THE DEVELOPMENT OF AN INTEGRATED PLATFORM FOR TRANSIENT PRODUCTION OF RECOMBINANT BIOPHARMACEUTICALS

A thesis submitted for the degree of Doctor of Philosophy to

The University Of Sheffield.

The University of Sheffield
Department of Chemical and Biological Engineering

Camille Roger Jean Segarra
October 2012

Declaration

I, Camille Roger Jean Segarra declare that, unless stated otherwise in the text, I am the sole author of this thesis and that the research presented within is the result of my own efforts and achievements. I confirm that this work has not been submitted for any other degrees.

Abstract

Biopharma industry currently works on shortening the time required to manufacture drug candidates to support early preclinical testing, and reduces the failure rate by early screening of the molecules with the best potential to reach the market later on. This thesis focused on the development and integration of three aspects of manufacturing to contribute in achieving these objectives. The first one looked at the way to approach experimentation. The use of DOE was generalised to the whole development and the creation of a library of statistical tools to develop an algorithm for the analysis of the commonly used Central composite and Box-Behnken designs. This algorithm proved to be as efficient as commercially available statistical softwares but presents the advantage of automating the analysis of the DOE designs. The second aspect consisted in developing viable manufacturing steps such as an upstream PEI-mediated transfection process in CHO cells, capable of generating 100mg L^{-1} of product in less than 15 days, or a cation exchange purification platform using a Quality by Design approach. The third aspect focused on the integration of different processing steps to yield a whole integrated platform for the production, then purification of one gram of a reporter antibody. This platform proved to be a cost effective alternative to the development of a stable producing cell clone for the production of recombinant product. In addition, this research reveals that the adoption of a statistically driven approach to process development is as important as the implementation of innovative technologies to address the challenges ahead.

To my parents and my grandfather Jean.

Acknowledgements

Sincerest thanks to my supervisor Professor David James for giving me the opportunity to learn in his lab and for guiding and supporting me through the challenging moments of my PhD. I would also like to thank my industrial supervisor Dr. Peter Levison for his invaluable and pragmatic advice, and Dr. John Woodgate for his unconditional support during the last month of my PhD within the R&D Applications team. The completion of this thesis would not have been possible without the help of my co-workers to whom I would like to express my gratitude. Finally, I would like to thank the Biotechnology and Biological Sciences Research Council and Pall Corporation for the financial sponsorship of this project.

Contents

List of Tables

List of Figures

Acronyms

AEX anion exchange chromatography

ANOVA analysis of variance

BBD Box-Behnken design

BHK baby hamster kidney

CaPi calcium phosphate

CCD central composite design

CCF face-centred composite design

CEX cation exchange chromatography

CHO Chinese Hamster Ovary

COG cost of goods

CQA Critical Quality Attribute

CV column volumes

DBC dynamic binding capacity

DHFR dihydrofolate reductase

DOE Design Of Experiments

FDA Food and Drug Administration

GS glutamine synthetase

HCP Host Cell Protein

HEK human embryo kidney

HSPG heparan sulfate proteoglycans

IEX ion exchange chromatography

IgG immunogloulin G

kDa kilodalton

mAb monoclonal antibody

OFAT one factor at a time

PEI polyethylenimine

QbD Quality by Design

RD Research and Development

RSM Response Surface Methods

SEAP secreted alkalyne phosphatase

SEC size exclusion chromatography

SP HP SP Sepharose HP

SS sum of squares

SU single-use

TGE Transient Gene Expression

UF/DF ultrafiltration/diafiltration

Chapter 1

Introduction

This Chapter first introduces the current challenges the pharmaceutical industry is facing, and the future scenario for the manufacturing of biopharmaceuticals through the use of mammalian expression systems, bio-manufacturing processes, and experimentation methodologies.

One of the main goals of cell culture technology is to generate medical drugs as treatments for cancer, infectious and autoimmune diseases. The development of a new drug is a long and challenging process that lasts on average 10-12 years. It is typically divided into three major steps: discovery, preclinical development and clinical trial. From an economical point of view, preclinical phase is a screening step of high importance. It is first aimed at determining the most accurate product efficacy, stability and safety profiles. Based on these profiles, it evaluates if a molecule has a reasonably good chance to hit the market in the near future, *i.e.* if a molecule is worth the costly clinical trials that follow. Therefore, preclinical testing should be a high throughput and cost-effective step. This is not the case. The transition from preclinical development to the clinic is so challenging that it is often referred to "the valley of death"(1–3) (Figure 1.1). If biopharma companies do not hesitate to invest in drugs that make it to the clinical phase, potential investors hesitate to support risky and costly preclinical development (4). One of the reasons is that the industry still struggles to intensify their manufacturing capacity for early testing. Indeed, biopharma industry mostly relies on costly and time consuming development

of stable clones to generate the quantities required of drug candidates to support preclinical testing. Moreover, the current process infrastructure for cell culture and subsequent purification processes lacks flexibility, considerably increasing the costs and time-lines of manufacture. This thesis focuses on adapting or developing optimization methods, statistical tools and processing technologies to allow the rapid production of milligram to gram quantity of biopharmaceuticals to support preclinical trials. The production platform will use an upstream transient gene expression step, coupled with an adapted downstream train for efficient product purification. The whole platform will consist of disposable/reusable integrated modules to provide maximum flexibility, fast product production and turn over.

This introduction chapter provides a brief overview of the research context. First, the current use of mammalian cells for the production of biopharmaceuticals is described. Second, the current state and future of manufacturing platforms is presented. Finally, a brief overview on experimental methodology is described.

1.1 Mammalian expression systems for the production of recombinant biopharmaceuticals

1.1.1 On the biopharmaceuticals

Biopharmaceuticals can be defined as clinical reagents, vaccines and drugs generated by modern biotechnology. In contrast to traditional chemically synthesised drugs, biopharmaceuticals show attractive properties for the treatment of cancer, infectious and autoimmune diseases. Indeed, because they mimic the action of endogenous proteins native to the human body, they theoretically present lower toxicity, higher specificity and efficacy. The worldwide market for biopharmaceuticals has been, and is expected to grow at a 15-18% rate annually, well above overall economic growth rates including the pharmaceuticals market in general (6). The recombinant proteins market alone is estimated at $100 billion.

Therapeutic monoclonal antibodies (mAb) boast one of the most active and promising pipeline in the recombinant biopharmaceuticals industry. This is mainly

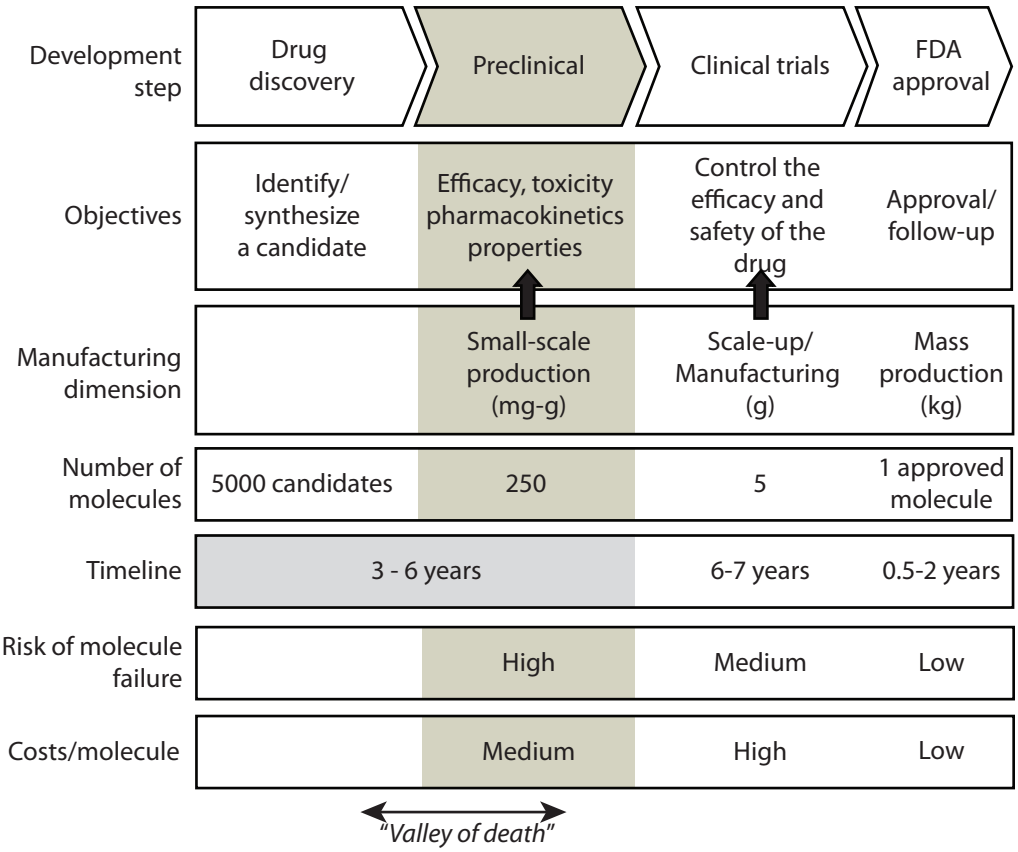


Figure 1.1: From the research to the market: the development of a biopharmaceutical. The development of a new drug is a multi-stage process. The translation of a molecule from early discovery and preclinical to clinical is particularly challenging. A way to overcome this bottleneck would be to shorten and reduce the costs associated with this translation process, encouraging potential investors to finance the testing of more molecules. Adapted from (4, 5).

