DEVELOPMENT OF A FLEXIBLE STABLE MAMMALIAN ANTIBODY EXPRESSION PIPELINE

JACQUELINE MARK KAR KEI

UNIVERSITI SAINS MALAYSIA

2023

DEVELOPMENT OF A FLEXIBLE STABLE MAMMALIAN ANTIBODY EXPRESSION PIPELINE

by

JACQUELINE MARK KAR KEI

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

February 2023

ACKNOWLEDGEMENT

I would like to express my greatest gratitude to my supervisor, Associate Professor Dr. Tye Gee Jun for his guidance during my master's journey. He has given me much inspiration as well as knowledge and training on becoming a good person and researcher. Thank you so much for the honour to work with you. I would also like to thank the administration and laboratory staff at Institute for Research in Molecular Medicine (INFORMM), Universiti Sains Malaysia (USM), for providing the necessary facilities and equipment as required for the conduct of research work.

My deepest gratitude to my family for their love and support whilst pursuing my masters. Their neverending unconditional encouragement had enabled me to overcome every obstacle in my postgraduate journey. Additionally, I would like to thank my colleagues and labmates, Dr. Sylvia Annabel Dass, Gan Shin Yi, Chin Ding Sheng, Matthew Wong Tze Jian, Liew HuaQiang and Ahmad Ismail Abdo, for their friendship and motivation throughout my research study. Special thanks to Dr. Lew Min Han for his guidance and help especially in the early phase of my research.

Furthermore, I would like to thank the Ministry of Higher Education Malaysia for Fundamental Research Grant Scheme (FRGS/1/2020/STG01/USM/02/12) and Ministry of Higher Education Malaysia for Higher Institution Centre of Excellence (HICoE: 311/CIPPM/4401005).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

℃	Celsius
3’	End of DNA at carbon-3 position in its sugar ring structure
5’	End of DNA at carbon-5 position in its sugar ring structure
bp	Base pair
g	Gram
hr	Hour
kbp	Kilo base pair
kDa	Kilo Dalton
L	Liter
M	Molar
mg	Milligram
mL	Milliliter
mm	Millimeter
mM	Millimolar
ng	Nanogram
p	*p*-value
t	Time-point
™	Trade mark
V	Volt
x	Multiply
α	Alpha
β	Beta

γ	Gamma
µg	Microgram
µL	Microliter
µm	Micrometer

LIST OF ABBREVATIONS

2xYT	2 x Yeast Extract Tryptone
ABTS	2,2’-azino-bis(3-ethylbenzothaizoline-6-sulfonic acid)
ANOVA	Analysis of variance
APS	Ammonium Persulfate
BLAST	Basic Local Alignment Search Tool
BSA	Bovine serum albumin
CO_2	Carbon dioxide
DAB	3,3’-diaminobenzidine
DdH_20	Double distilled water
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
FBS	Fetal bovine serum
Fc	Fragment crystallizable region of antibody
FcγR	Fc gamma receptor
FRT	Flp recognition target
GOI	Gene of interest
HRP	Horseradish peroxidase
Ig	Immunoglobulin
IgA1	Immunoglobulin isotype A class 1
IgD	Immunoglobulin isotype D

IgE	Immunoglobulin isotype E
IgG1	Immunoglobulin isotype G class 1
IgG2	Immunoglobulin isotype G class 2
IgG4	Immunoglobulin isotype G class 4
Igk	Immunoglobulin kappa chain
IV	Intermediary Vector
NaH_2PO_4	Sodium dihydrogen phosphate
$NaHCO_3$	Sodium bicarbonate
NaOH	Sodium hydroxide
NCBI	National Center for Biotechnology Information
PBS	Phosphate buffered saline
PBST	Phosphate buffered saline with Tween detergent
PCR	Polymerase chain reaction
RE	Restriction enzyme
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SFM	Serum-free media
TEMED	Tetramethylethylenediamine
TGE	Transient gene expression
Tris-HCI	Trisaminomethane hydochloride
UV	Ultraviolet
WHO	World Health Organisation

LIST OF APPENDICES

APPENDIX Q NUCLEOTIDE ALIGNMENT OF IGK SIGNAL PEPTIDE BASED ON NUCLEOTIDE QUERY NCBI BLAST

PEMBANGUNAN ALIRAN KERJA FLEKSIBEL UNTUK PENGHASILAN ANTIBODI MAMALIA SECARA STABIL

ABSTRAK

Antibodi mamalia adalah bahan biologi yang bertujuan untuk kegunaan biofarmaseutikal kerana kekhususan tinggi dan ketoksikan yang rendah. Protein ini sesuai untuk tujuan terapeutik dan diagnostik dan permintaannya yang semakin meningkat memerlukan barisan sel produktiviti yang tinggi. Sel mamalia boleh menghasilkan antibodi secara fana atau stabil. Walaupun sistem penghasilan fana mempunyai faedah pengeluaran pesat, ia lebih sesuai untuk penghasilan protein jangka pendek dan fasa awal ujikaji antibodi. Sementara itu, sistem penghasilan yang stabil adalah pilihan untuk pengeluaran protein berskala besar untuk jangka panjang kerana keupayaannya untuk dihasilkan semula, berskala dan boleh dipercayai tanpa variasi kelompok ke kelompok. Untuk menyelaraskan pembangunan sel yang stabil dan menghapuskan variasi klon, rekombinasi spesifik tapak dengan sel yang telah diubahsuai untuk mencapai integrasi genomik tertentu. Di sini aliran ekspresi antibodi mamalia stabil yang fleksibel telah dihasilkan menggunakan Sistem Flp-In dan sel Flp-In-293 dari Invitrogen. Vektor peralihan antara paparan phage dan sistem ekspresi fana kepada stabil dan vektor dengan bahagian Fc antibodi telah dihasilkan untuk menghasilkan antibodi format penuh dengan Fc yang berbeza. Prosedur dalam pembangunan sel mamalia penghasilan yang stabil telah dioptimumkan agar masa dan usaha dapat dikurangkan. 100 µg/ml antibiotik pilihan adalah kepekatan optimum untuk menghapuskan untransfektan dalam masa 2 minggu. Berkenaan dengan ekspresi, media yang mengandungi serum adalah media optimum untuk ekspresi

antibodi dengan tempoh ekspresi selama 10 hari dalam kultur kelompok. Walaupun menggunakan recombinasi tapak spesifik untuk penjanaan sel yang stabil mempunyai banyak faedah terutamanya dari segi kebolehskalaannya, ia tetap ada cabaran masa, kos dan tenaga kerja. Ini menyebabkan keperluan pengoptimuman dan aliran penghasilan antibody yang stabil. Secara ringkasnya, aliran penghasilan protein stabil yang dioptimumkan, fleksibel dan diperkemaskan telah dibangunkan untuk menghasilkan antibodi manusia dengan keluaran yang tinggi dan penilaian yang lebih baik dengan mudah dan tanpa variasi kelompok ke kelompok.

