Doctoral Dissertation

# Modeling of a CVA6 virus-like particle vaccine production process using CHO cell culture

Zhou Xing

December 2023

Graduate School of Engineering,
Osaka University

# Table of Contents

## Abbreviations

| | |
|---|---|
| ANN | artificial neural network |
| CHO | Chinese hamster ovary |
| COVID-19 | coronavirus disease of 2019 |
| CV | coefficient of variation |
| CVA16 | coxsackievirus A16 |
| CVA6 | coxsackievirus A6 |
| dsDNA | double stranded DNA |
| DT | digital twin |
| EMA | European Medicines Agency |
| EVA71 | Enterovirus A71 |
| FBA | flux balance analysis |
| FDA | Food and Drug Administration |
| GMP | Good Manufacturing Practice |
| GP | gaussian process |
| GSK | GlaxoSmithKline |
| HFMD | hand foot and mouth disease |
| LDH | lactate dehydrogenase |
| mAb | monoclonal antibody |
| MC | Monte Carlo |
| MDCK | Madin-Darby canine kidney |
| MERS-coronavirus | Middle East respiratory syndrome coronavirus |
| MFA | metabolic flux analysis |
| NTPs | nucleoside triphosphates |
| PAT | process analytical technology |
| PCA | principal component analysis |
| PLS | partial least squares |
| PTM | post-translational modification |

| | |
|---|---|
| SARS-CoV-2 | Severe acute respiratory syndrome coronavirus 2 |
| STR | stirred tank bioreactor |
| SVM | support vector machine |
| VLP | virus-like particle |

## Nomenclatures

| | |
|---|---|
| $c_{Amm}$ | ammonia concentration (mM) |
| $c_{Amm,cr}$ | critical ammonia concentration for specific death rate (mM) |
| $c_G$ | glucose concentration in the glucose supplement (mM) |
| $c_{Glc}$ | glucose concentration (mM) |
| $c_{Gln}$ | glutamine concentration (mM) |
| $c_{Lac}$ | lactate concentration (mM) |
| $c_M$ | glucose concentration in the feed medium (mM) |
| $k_D$ | specific death rate ($h^{-1}$) |
| $k_{D,dead}$ | death rate during dead phase ($h^{-1}$) |
| $k_{D,min}$ | minimum death rate ($h^{-1}$) |
| $k_{DL}$ | specific lysis rate from dead cells ($h^{-1}$) |
| $k_{DL,32}$ | specific lysis rate from dead cells at 32 °C ($h^{-1}$) |
| $k_L$ | specific lysis rate from viable cells ($h^{-1}$) |
| $k_{L,32}$ | specific lysis rate from viable cells at 32 °C ($h^{-1}$) |
| $K_{D_{Amm}}$ | constant for cell death due to ammonia accumulation (mM) |
| $K_{Lac}$ | Monod kinetic constant for lactate uptake (mM) |
| $m_{Glc}$ | specific glucose consumption rate for cell maintenance (mol $cell^{-1}$ $h^{-1}$) |
| $q_{Amm}$ | specific ammonia production rate (mol $cell^{-1}$ $h^{-1}$) |
| $q_{Amm,6.75}$ | specific ammonia production rate at pH of 6.75 (mol $cell^{-1}$ $h^{-1}$) |
| $q_{Amm,7.15}$ | specific ammonia production rate at pH of 7.15 (mol $cell^{-1}$ $h^{-1}$) |
| $q_{Lac}$ | specific lactate uptake rate (mol $cell^{-1}$ $h^{-1}$) |
| $q_{Lac,max}$ | maximum lactate uptake rate (mol $cell^{-1}$ $h^{-1}$) |

| | |
|---|---|
| $q_{VLP}$ | specific intracellular VLP accumulation rate (g cell$^{-1}$ h$^{-1}$) |
| $r$ | intrinsic growth rate (h$^{-1}$) |
| $r_{max}$ | maximum intrinsic growth rate (h$^{-1}$) |
| $t$ | time (h) |
| $V$ | working volume (L) |
| $V_{F,i}$ | volume of feed medium at ith feeding (L) |
| $V_{G,i}$ | volume of glucose supplement at ith feeding (L) |
| $VLP_{intra}$ | intracellular VLP content (g cell$^{-1}$) |
| $VLP_{max}$ | maximum intracellular VLP content (g cell$^{-1}$) |
| $X_d$ | dead cell density (cells L$^{-1}$) |
| $X_l$ | lysed cell density (cells L$^{-1}$) |
| $X_t$ | total cell density (cells L$^{-1}$) |
| $X_{t,max}$ | maximum total cell density (cells L$^{-1}$) |
| $X_v$ | viable cell density (cells L$^{-1}$) |
| $Y_{Amm/Gln}$ | yield coefficient of ammonia production to glutamine uptake (-) |
| $Y_{Lac/Glc}$ | yield coefficient of lactate production to glucose uptake (-) |
| $Y_{X_v/Amm}$ | yield coefficient of cell proliferation to ammonia uptake (cells mol$^{-1}$) |
| $Y_{X_v/Glc}$ | yield coefficient of cell proliferation to glucose uptake (cells mol$^{-1}$) |
| $Y_{X_v/Gln}$ | yield coefficient of cell proliferation to glutamine uptake (cells mol$^{-1}$) |

## Table of Figures

## Table of Tables

# Chapter 1:
# Introduction

# 1. Chapter 1: Introduction

## 1.1. Hand foot and mouth disease (HFMD)

### 1.1.1. Overview of HFMD

Hand foot and mouth disease (HFMD) is an infectious illness caused by enterovirus, which is a non-enveloped virus. It can be transmitted in multiple ways such as contact with the feces of an infected person (Takahashi et al., 2016). HFMD commonly occurs in children and the symptoms of it are usually mild including fever and rashes as shown in **Figure 1** (Aswathyraj et al., 2016). It is a self-limiting disease and can be recovered within a few days without any treatment. However, in some patients, severe complications may happen and even lead to death.

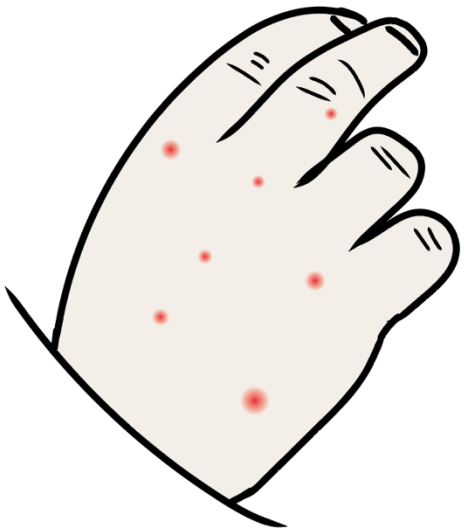

**Figure 1**: Sketch of rashes on a hand caused by HFMD

### 1.1.2. Outbreaks of HFMD

Many outbreaks of HFMD have been reported worldwide since 1970s (Aswathyraj et al., 2016). For example, two outbreaks were reported in Australia and Japan in 1973 (Ishimaru et al., 1980; Kennett et al., 1974). In recent years, small and large pandemics of HFMD were also found in Singapore, China, France, and USA (J. Hong et al., 2022; Kimmis et al., 2018; Mirand et al., 2021; Wu et al., 2010). Between 2008 and 2014, over 10 million cases of HFMD were reported by China with around 3000 deaths and a death rate of 0.03% (Esposito and Principi, 2018). Therefore, this disease poses a problem to public health.

### 1.1.3. Serotypes of the pathogen

In the enterovirus family, there are a few serotypes that are typically responsible for the outbreaks of HFMD. Enterovirus A71 (EVA71) is one of the famous pathogens that can cause HFMD and the first report of it can date back to 1969 in California, USA (Puenpa et al., 2019). EV71 still accounts for some outbreaks in recent decades. For example, about 120 cases of HFMD caused by EVA71 were identified by a 5-month investigation in 2013 in Sydney, Australia (Zander et al., 2014). Coxsackievirus A16 (CVA16) is another important agent of HFMD. A number of countries including Japan, Vietnam, Thailand, Singapore, and China have reported outbreaks due to CVA16 (Mao et al., 2014).

In addition to the two aforementioned serotypes, coxsackievirus A6 (CVA6) also plays a significant role in the HFMD and it is gaining more attention in recent years for several reasons. The incidence rate of HFMD caused by CVA6 increased resulting that CVA6 becomes a major pathogen of the HFMD (Kimmis et al., 2018). Besides, the course of HFMD associated with CVA6 is more severe and CVA6 is capable of causing HFMD among adults (Ramirez-Fort et al., 2014). Besides, other serotypes such as coxsackievirus A4, A10, and A16 also exist and account for some cases of HFMD.

### 1.1.4. Treatment and prevention measures

Currently, there is no specific treatment for HFMD. As a preventive measure, the vaccine is a promising method to fight against HFMD. An EVA71 vaccine was developed and launched in

China since 2016 which is the only vaccine available in the world and resulted in better control of HFMD in China (J. Hong et al., 2022). However, this vaccine failed to provide cross-protection against other serotypes (Takahashi et al., 2016). Therefore, vaccines targeting other serotypes of enterovirus are still needed in order to achieve a better control of HFMD.

## 1.2. Vaccines

### 1.2.1. Overview of vaccines

Vaccines are biological preparations that are administered into the body in order to prevent specific diseases. After the vaccination, they help the body to generate immunity in case of exposure to the real pathogen in the future. The vaccine has a history of over 200 years. Edward Jenner inoculated an 8-year-old boy in May 1796, which is known as the first inoculation of a vaccine (Riedel, 2005). In the 19$^{th}$ and 20$^{th}$ centuries, numerous human vaccines were developed targeting different diseases such as influenza, yellow fever, and hepatitis A (Plotkin, 2014). In the 21$^{st}$ century, vaccines are still playing important roles in fighting against infectious diseases. Most recently, billions of people were vaccinated during the pandemic of coronavirus disease of 2019 (COVID-19) and better control of the disease was achieved. Therefore, vaccines are still of great importance for the public health.

### 1.2.2. Types of vaccines

In the early stage of vaccine development, live attenuated or inactivated viruses were commonly used as vaccines. With the development of technology, several other types of vaccines were invented. As shown in **Figure 2**, there are different types of vaccines available now including live attenuated/inactivated vaccine, viral vector vaccine, subunit vaccine, and nucleic acid vaccine (DNA/RNA). In addition, some examples of these vaccines are listed in **Table 1** below as well as other information about them such as targeting disease and production system. For example, YF-VAX® is a live attenuated vaccine targeting yellow fewer produced by Sanofi. These vaccines can protect humans from infectious diseases by triggering immune responses through similar or different mechanisms.

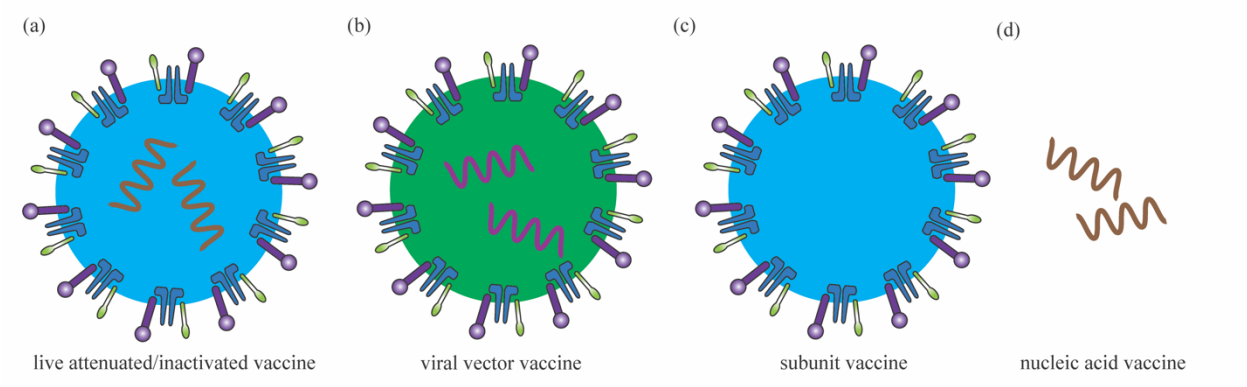


**Figure 2**: Different types of vaccines. **(a)** Live attenuated/inactivated vaccine. **(b)** Viral vector vaccine. **(c)** Subunit vaccine. **(d)** Nucleic acid vaccine (DNA/RNA vaccine)

**Table 1**: Examples of different types of vaccine and corresponding information.

| Vaccine name | Type of vaccine | Targeting disease | Production system | Manufacturer |
|---|---|---|---|---|
| YF-VAX® | live attenuated vaccine | yellow fewer | chicken embryo | Sanofi |
| Optaflu® | inactivated vaccine | influenza | Madin-Darby canine kidney (MDCK) cell | Novartis |
| Ervebo® | viral vector vaccine | Ebola | Vero cell | Merck Sharp & Dohme Corp. |
| HEPLISAV-B® | subunit vaccine | Hepatitis B | yeast cell | Dynavax |
| Comirnaty® | mRNA (nucleic acid vaccine) | COVID 19 | in vitro transcription | BioNTech & Pfizer |

1.2.2.1. Live attenuated/inactivated vaccine

Live attenuated vaccines use live viruses that have been attenuated to prevent them from causing diseases. Similarly, inactivated vaccines avoid the risk of causing diseases by killing viruses with chemical or physical methods. These types of vaccines have been used for a relatively long time compared with others.

An advantage of live attenuated and inactivated vaccines is that they can trigger the immune response which is similar to the one triggered by natural infection. Compared with live attenuated vaccines, inactivated vaccines cannot provide strong protection and therefore usually require adjuvant or multiple doses. Another drawback of inactivated vaccines is the risk of incomplete inactivation that might cause diseases in patients. In spite of these drawbacks, inactivated vaccines are still popular. For example, BBIBP-CorV, which is an inactivated vaccine against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) produced by Sinopharm, has been shown to be effective in preventing infection and death caused by COVID-19 (Rearte et al., 2022).

#### 1.2.2.2. Viral vector vaccine

Instead of using the original pathogen, an additional virus encoding target antigen is used as a viral vector vaccine. A wide range of viruses including adenoviruses, lentiviruses, and poxviruses were genetically engineered to work as viral vector vaccines.

The viral vector used for expressing antigen can be either non-replicable or replicable but low-pathogenic. Both replicable and non-replicable viral vector vaccines can induce strong immunogenicity (Travieso et al., 2022). In terms of replicable viral vector vaccines, since viral vectors can replicate spontaneously, they require less dosage and can trigger immune responses lasting for a longer time. The gene of viral vectors can be modified to be replication-deficient resulting in non-replicable viral vectors. Compared with replicable viral vector vaccines, non-replicable viral vector vaccines are generally safer. During the pandemic of COVID 19, Oxford and AstraZeneca codeveloped AZD1222 which is a viral vector vaccine using a non-replicable viral vector (chimpanzee adenovirus).

#### 1.2.2.3. Subunit vaccine

Subunit vaccines, as indicated by the name, only contain subunits of viruses such as purified antigens alone instead of whole pathogens. The immune response can be triggered by subunit vaccines as well depending on the antigen used (Vartak and Sucheck, 2016).

Virus-like particle (VLP) vaccines are another example of the subunit vaccine. As shown by **Figure 2(c),** VLP vaccines resemble the protein coats of specific viruses without containing the genetic substance. They are able to trigger both humoral and cellular immune responses (Fuenmayor et al., 2017). This type of vaccine is generally safer than others since it is deficient in genetic material. There are several VLP vaccines that have been approved such as Cervarix® and Mosquirix™, while some other candidates are undergoing different phases of clinical trials (Nooraei et al., 2021). Therefore, VLP vaccines provide another promising choice for vaccine production.

#### 1.2.2.4. Nucleic acid vaccine (DNA/RNA vaccine)

The mechanism of inducing immune responses by DNA and RNA vaccines is different from the aforementioned vaccines. As genetic materials, DNA and RNA carrying genetic information of specific antigens are administered so that antigens can be produced by cells. As a result, the immune response can be triggered.

Message RNA (mRNA) vaccines have played a key role in the war against SARS-CoV-2. BNT162b2 and mRNA-1273, which are two mRNA vaccines produced by Pfizer-BioNTech and Moderna respectively, have over 77% effectiveness against symptomatic infections caused by SARS-CoV-2 (Chuenkitmongkol et al., 2022).

In contrast to the success of mRNA vaccines, there is no DNA vaccine approved by regulatory agencies at present. However, many candidates of DNA vaccine are undergoing different clinical trial phases (Gary and Weiner, 2020). Compared with mRNA vaccines, DNA vaccines have the

advantage in ease of storage and transportation because mRNA is easy to be degraded while DNA is relatively stable due to its double helix structure.

### 1.2.3. Production of vaccines

Conventionally, vaccines are manufactured using an egg-based production process. In recent years, there has been a transition from egg-based production to cell-based production or even cell-free (in vitro transcription for mRNA) production process. These manufacturing processes have their own benefits, limitations, and challenges which are discussed in the following section.

#### 1.2.3.1. Egg-based production

In egg-based production, fertilized hens' eggs are inoculated with live viruses so that viruses can replicate during embryo development. It was the most common method in the past. The production process has been optimized for decades. Therefore, egg-based production has a higher downstream yield compared with cell-based production.

On the other hand, there are also some limitations of egg-based vaccine production. The efficiency of egg-based production is low and sometimes it may need 2 eggs for one dose of vaccine (Tree et al., 2001). Since a large number of eggs are required, the egg supply has to be carefully planned and guaranteed in advance, which might take a few months for preparation. Besides, when a pandemic happens, egg-based production is unsatisfactory to meet the demand for vaccines. Other limitations of egg-based production include inevitable contamination due to the nature of eggs as well as risk for egg allergy patients (Hütter et al., 2013).

#### 1.2.3.2. Cell-based production

Instead of inoculating viruses to eggs, the cell-based process produces vaccines by either inoculating cells or expressing subunits of viruses with cells that are genetically modified. Compared with the egg-based production process, cultivating well-characterized cell lines with chemically defined media improves the batch-to-batch consistency of the product and reduces the

risk of contamination. Cell-based vaccine production also allows a faster response to pandemics. Various types of cells can be used for vaccine production and are discussed as follows.

#### 1.2.3.2.1. Microbial expression system

Microorganisms such as E. coli and yeast are preferable choices for biopharmaceutical production. Some of the subunit vaccines are produced by the microbial system. For instance, an E. coli-derived VLP vaccine was developed targeting hepatitis E (Li et al., 2015). Besides, yeast is also capable of producing the VLP vaccine, which makes it a popular workhorse for the expression of various antigens to assemble subunit vaccines such as hepatitis B virus (HBV) and Human papillomavirus (HPV) (Roohvand et al., 2017).

Microorganisms have several features that are favorable for the expression of vaccines including fast-growing, lower cost, and high yield. Besides, the production process is easy to scale up. However, due to their relatively simple structure, they are not capable of performing complex post-translational modifications (PTMs) which limits the types of vaccines that can be produced by them.

#### 1.2.3.2.2. Baculovirus-insect cell expression system

Baculovirus-insect cell expression system is another choice of vaccine production mainly for manufacturing VLP vaccines. Briefly, recombinant baculoviruses are constructed to carry the genes of interest (GOI) and subsequently utilized to infect insect cells so that VLPs can be harvested by the end of cell culture.

An advantage of baculovirus is its large genome size, which allows the insertion of large amounts of foreign genes (Van Oers, 2006). The baculovirus-insect production process also shows a few desired features for commercial manufacturing such as scalability and low cost (Felberbaum, 2015). However, PTMs such as glycosylation and phosphorylation produced by this platform can be different from those produced in mammalian cells (Brooks, 2004). In spite of the drawback, there are still many VLP vaccine candidates produced by the baculovirus-insect platform that are

undergoing different phases of clinical trials, and some of them have already been approved (Contreras-Gómez et al., 2014).

#### 1.2.3.2.3. Mammalian cell-based production system

Mammalian cells are prevalent options for the production of biopharmaceuticals because they are capable of performing complex PTMs. In the vaccine production, mammalian cells are also used. They can be host cells for virus infection and replication. In addition, they can also work as an expression platform to express antigens or VLPs that require complex PTMs. However, compared with other production systems, mammalian cells are characterized as slow-growing and the culture media is more expensive. In addition, the scale-up of their production process is challenging as well. Therefore, the process development of mammalian cell-based vaccine production is more time-consuming, expensive, and difficult.

As listed in **Table 1**, MDCK and Vero cells are used for the production of influenza and Ebola vaccines respectively. They are two prevalent mammalian host cells for vaccine production. Virus isolates from MDCK and Vero cell cultures are able to meet the many requirements including antigenic and genetic stability after several passages as well as high titer (Donis et al., 2014). In addition to MDCK and Vero cells, Chinese hamster ovary (CHO) cells are another important cell line, which will be introduced in detail in the later part. They can also be used for the expression of vaccines such as VLPs.

### 1.2.3.3. Cell-free production (*in vitro* transcription)

In vitro transcription is used for the upstream production of mRNA vaccines without the involvement of cells. As shown in **Figure 3** below, similar to the large-scale cell culture, bioreactors are applied for conducting in vitro transcription. DNA and nucleoside triphosphates (NTPs) work as templates and substrates respectively while the reaction is catalyzed by RNA polymerase. Other process parameters such as temperature, pH, and cofactor need to be optimized and controlled as well.

Capping for the 5' terminal of the mRNA after transcription is an important modification because it can initiate translation (Rosa et al., 2021). It can be done with the help of a commercial capping reagent CleanCap® or through an enzymatic reaction using the vaccinia capping system (Shuman, 2015).

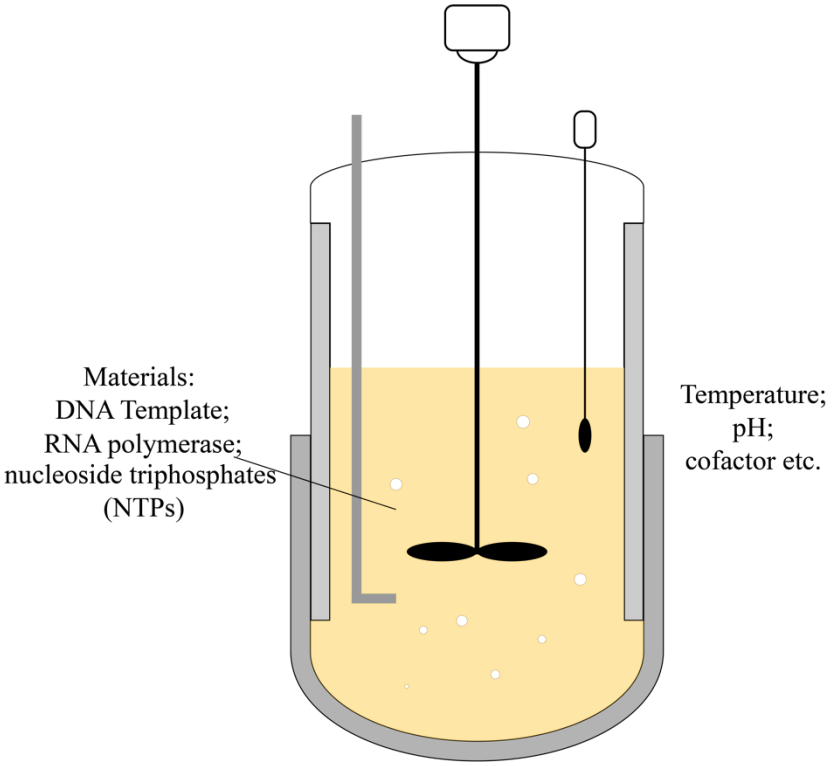


**Figure 3**: In vitro transcription for the mRNA vaccine production.

Two of the main advantage of mRNA vaccines is the low cost of their production and short process duration. In terms of upstream processing, raw materials for in vitro transcription are much cheaper than that of cell culture which uses media containing many components for the nutrient supplement of cells. Besides, the process control and scale-up are easier to achieve because the process is free of cells. In addition, the upstream production of mRNA only takes a few hours while it might take several weeks for a batch of mammalian cell culture. According to the estimation of Kis et al. (2020), the production scale of mRNA vaccines required to meet the global demand for COVID-19 vaccination is substantially smaller than that of conventional vaccines. Meanwhile, downstream processing can also benefit from the cell-free manner. There is no need to remove host cell proteins that generate a great burden for downstream purification.

The main challenge of mRNA vaccine production also comes from downstream. The delivery of mRNA vaccines mainly depends on lipid nanoparticles (LNPs), which generate an additional unit operation LNP encapsulation formulation in the downstream process. This unit operation is still not cost-effective and needs to be optimized at present (Rosa et al., 2021).

## 1.3. Chinese hamster ovary (CHO) cell line

### 1.3.1. General information of CHO cell

Chinese hamster ovary (CHO) cells, as indicated by their name, were derived from the ovary of a Chinese hamster in the late 1950s (Wurm and Wurm, 2021). They have become the dominant workhorse for the production of biopharmaceuticals since the 1980s when suspension culture of immortalized mammalian cells with large-scale bioreactors started to be used (Wurm and De Jesus, 2016).

As shown by **Figure 4**, between 1987 and 2021, CHO cell lines are responsible for 81% of approved biopharmaceuticals that are produced by mammalian cells (Al-Majmaie et al., 2021). Compared with CHO cell lines, other cell lines only account for a small amount of biopharmaceuticals. For example, HEK293, a cell line derived from humans, only occupies 2% of the approved biopharmaceuticals. Currently, there are several well-established CHO cell lines available such as CHO-S, CHO-K1, and CHO-DG44.