due to the robust and flexible nature of the immunoglobulin molecule, as well as advances in molecular biology. Indeed, since the approval of the first therapeutic mAb Muromonab-CD3 in 1986, mAb formats have evolved from purely murine sequence, to "chimeric" and later, "humanised" forms containing human sequences. New engineered mAbs present a generally reduced immunogenicity, increased stability and efficacy(7, 8). Recombinant DNA technology allows for the modification of existing DNA sequence and production of an almost unlimited library of new molecules (9, 10).

1.1.2 Mammalian cells as a choice of expression system

Over the past few years, cells from various origins have been used as hosts for the production of biopharmaceuticals including bacteria, yeast, insect cells, plant and mammalian cells (11). But in pharmaceutical industry, product quality represents an overriding concern, and mammalian cells became the preferential host for the production of biopharmaceuticals because of their capacity to perform protein folding, assembly and post-translational modifications similar to that found in humans (12).

While several recombinant biopharmaceuticals have been expressed with baby hamster kidney (BHK) (13), human embryo kidney (HEK) (14), NSO mouse myeloma cells (15) or more recently the PerC6 or HKB11 cell-lines (16, 17), Chinese Hamster Ovary (CHO) cell-line remains today the standard platform for several reasons. Firstly, CHO cells are considered safe for the production of therapeutic proteins as most human viruses are unable to replicate in these cells (18). Secondly, the history of approval of numerous recombinant proteins in CHO cells is an advantage in the current stringent regulatory environment. Finally, in a context where biopharmaceutical companies tend to minimise the development risks to a maximum, the expertise and knowledge amassed over the past two decades ensure that CHO cells will remain the industry's workhorse for therapeutic recombinant production for the foreseeable future.

For industrial purpose, suspension cell-culture processes are preferred to adher-

ent cell culture because of the well-understood principles of scaling parameters, space savings and the ease of process control in homogeneous systems (19). The media in which the host cells will be cultivated is of primary concern. Biopharmaceuticals must be expressed in animal-derived component free and if possible, chemically defined media. The use of chemically defined media reduces the risks of virus contamination, and results in greater consistency between batches, as well as easier to implement downstream applications. Chemically-defined media is today a standard for the production of clinical proteins (20).

At the moment, the quantities of product required to support early stage preclinical studies are generated using stably expressing cell-lines. Practically, the development of stably expressing cell clones consists of the transfection of the cells with a gene of interest cloned into a plasmid vector, then the integration of the gene of interest into the host cell genome, and finally the isolation and characterisation of the clones. The major drawback of this method is that specific productivity of the vast majority of the recovered stable cell clones is low. Hundreds to thousands of clones need to be screened with respect to growth and productivity characteristics. The process can be accelerated by selection, using a marker. In this case, the gene of interest is flanked with a gene coding for a vital enzyme. A well-known example is the dihydrofolate reductase (DHFR) system. DHFR catalyses the formation of a precursor involved in the synthesis of purines, pyrimidines and glycine. DHFR can be used in recessive selection for which a modified CHO cell-line lacking DHFR activity is used. DHFR selection can also be applied to a normal CHO cell-line cultivated in a depleted medium. The stably transfected cells will survive in culture while the one lacking the enzyme activity will die. The glutamine synthetase (GS) system is also commonly used. Most mammalian cells have naturally low levels of endogenous GS and require exogenous glutamine to grow. The GS catalyses the synthesis of glutamine from glutamate and ammonia. The co-transfection of the GS gene therefore allows a glutamine independent growth. The two systems can also speed-up the process of clone selection by imposing stringent conditions of survival. Indeed, cultivating the cells with increasing concentration of the DHFR or GS en-

zyme inhibitors, methotrexate or methionine sulphoxamine respectively, ensure to select cell clones that can titrate the inhibitor, *i.e.* have several copies of the DHFR or GS coding gene, and therefore several copies of the gene of interest (21–23). Even if these methods, coupled with process automation, helped reduce the development time-lines associated with clone selection (24), individual clones still need to be cultivated for several months to exclude stability problems. Indeed, inherent instability of CHO cell-lines at genetic and epigenetic levels, can result in progressive loss of specific productivity over the generations (25). If at the moment, having a stable clone available is a pre requisite for phase II and III of clinical trials, there is a need for alternative technologies to speed-up the manufacture of milligrams to gram quantities to support early preclinical phase.

1.1.3 Transient gene expression for the production of biopharmaceuticals

The introduction of foreign genetic material by transfection into mammalian cells on a transient basis, *i.e.* without stable integration of the plasmid into the host cell genome, has been used as routine procedure at small scale in research for decades. However, its use as a method for production of recombinant proteins beyond the laboratory scale is relatively recent (26). Transient gene expression (TGE) represents an interesting alternative to stable expression as this approach present several advantages: a short time frame for the generation of product, applicability to a wide range of host cell-lines, simple plasmid vector constructs and product consistency as the production time-frame is relatively short. Early attempts to use this technique highlighted several challenges: low yields, a relatively high quantity of genetic material required at transfection, and the short life-span of the transfected culture (Figure 1.2).

Various transfection techniques exist. The advantages and disadvantages of the main ones used are detailed in Table 1.1 and in the text below. Few physical means have been successfully used to transfect mammalian cells. The majority of the methods (particle mediated gene transfer, needle injection, jet injection) present severe

