were observed.

EBNA1-HBc expression was induced by IPTG of 0.5 mM at three (3) different cell densities, OD_{600} of 5, 10 and 20. The post-induction temperature was changed from 37 to 32 ^{o}C. Samples for protein expression analyses were harvested 30 min before induction and for every 1 h for the next 6 h post-induction. Optical cell density of each sample was measured to determine dry cell weight using spectrometer. The cell culture samples were centrifuged at 12000 rpm, 20 ^{o}C, 10 min, washed with deionized water then stored at minus 20 ^{o}C.

5.2.4 Fed-batch cultivations

Initially cultures were conducted in batch mode. The procedure is described in section 5.2.3. Before switching to fed-batch, dissolved oxygen was used as an indicator for the depletion of carbon source. A pulse rises in DO value indicated the depletion of carbon source and feeding is initiated. The supplementary media contained 400 g L^{-1} glycerol, 120 g L^{-1} tryptone and 240 g L^{-1} yeast extract, was sterilized at 121 ^{o}C for 20 min before adding kanamycin sulfate. Two (2) feeding strategies, namely, constant volume feeding (non-feedback control) and DO-stat feeding (feedback control), were employed.

To achieve comparable results, the feed amount should be equal between two (2) feeding methods, DO-stat based feeding strategy was conducted first. With DO-stat feeding, the agitation speed was increased gradually from 400 to 700 rpm then fixed at 700 rpm. The aeration was set at 0.5 VVM throughout the cultivation. After the pulse rose in DO value, every time the DO value increased over 50 %, a dose of 2 to 3 mL of supplementary media was injected to maintain the DO value at 30 %. The mass of the feeding bottle was weighted every 30 min using an electronic weighing balance to determine the amount of feeding solution added to the culture.

With constant feeding, the DO was maintained at 30 % by cascade mode of agitation speed (300 to 800 rpm) and aeration rate (0.2 to 1 VVM). The feed rate (mL min^{-1}) was calculated based on the amount of feed used in DO-stat and 6 h post-induction time.

Cultures were induced with IPTG at final concentration of 0.5 mM at OD_{600} of 20, then continued for six (6) h at 32 ^{o}C. During induction phase, samples of 25 mL culture media were collected every 1 h for protein analysis. Samples were centrifuged 12000 rpm, 20 ^{o}C, 10 min, washed with deionized water then the cell pellet was stored at minus 20 ^{o}C.

5.2.5 Cell disruption and protein concentration assay

Ultra-sonication was used to extract the intracellular protein. Cell disruption was followed as is described in section 3.2.3. Protein concentration was determined by Bradford assay (Bradford, 1976). Bovine serum albumin (Sigma-Aldrich, USA) was used protein standard. All measurements of protein concentration were performed in duplicate.

5.2.6 NAGE and SDS-PAGE

Native agarose gel electrophoresis was used to separate capsid of HBc-based VLP from crude lysate. The detail procedure is followed as is described in section 4.2.5. The purity of HBc-based VLP were estimated on SDS-PAGE gel (Polyacrylamide 12 w/v %). A detailed SDS-PAGE procedure has been also mentioned in section 3.2.4. The component of TAE buffer, NAGE-loading buffer, SDS- loading buffer, SDS-PAGE running buffer, staining buffer and staining buffer was listed in Appendix B.

5.2.7 Protein purification by ammonium sulphate precipitation

The clarified crude lysate was precipitated with solid $NH_4(SO_4)_2$ (Chem-supply, Australia) at the final concentration of 1 M at 25 ^{o}C for 30 min. The protein precipitate was collected using centrifugation (9000 rpm, 20 ^{o}C, 15 min) and the supernatant was discarded. Protein pellet was solubilised in buffer containing 20 mM Glycine-NaOH (Chem-supply, Australia), 4 M Urea (Chem-supply, Australia) and pH 9.0 for 12 h at 4 ^{o}C. The solution containing the target protein was dialyzed against 20 mM Tris-HCl buffer at pH 7.4 for 48 h. Fractions from each step were collected to analyse by SDS-PAGE.

5.2.8 Transmission electron microscopy

To confirm VLP structure of purified EBNA1-HBc after dialysis, purified EBNA1-HBc was diluted in 20 mM Tris-HCl, pH 7.4 to the concentration of 0.3 mg mL^{-1}. The carbon-Formvar coated copper grids was glow discharged for 15 s using Solarus Plasma Cleaning System (Gatan, Inc). 5 µl of protein sample at concentration was deposited on the grid for 2 min and touched dry by blotting paper. The grid was negatively stained with 5 µl of 2 % (v/v) uranyl acetate aqueous solution for 2 min then blotted dry. The grids were examined using Fei Tecnai G2 Spirit TEM (FEI™, Japan), followed the instructions of Adelaide Microscopy Center, Australia.

5.3 Results

5.3.1 Media screening for EBNA1-HBc VLP production

After obtained the optimal combination of process parameters from response surface model in previous chapter, two (2) post-induction time and four (4) different media compositions, as is listed in Table 5-1, were screened to boost the soluble expression of EBNA1-HBc VLP in shake-flask cultivation. Fig. 5-1 presents the soluble production of target protein in four (4) media between two (2) post-induction time in terms of growth curve, volumetric yield and cellular yield.

Fig. 5-1A demonstrates the growth curves of *E. coli* in LB, TB, SB and 2YT. The final cell density achieved in LB media was the lowest (OD_{600} of 3.4 for 6 h and OD_{600} of 4.5 for 9 h). Meanwhile other media were able to support *E. coli* culture to higher final cell density, particularly, in TB, OD_{600} of 4.5 and OD_{600} of 6.3; and in SB, OD_{600} of 4.5 and OD_{600} of 6.7, regarding the post-induction period of 6 h and 9 h, respectively.

Fig. 5-1B presents the estimation of EBNA1-HBc VLP in clarified crude lysate. Ferritin nanoparticle (approximately 500 kDa) with designated mass was used as internal maker as is shown in lane 1 to lane 4. Moreover, the amount of EBNA1-HBc after 6 h and after 9 h expression is shown in lane 5 to 8 and Lane 9 to 12, respectively. The sequential order from left to right is yield in LB, TB, SB and 2YT. The EBNA1-HBc VLP were separated from soluble aggregates and impurities.

Fig. 5-1C demonstrates the volumetric yield of soluble protein by different culture media at 6 h and 9 h post-induction time. Yield achieved from LB media was the lowest (250 mg L^{-1} and 276 mg L^{-1} of culture media), in comparison with those achieved from

other media. For the expression of 6 h, TB, SB and 2YT obtained a similar volumetric yield in range of 365 to 375 mg L^{-1}, however, in case of expression of 9 h, the volumetric yield acquired from 2YT media decreased from 375 to 304 mg L^{-1} and those yield acquired from TB and SB increased slightly to the range of 392 to 398 mg L^{-1}. Fig. 5-1D illustrates the cellular yield (mg g^{-1}), which was evaluated by the ratio between volumetric yields and biomass as DCW. Overall, the average cellular yield obtained after 6 h expression (191 mg g^{-1}) was higher those after 9 h expression (156 mg g^{-1}). It is found that the cellular yields obtained from four (4) different media were approximately equal regardless the post-induction time. The cellular yields of 6 h induction varies from 181 to 203 mg g^{-1}, and those of 9 h induction varies from 146 to 166 mg g^{-1}.

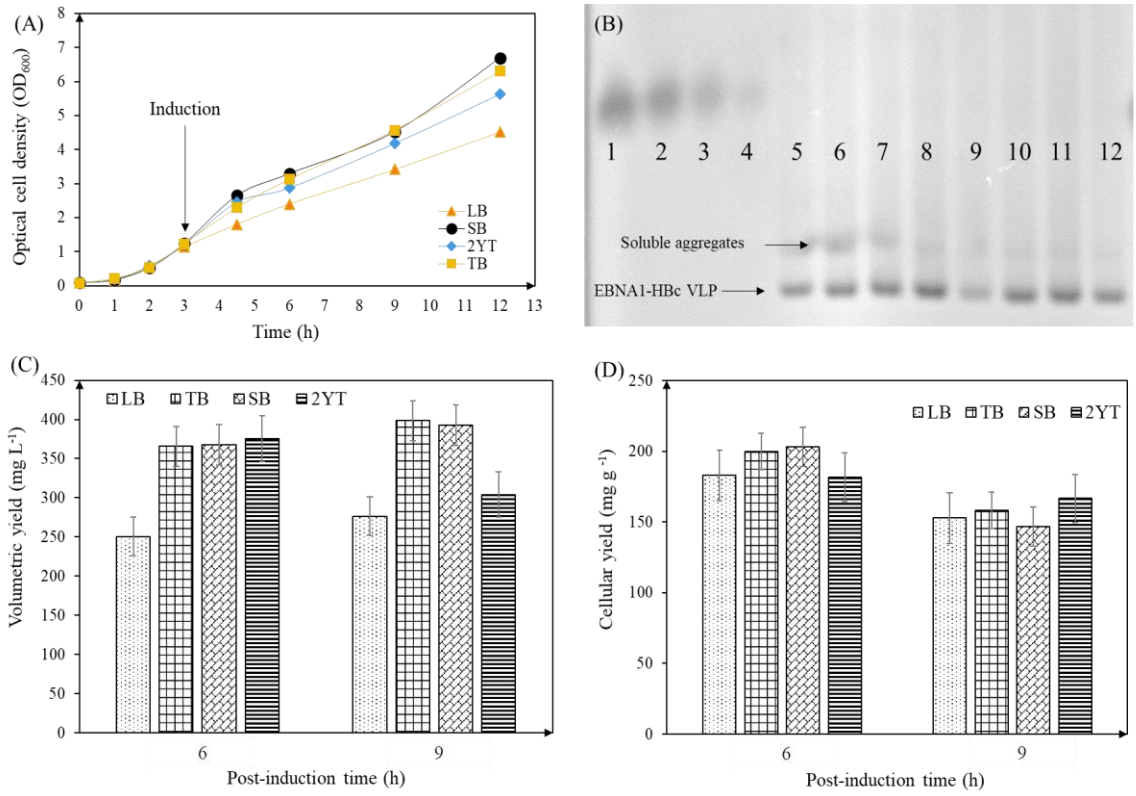

Fig. 5-1: Growth curves of *E. coli* and soluble yields of EBNA1-HBc in four media and two post-induction time. (A) The growth curves in LB (▲), SB (●), 2YT (♦) and TB (■); (B) the quantification of EBNA1-HBc in clarified crude lysate by NAGE. Lane 1 to 4, ferritin as internal marker. Lane 5 to 8, amount of EBNA1-HBc after 6 h expression; Lane 9 to 12, amount of EBNA1-HBc after 9 h expression. Sequential order from left to right: LB, TB, SB and 2YT; (C) The volumetric yields and (D) the cellular yields of EBNA-HBc expressed in LB, TB, SB and 2YT from left to right.

From these data revealed in Fig. 5-1, TB and SB were more effective than 2YT and LB, suggesting that the nutrient richness of medium and induction period have apparent effects on the volumetric yield of soluble and correct folded of EBNA1-HBc VLP in *E. coli*. Between TB and SB, in our study, only TB media composition has glycerol, which is an important carbon source to the survival of *E. coli* (Kram and Finkel, 2015), for this reason, further experiments in fermenter were conducted in TB media with 6 h post-induction time.

5.3.2 Purification using ammonium sulphate precipitation

The molecular weight of fully assembled VLP is significantly larger than other host cell protein, providing a pathway to high recovery rate by primary purification method, ammonium sulphate precipitation (Freivalds et al., 2011). Protein sample harvested from fermenter-scale cultivation was incubated with 1M ammonium sulphate at 25 ^{o}C for 30 min. It is feasible to collect 90 % of EBNA1-HBc VLP in protein precipitate with the purity of 80 %. After solubilised the precipitate in 4 M urea and dialyzed against Tris-HCl buffer, the overall recovery was 78 % with purity of 95 %. These measurements were acquired from densitometry of SDS-PAGE image, as is presented in Fig. 5-2A. EBNA1-HBc was located at at 23 kDa.

Moreover, TEM is used to confirm the correct HBc VLP structure after dialysis, as is presented in Fig. 5-2B. EBNA1-HBc folded into two classes based on their triangular number, T=3 (diameter of 32 nm) and T=4 (diameter of 35 nm), consists of 180 monomers and 240 monomers, respectively. The evidence shows that *E. coli*-derived EBNA1-HBc produced in fermenter-scale still maintains the correct VLP conformation.

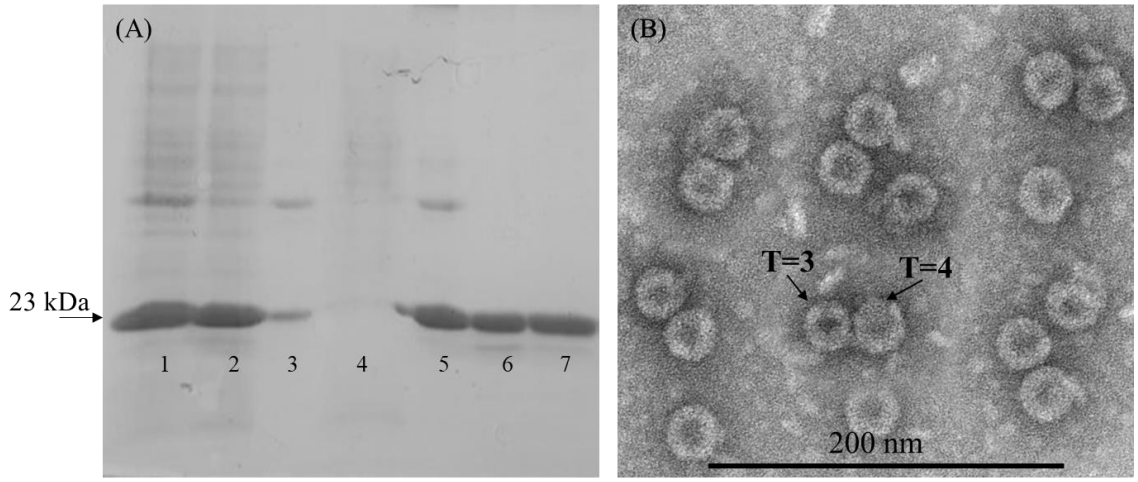

Fig. 5-2: The purification of EBNA1-HBc using ammonium sulphate precipitation. (A) Purity of EBNA1-HBc determined by SDS-PAGE. Lanes 1, 2 and 3 are the total cell protein, clarified crude lysate and inclusion bodies, respectively. Lanes 4 and 5 are ammonium sulphate supernatant and ammonium sulphate precipitate, Lanes 6 and 7 are sample after solubilisation in urea and after dialysis; (B) TEM image of correct HBc VLP structure after dialysis.

5.3.3 Influence of cell density at induction in batch cultivation

Samples taken from specific culture were lysed and purified using ammonium sulphate precipitation. The yield of EBNA1-HBc reported here has high purity level (> 95 %) as stated in section 5.3.2. It is established that the dry cell weight is equal to 0.4 g L^{-1} OD^{-1}.

Fig. 5-3 illustrates the expression of EBNA1-HBc when induced at three (3) different cell density, including OD_{600} of 5, 10 and 20. Overall, the results emphasized the cell density at induction has direct influence on the volumetric yield of EBNA1-HBc in *E. coli*. It is found that induction at early exponential phase (OD_{600} of 5) achieved highest volumetric yield and cellular yield, 1300 mg L^{-1} and 136 mg g^{-1}, respectively. The higher cell density at induction was, the lower the yields was. The lowest yield obtained were from the cultivation induced at OD_{600} of 20, as 418 mg L^{-1} and 26 mg g^{-1}.