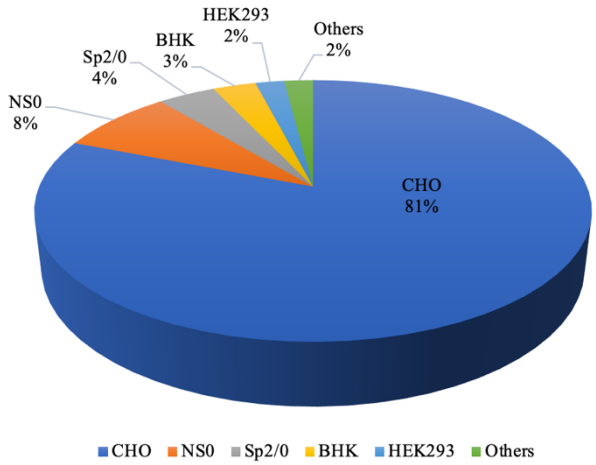


**Figure 4**: Percentage of various types of commercially used mammalian cell lines from 1987 to 2021 based on Al-Majmaie et al. (2021)

Monoclonal antibodies (mAbs) are one of the most important biopharmaceutical products. The market size of mAbs increased dramatically in recent decades. The market size was around $0.25 billion in 1997 initially while it is predicted to be $300 billion by 2025 (El Abd et al., 2022). CHO cells are also responsible for the production of mAbs.

More concretely, **Table 2** shows a few examples of the mAbs produced by CHO cells (information from the European Medicines Agency (EMA) and the United States Food and Drug Administration (FDA)). Numerous diseases can be treated with mAbs. For instance, Trastuzumab is a humanized IgG1 that can be used for the treatment of metastatic breast cancer whose tumor overexpresses human epidermal growth factor receptor 2 (HER2).

In addition to the production of mAbs, CHO cells can also be applied for the production of recombinant proteins for other purposes. For example, Nyon et al. (2018) established a stable CHO cell line expressing the antigen of Middle East respiratory syndrome coronavirus (MERS-coronavirus) that can work as a subunit vaccine. Therefore, CHO cells are versatile production platforms for the biopharmaceutical industry.

**Table 2**: Examples of mAbs produced by CHO cells.

| Product name | Targeting disease | Manufacturer |
|---|---|---|
| Brodalumab | Moderate to severe plaque psoriasis | AstraZeneca |
| Mepolizumab | Severe eosinophilic asthma, hypereosinophilic syndrome and Churg-Strauss Syndrome | GlaxoSmithKline (GSK) |
| Trastuzumab | HER2-positive metastatic breast cancer | Roche |
| Vedolizumab | Ulcerative colitis and Crohn's Disease | Takeda Pharma |

Besides, CHO cells have the potential to express VLP vaccines as well because VLPs are also a form of recombinant proteins. However, reports of producing VLP vaccines using CHO cell culture are still deficient.

### 1.3.2. Advantages of CHO cell lines

CHO cell lines have many advantages over other expression platforms, which makes them the most favorable choice in the commercial production of biopharmaceutical products. These advantages will be discussed in the following section.

One of the main benefits of using CHO cells is their ability to grow in chemically defined serum-free media (Bandaranayake and Almo, 2014). Since the media is chemically defined, the batch-to-batch variation of raw materials can be reduced and therefore enable higher batch-to-batch consistency of the product. Besides, serum-free media also outperforms serum-containing media by reducing the burden of downstream purification because there are more components such as proteins in the serum that need to be removed during purification steps. Serum-free media is also favorable for regulatory agencies as no unknown material is contained (Lai et al., 2013).

Another advantage of CHO cells is their ability to adapt to suspension culture. Generally, there are two types of cell culture, namely adherent culture and suspension culture. Adherent culture refers to the cultivation in which cells attach to the surface of a vessel while cells are freely suspended in agitated media in suspension culture. Suspension culture outperforms adherent culture in terms of scalability (Kim et al., 2012). The scale of suspension culture can be achieved by increasing the size of the bioreactor. However, for adherent culture, the surface-to-volume ratio decreases as the size of the bioreactor is increased. In other words, there is less surface area per cell available, which is adverse to cell growth. To overcome this limitation, microcarriers can be added to increase the surface area. However, the usage of microcarriers also generate additional cost and these microcarriers need to be removed in downstream, which makes adherent culture less cost-effective.

After development for decades, the yield of CHO-based production has been improved significantly resulting in a relatively high yield among mammalian cells. Thanks to the development of media compositions, gene engineering, and advanced process control, the volumetric productivity of CHO culture with bioreactors increased from 50-100 mg/L in the 1980s to 1-10 g/L now (Wurm and De Jesus, 2016). Other novel techniques such as gene amplification technology also contribute to the high yield of CHO cells (Lai et al., 2013).

Moreover, the safety and efficacy of the product can also benefit from the CHO expression platform. More specifically, most human viruses are not capable of replicating in CHO cells. As a result, the biosafety risk is reduced (Lai et al., 2013). Moreover, CHO cells can produce proteins with complex human-like PTMs which are critical for the efficacy of the pharmaceutical products (O'Flaherty et al., 2020).

### 1.3.3. Challenges of CHO cell lines in biopharmaceutical production

In spite of the aforementioned advantages, the application of CHO cell lines is still facing a number of challenges that need to be overcome.

Similar to other mammalian cell lines, CHO cell lines feature longer doubling time compared with non-mammalian cell lines. Besides, the medium for CHO culture is more expensive. Therefore, the production process as well as process development are more time-consuming and expensive than other expression platforms.

In addition, the scale-up of CHO cell lines can be challenging as CHO cells are shear sensitive. In large-scale production, to achieve a homogeneous environment in the bioreactor, agitation is necessary. On the other hand, shear stress generated by the impeller agitation can negatively affect cell growth and production.

Gene instability is another aspect worth considering when using CHO-based production. Due to the genome plasticity of CHO cells, these cells suffer from chromosomal rearrangements (Dahodwala and Lee, 2019). Such rearrangements might lead to unstable product titer and quality, which are not desired by regulatory agencies and are adverse to drug approval.

## 1.4. Modeling of bioprocess

### 1.4.1. General information on bioprocess modeling

Mathematical modeling of bioprocess refers to describing phenomena in bioprocess with mathematical languages. Bioprocess benefits from modeling in many aspects including improvement of cost efficiency, risk mitigation, and realization of automation. Besides, thanks to the development of machine learning, many advanced algorithm is developed and applied to model bioprocess. Therefore, modeling of bioprocess is a popular field now.

Currently, bioprocess is moving toward industry 4.0 which requires the digitalization of factories to enable intelligent manufacturing. Digital twin (DT) is a key part of the implementation of Industry 4.0. A brief workflow of digital twin adapted from Gargalo et al. (2020) is shown in **Figure 5**. In the beginning, data in the real world is collected by sensors and then transferred to the database. After that, data can be analyzed and used to update models subsequently. The newly updated models are able to make decisions based on both the current status of the real world and

prior knowledge such as historical data in the data space. These models can perform multiple tasks such as enabling process optimization in silico and computing optimal control signals. After the computation based on updated models in the virtual world, actuators such as controllers will be applied to affect the real world and new measurements will be collected in response to the actions. This iteration will continuously and automatically facilitate bioprocess in terms of process optimization, decision-making, and online process control. The availability and quality of models are key to the success of the implementation of the DT. Consequently, modeling will continue playing important roles in the bioprocess in the foreseeable future.

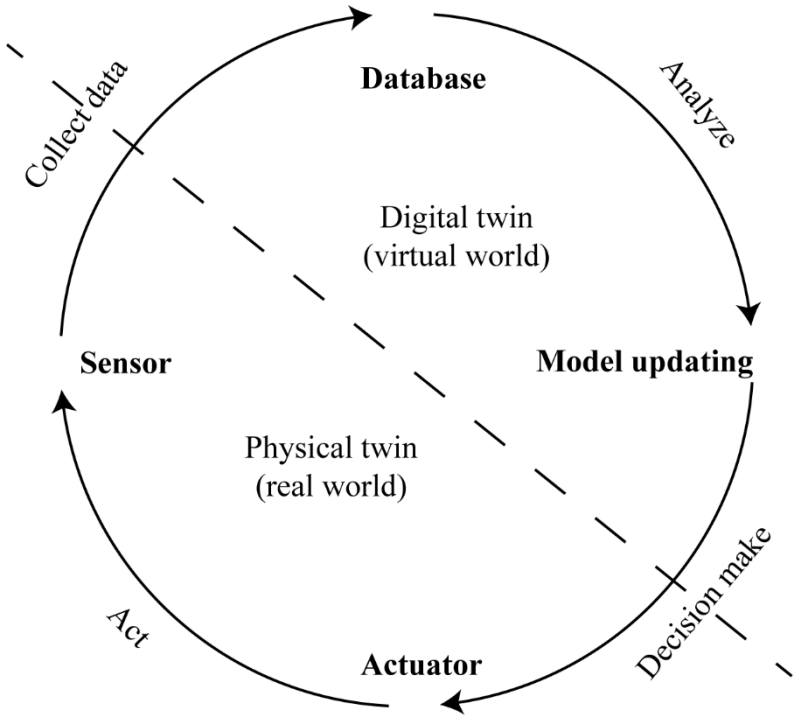


**Figure 5**: Brief workflow of digital twin (adapted from Gargalo et al. (2020)).

### 1.4.2. Different types of models in bioprocess

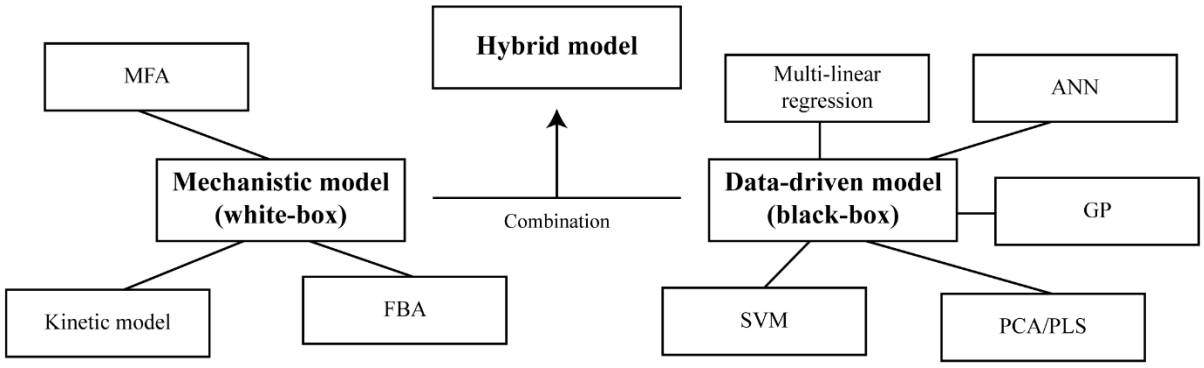


**Figure 6**: Various types of models applied in the upstream bioprocess. MFA: metabolic flux analysis; FBA: flux balance analysis; ANN: artificial neural network: GP: Gaussian process; SVM: support vector machine; PCA: principal component analysis; PLS: partial least squares.

In general, models in bioprocess can be classified into three categories. Mechanistic models, which are also called white-box models, are based on biological, chemical, and physical knowledge to explain the phenomenon in bioprocess. In contrast, the data-driven model is considered as a black box that purely depends on the data without consideration of the mechanism. These two types of models both have their own strengths. To combine their strength together, hybrid models are also proposed to facilitate the development of bioprocess. A few examples of various types of models applied in the upstream bioprocess are shown in **Figure 6** such as kinetic models. Besides, machine learning algorithms such as principal component analysis and partial least squares also serve as powerful tools in the modeling of bioprocess. Some of these examples are discussed in the following section.

#### 1.4.2.1. Kinetic model

Kinetic models are a type of mechanistic model. They are derived based on biological, chemical, and physical knowledge. They are able to reveal the kinetics of upstream cell culture by building up relationships between cells, metabolites, and products. They are able to generate a deeper understanding of the process and enable quantitative analysis of the data (N. P. Shirsat et al., 2015).

This type of model is widely applied for the facilitation of the optimization of bioprocess. For example, Xu et al., (2019) used a kinetic model to optimize the temperature shift strategies of

mAb1-producing CHO cell culture and increased the final product titer by around 10% after 14-day cultivation. In addition, a large kinetic model integrating upstream cell culture and downstream purification was developed for the comparison between batch and continuous mode for mAb production at a system level (Badr et al., 2021). A model-based optimization was applied to optimize the seeding strategy as well (Kern et al., 2016).

In addition, online process monitoring and control are desired from the standpoint of regulatory agencies through the process analytical technology (PAT) initiative. Kinetic models can also contribute to better monitoring and control of bioprocess. Frahm et al. (2002) constructed a controller as a valuable tool for bioprocess control with the help of the kinetic model built by them. Besides, kinetic models can also be applied to facilitate the scale-up of bioprocess (Arndt et al., 2021). Therefore, kinetic modeling is a powerful and versatile tool in bioprocess. Compared with data-driven models, these models generally require less data and can perform extrapolation better because prior knowledge and the underlying mechanism have been integrated into the model structure.

#### 1.4.2.2. Data-driven models

In contrast to models based on mechanism, data-driven models purely depend on the data itself serving as black boxes. Different data-driven models have their own strengths and are applied in different scenarios. Models can be constructed even when prior knowledge and underlying mechanisms are absent.

With the development of analytical methods in recent years, more and more data have been revealed such as multi-omics and spectrum. This kind of data features highly dimensional but limited sample size available due to the budget. Principal component analysis (PCA) and partial least squares (PLS) are two suitable modeling methods in such kind of situation. Both of them can handle data with a huge number of features. PCA can reduce the dimension of data in order to visualize, analyze, and group the data set. PCA has been widely used in multi-omics data analysis (Meng et al., 2014; Verhoeckx et al., 2004). Similarly, PLS is another tool for dimension reduction but as a regression algorithm. They can build up mathematical relationships between highly

dimensional inputs and user-defined outputs. A novel media optimization strategy was developed based on PCA and PLS (J. K. Hong et al., 2022). PLS can also combine with Raman spectroscopy to monitor the bioprocess (Esmonde-White et al., 2017).

Artificial neural network (ANN) is the dominant algorithm in the field of machine learning. The application of ANN in bioprocess is widely explored as well. For example, ANN was used to predict protein glycosylation produced by CHO cell culture (Kotidis and Kontoravdi, 2020). However, ANN requires a large amount of data because it has many parameters to tune, which is difficult in bioprocess due to the cost of conducting experiments. Another challenge is their poor performance in extrapolation (Mowbray et al., 2021).

#### 1.4.2.3. Hybrid models

In bioprocess, the underlying mechanism is not always clearly determined. In addition, the data availability is quite limited. Therefore, hybrid models that combine mechanistic and data-driven models are developed. Data-driven models are applied when a mechanism is not available while mechanistic models can reduce the data size required and achieve better extrapolation. For instance, Luo et al. (2021) built up a hybrid model by combining a kinetic model with a data-driven model to model the effects of amino acids and copper addition during CHO cell culture on mAb productivity and glycosylation.

### 1.4.3. Methodologies for model-based multicriteria optimization

Since models are able to provide predictions, they are usually used to perform optimization for the bioprocess. In the field of bioprocess, there are many objectives to be optimized including product yield, product quality, cost efficiency, process consistency, and risk mitigation. To perform multicriteria optimization, following methodologies can be potentially integrated to model simulation.

#### 1.4.3.1. Desirability methodology

Desirability methodology is a technique used in engineering. It is able to optimize parameters to meet several user-defined criteria simultaneously. For instance, Möller et al. (2019) optimize an antibody production process in CHO cells by maximizing viable cell density and antibody concentration as well as minimizing lactate concentration with the help of desirability methodology. To conduct desirability, objectives need to be defined first. Based on requirements, goals of different objectives including maximizing, minimizing, and constraining to specific range were selected. According to user-defined desirability function and bounds, individual desirability can be computed and normalized. Since different individual desirability value is normalized, desirability methodology weights various objectives at different scales the same. These individual desirability values will be multiplied together generating an overall desirability to optimize parameters in the process.

#### 1.4.3.2. Pareto optimization

Pareto optimization is another multicriteria optimization methodology which is usually used in economics. This method can be applied to both continuous and discrete problems. The main concept of Pareto optimization is to find solutions whose at least one single objective cannot be improved without worsening any other objectives. The main limitation of Pareto optimization is that it would provide multiple solutions when trade-offs exist between objectives. It is confusing for the decision-maker since it cannot provide a dominating solution for the given problem. Therefore, in this study, desirability methodology was selected instead of Pareto optimization to balance the trade-off between product yield and quality in chapter 2. Besides, Pareto optimization is computationally expensive, which is another limitation of it.

## 1.5. Overview of this thesis and relationship between chapters

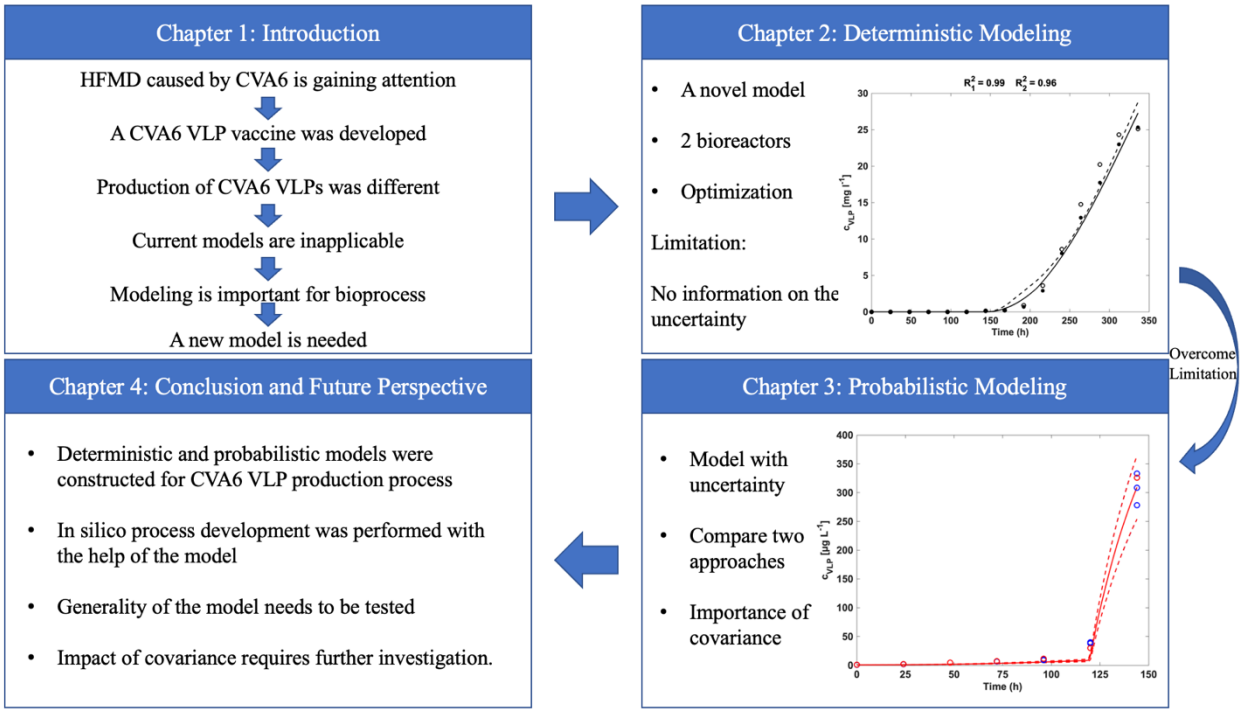


**Figure 7**: Outline of this thesis

Outline of this thesis is shown in **Figure 7**. In Chapter 1, the background, importance, objectives, and novelty of this research were introduced. As mentioned before, HFMD caused by CVA6 has become prevalent in recent years. Also, CVA6 is able to affect adults. However, vaccines targeting CVA6 are still deficient and the vaccine targeting another serotype cannot provide cross-protection. Therefore, a CVA6 vaccine needs to be developed.

By comparing different types of vaccines available now, the VLP vaccine was chosen because of its improved safety over conventional vaccines. To produce the CVA6 VLP vaccine, CHO cell culture was used due to the aforementioned merits of CHO cells. CHO cells are capable of performing human-like PTMs. Besides, they can grow in the chemically defined serum-free medium in a free suspension manner, which is desired for industrial production. Moreover, most human viruses cannot propagate in CHO cells reducing the biosafety risk.

To produce CVA6 VLPs in CHO cells efficiently, a kinetic model is desired especially during the era of Industry 4.0. Kinetic models can enhance the understanding of the production process.

They can also guide process optimization. Kinetic models also have the potential to be combined with data-driven to perform hybrid modeling. They are able to contribute DT as well. Moreover, compared with data-driven models, they require less data. Even though a lot of kinetic models have been constructed for CHO cell culture, these models failed to capture the kinetics of VLP production using CHO cell culture. More concretely, previous models used a parameter termed specific productivity to correlate product concentration with viable cell density. However, for non-enveloped VLP production, cell lysis is needed to release the intracellular product (Cervera and Kamen, 2018). Based on our observations, VLP concentration is strongly correlated with cell death and lysis instead of viable cells. Therefore, I constructed a novel kinetic model for the CVA6 VLP production process in CHO cell culture. To the best of my knowledge, this is the first time that a model for VLP production was reported.

In Chapter 2, I developed the protocol for cell lysis quantification in the batch culture of CVA6 VLP-producing cells firstly because cell lysis was important for the product release and it cannot be detected by image-based analyzers directly. Then, thanks to the samples provided by Dr. Thao Bich Nguyen, I performed cell lysis quantification for the fed-batch culture of CVA6 VLP-producing CHO cells with bioreactors. After combining cell lysis data with the cell culture data kindly provided by Dr. Thao, I developed a novel kinetic model for the CVA6 VLP vaccine production process in CHO cells based on my understanding of the process and equations from previous studies. Then, the kinetic model was trained with experimental data from 2 bioreactors and its performance was evaluated. After that, a series of analysis was performed to facilitate the process development of the CVA6 VLP vaccine *in silico*. Finally, the usefulness of this kinetic model was discussed.

During the procedure of data analysis and model construction, the limitation of this deterministic model was noticed as well. I found that even under the same culture conditions, cell culture performances differed. It is common because there are a lot of uncertainties involved in the cell culture such as experimental error, variations in raw materials as well as measurement error. Batch-to-batch consistency is an important aspect of the process validation of bioprocess in order to comply with Good Manufacturing Practice (GMP). Thus, quantification of the batch-to-batch variations is highly desired. The batch-to-batch variation cannot be captured by deterministic

models, which is a limitation of deterministic models because it can only simulate a single value at a given point without providing any information on uncertainties. Therefore, I move to probabilistic modeling in Chapter 3 to quantify uncertainties of the CVA6 VLP vaccine production process.

Consequently, in Chapter 3, I perform quantification of uncertainties between cell cultures for the CVA6 VLP vaccine production with the help of the previously constructed model. To do so, I conducted 4 batch cell cultures in parallel, 3 for model calibration and 1 for model test. After acquiring the data, I updated the model to fit the batch culture. Then, two approaches namely frequentist approach and Bayesian approach were applied based on the newly applied model to quantify the uncertainties of cell culture performances. Probabilistic models were constructed providing simulation of variables including both mean and distribution at different time points. The uncertainty level between batch and fed-batch culture was also compared. Finally, two approaches were compared and their strengths and limitations were discussed.

Finally, in Chapter 4, the main findings of this research were summarized and their potential applications in the facilitation of bioprocess development were discussed. Besides, a discussion about the limitations of this work was included as well. Finally, the future perspective was outlooked.

Chapter 2:

# Deterministic modeling of a CVA6 VLP vaccine production process using CHO cell culture

# 2. Chapter 2: Deterministic modeling of a CVA6 VLP vaccine production process using CHO cell culture

## 2.1. Introduction

As introduced in Section 1.4.2.1., kinetic models are powerful and versatile tools that can serve bioprocess in terms of process optimization, monitoring, control as well as scale-up. Besides, they are also key components for enabling DT. Thus, a kinetic model for the production process of the CVA6 VLP vaccine in CHO cells is highly desired. Conventionally, product concentration in the cell culture is directly correlated to viable cell density through a parameter termed specific productivity. Specific productivity represents a unit mass of the product produced by a single cell within a unit time. However, for the production of non-enveloped VLPs, a cell lysis step is required to release the product (Cervera and Kamen, 2018). Besides, based on previous observations, VLP concentration in the CHO cell culture was strongly correlated to cell death and lysis, instead of living cells. Therefore, I constructed a novel kinetic model of CVA6 VLP vaccine production in CHO cell culture. Then, I evaluated the constructed model and performed a series of analysis to facilitate *in silico* process development of the CVA6 VLP production process.

Since VLP concentration is correlated to cell death and lysis, it is necessary to quantify the cell lysis during cultivation. At present, cell counting depends on image-based analysis that cannot count lysed cell. Therefore, indirect methods for cell lysis quantification are needed. According to the research of Klein et al. (2015), double stranded DNA (dsDNA) and lactate dehydrogenase (LDH) can be used as indicators for the cell death and lysis quantification. As shown by **Figure 8**, cells may undergo apoptosis, autophagy or necrosis becoming dead cells or lysed cells. Also, dead cells can be lysed and become cell debris. During both procedure of cell death and lysis, intracellular components are released including dsDNA and LDH. As a result, cell death and lysis can be quantified indirectly by measuring dsDNA or LDH concentration in the supernatant of cell broth. Based on this mechanism, I started with the batch culture of CVA6 VLP-producing CHO cells to test two indicators. CHO cells were cultivated for one week and samples of medium

supernatant were taken. Subsequently, samples were incubated at 37 °C for several days prior to dsDNA and LDH measurements in order to determine the thermal stability of indicators.