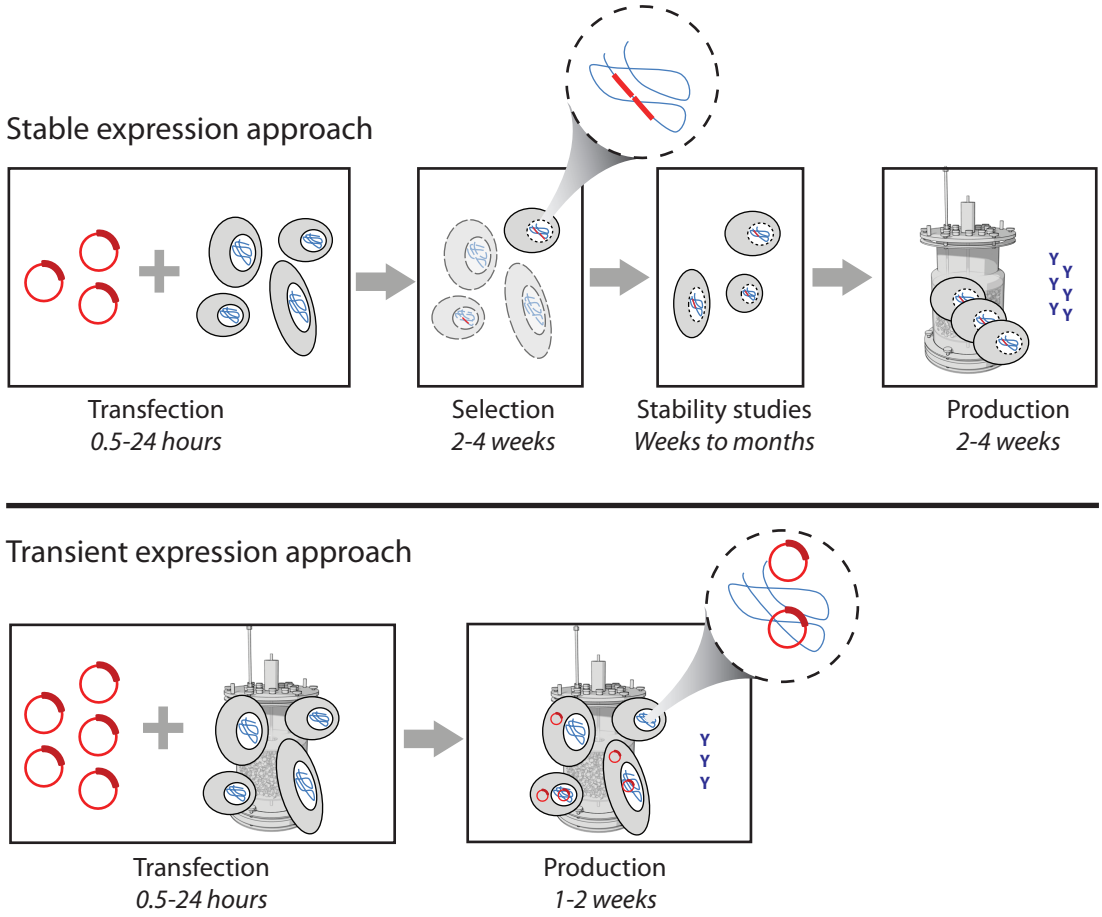


Figure 1.2: Comparison of stable and transient approach to recombinant protein expression. Transient gene expression represents an interesting alternative to stable expression as this approach allows for the rapid generation of milligrams to gram quantities of product, in days rather than months. In transient expression the foreign DNA is maintained as an extra-chromosomal unit within the cell nucleus.
Adapted from (27).

drawbacks such as a limited transfection efficiency (number of cells transfected), a relatively short duration of transgene expression and/or the complexity of the method itself (28). The most efficient physical method is electroporation-mediated gene delivery. Widely used at the millilitre scale, electroporation can only be performed on a high concentrated pool of cells, previously washed and re-suspended in a pulsing buffer, and is therefore difficult to apply at large scale (29). Cation chloride salts, especially calcium phosphate salts, have been used for several decades to transfect mammalian cells (30). This method is based on the co-precipitation of the DNA with calcium phosphate (CaPi). This complex is then added to the cell suspension. It has been shown that CaPi induces the endocytosis of the com-

plex and protects the DNA from degradation by intracellular nucleases. However, the time-sensitive nature of the transfection protocol makes the implementation at large scale a challenge (31). Most recently, other salts have been successfully used to transfect DNA in mammalian cells (32). However, the fact that the presence of serum is still a pre requisite, makes the method unsuitable for the production of clinical products (33–35).

Several types of viruses have been engineered for *in vitro* gene-delivery in mammalian cells (28). However, most of the systems have their own advantages and disadvantages and are suited for specific cell types and/or applications, making difficult the generalisation of one viral carrier for the production of recombinant proteins. More importantly, the use of the majority of these viral vectors has not been approved by the regulatory authorities due to their immunogenicity. Recently, recombinant baculovirus vectors engineered to contain mammalian cell-active promoter elements (known as the BacMam system from Invitrogen), have been used successfully for transient gene delivery in a broad spectrum of mammalian cells including CHO cell-lines (36). In contrast to other commonly used viral vectors, baculoviruses have the unique property of replicating in insect cells, while being incapable of initiating a replication cycle and producing infectious virus in mammalian cells. The viruses can be readily manipulated, accommodate large insertions of foreign DNA, initiate little to no microscopically observable cytopathic effect in mammalian cells and present a good biosafety profile. However, the production, isolation and amplification of the vector still represents a lengthy, costly and challenging process, which limits its application for large scale production of recombinant pharmaceuticals (37).

Cationic lipids and polymers are commonly used to transfect mammalian cells. Although the transfection efficiency is significantly lower than for viral carriers, cationic vehicles present several advantages including safety, low immunogenicity and easy-to-use properties. The success of transfection relies on the capacity of the carrier to encapsulate and/or condense the DNA and to transport it through the natural barriers of the cell to the nucleus. Cationic lipids, and more precisely

synthetic amphiphiles, represent the most used carriers for small scale research experiments. DNA is firstly mixed with the positively charged lipid carrier. The DNA can be encapsulated or directly bound to the lipid. The complex will then interact with the negatively charged cell membrane. The complex can be internalised by endocytosis and the DNA released during the liposome trafficking, and/or the DNA can be directly translocated into the cytoplasm after fusion of the lipid with the cell membrane (38–40). However most of the cationic lipids present two main drawbacks: a cell cytotoxicity at high concentration, and a high cost, minimizing their use for large scale production.

Synthetic polycations proved to be an interesting alternative to other transfection vehicles as they are cheaper, relatively efficient, safe and ease to use, making them suitable for large scale transfection. The most promising to date is the linear polyethylenimine (PEI). Its properties are presented below.

Table 1.1: Comparison of gene transfer methods

		Advantages	Disadvantages
Mechanical Means	Electroporation Particle mediated Needle injection Jet injection	Applicable to any cell type Efficient transfection	High cell mortality rate Complex methods/protocols Not applicable at litre scale
Co-Precipitation	$CaCl_2$ CaP_i	Inexpensive Applicable to any cell type Applicable at large scale	Complex protocol Need of serum
Virus Carrier	Baculovirus Parvovirus Adeno virus Herpes simplex virus Lentivirus Retrovirus Alphavirus	Broad host cell range Efficient transfection	Strong immunogenicity (excepted baculovirus) Costly and time consuming production Limitation to small DNA inserts
Cationic lipids	Liposomes Lipid particles	Safe, stable and biodegradable Efficient transfection	Expensive Cytotoxic
Cationic polymers	Synthetic Polysaccharides	Safe and stable Biodegradable (Polysaccharide only) Applicable at large scale Applicable to any cell type Inexpensive Medium to good transfection efficiency	Relatively complex protocol Cytotoxic (Synthetic only)

1.1.4 PEI-mediated transient gene expression

The production of biopharmaceuticals at large scale requires a DNA carrier to be GMP compatible, stable and cost-effective. Moreover, the carrier must efficiently tranport the plasmid DNA through the natural barriers of the cell, protect it from the cell's natural desoxyribonucleases, and once within the cell nucleus, not interfere with the DNA transcription process. PEI chemical is so far the most promising DNA carrier as it presents all the characteristics cited above, and has been used in the past at litre scale for the production of biopharmaceuticals (41–44).

1.1.4.1 Mechanism of transfection using PEI

Even if some discrepancies exist in literature with respect to the mechanisms involved during cell transfection, there is no doubt that the relatively high efficiency of PEI relies on its high cationic charged density potential (Figure 1.3).

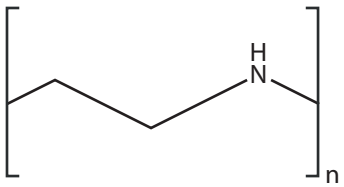


Figure 1.3: Linear PEI structure. Linear PEI contains protonable secondary amino groups every third atom.

PEI molecules are able to interact electrostatically with negatively charged DNA, to condense it, and form stable particles called polyplexes. The mechanism of the following internalisation of the complexes within the cell is subject to controversy. Some studies proved the polyplexes could be internalised via clathrin and caveolae mediated endocytosis (45, 46). Others showed that it could occur via interaction with negatively charged heparan sulfate proteoglycans (HSPG) on the outer side of the cell membrane. Using a proteoglycan-deficient CHO cell-line Payne *et al.* observed a 3.6 times reduction in transfection efficiency compared to a normal CHO cell-line (47). Kopatz *et al.* explored the mechanism further and found that. subsequent to the interaction of polyplexes with the cell membrane, HSPGs diffuse laterally to cluster in cholesterol rich rafts, and trigger an actin-mediated phagocy-

tosis mechanism (48).