DEVELOPMENT OF A FLEXIBLE STABLE MAMMALIAN ANTIBODY EXPRESSION PIPELINE

ABSTRACT

Mammalian antibodies are promising tools for biopharmaceutical use because of their high specificity and low toxicity. These proteins are favorable for both therapeutic and diagnostic purposes and its ever-growing demand calls for high productivity cell lines. Mammalian cell lines can express antibodies transiently or stably. Although the transient expression system has the benefits of rapid production, it is more suitable for short-term protein production and the initial phase of antibody testing. Meanwhile, a stable expression system is the go-to for long-term mass protein production because of its ability to express proteins in a reproducible, scalable, and reliable manner with no batch-to-batch variation. To streamline the development of stable cells and remove clonal variation, site-specific recombinase with pre-engineered cell line is employed to attain specific genomic integration. Here a flexible stable mammalian antibody expression pipeline was developed using Invitrogen's Flp-In System and parental Flp-In-293 cell line. Transition vectors between phage display and transient to stable expression systems and vectors with antibody Fc region were generated for expression of full format antibodies of different Fc. Procedures in the development of stable expressing mammalian cells were optimized in order to reduce the time and effort otherwise required. 100 µg/ml of selection antibiotic is the optimal concentration to eliminate untransfectants within 2 weeks. In regards to expression, serum-containing media is the optimal media for antibody expression with 10 days of expression period in a batch culture. This is as although utilizing site-specific

recombination for stable cell generation has many benefits especially in terms of its scalability, the challenges of time, cost, and labor remains, which brings forth the importance of optimizations and pipelines. In summary, an optimized, flexible, and streamlined stable expression pipeline was developed to produce human antibodies with high yield and improved evaluation with ease and without batch-to-batch variation.

CHAPTER 1
INTRODUCTION

1.1 Background of study

Demand for monoclonal antibodies has only seen a steady increase throughout the years as the novel treatment for diseases and conditions (Liu, 2014). By possessing high specificity and low toxicity, antibodies revolutionized biotechnology and lay the foundation in molecular medicine (Salazar et al., 2017). Growing numbers of this revolutionary biological products and biosimilars have thus intensified the demand for high productivity production cell lines.

Antibodies can be expressed in mammalian cell lines transiently or stably by integrating expression constructs into the host's genome. Transient expression has a short production time frame but generates a lower yield over time (Kim et al., 2020). This is as transiently transfected cells does not possess stable genetic and protein expression characteristics as while they are able to express the foreign gene, it is not integrated into the genome (Lee et al., 2019). The foreign gene will then be lost through cell division or other factors. Although this approach is the most frequently employed technique for functional studies, it is unapt for most structural studies and large-scale productions, leading it to only be used in short-term experiments (Kim & Eberwine, 2010). Furthermore, transiently transfected cells may express inconsistently throughout the population.

Stable cell lines may take several months to be generated but can express a large amount of protein of interest as the cell lines reproduce indefinitely and continue to express the transgene. This is because stable cell lines can carry and pass on the same modified genetic characteristics to their future generations as transfected DNA

is stably integrated into the host genome (Kim & Eberwine, 2010). As the DNA is permanently expressed in the cell's genome, the expression of the exogenous gene is stable (Lo et al., 2017). Using stable cell lines, the disparity and variation in expression levels associated with repeated transient transfection can be eradicated. Thus, stable expression system is the standard practice for long-term genetic studies or industrial production of highly specific proteins.

The long development process of stable cell lines can be streamlined with a common strategy in which pre-engineered cell lines are used because it will improve genomic integration of the transgene and a popular solution is site-specific recombination using FLP recombinase (Szczesny et al., 2018). It is known as the FLP-In System where the parental cell's genome contains a FRT sequence which is recognized by Flp recombinase (Craig, 1988; Sauer, 1994). Using the pre-engineered parental cells allows for a straightforward process for the Flp vectors to integrate the gene of interest into the genome of mammalian cell lines and thus granting generation of stable cell lines at a very specific site (O'Gorman et al., 1991).

This study is focused on the development of a flexible stable mammalian antibody expressing pipeline. This enables the development of an easy phage display to transient to stable system transition and an optimized stable expression system for producing human antibodies with high yield and improved evaluation with ease. Our lab has a working transient expression system which is useful for initial screening studies but not for mass production. A transition vector (Intermediary Vector) is necessary to shift expression systems in an easy and convenient way. The Intermediary Vector (IV) was modified from pET-28a(+) plasmid and signal peptide gene sequence in IV allows for convenient harvest of stably expressed antibodies. This transition vector paired with stable expression vectors possessing human antibody fragment

crystallizable (Fc) region (pcDNA5/FRT with Fc region) allows for expression of full-format antibodies of different Fc as described in **Figure 1.1**. Using commercial Flp-In System and Flp-In-293 cell line from Invitrogen (Thermo Fisher, United States), stable expressors capable of high yield can be generated efficiently as described in **Figure 1.2**. The transfection and selection of stable cells as well as the expression and purification of the produced antibody was then optimized to generate a streamlined pipeline for generation of antibodies from stable expression thus reducing time and effort otherwise required.