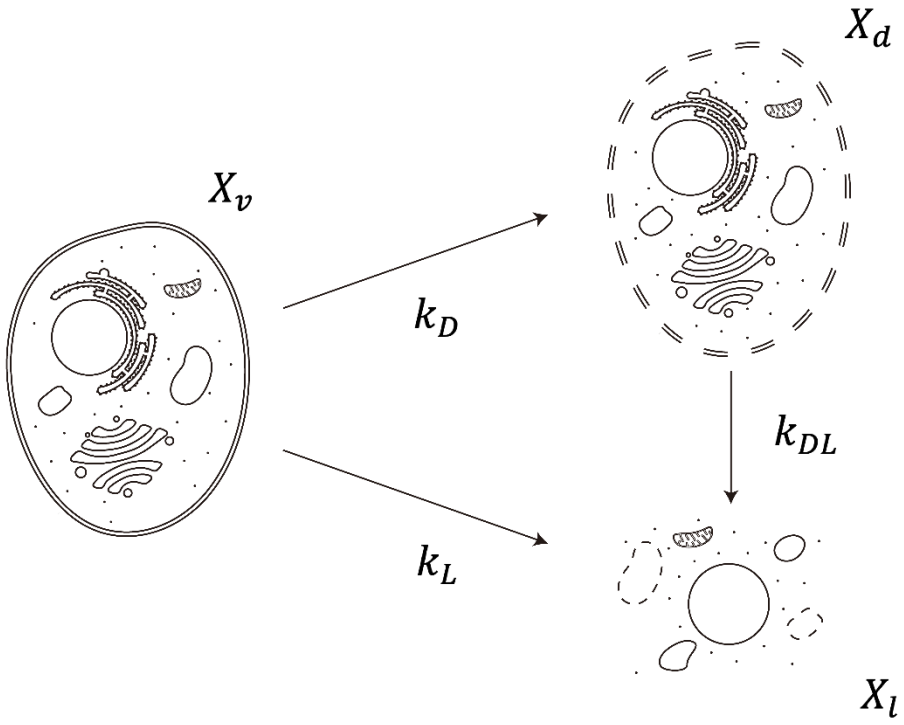


**Figure 8**: Cell death and lysis pathway with corresponding kinetic parameters (adapted from Kroll et al., 2017).

After that, fed-batch cultivations of CVA6 VLP-producing CHO cells with two bioreactors were performed by Dr. Thao Bich Nguyen. Cells were cultivated for two weeks and samples were taken every 24 h for further analysis including cell counting, measurements of metabolites concentrations and VLP concentration measurement, which is also conducted by Dr. Thao. In addition, similar to the batch culture, I performed dsDNA and LDH measurements for cell lysis quantification. During the cell culture, temperature shift was performed to improve the yield. Besides, cells were cultured with or without pH downshift in order to determine the effects of pH downshift.

Subsequently, I constructed a kinetic model for the CVA6 VLP vaccine production process in CHO cells based on the data acquired by myself and kindly provided by Dr. Thao. To do so, equations were either adapted from previous literature or newly proposed based on my

understanding of the process. Then, the accuracy of this model was evaluated. After that, sensitivity analysis was performed to examine the impact of small variations in raw materials or model parameters on the VLP yield. Moreover, cell culture performances under different pH downshift timing were predicted by the model to optimize the VLP yield. Finally, multiple-objective optimization was conducted by combining desirability methodology with the model simulation in order to balance the trade-off between product yield and quality.

## 2.2. Materials and methods

### 2.2.1. Batch culture for cell lysis indicator selection

#### 2.2.1.1. Cell line, medium, and culture conditions

The cell line used in this research is kindly provided by BIKEN Group Japan. It is a CHO cell line expressing CVA6 VLPs. To cultivate cells, CD FortiCHO™ medium (Thermo Fisher Scientific, Waltham, MA, USA) was used supplemented with 8 mM glutamine (Fujifilm Wako, Osaka, Japan) initially. Cells were thawed from frozen first and cultured with shaking flasks. A humidified incubator was utilized for the culture operated at 37 °C, 8% CO2, and 140 rpm. Cells were passaged every 72- 96 h before the full-duration culture for one week. During the full-duration cultivation, samples were taken every 24 h for subsequent thermal stability tests. The supernatant of samples was collected by centrifuging samples at 14000 g for 5 minutes for the cell removal.

#### 2.2.1.2. Thermal stability test and cell lysis indicator measurements

To investigate whether cell lysis indicators were degraded at 37 °C during the cell culture or not, thermal stability test was conducted. Supernatant samples taken at 72, 96, 120, and 144 h were incubated at 37 °C for 0 h, 24 h, 48 h and 72 h.

After that, dsDNA and LDH concentrations were measured for the incubated samples. Briefly, dsDNA measurement was performed with a Quant-iT PicoGreen dsDNA Assay Kit (Thermo

Fisher Scientific) following the protocol. Samples were diluted and mixed with the assay reagent in a 96-well plate. Similarly, $1 \times 10^5$ μg/L Lambda DNA Standard was diluted and mixed with the assay reagent in the same plate to plot the standard curve. After a two-minute room temperature incubation, fluorescence at 492/530 nm was measured by a VICTOR Nivo Multimode Microplate Reader (PerkinElmer, Waltham, MA, USA).

In terms of LDH, a Lactate Dehydrogenase Activity Assay Kit (Sigma-Aldrich, MO, USA). In brief, samples were diluted properly and mixed with LDH assay reagent in a 96-well plate. 1.25 mM NADH standard was prepared in the same 96-well plate as well. After incubating for 2 minutes, absorbance at 450 nm was measured. Then, samples and the standard were incubated at 37 °C for 5 minutes prior to the absorbance measurement at 450 nm again. This step was repeated several times until the absorbance was beyond the measurement range.

### 2.2.2. Fed-batch culture with two bioreactors

#### 2.2.2.1. Cell line, medium, and preculture

Cell line, medium, and culture conditions of the preculture were the same as described in Section 2.2.1.1. Cells were thawed and expanded in shaking flasks. Before the inoculation to bioreactors, cells were subcultured every 72-96 h.

#### 2.2.2.2. Fed-batch laboratory-scale bioreactor cultivation

Two 2 L stirred tank bioreactors (STR1 and STR2, ABLE Biott, Tokyo, Japan) were used to perform cell culture in a fed-batch manner. Initial cell concentration was set at $4 \times 10^8$ cells/L and the working volume was 0.8 L. Dissolved oxygen was controlled at 30% by sparging pure $O_2$. Impeller speed was 80 rpm. Besides, 0.1 L/min and $8 \times 10^{-3}$ L/min were set for the air and $CO_2$ flow respectively. Temperature was kept at 37 °C prior to the temperature shift to 32 °C after 120 h. To achieve the pH control, $CO_2$ and 1 M $NaHCO_2$ were used to adjust the pH. In order to investigate the effect of pH shift on the VLP production, cell cultures were conducted with and without a pH downshift. Specifically, in the control experiment (STR1), pH was kept at 7.15

throughout the cell culture. In the treatment experiment (STR2), pH was controlled at 7.15 during the first 720 h and changed to 6.75 until the end of cell culture.

To avoid foaming during the cell culture, FoamAway™ (Thermo Fisher Scientific) was added to bioreactors. Since 72 h, $1.6 \times 10^{-2}$ L of feeding medium accounting for 2% of the initial working volume was fed into bioreactors daily. A 300 g/L glucose solution was also added in order to maintain the glucose concentration above 2 g/L. During the cultivation, $4 \times 10^{-3}$ L of sample was taken from STR1 and STR2 for subsequent analytical methods.

#### 2.2.2.3. Cell counting and metabolite measurements

A ViCell machine (Beckman Coulter, Brea, CA, USA) was utilized to quantify viable cell concentration and viability. In addition, concentrations of metabolite (glucose, glutamine, lactate, and ammonia) were measured by a BioProfile 400 Automated Chemistry Analyzer (Nova Biomedical, Waltham, MA, USA).

#### 2.2.2.4. dsDNA and LDH thermal stability test and quantification

Similar to the procedure described in Section 2.2.1.2., the thermal stability of dsDNA and LDH in the samples from fed-batch culture in bioreactors was studied. Samples taken at 72, 120, 144, and 240 were incubated at 37 °C for 0, 24, 48, and 72 h for dsDNA quantification. Samples taken at 192, 240, 288, and 336 h were incubated at 37 °C for 0, 24, 48, 72, and 96 h for LDH quantification.

Measurement of dsDNA followed the same procedure as described in Section 2.2.1.2. It was assumed that cell growth stopped after the exponential phase. Thus, based on mass balance, intracellular dsDNA content can be determined by the slope of plotting dsDNA concentration versus viable cell density. The sum of lysed and dead cell concentration can be computed by dividing dsDNA in the medium by intracellular dsDNA content. Besides, dead cell concentration can be calculated with viable cell density and viability. Therefore, lysed cell concentration was calculated by subtracting dead cell concentration from the sum of dead and lysed cell concentration.

LDH concentration was measured with CytoTox 96® Non-Radioactive Cytotoxicity Assay (Promega, WI, USA). Briefly, supernatant samples were diluted properly and mixed with assay reagent in a 96-well plate. Fresh medium was mixed with the reagent for the background signal. After 30 minutes of incubation, stop solution was added to each well following the protocol. Then, the absorbance at 492 nm was measured by a VICTOR Nivo Multimode Microplate Reader.

#### 2.2.2.5. CVA6-VLP measurement

To quantify CVA6 VLP concentration, a sandwich enzyme-linked immunosorbent assay (ELISA) protocol developed by Biken Group Japan was used. Briefly, an anti-CVA6 monoclonal antibody was used to coat 96-well plates prior to the overnight incubation at 4 °C. To block non-specific binding, EzBlock Chemi blocking buffer (Atto, Tokyo, Japan) was added to each well, and plates were incubated at 37 °C for 1 h. Subsequently, washing buffer 0.05% (v/v) Tween 20 in phosphate-buffered saline) was utilized to wash plates thrice. Then, standard or samples were added to each well and incubated at 37 °C for 1h, followed by the same washing procedure again. After that, the addition of a secondary antibody, A horseradish peroxidase-conjugated anti-CVA6 monoclonal antibody, was performed and plates underwent incubation for 1 h at 37 °C followed by three time washes. To visualize the binding, a TMB solution (Surmodics, MN, USA) was added and another 15-minute incubation was conducted at room temperature. BioFX 650 nm Liquid Stop Solution for TMB Microwell substrates (Surmodics) was used as a stop solution to stop the reaction. Finally, a microplate reader (Corona Electrics, Ibaraki, Japan) was applied to measure the absorbance at 650 nm.

### 2.2.3. Model construction for the fed-batch culture

To construct a kinetic model for the CVA6 VLP production, I used ordinary differential equations to describe relationships between cells, major metabolites and VLP concentrations. Volume changes happened frequently during the cell culture. More concretely, volume can be changed due to the addition of feeding medium (16 mL per 24 h), sampling (20 mL per 24 h), and glucose supplement (when necessary). However, these changes were negligible compared with the

total working volume (800 mL). Besides, the difference of model simulation with and without dilution factor was found to be small. Therefore, in the following model construction, it was assumed that there was no volume change in the bioreactor. To take the effect of dilution into consideration, a dilution term could be used. This dilution term was computed by a dilution factor, which is given by dividing volume changes by the working volume, multiplied by corresponding variables as reported by Craven et al. (2014)

$$\frac{dX_t}{dt} = rX_t \tag{2-1-a}$$

$$r = r_{max} \cdot \left(1 - \frac{X_t}{X_{t,max}}\right) \tag{2-1-b}$$

To model total cell density ($X_t$), a logistic equation was used (Martínez et al., 2020; N. Shirsat et al., 2015). This model contains only two parameters ($r_{max}$ for maximum intrinsic growth rate and $X_{t,max}$ for maximum total cell density) so that the risk of overfitting can be reduced especially when the sample size for modeling was limited. Besides, following this equation, total cell concentration kept constant after the exponential phase complying with the previous assumption made in Section 2.2.2.4.

$$\frac{dX_d}{dt} = k_D X_v - k_{DL} X_d \tag{2-2-a}$$

If ($c_{Amm}$>$c_{Amm,cr}$ during dead phase):

$$k_D = k_{D,dead} \cdot \frac{K_{D_{Amm}} + c_{Amm} - c_{Amm,cr}}{K_{D_{Amm}}} \tag{2-2-b}$$

Else:

$$k_D = k_{D,min} \tag{2-2-c}$$

$$\frac{dX_l}{dt} = k_L X_v + k_{DL} X_d \tag{2-3}$$

$$\frac{dX_v}{dt} = rX_t - (k_D + k_L) \cdot X_v \tag{2-4}$$

As shown in **Figure 8**, cell death and lysis pathways were simulated with their corresponding parameters. Viable cells ($X_v$) can become dead cells ($X_d$) or lysed cells directly ($X_l$). Besides, dead cells may also break turning into cell debris. Kinetic equations were built based on equations that were proposed by previous studies (Kontoravdi et al., 2007; Kroll et al., 2017). Some changes were made to fit the cell behavior in our case. According to our observation, dead cells increased significantly when the ammonia concentration was higher than the critical concentration ($c_{Amm,cr}$ = 5 mM) during the dead phase. Two constant parameters ($k_D$ and $k_L$) were used to describe the behavior of lysed cell density that were generated by viable cells or dead cells, respectively. Viable cell concentration was computed by subtracting lysed cell density from total cell density.

$$\frac{dc_{Glc}}{dt} = -(rX_t - (k_D + k_L)X_v)/Y_{X_v/Glc} - m_{Glc}X_v + \frac{V_{F,i}c_M}{V} + \frac{V_{G,i}c_G}{V} \quad (2\text{-}5)$$

Before 120 h:

$$\frac{dc_{Lac}}{dt} = \left((rX_t - (k_D + k_L)\cdot X_v)/Y_{X_v/Glc} - m_{Glc}X_v\right)\cdot Y_{Lac/Glc} \quad (2\text{-}6\text{-a})$$

After 120 h:

$$\frac{dc_{Lac}}{dt} = \left((rX_t - (k_D + k_L)\cdot X_v)/Y_{X_v/Glc} - m_{Glc}X_v\right)\cdot Y_{Lac/Glc} - q_{Lac}X_v \quad (2\text{-}6\text{-b})$$

$$q_{Lac} = q_{Lac,max}\cdot\frac{c_{Lac}}{K_{Lac} + c_{Lac}} \quad (2\text{-}6\text{-c})$$

where $V_{F,i}$ and $V_{G,i}$ represent the volumes of feed medium and glucose solution added at $i$th feeding. Besides, glucose concentrations in the feed medium and glucose solution are nomenclated as $c_M$ and $c_G$, and $V$ is the working volume. The equation simulating glucose concentration (Equation 2-5) was adapted from (Xing et al., 2010). In this equation, glucose consumption was due to cell growth and maintenance while the addition of feed medium and glucose solution raises glucose level. Lactate switch is common in CHO cells. However, the mechanism is still unclear (Hartley et al., 2018). Based on our observations, cells produced lactate before 120 h and switched to lactate production after 120 h. In order to model lactate production as a result of glucose

consumption, a yield coefficient was used ($Y_{Lac/Glc}$). The lactate consumption rate was simulated by a Monod-type term.

If ($c_{Gln}$ > 0 mM):

$$\frac{dc_{Gln}}{dt} = -(rX_t - (k_D + k_{DL}) \cdot X_v)/Y_{X_v/Gln} \quad (2\text{-}7\text{-a})$$

$$\frac{dc_{Amm}}{dt} = (rX_t - (k_D + k_{DL}) \cdot X_v)/Y_{X_v/Gln} \cdot Y_{Amm/Gln} \quad (2\text{-}8\text{-a})$$

If ($c_{Gln}$ = 0 mM):

$$\frac{dc_{Gln}}{dt} = 0 \quad (2\text{-}7\text{-b})$$

$$\frac{dc_{Amm}}{dt} = -(rX_t - (k_D + k_{DL}) \cdot X_v)/Y_{X_v/Amm} + q_{Amm}X_v \quad (2\text{-}8\text{-b})$$

Similar to the yield coefficient of glucose production, glutamine consumption rate was described as a result of cell growth. Besides, ammonia was produced when glutamine was available. If glutamine was used out, ammonia was consumed when cells kept growing. Therefore, I hypothesized that ammonia worked as an alternative nitrogen source during that period. Some other mammalian cell lines also showed this behavior (Lie et al., 2019). In addition to the consumption, the ammonia production rate due to cell maintenance was denoted by $q_{Amm}$. Its value can be different at various pHs.

After 144 h

If ($VLP_{intra} < VLP_{max}$ and pH = 7.15):

$$\frac{dVLP_{intra}}{dt} = q_{VLP} \quad (2\text{-}9\text{-a})$$

If ($VLP_{intra} = VLP_{max}$ or pH = 6.75):

$$\frac{dVLP_{intra}}{dt} = 0 \quad (2\text{-}9\text{-b})$$

$$\frac{dc_{VLP}}{dt} = VLP_{intra} \cdot (k_D + k_L) \cdot X_v \quad (2\text{-}10)$$

Based on observations in the past, cells started to produce VLPs efficiently after 144 h. Therefore, it was assumed that VLPs accumulated intracellularly after 144 h and the amount of it was denoted by $VLP_{intra}$ as shown by equations above. Additionally, the production of VLPs was found to be suppressed by pH downshift. Moreover, intracellular VLPs had a maximal amount, which was termed maximal intracellular VLP content ($VLP_{max}$)). Finally, cell death and lysis can release intracellular VLPs. As a result, the concentration of VLPs in the medium was computed by multiplying intracellular VLP content and dead and lysed cell concentration.

I used Matlab R2021a and Optimization Toolbox (The Math Works, Inc.) to perform the computational work for the model construction and subsequent in silico analysis. Ordinary differential equations listed above were solved by ode45 solver following the recommendation from Matlab. In order to determine parameters in these equations, weighted root-mean-square deviation (wRMSD) was selected as the cost function and the optimization was performed with fminsearch command in Matlab. To evaluate the model simulation, the coefficient of determination ($R^2$) was used (Colin Cameron and Windmeijer, 1997).

### 2.2.4. Local sensitivity analysis

After the model construction, a deeper understanding of the effect of model parameters and initial culture conditions on the final product yield is desired. Thus, local sensitivity analysis was performed. To do so, a 10% change in model coefficients and initial conditions was made independently and simulation was conducted with the constructed model. The responses of final VLP concentration to the changes in model parameters were recorded in the form of relative change. Model parameters of STR1 and STR2 obtained in the previous section were averaged when applicable because STR1 and STR2 shared the same operating and initial conditions before pH downshift. After the pH shift, parameters of STR2 were applied for the purpose of simulation. Therefore, the simulation result using averaged parameters before pH shift and parameters of STR2 after the pH shift was used as the standard for computing relative final VLP concentration changes.

### 2.2.5. Predictions under different pH shift timing

Predictions of cell culture performance under pH shift on various timing were carried out similarly to the simulation described in Section 2.2.4. Briefly, before pH shift, averaged model parameters of STR1 and STR2 were applied while model coefficients of STR2 were applied after the pH shift. A series of time-dependent variables including cell densities, metabolite concentrations as well as VLP concentration were predicted if pH shift was conducted at 144, 168, 192, 216, 240, 264, 288, 312, and 336 h (no pH shift). Based on predictions, further understanding of the effects of pH downshift was gained and the optimal pH shift timing maximizing final VLP concentration was determined.

### 2.2.6. Multiple objective optimization

For the production of biopharmaceuticals, final product yield is not the only target for optimization. Product quality is another aspect worth considering. I found that there exists a trade-off between product yield and quality. More concretely, high VLP concentration was accompanied with high ammonia concentration. However, high level of ammonia can suppress the expression of glycosylation genes in CHO cells (Chen and Harcum, 2006). Glycosylation is important for the VLP vaccines for triggering immune responses (Lavado-García et al., 2022). Besides, glycoprotein folding and VLP assembly, which are critical for the structure and function for the VLP, also rely on glycosylation (Chen and Lai, 2013). Thus, ammonia concentration was related to the product quality and high ammonia concentration should be avoided. So, in this research, VLP concentration was maximized because it was the product. At the same time, ammonia concentration was kept as low as possible as it had negative effect on the product quality. Therefore, I applied desirability methodology to balance product yield and quality (Möller et al., 2019). Briefly, desirability methodology computes individual desirability ($d_i(y_i)$) of different objectives separately at first and combines individual desirability together. Individual desirability ranges between 0 and 1 and a desirability value close to 1 is preferred. VLP concentration is one of the objectives to be maximized. Based on customized upper ($U_i$) and lower boundaries ($L_i$), its desirability can be computed by Equation 2-11-a. By contrast, ammonia concentration is the objective to be minimized because it has adverse effects on the product quality. Its individual

desirability was given by Equation 2-11-b. In the end, an overall desirability was calculated by multiplying the desirability values of VLP and ammonia concentrations as shown by Equation 12. An overall desirability value close to 1 is favorable, which was a quantitative criterion for balancing product yield and quality.

$$d_i(y_i) = \begin{cases} 0 & if\ y_i < L_i \\ \left(\frac{y_i - L_i}{U_i - L_i}\right) & if\ L_i < y_i < U_i \\ 1 & if\ y_i > U_i \end{cases} \quad (2\text{-}11\text{-a})$$

$$d_i(y_i) = \begin{cases} 1 & if\ y_i < L_i \\ \left(\frac{U_i - y_i}{U_i - L_i}\right) & if\ L_i < y_i < U_i \\ 0 & if\ y_i > U_i \end{cases} \quad (2\text{-}11\text{-b})$$

$$D = \prod_{i=1}^{n} d_i(y_i) = d_1(y_1) \times d_2(y_2) \cdots \times d_n(y_n) \quad (2\text{-}12)$$

## 2.3. Results and discussion

### 2.3.1. Thermal stability test of cell lysis indicators in batch and fed-batch culture

dsDNA has shown to be a good indicator for cell lysis quantification because it is stable after incubating at 37 °C for several days. As shown in **Figure 9**, concentrations of dsDNA in supernatant samples of batch culture remained relatively stable after 72 h of incubation. Even though some fluctuations in the dsDNA concentrations can be noticed, there was no clear trend of degradation. These fluctuations might be due to experimental and measurement errors. In addition, from **Figure 9**, it can also be found that dsDNA concentration increased during the cultivation, especially at the end of cell culture. That was due to the cell death by the end of cultivation releasing intracellular DNA content.

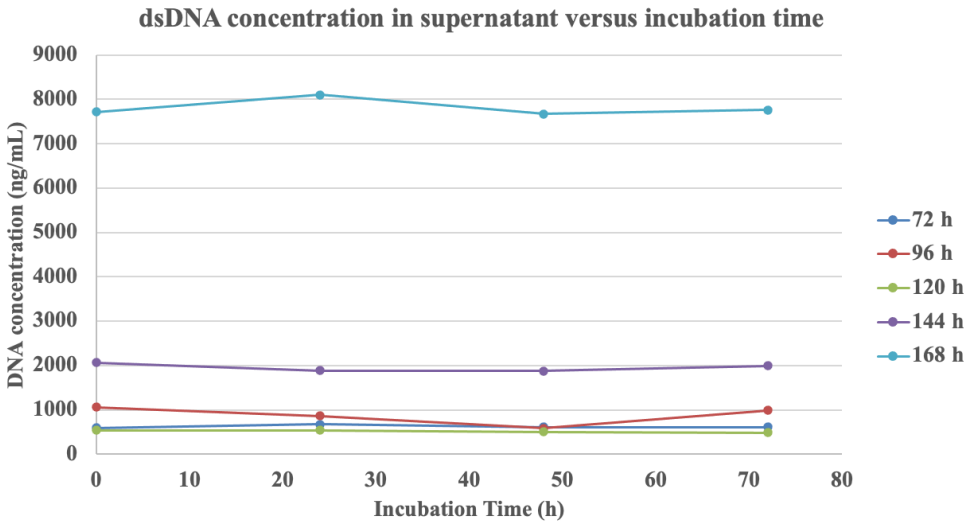


**Figure 9**: dsDNA concentrations in supernatant samples of batch culture taken on different time after 72 h of incubation at 37 °C

The same conclusion was found in the samples of fed-batch culture (**Figure 10**). Some fluctuations in dsDNA concentrations can be detected, which might be due to experimental and measurement errors. However, the dsDNA concentrations were almost constant at 37 °C in general. Therefore, dsDNA is capable of acting as an indicator for cell lysis quantification as it cannot be degraded during the cell culture.

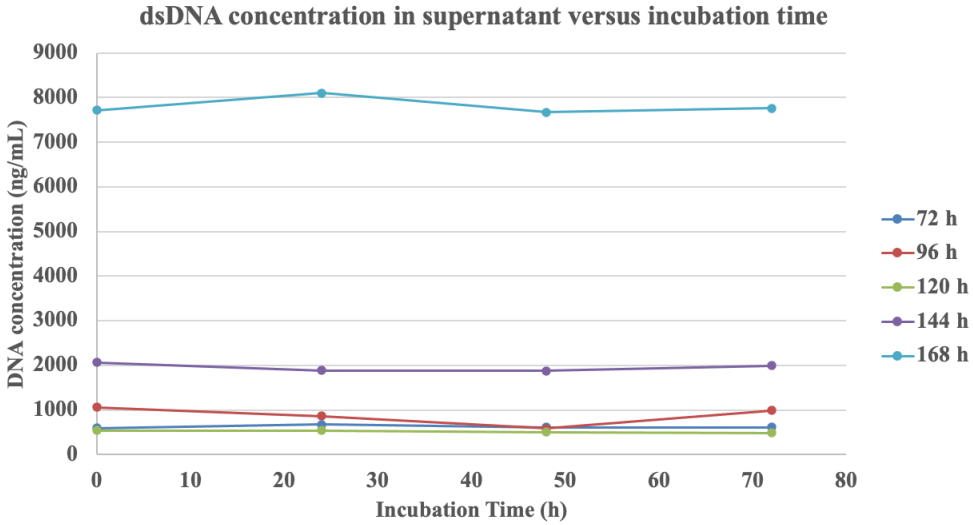


**Figure 10**: dsDNA concentrations in supernatant samples of fed-batch culture with two bioreactors (STR1 and STR2) taken on different time after 72 h of incubation at 37 °C

However, LDH had a different behavior. In the thermal stability test of samples from batch culture, LDH concentrations were not stable at 37 °C (data not shown). In the fed-batch culture with bioreactors, there was a clear degradation trend of LDH after incubation (**Figure 11**). The absorbance of supernatant samples after LDH assay decreased gradually after incubation. Therefore, LDH is not capable of being an indicator for estimating dead and lysed cell density precisely because it can be degraded during cultivation.