The temperature of the culture was adjusted from 37 ^{o}C to 32 ^{o}C at the time of induction. From Fig. 5-3A, 5-3B and 5-3C, after induced by IPTG, cells were continued to grow at lower growth rate for first 2 h, then were suppressed from hour 2 to hour 5. It is suggested that IPTG lost its suppression property from hour 5 to hour 6 observed by the

increment of biomass, as is presented in Fig. 5-3A and Fig. 5-3C. Besides that, the soluble protein reached their maximum volumetric yield after 4 h of post-induction.

To estimate the ratio between soluble and total EBNA1-HBc expressed, cell samples harvested at the end of the cultivation were lysed. The crude lysate was further centrifuged to separate 1) soluble intracellular protein and 2) inclusion bodies. Both fractions were analysed using 12 % reducing SDS-PAGE. From Fig. 5-3D, lane 1, 3 and 5 showed a high-intensity band at 23 kDa. By scanning the densitometry of these bands against those in lane 2, 4, and 6, it is estimated that the ratio of soluble/total EBNA1-HBc expressed in fermenter scale was more than 90 %, showing that more than 90 % of EBNA1-HBc expressed in soluble form.

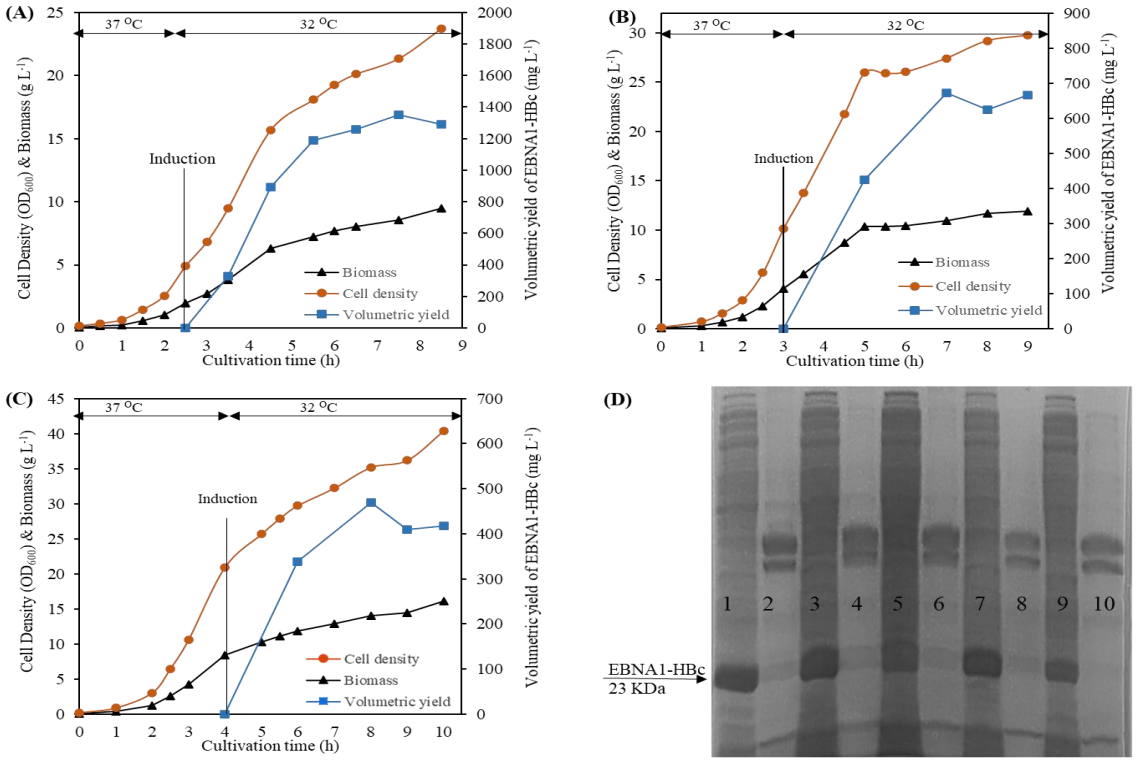

Fig. 5-3: Expression of EBNA1-HBc in *E. coli* induced at three different cell density. Optical cell density (OD_{600}) (●), biomass (g L^{-1}) (▲), and volumetric yield (mg L^{-1}) (■). (A) Culture induced at OD_{600} of 5; (B) Culture induced at OD_{600} of 10; (C) Culture induced at OD_{600} of 20. (D) Expression forms (product folding state) confirmed by SDS-PAGE. Lanes 1-2, Lane 3-4, and Lane 5-6, soluble and insoluble fraction of final sample from culture that was induced at OD_{600} of 5, 10, and 20, respectively. Lanes 7-8, Lanes 9-10, soluble and insoluble fraction of final sample from culture that was induced at OD_{600} of 20 with constant feeding and DO-stat feeding, respectively.

5.3.4 Influence of feeding strategy in fed-batch cultivation

Batch cultivation has limitation as the amount of nutrient is fixed during the operation. It is feasible that cells have utilized all the essential nutrient to generate biomass, leading to poor productivity. To overcome this drawback, fed-batch cultivation was employed to supply more nutrient for EBNA1-HBc production.

Fig. 5-4 presents the cultivation and expression of EBNA1-HBc in fed-batch mode. For accurate comparison between two strategies, DO-stat based feeding was conducted first to measure the amount of supplementary media, which was approximately 250 mL. Therefore, 250 mL of supplementary media was supplied at constant feeding rate of 0.7 mL min^{-1} for six (6) h of induction time.

As are illustrated in Fig. 5-4B and Fig. 5-4D, the feeding profiles were similar to each other. Since the DO level of constant feeding strategy was controlled in cascade mode by adjusting agitation and aeration rate, the DO level of culture was consistent at 30 %, however, the DO level of culture using DO-stat feeding strategy was more fluctuated because it was controlled by the addition of feed solution. Overall, Fig. 5-4A and Fig. 5-4C demonstrates the continuous production of EBNA-HBc over 6 h. Both volumetric and cellular yields of these two fed-batch fermentations were higher than those yield in batch mode. It is remarkable that constant feeding strategy was able to boost the volumetric yield from approximately 400 mg L^{-1} to 1800 mg L^{-1}, an increase of 4.3-fold, meanwhile, DO-stat based feeding strategy only increased the productivity by 1.7-fold, as compared in Table 5-2.

Besides that, the final cell density achieved from constant feeding was higher than those achieved from DO-stat feeding by 2 g L^{-1}. Another noteworthy point is the results from SDS-PAGE, as is presented in Lane 7 to 10 of Fig. 5-3D. Most of the target protein (> 95 %) was expressed as soluble form in both cases of feeding strategy shown in lane 7 and lane 9.

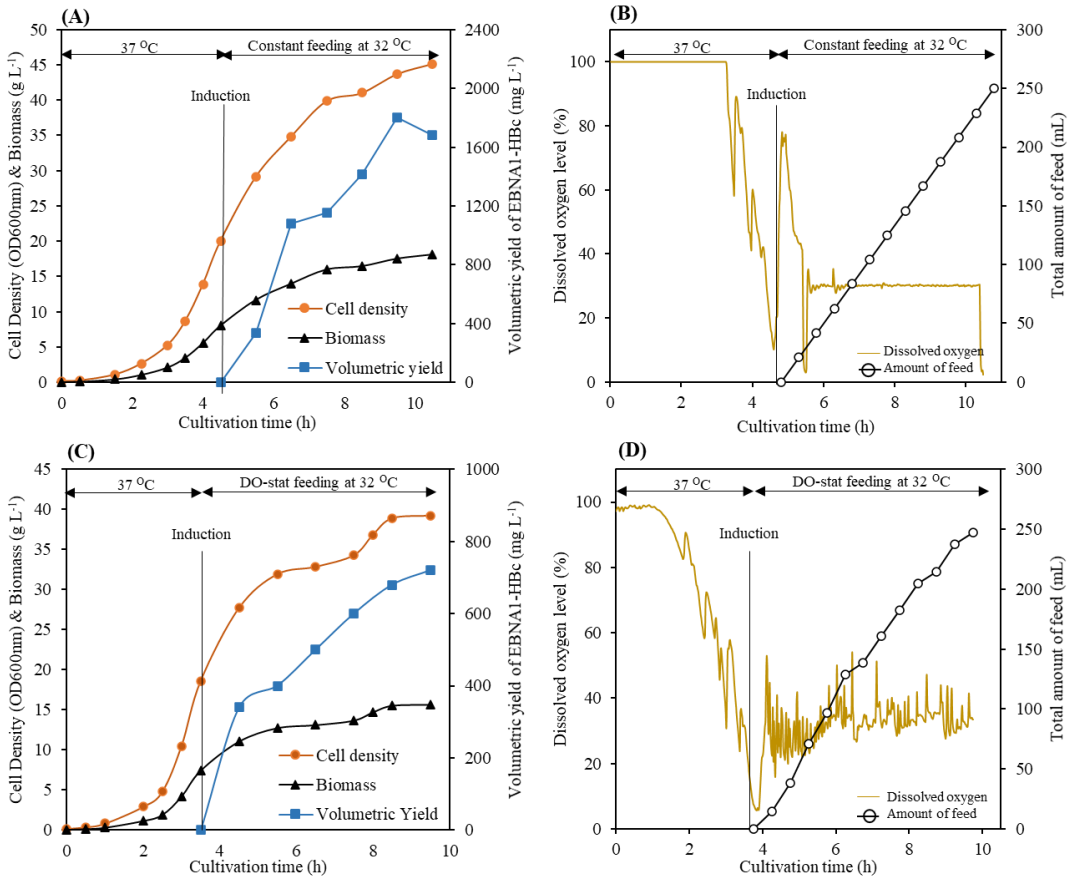

Fig. 5-4: Cultivation and expression of EBNA1-HBc in *E. coli* induced at OD_{600} of 20 with DO-stat feeding and constant feeding. (A) and (C) Biomass and productivity from constant feeding and DO-stat feeding, respectively; Optical cell density (OD_{600}) (●), biomass (g L^{-1}) (▲), volumetric yield (mg L^{-1}) (■); (B) and (D) Dissolved oxygen level and amount of feed using the constant feeding and DO-stat feeding strategy, respectively; Dissolved oxygen level (%) (-) and total amount of feed (mL) (o).

Table 5-2: EBNA1-HBc expression in *E. coli* in batch and fed-batch cultivations.

	Induction at OD_{600} of 5	**Induction at OD_{600} 10**	**Induction at OD_{600} of 20**	**Induction at OD_{600} of 20**	**Induction at OD_{600} of 20**
Type of cultivation	Batch	Batch	Batch	Fed-batch	Fed-batch
Feeding strategy	N/A	N/A	N/A	DO-stat feeding	Constant feeding
Time of culture (h)	8.5	9	10	9.5	10.5
Final cell density (OD_{600})	23.7	29.8	40.4	39.8	45.2
Final biomass conc. (DCW g L^{-1})	9.5	11.9	16.2	15.9	18.1
Soluble/Total EBNA1-HBc (%)	>90	>90	>90	>95	>95
Before-induction specific growth rate (h^{-1})					
Pre-induction phase	1.3	1.3	1.1	1.3	1.1
At the time of induction	1.3	1.1	0.7	1.1	0.7
Post-induction specific growth rate (h^{-1})					
Post-induction (hour 0 to hour 2)	0.58	0.5	0.17	0.27	0.27
Post-induction (hour 2 to hour 6)	0.1	0.04	0.07	0.07	0.04
Overall (6 h)	0.25	0.16	0.1	0.1	0.12
Volumetric yield (mg L^{-1})	1290	666	418	700	1800
Cellular yield YP/X (mg g^{-1})	136	56	26	44	99
Volumetric production rate (mg L^{-1} h^{-1})	151.8	74	41.8	73.7	171.4
Cellular production rate (mg g^{-1} h^{-1})	16	6.2	2.6	4.6	9.4
Soluble protein expression level (%)	31.3	21.3	8.9	15	32.1
Volumetric yield increase (fold)	3.1	1.6	1.0	1.7	4.3
Cellular yield increase (fold)	5.2	2.2	1.0	1.7	3.8
EBNA1-HBc protein mass (g) in volume of culture (L)	**6.8 g** in 5.2 L	**3.4 g** in 5.2 L	**2.1 g** in 5.2 L	**3.8 g** in 5.45 L	**9.8 g** in 5.45 L

5.3.5 Media cost per protein analysis

To evaluate the process economics, the direct production cost was broken down to the cost of raw materials (antibiotic, inducer and media composition), and utilities (Cardoso et al., 2020). Here, the production cost was focused on the cost of media per protein.

As is presented in Table 5-3, media cost per protein (AU$ per gram of protein) was evaluated for five (5) cultivation based on the amount of media ingredient, inducer and antibiotic used. In batch cultivation, media cost per protein was $10.8, $22.1, and $35.9 in cultivations that induced at OD_{600} of 5, 10 and 20, respectively. The estimation show that induction at early exponential phase (OD_{600} of 5) was most cost-effective in terms of media cost. It reduced the media cost per protein by 70 %, compared with those in the cultivation that induced at (OD_{600} of 20). In fed-batch cultivation, the media expense per protein was $25.0 and $9.7 for DO-stat feeding strategy and constant feeding strategy, showing that constant feeding strategy has significant benefits in media cost per protein. Details of media ingredient cost per unit (AUD$ per g) can be found in Appendix C – Table C-1.

Table 5-3: Media cost per protein analysis.

		Cultivation 1		**Cultivation 2**		**Cultivation 3**		**Cultivation 4**			**Cultivation 5**		
	Units	Initial medium	Expenses ($)	Initial medium	Expenses ($)	Initial medium	Expenses ($)	Initial medium	Additional medium	Expenses ($)	Initial medium	Additional medium	Expenses ($)
Volume	L	5.2		5.2		5.2		5.2	0.25		5.2	0.25	
Glycerol	g L^{-1}	5	0.9	5.0	0.9	5.0	0.9	5.0	400.0	4.2	5.0	400.0	4.2
Yeast extract	g L^{-1}	24	20.5	24.0	20.5	24.0	20.5	24.0	240.0	30.3	24.0	240.0	30.3
Tryptone	g L^{-1}	12	14.9	12.0	14.9	12.0	14.9	12.0	120.0	22.0	12.0	120.0	22.0
K_2HPO_4	g L^{-1}	12.5	7.9	12.5	7.9	12.5	7.9	12.5	0.0	7.9	12.5	0.0	7.9
KH_2PO_4	g L^{-1}	2.3	1.1	2.3	1.1	2.3	1.1	2.3	0.0	1.1	2.3	0.0	1.1
Kanamycin sulfate	g L^{-1}	0.05	3.1	0.05	3.1	0.05	3.1	0.05	0.05	3.3	0.05	0.05	3.3
IPTG (0.5 mM)	g L^{-1}	0.12	24.6	0.12	24.6	0.12	24.6	0.12	0.0	24.6	0.12	0.0	24.6
H_3PO_4 (21 % w/v)	mL	27	0.3	40.0	0.4	45.0	0.4	74.0		0.7	36.2		0.3
NH_4OH (12.5% w/v)	mL	10	0.2	115.0	1.8	123.0	1.9	72.0		1.1	78.3		1.2
Total cost	AUD$		**73.4**		**75.1**		**75.3**			**95.2**			**95.0**
Induction point		OD_{600} of 5		OD_{600} of 10		OD_{600} of 20		OD_{600} of 20			OD_{600} of 20		
Type of cultivation		Batch		Batch		Batch		Fed-batch with DO-stat feeding			Fed-batch with constant feeding		
EBNA1-HBc protein mass	g	**6.8 g** in 5.2 L		**3.4 g** in 5.2 L		**2.1 g** in 5.2 L		**3.8 g** in 5.45 L			**9.8 g** in 5.45 L		
Medium cost per protein	AUD$ per g	**10.8**		**22.1**		**35.9**		**25.0**			**9.7**		

5.4 Discussion

Hepatitis B core is one of the most attractive platforms to display foreign epitope of interest (Roose et al., 2013). It has been a challenge to produce high yield of soluble HBc VLP as HBc VLP tends to be expressed as inclusion bodies (Suffian et al., 2017). Results from Chapter 4 revealed that it is feasible to improve both solubility and yield of *E. coli*-derived EBNA1-HBc by optimisation of the process factors. As the cell density in fermenter cultivation is significantly greater than cell density in shake flask cultivation, the influence of cell density at induction point should be determined. In addition to process factors, media optimisation plays a vital role in recombinant production as well.