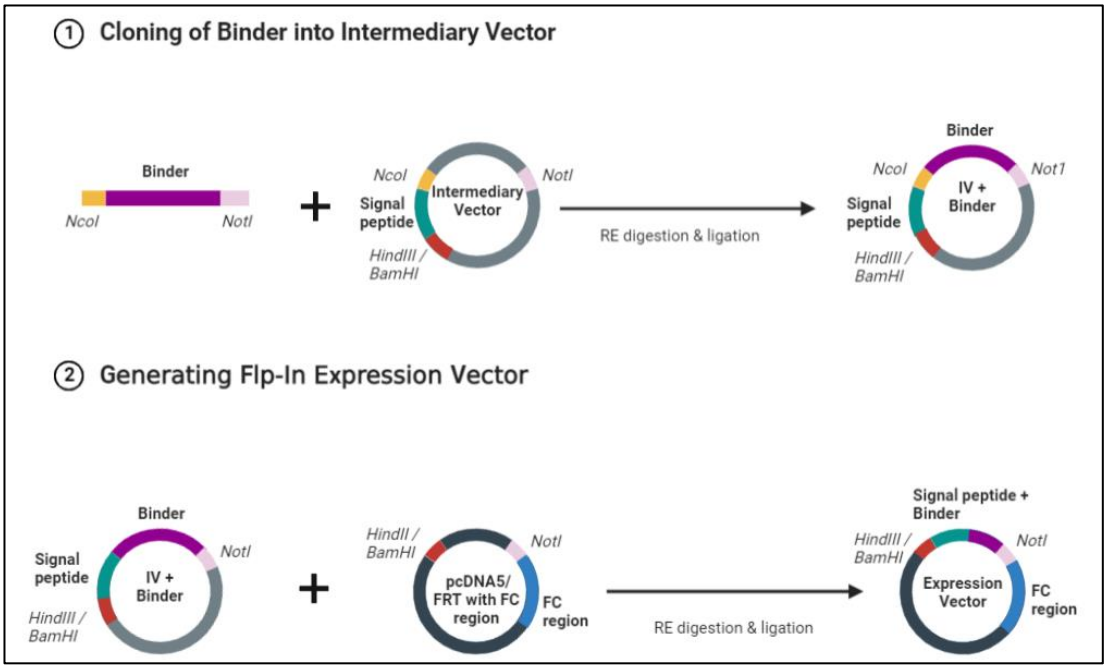

Figure 1.1 Strategy approach of flexible transition of binder from transient into stable expression system (Created with Biorender.com)

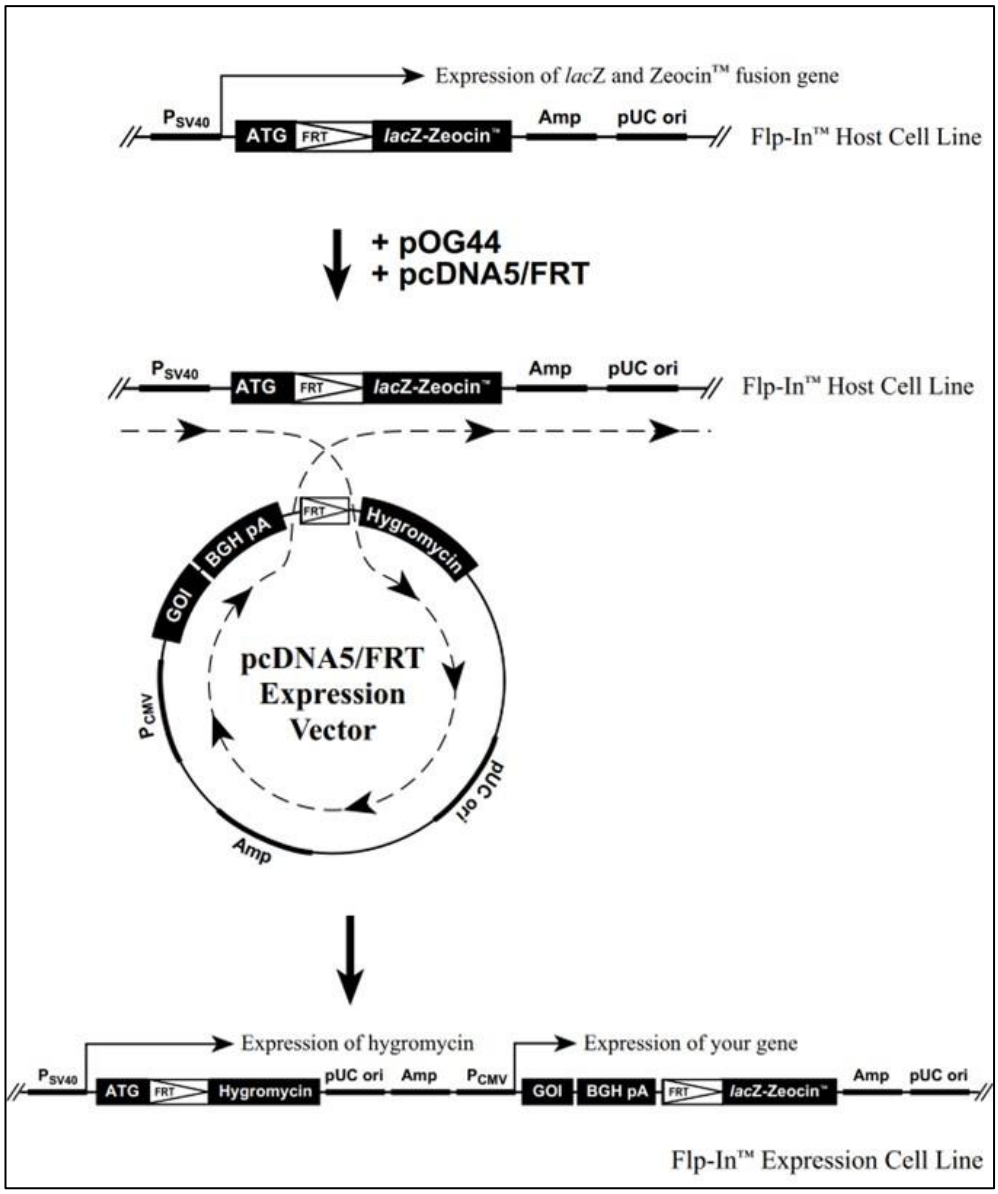

Figure 1.2 An overview of stable cell line generation using Flp-In System (Adapted from Flp-In System, Invitrogen).

1.2 Problem statement

Flexible transient expression system generated is useful for initial phase of recombinant antibody development but low productivity, high cost, and batch-to-batch variation thus resulting in this system being less desirable for mass production. Stable cell conversion improves the yield and evaluation of the generated antibodies. However, there was no flexible transition vector for transient to stable expression.

1.3 Research Objectives

i. To modify and generate an intermediary vector between transient and stable expression systems, as well as pcDNA5/FRT vectors that are ready for cloning of binders with different human Fc regions

ii. To optimize selection of stable cells producing model antibody and purification of said antibody

iii. To evaluate the stability of antibody-expression of the stable cells generated

1.4 Significance of Research

Over the past few decades, the large demand for antibodies has led to its endless advancement and commercialization. Transient expression systems are appropriate for initial antibody development but a shift to stable expression systems for mass production is necessary. In efforts to reduce time consumption and clonal variation when generating stable cells, site-specific recombination is a suitable strategy. However, the challenges of the process being time, cost, and labor consuming remains. Therefore, optimizations and pipelines must be expanded to accommodate this. This pipeline utilizing the commercial Flp-In System allows human antibodies to be produced in a flexible, streamlined manner and thus significantly reducing the necessary time, cost, and labour.