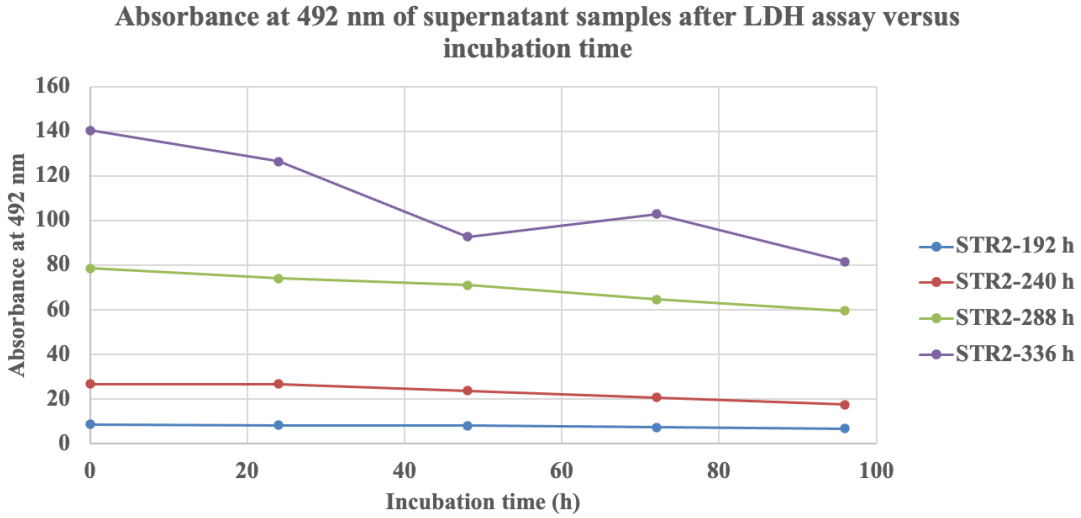


**Figure 11**: Absorbance at 492 nm of STR2 supernatant samples taken on different time after LDH assay versus incubation time

To summarize, dsDNA has shown good thermal stability while LDH was not stable at cell culture conditions. Therefore, cell lysis quantification was performed with dsDNA as the indicator using the method described in Section 2.2.2.4.

### 2.3.2. Simulation results

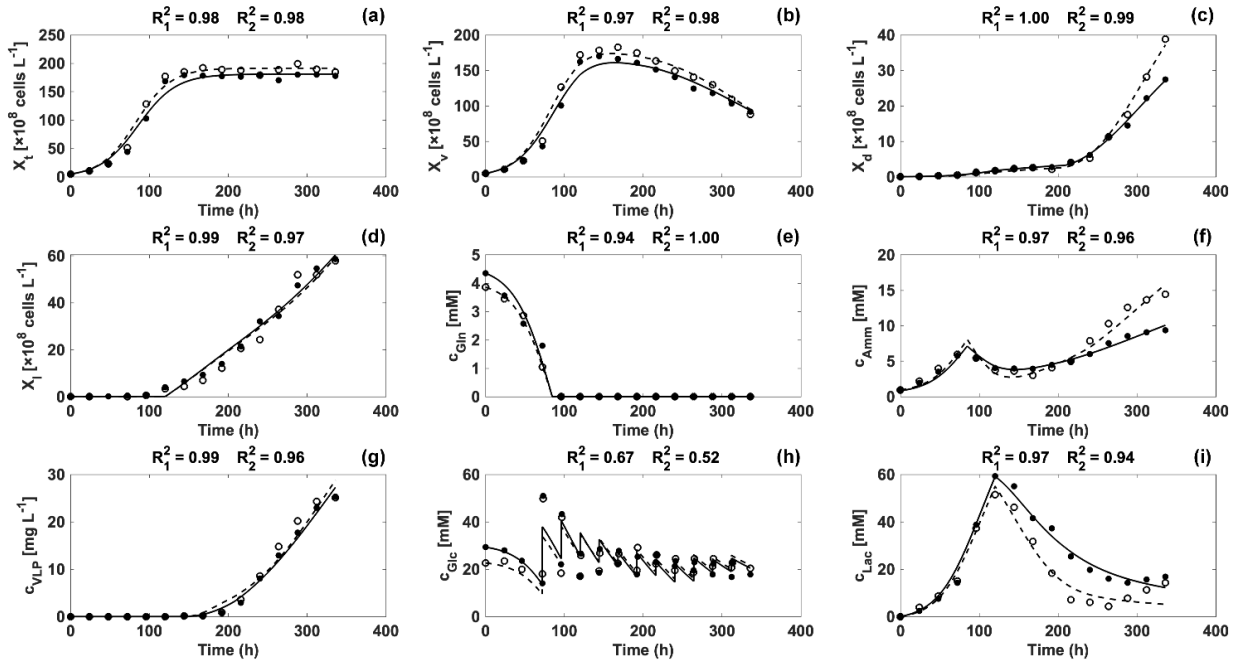


**Figure 12**: Comparison between model simulations of bioreactor 1 (solid line) and bioreactor 2 (dash line) (Edward, 2022) versus experimental data of bioreactor 1 (closed circle) and bioreactor 2 (open circle) for nine variables (**a–i**). $R^2$ was used to evaluate the goodness of fit

Parameters in the equations listed in Section 2.2.3. were determined by minimizing wRMSE as described before. After determining parameters for STR1 and STR2 separately, the results of the simulation were plotted as shown in **Figure 12** above in comparison with experimental data. Besides, some parameters are listed in **Table 3** below.

In **Figure 12a**, total cell densities in STR1 and STR2 were modeled precisely ($R^2 = 0.98$). Two bioreactors had similar profiles in total cell density because they were operated under the same conditions before 240 h and cell growth happened during this period. Some deviations can be found for the data points before 144 h, which are due to the simplicity of the model. By comparing model parameters related to total cell density ($r_{max}$ and $X_{t,\,max}$) of STR1 and STR2 in **Table 3**, it was found that these parameters are close. However, values are not exactly the same, which is due to batch-to-batch variation.

Besides, viable cell densities (**Figure 12b**), dead cell densities (**Figure 12c**) and lysed cell densities (**Figure 12d**) fitted experimental data well ($R^2$>0.95) based on model equations. Profiles

of viable cell density and lysed cell density were comparable between the two bioreactors. In contrast, cell death in STR2 was accelerated after pH shift resulting in a larger amount of dead cells in STR2 over STR2. This phenomenon can be explained by the faster accumulation of ammonia. Ammonia is known to be toxic to cells. As shown in **Figure 12f**, ammonia concentrations throughout the cell culture were modeled successfully. At first, ammonia was produced as a result of glutamine consumption. When glutamine was not available, ammonia was used as an alternative nitrogen source for cells and consumed until the cell growth stopped. Similar behavior was reported in other mammalian cell lines (Lie et al., 2019). After that, ammonia was produced again due to the cell maintenance (Xu et al., 2019). According to the specific ammonia production rate at 6.75 ($q_{Amm,6.75}$) and 7.15($q_{Amm,7.15}$) listed in **Table 3**, it was found that specific ammonia production rate was increased dramatically after the pH shift. As a result, ammonia concentration was higher in STR2, which is in line with the observation reported previously (Lee et al., 2021). Finally, the higher ammonia concentration led to faster cell death in STR2. In general, the changes in ammonia concentration were modeled successfully. However, some deviations between the real data and simulation can be noticed in the late stage of cell culture, which was due to the simplicity of the model. Besides, based on the change in specific ammonia production rate, the effect of pH shift was quantified, which improves the understanding of the process.

In addition, lysed cell density was simulated with high accuracy as well (**Figure 12d**). According to data points and profiles of simulation of STR1 and STR2, lysed cell densities in the two bioreactors were similar. Similarly, related model parameters ($k_{DL,32}$ and $k_{L,32}$) were comparable between the two bioreactors suggesting that pH shift hardly had any effects on the cell lysis. **In Figure 12e**, glutamine concentrations were simulated accurately as well. Two bioreactors have shown similar profiles and some variations can be due to experimental and measurement errors.

The equations gave a good estimation of VLP concentration in two bioreactors ($R^2$>0.95). Based on our previous observations, the release of VLPs happened after 144 h of cell culture. As mentioned before, STR2 had more dead cells than STR1. However, VLP concentrations were still comparable in the two bioreactors. Based on the $VLP_{max}$ listed in **Table 3**, this phenomenon can be explained. $VLP_{max}$ is the maximum intracellular VLP content in a single cell. It was $3.85\times10^{-12}$ g cells$^{-1}$ in STR1 and $3.39\times10^{-12}$ g cells$^{-1}$ in STR2. This difference came from the batch-to-batch

variation which is due to the nature of cells instead of pH shift. As indicated by $t_{VLP,cr}$ which is the time required for VLP accumulation to reach its maximum, intracellular VLPs can achieve the peak before the pH shift. Therefore, compared with the control group, the treatment group (STR2) had higher dead cell density and lower value of $VLP_{max}$ resulting in insignificant difference in VLP yield. In addition, the VLP concentrations were modeled accurately with the newly proposed equation which correlates cell death and lysis with VLP concentration (Equation 2-10). Therefore, this equation can potentially be used to model production processes whose product release depends on cell death and lysis such as other non-enveloped VLPs.

Finally, concentrations of glucose (**Figure 12h**) and lactate (**Figure 12i**) were modeled with the proposed kinetic model. In terms of glucose, this model can only provide rough estimations of its concentration ($R^2$ = 0.67 and 0.52 for STR1 and STR2, respectively). These deviations were caused by experimental and measurement errors. Besides, model simplicity is responsible for the deviations. Lactate concentrations were described by the model with reasonable accuracy. Lactate was produced by cells before 120 h and switched to consumption after 120 h. The pH shift hardly had any impact on the lactate concentration as STR1 and STR2 showed similar profiles.

According to the simulation results, model equations proposed in this research were capable of simulating experimental data of CVA6 VLP vaccine production process. Several model parameters were assumed to be constant including the yield constants and specific consumption or production rates of metabolites. Using constants instead of Michaelis Menten type term was beneficial for keeping the simplicity of the model. Bioprocess models are overparameterized, which has the risk of overfitting. Therefore, I used as few model parameters as possible. These parameters turned out to be capable of modeling the process with high accuracy. In terms of the equation for the lactate, a Monod type term was used because a single constant was not enough to capture the behavior of lactate concentration. Another method to avoid overfitting is using regularization term. More concretely, because two bioreactors were operated under the same conditions before 240 h, their parameters can be used to regularize each other. The cost function for parameter determination can be rewritten to minimize the distance between simulation and experimental data as well as the difference between parameters from STR1 and STR2 at the same time. Model parameters can also be shared for STR1 and STR2. However, it will also lose the

precision of the simulation because of batch-to-batch variations. It should be noted that there exists a tradeoff between precision and generality for the purpose of modeling.

There are still several limitations of this model. This model can only simulate the processes whose product release relies on cell death and lysis because the equation of product concentration was correlated to cell death and lysis. Besides, since different cell lines have their own limiting substrates or toxicities, model equations of cell growth, death, lysis, and metabolite production and consumption are not generally applicable to all CHO cell lines. To apply this model to other cell lines, equations need to be adapted and customized according to the cell behavior. Moreover, the effects of dilution were not taken into consideration in the model construction. To enable the modeling of fed-batch mode, an additional dilution term should be added to each equation.

**Table 3**: Model parameters of STR1 and STR2.

| Parameter | STR1 | STR2 |
|---|---|---|
| $r_{max}$ ($h^{-1}$) | 0.042 | 0.044 |
| $X_{t,\,max}$ ($10^8$ cells $L^{-1}$) | 181.1 | 191.5 |
| $k_{D,\,min}$ ($10^{-4}$ $h^{-1}$) | 2.5 | 1.7 |
| $k_{DL,\,32}$ ($10^{-3}$ $h^{-1}$) | 8.8 | 6.9 |
| $k_{L,\,32}$ ($10^{-3}$ $h^{-1}$) | 1.4 | 1.3 |
| $k_{D,dead}$ ($10^{-4}$ $h^{-1}$) | 5.1 | 3.3 |
| $K_{D_{Amm}}$ (mM) | 0.59 | 0.60 |
| $Y_{X_{v/Gln}}$ ($10^{12}$ cells $mol^{-1}$) | 1.90 | 2.45 |
| $Y_{Amm/Gln}$ (-) | 1.45 | 1.84 |
| $Y_{X_{v/Amm}}$ ($10^{12}$ cells $mol^{-1}$) | 1.74 | 1.08 |
| $q_{Amm,\,7.15}$ ($10^{-15}$ mol $cell^{-1}$ $h^{-1}$) | 0.95 | 1.65 |
| $q_{Amm,\,6.75}$ ($10^{-15}$ mol $cell^{-1}$ $h^{-1}$) | N.A. | 2.66 |

| | | |
|---|---|---|
| $VLP_{max}$ ($10^{-12}$ g cell$^{-1}$) | 3.85 | 3.39 |
| $q_{VLP}$ ($10^{-13}$ g cell$^{-1}$ h$^{-1}$) | 0.59 | 1.08 |
| $t_{VLP,cr}$ (h) | 209.5 | 175.3 |

### 2.3.3. Local sensitivity analysis

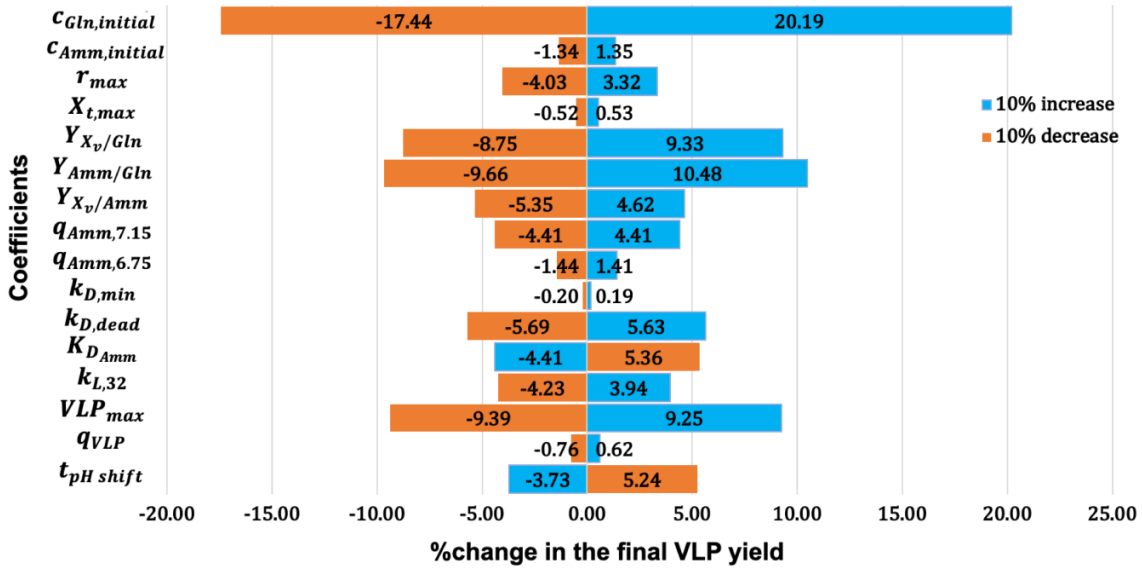


**Figure 13**: Results of local sensitivity analysis shown by the relative change in the final VLP yield in response to a 10% change in initial conditions and model parameters.

In the previous, it was found that there was batch-to-batch variation even though two bioreactors were operated under the same conditions before 240 h. Therefore, in order to further understand how the CVA6 VLP production process responds to small variations in the initial conditions, sensitivity analysis was carried out. Besides, this analysis also determined the impact of small changes in the model parameters on the final product yield. To do so, the analysis was conducted in a one-way manner by increasing or decreasing one factor by 10% independently while keeping other factors constant (Qian and Mahdi, 2020). Then, the final VLP concentration on 240 h was recorded.

As shown in **Figure 13**, a tornado plot was used to visualize the results of sensitivity analysis. The initial glutamine concentration was found to have a huge impact. A 10% increase in the initial glutamine concentration was predicted to lead to a 20.19% increase in the VLP concentration on 336 h. Higher glutamine concentration will result in an elevated level of ammonia concentration which could speed up cell death. As a result, more VLPs can be released from dead cells giving a higher yield. However, the VLP yield is not sensitive to the changes in the initial ammonia concentration because the numerical value of it was too low. To summarize, the sensitivity analysis for the initial conditions has identified glutamine supplementation in the medium as an important material attribute that needs to be controlled carefully. Conversely, the variance in the initial ammonia concentration cannot affect the final product yield much.

Similarly, sensitivity analysis was also carried out to determine the influence of model parameters. The decrease in $r_{max}$ can lead to a lower final product yield since $r_{max}$ is related to cell growth and the decrease of it will reduce the total cell number which reduces dead and lysed cell density subsequently. Besides, the decrease in $X_{t,max}$ also has a negative effect on the VLP yield, but its impact is limited compared with the decrease in $r_{max}$. The final VLP concentration is also affected by the yield coefficients as changes in these coefficients will change the ammonia concentration which affects the rate of cell death. More concretely, a 10% independent increase in $Y_{X_v/Gln}$, $Y_{Amm/Gln}$, or $Y_{X_v/Amm}$ would improve VLP concentration by 9.33%, 10.48%, and 4.62%, respectively. In addition, ammonia concentration can also be affected by specific ammonia production rate which means the amount of ammonia produced by a single cell in a given period. This rate was changed at different pHs. Increases in $q_{Amm,7.15}$ and $q_{Amm,6.75}$ will accelerate the accumulation of ammonia. As a result, dead cell density will increase and VLP concentration is supposed to be higher. Besides, changes in final VLP yield could also be induced by 10% changes in model parameters that are related to cell death and lysis. Since most dead and lysed cells were generated during the late stage of cell culture, $K_{D,min}$ only had quite limited impact on VLP yield because it was the specific death rate during the early stage instead of the late stage. By contrast, 4%–6% difference in the final VLP concentration can be generated in response to 10% variances in $k_{D,dead}$, $K_{D_{Amm}}$, and $k_{L,32}$. Finally, there are two model parameters that are directly related to VLP production, namely $VLP_{max}$ and $q_{VLP}$. The final VLP concentration is sensitive to the change in $VLP_{max}$ as a 10% increase in $VLP_{max}$ could improve VLP yield by 9.25%. Conversely, the

effect of $q_{VLP}$ is limited. In addition, VLP concentration on 336 h was also influenced by the timing of pH shift. Based on the result of local sensitivity analysis, it was suggested that VLP yield can be improved by shifting the pH on the earlier time.

Results of sensitivity analysis suggested that the VLP yield can be improved by changing the initial culture conditions and model parameters, especially glutamine concentration. However, increasing the initial glutamine concentration would also lead to high ammonia concentration which has negative effect on product quality as mentioned before. In addition, it was important to verify the result of sensitivity analysis based on model simulation by performing additional experiments. However, changing one single model parameter independently without changing other parameters was quite difficult, which hindered the validation of results.

2.3.4. Effect of the pH shift timing on VLP production

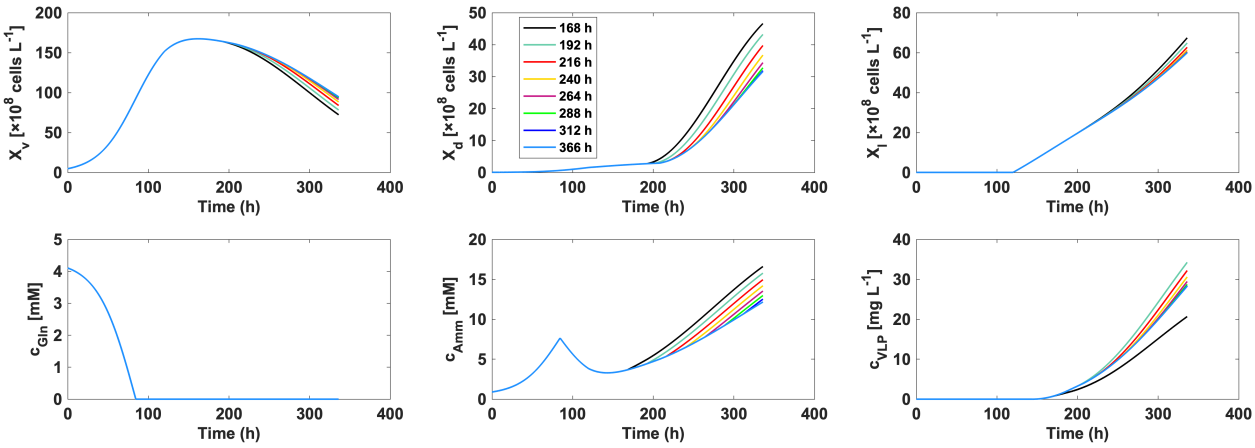


**Figure 14**: Prediction of experimental performance of the VLP vaccine production process with various pH shift timings for six variables (**a–h**). 168 h (black), 192 h (deep green), 216 h (red), 240 h (yellow), 264 h (purple), 288 h (light green), 312 h (deep blue), and 336 (light blue) denote the time when a pH shift is performed.

In the result of local sensitivity analysis, it was found that VLP yield could be improved by shifting the pH earlier. Therefore, to further investigate the impact of pH shift timing on the cell culture performance, predictions were conducted based on the calibrated model. The results are displayed in **Figure 14**. Prediction of ammonia concentration over time under different pH shift

timing was shown in **Figure 14e**. It was noticed that the production of ammonia was accelerated after the pH shift. As a result, higher dead cell density can be observed in the cell culture with pH shift on earlier time as shown in **Figure 14b**. Besides, lysed cell density was also elevated slightly if the pH shift was conducted on earlier time (**Figure 14c**). Glutamine consumption was hardly affected by the pH shift timing based on **Figure 14d**. Finally, the VLP concentration throughout the cell culture was predicted as well. Even though the cell culture with pH shift at 168 h had the highest dead and lysed cell density, the product yield of it was still low. This phenomenon can be explained by the suppression effect of pH shift on the VLP accumulation. Based on our previous observation, after the pH downshift, intracellular VLP accumulation would be stopped and the $VLP_{max}$ could not be reached. Therefore, in spite of the highest dead and lysed cell density, a pH shift at 168 h did not produce the most VLP. By contrast, if pH shift was performed at 192 h or later time, $VLP_{max}$ can be reached. Therefore, cell culture with a pH shift at 192 h can generate the highest VLP yield because $VLP_{max}$ is achievable and it had a higher number of dead and lysed cells. Compared with the cell culture without a pH shift, cell culture with a pH shift on at 192 h could improve the final product concentration at 336 h by approximately 20%.

It would be beneficial to validate the model prediction by performing additional experiments. However, due to technical difficulties, model validation was absent. Besides, it was also difficult to verify the simulation in the presence of batch-to-batch variation. But the result of prediction was still in line with previous studies. More specifically, several reports have revealed that pH downshift could lead to elevated ammonia concentration (Lee et al. 2021; Trummer et al. 2006). In addition, ammonia was known to be toxic for cells (Schneider 1996) and it can induce cellular apoptosis (Wang et al.2018). Therefore, accelerated cell death due to high ammonia concentration caused by pH downshift predicted by the kinetic model was reasonable.

To further improve the VLP yield, pH shift strategy can not only be optimized by changing the pH shift timing. The optimal pH value also needs to be determined. However, in this research, only two pH values (pH = 7.15 and 6.75) were investigated. To optimize the pH value, cell culture at different pHs needs to be explored. pH-dependent model parameters can be determined. Mathematical functions between pH value and pH-dependent model parameters would be built by

linear regression or more advanced machine learning algorithms, enabling hybrid modeling. Based on this procedure, the optimal pH shift value can be predicted and determined.

### 2.3.5. Multiple objective optimization

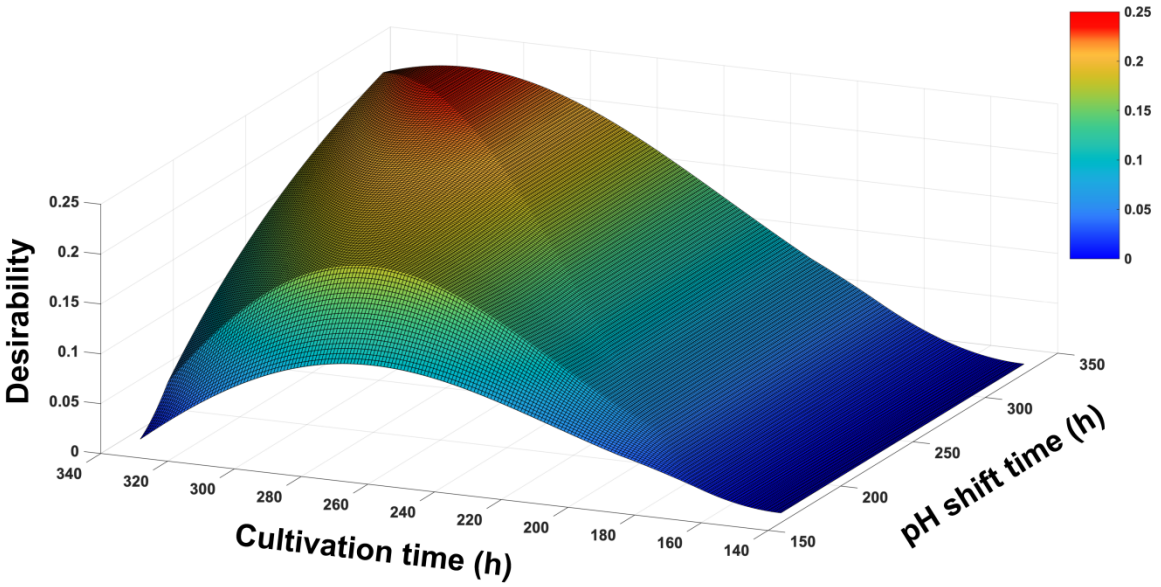


**Figure 15**: Surface plot of desirability as a function of cultivation and pH shift timings.

For the production of biopharmaceuticals, product yield is one of the important objectives but not the only one. Product quality is another aspect worth consideration. In the previous section, shifting pH to a lower value during the cell culture led to higher ammonia concentration, which increased the number of dead cells and therefore improved the VLP yield. However, at the same time, ammonia concentration needs to be carefully controlled because a high ammonia concentration can suppress the expression of glycosylation genes in CHO cells, which inhibits the glycosylation of the recombinant proteins produced by cells (Chen and Harcum, 2006). As a type of recombinant protein, glycosylation of VLPs also plays a critical role. A previous study has reported that glycosylation is important for triggering the immune response for VLP vaccines (Lavado-García et al., 2022). In addition, glycosylation plays important roles in the function and structure of VLPs including glycoprotein folding and VLP assembly subsequently (Chen and Lai, 2013). Therefore, a high ammonia concentration needs to be avoided during the cell culture. Thus, another objective of the production of CVA6 VLP vaccines is to minimize the ammonia concentration when cultivating cells and there exists a trade-off between product yield and quality.