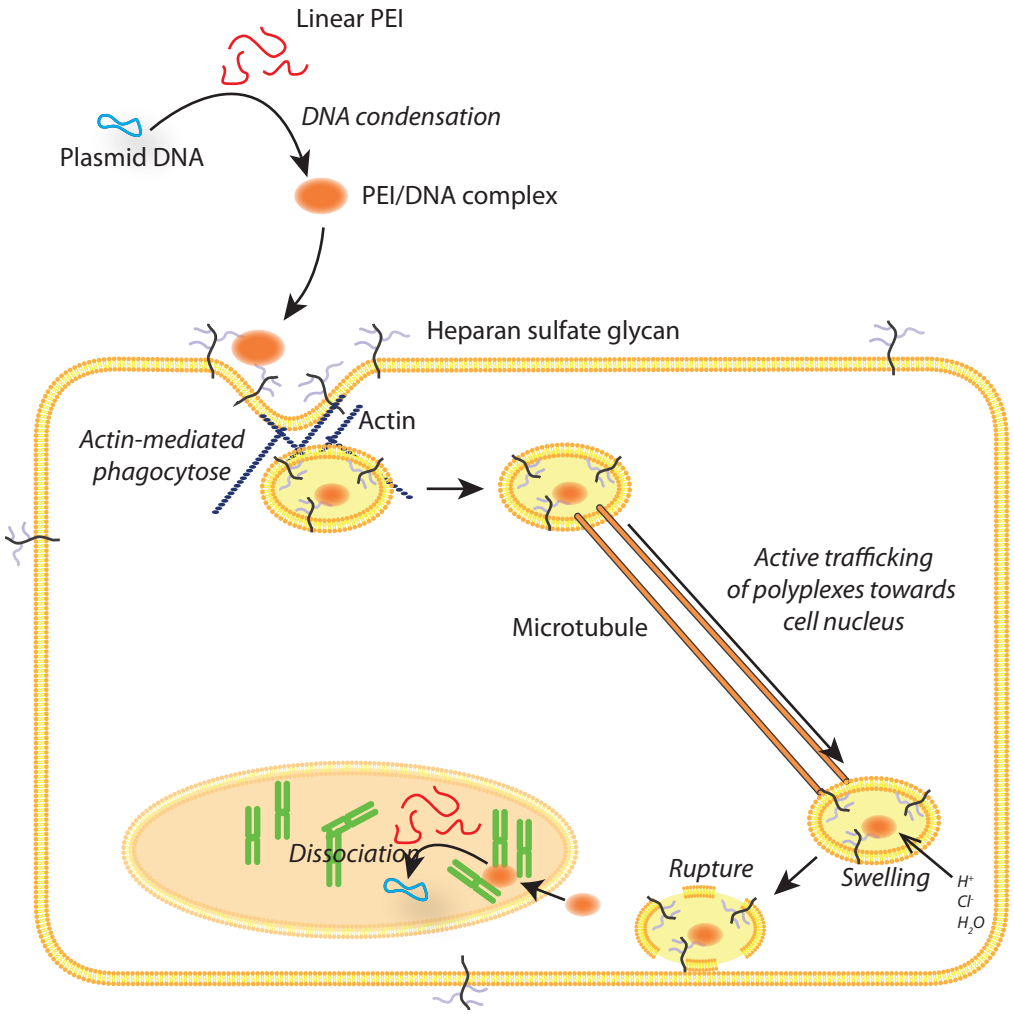


Figure 1.4: PEI/DNA polyplexes transfection and trafficking within mammalian cells. PEI molecules are able to interact electrostatically with negatively charged DNA to condense it, and form stable particles called polyplexes. Polyplexes are then internalised following an heparan sulphate mediated endocytosis. Trapped in endosomes, polyplexes migrate towards the cell nucleus via a protein driven transport on microtubules. Polyplexes are then liberated in the cytoplasm, by rupture of the endosomes, due to an influx of water caused by the the proton sponge effect of polyplexes.Following their translocation into the cell nucleus, PEI molecules will exchange from plasmid DNA to surrounding chromatin.

Trapped in cellular endosomes, it is believed that DNA degradation is prevented by its tight association with molecules of PEI (49). While the work of Suh *et al.* shows that complexes are transported through the cell's cytoplasm to the perinuclear zone, within minutes, by a protein driven transport on microtubules (50), Payne *et al.* contradicts this mechanism by showing that intact microtubules are not a pre requisite to polyplexes trafficking to late endosomes, and that polyplexes trafficking

could mostly be directed to late endosomes by their initial HSPG receptors (47). Polyplexes are then liberated in the cytoplasm by a mechanism called the proton sponge effect. PEI specificity over other polycations is a consequence of the high number of protonable amino groups which results in a greater buffering effect in endosomes. The pumping of protons, along with the concurrent influx of chloride ions to maintain charge neutrality, increases ionic strength in the endosome. This results in a swelling and a rupture of the endosome (51). Moreover, the proton sponge effect would contribute to increase the endosomal pH and therefore inactivate most of the endosomes nucleases. The relative proportion of PEI/DNA complexes escaping the lysosomal trafficking pathway is relatively unknown and is cell-line dependent. It seems that PEI/DNA complexes can be found in different forms in the cytosol: captured in an endosome, bound to membrane fragments of a burst endolysosome, or completely free within the cell cytoplasm (51). The mechanism of DNA internalisation in the nucleus remains unclear. A controversy exists regarding a potential cell competency for complex internalisation into the nucleus. Indeed, according to some authors, the transition through mitosis, with the concomitant transient loss of nuclear membrane integrity, is likely to be a prerequisite for nuclear entry of endosomally derived complexes and expression of transgenes (52–54). On the contrary, some studies showed that PEI/DNA complexes are translocated into the nucleus without dependence on membrane integrity or cell mitosis (55–57). The fact that post mitotic cells such as neurons could be transfected suggest that non-mitosis dependent entry of plasmid into the nucleus is a reality. Pollard *et al.* showed that PEI could promote transgene delivery to the nucleus via electrostatic interactions with anionic lipids present at the nucleus surface (58). The presence on the DNA plasmid of specific sequence, such as a SV40 promoter, showed to promote DNA import via the nuclear localisation signal import machinery (59, 60). The whole PEI/DNA complex is thought to enter in the nucleus. The polyplexes will dissociate following an exchange of the PEI polymers with surrounding chromatin and competitive interaction with naturally occurring nuclear polyamines (61). The complete dissociation of DNA/PEI is not, however, necessarily a prerequisite prior

to transcription (62). The principals mechanisms identified to date are summarised in Figure 1.4.The several mechanistic discrepancies reported in the literature may come from the fact that studies have been carried out using different mammalian cell-lines. Indeed, Von Gersdorff, for example, described a cell-line specific preferred route for polyplexes internalisation (46).

1.1.4.2 Optimisation in mammalian cells

Since Schlaeger and Christensen have demonstrated the usefulness of PEI for large scale transfection, numerous studies focused on improving the method to increase protein titres. Titres per litre now range in the milligrams to gram category whereas a decade ago it was only possible to produce microgram to milligram quantities of protein. Optimisation studies focused on different aspects of transient production: (i) the transfection process and (ii) the modulation of the transgene expression during the culture.

The success of PEI as a transfection vehicle lies in its ability to condense DNA to form polyplexes. The mechanisms underpinning the formation of polyplexes have been investigated in depth. It was found that the formation of polyplexes was a kinetic reaction influenced by the size of the DNA vector, the molecular weight and shape of the PEI molecule, the relative quantities of PEI and DNA, as well as the medium in which the complexation occurs (63). Both the size and shape of the polyplexes affect transfection efficiency (64). PEI molecules can be found in linear or branched forms and in various molecular weights ranging from a few to several thousands kilodaltons (kDa). The nature of PEI molecules is thought to influence the rate of association and dissociation with DNA molecules as well as the characteristics of PEI/DNA polyplexes. This is of primary importance as a weak association will result in the degradation and/or loss of DNA molecules through trafficking towards the cell nucleus, large polyplexes will struggle to be endocytosed, and too strong an association will probably prevent the dissociation of PEI/DNA polyplexes and limit the transcription of DNA within the cell nucleus.