CHAPTER 2
LITERATURE REVIEW

2.1 Value of Antibody Biopharmaceuticals

The rise and demand for global biopharmaceuticals and therapies are skyrocketing. This is due to the incessant advances in our grasp of diseases at cellular and molecular level, besides the unmet medical demands of chronic diseases (O'Flaherty et al., 2020). Small molecules were once the backbone of the pharmaceutical industry, but antibodies have replaced them (Shepard et al., 2017). The popularity of antibody pharmaceuticals stem from their ability to provide a cost effective, highly specific treatment with fewer side effects while being a potential cure. This makes them the most common treatment for a variety of diseases over the last 25 years (Ecker et al., 2015; Lu et al., 2020). Antibodies are naturally produced by B cells in the human body but can be recombinantly synthesized in the lab. Mass production of pure, synthetic antibodies were first made possible by Köhler and Milstein's hybridoma technique, first proposed in 1975, thus considerably increasing basic research and the possibility for therapeutic usage (Köhler, G., Milstein, 1975; Lu et al., 2020). An acute transplant rejection immunosuppressant, the muromonab-CD3 (Orthoclone OKT3) was the first FDA approved commercialized therapeutic monoclonal antibody (mAb) in 1986 (Sgro, 1995). Currently, no less than 570 antibody pharmaceuticals have been investigated in clinical trials across the world, with 79 therapeutic mAbs approved by the US Food and Drug Administration (US FDA), which is a testament to how far antibody biologics have come throughout time (Ecker et al., 2015; Lu et al., 2020).

Antibodies have a hand in effective disease management as diagnostic tools by detecting molecules and measuring biological markers in blood. Their specificity allows for a rapid evaluation of the molecular phenotype of numerous tissues, making mAbs essential in pathology diagnosis and basic laboratory research (Weiner, 2015). Furthermore, antibodies can be used to track existing and previous diseases, as they are immune proteins and therefore can track the progression of an infectious immune response while being superior to pathogen detection due to their durability (Alter & Seder, 2020). Its adoption in diagnostic use is based on a variety of factors, including drug safety profiles, technological breakthroughs, development, and market performance, to name a few (Siddiqui, 2010).

Paul Ehrlich first postulated antibody as a "magic bullet" for specific cell targeting about a century ago (Heynick, 2009; Tan & Grimes, 2010). The creation of recombinant monoclonal antibodies (mAbs) has given treatments a new lease on life, as they have a known, specified sequence, no batch-to-batch variation, a long serum half-life, and can be manufactured in commercially available expression methods in addition to their rapid discovery process, regulatable precision, remarkable specificity, and low toxicity that decreases the risk of safety issues in human clinical trials (Ecker et al., 2015; Lua et al., 2018; Panawala, 2017; Salazar et al., 2017; Shepard et al., 2017). Antibodies can be used in therapeutics to nullify poisons or cytokines, inhibit receptors, attach to cells, and influence the immune system of the host (Kuhn et al., 2016).

MAb treatments are one of the most well-known and fast expanding clinically licenced pharmaceuticals in this timeframe. They are so widely utilised that they account for most of the top therapeutic proteins, as well as approximately 50% of all therapeutic protein sales (Carter & Lazar, 2018; O. Yang et al., 2019). Their popularity

in oncology, autoimmunity, as well as chronic inflammatory and infectious diseases, originates from their ability to target vulnerable sites and deliver themselves onto target surface proteins and into cells by antigen recognition (Abdollahpour-Alitappeh et al., 2019; Nasiri et al., 2018; Saphire et al., 2018). Therapeutic antibodies are deemed to be a viable treatment option for a variety of infectious disease targets, including Ebola, human cytomegalovirus, and others (Salazar et al., 2017). Additionally, their great specificity facilitates in the analysis of host-virus interaction and viral pathogenesis processes, thus contributing to the research of viral infection and therapies (Salazar et al., 2017). Furthermore, mAbs are widely used in cancer immunotherapy to inhibit suppressive or stimulate positive immune checkpoints (Marhelava et al., 2019).

2.2 Anti-filarial recombinant antibody

Lymphatic Filariasis (LF) was declared a public health problem by WHO in 1997 (WHA, 1997). It is a mosquito-borne parasitic disease also known as elephantiasis. LF persists in poor, marginalized communities in tropical and sub-tropical regions including Malaysia (Al-Abd et al., 2014). This disease is often overlooked because there is little potential for profit despite its dire consequences. LF can disfigure, debilitate, and even kill its victims. Therefore, there was an urgent need for better treatment and prevention, but the concerned communities have little access to them. The ‘Global Program to Eliminate Lymphatic Filariasis’ (GPELF) was launched in 1998 to eradicate this disease as a public health crisis by 2020 (Molyneux & Zagaria, 2002). In efforts to eliminate this disease, diagnostic tools were produced, and the most practical diagnostic assays are field-applicable and provide rapid results

to allow timely strategic decisions. Although most countries had successfully eradicated LF as a public health problem, diagnostic tools are still necessary to ensure the disease does not re-emerge.

One of the diagnostic tests for LF is Brugia Rapid, which is manufactured by Reszon Diagnostics International Sdn. Bhd. (Selangor, Malaysia). It is a lateral flow rapid test using BmR1 recombinant protein as the test line to detect brugian filariasis (Rahumatullah et al., 2020). Quality control (QC) is essential during the manufacture of diagnostic tools to ensure sensitivity and specificity. The significant reduction in LF patients has caused patient samples to be diminished. The efforts to address this issue led to usage of recombinant mAbs as a QC reagent. Using phage display library, recombinant mAbs sensitive and specific to BmR1 antigen were isolated (Rahumatullah et al., 2017). The production of anti-filarial IgG antibodies for the QC assessment of Brugia Rapid is currently reliant on transient mammalian or bacterial expression systems (Abdo et al., 2022; Abdul Rahman et al., 2007).

2.3 Antibody formats

2.3.1 Human Immunoglobulin Isotypes

Recombinant antibodies come in multiple molecular formats. There are five isotypes of human immunoglobulin which are IgA, IgD, IgE, IgG, and IgM. IgG is the most popular where all commercial mAbs are in IgG formats and the majority are IgG1 (Grilo & Mantalaris, 2019). However, there is an incline in non-IgG formats being tested in preclinical trials as differing isotypes regulates their effector function differently (Carter & Lazar, 2018).

IgA is able to reduce tumours and thus has the potential as anti-cancer therapeutics (van Tetering et al., 2020). IgD CH variants can release immunoactivity, pro-inflammatory, and antimicrobial mediators (Chen & Cerutti, 2011; Lua et al., 2018). IgE isotypes are able to activate different effector cell population using FcRs from monocytes and macrophages (Karagiannis et al., 2017). IgM antibody molecules stimulates complements which makes it pro-inflammatory besides having a higher avidity and steric hindrance, leading to specialized immune functions (Iwasaki & Yang, 2020; Klingler et al., 2020; Lua et al., 2018; Shen et al., 2019).