In order to make decisions for this tradeoff quantitatively, I combined the desirability methodology with the model simulation. I customized a desirability function to maximize the product concentration while keeping the ammonia concentration as low as possible simultaneously.

To visualize the result of multiple-objective optimization using desirability methodology, a surface plot was applied as shown in **Figure 15**. The value of desirability was indicated as a function of cultivation time and pH shift time by the height of the surface as well as the color. Generally, at a given pH shift time, the desirability value increased gradually at first during the cell culture because VLPs were released. However, its value slightly dropped by the end of cell culture because of the elevated ammonia concentration. Besides, desirability can also be improved by shifting pH later. It is noteworthy that there was a diagonal on the surface. This diagonal represents that the pH shift time equals the culture time. If the pH shift time was later than the cultivation time, the result would be the same. Based on the result of the desirability methodology, it was predicted that the cell culture without a pH shift stopped at 311 h was the most desired operating strategy to balance the trade-off between product yield and quality.

## 2.4. Conclusion

In this chapter, I proposed a novel kinetic model of the CVA6 VLP vaccine production process using CHO cell culture since the current models are not capable of modeling the production process. Based on our previous observations, the production of CVA6 VLP was strongly correlated to cell death and lysis. Therefore, to construct this model, I developed a protocol for the quantification of cell lysis firstly because it cannot be measured directly using image-based analyzers. Two candidates of cell lysis indicators namely LDH and dsDNA were tested. For the samples taken from batch culture, dsDNA was found to be stable at the temperature of cell culture (37 °C). However, the concentration of LDH was not constant after the incubation at 37 °C. Similar behaviors were observed in the samples from fed-batch cultures in bioreactors. The concentration of dsDNA remained stable after incubating for 96 h in the samples from bioreactors while there was a clear trend of degradation in the LDH concentration after 96 h of incubation. Therefore, dsDNA was selected as the indicator for determining lysed cell density.

After acquiring the information on cell lysis and combining it with the data of cell culture kindly provided by Dr. Thao, I constructed a kinetic model based on the data. Model equations were proposed based on my understanding of the process or by adapting previous equations to fit the process. The results of the simulation have shown that the constructed model was able to describe the kinetics of most process variables with high quality. Concretely, total, viable, dead, and lysed cell densities during cultivation were simulated with high accuracy. Besides, time-dependent concentrations of glutamine, ammonia, and lactate were modeled precisely as well. Nevertheless, due to the simplicity of the model as well as experimental and measurement errors, the constructed model was unable to describe the changes in glucose concentration accurately. Finally, the newly proposed equation correlating the VLP concentration with cell death and lysis was qualified to model the CVA6 VLP concentration precisely. Additionally, model parameters also improved the understanding of the process. Specifically, the pH downshift accelerated ammonia accumulation during the cultivation. This effect was described quantitatively through a model parameter named specific ammonia production rate. In addition, even though STR2 had higher dead cell concentration, the VLP yield was not improved much. This phenomenon can be explained by the difference in the model coefficient called maximum intracellular VLP content, which was due to batch-to-batch variation.

Subsequently, a series of analysis was performed with the help of the constructed model in order to facilitate the process development of CVA6 VLP production. According to the result of the sensitivity analysis, the impact of variances in the initial conditions and model parameters on the final VLP concentration was determined. The initial glutamine concentration was suggested to be an important material attribute that can affect the final product yield considerably. Therefore, glutamine supplementation needs to be carefully designed and controlled. Also, the final VLP yield was sensitive to the small variance in the maximum intracellular VLP content. Since sensitivity analysis revealed it was possible to improve the VLP yield by shifting the pH on earlier time, predictions were conducted if the pH shift was performed on different timing. Based on the predictions using the calibrated model, the optimal pH shift timing was 192 h so that the VLP concentration by the end of cell culture could be maximized. However, it was found that a tradeoff between product yield and quality existed. Therefore, to perform multiple objective optimization, desirability methodology was applied and integrated with the model simulation. The most desired

operating condition that can balance the trade-off between product yield and quality was predicted to be the cell culture without a pH shift at 311 h. To summarize, the constructed model is a versatile tool to assist the process development of CVA6 VLP vaccine production. It enabled *in silico* process development by guiding the decision making in a quantitative manner.

In terms of future perspective, the newly proposed equation that correlates VLP concentration with cell death and lysis provides an opportunity to model other production processes whose product could not be released such as other non-enveloped VLP vaccines. Besides, this model also has the potential to be integrated into data-driven modeling such as artificial neural network enabling hybrid modeling which can further model the effects of different pHs (Narayanan et al., 2019). It can also be combined with downstream models so that the optimization can be performed in silico at a system level instead of optimizing unit operations independently. In addition, the constructed model could also contribute to the implementation of DT as modeling is one of the key components in the concept of DT. As the biopharmaceutical industry is moving towards industry 4.0, DT is highly desired to achieve online process monitoring and control and therefore enable more advanced manufacturing in the biopharmaceuticals (Cardillo et al., 2021; von Stosch et al., 2021).

The limitation of this part of the research is that this deterministic model can only simulate or predict a single value at each time point without providing any information on the uncertainty. In the field of bioprocess, there are a few sources of uncertainties in the process. These uncertainties may come from the nature of cells, variations in raw materials, fluctuations of controllers, as well as experimental and measurement errors. As a result of these uncertainties, the cell culture performance of parallel experiments can be different even if materials and operating conditions are kept the same as much as possible. In Section 2.3.2., it was also found that even though two bioreactors were operated under the same conditions before 240 h, their performance still had some deviations. More importantly, the variation in $VLP_{max}$, which is due to the nature of cells, has resulted in insignificant improvement in the final VLP yield over the control group even though STR2 had more dead and lysed cells. Therefore, a probabilistic model that can provide both mean and variance of model parameters is needed to provide quantitative information on the uncertainty of the production process of the CVA6 VLP vaccine. Thus, in the next chapter, I constructed a probabilistic model to quantify the uncertainty of the CVA6 VLP vaccine production process.

Chapter 3:

# Probabilistic modeling of a CVA6 VLP vaccine production process using CHO cell culture

# 3. Chapter 3: Probabilistic modeling of a CVA6 VLP vaccine production process using CHO cell culture

## 3.1. Introduction

In the field of biopharmaceutical production, batch-to-batch variation means that even production process is operated under the same conditions using the same materials and equipment as much as possible, the performance of the production process such as product yield and quality can be still different among different batches. This variation is inevitable since there are many uncertainties involved in the bioprocess contributing to the variation. For instance, there can be variations in the raw materials such as media composition. Also, fluctuations in controllers may happen due to the nature of the control system. These changes can lead to differences in the final products. Besides, experimental, human, and measurement errors are sources of uncertainties in bioprocess as well. From the standpoint of regulatory agencies, batch-to-batch variation needs to be carefully monitored and controlled in order to avoid unqualified products that might be harmful to the patients.

Therefore, GMP regulates the production process of drug substances and should produce products with high quality consistently. Thus, the batch-to-batch variation should be studied in order to be in compliance with GMP. Probabilistic models are useful tools for studying batch-to-batch variation. Different from deterministic models which can only provide a single value of variables at a given time point, probabilistic models can provide probability distribution throughout the process. The simulation generated by deterministic models can be considered as the average of cell culture performance. This averaged result is enough for the purpose of optimization, but it is not qualified for online process control in some cases. More concretely, for instance, during the CHO cell culture, glucose concentration is generally kept above 2 g/L. For predictive and online control, deterministic models can only predict when the glucose is lower than the set point on average, but outliers may occur sometimes. By contrast, probabilistic models can provide uncertainty information allowing decision-making based on the worst case to avoid negative effects caused by outliers. In the bioprocess design, there are scenarios that require

consideration based on the worst case to eliminate the risk of batch failure because batch failure is expensive. To enable the consideration based on the worst case, probabilistic models are needed. Besides, probabilistic models take stochasticity into consideration, they are highly desired for the implementation of DT because they can represent the real world better. In addition, as mentioned in the previous chapter, the constructed deterministic model was unable to provide information about variance between batches and the variation in $VLP_{max}$ resulted in insignificant improvement in the final VLP concentration in STR2. Thus, in this chapter, I performed probabilistic modeling for the CVA6 VLP production process in CHO cells. Different from the fed-batch culture using bioreactors in the previous chapter, data was collected from batch culture of CVA6 VLP-producing CHO cells using shaking flasks. The reason for choosing batch culture in shaking flasks was that the number of studies discussing probabilistic modeling in bioprocess is quite limited currently and it was better to start with the process which contained fewer uncertainties and was easier to quantify. Compared with fed-batch culture using bioreactors, batch culture using shaking flasks was free of uncertainties from the addition of feeding media and fluctuations of several control systems such as pH and DO control. Therefore, cell cultures were performed in parallel using shaking flasks in a batch mode.

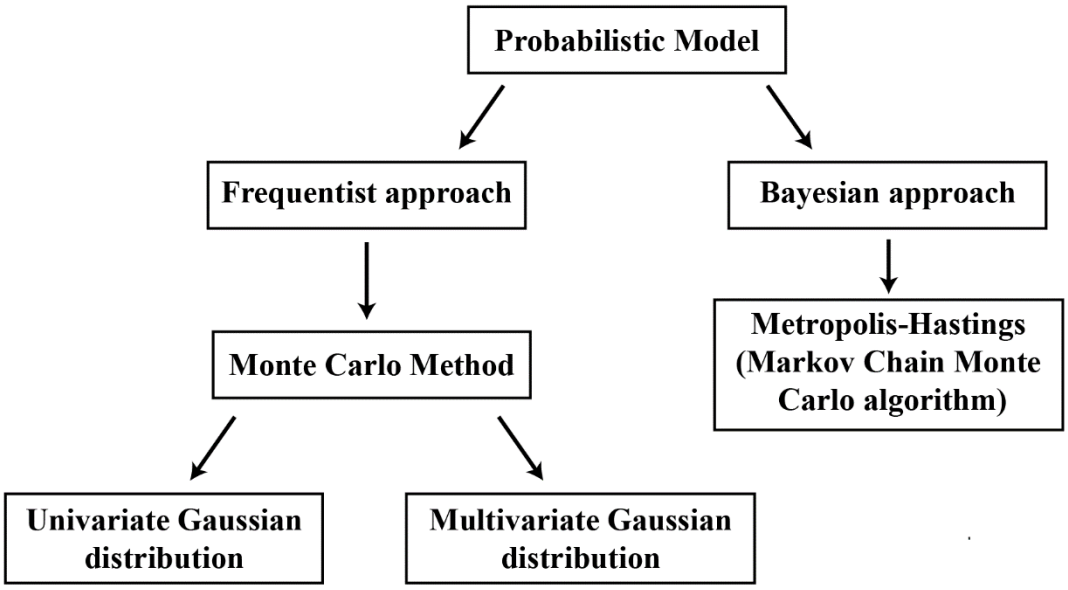


**Figure 16**: Frequentist and Bayesian approaches for performing probabilistic modeling.

In terms of computational work after the data collection, two approaches could be applied to perform probabilistic modeling namely frequentist and Bayesian approaches as shown in **Figure 16** (Hernández Rodríguez and Frahm, 2020). Frequentist approach is purely based on the data while Bayesian approach allows the use of both data and prior knowledge when performing the parameter estimation. To the best of my knowledge, there is no research comparing the performance of two approaches in the field of bioprocess and it is difficult to evaluate which one is better in advance. Therefore, in this chapter, I tried to use both methods and compared the advantages and disadvantages of them.

To conduct frequentist approach, sampling techniques are commonly used. So, I applied the Monte Carlo (MC) method under the assumption that the data at each time point either followed univariate or multivariate Gaussian distribution. Subsequently, sampling was carried out to generate a large data set following the computed distributions. Then, parameter estimation was performed given the generated data set, and the performance of univariate and multivariate Gaussian distribution was compared and discussed. In terms of Bayesian method, Metropolis-Hasting, which is a Markov Chain Monte Carlo method, can be used to provide a numerical solution for estimating the uncertainties of parameters (Hernández Rodríguez and Frahm, 2020). Briefly, the posterior distribution of parameters can be computed by using the prior distribution of parameters and observations (the data) through Bayesian theorem and a more detailed description will be provided in the next section. After trying the methods mentioned above, their performance was compared, and limitations were discussed.

## 3.2. Materials and methods

### 3.2.1. Flask culture

In order to acquire data for the probabilistic model construction, cell cultures were performed using 4 shaking flasks in parallel. Materials and culture conditions were kept the same as listed in Section 2.2.1.1. Briefly, CVA6 VLP-producing CHO-S cells provided by BIKEN Group Japan were cultivated. CD FortiCHO™ medium (Thermo Fisher) and 0.5 L Optimum Growth® Flasks (Thomson Instrument, CA, USA) with 0.1 L working volume were used for the cultivation. 8 mM

glutamine (Fujifilm Wako) was supplied initially. A humidified incubator (Climo-Shaker, Kuhner, Switzerland) operated at 37 °C, 8% CO2, and 140 rpm was applied.

### 3.2.2. Analytical methods

Samples of cell cultures were taken every 24 h. A ViCell machine (Beckman Coulter) was utilized to perform cell counting for the samples. The supernatant of samples was collected by centrifuging samples at 14000 g for 5 minutes. Measurement of dsDNA concentrations in the supernatant was carried out using the Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific) following the procedure described in 2.2.1.2. A YSI 2900 analyzer (YSI Life Sciences, Yellow Springs, USA) was used to conduct measurements for glucose, lactate, glutamine, and ammonia concentrations. In terms of CVA6 VLP quantification, a sandwich ELISA protocol developed by Biken Group Japan was used. The procedure was described in Section 2.2.2.5.

### 3.2.3. Model update for batch culture

Cell culture using shaking flasks in a batch manner had shown some different behaviors compared with cultivation using bioreactors in a fed-batch mode. There were many potential reasons accounting for the phenomenon. For instance, nutrient depletion was one of the reasons as nutrients such as glucose can be used out without additional supplements in the batch culture leading to steep cell death by the end of cell culture. Besides, shear stresses experienced by cells were also different as impellers were used in the bioreactors. Therefore, the kinetic model was updated to fit the behaviors of batch culture using flasks of CVA6 VLP-producing cells.

$$\frac{dX_t}{dt} = rX_t \tag{3-1-a}$$

$$r = r_{max} \cdot \left(1 - \frac{X_t}{X_{t,max}}\right) \tag{3-1-b}$$

where $X_t$ represents total cell density; $r$ stands for intrinsic growth rate and $r_{max}$ denotes maximum intrinsic growth rate; $X_{t,max}$ is maximum total cell density.

The same logistic equations (Equation 2-1-a and 2-1-b) were also applied in describing the concentration of total cell density in the batch culture. These equations were found to be qualified for simulating total cell density in the fed-batch culture with only two model parameters reducing the risk of overfitting.

$$\frac{dX_d}{dt} = k_D X_v \tag{3-2-a}$$

Before 120 h:

$$k_D = k_{D,min} \tag{3-2-b}$$

After 120 h:

$$k_D = k_{D,dead} \tag{3-2-c}$$

$$\frac{dX_v}{dt} = rX_t - k_D \cdot X_v \tag{3-3}$$

where $X_d$ denotes dead cell density and $X_v$ stands for viable cell density; $k_D$ is specific death rate; $k_{D,min}$ represents minimum death rate; $k_{D,dead}$ stands for death rate during dead phase.

Based on the results of dsDNA measurements and cell counting, it was found that most dsDNA was released by dead cells and very few lysed cells were detected. Therefore, lysed cells were not taken into consideration in the modeling of batch culture, and terms related to lysed cells were removed from model equations. Besides, in the batch culture, cells died steeply after 120 h potentially due to the depletion of nutrients. Therefore, two death rates ($k_{D,min}$ and $k_{D,dead}$) were used as kinetic parameters to model the cell death. $k_{D,dead}$ was a constant since many factors might contribute to the cell death at the end of batch culture, which was difficult to model. Similar to Equation 2-4, viable cell concentration was given by the difference between total and dead cell density.

If ($c_{Glc}$ > 0 mM):

$$\frac{dc_{Glc}}{dt} = -(rX_t - k_D X_v)/Y_{X_v/Glc} - m_{Glc} X_v \quad (3\text{-}4\text{-a})$$

If ($c_{Glc}$ = 0 mM):

$$\frac{dc_{Glc}}{dt} = 0 \quad (3\text{-}4\text{-b})$$

Before 72 h:

$$\frac{dc_{Lac}}{dt} = \left((rX_t - k_D \cdot X_v)/Y_{X_v/Glc} - m_{Glc} X_v\right) \cdot Y_{Lac/Glc} \quad (3\text{-}5\text{-a})$$

After 72 h:

$$\frac{dc_{Lac}}{dt} = \left((rX_t - k_D \cdot X_v)/Y_{X_v/Glc} - m_{Glc} X_v\right) \cdot Y_{Lac/Glc} - q_{Lac} X_v \quad (3\text{-}5\text{-b})$$

$$q_{Lac} = q_{Lac,max} \cdot \frac{c_{Lac}}{K_{Lac} + c_{Lac}} \quad (3\text{-}5\text{-c})$$

where $c_{Glc}$ and $c_{Lac}$ denote glucose and lactate concentrations, respectively; $Y_{X_v/Glc}$ represents yield coefficient of cell proliferation to glucose uptake; $m_{Glc}$ is specific glucose consumption rate for cell maintenance; $Y_{Lac/Glc}$ stands for yield coefficient of lactate production to glucose uptake and $q_{Lac}$ is specific lactate uptake rate; $q_{Lac,max}$ denotes maximum lactate uptake rate and $K_{Lac}$ denotes Monod kinetic constant for lactate uptake.

For the purpose of modeling glucose and lactate concentrations in batch culture, equations used in Section 2.2.3 were modified slightly to fit the behavior of cells in batch culture. Glucose consumption was modeled as a result of cell growth and maintenance when glucose was available. After glucose was depleted, the concentration of it did not change anymore since no feeding medium or glucose supplement was added. Lactate concentration was modeled using the same equations as Equation 2-6-a, 2-6-b, and 2-6-c without cell lysis terms. The time of the lactate switch was changed to 72 h according to the experimental observation.

If ($c_{Gln}$ > 0.3 mM):

$$\frac{dc_{Gln}}{dt} = -(rX_t - k_D \cdot X_v)/Y_{X_v/Gln} \quad (2\text{-}6\text{-a})$$

$$\frac{dc_{Amm}}{dt} = (rX_t - k_D \cdot X_v)/Y_{X_v/Gln} \cdot Y_{Amm/Gln} \quad (2\text{-}7\text{-a})$$

If ($c_{Gln}$ = 0.3 mM):

$$\frac{dc_{Gln}}{dt} = 0 \quad (2\text{-}6\text{-b})$$

$$\frac{dc_{Amm}}{dt} = -(rX_t - k_D \cdot X_v)/Y_{X_v/Amm} + q_{Amm}X_v \quad (2\text{-}7\text{-b})$$

where $c_{Gln}$ and $c_{Amm}$ are glutamine and ammonia concentrations, respectively; $Y_{X_v/Gln}$ is yield coefficient of cell proliferation to glutamine uptake and $Y_{Amm/Gln}$ is yield coefficient of ammonia production to glutamine uptake; $Y_{X_v/Amm}$ denotes yield coefficient of cell proliferation to ammonia uptake and $q_{Amm}$ denotes specific ammonia production rate.

Changes in glutamine and ammonia concentrations were simulated using similar equations as those described in Section 2.2.3. Similarly, terms related to cell lysis were removed. Besides, it was found that glutamine concentration stopped dropping after reaching 0.3 mM, approximately. This phenomenon was possibly due to the error of the measurement instrument. Therefore, the threshold of glutamine concentration was set to 0.3 mM in the case of batch culture.

$$\frac{dc_{VLP}}{dt} = VLP_{max} \cdot k_D \cdot X_v \quad (3\text{-}8)$$

where $VLP_{max}$ denotes the maximum amount of VLPs accumulated intracellularly and $c_{VLP}$ is VLP concentration in the medium.

Finally, the concentration of VLP in the medium was given by multiplying $VLP_{max}$ with dead cell density. Different from fed-batch mode, in the batch culture, cells had less time for the VLP accumulation since the culture duration was shorter. Therefore, in this case, $VLP_{max}$ was a constant in order to simplify this model.

After the kinetic model was updated, parameters and their uncertainties were estimated under the framework of either frequentist approach or Bayesian approach.

### 3.2.4. Frequentist approach for uncertainty quantification

In order to implement frequentist approach, the MC sampling technique was used to estimate the distribution of model parameters by assuming Gaussian distribution of data at each time point. Gaussian distribution was selected because observations can be affected by several factors such as experimental and measurement error and these factors are random and independent to each other. If more data is available, the distribution of observations can be further investigated. In this case, three data points were used to calculate their mean, variance, and covariance defining a Gaussian distribution at the given time point. Then, a large number of samples were generated by randomly sampling from the computed distribution. Based on the generated samples, model fitting was performed using the same method described in Section 2.2.3 many times, and parameters of each run were recorded so that the uncertainties of model parameters could be estimated.

More concretely, data from 4 flasks was divided into 2 groups, 3 flasks for the training set and 1 flask for the test set. The flask for the test set was selected randomly. The training set was used for determining model parameters and their uncertainties respectively and the test set served as unseen data to evaluate the performance of frequentist approach. Based on the data of 3 flasks from the training set, distributions of variables such as viable cell density and glucose concentration at each time point can be estimated under the assumption of either univariate or multivariate Gaussian distribution.

$$p(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{3-9}$$

where $p$ is the probability density function; $x$ are variables at each time point; $\mu$ denotes mean and $\sigma$ denotes standard deviation.

The formula of univariate Gaussian distribution was shown by Equation 3-9 above. The probability density function of variables at each time point was defined by two parameters, mean and standard deviation. In this case, variables at each time point were assumed to be independent to each other. Given three data points of variables at each time point in the training set, mean and standard deviation of the univariate Gaussian distribution were determined to define the

distribution. After distributions of variables at different time points were defined, 100,000 samples were generated following defined distributions using randn command in the Matlab. Then, based on the artificially generated samples, model fitting was conducted following the same method described in Section 2.2.3. wRMSD was minimized with the help of fminsearch command in the Matlab. After 100,000 runs, the parameters and results of the model simulation were recorded for the subsequent evaluation.

By contrast, multivariate Gaussian distribution takes covariance between data points into consideration. Its formula was shown below.

$$p(x;\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} e^{-(\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))} \tag{3-9}$$

$$x = (x_1, x_2 \cdots x_n)$$

$$\Sigma = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix}$$

$$c_{ij} = Cov(x_i, x_j) = E\left[(x_i - E(x_i))(x_j - E(x_j))\right]$$

where $x$ is a matrix containing variables at different time points; $\mu$ denotes a vector containing means of variables at various time points; $\Sigma$ represents covariance matrix; $n$ stands for the number of variables in $x$ matrix.

Considering the data in the cell culture, data points are not independent to each other. For instance, the consumption of glucose is a cumulative process. More specifically, if the glucose concentration is high on the previous sampling time, it is more likely to be high on the subsequent sampling time. Therefore, it is important to investigate the difference between assuming multivariate and univariate Gaussian distribution when applying the MC technique for estimating the uncertainties in parameters of bioprocess kinetic models. To the best of my knowledge, such kind of comparison has not been reported by other studies. Thus, I also generated samples following multivariate Gaussian distribution estimated based on the same data set and compared its performance with the result from univariate Gaussian distribution assumption. To do so, the

covariance matrix of each variable was computed respectively using the cov command in Matlab. Then, covariance matrices and means of variables were used as inputs to generate 100,000 samples randomly following the estimated multivariate Gaussian distribution using mvnrnd command. Similar to the previous part, model fitting was conducted for 100,000 times and results of each run were recorded. To compare the difference between the assumption of univariate and multivariate, distributions of parameters and model simulation undertook a two-sample t-test using ttest2 command in Matlab and $p<0.05$ was chosen as the significance level. The performance of probabilistic modeling was evaluated using the within-band score as shown by the Equation 3-10 (Hernández Rodríguez and Frahm, 2020). A higher within-band score reflects better performance of model simulation as the model predicts unseen data points well.

$$within - band\ score = \frac{n_{within-band}}{n_{total}} \times 100\% \quad (3\text{-}10)$$

where $n_{within-band}$ is the number of data points in the training or test set that located within the 95% prediction interval of model simulation; $n_{total}$ is the total number of data points.