Screening of the growth media is an effective method to boost the cell density and production of recombinant protein (Islam et al., 2007). Four (4) complex media, including LB, TB, SB and 2YT, were used to evaluate the expression of EBNA1-HBc in 6 h and 9 h of post-induction time. From Fig. 5-1, the volumetric yield achieved from LB was lower than other yields achieved from other media, this could be explained by the nutrient availability and biomass. Since the cellular yields achieved from these media were approximately equal, the increase in volumetric yield came from the higher biomass. TB was able to increase the volumetric yield 1.45-fold, in comparison to yield in LB media. Among four media, only TB has carbon source. Liu et al. (2005) reported that the growth rate of *E. coli* on glycerol during exponential phase without any recombinant gene is 0.54 h^{-1}, less than growth rate on glucose, which is 0.97 h^{-1}. It is reported that *E. coli* does not produce significant amount of acetate when culturing on glycerol (Lee, 1996) and slow utilization of glycerol in *E. coli* is found to reduce the accumulations of acetate (Martinez-Gomez et al., 2012). These studies support the argument that glycerol is beneficial for EBNA1-HBc production.

In Chapter 4, heat precipitation and NAGE has been used to evaluate the soluble expression level of different expression condition. The purpose of using heat precipitation is to provide a clean feed for HPSEC to determine the expression level. However, as stated in Chapter 4 – Discussion, the interference of nucleic acid makes the measurement less accurate. Because of the significant difference in quantity, it will be impractical to use heat to remove host cell protein in terms of cost and efficiency. Another key factor is the limit of detection in NAGE, sample needs to be diluted so the protein concentration of sample is in range of detection, increasing the deviation in measurement. Therefore, an alternative primary recovery step in purification process is necessary to prepare high quality sample for further determination of soluble protein expression level in fermenter scale cultivation. Ammonium

sulphate precipitation and urea resuspension is proved to maintain the structure of the assembled VLP with the high protein recovery yield and purity (Zhang et al., 2020). It is established based on the complete disassembly of HBc VLP particle into assembly unit while maintaining the primary conformation of the assembly subunit (Zhang et al., 2020). The application of ammonium sulphate precipitation and urea resuspension for EBNA1-HBc VLP achieved the protein recovery yield of 78 % with purity of 95 %. This method provides an accurate means to determine the soluble expression of EBNA1-HBc.

It is established that biphasic cultivation strategy can be used to optimise the protein expression (Kaisermayer et al., 2016), culture was cultured at 37 ^{o}C to gain optimal biomass before culturing at 32 ^{o}C for protein expression. After conducted three (3) batch cultivation with different induction points, it is found that most of the target protein was expressed as soluble form, as is presented in Fig. 5-3D. Cell density at induction point has been proved to have significant impact on the soluble expression of EBNA1-HBc (Fig. 5-3A, 5-3B and 5-3C). These findings were consistent to the results that achieved in shake-flask fermentation. The higher cell density at induction point is, the lower yield is. The reason behind this could come from the nutrient availability and accumulation of by-product, such as acetate (Lee, 1996; Choi et al., 2006). Because recombinant protein synthesis demands a large amount of resource in both energy and nutrient, when culture enters the later stage of exponential phase, amino acid starvation is often occurred during the expression, making it hard for cells to utilise the remaining nutrient for protein production (Tripathi et al., 2009).

As is presented in Table 5-2, the volumetric yield of culture induced at OD_{600} of 5 is 1.6-fold and 3.1-fold, compared to culture that induced at OD_{600} of 10 and 20, respectively. Moreover, shake-flask cultivation when culture is induced at 33 ^{o}C, OD_{600} of 1.2 with 0.5 mM of IPTG achieves the optimal volumetric and cellular yield. For comparison between shake-flask cultivation and batch cultivation in fermenter, the volumetric yield increased from 272 to 1300 mg L^{-1} (4.8-fold) and cellular yield decreased from 210 to 136 mg g^{-1}. It has been observed that HCDC often achieves less cellular yield than protein expression in shake-flask (Jeong and Lee, 1999). Even though these reported yields here are bulk production, it is significantly higher than other study in terms of *E. coli* derived-HBc expression. Despite of the different in nature of epitope and insertion site of HBc, in the study of Zhang et al. (2020), the yield of purified HBc183-MAGE is recorded at 43 mg L^{-1}. Moreover, in the study of Yap et al. (2009), HBc is attached to His-tag and octapeptide of β-galactosidase at N-terminus, to form His-HBcAg and His-β-HBcAG, which had volumetric yield of 20 mg L^{-1} and 70 mg L^{-1}, respectively.

It is suggested that the nutrient was not enough for protein production when culture reached to OD_{600} of 20. To overcome this limitation, fed-batch mode is used to supply the essential nutrient. However, feeding strategy and the choice of carbon source are important to boost the production rate since overfeeding can lead to the accumulation of acetate or high flux of carbon through glycolysis (Holms, 1986). As glycerol has been proved to be more effective than glucose-based cultivation (Pflug et al., 2007), supplementary media consists of glycerol, yeast extract and tryptone is introduced to the cultures induced at OD_{600} of 20 based on constant feeding and DO-stat feeding. The advantages of DO-stat feeding strategy include the ability to scale up, particularly for aerobic cultivation, since the configuration of process parameters did not change much to maintain dissolved oxygen level, compare to other feeding approaches where process parameters were controlled in cascade mode (Farrell et al., 2015). DO-stat feeding has shown its effectiveness in production of soluble β-galactosidase by de Andrade et al. (2019).

It is observed that the growth curves, feeding amount and feeding profile were similar for the two feeding strategies, which questions why the yield obtained from constant feeding at 0.7 mL min^{-1} was 2.6-fold higher than the yield obtained from DO-stat feeding strategy. This negative effect can be explained by the carbon limitation between feeding pulses as the pump control is set as "on" and "off". Kaleta et al. (2013) reports that carbon source is required to produce protein and direct the ATP generation. Carbon limitation leads to the change in the metabolism pathway of *E. coli*, eventually decreasing the productivity. This result suggests that DO-stat feeding strategy is not an effective method for EBNA1-HBc production if the pump system is set as on/off mechanism. This strategy can be used to boost the biomass to high cell density culture but not for protein production. Even though DO-stat feeding is not suitable for EBNA1-HBc production, the feeding profiles and amount of feed are useful to develop an alternative feeding approach, constant feeding. By this approach, the carbon source is supply continuously at a rate that prevents overflow metabolism.

It is remarkable that production rate achieved in gram-per-litter scale. However, in comparison between batch induced at OD_{600} of 5 and fed-batch induced at OD_{600} of 20 with constant feeding, the volumetric yield only increased from 1300 to 1800 mg L^{-1}. It raises the question in production cost. As is shown in Table 5-3, EBNA1-HBc cost per media (AU$ per gram of protein) was evaluated for batch induced at OD_{600} of 5 and the fed-batch induced at OD_{600} of 20 with constant feeding, which is $10.8 and $9.7, respectively. By using fed-batch mode, EBNA1-HBc cost per media was less than 10 %.

While the constant feeding has been shown to be effective to produce EBNA1-HBc in *E. coli*. The induction at higher cell density (OD_{600} of 40 to 100) should be determined in future development to have a complete induction profile for expression of HBc VLP in *E. coli*.

5.5 Chapter summary and conclusions

The following important factors emerge from this chapter:

1. TB and SB media were more advanced, compare with other complex media such as LB and 2YT media. TB media is the most advantageous media for soluble production of EBNA1-HBc in 6 h.

2. EBNA1-HBc VLP was expressed as soluble intracellular protein in both batch and fed-batch operation. The solubility of the expression is > 90 %. The effect of cell density at induction is consistent between shake-flask and fermenter-scale cultivation. The productivity of the operation decreases as the cell culture is induced at higher cell density.

3. At OD_{600} of 20, the cultivation is hypothesised to be in carbon limitation. In comparison with batch cultivations induced at OD_{600} of 20, fed-batch cultivation with DO-stat feeding improves the volumetric yield 1.7-fold and fed-batch cultivation with constant feeding boosts the volumetric yield to 4.3-fold. The greatest yield achieved was 1800 mg L^{-1}.

4. In terms of media cost per protein (AUD$ per gram of protein), fed-batch cultivation with constant feeding saves 10 %, compared with batch cultivation induced at OD_{600} of 5.

In the next chapter Chapter 6, the conclusions arising from this research and the future developments that can be made to boost the soluble production of *E. coli*-derived HBc VLP from shake-flask to fermenter-scale cultivation are presented.

CHAPTER 6

CONCLUSIONS AND FUTURE DEVELOPMENT

6.1 Conclusions

The main aim of this research is to apply statistical approach to determine the effects of process factors, eventually boosting the soluble production of chimeric HBc-VLP in microbial system. It includes the screening and optimisation of two (2) protein models in small-scale fermentation: 1) HBc-VLP carrying epitope HCV, and 2) HBc-VLP carrying epitope EBNA1. Cultivation in fermenter-scale was also optimised based on the findings from small-scale (shake-flask cultivation).

The key findings of this thesis are:

1. A statistical approach offers a platform to identify the most important factors and interactions. Results permit optimisation. This approach is more advanced than one-variable-at-a time (OFAT) methods. The results achieved from statistical approach can be applied into fermenter-scale cultivation and to recombinant protein bioprocessing.

2. Soluble production of HCV- HBc was improved using fractional factorial design (FFD). The results from FFD ranked the most important process factors in order: 1) cell density at induction, 2) post-induction rotation speed of shaker, and; 3) post-induction temperature. The greatest soluble volumetric and cellular yields achieved from FFD were 84.4 mg L^{-1} of culture media and 89.7 mg g^{-1} DCW. It was about 65 to 70, % of total HCV-HBc expressed. This outcome was obtained from expression conditions of 0.8 mM of IPTG, 30 ^{o}C, 250 rpm and induction point at OD_{600} of 0.8. The pH and presence of Mg^{2+} ion did not show statistically significant impacts on soluble production of HCV-HBc.

3. Fractional factorial design was performed to identify significant process factors. Response surface methodology (RSM) with central composite design (CCD) was conducted to optimise the soluble production of EBNA1-HBc. The important process factors were identified, they are 1) cell density at induction, 2) post-induction rotation speed of shaker, and; 3) post-induction temperature. From the "point optimisation" feature of RSM, the optimal combination of post-induction process factors are 33 ^{o}C, 235 rpm, and induction point of OD_{600} of 1.2 with 0.5 mM of IPTG. The yields were measured at 272.0 mg L^{-1} of culture media and 210.5 mg g^{-1} DCW.

4. Terrific Broth was found to be the most advantageous media for soluble production of EBNA1-HBc in *E. coli*. Induction time of 6 h had greater soluble yield than induction time of 9 h.

5. Because of the differences in cell density, to confirm the effects of process factors and their optimal combination, fermenter-scale cultivations were conducted. Findings demonstrate that the influence of cell density at induction on soluble production of EBNA1-HBc in batch operation is consistent between shake-flask and fermenter-scale. The soluble yields decreased as the culture was induced at greater cell density. Induction at OD_{600} of 5 achieved greatest soluble yield, 1290 mg L^{-1} of culture media.

6. Two (2) fed-batch operations were conducted following DO-stat-based feeding and constant feeding in which it was observed that DO-stat feeding improved volumetric soluble production of culture induced at OD_{600} of 20 1.7 x times, whilst constant feeding boosted the volumetric productivity 4.3 x times. It achieved gram-per-litre production scale, particularly 1800 mg L^{-1} of culture media However, the reason why DO-stat is not effective as constant feeding is not presently understood. It is hypothesised that carbon limitation between feeding pulses affects the metabolism pathway of how cells utilize their energy and convert it into biomass or metabolic product. Further investigation is required to confirm this hypothesis; however, this is beyond the scope of the present research.

7. Significantly most of the EBNA1-HBc in fermenter-scale cultivation is expressed as soluble intracellular protein. The solubility of expression is significantly high, greater than 90 %.

8. Fed-batch cultivation with constant feeding saved 10 % in media cost compared with batch cultivation that induced at OD_{600} of 5.

9. It was found that clarified crude lysate containing EBNA1-HBc can be treated with heat at 70 ^{o}C to remove host cell protein (impurities). Heat precipitation and size exclusion chromatography offer an alternative approach to quantitatively measure the fully assembled HBc-VLP in clarified crude lysate.

10. Although full-length HBc-VLP is a versatile vaccine scaffold, the choice of epitope should be carefully considered to achieve productive soluble expression. Considering the

insertion at N-terminal of the HBc protein, it is shown that EBNA1-HBc had greater yield than HCV-HBc in terms of soluble expression. In comparison with HBc carrier (without epitope), HCV-HBc VLP has less soluble yield than HBc VLP in the same expression condition. It is hypothesised that the structure of epitope has influence on the formation of HBc VLP in *E. coli* because HCV epitope has a long amino acid sequence (44 aa) and α-helix structure and EBNA1 epitope has 11 aa and random coils. However, there is insufficient data to conclude this hypothesis, the investigation should be expanded to the insertion at MIR region and C-terminal.

It is concluded that the statistical approach offers an advanced platform to optimise process factors in microbial expression system in small scale fermentation. FFD coupled with RSM is generalizable to a range of recombinant protein bioprocesses, not just only HBc-VLP. It could provide a new process design tool to boost the production of vaccine candidate at both synthesis and analysis stages.

Findings from this research work promote the use of microbial system to synthesise HBc vaccine candidate. These are going to aid a detailed understanding of fermentation process factors that contribute to the soluble production of HBc-VLP.

This research work is original. Results obtained from this research work will be beneficial to recombinant protein bioprocessing industry.

6.1.1 Limitation of this work

There are present limitations of this research that need to be determined in future work, namely 1) quantitative variation of complex media between batches, 2) an explanation of cell metabolic activity to answer why particular setting is the optimal combination of factors, and; 3) the reason why carbon limitation from DO-stat-feeding limits the production.

6.2 Recommendations for future research

Importantly, the success of this research shows that this novel statistical approach can be applied to boost the soluble production of recombinant protein. It is recommended that statistical approach should be used to enhance the quality of design by the International Conference on Harmonization (ICH) (2009) for pharmaceutical development.