Lastly, there is the Y-shaped IgG, popular for its ability to neutralize pathogens and activate immune cells (Lua et al., 2018; van Tetering et al., 2020). The IgG subclasses have different constant regions that influence the effector function. Thus, each subclass has differing half-lives, antigen binding, immune complex formation, and complement activation. IgG1 prevents recurrent infections as well as lyse virions and infected cells, IgG2 defends against bacterial infections, IgG3 cause inflammation, and IgG4 defends against antigens in a non-infectious environment (Baba et al., 2000; Gogesch et al., 2021; Vidarsson et al., 2014).

2.3.2 Advantages of Fc domain

Recombinant mAbs are generally generated by phage-display libraries encoding single-chain variable antibody fragments (scFv) and fragment antigen binding fragments (Fabs) (Kuhn et al., 2016). ScFvs are VH and VL domains connected by a short flexible polypeptide linker while Fabs are heterodimers consisting of light chains (VL and CL) and the first two domains of the heavy chain (VH and CH1) (Doerner et al., 2014; Lu et al., 2020). As described in **Figure 2.1**, the

two antibody fragments are derivatives of the classical mAb. These fragments are the most popular class of binders in diagnostics (Chin et al., 2017). Solitary scFv and Fabs are popular tools for their small size allows easy penetration into tissues but these antibody fragments have extremely short half-lives obstructing its use in vivo (C. Yang et al., 2018). Their short half-lives may be useful for radiolabeled imaging purposes but impedes its use in therapeutics (Xenaki et al., 2017).

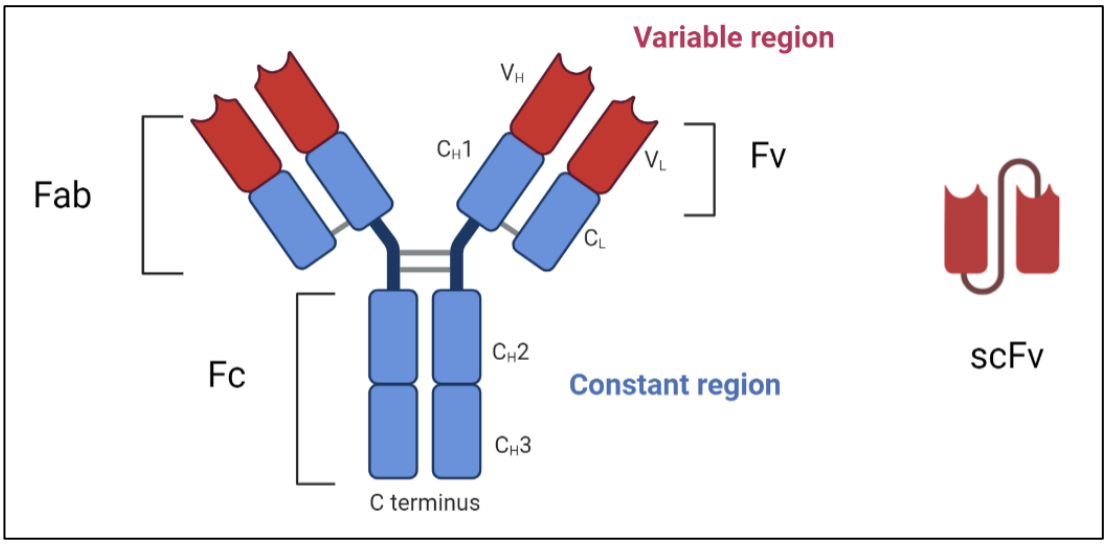

Figure 2.1 Derivatives of classical mAb (Created with Biorender.com)

Addition of Fc (crystallizable fragment) domains, specifically those of human IgG1, IgG2, and IgG4, to antibody fragments can substantially increase their serum half-life from a few minutes to 20 days (Beck et al., 2008; Xenaki et al., 2017). This is as Fc regulates mediator functions such as ADCC, CDC, Ab-dependent cellular phagocytosis, and interaction with neonatal Fc receptor (FcRn) which is responsible for the long half-life (Lim et al., 2021; Lobner et al., 2017).

A scFv or Fab gene can be reformatted to a full mAb by cloning the variable regions into expression vectors with antibody constant regions (Alfaleh et al., 2020).

Fc regions can be fused with these fragments to form scFv-Fc or Fab-Fc formats as well. The presence of a Fc in an antibody molecule elevates its biological and physicochemical properties such as protein stability, binding affinity, and aggregation resistance, thus increasing its advantages for clinical and preclinical use (Czajkowsky et al., 2012; Jørgensen et al., 2014; C. Yang et al., 2018). Additionally, the presence of Fc allows for easier purification and increased solubility (Lim et al., 2021). Fc domain provides numerous labelling sites for signalling molecules which is an additional advantage for diagnostic assays (Abdo et al., 2022). The most popular format for therapeutics are IgGs whether as a full format mAb or IgG-scFv (Almagro et al., 2019; Krishnamurthy & Jimeno, 2018).

Although Fcs are more commonly known for their roles in therapeutics, they prove to be an asset in diagnostic assays as well. During the development of a diagnostic tool for lymphatic filariasis, Brugia Rapid, a scFv protein which had shown antigen reactivity with ELISA, did not exhibit reactivity when tested with Brugia Rapid (Rahumatullah et al., 2020). However, fusion of a IgG4 Fc to the scFv improved its diagnostic sensitivity and a positive reaction was obtained from the kit. Furthermore, there was a recent diagnostic assay for SARS-CoV-2 which was developed using scFv-Fc antibodies (Kim et al., 2021).

2.4 Antibody Expression Systems

2.4.1 Cell Lines

Host cell selection is the most significant and effective approach to improve volumetric cell output while reducing manufacturing costs (Jiang et al., 2019). However, although obtaining high quantity of yield at a reduced cost is desirable, it is

also essential to examine the quality of the synthesized proteins. This is as the type and degree of the product structure, primarily glycosylation, can be highly impacted by the expression system chosen (O'Flaherty et al., 2020). Continued advancement, defined structures for protein expression besides strain engineering instruments to incorporate recombinant gene into host genome and develop post-translational modifications are all important characteristics of a model antibody expression host (Matthews et al., 2017). Furthermore, intrinsic properties that include established distinct productivities, absence of co-secreted proteases, amount of host cell proteins in the proteome, and increased protection from the absence of mammalian viral communicables are also factors to consider for choice of host (Jiang et al., 2019). Some other factors are process flexibility, such as the use of more economical feeds, as well as hardiness to shear stress, pH, and temperature. (Jiang et al., 2019).