So far, transfecting a cell culture using PEI can be done in two ways. The

most common approach involves incubating the PEI and DNA molecules for a given amount of time before adding them to the culture (65). However, Backliwal *et al.* have shown that good transfection efficiency could be achieved by directly adding PEI, then DNA, to a highly concentrated cell culture (66). Linear 25kDa PEI molecules are so far judged as being the most efficient form of PEI. However, several laboratories are modifying parental PEI molecules to further improve its efficacy as a gene delivery vehicle. Commercially available jetPEI® (Polyplus) shows higher transfection of small interfering RNA (67). Conjugated PEI with tween 85 or lipoic acid showed to efficiently mediate non toxic gene delivery (68, 69). Indeed the main drawback of PEI is its cytotoxicity. In the range of generally used concetrations, PEI both induce cell membrane destabilisation and promote apoptosis via activation of mitochondrially mediated apoptotic program. Free PEI molecules seem to be more cytotoxic than when associated with DNA. Therefore, it is not surprising that concentrations of PEI and DNA, the ratio PEI/DNA, and the cell concentration at transfection are considered as the most critical factors for efficient transfection and have been subject to numerous optimisation studies (34, 35, 53, 70–74). As shown in Table 1.2, the optimal values of these critical factors significantly differ from one study to another. It can be explained by the fact that PEI-mediated transfection is a complex process in which a large number of factors are interacting. Except for Thompson *et al.*, all the studies presented have been carried out by optimizing one factor at a time, meaning that factors interaction have been ignored. It is highly probable that the identified optimal values for each factor depend on the order in which factors have been optimised. As shown by Thompson *et al.*, it seems that the specific characteristics of the CHO cell-line may also play a large role in the variations observed. On another level, Tait *et al.* increased transfection efficiency and subsequent protein expression by synchronizing CHO cells in G2/M phase using the microtubule polymerizing agent nocodazole. The proposed mechanism is that during mitosis, more polyplexes could enter the cell nucleus thanks to the loss of membrane integrity (53).

Table 1.2: Summary of basal factors concentrations used or optimised for transfection in CHO cells using 25kDa linear PEI.

Cell line	DNA:PEI Ratio (w/w)	[PEI] (μg mL^{-1})	[DNA] (μg mL^{-1})	[Cell] (μg mL^{-1})	Medium	Reporter Protein	Reference
CHO K1SV	1:2.7	2.7	1	0.2	*DMEM*	*SEAP*	[77]
CHO-S	1:1.35	1.35	1	0.2	*CHO-SFM II*	*Luciferase*	[53]
CHO-S	1:3	3	1	1	*Mix**	*GFP*	[71]
CHO DG44	1:3	7.5	2.5	2	*ProCHO5*	*mAb*	[74]
CHO DG44	1:2	5	2.5	0.5	*RPMI 1640*	*GFP*	[42]
CHO DG44	1:2	7	3.5	2	*CHOM*	*SEAP*	[72]
CHO DG44	1:5	12.5	2.5	2	*ProCHO5*	*GFP*	[34]
CHO L	1:2	22.6	11.2	2.1	*CD-CHO*	*SEAP*	[73]
CHO M	1:1.1	13	11.4	1.6	*CD-CHO*	*SEAP*	[73]
CHO-S	1:3	9	16.3	2.5	*CD-CHO*	*SEAP*	[73]
CHO DG44	1:3	0.625	0.1875	4	*ProCHO5*	*mAb*	[70]

* Mix: 25% CD-CHO/ 75% DMEM

An inherent limitation in TGE is the premature cell death caused by the stress the cells undergo during the process. Indeed, transfection protocols generally involve several dilutions and/or cell concentration steps, plus the addition of cytotoxic PEI and exogenous DNA material. As a result, transient production processes tend to be shorter than other culture processes. Moreover, the early cell death may decrease the overall yield, while the lysis of cells may liberate cellular enzymes, such as glycosidases, proteases or reductases, potentially affecting the product quality. A way to overcome this problem is to maintain the viable cells for a longer period. Culture process designs, such as fed-batch or perfusion cultures, limit by-product accumulation and continuously supply the culture with nutrients to extend the lifespan of the culture. These processes can be applied to transfected cultures. Using a perfusion mode, Sun *et al.* successfully extended the culture phase post transfection, significantly improving protein titres compared to a normal batch culture (75).

Another approach is to modify the serum-free culture medium by supplementing it with different amino acids, growth factors or peptones. Pham *et al.* reported a two fold increase in secreted alkaline phosphatase transiently expressed in HEK 293 host cells after the addition of a casein peptone 24 hours post-transfection (76). This effect was shown to be time, concentration and cell-line dependent. The addition post transfection of some peptones promoted gene expression but resulted, in some cases, in an enhancement of cell growth at the expense of protein expression. Stettler successfully used a peptone supplementation method to enhance TGE in large scale CHO cell culture (74). However, the detrimental effect that peptones on product purity needs to be considered. Mild-hypothermia also proved to be a valid option to maintain cell viability in culture by promoting the accumulation of cells in the G1 phase of the cell cycle with an increased cell size, a reduced cellular metabolism, a greater cell viability due to decreased accumulation of waste products, and a potentially, partially inhibited PEI-mediated cytotoxicity (77, 78). Another limitation to high yielding cultures is the degradation and dilution of the transgene during the culture. It seems that mild-hypothermia also contributes to maintain high steady-state levels of transgene mRNAs. On another level, *in vivo* and *in vitro* studies have

demonstrated a rapid decrease in transcription from extra-chromosomal DNA prior to the physical loss of the DNA from the nucleus, as the result of association with histone deacetylases (79). Histone deacetylases promote DNA condensation within the cell nucleus and therefore inhibit DNA transcription. Strategies have been developed to overcome these epigenetic pathways using specific histone deacetylase inhibitors such as sodium butyrate or valproic acid (80). Valproic acid has been approved by the FDA in the past and is relatively inexpensive. Therefore its use can be extended to the production of biopharmaceuticals in large scale culture.

1.2 The future of manufacturing processes

The biopharmaceutical industry has long been characterised by high research and development costs. Until recently, companies managed to sustain high profit margins thanks to few blockbuster medicines. This is no longer the case. First the "blockbuster" paradigm does not prevail anymore and we will more likely see, in the future, more products with shorter life cycles. Second, numerous pressures including governmental regulations, competition from generics, and shortening of blockbuster patent protection times, incite biopharmaceutical companies to reduce the time, costs and lack of flexibility of current manufacturing platforms, as well as the risks of product failure during late clinical trials. So far several have been and are being investigated:

- Limit unexpected, unforeseen issues during clinical trials by developing a good knowledge of the product and the manufacturing process early on.
- develop/use tools for both high-throughput and better characterisation of candidates
- increase manufacturing flexibility for fast product turn-over
- streamline processing units for more straight-forward manufacturing

1.2.1 Product quality as a driver for faster process development

Ten years ago, the Food and Drug Administration (FDA) published a draft paper on the good manufacturing practices for the 21st century (81). The document specified that companies should build quality, safety and efficacy into their products as early as possible. This concept became known as Quality by Design (QbD). A more recent report (82) mentions that adopting a QbD approach to process development could:

- Streamline product development
- Simplify regulatory compliance and increase flexibility

- Fasten improvements to product manufacturing.

But how exactly? The easiest way to understand the advantages of QbD is to look at the main reasons of a product development failure. Success or failure are defined according to three dimensions: time, cost and product quality. In other words, a project fails if the product fails at meeting initial expectations, and/ or if the development takes too long or is too costly (83). Generally, time and costs are overrun because the product fails at meeting all the expectations and induces manufacturing process readjustments. The cost of failure is huge, not only financially, but also from an ethical and reputation point of view. Prevention against failure lies on the extent to which a product quality risk can be anticipated and managed early on. Yet, the quality of a biological product depends to a large extent on the design of the manufacturing process. Biopharmaceuticals, and especially antibodies, are complex molecules, expressed in a complex environment. Therefore, compared to chemically synthesised products, biopharmaceuticals are subject to numerous contaminants and, potentially, structural alterations. In the case of antibodies, protein folding depends on low energy hydrogen bonds, and minor changes in the expression environment may generate structural variants. Yet, the efficacy but also safety of the molecule is conditioned by its structure. For example, changes in folding may affect receptor binding and/or signaling. Abnormally folded proteins may impact on immunogenicity through product aggregation or fragmentation. Natural post translational modifications can also be altered (Figure 1.5). Those modifications, as well as the presence of process-related impurities, may be critical with respect to product pharmacokinetics, activity and, in some cases, safety. Typical quality attributes for a monoclonal antibody are presented in Table 1.3.