Because the soluble production of EBNA1-HBc has been significantly improved, it is essential to investigate the expression of HBc-VLP at greater cell density (OD_{600} of 40 to 100) to complete the induction profile. Instead of using yeast extract, a combination glucose–glycerol mixture and glutamate could minimise the variation in yeast extract and improve protein production (Chiang et al., 2020). A study on how carbon limitation affects protein production in DO-stat feeding is needed in future development.

Based on the experimental results, the native structure of inserted epitope influences the soluble expression yield of HBc VLP, particularly, epitope that has secondary structure (HCV) achieved less yield than simple epitope (EBNA1). Because the influence of epitope on the formation of HBc VLP in *E. coli* is not well-studied, it is recommended to use Molecular Dynamic simulation software to identify how the folding of epitope influence on HBc VLP formation.

It is suggested that the empirical equation from RSM has potential to be coupled with process simulation software e.g. SuperPro Designer (Intelligen, Inc) to produce a more powerful design and assessment tool for the biopharmaceutical industries.

APPENDIX A – A definition of some important terms used in this research

Virus-like particle	Virus-like particle is a multi-protein complex, which assembles into native structure of viral capsid without viral genome; VLP forms a high-density of repetitive surface structure (spikes) (Fuenmayor et al., 2017).
Hepatitis B core protein	The structural of Hepatitis B virus monomeric protein, consists of 183 to 185 amino acids, based on the subtypes, *ayw* and *adw* (Pumpens and Grens, 1999).
Recombinant protein expression	The production of biological pharmaceutical product or specific protein i.e. vaccine, insulin. A manipulated gene is constructed and transferred into an expression host cell. Cultivation of transformed cell is conducted to obtain the protein of interest.
Soluble intracellular protein expression	The recombinant proteins are expressed and folded in correct structure in cytoplasm. After cell disruption, the target protein is detected in soluble cytoplasmic fraction (Sorensen and Mortensen, 2005).
Proto-aggregate (soluble aggregate)	Newly synthesised protein folding intermediates fails to achieve the native folding, they will be degraded or form into small proto-aggregates (soluble aggregates) (Schrodel, and de Marco, 2005)
Inclusion bodies	The protein of interest is expressed as insoluble aggregates because of mis-folding. These aggregates are often biologically inactive (Markossian and Kurganov, 2004).
Cell lysis/ cell disruption	An operation in which cell pellet is suspended in lysis buffer. Cells is lysed using ultra sonication or high-pressure homogeniser. Crude lysate is obtained after operation.
Crude lysate	A protein solution that is obtained after cell disruption process. It contains cell debris, soluble host cell protein, soluble target protein, soluble aggregates and inclusion bodies.
Clarification	An operation that is used to separate soluble protein, inclusion bodies and cell debris by centrifugation.
Clarified crude lysate	The supernatant fraction of protein sample achieved from clarification. It contains soluble host cell protein, soluble target protein, soluble aggregates.

Fermentation/ cultivation	The upstream process of recombinant protein production. It includes two steps: 1) generation of biomass and 2) induction. It can be operated in small-scale (shake-flask) or large-scale (fermenter).
Fermentation process factors	Process factors that belong to cultivation i.e. temperature, pH, dissolved oxygen level, etc.
Induction	Cells carry the plasmid of interest to synthesise target protein. The step of adding inducer i.e. IPTG or lactose to activate the expression.
Feeding strategy	The operation that is used to supply extra nutrient to the culture, switching from batch mode to fed-batch mode. It includes pH-stat, DO-stat, constant feeding, linear, etc.
Optical cell density	A unit that corresponds to the cell population. Optical cell density is measured by spectrometer at wavelength 600 nm. It is used as a correlation to dry cell mass. It is found in this study that biomass concentration equal to 0.4 times optical density.
One factor at a time (OFAT)	A conventional research method, also known as, one variable at a time. A set of experiment is conducted by changing the levels of a factor while other factors and their level are unchanged (Frey et al., 2013).
Statistical approach	A novel method that is used to screening and optimisation of process variables. This approach is based on exploration of the relationships between variables (categorical and continuous) and experimental response to identify the factors that significantly influenced the response and predict the optimal combination of factors (Snedecor and Cochran, 1989).
Fractional factorial design (FFD)	A design that is used for screening purpose. It is considered as a subset of factorial design, usually half or quarter of the full factorial design. It is suitable when the variable larger than five (5). It can fit the experimental data to first-order equation (Montgomery, 2012)
Response surface methodology (RSM)	A design that is used for optimisation purpose. The objective can be single or multiple responses. It is often used for continuous factors and there are two main types, central composite design and Box–Behnken design. (Montgomery, 2012)
Point optimisation	A feature from response surface model in which the optimal combination of process factors is obtained using desirability function or graphical method.

APPENDIX B – Buffer composition in-use for Chapters 3, 4 and 5

Cell lysis buffer	A buffer that is prepared for cell disruption operation. Cell pellet is suspended in this lysis at desired ratio. Composition: i. 20 mM of Tris base ii. 3 mM of EDTA iii. 0.1 % (v/v) Triton X-100 iv. pH 8.0, adjusted by HCl
Size exclusion chromatography mobile phase	A buffer that is used as mobile phase for size exclusion chromatography, particularly, TSK-G5000 GW_{XL} column. i. For HCV-HBc: 50 mM phosphate buffer at pH 7.4 ii. For EBNA1-HBc: 50 mM Tris-HCl buffer at pH 7.4
SDS-PAGE sample loading buffer (5X)	A buffer that is mixed with protein sample at ratio 4:1. The prepared sample is heated on heat block before place in gel. Composition: i. 0.25 M Tris-base ii. 0.25 % (w/v) Bromophenol blue iii. 0.5 M dithiothreitol (DTT) iv. 50 % (v/v) Glycerol v. 10 % (v/v) sodium dodecyl sulfate (SDS) vi. pH 6.8, adjusted by HCl
SDS-PAGE running buffer (10X)	A buffer that is used as electrophoresis buffer. Buffer is diluted to 1X when needed. Composition in 1000 mL of water: i. 30.0 g of Tris base ii. 144.0 g of glycine iii. 10.0 g of SDS iv. pH 8.3
Coomassie Blue staining solution	A solution that is used to dye the gel after gel electrophoresis. Gel is often stained for 30 min. Composition: i. 0.1 % (w/v) Coomassie blue R350 ii. 20 % (v/v) methanol iii. 10 % (v/v) acetic acid
De-staining solution	A solution that is used to remove dye from the gel. After de-staining, protein band on the gel will appear while the other area of the gel is transparent. Composition in 1000 mL of water: i. 10 % (v/v) acetic acid ii. 10 % (v/v) ethanol

TAE (10X)	A buffer that is used in native agarose gel electrophoresis (NAGE). TAE stands for Tris – Acetate – EDTA. Solution is diluted to 1X when needed Composition in 1000 mL of TAE 10X: i. 48.5 g of Tris base ii. 11.4 mL glacial acetic acid iii. 20 mL of 0.5 M EDTA at pH 8.0
NAGE sample loading buffer (5X)	A buffer that is mixed with protein sample at ratio 4:1. The prepared sample is placed at 25 ^{o}C for 20 min before it is placed in gel. Composition: i. 0.25 M Tris-base ii. 0.5 % (w/v) Bromophenol blue iii. 50 % (v/v) Glycerol iv. pH 6.8, adjusted by HCl
Urea suspension buffer	A buffer that is used to solubilise the protein precipitate from ammonium sulfate precipitation. Composition: i. 20 mM Glycine-NaOH ii. 4 M Urea iii. pH 9.0, adjusted by NaOH
Dialysis buffer	A buffer that is used for buffer exchange. Composition: i. 20 mM of Tris base ii. pH 7.4, adjusted by HCl

APPENDIX C – Supporting data

Fig. C-1 presents the calibration curve that was used in protein concentration assay. Bovine serum albumin with the concentration of 0.1, 0.25, 0.5, 0.75, mg mL^{-1}, was used to plot the curve. Linear trend was fitted to generate the equation: y = 1.13x + 0.11, with R^2 = 0.99; where y is the protein concentration (mg mL^{-1}), and x is the absorbance from spectrometer (AU). In case of the measured absorbance is out of range, protein sample is diluted.

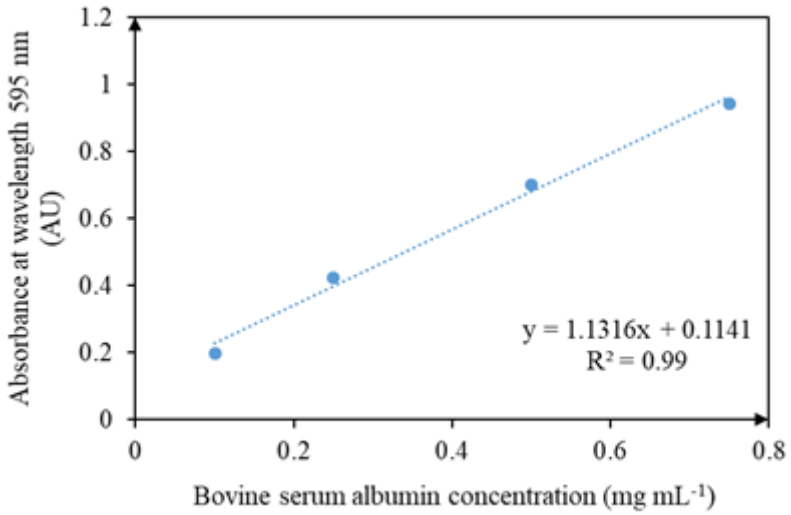

Fig C-1: The calibration curve for protein concentration assay using Bradford reagent.

Fig C-2 presents the calibration curve that is used to convert optical cell density to dry cell concentration. Sample that had optical cell density of 5 was diluted into samples that had OD_{600} of 1, 1.5, 2.5, 3.5 and 4.5. These sample was heated for 24 h at 60 ^{O}C to remove water. The dry cell was weighted on analytical scale. Linear trend was fitted to generate the equation: $y = 0.40x$, with $R^2 = 0.99$; where y is the dry cell concentration (g L^{-1}), and x is the optical cell density measured from spectrometer at wavelength 600 nm (OD_{600}).

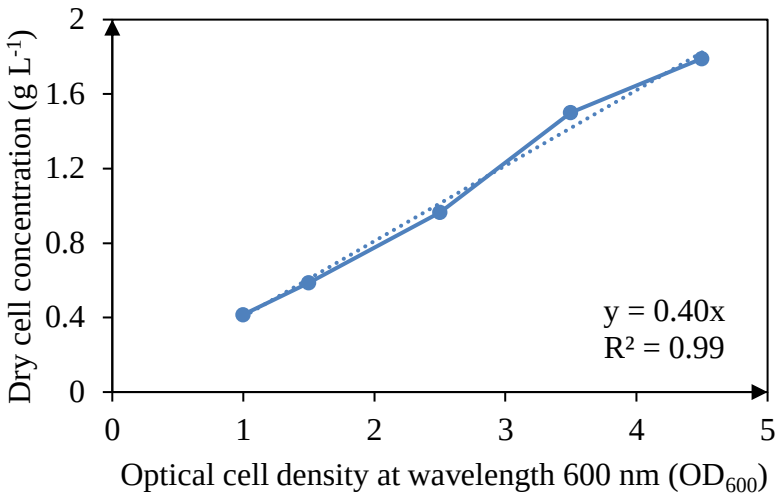

Fig. C-2: The correlation curve between optical cell density and dry cell concentration.

The price of media ingredient, IPTG inducer, and antibiotic are listed in Table C-1. Price was obtained from supplier websites (last accessed on Oct. 30, 2020, 10.00 h), they are:

1. Chem Supply Australia: https://www.chemsupply.com.au
2. Oxoid Thermo Fisher: https://www.thermofisher.com
3. Gibco Thermo Fisher: https://www.thermofisher.com
4. Invitrogen Thermo Fisher: https://www.thermofisher.com.

Price of each items are converted to AUD$ per g or AUD$ per mL^{-1} to calculate the media expense in Table 5-3. The expense of media (AUD$) and media cost per protein (AUD$ g^{-1}) are calculated following Equation C-1 and Equation C-2.

$$\text{The expense of media (AUDS)} = (X_O \times Y_O + X_{\text{additional}} \times Y_{\text{additional}}) \times Z \qquad \textbf{(Eq. C-1)}$$

where X_o is the initial medium volume (L), $X_{\text{additional}}$ is the additional medium volume (L), Y_O is the initial media ingredient concentration (g L^{-1}), $Y_{\text{additional}}$ (g L^{-1}) is the additional media ingredient concentration (g L^{-1}), and Z is the cost per unit (AUD$ g^{-1}).

$$\text{The media cost of protein (AUD\$ } g^{-1}) = \frac{\text{The expense of media (AUD\$)}}{\text{Protein mass (g)}} \qquad \textbf{(Eq. C-2)}$$

Table C-1: Media ingredients price.

	Supplier	**AUD$ g^{-1}**
Glycerol	Chem Supply	0.033
Yeast extract	Oxoid Thermo Fisher	0.164
Tryptone	Oxoid Thermo Fisher	0.238
K_2HPO_4	Chem Supply	0.122
KH_2PO_4	Chem Supply	0.092
Kanamycin sulfate	Gibco Thermo Fisher	11.96
IPTG	Invitrogen Thermo Fisher	39.80
	Supplier	**AUD$ mL^{-1}**
H_3PO_4 (21 % w/v)	Chem Supply	0.0093
NH_4OH (12.5 % w/v)	Chem Supply	0.0154

The cost of ingredients was checked on the supplier website on 30th Oct 2020

REFERENCES

Abou-Taleb, K., Galal, G., 2018. A comparative study between one-factor-at-a-time and minimum runs resolution-IV methods for enhancing the production of polysaccharide by *Stenotrophomonas daejeonensis* and *Pseudomonas geniculate*. Annals of Agricultural Sciences, 63 (2), 173–180.

Afzal, M., Minor, P., 2002. Vaccines, Crohn's disease and autism. Molecular Psychiatry, 7, S49–S50.

Alqueres, S.M.C., Branco, R.V., Freire, D.M.G., Alves, T.L.M., Martins, O.B., Almeida, R.V., 2011. Characterization of the recombinant thermostable lipase (Pf2001) from *Pyrococcus furiosus*: Effects of thioredoxin fusion tag and Triton X-100. Enzyme Research, 2011, 1–7.

Altmann, M., Pich, D., Ruiss, R., Wang, J., Sugden, B., Hammerschmidt, W., 2006. Transcriptional activation by EBV Nuclear Antigen 1 is essential for the expression of EBV's transforming genes. Proceedings of the National Academy of Sciences, 103 (38), 14188–14193.

Alvarez-Lajonchere, L., Duenas-Carrera, S., 2012. Complete definition of immunological correlates of protection and clearance of Hepatitis C Virus infection: A relevant pending task for vaccine development. International Reviews of Immunology, 31, 223 - 242.

Amorij, J., Huckriede, A., Wilschut, J.H., Frijlink, W., Hinrichs, W.L.J., 2008. Development of stable influenza vaccine powder formulations: Challenges and Possibilities. Pharmaceutical Research, 25, 1256–1273.

Anraku, Y., 1988. Bacterial Electron Transport Chains. Annual Review of Biochemistry, 57 (1), 101–132.

Arora, U., Tyagi, P., Swaminathan, S., Khanna, N., 2012. Chimeric Hepatitis B core antigen virus-like particles displaying the envelope domain III of dengue virus type 2. Journal of Nanobiotechnology, 10 (30).