Recombinant proteins can be expressed on a variety of platforms, and several expression systems were established. Bacteria are the most well-known workhorse, generally *Escherichia coli* for their simple and low cost expression (Assur et al., 2012; Dalton & Barton, 2014). However, the toxicity of the foreign protein to the host, differences in protein folding machinery, and the lack of post-translational modifications including disulphide bonds and unique glycosylation patterns in bacteria render prokaryotes unsuitable for functional therapeutic antibody expression (Assur et al., 2012; Dalton & Barton, 2014). Yeast expression systems offer numerous advantages where they are able to produce extremely large yields of secreted recombinant proteins while being relatively economical and time efficient. However, yeast systems produce a number of proteases which are able to degrade proteins, and although protease-deficient cell lines are available, yeast systems, notably *Pichia* and *Saccharomyces*, generate high-mannose type N-glycosylations and are unable to

produce complicated intricacies of human proteins making them unfeasible for routine protein production (Dalton & Barton, 2014). Insect or baculovirus vector systems are the primary method for producing difficult cytosolic proteins that cannot be effectively generated in prokaryotic hosts, but in the expression of secreted mammalian proteins, this system is impractical due to its high cost, inability to produce complex sialylated glycans as in humans, tediousness, short storage life, and low yield while necessitating numerous host factors such as glycosylation machinery, folding chaperones, and disulphide isomerases for complex proteins (Dalton & Barton, 2014). All in all, as described in **Figure 2.2**, the aforementioned hosts develop glycan structures that are fundamentally different from humans, posing the potential of unfavorable immunogenicity in therapies (O'Flaherty et al., 2020). Furthermore, the progression of these systems is fragmented, negatively impacting evaluations of their prospective benefits compared to mammalian cell lines (Jiang et al., 2019).

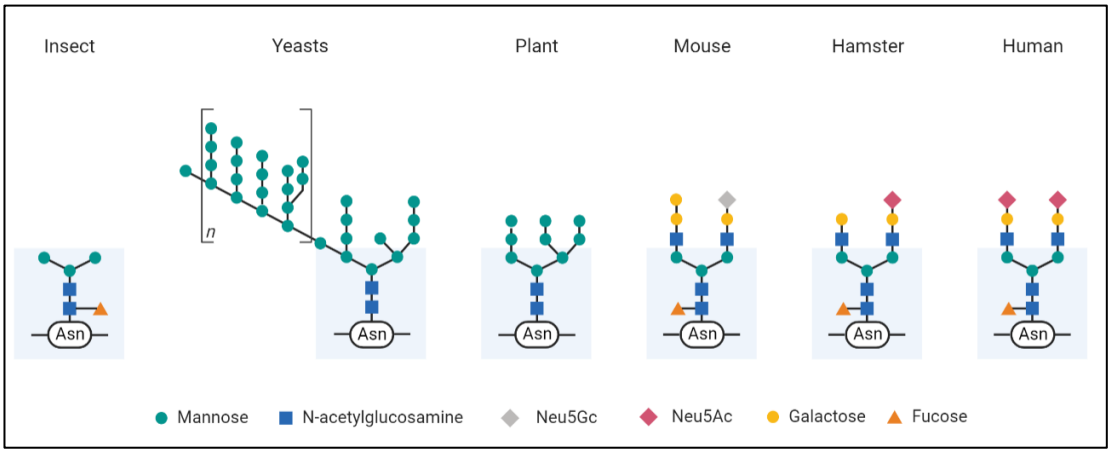

Figure 2.2 N-linked Glycoprotein Structural Differences between expression Cell Types (Created with Biorender.com)

Consequently, mammalian cell lines are the rational choice models for the synthesis of therapeutic recombinant proteins because they exhibit the closest match to the native environment for ectopically produced mammalian genes and hence have an intrinsic aptitude to produce functional mammalian antibodies (Assur et al., 2012). In addition to being able to generate high quantities of therapeutic antibodies and flexibility in large-scale production systems, these hosts undertake appropriate protein folding, post translational modifications, and glycosylation similarly to those in humans, reducing the probability of rejection (Dangi et al., 2018; Wells & Robinson, 2017). Glycosylation in antibodies is an essential element to consider because it has a great deal of variation and complexity (Kaur, 2021). Glycosylation defines the structure and effector functions of antibodies such as ADCC and CDC, as well as their pharmacokinetic/pharmacodynamic properties (Kaur, 2021). Mammalian cell lines are the primary source of recombinant proteins on the market, accounting for more than two-thirds of all recombinant proteins on the market (Lalonde & Durocher, 2017).

In 1956-1957, the acquisition of the first CHO cell line from Chinese Hamster has prompted the emergence of various lineages with considerable genetic changes (Lewis et al., 2013; Wurm & Hacker, 2011). CHO cell lines are by far the most commonly used host cells in industrial production and were first chosen because they can grow robustly and quickly to high densities in bioreactors, produce large amounts of recombinant protein, are considered safe as they have elevated resistance to human pathogenic viruses, and have the inherent capabilities to perform human compatible post translational modifications (Dahodwala & Lee, 2019; Wurm & Hacker, 2011). Additionally, they have a genome with high plasticity, which, while it may offer flexibility to allow easy transfer of foreign DNA into its genome, their genetic instability and heterogeneity cause regular genotypic variation and phenotypic

instability, as well as cell line instability and issues regarding process reproducibility and consistency (Baik & Lee, 2017). This problem causes genomic instability throughout time, resulting in a decline in recombinant protein output over lengthy periods of culture, resulting in titer and product quality instability (Dahodwala & Lee, 2019). Usage of mouse and hamster derived cell lines for recombinant mAb expression, including the aforementioned CHO cell lines has its drawbacks, because recombinant proteins expressed from non-human cell lines have glycan structures differing from those found in human cells, e.g. N-glycolylneuraminic acid (Neu5Gc) and galactose-alpha 1,3-galactose group, which may evoke immunogenic responses in human cells where antibodies are formed against these two glycan structures (Dou et al., 2021). Moreover, at certain concentrations, a-1,3-galactose linkage can cause anaphylaxis (Bosques et al., 2010). Murine cells generate fucosylated IgG-G0 glycoforms that, when clustered, can activate complement via MBL binding (Arnold et al., 2007). These cell lines have been shown to produce glycan structures that are not present in humans (Melville & Estes, 2013). There are also variances in the synthesis of sialic acid. As described in **Figure 2.2**, non-human mammalian cells are able to synthesise complex human-type glycans but are capped differently with Neu5Gc and N-acetylneuraminic acid (Neu5Ac) while glycans produced by human cells are only capped with Neu5Ac. Neu5Gc production is prevented by a naturally occurring mutation in the CMAH gene, however this glycan moiety has been identified in various biologics produced from mouse cell lines, which impacts antibody production and contributes to serum sickness in humans, among other pathophysiological ramifications (Noguchi et al., 1995; Yu et al., 2016). Furthermore, xenoantigenic gal-a-gal linkages in murine-expressed proteins might trigger hypersensitivity reactions, increasing the risk of murine expressed antibody-based

therapies (Chung et al., 2008). As a result, murine expressed biologics, particularly glycan structures must be assessed for quality assurance, however, these concerns may lead to good proteins being excluded.