In addition, for the comparison between uncertainties of parameters from batch culture using flasks and fed-batch culture using bioreactors, the coefficient of variation (CV) was applied. CV allows comparing uncertainties across different scales by dividing standard deviation by mean as shown by Equation 3-11 below.

$$CV = \frac{\sigma}{\mu} \quad (3\text{-}11)$$

### 3.2.5. Bayesian approach for uncertainty quantification

The basis of Bayesian approach is the famous Bayes' theorem as shown by Equation 3-12.

$$p(\theta|y) = \frac{p(y|\theta) \cdot p(\theta)}{p(y)} \quad (3\text{-}12)$$

where $\theta$ represents a vector containing model parameters; $y$ is a vector containing observations (data); $p$ denotes probability density function; $p(\theta|y)$ means the conditional probability density function of $\theta$ when $y$ is given (also called posterior distribution) ; $p(y|\theta)$ denotes the conditional probability distribution of the occurrence of $y$ when $\theta$ is fixed (also call likelihood function).

Based on this equation, the distribution of model parameters can be estimated when data is given. More importantly, one of the advantages of Bayesian approach is that it allows the integration of prior distribution into the computation. Prior knowledge of parameters can be quantified first and then used to contribute to the computation. However, in this case, no prior knowledge was available since it was a novel model. Therefore, prior distribution of parameters was assumed to be uniformly distributed meaning that no prior information was provided.

This equation can be further simplified because $p(y)$ is a constant which constrains the probability distribution integrated to 1. Therefore, Equation 3-12 can be updated. The posterior distribution of $\theta$ is proportional to the product of its prior distribution and the likelihood function as shown by Equation 3-13 below.

$$p(\theta|y) \propto p(y|\theta) \cdot p(\theta) \tag{3-13}$$

However, considering that $p(y|\theta)$ and $p(\theta)$ are two distributions. It is computationally expensive to solve the product. Alternatively, some algorithms which can approximate the solution are available. Metropolis-Hastings algorithm was selected to give an approximation of the posterior distribution of model parameters since it is a general method and can solve high-dimensional problems (Hernández Rodríguez et al., 2019). The pseudocode of performing Bayesian approach using Metropolis-Hastings algorithm was provided as follows.

Pseudocode of Metropolis-Hasting algorithm for model parameter uncertainty estimation

1: Give initial guess of model parameters $\theta^0$

2: **For** i = 0, 1, …, 100,000 **do**

3: $\theta' = \theta^0$

4: Sample $\theta^* \sim p(\theta)$

5: Sample $a \sim U(0,1)$

6: Compute $b = \frac{L(y|\theta^*)\cdot p(\theta^*)}{L(y|\theta\prime)\cdot p(\theta')}$

7: $c = \min(1, b)$

8: $If\ a < c$: $\theta' = \theta^*$

9: $Else$: $\theta' = \theta'$

10: Record $\theta'$ and model simulation

11: **End for**

After the initialization by giving an initial guess of parameters to the current parameters ($\theta'$), new parameters ($\theta^*$) were proposed. For the parameter uncertainty estimation using data from flask 1 in the training set, prior distribution was assumed to follow uniform distribution since no prior knowledge was available. After the parameter estimation, the posterior distribution of flask 1 was used as the prior distribution for the parameter estimation using flask 2. Similarly, the procedure of flask 3 used the posterior distribution of flask 2 as the prior distribution. Based on the proposed parameters, model simulation was performed using the updated kinetic model, and simulation results at time points that experiment samples were taken were collected as a matrix $y^*$. Then the likelihood can be computed using the following equation.

$$L(y|\theta) = e^{(-\frac{1}{2}\sum_i\sum_j(\frac{y_{i,j}-y^*_{i,j}}{\sigma_{i,j}})^2)} \quad (3\text{-}14)$$

where $y_{i,j}$ is the experimental data of $j$ variable at time $i$ and $y^*_{i,j}$ is the corresponding simulation result; $\sigma_{i,j}$ is the standard deviation of $j$ variable at time $i$ computed based on data in the training set.

The product of likelihood and prior probability of proposed parameters was divided by the product of likelihood and prior probability of current parameters generating the ratio $b$. Then the ratio $b$ was further refined to make sure that it was not greater than 1 giving the new value $c$. A random number $a \sim U(0|1)$ was used as the criterion. If $b$ is greater than $a$, the proposed model parameters would be accepted and replace current parameters. Otherwise, current parameters would be kept. Finally, current parameters and their simulation result will be recorded to estimate the posterior distribution.

## 3.3. Results and discussion

### 3.3.1. Results of frequentist method

#### 3.3.1.1. Random sample generation

Before performing model fitting, random samples generated based on the estimated univariate or multivariate Gaussian distribution were checked and compared. In the case of univariate Gaussian, samples of a variable at different time points were independently distributed. By contrast, samples generated by multivariate Gaussian distribution had covariance if data points were from the same variable at different time points. The result of the covariance matrix supported that the covariance between data points existed. In the covariance matrix, the diagonal entries are the variance of data points while non-diagonal entries are the covariance between data points. If these covariances are 0, it means that no covariance is found, and data points are independent to each other. Based on the given data, I found that covariances between data points from the same variable were nonzero (data not shown). Therefore, it was necessary to take covariance into consideration when implementing the MC technique. To further understand the impact of covariance, data points generated with univariate and multivariate Gaussian distributions were visualized. Glucose concentrations at 96 h and 120 h were used as an example as shown in **Figure 17**.

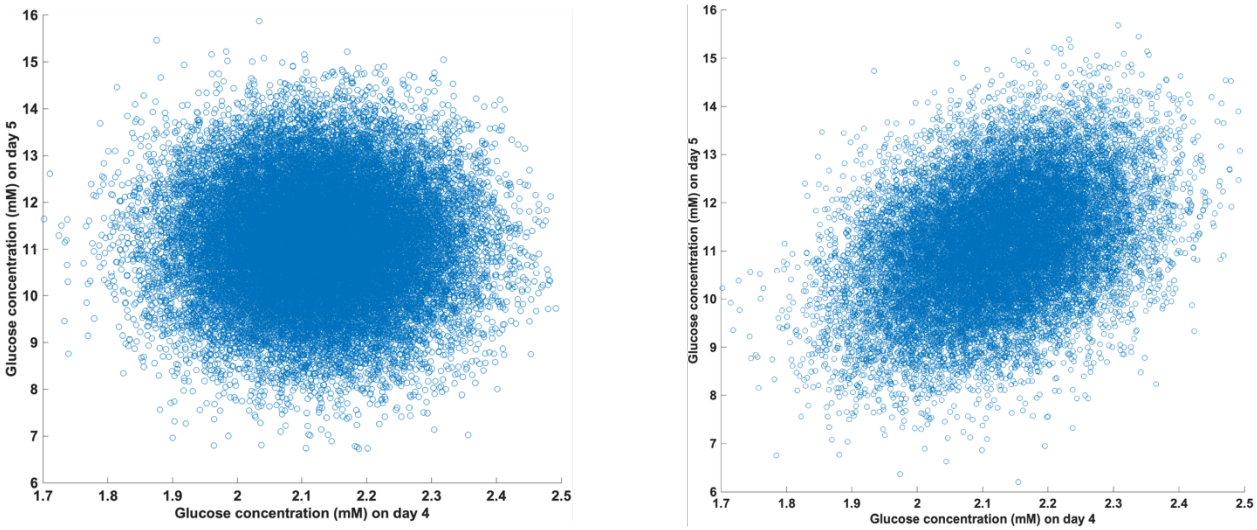


**Figure 17**: Data points of glucose concentration at 96 h versus 120 h generated by univariate (left) and multivariate (right) Gaussian distribution.

In **Figure 17**, on the left side, samples of glucose concentration at 96 h versus 120 h based on univariate Gaussian distribution were plotted. The points roughly formed a circle implying that glucose concentration at 96 h was independent to that at 120 h. More concretely, if vertical slices were taken at different glucose concentrations at 96 h, the glucose concentrations at 120 h followed the same Gaussian distribution. By contrast, points on the right side formed an ellipse roughly indicating the impact of covariance. If vertical slices were taken at different glucose concentrations at 96 h, it can be found that the distributions of those concentrations at 120 h were changed. If glucose concentration at 96 h was high, it was more likely that glucose concentration at 120 h was high. The same conclusion can also be drawn if taking horizontal slices at different glucose concentrations at 120 h.

Considering the scenario in the experiment. Glucose consumption is a cumulative procedure. When glucose concentration at specific time is high, it means the start point of glucose consumption is high and therefore glucose concentration at the later time would be high with higher possibility. Besides, high glucose concentration could also be due to low glucose consumption rate during previous period, which also implies low glucose consumption in the future resulting in high glucose concentration at later time.

Therefore, by comparing samples generated by univariate and multivariate Gaussian distribution, the importance of covariance was confirmed. Cell culture is a cumulative process and data points are not independent to each other. However, one of the limitations of this study was that the data set provided might be not large enough to estimate multivariate Gaussian distribution because only 3 data points were provided at the same time point. Besides, in this case, the covariance was only assumed within a variable. Covariance across different variables was not taken into account. For instance, when viable cell density is high, whether glucose concentration is likely to be low or not since more cells are consuming the glucose. Thus, covariance across different variables needs to be discussed as well when a huge amount of data is available.

#### 3.3.1.2. Uncertainty quantification of model parameters and process variables

Based on generated samples, model fitting was performed 100,000 times for either univariate or multivariate Gaussian distribution. After that, model parameters and simulation distributions were recorded. **Table 4** below lists the means and standard deviations of some model parameters. With the help of the two-sample t-test, statistically significant differences between parameter distributions using samples from univariate and multivariate Gaussian distributions were determined in several model parameters and labeled. It should be noted that values in the table were rounded off. Even though some model parameters in the table showed the same mean and standard deviation, they still had small differences in their mean and standard deviation in fact. Besides, since the model fitting has been conducted many times, the number of samples used for two-sample t-test was quite large. Therefore, small differences in mean and standard deviation between univariate and multivariate methods could still lead to statistically significant differences. It was found that distributions of $X_{t,max}$, $k_{D,min}$, $k_{D,dead}$ and $Y_{X_{v/Gln}}$ estimated based on univariate and multivariate Gaussian distribution samples were significantly different suggesting that two different sample generation methods can lead to different distribution of model parameters in some cases. Meanwhile, no significant difference was determined in the distributions of $r_{max}$, $Y_{Amm/Gln}$ and $VLP_{max}$, which indicates that distributions of some estimated model parameters were not influenced by covariance of variables at different time points.

**Table 4**: Model parameters including both means and standard deviations estimated by frequentist approach based on samples taken from univariate or multivariate Gaussian distribution. Significant differences between parameter distributions from univariate and multivariate Gaussian distributions were labeled by * ($p<0.05$).

| Parameter | Univariate (mean±s.d.) | Multivariate (mean±s.d.) |
|---|---|---|
| $r_{max}$ ($10^{-2}$ $h^{-1}$) | 5.12±0.17 | 5.12±0.18 |
| $X_{t,max}$ ($10^{10}$ cells $L^{-1}$)* | 1.27±0.05 | 1.27±0.05 |
| $k_{D,min}$ ($10^{-4}$ $h^{-1}$)* | 2.44±0.22 | 2.44±0.21 |
| $k_{D,dead}$ ($10^{-2}$ $h^{-1}$)* | 3.25±0.28 | 3.25±0.26 |
| $Y_{X_{v/Gln}}$ ($10^{12}$ cells $mol^{-1}$)* | 1.03±0.06 | 1.03±0.06 |
| $Y_{Amm/Gln}$ (-) | 1.26±0.05 | 1.26±0.05 |
| $VLP_{max}$ ($10^{-14}$ g $cell^{-1}$ ) | 4.54±0.47 | 4.54±0.47 |

To further investigate the performance of simulation with uncertainties using frequentist approach, the simulation of variables based on univariate or multivariate Gaussian distribution was plotted. Besides, within-band score was also listed in **Table 5** quantifying the performance of each method in the training set or test set.

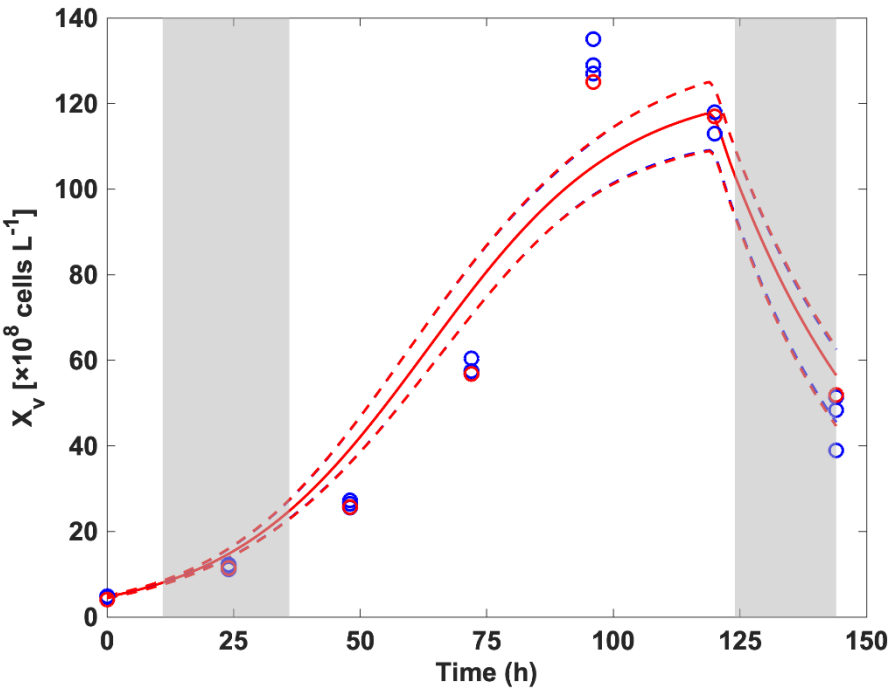


**Figure 18**: Simulation result with uncertainty of viable cell density using frequentist approach. Experimental data in the training set was plotted by blue dots and that in the test set was plotted by red dots. Red and blue solid lines (overlapped) are means of simulation based on generated samples from univariate and multivariate Gaussian distribution respectively. Red and blue dashed lines contain the 95% prediction intervals of simulation based on univariate and multivariate Gaussian distribution respectively. The gray area indicates that the distributions of simulation results using univariate and multivariate Gaussian distributions show statistically significant difference according to two-sample t-test ($p<0.05$).

As shown in **Figure 18**, the simulation result of viable cell density using frequentist approach was plotted as well as the experimental data. 95% prediction intervals were indicated by dashed lines. Based on the graph, it can be found that either univariate or multivariate Gaussian distribution methods could not contain most of the data points in both training and test sets. The main reason is the simplicity of the kinetic model. To improve the accuracy, increasing the number of model parameters would be helpful. For instance, using Monod-type equations would be beneficial for fitting the experimental data. However, it is noteworthy that there exists a trade-off between generality and precision (N. P. Shirsat et al., 2015). In this case, the accuracy was sacrificed to guarantee the generality. Besides, as shown in **Table 5**, the within-band score was also low indicating that the kinetic model should be further improved.

**Table 5**: Within-band score of frequentist approach based on univariate or multivariate Gaussian distribution in the training set or test set.

| | Training-univariate | Training-multivariate | Test-univariate | Test-multivariate |
|---|---|---|---|---|
| $X_t$ | 0.38 | 0.38 | 0.29 | 0.29 |
| $X_v$ | 0.38 | 0.38 | 0.29 | 0.29 |
| $X_d$ | 0.29 | 0.29 | 0.14 | 0.14 |
| $c_{Gln}$ | 0.29 | 0.29 | 0.14 | 0.14 |
| $c_{Amm}$ | 0.43 | 0.48 | 0.14 | 0.14 |
| $c_{Glc}$ | 0.90 | 0.95 | 0.86 | 1.00 |
| $c_{Lac}$ | 0.57 | 0.57 | 0.57 | 0.57 |
| $c_{VLP}$ | 0.29 | 0.29 | 0.14 | 0.14 |

In addition, dead cell density was also simulated and plotted in the same way as shown by **Figure 19**. Simulation using univariate and multivariate Gaussian distribution has shown significant difference between 96 h and 121 h suggesting that the covariance from 96 h to 120 h had an impact on the simulation using frequentist approach. The within-band scores were also low in both training and test sets. One of the possible reasons is that the specific death rate was a constant in the differential equation. Changing the specific death rate from a constant to a Monod-type term would be helpful for the model accuracy. However, increasing the number of model parameters could also lead to overfitting. This model was able to capture all of the data points at the end of cell culture, which was the most important part since most of the product was released during this period. Therefore, even though the precision was not high, it was still acceptable. Another reason for the deviation between the data points and prediction intervals was that the change of specific death rate was selected to be 120 h arbitrarily. Due to the sampling frequency, it was difficult to determine the exact time when cells started to die dramatically.

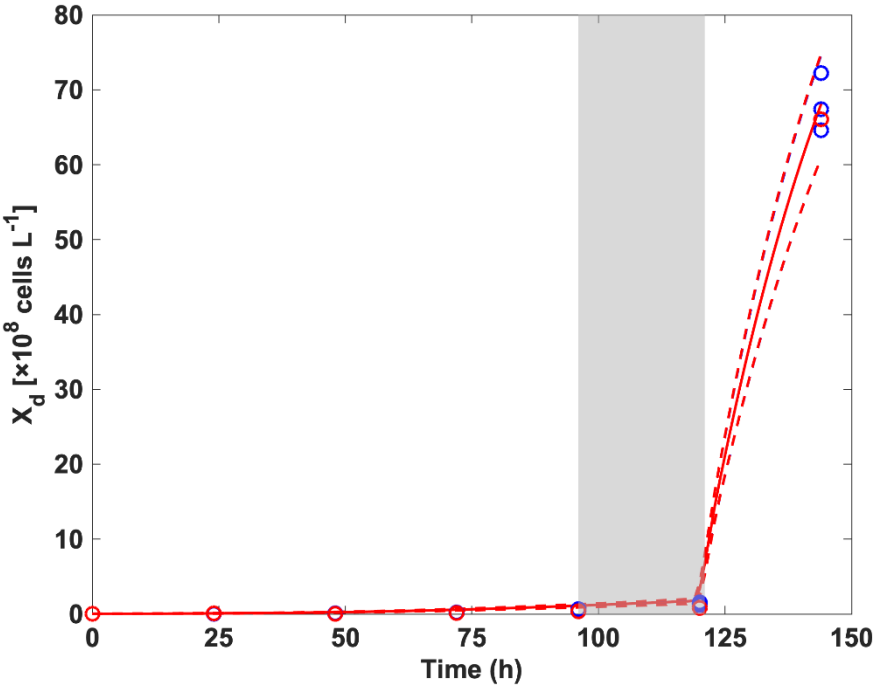


**Figure 19**: Simulation result with uncertainty of dead cell density using frequentist approach. Experimental data in the training set was plotted by blue dots and that in the test set was plotted by red dots. Red and blue solid lines (overlapped) are means of simulation based on generated samples from univariate and multivariate Gaussian distribution respectively. Red and blue dashed lines contain the 95% prediction intervals of simulation based on univariate and multivariate Gaussian distribution respectively. The gray area indicates that the distributions of simulation results using univariate and multivariate Gaussian distributions show statistically significant difference according to two-sample t-test ($p<0.05$).

In contrast to the previous simulations whose prediction intervals were unable to contain most of the data points, frequentist approach has shown high accuracy in modeling the glucose concentration during the cell culture. The result of the glucose simulation is plotted in **Figure 20** below. In **Figure 20**, it could be found that almost all of the data points were contained by the prediction intervals from both univariate Gaussian and multivariate Gaussian. As indicated by the gray area, simulation results from different Gaussian distributions showed significant difference from the beginning of cell culture until 100 h approximately. This statistically significant difference suggested the importance of taking covariance between data points into account. Besides, simulation using multivariate Gaussian distribution generated a slightly larger prediction interval (blue dashed lines) over that generated by univariate Gaussian distribution (red dashed

lines) and therefore contained more data points. As a result, multivariate Gaussian gave higher within-band scores in both the training group and test group than univariate Gaussian, which can be found in **Table 5**.

Therefore, the result of the glucose concentration simulation indicated that multivariate Gaussian distribution performed better than univariate Gaussian when applying frequentist approach for modeling glucose concentration. Besides, as shown in **Table 5**, multivariate Gaussian provided better performance in the simulation for glucose as well as ammonia and comparable performance in other variables. So, this result suggested that compared with assuming independently distributed data points, considering covariance between data points could better reflect the real world. As mentioned before, cell culture is a cumulative process. The current data point depends on previous data points. Thus, covariance between data points should not be ignored when applying not only frequentist approach for uncertainty quantification but also other sampling procedures. More concretely, with the development of machine learning, many powerful modeling methods that could potentially be applied in bioprocess have been proposed such as ANN. However, such kind of machine learning algorithms require a huge amount of data, which is unrealistic in the field of bioprocess because of the high cost of experiments. Data augmentation can be utilized to generate data required for driving machine learning algorithms based on a few observations by assuming Gaussian distribution and performing sampling processes. According to the result of this research, assuming multivariate Gaussian distributions would be more appropriate, which can provide better modeling performance.

The limitation of this part was that covariance was only assumed within a single variable. Nevertheless, it should be further discussed whether covariance exists across different variables because variables in cell culture can influence each other. For instance, at the current data point, if viable cell density is high, the glucose consumption would be fast resulting in lower glucose concentration in the next data point. In this research, the number of replications was not large enough to estimate covariance across different variables. However, such kind of covariance is worth considering when applying sampling techniques or data augmentation.

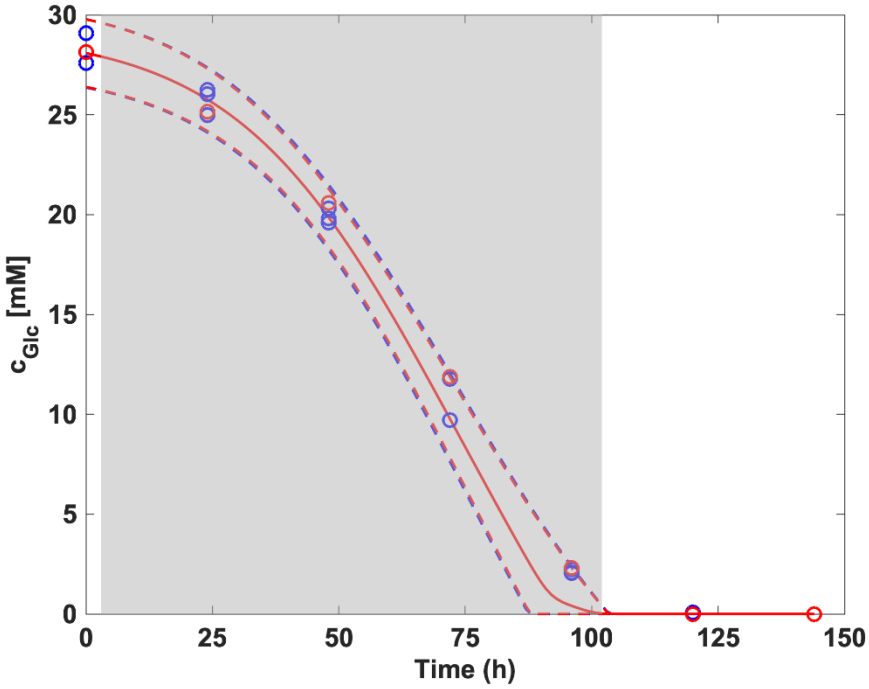


**Figure 20**: Simulation result with uncertainty of glucose concentration using frequentist approach. Experimental data in the training set was plotted by blue dots and that in the test set was plotted by red dots. Red and blue solid lines (overlapped) are means of simulation based on generated samples from univariate and multivariate Gaussian distribution respectively. Red and blue dashed lines contain the 95% prediction intervals of simulation based on univariate and multivariate Gaussian distribution respectively. The gray area indicates that the distributions of simulation results using univariate and multivariate Gaussian distributions show statistically significant difference according to two-sample t-test ($p<0.05$).

**Figure 21** shows the simulation of VLP concentration using two distributions. The uncertainties were quantified with 95% prediction intervals. Gray area displayed the period when simulation results using multivariate and univariate Gaussian distributions had statistically significant difference. The gray period was quite short suggesting that covariance was not so important in this case. The accuracy of modeling VLP concentration was low in both methods because only the last data point was contained by the prediction intervals in **Figure 21**. Also, in **Table 5**, within-band scores of VLP concentration in both training set and test set were low. The main reason was that the model could not simulate dead cell density with high accuracy in the previous part. VLP concentration was linearly correlated with dead cell density. As a result, the model simulation failed to contain data points at the beginning of cell culture. Another reason was the nature of the

cost function. wRMSD was selected as the cost function to be minimized when performing computation. This cost function weighs data points with large numerical values more than data points with small numerical values. As a result, the final VLP concentrations at 144 h were modeled within prediction intervals while other data points fell outside of the intervals. Considering that most of the VLPs released by the end of cell culture and final product concentration is the most important, this simulation results were acceptable even though their within-band scores were low.