Aston-Deaville, S., Carlsson, E., Saleem, M., Thistlethwaite, A., Chan, H., Maharjan, S., Facchetti, A., Feavers, I.M., Alistair Siebert, C., Collins, R.F., Roseman, A., Derrick, J.P., 2020. An assessment of the use of Hepatitis B Virus core protein virus-like particles to display heterologous antigens from Neisseria meningitidis. Vaccine, 38 (16), 3201–3209.

Bachmann, M.F., Jennings, G.T., 2010. Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns. Nature Reviews Immunology, 10, 787–796.

Balderas, H.V.E., Paz Maldonado, L.M.T., Rivero, E.M., Barba de la Rosa, A.P., Ordonez A.L.G., Rodriguez, A.D.L., 2008. Optimization of human interferon gamma production in *Escherichia coli* by response surface methodology. Biotechnology and Bioprocess Engineering, 13, 7–13.

Baneyx, F., Mujacic. M., 2004. Recombinant protein folding and misfolding in *Escherichia coli*. Nature Biotechnology, 22 (11), 1399–1408.

Beigi, L., Karbalaei-Heidari, H.R., Kharrati-Kopaei, M., 2012. Optimization of an extracellular zinc-metalloprotease (SVP2) expression in *Escherichia coli* BL21 (DE3) using response surface methodology. Protein Expression and Purification, 84 (1), 161–166.

Bellier, B., Klatzmann, D., 2013. Virus-like particle-based vaccines against hepatitis C virus infection. Expert Review of Vaccines, 12 (2), 143–154.

Bettenbrock, K., Bai, H., Ederer, M., Green, J., Hellingwerf, K., Holcombe, M., Kunz, S., Rolfe, M., Sanguinetti, G., Sawodny, O., Sharma, P., Steinsiek, S., Poole, R., 2014. Towards a systems level understanding of the oxygen response of *Escherichia coli*, Advances in Microbial Systems Biology, 64, 65–114.

Bezerra, M., Santelli, R., Oliveira, E., Villar, L., Escaleira, L., 2008. Response surface methodology (RSM) as a tool for optimization in analytical chemistry. Talanta, 76 (5), 965–977.

Birnbaum, F., Nassal, M., 1990. Hepatitis B virus nucleocapsid assembly: Primary structure requirements in the core protein. Journal of Virology, 64, 3319–3330.

Blake, N., Haigh, T., Shaka'a, G., Croom-Carter, D., Rickinson, A., 2000. The importance of exogenous antigen in priming the human CD8+ T cell response: Lessons from the EBV Nuclear Antigen EBNA1. The Journal of Immunology, 165 (12), 7078-7087.

Blake, N., Lee, S., Redchenko, I., Thomas, W., Steven, N., Leese, A., Steigerwald-Mullen, P., Kurilla, M.G., Frappier, L., Rickinson, A., 1997. Human CD8+ T cell responses to EBV EBNA1: HLA class I presentation of the (Gly-Ala)-containing protein requires exogenous processing. Immunity (Cambridge, Mass.), 7 (6), 791–802.

Borisova, G.P., Kalis, I., Pushko, P.M., Tsibinogin, V.V., Loseva, V., 1988. Genetically engineered mutants of the core antigen of the human hepatitis B virus preserving the ability for native self-assembly. Doklady Akademii nauk SSSR, 298 (6), 1474–1478.

Boulant, S., Vanbelle, C., Ebel, C., Penin, F., Lavergne, J.P., 2005. Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features. Journal of Virology, 79 (17), 11353–11365.

Box, G., Behnken, D., 1960. Some new three level designs for the study of quantitative variables. Technometrics, 2 (4), 455-475.

Box, G., Cox, D., 1964. An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological), 26, 211-252.

Box, G., Hunter, J., Hunter, W., 2005. Statistics for experimenters: design, innovation, and discovery, second ed. Wiley-Interscience, Hoboken, N.J. ISBN: 978-0-471-71813-0.

Box, G., Wilson, K., 1951. On the experimental attainment of optimum conditions, Journal of the Royal Statistical Society. Series B (Methodological), 13 (1), 1-45.

Bradford, M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry, 72 (1-2), 248–254.

Buonaguro, L., Tagliamonte, M., Tornesello, M.L., Buonaguro, F.M., 2011. Developments in virus-like particle-based vaccines for infectious diseases and cancer. Expert Review of Vaccines, 10 (11), 1569–1583.

Burgess, R., 2009. Refolding Solubilized Inclusion Body Proteins. In Burgess, R., Deutscher, M., (Eds). Methods in Enzymolog, Academic Press, 463, 259–282, ISBN 9780123745361

Calvaruso, V., Petta, S., Craxi, A., 2018. Is global elimination of HCV realistic. Liver international: Official Journal of the International Association for the Study of the Liver, 38 Suppl 1, 40–46.

Cardoso, V.M., Campani, G., Santos, M.P., Silva, G.G., Pires, M.C., Goncalves, V.M., de C. Giordano, R, Sargo, C.R., Horta, A.C.L., Zangirolami, T.C., 2020. Cost analysis based on bioreactor cultivation conditions: Production of a soluble recombinant protein using *Escherichia coli BL21*(DE3). Biotechnology Reports (Amsterdam, Netherlands), 26, e00441–e00441.

Carrio, M., Villaverde, A., 2002. Review of construction and deconstruction of bacterial inclusion bodies. Journal of Biotechnology, 96 (1), 3–12.

Caselmann, W.H., Alt, M., 1996. Hepatitis C virus infection as a major risk factor for hepatocellular carcinoma. Journal of Hepatology, 24 (2 Suppl), 61–66.

Chain, B.M., Myers, R., 2005. Variability and conservation in hepatitis B virus core protein. BMC Microbiology, 5, 33.

Chen, R, 2012. Bacterial expression systems for recombinant protein production: *E. coli* and beyond. Biotechnology Advances, 30, 1102-1107.

Chen, S., Zheng, D., Li, C., Zhang, W., Xu, W., Liu, X., Fang, F., Chen, Z., 2015. Protection against multiple subtypes of influenza viruses by virus-like particle vaccines based on a hemagglutinin conserved epitope. BioMed Research International, 901817.

Chiang, C., Hu, M., Chao, Y., 2020. A strategy to improve production of recombinant proteins in *Escherichia coli* based on a glucose–glycerol mixture and glutamate. Journal of Agricultural and Food Chemistry, 68 (33), 8883–8889.

Chien, D.Y., Arcangel, P., Medina-Selby, A., Coit, D., Baumeister, M., Nguyen, S., George-Nascimento, C., Gyenes, A., Kuo, G., Valenzuela, P., 1999. Use of a novel Hepatitis C Virus (HCV) major-epitope chimeric polypeptide for diagnosis of HCV infection. Journal of Clinical Microbiology, 37 (5), 1393–1397.

Choi, J.H., Keum, K.C., Lee, S.Y., 2006. Production of recombinant proteins by high cell density culture of *Escherichia coli*. Chemical Engineering Science, 61, 876–885.

Choo, Q.L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W., Houghton, M., 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science (New York, N.Y.), 244 (4902), 359–362.

Christensen, D., Orr, J., Rao, C., Wolfe, A. 2017. Increasing growth yield and decreasing acetylation in *Escherichia coli* by optimizing the carbon-to-magnesium ratio in peptide-based media. Applied and Environmental Microbiology, 83 (6), AEM.03034-16.

Chu, X., Li, Y., Long, Q., Xia, Y., Yao, Y., Sun, W., Huang, W., Yang, X., Liu, C., Ma, Y. 2016. Chimeric HBcAg virus-like particles presenting a HPV 16 E7 epitope significantly suppressed tumor progression through preventive or therapeutic immunization in a TC-1-grafted mouse model. International Journal of Nanomedicine, 11, 2417-2429.

Chua, L.H., Tan, S.C., Liew, M.W., 2018. Process intensification of core streptavidin production through high-cell-density cultivation of recombinant *E. coli* and a temperature-based refolding method. Journal of Biotechnology, 276-277, 34–41.

Chuan, Y., Lua, L., Middelberg, A., 2008. High-level expression of soluble viral structural protein in *Escherichia coli*. Journal of Biotechnology, 134, 64-71.

Chuan, Y., Wibowo, N., Lua, L., Middelberg, A., 2014. The economics of virus-like particle and capsomere vaccines. Biochemical Engineering Journal, 90, 255-263.

Clark, D.P., 1989. The fermentation pathways of *Escherichia coli*. FEMS Microbiology Reviews, 5, 223–234.

Clarke, B.E., Newton, S.E., Carroll, A.R., Francis, M.J., Appleyard, G., Syred, A.D., Highfield, P.E., Rowlands, D.J., Brown, F., 1987. Improved immunogenicity of a peptide epitope after fusion to hepatitis B core protein. Nature (London), 330 (6146), 381–384.

Cohen, B.J., Richmond, J.E., 1982. Electron microscopy of hepatitis B core antigen synthesized in *E. coli*. Nature (London), 296 (5858), 677–678.

Cohen, J.I., 2018. Vaccine development for Epstein-Barr virus. Human Herpesviruses, 1045, 477–493.

Cohen, J.I., Fauci, A.S., Varmus, H., Nabel, G.J., 2011. Epstein-Barr Virus: An important vaccine target for cancer prevention. Science Translational Medicine, 3 (107), 107fs7.

Contiero, J., Beatty, C.M., Kumari, S., DeSanti, C.L., Strohl, W.R., Wolfe, A.J., 2000. Effects of mutations in acetate metabolism in high-cell-density growth of *Escherichia coli*. Journal of Industrial Microbiology and Biotechnology, 24, 421-430.

Cox, M.M., Izikson, R., Post, P., Dunkle, L., 2015. Safety, efficacy, and immunogenicity of Flublok in the prevention of seasonal influenza in adults. Therapeutic Advances in Vaccines, 3 (4), 97–108.

Crowther, R., Kiselev, N., Bottcher, B., Berriman, J., Borisova, G., Ose, V., Pumpens, P., 1994. Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. Cell, 77, 943-950.

D'Aoust, M., Lavoie, P., Couture, M., Trepanier, S., Guay, J., Dargis, M., Mongrand, S., Landry, N., Ward, B., Vezina, L., 2008. Influenza virus-like particles produced by transient expression in *Nicotiana benthamiana* induce a protective immune response against a lethal viral challenge in mice. Plant Biotechnology Journal, 6, 930-940.

Dahari, H., Feinstone, S.M., Major, M.E., 2010. Meta-analysis of Hepatitis C Virus vaccine efficacy in chimpanzees indicates an importance for structural proteins. Gastroenterology New York, N.Y. 1943, 139 (3), 965–974.

De Andrade, B.C., Migliavacca, V.F., Okano, F.Y., Grafulin, V.Y., Lunardi, J., Roth, G., de Souza, C.F.V., Santos, D.S., Chies, J.M., Renard, G., Volpato, G., 2018. Production of recombinant β -galactosidase in bioreactors by fed-batch culture using DO-stat and linear control. Biocatalysis and Biotransformation, 37 (1), 3–9.

De Boer, H.A., Comstock, L.J., Vasser, M. 1983. The tac promoter: a functional hybrid derived from the trp and lac promoters. Proceedings of the National Academy of Sciences of the United States of America, 80 (1), 21–25.

De Groot, N.S., Ventura, S., 2006. Effect of temperature on protein quality in bacterial inclusion bodies. FEBS Letters, 580 (27), 6471–6476.

Derringer, G., Suich, R., 1980. Simultaneous optimization of several response variables. Journal of Quality Technology, 12 (4), 214-219.

Deuschle, U., Kammerer, W., Gentz, R., Bujard, H., 1986. Promoters of *Escherichia coli*: a hierarchy of *in vivo* strength indicates alternate structures. The EMBO Journal, 5 (11), 2987–2994.

Dishlers, A., Skrastina, D., Renhofa, R., Petrovskis, I., Ose, V., Lieknina, I., Jansons, J., Pumpens, P., Sominskaya, I., 2015. The Hepatitis B Virus core variants that expose foreign C-Terminal insertions on the outer surface of virus-like particles. Molecular Biotechnology, 57, 1038-1049.

Eiteman, M.A., Altman, E., 2006. Overcoming acetate in *Escherichia coli* recombinant protein fermentations. Trends in Biotechnology, 24, 530-536.

Epstein, M.A., 1976. Epstein-Barr virus--is it time to develop a vaccine program. Journal of the National Cancer Institute, 56 (4), 697–700.

Fahnert, B., Lilie, H., Neubauer, P., 2004. Inclusion bodies: Formation and utilisation, physiological stress responses in bioprocesses. Advances in Biochemical Engineering, Biotechnology, 89, 93-142.

Farrell, P., Sun, J., Champagne, P., Lau, H., Gao, M., Sun, H., Zeiser, A., D'amore, T., 2015. The use of dissolved oxygen-controlled, fed-batch aerobic cultivation for recombinant protein subunit vaccine manufacturing. Vaccine, 33 (48), 6752–6756.

Farrell, P., Sun, J., Gao, M., Sun, H., Pattara, B., Zeiser, A., D'amore, T., 2012. Development of a scaled-down aerobic fermentation model for scale-up in recombinant protein vaccine manufacturing. Vaccine, 30 (38), 5695–5698.

Farrell, P.J. 2019. Epstein-Barr virus and cancer. Annual Review of Pathology, 14 (1), 29–53.

Faulkner, E., Barrett, M., Okor, S., Kieran, P., Casey, E., Paradisi, F., Engel, P., Glennon, B, 2006. Use of fed-batch cultivation for achieving high cell densities for the pilot-scale production of a recombinant protein (*Phenylalanine Dehydrogenase*) in *Escherichia coli.* Biotechnology Progress, 22 (3), 889–897.

Fenner, F., Henderon, D.A., Arita, I., Jezek, Z., Ladnyi, I.D., 1988. Smallpox and its eradication. World Health Organization, Geneva.

Fernandez-Castane, A., Vine, C. E., Caminal, G., Lopez-Santin, J., 2012. Evidencing the role of lactose permease in IPTG uptake by *Escherichia coli* in fed-batch high cell density cultures. Journal of Biotechnology, 157 (3), 391–398.

Ferreira, S.L., Bruns, R.E., da Silva, E.G., Dos Santos, W.N., Quintella, C.M., David, J.M., de Andrade, J.B., Breitkreitz, M.C., Jardim, I.C., Neto, B.B., 2007. Statistical designs and response surface techniques for the optimization of chromatographic systems. Journal of chromatography. A, 1158 (1-2), 2–14.

Ferrer, M., Chernikova, T.N., Yakimov, M.M., Golyshin, P.N., Timmis, K.N., 2003. Chaperonins govern growth of *Escherichia coli* at low temperature. Nature Biotechnology, 21, 1266-1267.

Freivalds, J., Dislers, A., Ose, V., Pumpens, P., Tars, K., Kazaks, A., 2011. Highly efficient production of phosphorylated hepatitis B core particles in yeast *Pichia pastoris*. Protein Expression and Purification, 75 (2), 218–224.

Frey, D.D., Engelhardt, F., Greitzer, E.M., 2003. A role for 'one-factor-at-a-time' experimentation in parameter design. Research in Engineering Design, 14 (2), 65–74.

Fried, M.W., Shiffman, M.L., Reddy, K.R., Smith, C., Marinos, G., Goncales, F.L., Haussinger, D., Diago, M., Carosi, G., Dhumeaux, D., Craxi, A., Lin, A., Hoffman, J., Yu, J., 2002. PEGinterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. The New England Journal of Medicine, 347 (13), 975–982.