Immunogenicity can be prevented by using human cell lines to synthesize recombinant proteins. Human cell lines, on the other hand, do not produce immunogenic glycan but are vulnerable to human virus infection, necessitating a sequence of viral inactivation steps when the cell lines are to be employed to construct biotherapeutics and vaccines (O'Flaherty et al., 2020). The Human Embryonic Kidney 293 (HEK 293) cell line is a human cell line that has been transfected with viral DNA and is capable of producing human glycan profiles. Whilst its efficacy is limited by low levels of transgene expression, it is the most widely used human cell line for variations of recombinant therapeutic proteins expression (Dou et al., 2021). Both CHO and HEK cell lines are exceptionally stable, receptive to genetic modification, and can expand and proliferate rapidly, suggesting a strong capability to synthesize proteins (Roobol et al., 2020). Human-derived cell lines, on the other hand, have additional advantages. After CHO cell lines, HEK cell lines are the most utilised cell lines.

2.4.2 Transient and Stable Expression

The first phase of the overall process of a recombinant protein expression in a mammalian expression system is transfection of a recombinant gene with structural components allowing transcription being integrated into the preferred cell line via a plasmid vector , which is then followed by selection for successfully transfected cells for clone generation (O'Flaherty et al., 2020). Transient or stable antibody expression

can be performed in mammalian cells by incorporating expression constructs into the genome. These two expression systems have their own advantages and disadvantages as described in **Table 2.1**.

Table 2.1 Comparison between transient and stable expression

	Transient expression	**Stable expression**
Expression period	Short-term, limited time	Long-term, sustained
Production time	Short	Long (up to several months)
Plasmid integration into genome	Non-integrated	Integrated
Usage	Short-term experiments (functional studies, and to evaluate gene activity or regulation)	Mostly used for long-term experiments or continuous production (structural studies and large-scale productions)
Yield	Lower over time	Higher over time
Expression	Inconsistent	Consistent, in a stable manner, and with regulatory familarity
Transfection	Chemical methods or electroporation	Chemical methods or viral vectors followed by antibiotic or drug selection
Genetic alteration	Temporary and the transfected gene is not passed onto the cells' future generations	Permanent and the transfected gene is passed onto the cell's future generations
Copy number of transfected plasmid	High	Single or low

Transient gene expression (TGE) has a limited time frame for antibody expression which results in a reduced yield but has the advantage of quick production. Transiently transfected cells lack stable genetic and protein expression properties, as the transgene encoded plasmid is not merged into the genome. TGE is based on the notion that expression constructs are transfected into cell nuclei as non-integrated plasmids and therefore the foreign gene is gradually not replicated and is lost due to cell division, degradation, and perhaps other factors (Lee et al., 2019). This system is

commonly and efficaciously used for functional studies such as biochemical and preclinical analysis, as well as testing for novel expression strategies, particularly in evaluation of the vector construct and molecular candidates. However, TGE is unsuitable for most structural studies as well as large-scaled production and therefore primarily used for short-term experiments (Park et al., 2021). Even so, in past few years, TGE systems have progress to where there is an a substantial increase of recombinant protein yield, although it is still fairly low in comparison to that of stable gene expression systems (Lee et al., 2019). For TGE, HEK 293 are the preferred hosts for their flexible conversion to suspension cultivation under serum-free conditions, high transfection performance, high productivity, ability to grow to high cell densities, and inclusion of an episomal replication mechanism, all in all, leading to a higher yield (Gutiérrez-Granados et al., 2018; Park et al., 2021).

Transiently transfected cells can also express in an inconsistent manner throughout the population as minor differences in culture environment, cell quality, medium components, and transfection produce significant differences in glycan content, even when generated under similar parameters (Saphire et al., 2018). Another disadvantage is that in transient transfection assays, gene regulatory regions and promoters operate inconsistently, which could be due to the fact that a large number of long-range enhancer and silencer components fail to function except if integrated into genomic DNA or deficient chromatin formation, making TGE unapt for studies regarding regulatory region impact on gene expression (Jensen et al., 2020). All in all, stable expression can eliminate discrepancy and volatility in expression levels applicable to transient expression, and thus making them crucial and the typical procedure for long-term genetic research as well as commercial manufacturing of highly specialised proteins.

Stable clones require tedious and long production, but can express large amounts of protein of interest with consistent, stable quality, and regulatory familiarity because the cell line genome is stably integrated with the expression construct and therefore can reproduce indefinitely and continue to express the transgene in its future generations (Lalonde & Durocher, 2017; Lo et al., 2017). However, stable cell lines are made up of a single gene copy number of transgene, reducing their yield in comparison to transient cell lines which allows a higher amount of DNA to be transfected (Schlicht et al., 2021). The oversupply of antibody medicines in today's world necessitates the development of far better alternatives, and the patent expiration of numerous "blockbuster" biologics prompts the development of several biosimilar antibodies, inciting the increased appeal antibodies of high quality and quantity produced in a timesaving manner (Carter & Lazar, 2018). Stable cell lines theoretically check all of these boxes, since they can be used for long-term manufacturing, are cost-efficient and produce a high yield of stable antibodies to meet the demand. Hence, stable expression system is commonly used for production of a variety of antibody formats as described in **Table 2.2**.