QbD represents a systematic and rational approach to process development that encompasses predefined objectives, and emphasises the link between product quality and process design based on quality risk assessments (86). QbD innovation lies in directly promoting and proving the quality of the product during the development, by characterizing the product and the process early on, to avoid failure later on (87, 88). QbD should be implemented wherever possible, from the development of a

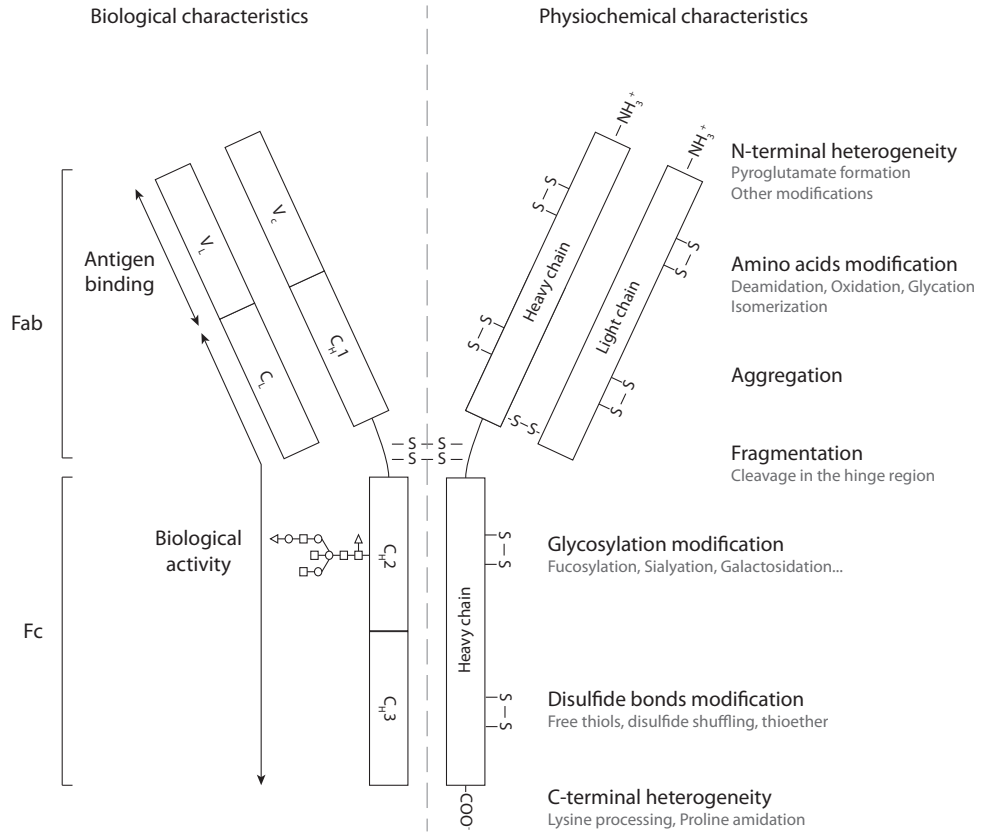


Figure 1.5: Antibodies structure is subject to potential structural modifications in culture.
V_L: variable region on light chain; C_H: constant region on heavy chain. Adapted from (84).

single processing step to the whole manufacturing process. The QbD approach to the development of one single processing unit is detailed in Figure 1.6. Each processing unit's objectives should be aimed at improving product quality while maintaining good enough process performance. The first step is therefore to define the Critical Quality Attribute (CQA) of the product: the "physical, chemical, biological or microbiological properties or characteristics that should be within an appropriate limit, range or distribution to ensure desired product quality" (86). At first, a risk assessment is aimed at establishing the potential interactions between the process specific CQAs and the desirable process performance markers with the controllable process parameters. In other words, this risk assessment serves to build a rational experimental strategy. Process characterisation can therefore be conducted. The experimental design evolves with the first experimental results until the process is fully characterised. Finally a design space is chosen, *i.e.* critical operational param-

Table 1.3: Typical quality attributes for a monoclonal antibody. From (85)

Products variants	Process impurities	Drug formulation attributes
Aggregation	Microbiological	Foreign particles
Conformation	Virus	Clarity
C-terminal lysine	DNA	Color
Deamided isoforms	HCP	Osmolality
Disulfide bonds	Protein A	pH
Fragmentation	Buffer components	Concentration
Glycation	Medium components	Potency
Glycosilation		Volume
Oxydation		
Thioether link		

eter ranges where all CQAs are within an acceptable range. Because most of the experiments are conducted using Design Of Experiments (DOE), it is necessary to perform a validation step to control that predicted values are correct. QbD therefore presents many advantages. Because of the extended amount of knowledge gained early on in the development, the risk of introducing the "wrong" candidate into clinical trials is reduced. Moreover, the fact that the quality risk assessment is based on previous experience and data available within the literature can significantly speed up the screening of candidates by ruling out the ones that will most probably fail at a later date. Finally, QbD allows for the development of well characterised and therefore flexible process platforms for which the impact of a change in process parameters on product quality is known and detailed. Still in its infancy, the evolution of the number of recent publications highlights the importance this methodology is gaining in our industry (Figure 1.7).

1.2.2 Current processing options for the manufacture of antibodies

Monoclonal antibodies represent a unique class of biopharmaceuticals. Their homologous structure, combined with the nearly universal use of mammalian cells as expression hosts, make possible the harmonisation of manufacturing around base platforms that can accommodate slight product variations. For these reasons, mAbs represent really attractive products for biopharmaceutical companies as they are associated with rapid process development time-lines, and relatively cheap costs of production. In fact, the relatively high price of this family of products on the mar-

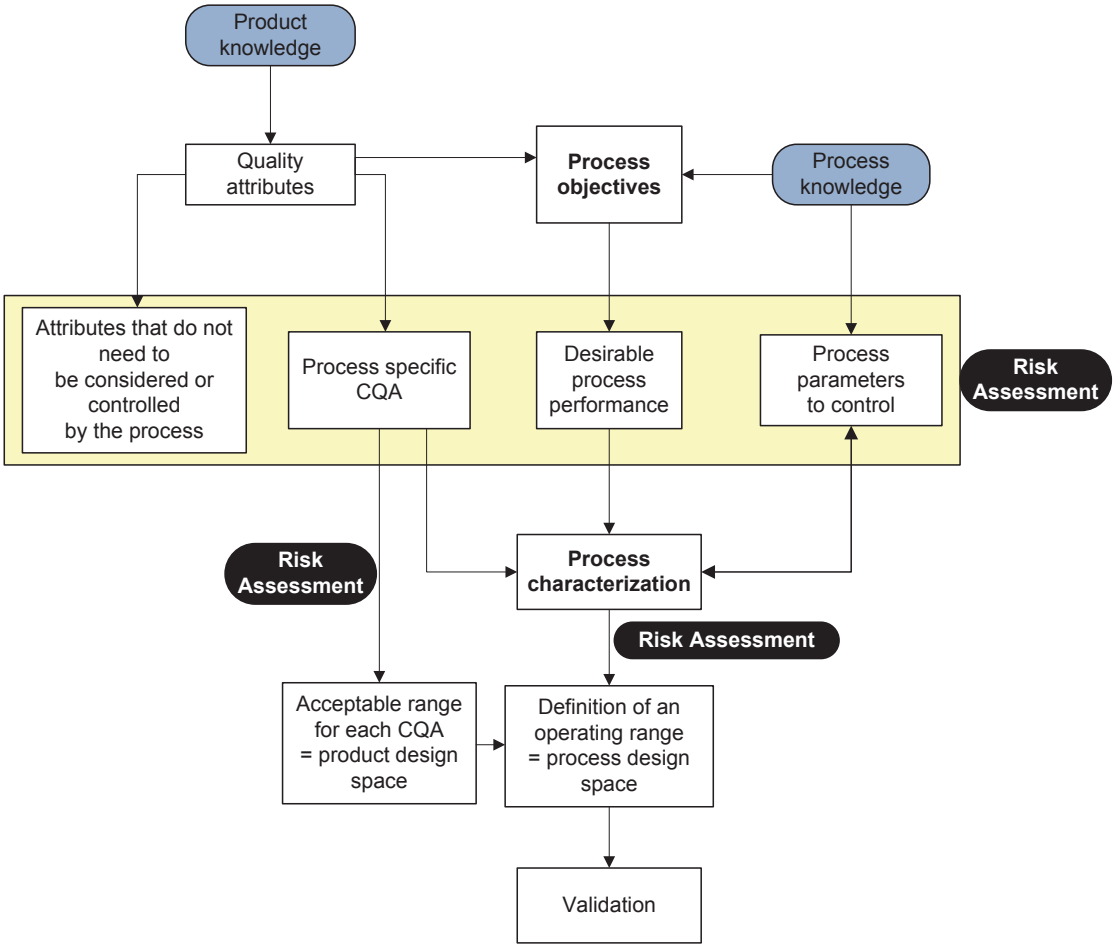


Figure 1.6: QbD approach to the development of a process step. QbD is a rational approach towards process development. It is based on systematic risk assessments through the development to efficiently characterise, then control the critical operational parameters to ensure the desired product quality.