Fuenmayor, J., Godia, F., Cervera, L., 2017. Production of virus-like particles for vaccines. New Biotechnology, 39, 174-180.

Gasser, B., Saloheimo, M., Rinas, U., Dragosits, M., Rodriguez-Carmona, E., Baumann, K., Giuliani, M., Parrilli, E., Branduardi, P., Lang, C., Porro, D., Ferrer, P., Tutino, M.L., Mattanovich, D., Villaverde, A., 2008. Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. Microbial Cell Factories, 7, 11.

Geldmacher, A., Skrastina. D., Petrovskis, I., Borisova, G., Berriman, J.A., Roseman, A.M, Crowther, R.A., Fischer, J., Musema, S., Gelderblom, H.R., Lundkvist, A., Renhofa, R., Ose, V., Kruger, D.H., Pumpens, P., Ulrich, R., 2004. An amino-terminal segment of Hantavirus nucleocapsid protein presented on hepatitis B virus core particles induces a strong and highly cross-reactive antibody response in mice. Virology, 323 (1), 108-119.

Gillam, F., Zhang, C., 2018. Epitope selection and their placement for increased virus neutralization in a novel vaccination strategy for porcine epidemic diarrhea virus utilizing the Hepatitis B virus core antigen. Vaccine, 36, 4507–4516.

Glasel, J.A, 1995. Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. BioTechniques, 18, 62–63.

Glick, B.R., 1995. Metabolic load and heterologous gene expression. Biotechnology Advances, 13 (2), 247-261.

Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K., Riggs, A.D., 1979. Expression in *Escherichia coli* of

chemically synthesized genes for human insulin. Proceedings of the National Academy of Sciences of the United States of America, 76 (1), 106–110.

Greenwood B., 2014. The contribution of vaccination to global health: past, present and future. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 369 (1645), 20130433.

Gregson, A.L., Oliveira, G., Othoro, C., Calvo-Calle, J.M., Thorton, G.B., Nardin, E., Edelman, R., 2008. Phase I trial of an alhydrogel adjuvanted hepatitis B core virus-like particle containing epitopes of *Plasmodium falciparum* circumsporozoite protein. PloS one, 3 (2), e1556.

Grgacic, E.V., Anderson, D.A., 2006. Virus-like particles: passport to immune recognition. Methods, 40, 60–65.

Grossman, T., Kawasaki, E., Punreddy, S., Osburne, M., 1998. Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. Gene, 209 (1-2), 95–103.

Guan, L., Kaback, H.R., 2006. Lessons from lactose permease. Annual Review of Biophysics and Biomolecular Structure, 35 (1), 67–91.

Guzman, L.M., Belin, D., Carson, M.J., Beckwith, J., 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. Journal of Bacteriology, 177 (14), 4121–4130.

Handley, L.M., Mackey, J.P., Buller, R.M.L., Bellone, C.J., 2007. Orthopoxvirus vaccines and vaccination. Birkhauser, Basel, 329–353.

Hempfling, W.P., Mainzer, S.E., 1975. Effects of varying the carbon source limiting growth on yield and maintenance characteristics of *Escherichia coli* in continuous culture. Journal of Bacteriology, 123 (3), 1076–1087.

Henkel, S.G., Ter Beek, A., Steinsiek, S., Stagge, S., Bettenbrock, K., de Mattos, M.J., Sauter, T., Sawodny, O., Ederer, M., 2014. Basic regulatory principles of *Escherichia coli*'s electron transport chain for varying oxygen conditions. PloS one, 9 (9), e107640.

Hoft, D.F., Lottenbach, K.R., Blazevic, A., Turan, A., Blevins, T.P, Pacatte, T.P., Yu, Y., Mitchell, M.C., Hoft, S.G., Belshe, R.B, 2017. Comparisons of the humoral and cellular immune responses induced by live attenuated influenza vaccine and inactivated influenza vaccine in adults. Clinical and Vaccine Immunology, 24 (1), e00414-16.

Holmes, J., Thompson, A., Bell, S., 2013. Hepatitis C - an update. Australian Family Physician, 42 (7), 452–456.

Holms, W.H., 1986. The central metabolic pathways of *Escherichia coli*: Relationship between flux and control at a branch point, efficiency of conversion to biomass, and excretion of acetate. Current Topics in Cellular Regulation, 28, 69-105.

Horzinek, M.C., 2011. Rinderpest: the second viral disease eradicated. Veterinary Microbiology, 149 (3-4), 295–297.

Hu, J., Wang, F., Liu, C., 2015. Development of an efficient process intensification strategy for enhancing Pfu DNA polymerase production in recombinant *Escherichia coli*. Bioprocess and Biosystems Engineering, 38 (4), 651–659.

Hunke, S., Betton, J.M., 2003. Temperature effect on inclusion body formation and stress response in the periplasm of *Escherichia coli*. Molecular Microbiology, 50, 1579– 1589.

Ibanez, L.I., Roose, K., De Filette, M., Schotsaert, M., De Sloovere, J., Roels, S., Pollard, C., Schepens, B., Grooten, J., Fiers, W., Saelens, X., 2013. M2e-displaying virus-like particles with associated RNA promote T helper 1 type adaptive immunity against influenza A. PloS one, 8 (3), e59081.

International Conference on Harmonization (ICH), 2009. Pharmaceutical development Q8 (R2). IFPMA, Geneva. https://database.ich.org/sites/default/files/Q8_R2_Guideline.pdf [last accessed Dec. 02, 2020, 12.00 h].

Islam, R., Tisi, D., Levy, M., Lye, G., 2007. Framework for the rapid optimization of soluble protein expression in *Escherichia coli* combining microscale experiments and statistical experimental design. Biotechnology Progress, 23 (4), 785–793.

Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heyneker, H.L., Bolivar, F., Boyer, H.W., 1977. Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. Science, N.Y., 198 (4321), 1056–1063.

Jeong, K.J., Lee, S.Y., 1999. High-level production of human leptin by fed-batch cultivation of recombinant *Escherichia coli* and its purification. Applied and Environmental Microbiology, 65, 3027–3032.

Jevsevar, S., Gaberc-Porekar, V., Fonda, I., Podobnik, B., Grdadolnik, J., Menart, V., 2005. Production of nonclassical inclusion bodies from which correctly folded protein can be extracted. Biotechnology Progress, 21, 632-639.

Ji, M., Xie, X., Liu, D., Yu, X., Zhang, Y., Zhang, L., Wang, S., Huang, Y., Liu, R., 2018. Hepatitis B core VLP-based mis-disordered *tau* vaccine elicits strong immune response and alleviates cognitive deficits and neuropathology progression in *Tau*. P301S mouse model of Alzheimer's disease and frontotemporal dementia. Alzheimer's Research & Therapy, 10 (55).

Jia, B., Jeon, C.O., 2016. High-throughput recombinant protein expression in *Escherichia coli*: current status and future perspectives. Open Biology, 6 (8), 10196.

Kaisermayer, C., Reinhart, D., Gili, A., Chang, M., Aberg, P.M., Castan, A., Kunert, R., 2016. Biphasic cultivation strategy to avoid Epo-Fc aggregation and optimize protein expression. Journal of Biotechnology, 227, 3–9.

Kaleta, C., Schauble, S., Rinas, U., Schuster, S., 2013. Metabolic costs of amino acid and protein production in *Escherichia coli*. Biotechnology Journal, 8 (9), 1105–1114.

Karpenko, L.I., Ivanisenko, V.A., Pika, I.A., Chikaev, N.A., Eroshkin, A.M., Veremeiko, T.A., Ilyichev, A.A., 2000. Insertion of foreign epitopes in HBcAg: how to make the chimeric particle assemble. Amino Acids, 18 (4), 329–337.

Kaur, J., Kumar, A., Kaur, J., 2018. Strategies for optimization of heterologous protein expression in *E. coli*: Roadblocks and reinforcements. International Journal of Biological Macromolecules, 106, 803–822.

Kazaks, A., Balmaks, R., Voronkova, T., Ose, V., Pumpens, P., 2008. Melanoma vaccine candidates from chimeric hepatitis B core virus-like particles carrying a tumor-associated MAGE-3 epitope. Biotechnology Journal, 3 (11), 1429–1436.

Kenney, J. M., von Bonsdorff, C. H., Nassal, M., & Fuller, S. D., 1995. Evolutionary conservation in the hepatitis B virus core structure: comparison of human and duck cores. Structure (London, England : 1993), 3 (10), 1009–1019.

Kesik-Brodacka, M., Romanik, A., Mikiewicz-Sygula, D., Plucienniczak, G., Plucienniczak, A., 2012. A novel system for stable, high-level expression from the T7 promoter. Microbial Cell Factories, 11, 109.

Kim, B.S., Lee, S.C., Lee, S.Y., Chang, Y.K., Chang, H.N., 2004. High cell density fed-batch cultivation of *Escherichia coli* using exponential feeding combined with pH-stat. Bioprocess and Biosystems Engineering, 26 (4), 283.

Kim, H.J., Kwon, Y.D., Lee, S.Y., Kim, P., 2012. An engineered *Escherichia coli* having a high intracellular level of ATP and enhanced recombinant protein production. Applied Microbiology and Biotechnology, 94 (4), 1079–1086.

Kim, Y., Hipp, M., Bracher, A., Hayer-Hartl, M., Ulrich-Hartl, F., 2013. Molecular chaperone functions in protein folding and proteostasis. Annual Review of Biochemistry, 82 (1), 323–355.

Kimata, K., Takahashi, H., Inada, T., Postma, P., Aiba, H., 1997. cAMP receptor protein–cAMP plays a crucial role in glucose–lactose diauxie by activating the major glucose

transporter gene in *Escherichia coli*. Proceedings of the National Academy of Sciences, 94 (24), 12914-12919.

Koletzki, D., Lundkvist, A., Sjolander, K.B., Gelderblom, H. R., Niedrig, M., Meisel, H., Kruger, D. H., and Ulrich, R., 2000. Puumala (PUU) hantavirus strain differences and insertion positions in the hepatitis B virus core antigen influence B-cell immunogenicity and protective potential of core-derived particles. Virology, 276, 364-375.

Konz, J.O., King, J., and Cooney, C.L., 1998. Effects of oxygen on recombinant protein expression. Biotechnology Progress, 14, 393-409.

Kovarova, K., Zehnder, A.J., Egli., T., 1996. Temperature-dependent growth kinetics of *Escherichia coli* ML 30 in glucose-limited continuous culture. Journal of Bacteriology, 178 (15), 4530–4539.

Kram, K.E., Finkel, S.E., 2015. Rich medium composition affects *Escherichia coli* survival, glycation, and mutation frequency during long-term batch culture. Applied and Environmental Microbiology, 81 (13), 4442–4450.

Ladd-Effio, C., Baumann, P., Weigel, C., Vormittag, P., Middelberg, A., Hubbuch, J., 2016. High-throughput process development of an alternative platform for the production of virus-like particles in *Escherichia coli*. Journal of Biotechnology, 219, 7–19.

Lederman, E.R., Davidson, W., Groff, H.L., Smith, S.K., Warkentien, T., Li, Y., Wilkins, K.A., Karem, K.L., Akondy, R.S., Ahmed, R., Frace, M., Shieh, W.J., Zaki, S., Hruby, D.E., Painter, W.P., Bergman, K.L., Cohen, J.I., Damon, I. K., 2012. Progressive vaccinia: case description and laboratory-guided therapy with vaccinia immune globulin, ST-246, and CMX001. The Journal of Infectious Diseases, 206 (9), 1372–1385.

Lee, S.Y., 1996. High cell-density culture of *Escherichia coli*. Trends Biotechnology, 14, 98–105.

Lenth, R.V., 1989. Quick and easy analysis of unreplicated factorials. Technometrics, 31, 469–473.

Li, J., Jaitzig, J., Lu, P., Sussmuth, R. D., Neubauer, P., 2015. Scale-up bioprocess development for production of the antibiotic valinomycin in *Escherichia coli* based on consistent fed-batch cultivations. Microbial Cell Factories, 14, 83.

Li, M., Meng, X., Diao, E., Du, F., Zhao, X., 2012. Productivity enhancement of S -adenosylmethionine in *Saccharomyces cerevisiae* using n -hexadecane as oxygen vector. Journal of Chemical Technology & Biotechnology, 87 (10), 1379–1384.

Li, Z., Wei, J., Yang, Y., Liu, L., Ma, G., Zhang, S., Su, Z., 2018. A two-step heat treatment of cell disruption supernatant enables efficient removal of host cell proteins before

chromatographic purification of HBc particles. Journal of Chromatography A, 1581-1582, 71-79.

Liew, M., Rajendran, A., Middelberg, A., 2010. Microbial production of virus-like particle vaccine protein at gram-per-litre levels. Journal of Biotechnology, 150, 224-231.

Lim, H., Jung, K., Park, D., Chung, S., 2000. Production characteristics of interferon-α using an l -arabinose promoter system in a high-cell-density culture. Applied Microbiology and Biotechnology, 53 (2), 201–208.

Lindner, S.E., Sugden, B., 2007. The plasmid replicon of Epstein–Barr virus: Mechanistic insights into efficient, licensed, extrachromosomal replication in human cells. Plasmid, 58 (1), 1–12.

Liu, M., Durfee, T., Cabrera, J.E., Zhao, K., Jin, D.J., Blattner, F.R., 2005. Global transcriptional programs reveal a carbon source foraging strategy by *Escherichia coli*. The Journal of Biological Chemistry, 280 (16), 15921–15927.

Liu, X., Xu, J., Xia, J., Lv, J., Wu, Z., Deng, Y., 2016. Improved production of citric acid by *Yarrowia lipolytica* using oleic acid as the oxygen-vector and co-substrate. Engineering in Life Sciences, 16 (5), 424–431.

Lobaina, Y., Aguiar, J., Penton, E., Aguilar, J.C., 2015. Demonstration of safety, immunogenicity and evidences of efficacy of the therapeutic vaccine candidate HeberNasvac and characterization of chronic hepatitis B patient populations. Biotecnologia Aplicada, 32, 3511–3513.

Losen, M., Frolich, B., Pohl, M., Buchs, J., 2004. Effect of oxygen limitation and medium composition on *Escherichia coli* fermentation in shake-flask cultures. Biotechnology Progress, 20, 1062–1068.

Ma, X., Fan, D., Shang, L., Cai, Q., Chi, L., Zhu, C., Mi, Y. Luo, Y., 2010. Oxygen transfer rate control in the production of human-like collagen by recombinant *Escherichia coli*. Biotechnology and Applied Biochemistry, 55, 169-174.

Maachupalli-Reddy, J., Kelley, B.D., Clark, E.D.B., 1997. Effect of inclusion body contaminants on the oxidative renaturation of hen egg white lysozyme. Biotechnology Progress, 13 (2), 144–150.

Mak, T.W., Saunders, M.E., 2006, Vaccines and clinical immunization. In Mak, T.W., Saunders, M.E., (Eds). The immune response, Academic Press, 695-749, ISBN: 9780120884513.

Mandenius, C., Brundin, A., 2008. Bioprocess optimization using design-of-experiments methodology. Biotechnology Progress, 24 (6), 1191-1203.

Marini, G., Luchese, M.D., Argondizzo, A.P., de Goes, A.C., Galler, R., Alves, T.L., Medeiros, M.A., Larentis, A.L., 2014. Experimental design approach in recombinant protein expression: determining medium composition and induction conditions for expression of pneumolysin from *Streptococcus pneumoniae* in *Escherichia coli* and preliminary purification process. BMC Biotechnology, 14, 1.