Table 2.2 Examples of antibody molecules produced using stable expression system

Platform	Antibody name	Antibody type	Reference
NS0	anti-TNFα (Infliximab)	mAb	(Bueno-Soler et al., 2022)
CHO	anti-FUT8	Intrabody	(Joubert et al., 2022)
CHO-DG44	anti-TNFα (adalimumab)	mAb	(Doan et al., 2021)
CHO-K1	anti-TNFα (adalimumab)	mAb	(Gupta et al., 2021)
CHO	anti-TNFα (Haidalimumab)	mAb	(Liao et al., 2021)
CHO-K1	Anti-PCSK9	mAb	(Mahboudi et al., 2021)
MC38	anti-PD-L1	Blocking antibody	(Buñuales et al., 2021)
HepG2	ER-ADκ	scFab	(Wu et al., 2020)
THP-1	FcγRIIa (CD32), FcγRI (CD64) and FcαR (CD89)	Fc fragment	(W. S. Lee et al., 2021)
CHO-K1	CD137, PD-L1	Bispecific antibody	(Geuijen et al., 2021)

CHO-K1	PMab-237	mAb	(Kato et al., 2019)

Furthermore, in terms of efficiency and efficacy, stably expressed proteins are not inferior to their transient counterparts. Stably expressed antibodies bypass batch-to-batch variation found in transiently expressed antibodies with no structural variation between these two, as judged by biochemical, antigenicity, electron microscopy, and differential scanning calorimetry analyses (Chung et al., 2014; Li et al., 2007; Mason et al., 2012; Rajendra et al., 2017; Wang et al., 2019). It was noted that stably expressed proteins only have a marginally higher level of expression and proper assembly (Roobol et al., 2020; Wang et al., 2019). However with time, the disparity between expression levels of these two systems becomes wider as transient system is limited to milligrams of protein while stable systems are able to generate grammes (Rajendra et al., 2017). Regardless, the expression level is mostly dependent on promoter used rather than transient or stable system utilised (Bayat et al., 2018; Dou et al., 2021; Xia et al., 2006).

Stable cell lines are created by integrating an expression construct that expresses recombinant DNA into the host's genome (Lo et al., 2017). Protein expression levels are dependent on the number of integrants per cell, the transgene locations, and respective cell differences to name a few (Assur et al., 2012). Therefore, to obtain a high-proliferating and productive cell line, screening for high-producing cells is vital.

In the formation of recombinant cell line, it is customary for the gene of interest (GOI) to be randomly incorporated into the host's genomic DNA which brings about unpredictable site and transgene copy number in the genome (Lee et al., 2019; Wirth & Hauser, 2004). Although, the expression of the randomly integrated transgene is

dependent on its genomic surroundings, making it difficult to monitor (Chen et al., 2013). This phenomenon is known as the position effect and is caused by clonal variation. Random integration of transgenes limits predictive value, process efficiency, and cost-efficient therapeutic glycoprotein production while ceasing the ability to evaluate wild-type and mutant reporters directly (Jensen et al., 2020; Shin et al., 2020). Clonal variation brings about inconsistent genomic backgrounds, from variation of chromosome structure, specific coding/noncoding genomic regions, to epigenetic regulation that can effect heritable functional variations between cells such as specific growth rate, surface glycan content, and biosynthetic capacity, thus compromising the expression and stability of transgenes, particularly in long-term cultivation (Lee et al., 2019). Random integration causes the copy number of an integrated gene to be ambiguous and varies from cell to cell, resulting in a heterogeneous population of cells. Moreover, when there are multiple integrants in a cell, it will result in overexpression of recombinant antibodies that affect protein production as well as assert a major strain on host cells (Lo et al., 2017).

2.4.3 Site specific Recombinase

To achieve the objective of eradicating clonal variance whilst producing stable, predictable, quantitative, and efficient high-expressing cell lines, a versatile, self-sufficient strategy known as site-specific recombinase can be employed. These enzymes can modify genomes without aid from the host DNA repair machinery besides not creating double-strand breaks in DNA (Shah et al., 2015). This technology facilitates identical chromosomal integration sites in all targeted cells, thus limiting the copy number of transgenes to one, and hence causing the expression level to be highly

predictable, and to achieve this, expression cassettes are exchanged in a previously tagged site (Wirth & Hauser, 2004). Site-specific recombinases can modify genomic DNA with great efficiency and fine spatiotemporal control by introducing GOI into the chromatin in a transcriptionally active region or genomic hot spot (Tian et al., 2020; Tian & Zhou, 2021).

Furthermore, the respective recombinases possess high specificity and this quality makes them extremely useful tools for genetic engineering as they enable directed operations in vivo on large DNA molecules such as eukaryotic chromosomes (Schweizer, 2003). These systems do not have a bias for specific cell types, but require "landing pads" on genetically modified master cell lines which are chromosomal targets that are precisely positioned recombinase recognition sites (Phan et al., 2017). It is a type of genome editing that generates high producing, controllable, and predictable stable cells in a time-efficient manner by reducing the time required for cell line screening (Melville & Estes, 2013). Site-specific recombinases are essential genome engineering techniques due to their extraordinary capacity to mediate deletions, insertions, inversions, and translocations of genomic DNA in living organisms, although it may be reliant on their targets' locations and relative orientations (Shah et al., 2015; Tian & Zhou, 2021). Site-specific recombination is favoured over homologous recombination because, while homology can be used for integration of transgenes into the genome of certain cell types, such as embryonic stem cells, it is less effective on cultured cell lines used for genetic engineering in routine usage (Schlake & Bode, 1994). Site-specific targeted gene transfer in mammalian cell lines involves the genetic toolbox of phage or yeast derived auxiliary enzyme machinery such as Cre/Loxp from coliphage P1, Flp/FRT from 2μm plasmid of *Sacchromyces cerevisiae*, and φC31/*att* from *Streptomyces* phage φC31, which relies

on the random integration of a reporter gene to locate genomic hot spots (Melville & Estes, 2013; Schlake & Bode, 1994).

2.4.4 Flp-In System

The Flp-FRT recombination system is encoded by the 2μ plasmid from *Saccharomyces cerevisiae.* Flippase (Flp) recombinase catalyzes efficient recombination by catalyzing DNA rearrangements through engineered genomic Flp recognition target (FRT) sites to achieve precise gene integration (Li et al., 2020; Srivastava, 2021). A FRT site is made up of three 13 bp long symmetric elements where two are present in inverse orientation and flank an 8 bp spacer of which two of these sites are necessary in an event of recombination four monomers of the Flp enzyme (Jakobsen & Bjergbæk, 2018). The interaction between the two sites is initiated by strand cleavage by the active-site tyrosine in Flp (Tyr343) to result in the formation of nicks at specific nucleotides in the Flp site, resulting in a staggered break and exchange where recombination occurs, with the Flp recombinase attached to DNA via a free 5'hydroxyl group and a 3'phosphotyrosine linkage and its residue used for protein attachment to either DNA strand (Gronostajski & Sadowski, 1985; Jakobsen & Bjergbæk, 2018).

This technology is commonly used to regulate gene expression in both germline and somatic tissue by modifying DNA elements with FRT sites previously integrated into the genome or inserted into extrachromosomal arrays, or by inserting DNA sequences at particular locations in a genome (Nonet, 2020). The Flp-FRT system also has the advantage of being more versatile than the other specific recombinase systems because it is the least restricted in terms of host range and can be