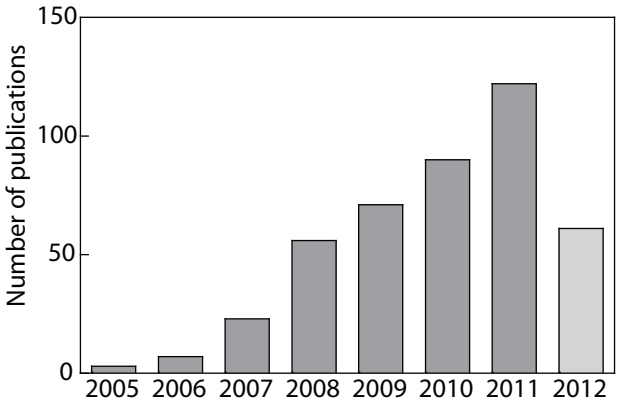


Figure 1.7: Evolution of the number of publication on QbD between 2005 and 2012. Study was conducted using SciVerse Scopus database on the 4th July 2012 with "Quality by design" keyword in article title, abstract or keyword within the Life Sciences subject area.

.

ket is justified by the high research and development costs that led to their release on the market.

The last two decades have seen the emergence of generic production platforms for the manufacturing of mAbs. Those are presented in Figure 1.8. Despite a relative uniformity with respect to the global processing chain, several steps are in fact subject to a lot of variations between products and/or companies, upstream cell culture probably being the step encompassing and leading to the greatest variety. Indeed, numerous different culture medium, mammalian cell hosts, product type as well as processing technology and process design, assure a great diversity of cell culture broth in terms of products and product contaminants. If the clarification step is relatively unaffected by the nature of product expressed in the cell broth, a feed containing a relatively high amount of contaminants can cause issues such as filtration membrane fouling or extended/complicated centrifugation protocol. Since no single chromatography step can achieve the product purity required for biopharmaceuticals, a multi-step purification process scheme is required. Capture by Protein A became a standard as the high specificity of Protein A for binding IgG makes it a very robust process, relatively insensitive to variations in the feed composition. The following intermediate and polishing steps are generally performed using ion-exchange as the technique is highly effective at removing low concentrated products such as DNA, aggregates and host cell proteins (HCP). The ICH Q5A guidance document recommends the use of at least two orthogonal steps for viral clearance (89). The first one is usually performed right after the Protein A binding step as the product is eluted in a low pH buffer. Indeed, low pH treatment has shown to successfully inactivate many enveloped viruses. The second step involves mechanical separation by a filter. Finally the product is concentrated and formulated in a buffer where the product will be stable *e.g.* a histidine buffer. The main techniques involved in mAb processing are presented below, with the exception of upstream cell culture processes as those have been described in previous sections.

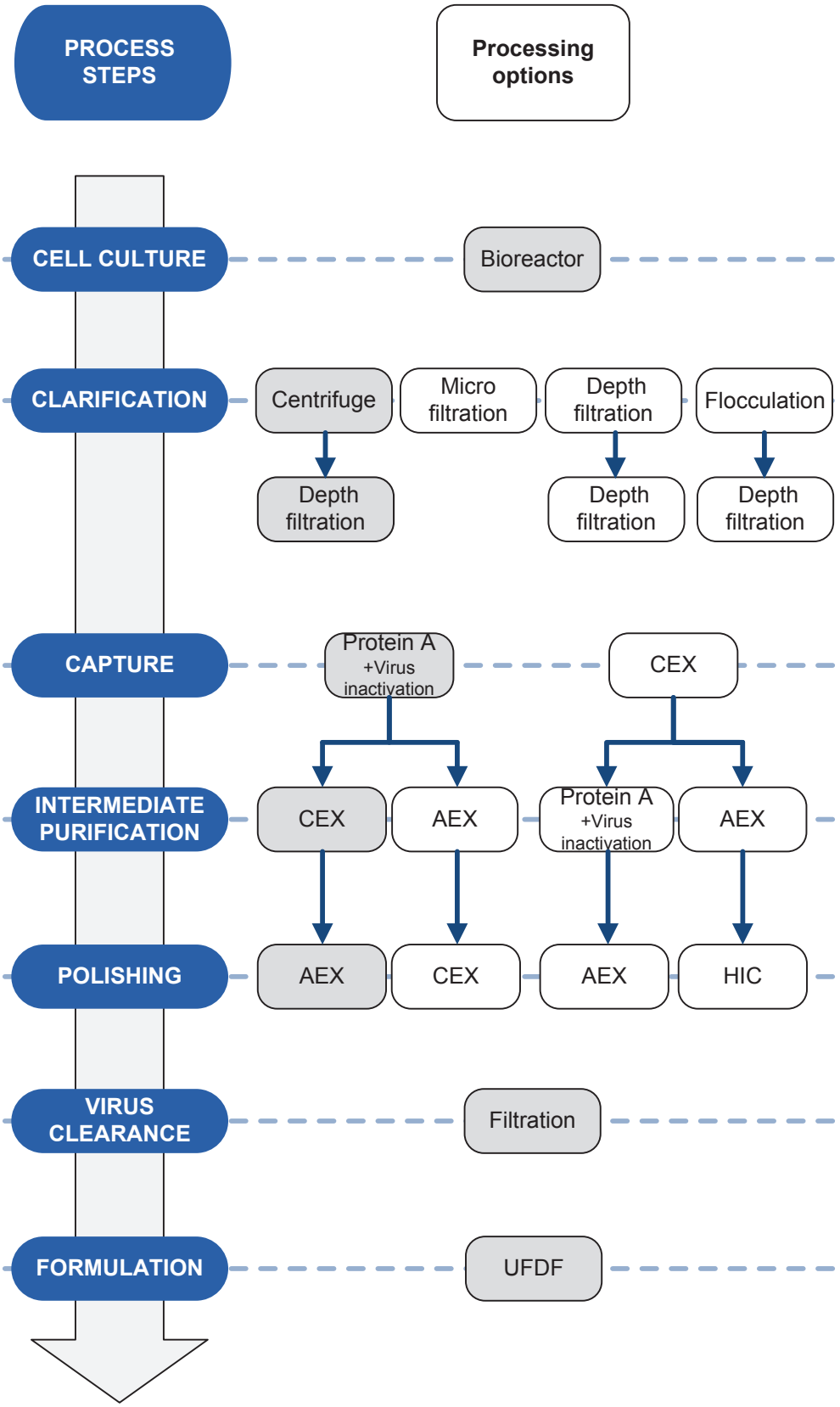


Figure 1.8: Processing schemes for the production of mAbs. The typical mAb processing options are highlighted in grey.

1.2.2.1 Clarification

The objective of clarification is to separate the product of interest from a solution concentrated in cells, debris, particles and colloids. Centrifugation and microfiltration are the primary techniques used today in industry. While depth filtration can be used as a harvest method it is more common to employ this technique as an additional clarification step. For cell culture broths rich in nutrients, minimising the duration of the harvesting step to a few hours is of primary importance to prevent bacterial growth during the process. Centrifuges accelerate the settling that normally occurs during sedimentation by applying a centrifugal force. Centrifugation can be used at every processing scale from millilitres to thousands of litres. Most large scale applications (hundreds to thousands of litres) use disk-stack centrifuges. Conventional lab centrifuges can effectively be used for scales up to 100 litres. The main drawback of a centrifuge is its inability to separate contaminants with a molecular weight close to that of the product of interest. As a result, centrifugation processes are often followed by an additional depth filtration step.