Markossian, K., Kurganov, B., 2004. Protein folding, misfolding, and aggregation - Formation of inclusion bodies and aggresomes. Biochemistry (Moscow), 69 (9), 971-984.

Martinez-Gomez, K., Flores, N., Castaneda, H.M., Martinez-Batallar, G., Hernandez-Chavez, G., Ramirez, O.T., Gosset, G., Encarnacion, S., Bolivar, F., 2012. New insights into *Escherichia coli* metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol. Microbial Cell Factories, 11 (1), 46.

Mayer, C.E., Geerlof, A., Schepers, A., 2012. Efficient expression and purification of tag-free Epstein–Barr virus EBNA1 protein in *Escherichia coli* by auto-induction. Protein Expression and Purification, 86 (1), 7–11.

McCarthy, B.J., 1962. The effects of magnesium starvation on the ribosome content of *Escherichia coli*. Biochimica et Biophysica Acta, 55, 880–888.

Mears, L., Stocks, S.M., Sin, G., Gernaey, K.V., 2017. A review of control strategies for manipulating the feed rate in fed-batch fermentation processes. Journal of Biotechnology, 245, 34–46.

Miao, F., Kompala, D., 1992. Overexpression of cloned genes using recombinant *Escherichia coli* regulated by a T7 promoter: I. Batch cultures and kinetic modelling. Biotechnology and Bioengineering, 40 (7), 787–796.

Middelberg, A., Rivera-Hernandez, T., Wibowo, N., Lua, L., Fan, Y., Magor, G., Chang, C., Chuan, Y., Good, M., Batzloff, M., 2011. A microbial platform for rapid and low-cost virus-like particle and capsomere vaccines. Vaccine, 29, 7154-7162.

Mihailova, M., Boos, M., Petrovskis, I., Ose, V., Skrastina, D., Fiedler, M., Sominskaya, I., Ross, S., Pumpens, P., Roggendorf, M., Viazov, S., 2006. Recombinant virus-like particles as a carrier of B- and T-cell epitopes of hepatitis C virus (HCV). Vaccine, 24, 4369-4377.

Mogk, A., Mayer, M.P., Deuerling, E., 2002. Mechanisms of protein folding: Molecular chaperones and their application in biotechnology. ChemBioChem, 3 (9), 807–814.

Montague, N.P., Thuenemann, E.C., Saxena, P., Saunders, K., Lenzi, P., Lomonossoff, G.P., 2011. Recent advances of Cowpea mosaic virus-based particle technology, Human Vaccines, 7 (3), 383-390.

Montgomery, D.C., 2012. Design and analysis of experiments, eight ed. John Wiley & Sons, Inc. New York: Wiley. ISBN 978-1-118-14692-7.

Muhlmann, M., Forsten, E., Noack, S., Buchs, J., 2017. Optimizing recombinant protein expression via automated induction profiling in microtiter plates at different temperatures. Microbial Cell Factories, 16.

Muntari, B., Amid, A., Mel, M., Jami, M., Salleh, H., 2012. Recombinant bromelain production in *Escherichia coli*: process optimization in shake flask culture by response surface methodology. AMB Express, 2(1), 1–9

Murphy, T.E., Tsui, K., Allen, J.K., 2005. A review of robust design methods for multiple responses. Research in Engineering Design, 16, 118–132.

Murray, K., Shiau, A.L., 1999. The core antigen of Hepatitis B Virus as a carrier for immunogenic peptides. Biological Chemistry, 380 (3), 277-283,

Myers, R., Montgomery, D., Anderson-Cook, C., 2016. Response surface methodology: Process and product optimization using designed experiments, fourth ed., Wiley series in probability and statistics, Hoboken, New Jersey: Wiley. ISBN: 978-1-118-91601-8.

Naito, M., Ishii, K., Nakamura, Y., Kobayashi, M., Takada, S., Koike, K., 1997. Simple method for efficient production of hepatitis B virus core antigen in *Escherichia coli*. Research in Virology (Paris), 148 (4), 299–305.

Narta, U., Roy, S., Kanwar, S. S., Azmi, W., 2011. Improved production of L-asparaginase by *Bacillus brevis* cultivated in the presence of oxygen-vectors. Bioresource Technology, 102 (2), 2083–2085.

Naskalska, A., Pyre, K., 2015. Virus-like Particles as immunogens and universal nanocarriers. Polish Journal of Microbiology, 64 (1), 3–13.

Ng, M.Y.T., Tan, W.S., Abdullah, N., Ling, T.C., Tey, B.T., 2006. Heat treatment of unclarified *Escherichia coli* homogenate improved the recovery efficiency of recombinant hepatitis B core antigen. Journal of Virological Methods, 137 (1), 134–139.

Nielsen, D., Daugulis, A., Mclellan, P., 2003. A novel method of simulating oxygen mass transfer in two-phase partitioning bioreactors. Biotechnology and Bioengineering, 83 (6), 735–742.

Nierhaus, K., 2014. Mg^{2+}, K^{+}, and the ribosome. Journal of Bacteriology, 196 (22), 3817–3819.

Noguere, C., Larsson, A., Guyot, J., Bignon, C., 2012. Fractional factorial approach combining 4 *Escherichia coli* strains, 3 culture media, 3 expression temperatures and 5 N-terminal fusion tags for screening the soluble expression of recombinant proteins. Protein Expression and Purification, 84 (2), 204–213.

Nugent, M., Primrose, S., Tacon, W., 1983. The stability of recombinant DNA. Developments in Industrial Microbiology 24, 271-285.

Nuttall, J.J., Eley, B.S., 2011. BCG vaccination in HIV-infected children. Tuberculosis Research and Treatment, 712736.

Ockuly, R., Weese, M., Smucker, B., Edwards, D., Chang, L., 2017. Response surface experiments: A meta-analysis. Chemometrics and Intelligent Laboratory Systems, 164, 64–75.

Olaofe, O., Burton, S., Cowan, D., Harrison, S., 2010. Improving the production of a thermostable amidase through optimising IPTG induction in a highly dense culture of recombinant *Escherichia coli*. Biochemical Engineering Journal, 52 (1), 19-24.

Overton, T., 2014. Recombinant protein production in bacterial hosts. Drug Discovery Today, 19 (5), 590–601.

Papaneophytou, C., Kontopidis, G., 2012. Optimization of TNF-α overexpression in *Escherichia coli* using response surface methodology: Purification of the protein and oligomerization studies. Protein Expression and Purification, 86 (1), 35–44.

Papaneophytou, C., Kontopidis, G., 2014. Statistical approaches to maximize recombinant protein expression in *Escherichia coli*: A general review. Protein Expression and Purification, 94, 22-32.

Perez, E.M., Foley, J., Tison, T., Silva, R., Ogembo, J.G., 2017. Novel Epstein-Barr virus-like particles incorporating gH/gL-EBNA1 or gB-LMP2 induce high neutralizing antibody titers and EBV-specific T-cell responses in immunized mice. Oncotarget, 8 (12), 19255–19273.

Peternel, S., Komel, R., 2011. Active Protein aggregates produced in *Escherichia coli*. International Journal of Molecular Sciences, 12 (11), 8275–8287.

Pflug, S., Richter, S.M., Urlacher, V.B., 2007. Development of a fed-batch process for the production of the cytochrome P450 monooxygenase CYP102A1 from *Bacillus megaterium* in *E. coli*. Journal of Biotechnology, 129 (3), 481–488.

Phue, J., Shiloach, J., 2005. Impact of dissolved oxygen concentration on acetate accumulation and physiology of *E. coli* BL21, evaluating transcription levels of key genes at different dissolved oxygen conditions. Metabolic Engineering, 7, 353-363.

Pilarek, M., Glazyrina, J., Neubauer, P., 2011. Enhanced growth and recombinant protein production of *Escherichia coli* by a perfluorinated oxygen carrier in miniaturized fed-batch cultures. Microbial Cell Factories, 10, 50.

Porterfield, J.Z., Dhason, M.S., Loeb, D.D., Nassal, M., Stray, S.J., Zlotnick, A., 2010. Full-length Hepatitis B virus core protein packages viral and heterologous RNA with similarly high levels of cooperativity. Journal of Virology, 84 (14), 7174–7184.

Proffitt, A., 2012. First HEV vaccine approved. Nature Biotechnology, 30, 300.

Pu, L., Yao, Y., Dong, S.Q., Feng, Q., Ting, F.X, Xin, L.X, Li, B.S., 2018. Efficient humoral and cellular immune responses induced by a chimeric virus-like particle displaying the epitope of EV71 without adjuvant. Biomedical and Environmental Sciences, 31, (5): 343-350.

Puente-Massaguer, E., Lecina, M., Godia, F., 2020. Integrating nanoparticle quantification and statistical design of experiments for efficient HIV-1 virus-like particle production in High Five cells. Applied Microbiology and Biotechnology, 104 (4), 1569–1582.

Pumpens, P., Grens, E., 1999. Hepatitis B core particles as a universal display model: a structure-function basis for development. FEBS letters, 442 (1), 1–6.

Pumpens, P., Grens, E., 2001. HBV core particles as a carrier for B cell/T cell epitopes. Intervirology, 44, 98–114.

Pumpens, P., Ulrich, R., Sasnauskas, K., Kazaks, A., Ose, V., Grens, E., 2008. Construction of novel vaccines on the basis of the virus-like particles: hepatitis B virus proteins as vaccine carriers, in Khudyakov, Y., (Eds), Medicinal protein engineering. CRC Press Taylor & Francis Group, Boca Raton, FL, 205-248.

Pushko, P., Pumpens, P., Grens, E., 2013. Development of virus-like particle technology from small highly symmetric to large complex virus-like particle structures. Intervirology, 56, 141-165.

Qiao, L., Zhang, Y., Chai, F., Tan, Y., Huo, C., Pan, Z., 2016. Chimeric virus-like particles containing a conserved region of the G protein in combination with a single peptide of the M2 protein confer protection against respiratory syncytial virus infection. Antiviral Research, 131, 131-140.

Rakic, T., Kasagic-Vujanovic, I., Jovanovic, M., Jancic-Stojanovic, B., Ivanovic, D., 2014. Comparison of full factorial design, central composite design, and box-behnken design in chromatographic method development for the determination of fluconazole and its impurities. Analytical Letters, 47 (8), 1334–1347.

Ramchuran, S.O., Holst, O., Karlsson, E.N., 2005. Effect of post-induction nutrient feed composition and use of lactose as inducer during production of thermostable xylanase in *Escherichia coli* glucose-limited fed-batch cultivations. Journal of Bioscience and Bioengineering, 99 (5), 477–484.

Ramirez, O., Zamora, R., Espinosa, G., Merino, E., Bolívar, F., Quintero, R., 1994. Kinetic study of penicillin acylase production by recombinant *E. coli* in batch cultures, Process Biochemistry, 29 (3), 197–206.

Ravin, N.V., Blokhinaa, E.A., Kuprianova, V.V., Stepanovab, L.A., Shaldjanb, A.A., Kovalevab, A.A., Tsybalovab, L.M., Skryabin, K.G., 2015. Development of a candidate influenza vaccine based on virus-likeparticles displaying influenza M2e peptide into the immunodominantloop region of hepatitis B core antigen: Insertion of multiple copies ofM2e increases immunogenicity and protective efficiency. Vaccine, 33, 3392-3397.

Riedel, S., 2005. Edward Jenner and the history of smallpox and vaccination. Proceedings Baylor University. Medical Center, 18 (1), 21–25.

Rohe, P., Venkanna, D., Kleine, B., Freudl, R., Oldiges, M., 2012. An automated workflow for enhancing microbial bioprocess optimization on a novel microbioreactor platform. Microbial Cell Factories, 11, 144.

Roldao, A., Mellado, M.C.M., Castilho, L.R., Carrondo, M.J.T., Alves, P.M., 2010. Virus-like particles in vaccine development. Expert Review of Vaccines, 9 (10), 1149-1176.

Rols, J., Condoret, J., Fonade, C., Goma, G., 1990. Mechanism of enhanced oxygen transfer in fermentation using emulsified oxygen-vectors. Biotechnology and Bioengineering, 35 (4), 427–435.

Roose, K., Baets, S., Schepens, B., Saelens, X., 2013. Hepatitis B core–based virus–like particles to present heterologous epitopes. Expert Review of Vaccines, 12, 183-198.

Rosano, G.L., Ceccarelli, E.A., 2014. Recombinant protein expression in *Escherichia coli*: Advances and challenges. Frontier in Microbiology, 5, 172.

Roseman, A.M., Berriman, J.A., Wynne, S.A., Butler, P.J.G., Crowther, R.A., 2005. A structural model for maturation of the hepatitis B virus core. Proceedings of the National Academy of Sciences of the United States of America, 102, 15821–15826.

Roseman, A.M., Borschukova, O., Berriman, J.A., Wynne, S.A., Pumpens, P., Crowther, R. A., 2012. Structures of hepatitis B virus cores presenting a model epitope and their complexes with antibodies. Journal of molecular biology, 423 (1), 63–78.

Rozkov, A., Enfors, S., 2004. Analysis and control of proteolysis of recombinant proteins in *Escherichia coli*. Advances in Biochemical Engineering, Biotechnology, 89, 163–195.

Ruiss, R., Jochum, S., Wanner, G., Reisbach, G., Hammerschmidt, W., Zeidler, R., 2011. A virus-like particle-based Epstein-Barr virus vaccine. Journal of Virology, 85 (24), 13105–13113.

Sagmeister, P., Jazini, M., Klein, J., Herwig, C., Meyer, H., Schmidhalter, D.R., 2014. Bacterial suspension cultures, in: Industrial scale suspension culture of living cells. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, pp. 40–93.

Saibil H., 2013. Chaperone machines for protein folding, unfolding and disaggregation. Nature Reviews Molecular Cell Biology, 14, 630–642.

Sandoval-Basurto, E.A., Gosset, G., Bolivar, F., Ramirez, O.T., 2005. Culture of *Escherichia coli* under dissolved oxygen gradients simulated in a two-compartment scale-down system: metabolic response and production of recombinant protein. Biotechnology and Bioengineering, 89 (4), 453–463.

Schein, C.H., 1989. Production of soluble recombinant proteins in bacteria. Nature Biotechnology, 7, 1141–1148.

Schneider, C., Rasband, W., Eliceiri, K., 2012. NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9, 671-675.

Schodel, F., Moriarty, A.M., Peterson, D.L., Zheng, J.A., Hughes, J.L., Will, H., Leturcq, D.J., McGee, J.S., Milich, D.R., 1992. The position of heterologous epitopes inserted in hepatitis B virus core particles determines their immunogenicity. Journal of Virology, 66 (1), 106–114.

Schrodel, A., de Marco, A., 2005. Characterization of the aggregates formed during recombinant protein expression in bacteria. BMC Biochemistry, 6, 10.

Seifer, M., Standring, D.N., 1994. A protease-sensitive hinge linking the two domains of the hepatitis B virus core protein is exposed on the viral capsid surface. Journal of Virology, 68 (9), 5548–5555.

Sezonov, G., Joseleau-Petit, D., D'Ari, R., 2007. *Escherichia coli* physiology in Luria-Bertani Broth. The Journal of Bacteriology, 189 (23), 8746–8749.