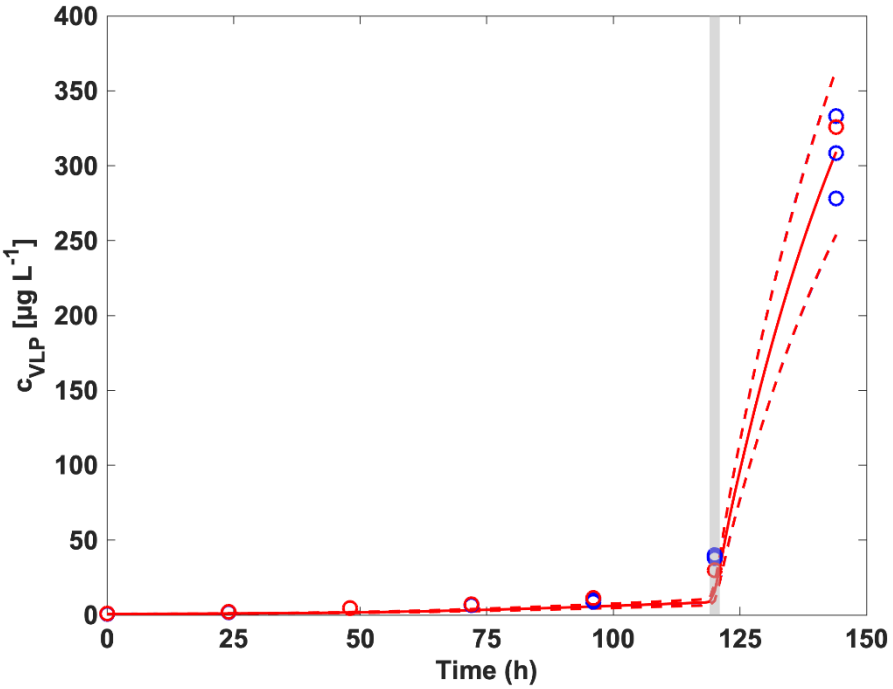


**Figure 21**: Simulation result with uncertainty of VLP concentration using frequentist approach. Experimental data in the training set was plotted by blue dots and that in the test set was plotted by red dots. Red and blue solid lines (overlapped) are means of simulation based on generated samples from univariate and multivariate Gaussian distribution respectively. Red and blue dashed lines contain the 95% prediction intervals of simulation based on univariate and multivariate Gaussian distribution respectively. The gray area indicates that the distributions of simulation results using univariate and multivariate Gaussian distributions show statistically significant difference according to two-sample t-test ($p<0.05$).

#### 3.3.1.3. Comparison of uncertainties between cultures with bioreactors and flasks

**Table 6**: Coefficient of variation (CV) for the comparison among model parameters from fed-batch culture with bioreactors and batch culture with flasks.

| | Fed-batch with bioreactors | Batch with flasks (univariate) | Batch with flasks (multivariate) |
|---|---|---|---|
| $r_{max}$ | 4.6% | 3.3% | 3.5% |
| $X_{t,max}$ | 4.0% | 4.1% | 4.1% |
| $VLP_{max}$ | 9.1% | 10.3% | 10.3% |

In order to compare the level of uncertainties between fed-batch culture using bioreactors and batch culture using flasks, it was inappropriate to compare standard deviations of model parameters directly because these model parameters were at different scales. Therefore, the coefficient of variation (CV) was utilized to enable comparison across different numerical scales as it normalized standard deviations with their means respectively.

As shown in **Table 6**, three model parameters were selected for the comparison because they were important parameters for cell growth and VLP production. It was found that fed-batch culture generally had higher cv (4.6%) over batch culture (3.3% and 3.5%) in $r_{max}$. This means the rate of cell growth was more consistent in the batch culture using flasks. Besides, in terms of batch culture, parameters computed by frequentist approach based on univariate and multivariate Gaussian distribution also had small difference in CV indicating the impact of covariance between data points. In addition, CVs of $X_{t,max}$ were comparable between cultures with bioreactors and flasks. So, the variation levels of maximum total cell density were similar in different culture modes. Finally, in terms of $VLP_{max}$ which is an important parameter for the VLP production, batch culture with flasks had slightly higher CV over that of fed-batch culture. This result supports the conclusion made in Chapter 2. In the previous chapter, it was concluded that even though STR2 had more dead and lysed cells, the VLP concentration was not improved much due to the batch-to-batch variation in $VLP_{max}$. The result in batch culture also confirmed that even cell cultures

were performed in parallel, variation in $VLP_{max}$ still existed and its uncertainty was comparable with that in fed-batch culture. Therefore, the conclusion made in the Chapter 2 was reasonable.

### 3.3.2. Result of Bayesian method

The computational procedure of Bayesian method was described in Section 3.2.5 using the Metropolis-Hasting algorithm. However, the computation cannot be performed appropriately, and a problem was found when running the script.

In the Metropolis-Hasting algorithm, a new set of model parameters would be proposed for each run. Then, the likelihood function can be calculated based on the proposed parameters and multiplied by the probability of proposing these parameters. After that, the product from proposed parameters is compared with that from current parameters deciding whether to replace current parameters with proposed parameters or not. However, to enable the comparison, the likelihood function should not be 0.

In this case, a few failure points at which the likelihood function became 0 were found. **Figure 22** shows an example of the failure point when the likelihood function was computed at the time point of 120 h. When a set of model parameters was proposed, a probability density function would be defined at a given point as shown by the blue peak and the probability of observing the experimental data could be computed. However, as shown in **Figure 22**, the data point at 120 h was quite far from the probability distribution, which meant the likelihood at that point was almost 0. This phenomenon was due to the low accuracy of the kinetic model. The overall likelihood was given by the product of the likelihood at each data point. Therefore, the comparison between proposed and current model parameters cannot be conducted because the likelihood functions were always zero and the script could not be run as expected.

This failed result suggested the limitation of Bayesian approach. It cannot work properly when the accuracy of the defined model structure is low. High-quality model is not always available, especially for modeling bioprocesses. If the model simulation deviates from experimental data, the likelihood function would be 0 leading to the failure of comparison, which is due to the nature of

defined probability distribution. By contrast, in the frequentist approach, wRMSD was used as the cost function. If the data points are far from the simulation, their distances would be summed up and the minimization of cost function can still be performed. Therefore, compared with Bayesian approach, frequentist approach is more suitable when the accuracy of the kinetic model is not high enough.

To improve the model accuracy, increasing the number of model parameters would be helpful. However, it is noteworthy that more model parameters can also lead to the risk of overfitting and loss of generality. The tradeoff between model accuracy and generality should be carefully considered in different scenarios.

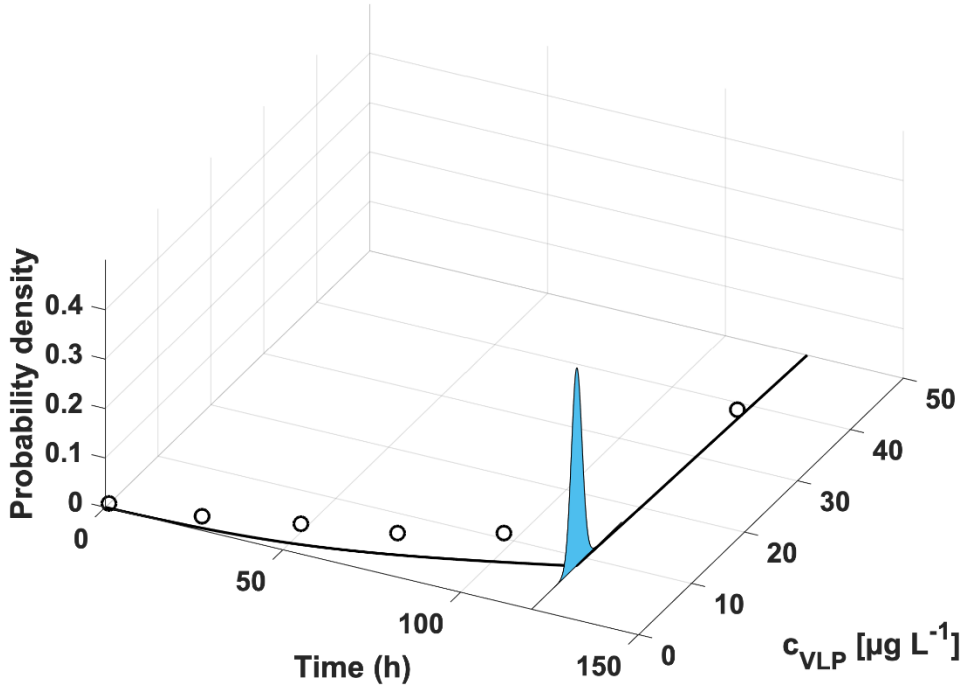


**Figure 22**: An example of the failure point when computing the likelihood function for the Metropolis-Hasting algorithm. Culture time (h) and VLP concentration ($\mu$g/L) were plotted by the x and y axis respectively. The probability density (likelihood function) of VLP concentration at 120 h, when model parameters were given, was plotted by the z-axis and shown by the blue peak.

## 3.4. Conclusion

In Chapter 2, a problem of batch-to-batch variation was noticed and the limitation of the deterministic model was realized because it was unable to provide information about the uncertainty of the process. Therefore, in this chapter, I intended to quantify the uncertainty of batch culture of CVA6 VLP-producing CHO cells using frequentist and Bayesian approaches. Besides, I wanted to compare two approaches in modeling uncertainties of the biopharmaceutical production process using CHO cells, which was not reported by other researchers before.

Data acquisition was carried out by cultivating 4 flasks of CVA6 VLP-producing CHO cells in batch mode in parallel. Samples were taken from flasks every 24 h and used for subsequent analysis including cell counting, metabolite concentration measurement, and VLP concentration quantification. Data from 3 flasks were used as the training set to train the model and the rest of the data was used as unseen data to test the performance of the model. The previously developed kinetic model in Chapter 2 was updated to fit the behavior of cells in the batch culture. This updated model was then utilized for conducting frequentist and Bayesian approaches.

Frequentist approach was based on the assumption that the distribution of data points followed Gaussian distribution. There were two choices of Gaussian distribution namely univariate and multivariate Gaussian distribution. Univariate assumed data at different time points were independent while multivariate assumed that there existed covariance between data points. To investigate the impact of covariance, both univariate and multivariate Gaussian distributions were applied to perform frequentist approach. Data generated based on univariate and multivariate Gaussian distribution has shown different behaviors. For example, if glucose concentration at 96 h was high, its concentration at 120 h was more likely to be high. This result indicated the impact of covariance between data points. Considering glucose consumption was a cumulative process, the result was reasonable. Besides, simulation results also had differences when using two types of Gaussian distributions. Results from multivariate Gaussian distribution generated larger prediction intervals in modeling glucose concentration and contained more experimental data points. Meanwhile, multivariate Gaussian distribution outperformed univariate in terms of within-band score in ammonia and glucose concentration meaning that multivariate Gaussian distribution

was a better choice for performing probabilistic modeling in this case and covariance between data points should be considered for better reflecting the real process. Besides, the finding about covariance can also be extended to other sampling techniques such as data augmentation. It is noteworthy that in this chapter, covariance was only assumed within every single variable. Covariance between different variables should be further investigated and discussed when more data is available. In addition to the finding about covariance, probabilistic models were developed using frequentist approach based on two types of Gaussian distributions. However, prediction intervals provided by model simulation failed to contain most of the data points in modeling several variables such as dead cell density and VLP concentration. Low accuracy of the kinetic model and the nature of the cost function selected (wRMSD) were the two main reasons responsible for the deviation between real data and simulation. However, the probabilistic model was still capable of simulating the final points of cell culture which was the most important period because most of the product was produced in this period. Therefore, the simulation results of probabilistic models were still acceptable.

In terms of Bayesian approach, the computation could not be carried out as expected using the Metropolis-Hasting algorithm because the likelihood function was 0 at several data points. It was caused by the low accuracy of the kinetic model since the simulation deviated from experimental data resulting in 0 probability. This failure implied the limitation of Bayesian approach. Compared with frequentist approach, Bayesian approach is incapable of developing probabilistic models when high-quality defined model structures are not available. In the field of bioprocess, high-quality models are not always available and there exists a tradeoff between accuracy and generality. Therefore, frequentist approach with multivariate Gaussian distribution might be a better choice for developing probabilistic models in CHO cells when the accuracy of kinetic models is low.

Chapter 4:

# Conclusions and future perspectives

# 4. Chapter 4: Conclusions and future perspectives

In recent years, many outbreaks of HFMD have been reported all over the world (Mirand et al., 2021; Puenpa et al., 2019; Solomon et al., 2010; Wu et al., 2010), becoming a threat to public health. Among different serotypes of enteroviruses, which are viruses causing HFMD, CVA6 is gaining attention because of its increasing incidence (Kimmis et al., 2018) and its ability to infect adults (Ramirez-Fort et al., 2014) not only children. VLP vaccines are a promising prevention measure against the CVA6 and as a novel type of vaccine, they are generally safer than conventional vaccines due to the absence of genetic materials (Nooraei et al., 2021). Therefore, the CVA6 VLP vaccine was developed.

To produce the CVA6 VLP vaccine, CHO cells were selected for their advantages including ease of scalability, capability of performing human-like PTMs (O'Flaherty et al., 2020), ability to grow in chemically defined serum-free medium (Lai et al., 2013), and low biosafety risk (Bandaranayake and Almo, 2014). As the industry of biopharmaceuticals is moving towards digitalization, many kinetic models have been developed as powerful and versatile tools to facilitate bioprocess development in many aspects such as reducing cost, saving time (Möller et al., 2019), and enabling quantitative decision-making. As a well-developed cell line, many kinetic models have been constructed for CHO cells. However, it was found that the production of CVA6 VLP vaccine using CHO cells has shown different behavior which cannot be modeled by the current kinetic models.

Therefore, in this research, I proposed a novel kinetic model for the production of the CVA6 VLP vaccine in CHO cells. To the best of my knowledge, this is the first kinetic model for the production of VLP vaccines in CHO cells. With the help of this kinetic model, I performed deterministic modeling in Chapter 2 and then conducted a series of analysis to facilitate the process development of CVA6 VLP vaccine production. After the analysis, a limitation of the deterministic model was noticed. Such kind of model cannot provide any information about the uncertainties of the process. Considering that batch-to-batch variation commonly exists in the bioprocess, it is important to quantify the uncertainty in the CVA6 VLP production process. Thus, I conducted

probabilistic modeling of CVA6 VLP production in Chapter 3. To do so, two mathematical approaches, namely frequentist approach and Bayesian approach, were utilized and compared. To the best of my knowledge, this is the first time that the comparison between frequentist and Bayesian approaches for modeling the production process using CHO cells was reported. After the comparison, the advantages and limitations of the two approaches were discussed.

In Chapter 2, I proposed a novel kinetic model for the production of CVA6 VLP vaccine in CHO cells and performed deterministic modeling. Based on the previous observations, the release of CVA VLPs strongly correlated with cell death and lysis. Therefore, the cell lysis needed to be quantified first because it cannot be detected by the image-based analyzers. So, I adapted a protocol that quantifies cell lysis indirectly by measuring dsDNA or LDH concentration proposed by a previous study (Klein et al., 2015). Based on the results of the thermal stability test, dsDNA was selected as the indicator because it was stable under the culture temperature while LDH was degraded at 37 °C. After the cell lysis quantification protocol was developed, cultivations of CVA6 VLP-producing CHO cells were carried out using 2 bioreactors in fed-batch mode. Different pH shift strategies were applied to investigate the impact of pH downshift. Experimental data was collected from samples taken from two bioreactors.

Subsequent to the data acquisition, the kinetic model was constructed by adapting equations proposed by previous studies or proposing new equations based on my understanding of the process. A novel equation that correlated VLP concentration with cell death and lysis was proposed. Simulation results have shown that the constructed model was able to simulate most variables of the production process such as viable cell density and VLP concentration with high accuracy ($R^2$>0.95). The newly proposed equation correlating VLP titer with dead and lysed cell density provides an opportunity of modeling other non-enveloped VLP production processes because the product release of non-enveloped VLPs relies on cell lysis (Cervera and Kamen, 2018). Moreover, according to our observations on other non-enveloped VLP production processes, the production of Enterovirus 71, Coxsackievirus A10, Coxsackievirus A16, and Norovirus VLPs have also shown a similar pattern, which can be modeled by the newly proposed equation. In addition to the simulation of VLP concentration, the kinetic model also quantified the impact of pH downshift. Specific ammonia production rate increased at low pH. As a result, higher ammonia concentration

was observed in STR2 resulting in accelerated cell death and higher dead cell density. However, product yield in STR2 was not improved over STR1 even though STR2 had more dead cells. This phenomenon was caused by the difference in $VLP_{max}$. Each cell in STR2 could contain less VLP content than STR1 resulting in insignificant improvement in the VLP yield. Considering that $VLP_{max}$ was reached before the pH shift, such difference was caused by the batch-to-batch variation due to the nature of cells. This finding also suggested the limitation of deterministic modeling which cannot provide information on the batch-to-batch variation and quantify uncertainties of the process.

After model parameters were determined based on experimental data, sensitivity analysis was conducted to investigate the impact of variations in the initial culture conditions and model parameters on the final VLP concentration. It was revealed that final VLP yield could be affected largely in response to the change in the initial glutamine concentration which was an important material attribute potentially. Besides, model parameters related to ammonia production and cell death such as $Y_{Amm/Gln}$ and $k_{D,dead}$ also had certain impact on the VLP production. The final VLP yield was also sensitive to the variation in $VLP_{max}$. It would be beneficial to validate the results of sensitivity analysis by experiments. However, it was challenging to change a single model parameter independently without changing others.

The result of sensitivity analysis also suggested that the VLP yield could be improved by shifting pH earlier than 240 h. To further determine the optimal pH shift timing, predictions were made with the help of the calibrated model. It was predicted that the highest VLP yield could be achieved by shifting pH on 192 h because pH downshift accelerated the accumulation of ammonia resulting in higher dead cell density. The model validation was absent due to the difficulty generated by the presence of batch-to-batch variation. However, the prediction result was still in line with previous reports. More concretely, several studies (Lee et al., 2021; Trummer et al., 2006) have reported that pH downshift could lead to higher ammonia levels in the culture of CHO cells. Besides, ammonia is known to be toxic to cells (Schneider, 1996) and can induce cellular apoptosis (Wang et al., 2018). Therefore, the prediction result was reasonable and logical.

According to the model prediction, it was noticed that high VLP yield was accompanied by high ammonia concentration. However, high ammonia concentration is also harmful to the product quality. Therefore, multiple objective optimization was combined with model simulation in order to balance product yield and quality. Based on the simulation, cell culture without pH shift at 311 h was the optimal operating condition that balances the tradeoff between product yield and quality. With the help of model simulation, the decision was made quantitatively. To summarize, the proposed kinetic model could work as a versatile tool facilitating bioprocess development in several aspects. However, the limitation of deterministic models was also realized. Deterministic models can only simulate a single value at a given time point without providing any information about the uncertainty. To overcome this limitation, probabilistic modeling was performed in the subsequent chapter.

The newly proposed model provides an opportunity to model production processes whose product release relies on cell death and lysis such as other non-enveloped VLP production in mammalian cells. The equation that correlates product concentration with dead and lysed cell density constructed in this research would be applicable to other processes. More specifically, according to observations on other non-enveloped VLPs in our laboratory, the production of Enterovirus 71, Coxsackievirus A10, Coxsackievirus A16, and Norovirus VLPs in CHO cells relies on cell lyses as well and the proposed equation can be applicable to simulate such kind of behavior. Therefore, subsequent work can investigate the generality of this model in simulating production processes of other non-enveloped VLPs in CHO cells. To adapt the model constructed in this research to other production processes, equations related to cell growth, death, lysis, as well as metabolite production and consumption need to be customized. Several boundaries were set arbitrarily in several equations based on the observation of cell behavior such as the time of lactate switch since its mechanism remains unclear. These boundaries also limit the generality of this model. This model was also limited under the assumption that volume change was negligible. To improve this model, an additional dilution term could be added to each corresponding variable to enable simulation in fed-batch mode. In terms of interpolation and extrapolation, mechanistic models are considered to perform well in extrapolation while black-box models are good at interpolation. As a mechanistic model, the kinetic model built in this study was expected to have a good extrapolation while its real performance should be further tested. Finally, this model

contributes to the implementation of digital twin through providing a new opportunity of modeling non-enveloped VLP production in CHO cells as models are critical components for the decision-making process in digital twin. The proposed model also has the potential to be combined with downstream models enabling optimization at the system level, which is also desired by the digital twin.

In Chapter 3, I carried out probabilistic modeling to quantify uncertainties of CVA6 VLP-producing CHO cell culture due to batch-to-batch variation. In bioprocess, there are a few sources of uncertainties such as variation of raw materials as well as experimental and measurement errors. For biopharmaceutical production, batch-to-batch consistency is critical in compliance with GMP. Therefore, studying uncertainties of the production process of CVA6 VLP was needed. To do so, 4 flasks of CVA6 VLP-producing CHO cells were cultivated in parallel in batch mode for the data acquisition. Data from 3 flasks were used for the model training and the remaining data from 1 flask worked as the test set. After the pre-processing of the data, the kinetic model constructed before was updated to fit cell behaviors in batch mode.

Subsequently, two mathematical approaches were applied to determine uncertainties of model parameters and therefore estimate uncertainties in the process. Frequentist approach was first used by utilizing MC sampling based on assumptions that data points followed either univariate or multivariate Gaussian distribution. Data generated based on two types of Gaussian distribution has shown different patterns indicating the existence of covariance between different data points. After data generation, parameter determination and model simulation were performed 100,000 times and their uncertainties were estimated by the distribution of simulation results. As a result, this probabilistic model was able to generate both means and 95% prediction intervals providing information about the uncertainty level of the process. However, in some variables of the process such as glutamine and VLP concentrations, the prediction intervals failed to contain most of the data points in both training set and test set because the accuracy of the updated kinetic model was not high enough. But they can still quantify uncertainties by the end of cell culture which was the most important period when most VLPs were released.

Another main finding of the result of frequentist approach was that simulation based on multivariate Gaussian distribution outperformed that of univariate Gaussian distribution. Such

result suggested the importance of taking covariance between data points into consideration when modeling bioprocess. As a cumulative process, data points collected from cell culture were not independent to each other but had covariance instead. This finding also provides a valuable point when performing other sampling techniques as such kind of techniques are required to drive large and powerful machine learning models.

Besides, uncertainties of several model parameters were compared between fed-batch culture using bioreactors and batch culture using flasks. It was found that uncertainty levels of model parameters were comparable between fed-batch and batch culture. More specifically, CVs of $VLP_{max}$ were comparable supporting the conclusion made in Chapter 2 that the difference in $VLP_{max}$ between STR1 and STR2 was due to batch-to-batch variation because a similar level of uncertainty was observed in batch culture even experiments were performed in parallel.

In addition to the frequentist approach, I also tried to apply Bayesian approach to determine uncertainties of model parameters using the Metropolis-Hasting algorithm. Nevertheless, the script cannot be run as expected because the accuracy of the kinetic model was low. In fact, in bioprocess, high-quality models are not always available. Therefore, if high-quality kinetic models are not available, frequentist approach might be a better choice compared with Bayesian approach for probabilistic modeling.

In terms of future perspectives, the generality of the proposed model needs to be further investigated. For example, the newly developed model can be applied to other non-enveloped VLP production processes to test its generality. Besides, more efforts should be made to improve the model accuracy in order to enable Bayesian approach. The kinetic model also has the potential to be integrated with machine learning algorithms enabling hybrid modeling. For instance, machine learning algorithms can be used to build up relationships between parameters of the kinetic model and different pH values if more data is available. One of the limitations of this research is that the performance of frequentist and Bayesian approaches was not compared successfully. Therefore, further study may be focused on investigating the advantages and disadvantages of two methods for probabilistic modeling of CHO cells. Another limitation of this research is that covariance between data points was assumed within a single variable due to the limited sample size. It would

be beneficial to study whether covariance between data points across different variables exists and whether such kind of covariance is critical for the data augmentation of bioprocess.

To summarize, probabilistic modeling is only applicable when the model for the production process has already been constructed. Besides, the accuracy of probabilistic models largely depends on the accuracy of the model for the process. For the frequentist approach, if high-quality models are available, it can generate prediction intervals providing precise information on the uncertainty of the process. It can also work when model accuracy is low, but the prediction intervals cannot contain most of observations. The Bayesian approach allows integration of prior knowledge into simulation, which is its unique advantage. However, when the high-quality models are deficient, it cannot run as expected. In addition, covariance was found to exist and play a role in model simulation. This finding is also valuable for the digital twin. To implement digital twins, advanced machine learning algorithms are necessary. Nevertheless, these algorithms require huge amounts of data, which is unpractical considering the expense of performing experiments. Therefore, data augmentation, which generates data artificially and is commonly used in machine learning in practice, is needed. This data augmentation in bioprocess can be performed by using sampling techniques under the assumption of Gaussian distribution. In such sampling procedures, covariance needs to be taken into account as it can affect the performance of simulation in order to guarantee high-quality data is generated, which is important for the model performance. Therefore, the finding about covariance between data points is valuable for the implementation of digital twin in the aspect of high-quality data generation.