Shalel-Levanon, S., San, K., Bennett, G., 2005. Effect of oxygen, and ArcA and FNR regulators on the expression of genes related to the electron transfer chain and the TCA cycle in *Escherichia coli*. Metabolic Engineering, 7, 364-374.

Shin, C., Hong, M., Lee, J., 1996. Oxygen transfer correlation in high cell density culture of recombinant *E. coli*. Biotechnology Techniques, 10 (9).

Siegele, D.A., Hu, J.C., 1997. Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations.

Proceedings of the National Academy of Sciences of the United States of America, 94, 8168–8172.

Silva, F., Queiroz, J.A., Domingues, F.C., 2012. Evaluating metabolic stress and plasmid stability in plasmid DNA production by *Escherichia coli*. Biotechnology Advances, 30 (3), 691–708.

Sivachandran, N., Wang, X., Frappier, L., 2012. Functions of the Epstein - Barr virus EBNA1 protein in viral reactivation and lytic infection. Journal of Virology, 86 (11), 6146-6158.

Sklodowska, J., Jakiela, S., 2017. Enhancement of bacterial growth with the help of immiscible oxygenated oils. RSC Advances. 7 (65), 40990-40995.

Skrastina, D., Petrovskis, I., Petraityte, R., Sominskaya, I., Ose, V., Lieknina, I., Bogans, J., Sasnauskas, K., Pumpens, P., 2013. Chimeric derivatives of hepatitis B virus core particles carrying major epitopes of the rubella virus E1 glycoprotein. Clinical and Vaccine Immunology, 20 (11), 1719–1728.

Smith, D.W., Sugden, B., 2013. Potential cellular functions of Epstein-Barr Nuclear Antigen 1 (EBNA1) of Epstein-Barr Virus. Viruses, 5 (1), 226–240.

Smith, M.L., Lindbo, J.A., Dillard-Telm, S., Brosio, P.M., Lasnik, A.B., McCormick, A.A., Nguyen, L.V., Palmer, K.E., 2006. Modified tobacco mosaic virus particles as scaffolds for display of protein antigens for vaccine applications. Virology, 34, 475.

Snedecor, G.W., Cochran, W.G., 1989. Statistical Methods. Iowa State University Press, Iowa, USA. ISBN: 0-8138-1561-6.

Sokal, E.M., Hoppenbrouwers, K., Vandermeulen, C., Moutschen, M., Leonard, P., Moreels, A., Haumont, M., Bollen, A., Smets, F., Denis, M., 2007. Recombinant gp350 vaccine for infectious mononucleosis: a phase 2, randomized, double-blind, placebo-controlled trial to evaluate the safety, immunogenicity, and efficacy of an Epstein-Barr virus vaccine in healthy young adults. The Journal of Infectious Diseases, 196 (12), 1749–1753.

Sominskaya, I., Skrastina, D., Dislers, A., Vasiljev, D., Mihailova, M., Ose, V., Dreilina, D., Pumpens, P., 2010. Construction and immunological evaluation of multivalent hepatitis B virus (HBV) core virus-like particles carrying HBV and HCV epitopes. Clinical and Vaccine Immunology, 17, 1027-1033.

Sominskaya, I., Skrastina, D., Petrovskis, I., Dishlers, A., Berza, I., Mihailova, M., Jansons, J., Akopjana, I., Stahovska, I., Dreilina, D., Ose, V., Pumpens, P., 2013. A VLP library of C-terminally truncated hepatitis B core proteins: Correlation of RNA encapsidation with a Th1/Th2 switch in the immune responses of mice. PLoS ONE. 8, e75938.

Sorensen, H.P., Mortensen, K.K., 2005. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. Microbial Cell Factories, 4 (1), 1–1.

Soriano, V., Young, B., Reau, N., 2018. Report from the International Conference on Viral Hepatitis – 2017. AIDS reviews, 20 (1), 58–70.

Spiro, S., Guest, J.R., 1991. Adaptive responses to oxygen limitation in *Escherichia coli*. Trends in Biochemical Sciences (Amsterdam. Regular Ed.), 16 (8), 310–314.

Stoll-Keller, F., Barth, H., Fafi-Kremer, S., Zeisel, M.B., Baumert, T.F., 2009. Development of hepatitis C virus vaccines: challenges and progress. Expert Review of Vaccines, 8 (3), 333–345.

Strandberg, L., Enfors, S.O., 1991. Factors influencing inclusion body formation in the production of a fused protein in *Escherichia coli*. Applied Environmental Microbiology, 57 (6), 1669-1674.

Studier, F.W., 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. Journal of Molecular Biology, 219 (1), 37–44.

Studier, F.W., 2005. Protein production by auto-induction in high density shaking cultures. Protein Expression and Purification, 41 (1), 207–234.

Studier, F.W., Moffatt, B., 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. Journal of Molecular Biology, 189 (1), 113–130.

Suffian, I.F.B.M., Garcia-Maya, M., Brown, P., Bui, T., Nishimura, Y., Palermo, A.R., Ogino, C., Kondo, A., Al-Jamal, K.T., 2017. Yield optimisation of Hepatitis B Virus Core particles in *E. coli* expression system for drug delivery applications. Scientific Reports, 7, 43160.

Sun, Q., Cao, L., Fang, L., Chen, C., Dai, J., Chen, L., Hua, Z., 2005. Expression, purification of human vasostatin120–180 in *Escherichia coli*, and its anti-angiogenic characterization. Protein Expression and Purification, 39 (2), 288–295.

Sun, X., Wang, Y., Dong, C., Hu, J., Yang, L., 2015. High copy numbers and N terminal insertion position of influenza A M2E fused with hepatitis B core antigen enhanced immunogenicity. Bioscience Trends, 9 (4), 221–227.

Swalley, S.E., Fulghum, J.R., Chambers, S.P., 2006. Screening factors effecting a response in soluble protein expression: formalized approach using design of experiments. Analytical biochemistry, 351 (1), 122–127.

Szenk, M., Dill, K.A., de Graff, A.M.R., 2017. Why do fast-growing bacteria enter overflow metabolism? Testing the membrane real estate hypothesis. Cell Systems, 5, (2), 95–104.

Tagliamonte, M., Tornesello, M.L., Buonaguro, F.M., Buonaguro, L., 2017. Chapter 11 - Virus-like particles. In Skwarczynski, M., Toth, I., (Eds.), Micro and nanotechnology in vaccine development, 205-214. Elsevier. ISBN: 978-0-323-39981-4.

Tan, W.S., McNae, I.W., Ho, K.L., Walkinshaw, M.D., 2007. Crystallization and X-ray analysis of the T=4 particle of hepatitis B capsid protein with an N-terminal extension. Acta Crystallographica. Section F, Structural Biology and Crystallization Communications, 63 (8), 642–647.

Taylor, G.A.R, 1994. Analysis of Experiments by Using Half-Normal Plots. Journal of the Royal Statistical Society. Series D (The Statistician), 43 (4), 529–536.

Tegel, H., Ottosson, J., Hober, S., 2011. Enhancing the protein production levels in *Escherichia coli* with a strong promoter, The FEBS Journal, 278 (5), 729–739.

Tree, J.A., Richardson, C., Fooks, A.R., Clegg, J.C., Looby, D., 2001. Comparison of large-scale mammalian cell culture systems with egg culture for the production of influenza virus A vaccine strains. Vaccine, 19, 3444–3450.

Tripathi, N., Babu, J., Shrivastva, A., Parida, M., Jana, A., Rao, P., 2008. Production and characterization of recombinant dengue virus type 4 envelope domain III protein. Journal of Biotechnology, 134 (3-4), 278–286

Tripathi, N., Sathyaseelan, K., Jana, A., Rao, P., 2009. High yield production of heterologous proteins with *Escherichia coli*. Defence Science Journal, 59 (2), 137-146.

Tritel, M., Resh, M., 2001. The late stage of human immunodeficiency virus Type 1 assembly is an energy-dependent process. Journal of Virology, 75, 5473-5481.

Tumban, E., Peabody, J., Peabody, D.S., Chackerian, B., 2013. A universal virus-like particle-based vaccine for human papillomavirus: longevity of protection and role of endogenous and exogenous adjuvants. Vaccine, 31 (41), 4647-4654.

Uhoraningoga, A., Kinsella, G.K., Henehan, G.T., Ryan, B.J., 2018. The goldilocks approach: A review of employing design of experiments in prokaryotic recombinant protein production. Bioengineering, 5 (4), 89.

van Dijk E., Hoogeveen A., Abeln S., 2015. The hydrophobic temperature dependence of amino acids directly calculated from protein structures. PLOS Computational Biology, 11, e1004277.

Ventura, S., Villaverde, A., 2006. Protein quality in bacterial inclusion bodies. Trends in biotechnology, 24 (4), 179–185.

Vera, A., Gonzalez-Montalban, N., Aris, A., Villaverde, A., 2007. The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. Biotechnology and Bioengineering, 96 (6), 1101–1106.

Vicente, T., Roldao, A., Peixoto, C., Carrondo, M.J., Alves, P.M., 2011. Large-scale production and purification of VLP-based vaccines. Journal of Invertebrate Pathology, 107 Suppl, S42–S48.

Vogel, M., Vorreiter, J., Nassal, M., 2005. Quaternary structure is critical for protein display on capsid-like particles (CLPs): Efficient generation of hepatitis B virus CLPs presenting monomeric but not dimeric and tetrameric fluorescent proteins. Proteins, 58, 478-488

von Wulffen, J., Sawodny, O., Feuer, R., 2016. Transition of an anaerobic *Escherichia coli* culture to aerobiosis: Balancing mRNA and protein levels in a demand-directed dynamic flux balance analysis. PLOS One, 11, e0158711.

Waegeman, H., De Lausnay, S., Beauprez, J., Maertens, J., De Mey, M., Soetaert, W., 2013. Increasing recombinant protein production in *Escherichia coli* K12 through metabolic engineering. New biotechnology, 30 (2), 255–261.

Walsh, G., Jefferis, R., 2006. Post-translational modifications in the context of therapeutic proteins. Nature biotechnology, 24 (10), 1241–1252.

Wang, Y.S., Ouyang, W., Liu, X.J., He, K.W., Yu, S.Q., Zhang, H.B., Fan, H.J., Lu, C.P., 2012. Virus-like particles of hepatitis B virus core protein containing five mimotopes of infectious bursal disease virus (IBDV) protect chickens against IBDV. Vaccine, 30 (12), 2125–2130.

Wechselberger, P., Sagmeister, P., Engelking, H., Schmidt, T., Wenger, J., Herwig, C., 2012. Efficient feeding profile optimization for recombinant protein production using physiological information. Bioprocess and Biosystems Engineering, 35 (9), 1637–1649.

Westbrook, A., Ren, X., Moo-Young, M., Chou, C., 2018. Application of hydrocarbon and perfluorocarbon oxygen vectors to enhance heterologous production of hyaluronic acid in engineered *Bacillus subtilis*. Biotechnology and Bioengineering, 115 (5), 1239–1252.

Wheeler, C.M., Bautista, O.M., Tomassini, J.E., Nelson, M., Sattler, C.A., Barr, E., 2008. Safety and immunogenicity of co-administered quadrivalent human papillomavirus (HPV)-6/11/16/18 L1 virus-like particle (VLP) and hepatitis B (HBV) vaccines. Vaccine, 26, 686–696.

Whiffin, V.S., Cooney, M.J., Cord-Ruwisch, R., 2004. Online detection of feed demand in high cell density cultures of *Escherichia coli* by measurement of changes in dissolved oxygen transients in complex media. Biotechnology and Bioengineering, 85, 422-433.

Wong, S.S., Webby, R.J., 2013. Traditional and new influenza vaccines. Clinical Microbiology Reviews, 26 (3), 476–492.

Wu, C., Hamada, M., 2009. Experiments: Planning, analysis, and optimization, second ed., Wiley series in probability and statistics, Hoboken, New Jersey: Wiley. ISBN: 978-0-471-69946-0.

Wu, C.L., Leu, T.S., Chang, T.T., Shiau, A.L., 1999. Hepatitis C virus core protein fused to hepatitis B virus core antigen for serological diagnosis of both hepatitis C and hepatitis B infections by ELISA. Journal of Medical Virology, 57 (2), 104–110.

Wynne, S., Crowther, R., Leslie, A., 1999. The crystal structure of the human hepatitis B virus capsid. Molecular Cell, 3 (6), 771–780.

Xu, H., 2018. Cochaperones enable Hsp70 to use ATP energy to stabilize native proteins out of the folding equilibrium. Scientific Reports, 8.

Yang, Q., Xu, J., Li, M., Lei, X., An, L., 2003. High-level expression of a soluble snake venom enzyme, gloshedobin, in *E. coli* in the presence of metal ions. Biotechnology Letters, 25 (8), 607–610.

Yang, Y., Li, H., Li, Z., Zhang, Y., Zhang, S., Chen, Y., Yu, M., Ma, G., Su, Z., 2015. Size-exclusion HPLC provides a simple, rapid, and versatile alternative method for quality control of vaccines by characterizing the assembly of antigens. Vaccine, 33, 1143-1150.

Yap, W.B., Tey, B.T., Ng, M.Y.T., Ong, S.T., Tan, W.S., 2009. N-terminally His-tagged hepatitis B core antigens: construction, expression, purification and antigenicity. Journal of Virological Methods, 160 (1), 125–131.

Yoon, K., Tan, W., Tey, B., Lee, K., Ho, K., 2012. Native agarose gel electrophoresis and electroelution: A fast and cost-effective method to separate the small and large hepatitis B capsids. Electrophoresis, 34, 244-253.

Zhang, S., Song, P., Li, S., 2018. Application of n-dodecane as an oxygen vector to enhance the activity of fumarase in recombinant *Escherichia coli*: role of intracellular microenvironment. Brazilian Journal of Microbiology, 49 (3), 662–667.

Zhang, Y., Liu, Y., Zhang, B., Yin, S., Li, X., Zhao, D., Wang, W., Bi, J., Su, Z., 2020. *In vitro* preparation of uniform and nucleic acid free hepatitis B core particles through an optimized disassembly-purification-reassembly process. Protein Expression and Purification, 105747.

Zhang, Y., Song, S., Liu, C., Wang, Y., Xian, X., He, Y., Wang, J., Liu, F., Sun, S., 2007. Generation of chimeric HBc proteins with epitopes in *E. coli*: Formation of virus-like

particles and a potent inducer of antigen-specific cytotoxic immune response and anti-tumor effect *in vivo*. Cellular immunology, 247 (1), 18–27.

Zhang, Y., Yin, S., Zhang, B., Bi, J., Liu, Y., Su, Z, 2020. HBc-based virus-like particle assembly from inclusion bodies using 2-methyl-2, 4-pentanediol. Process Biochemistry, 89, 233-237.

Zlotnick, A., Cheng, N., Conway, J., Booy, F., Steven, A., Stahl, S., Wingfield, P., 1996. Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. Biochemistry, 35 (23), 7412–7421.

Zlotnick, A., Cheng, N., Stahl, S.J., Conway, J.F., Steven, A.C., Wingfield, P.T., 1997. Localization of the C Terminus of the assembly domain of Hepatitis B Virus capsid protein: Implications for morphogenesis and organization of encapsidated RNA. Proceedings of the National Academy of Sciences, 94 (18), 9556–9561.

Zolgharnein, J., Shahmoradi, A., Ghasemi, J.B., 2013. Comparative study of Box–Behnken, central composite, and Doehlert matrix for multivariate optimization of Pb (II) adsorption onto Robinia tree leaves. Journal of Chemometrics, 27, 12-20.