## References

Al-Majmaie, R., Kuystermans, D., Al-Rubeai, M., 2021. Biopharmaceuticals Produced from Cultivated Mammalian Cells, in: Pörtner, R. (Ed.), Cell Culture Engineering and Technology, Cell Engineering. Springer International Publishing, Cham, pp. 3–52. https://doi.org/10.1007/978-3-030-79871-0_1

Arndt, L., Wiegmann, V., Kuchemüller, K.B., Baganz, F., Pörtner, R., Möller, J., 2021. Model-based workflow for scale-up of process strategies developed in miniaturized bioreactor systems. Biotechnol. Prog. 37, p.e3122. https://doi.org/10.1002/btpr.3122

Aswathyraj, S., Arunkumar, G., Alidjinou, E.K., Hober, D., 2016. Hand, foot and mouth disease (HFMD): emerging epidemiology and the need for a vaccine strategy. Med. Microbiol. Immunol. (Berl.) 205, 397–407. https://doi.org/10.1007/s00430-016-0465-y

Badr, S., Okamura, K., Takahashi, N., Ubbenjans, V., Shirahata, H., Sugiyama, H., 2021. Integrated design of biopharmaceutical manufacturing processes: Operation modes and process configurations for monoclonal antibody production. Comput. Chem. Eng. 153, 107422. https://doi.org/10.1016/j.compchemeng.2021.107422

Bandaranayake, A.D., Almo, S.C., 2014. Recent advances in mammalian protein production. FEBS Lett. 588, 253–260. https://doi.org/10.1016/j.febslet.2013.11.035

Brooks, S.A., 2004. Appropriate glycosylation of recombinant proteins for human use: Implications of choice of expression system. Mol. Biotechnol. 28, 241–256. https://doi.org/10.1385/MB:28:3:241

Cardillo, A.G., Castellanos, M.M., Desailly, B., Dessoy, S., Mariti, M., Portela, R.M.C., Scutella, B., von Stosch, M., Tomba, E., Varsakelis, C., 2021. Towards *in silico* process modeling for vaccines. Trends Biotechnol. 39, 1120–1130. https://doi.org/10.1016/j.tibtech.2021.02.004

Cervera, L., Kamen, A.A., 2018. Large-Scale Transient Transfection of Suspension Mammalian Cells for VLP Production, in: Picanço-Castro, V., Swiech, K. (Eds.), Recombinant Glycoprotein Production, Methods in Molecular Biology. Springer New York, New York, NY, pp. 117–127. https://doi.org/10.1007/978-1-4939-7312-5_10

Chen, P., Harcum, S., 2006. Effects of elevated ammonium on glycosylation gene expression in CHO cells. Metab. Eng. 8, 123–132. https://doi.org/10.1016/j.ymben.2005.10.002

Chen, Q., Lai, H., 2013. Plant-derived virus-like particles as vaccines. Hum. Vaccines Immunother. 9, 26–49. https://doi.org/10.4161/hv.22218

Chuenkitmongkol, S., Solante, R., Burhan, E., Chariyalertsak, S., Chiu, N.-C., Do-Van, D., Husin, M., Hwang, K.-P., Kiertiburanakul, S., Kulkarni, P.S., Lee, P.-I., Lobo, R.C., Nghia, C.H., Ong-Lim, A., Sivasampu, S., Suah, J.L., Tok, P.S.K., Thwaites, G., SEA Vaccine Effectiveness Expert Working Group, 2022. Expert review on global real-world vaccine effectiveness against SARS-CoV-2. Expert Rev. Vaccines 21, 1255–1268. https://doi.org/10.1080/14760584.2022.2092472

Colin Cameron, A., Windmeijer, F.A.G., 1997. An R-squared measure of goodness of fit for some common nonlinear regression models. J. Econom. 77, 329–342. https://doi.org/10.1016/S0304-4076(96)01818-0

Contreras-Gómez, A., Sánchez-Mirón, A., García-Camacho, F., Molina-Grima, E., Chisti, Y., 2014. Protein production using the baculovirus-insect cell expression system. Biotechnol. Prog. 30, 1–18. https://doi.org/10.1002/btpr.1842

Dahodwala, H., Lee, K.H., 2019. The fickle CHO: a review of the causes, implications, and potential alleviation of the CHO cell line instability problem. Curr. Opin. Biotechnol. 60, 128–137. https://doi.org/10.1016/j.copbio.2019.01.011

Donis, R.O., Chen, L.-M., Davis, C.T., Foust, A., Hossain, M.J., Johnson, A., Klimov, A., Loughlin, R., Xu, X., Tsai, T., Blayer, S., Trusheim, H., Colegate, T., Fox, J., Taylor, B., Hussain, A., Barr, I., Baas, C., Louwerens, J., Geuns, E., Lee, M.-S., Venhuizen, L., Neumeier, E., Ziegler, T., 2014. Performance characteristics of qualified cell lines for isolation and propagation of influenza viruses for vaccine manufacturing. Vaccine 32, 6583–6590. https://doi.org/10.1016/j.vaccine.2014.06.045

Edward Abraham. Dashline. Accessed August 9, 2022. https://www.mathworks.com/matlabcentral/fileexchange/1892-dashline

El Abd, Y., Tabll, A., Smolic, R., Smolic, M., 2022. Mini-review: The market growth of diagnostic and therapeutic monoclonal antibodies – SARS CoV-2 as an example. Hum. Antibodies 30, 15–24. https://doi.org/10.3233/HAB-211513

Esmonde-White, K.A., Cuellar, M., Uerpmann, C., Lenain, B., Lewis, I.R., 2017. Raman spectroscopy as a process analytical technology for pharmaceutical manufacturing and

bioprocessing. Anal. Bioanal. Chem. 409, 637–649. https://doi.org/10.1007/s00216-016-9824-1

Esposito, S., Principi, N., 2018. Hand, foot and mouth disease: current knowledge on clinical manifestations, epidemiology, aetiology and prevention. Eur. J. Clin. Microbiol. Infect. Dis. 37, 391–398. https://doi.org/10.1007/s10096-018-3206-x

Felberbaum, R.S., 2015. The baculovirus expression vector system: A commercial manufacturing platform for viral vaccines and gene therapy vectors. Biotechnol. J. 10, 702–714. https://doi.org/10.1002/biot.201400438

Frahm, B., Lane, P., Atzert, H., Munack, A., Hoffmann, M., Hass, V.C., Portner, R., 2002. Adaptive, model-based control by the open-loop-feedback-optimal (OLFO) controller for the effective fed-batch cultivation of hybridoma cells. Biotechnol. Prog. 18, 1095–1103. https://doi.org/10.1021/bp020035y

Fuenmayor, J., Gòdia, F., Cervera, L., 2017. Production of virus-like particles for vaccines. New Biotechnol. 39, 174–180. https://doi.org/10.1016/j.nbt.2017.07.010

Gargalo, C.L., De Las Heras, S.C., Jones, M.N., Udugama, I., Mansouri, S.S., Krühne, U., Gernaey, K.V., 2020. Towards the development of digital twins for the bio-manufacturing industry, in: Herwig, C., Pörtner, R., Möller, J. (Eds.), Digital Twins, Advances in Biochemical Engineering/Biotechnology. Springer International Publishing, Cham, pp. 1–34. https://doi.org/10.1007/10_2020_142

Gary, E.N., Weiner, D.B., 2020. DNA vaccines: prime time is now. Curr. Opin. Immunol. 65, 21–27. https://doi.org/10.1016/j.coi.2020.01.006

Gopaluni, R.B., Tulsyan, A., Chachuat, B., Huang, B., Lee, J.M., Amjad, F., Damarla, S.K., Kim, J.W., Lawrence, N.P., 2020. Modern machine learning tools for monitoring and control of industrial processes: A survey. IFAC-Pap. 53, 218–229. https://doi.org/10.1016/j.ifacol.2020.12.126

Hartley, F., Walker, T., Chung, V., Morten, K., 2018. Mechanisms driving the lactate switch in Chinese hamster ovary cells: Biotechnol. Bioeng. 115, 1890–1903. https://doi.org/10.1002/bit.26603

Hernández Rodríguez, T., Frahm, B., 2020. Digital seed train twins and statistical methods, in: Herwig, C., Pörtner, R., Möller, J. (Eds.), Digital Twins, Advances in Biochemical

Engineering/Biotechnology. Springer International Publishing, Cham, pp. 97–131. https://doi.org/10.1007/10_2020_137

Hernández Rodríguez, T., Posch, C., Schmutzhard, J., Stettner, J., Weihs, C., Pörtner, R., Frahm, B., 2019. Predicting industrial-scale cell culture seed trains–A Bayesian framework for model fitting and parameter estimation, dealing with uncertainty in measurements and model parameters, applied to a nonlinear kinetic cell culture model, using an MCMC method. Biotechnol. Bioeng. 116, 2944–2959. https://doi.org/10.1002/bit.27125

Hong, J., Liu, F., Qi, H., Tu, W., Ward, M.P., Ren, M., Zhao, Z., Su, Q., Huang, J., Chen, X., Le, J., Ren, X., Hu, Y., Cowling, B., Li, Z., Chang, Z., Zhang, Z., 2022. Changing epidemiology of hand, foot, and mouth disease in China, 2013−2019: a population-based study. Lancet Reg. Health - West. Pac. 20, 100370. https://doi.org/10.1016/j.lanwpc.2021.100370

Hong, J.K., Choi, D.-H., Park, S.-Y., Silberberg, Y.R., Shozui, F., Nakamura, E., Kayahara, T., Lee, D.-Y., 2022. Data-driven and model-guided systematic framework for media development in CHO cell culture. Metab. Eng. 73, 114–123. https://doi.org/10.1016/j.ymben.2022.07.003

Hütter, J., Rödig, J.V., Höper, D., Seeberger, P.H., Reichl, U., Rapp, E., Lepenies, B., 2013. Toward animal cell culture–based influenza vaccine design: Viral hemagglutinin *N*-glycosylation markedly impacts immunogenicity. J. Immunol. 190, 220–230. https://doi.org/10.4049/jimmunol.1201060

Ishimaru, Y., Nakano, S., Yamaoka, K., Takami, S., 1980. Outbreaks of hand, foot, and mouth disease by enterovirus 71. High incidence of complication disorders of central nervous system. Arch. Dis. Child. 55, 583–588. https://doi.org/10.1136/adc.55.8.583

Kennett, M.L., Birch, C.J., Lewis, F.A., Yung, A.P., Locarnini, S.A., Gust, D., 1974. Enterovirus type 71 infection in Melbourne.

Kern, S., Platas-Barradas, O., Pörtner, R., Frahm, B., 2016. Model-based strategy for cell culture seed train layout verified at lab scale. Cytotechnology 68, 1019–1032. https://doi.org/10.1007/s10616-015-9858-9

Kim, J.Y., Kim, Y.-G., Lee, G.M., 2012. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. Appl. Microbiol. Biotechnol. 93, 917–930. https://doi.org/10.1007/s00253-011-3758-5

Kimmis, B.D., Downing, C., Tyring, S., 2018. Hand-foot-and-mouth disease caused by coxsackievirus A6 on the rise. Cutis 102, 353–356.

Kis, Z., Kontoravdi, C., Shattock, R., Shah, N., 2020. Resources, production scales and time required for producing RNA vaccines for the Gglobal Pandemic Demand. Vaccines 9, 3. https://doi.org/10.3390/vaccines9010003

Klein, T., Heinzel, N., Kroll, P., Brunner, M., Herwig, C., Neutsch, L., 2015. Quantification of cell lysis during CHO bioprocesses: Impact on cell count, growth kinetics and productivity. J. Biotechnol. 207, 67–76. https://doi.org/10.1016/j.jbiotec.2015.04.021

Kontoravdi, C., Wong, D., Lam, C., Lee, Y.Y., Yap, M.G.S., Pistikopoulos, E.N., Mantalaris, A., 2007. Modeling amino acid metabolism in mammalian cells-toward the Development of a model library. Biotechnol. Prog. 23, 1261–1269. https://doi.org/10.1021/bp070106z

Kotidis, P., Kontoravdi, C., 2020. Harnessing the potential of artificial neural networks for predicting protein glycosylation. Metab. Eng. Commun. 10, e00131. https://doi.org/10.1016/j.mec.2020.e00131

Kroll, P., Eilers, K., Fricke, J., Herwig, C., 2017. Impact of cell lysis on the description of cell growth and death in cell culture. Eng. Life Sci. 17, 440–447. https://doi.org/10.1002/elsc.201600088

Lai, T., Yang, Y., Ng, S., 2013. Advances in mammalian cell line development technologies for recombinant protein production. Pharmaceuticals 6, 579–603. https://doi.org/10.3390/ph6050579

Lavado-García, J., Zhang, T., Cervera, L., Gòdia, F., Wuhrer, M., 2022. Differential *N* - and *O* -glycosylation signatures of HIV-1 Gag virus-like particles and coproduced extracellular vesicles. Biotechnol. Bioeng. 119, 1207–1221. https://doi.org/10.1002/bit.28051

Lee, A.P., Kok, Y.J., Lakshmanan, M., Leong, D., Zheng, L., Lim, H.L., Chen, S., Mak, S.Y., Ang, K.S., Templeton, N., Salim, T., Wei, X., Gifford, E., Tan, A.H., Bi, X., Ng, S.K., Lee, D., Ling, W.L.W., Ho, Y.S., 2021. Multi-omics profiling of a CHO cell culture system unravels the effect of culture pH on cell growth, antibody titer, and product quality. Biotechnol. Bioeng. 118, 4305–4316. https://doi.org/10.1002/bit.27899

Li, S.-W., Zhao, Q., Wu, T., Chen, S., Zhang, J., Xia, N.-S., 2015. The development of a recombinant hepatitis E vaccine HEV 239. Hum. Vaccines Immunother. 11, 908–914. https://doi.org/10.1080/21645515.2015.1008870

Lie, S., Wang, T., Forbes, B., Proud, C.G., Petersen, J., 2019. The ability to utilise ammonia as nitrogen source is cell type specific and intricately linked to GDH, AMPK and mTORC1. Sci. Rep. 9, 1461. https://doi.org/10.1038/s41598-018-37509-3

Luo, Y., Lovelett, R.J., Price, J.V., Radhakrishnan, D., Barnthouse, K., Hu, P., Schaefer, E., Cunningham, J., Lee, K.H., Shivappa, R.B., Ogunnaike, B.A., 2021. Modeling the effect of amino acids and copper on monoclonal antibody productivity and glycosylation: A modular approach. Biotechnol. J. 16, 2000261. https://doi.org/10.1002/biot.202000261

Mao, Q., Wang, Y., Yao, X., Bian, L., Wu, X., Xu, M., Liang, Z., 2014. Coxsackievirus A16: Epidemiology, diagnosis, and vaccine. Hum. Vaccines Immunother. 10, 360–367. https://doi.org/10.4161/hv.27087

Martínez, J.A., Bulté, D.B., Contreras, M.A., Palomares, L.A., Ramírez, O.T., 2020. Dynamic modeling of CHO cell metabolism using the hybrid cybernetic approach with a novel elementary mode analysis strategy. Front. Bioeng. Biotechnol. 8, 279. https://doi.org/10.3389/fbioe.2020.00279

Meng, C., Kuster, B., Culhane, A.C., Gholami, A.M., 2014. A multivariate approach to the integration of multi-omics datasets. BMC Bioinformatics 15, 162. https://doi.org/10.1186/1471-2105-15-162

Mirand, A., Cohen, R., Bisseux, M., Tomba, S., Sellem, F.C., Gelbert, N., Béchet, S., Frandji, B., Archimbaud, C., Brebion, A., Chabrolles, H., Regagnon, C., Levy, C., Bailly, J.-L., Henquell, C., 2021. A large-scale outbreak of hand, foot and mouth disease, France, as at 28 September 2021. Eurosurveillance 26, 2100978. https://doi.org/10.2807/1560-7917.ES.2021.26.43.2100978

Möller, J., Kuchemüller, K.B., Steinmetz, T., Koopmann, K.S., Pörtner, R., 2019. Model-assisted Design of Experiments as a concept for knowledge-based bioprocess development. Bioprocess Biosyst. Eng. 42, 867–882. https://doi.org/10.1007/s00449-019-02089-7

Mowbray, M., Savage, T., Wu, C., Song, Z., Cho, B.A., Del Rio-Chanona, E.A., Zhang, D., 2021. Machine learning for biochemical engineering: A review. Biochem. Eng. J. 172, 108054. https://doi.org/10.1016/j.bej.2021.108054

Narayanan, H., Sokolov, M., Morbidelli, M., Butté, A., 2019. A new generation of predictive models: The added value of hybrid models for manufacturing processes of therapeutic proteins. Biotechnol. Bioeng. 116, 2540–2549. https://doi.org/10.1002/bit.27097

Nooraei, S., Bahrulolum, H., Hoseini, Z.S., Katalani, C., Hajizade, A., Easton, A.J., Ahmadian, G., 2021. Virus-like particles: preparation, immunogenicity and their roles as nanovaccines and drug nanocarriers. J. Nanobiotechnology 19, 1-27. https://doi.org/10.1186/s12951-021-00806-7

Nyon, M.P., Du, L., Tseng, C.-T.K., Seid, C.A., Pollet, J., Naceanceno, K.S., Agrawal, A., Algaissi, A., Peng, B.-H., Tai, W., Jiang, S., Bottazzi, M.E., Strych, U., Hotez, P.J., 2018. Engineering a stable CHO cell line for the expression of a MERS-coronavirus vaccine antigen. Vaccine 36, 1853–1862. https://doi.org/10.1016/j.vaccine.2018.02.065

O'Flaherty, R., Bergin, A., Flampouri, E., Mota, L.M., Obaidi, I., Quigley, A., Xie, Y., Butler, M., 2020. Mammalian cell culture for production of recombinant proteins: A review of the critical steps in their biomanufacturing. Biotechnol. Adv. 43, 107552. https://doi.org/10.1016/j.biotechadv.2020.107552

Plotkin, S., 2014. History of vaccination. Proc. Natl. Acad. Sci. 111, 12283–12287. https://doi.org/10.1073/pnas.1400472111

Puenpa, J., Wanlapakorn, N., Vongpunsawad, S., Poovorawan, Y., 2019. The history of enterovirus A71 outbreaks and molecular epidemiology in the Asia-Pacific region. J. Biomed. Sci. 26, 1-11. https://doi.org/10.1186/s12929-019-0573-2

Qian, G., Mahdi, A., 2020. Sensitivity analysis methods in the biomedical sciences.

Ramirez-Fort, M.K., Downing, C., Doan, H.Q., Benoist, F., Oberste, M.S., Khan, F., Tyring, S.K., 2014. Coxsackievirus A6 associated hand, foot and mouth disease in adults: Clinical presentation and review of the literature. J. Clin. Virol. 60, 381–386. https://doi.org/10.1016/j.jcv.2014.04.023

Rearte, A., Castelli, J.M., Rearte, R., Fuentes, N., Pennini, V., Pesce, M., Barbeira, P.B., Iummato, L.E., Laurora, M., Bartolomeu, M.L., Galligani, G., Del Valle Juarez, M., Giovacchini, C.M., Santoro, A., Esperatti, M., Tarragona, S., Vizzotti, C., 2022. Effectiveness of rAd26-rAd5, ChAdOx1 nCoV-19, and BBIBP-CorV vaccines for risk of infection with SARS-CoV-2 and death due to COVID-19 in people older than 60 years in Argentina: a test-negative, case-control, and retrospective longitudinal study. The Lancet 399, 1254–1264. https://doi.org/10.1016/S0140-6736(22)00011-3

Riedel, S., 2005. Edward Jenner and the history of smallpox and vaccination. Bayl. Univ. Med. Cent. Proc. 18, 21–25. https://doi.org/10.1080/08998280.2005.11928028

Roohvand, F., Shokri, M., Abdollahpour-Alitappeh, M., Ehsani, P., 2017. Biomedical applications of yeast- a patent view, part one: yeasts as workhorses for the production of therapeutics and vaccines. Expert Opin. Ther. Pat. 27, 929–951. https://doi.org/10.1080/13543776.2017.1339789

Rosa, S.S., Prazeres, D.M.F., Azevedo, A.M., Marques, M.P.C., 2021. mRNA vaccines manufacturing: Challenges and bottlenecks. Vaccine 39, 2190–2200. https://doi.org/10.1016/j.vaccine.2021.03.038

Schneider, M., 1996. The importance of ammonia in mammalian cell culture. J. Biotechnol. 46, 161–185. https://doi.org/10.1016/0168-1656(95)00196-4

Shirsat, N., Mohd, A., Whelan, J., English, N.J., Glennon, B., Al-Rubeai, M., 2015. Revisiting Verhulst and Monod models: analysis of batch and fed-batch cultures. Cytotechnology 67, 515–530. https://doi.org/10.1007/s10616-014-9712-5

Shirsat, N.P., English, N.J., Glennon, B., Al-Rubeai, M., 2015. Modelling of mammalian cell cultures, in: Al-Rubeai, M. (Ed.), Animal Cell Culture, Cell Engineering. Springer International Publishing, Cham, pp. 259–326. https://doi.org/10.1007/978-3-319-10320-4_10

Shuman, S., 2015. RNA capping: progress and prospects. RNA 21, 735–737. https://doi.org/10.1261/rna.049973.115

Solomon, T., Lewthwaite, P., Perera, D., Cardosa, M.J., McMinn, P., Ooi, M.H., 2010. Virology, epidemiology, pathogenesis, and control of enterovirus 71. Lancet Infect. Dis. 10, 778–790. https://doi.org/10.1016/S1473-3099(10)70194-8

Takahashi, S., Liao, Q., Van Boeckel, T.P., Xing, W., Sun, J., Hsiao, V.Y., Metcalf, C.J.E., Chang, Z., Liu, F., Zhang, J., Wu, J.T., Cowling, B.J., Leung, G.M., Farrar, J.J., van Doorn, H.R., Grenfell, B.T., Yu, H., 2016. Hand, Foot, and Mouth Disease in China: Modeling epidemic dynamics of Enterovirus serotypes and implications for vaccination. PLOS Med. 13, e1001958. https://doi.org/10.1371/journal.pmed.1001958

Travieso, T., Li, J., Mahesh, S., Mello, J.D.F.R.E., Blasi, M., 2022. The use of viral vectors in vaccine development. Npj Vaccines 7, 75. https://doi.org/10.1038/s41541-022-00503-y

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. https://doi.org/10.1016/S0264-410X(01)00053-6

Trummer, E., Fauland, K., Seidinger, S., Schriebl, K., Lattenmayer, C., Kunert, R., Vorauer-Uhl, K., Weik, R., Borth, N., Katinger, H., Müller, D., 2006. Process parameter shifting: Part I. Effect of DOT, pH, and temperature on the performance of Epo-Fc expressing CHO cells cultivated in controlled batch bioreactors. Biotechnol. Bioeng. 94, 1033–1044. https://doi.org/10.1002/bit.21013

Van Oers, M.M., 2006. Vaccines for viral and parasitic diseases produced with Baculovirus vectors, in: Advances in virus research. 68, 193–253. https://doi.org/10.1016/S0065-3527(06)68006-8

Vartak, A., Sucheck, S., 2016. Recent Advances in Subunit Vaccine Carriers. Vaccines 4, 12. https://doi.org/10.3390/vaccines4020012

Verhoeckx, K.C.M., Bijlsma, S., De Groene, E.M., Witkamp, R.F., Van Der Greef, J., Rodenburg, R.J.T., 2004. A combination of proteomics, principal component analysis and transcriptomics is a powerful tool for the identification of biomarkers for macrophage maturation in the U937 cell line. PROTEOMICS 4, 1014–1028. https://doi.org/10.1002/pmic.200300669

von Stosch, M., Portela, R.M., Varsakelis, C., 2021. A roadmap to AI-driven *in silico* process development: bioprocessing 4.0 in practice. Curr. Opin. Chem. Eng. 33, 100692. https://doi.org/10.1016/j.coche.2021.100692

Wang, F., Chen, S., Jiang, Y., Zhao, Y., Sun, L., Zheng, B., Chen, L., Liu, Z., Zheng, X., Yi, K., Li, C., Zhou, X., 2018. Effects of ammonia on apoptosis and oxidative stress in bovine mammary epithelial cells. Mutagenesis 33, 291–299. https://doi.org/10.1093/mutage/gey023

Wu, Y., Yeo, A., Phoon, M.C., Tan, E.L., Poh, C.L., Quak, S.H., Chow, V.T.K., 2010. The largest outbreak of hand; foot and mouth disease in Singapore in 2008: The role of enterovirus 71 and coxsackievirus A strains. Int. J. Infect. Dis. 14, e1076–e1081. https://doi.org/10.1016/j.ijid.2010.07.006

Wurm, F.M., De Jesus, M., 2016. Manufacture of recombinant therapeutic proteins using Chinese hamster ovary cells in large-scale bioreactors: History, methods, and perspectives, in: Liu, C., Morrow, K.J. (Eds.), Biosimilars of Monoclonal Antibodies. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp. 327–353. https://doi.org/10.1002/9781118940648.ch11

Wurm, M.J., Wurm, F.M., 2021. Naming CHO cells for bio-manufacturing: Genome plasticity and variant phenotypes of cell populations in bioreactors question the relevance of old names. Biotechnol. J. 16, 2100165. https://doi.org/10.1002/biot.202100165

Xing, Z., Bishop, N., Leister, K., Li, Z.J., 2010. Modeling kinetics of a large-scale fed-batch CHO cell culture by Markov chain Monte Carlo method. Biotechnol. Prog. 26(1), 208-219. https://doi.org/10.1002/btpr.284

Xu, J., Tang, P., Yongky, A., Drew, B., Borys, M.C., Liu, S., Li, Z.J., 2019. Systematic development of temperature shift strategies for Chinese hamster ovary cells based on short duration cultures and kinetic modeling. mAbs 11, 191–204. https://doi.org/10.1080/19420862.2018.1525262

Zander, A., Britton, P.N., Navin, T., Horsley, E., Tobin, S., McAnulty, J.M., 2014. An outbreak of enterovirus 71 in metropolitan Sydney: enhanced surveillance and lessons learnt. Med. J. Aust. 201, 663–666. https://doi.org/10.5694/mja14.00014

## Publication List

**Xing, Z**., Nguyen, T.B., Kanai-Bai, G. Yamano-Adachi, N., Omasa, T.: Construction of a novel kinetic model for the production process of a CVA6 VLP vaccine in CHO cells. *Cytotechnology* (2023). https://doi.org/10.1007/s10616-023-00598-8

## Acknowledgement

I would like to deeply thank Professor Takeshi Omasa for being my supervisor for my PhD study at Osaka University. He is always helpful and knowledgeable. Every discussion is beneficial inspiring me to generate new ideas. His kindness also helps me to get through my PhD career. Studying in Omasa laboratory is a great joy of mine, which provides me opportunities to talents in biochemical engineering from different countries and many thanks to members in Omasa laboratory.

I would like to thank Dr. Thao Bich Nguyen as well. She kindly provided a part of experimental data in this research so that I could perform computational work. Besides, consultation with her also helped me a lot. I do appreciate it.

I would like to thank the Department of Biotechnology at Osaka University for providing this nice PhD program. I do enjoy not only excellent education and research but also wonderful campus life at Osaka University.

Finally, I would like to thank my family and my friends. I am always grateful to have so many people supporting me.