Microfiltration becomes generally useful when litres of broth need to be clarified. Compared to centrifugation, microfiltration generates a near particle-free harvest stream that requires minimal additional filtration. Microfiltration uses membranes with pore sizes ranging between 0.2 to 0.45μm. The choice of the pore size should be dictated by the composition of the feed. Indeed, membrane filtration is generally prone to fouling, with a progressive decline in membrane flux with time because of the accumulation of retained material at the membrane surface. The flux also decreases as the concentration of particulates/proteins in the feed increases. As the fluid is forced through the pores, a concentration gradient in particulates is created from the centre of the feed channel to the membrane. At high concentration, the viscosity at the membrane surface increases and a gel layer can form. Fouling occurs when the gel layer at the top of the membrane is no longer permeable. Fouling is often characterised by chemical and physical adsorption of particulates either on the membrane surface or within the pores. In other words, resistance to filtration depends on the permeability of the gel layer and the size of the membrane pores.

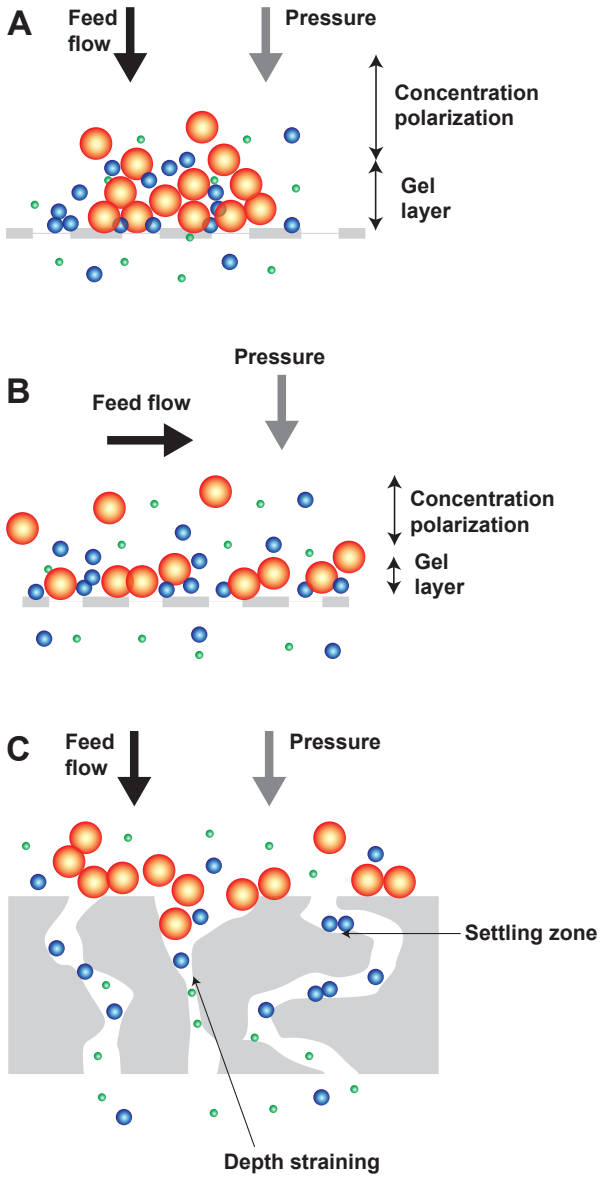


Figure 1.9: Comparison of different filtration modes. Membrane filtration is generally prone to fouling, with a progressive decline in membrane flux with time because of the accumulation of retained material at the membrane surface. Fouling occurs when the gel layer at the top of the membrane is no longer permeable. As opposed to normal flow filtration, tangential flow filtration consists of pumping the feed tangentially along the surface of the membrane, causing the particles to be swept along and reducing the thickness of the gel layer. Depth filtration differentiates by retaining particles throughout the width of a porous filter rather than just on the surface of a filtration layer. Large particles are mainly trapped in the membrane itself whereas small particles are retained by electrical and/or molecular interaction with the media.

As opposed to normal flow filtration, tangential flow filtration (TFF) consists of pumping the feed tangentially along the surface of the membrane, causing the particles to be swept along and reducing the thickness of the gel layer (Figure 1.9, A and B). Alternatives in flow configurations have been proposed to further mitigate the effects of the polarisation layer. For example, the solids accumulated at the membrane surface can be discharged back into the feed by using backflush systems, where the flow is momentarily and periodically reversed. Other alternatives consist in developing specific flow paths to initiate vortices that will sweep the membrane efficiently. The formation of the gel layer is a well known mechanism and involves two phases: a pressure dependent regime, where an increase in pressure results in an increase in flux and a pressure independent phase in which increasing the pressure will not increase the flux. Process efficiency can be significantly enhanced by adjusting the transmembrane pressure at the maximal value that still allows control of the flux. Therefore, and as opposed to centrifugation, the process often requires a fine development prior to routine operation and complicates the introduction of microfiltration as a multi-product platform.

Depth filtration differs from surface filtration in that particles are retained throughout the width of a porous filter rather than just on the surface of a filtration layer (Figure 1.9, C). Depth filtration operates on two different mechanisms: mechanical sieving and adsorption. Large particles are mainly trapped in the membrane itself whereas small particles are retained by electrical and/or molecular interaction with the media. Most depth filters used in biopharmaceutical processes are made of filter aids and cellulose fibre bound together by a polymeric sorbent that provides wet strength and presents cationic surface charges that will help retain colloidal particles. Depth filters usually consist of an series of lenticular disks assembled into a multi-stack housing. The different disks vary in terms of permeability and chemical properties. These disks are usually laid out to form a gradient, the media becoming progressively finer and denser along the passage of the solution.

Efficient and inexpensive clarification becomes more and more of a challenge as high titre cell culture processes yield a greater amount of solids, of all sizes

and density, that will foul the membrane, or even worse, escape coarse filtration and damage chromatography sorbents. An inexpensive solution consists in adding flocculants to the cell broth that will aggregate small particles and facilitate the clarification. Association between flocculants and particles is based generally on electrostatic interactions. By adding calcium chloride, then potassium phosphate to a cell broth, Coffman *et al.* managed to trap cell debris, host cell proteins and nucleic acid into large particles to yield a clear supernatant with the recovery of 95% of the antibody (90).

1.2.2.2 Capture by Protein A

Protein A is a 54kDa protein originally found in the cell wall of the bacteria *Staphylococcus aureus*. The molecule is structurally characterised by a linear series of five homologous antibody binding domains. Immunoglobulin G binds to the individual domains of Protein A via its Fc region at the junction between CH2 and CH3 domains. The Protein A-IgG interaction consists of hydrophobic interactions along with hydrogen bonding and salt bridges. The IgG class of antibody presents a highly conserved histidyl residue that aligns face to face with a complementary histidyl residue on Protein A. At physiological pH these residues are uncharged, promoting the IgG-Protein A interaction. The interaction is reinforced by the hydrophobic character of the histidine imidazole side chain (91). At low pH however, the residues are fully charged, hydrophobic and mutually repellent. Theoretically, IgG can bind to any of the five domains of Protein A. However in reality, a binding stoichiometry of 2 to 3.3 has been observed in a free solution (92). Capture by Protein A is used in a bind and elute chromatography mode. Despite the costs, Protein A high selectivity and capacity makes it the most used option for the capture of human immunoglobulin G1 and G2. Indeed, the Protein A capture step is a very robust operation. This ligand can accommodate with a wide pH range (2-11) and is able to refold after treatment with denaturing solutions. The random coil sequences that link the domains of Protein A are, however, sensitive to proteolytic cleavage and the Protein A itself can be denatured in alkaline conditions (93). It

becomes therefore compulsory to remove any leached Protein A from the product of interest to prevent any immunogenic reaction to the patient. The Protein A ligand source varies with respect to the mode of production. Natural wild Protein A can be obtained at high purity by lysostaphin digestion of the bacteria, followed by gel filtration then affinity chromatography on IgG-agarose sorbent. However, recombinant Protein A expression in *Escherichia coli* is now a preferred approach. Recombinant variants can be expressed with different features to enable directional coupling of the ligand to the solid phase sorbents. Regarding the latter, GE Healthcare has a market advantage for Protein A sorbents with the MabSelect Sure^{TM} product that can withstand strong alkaline conditions allowing the repeated use of 0.5M sodium hydroxide for sanitisation. Protein A sorbents available on the market mainly vary with respect to the solid phase matrix to which the Protein A is coupled. Matrix composition but also bead and pore size, affect sorbents compressibility, chemical resistance, mass transfer properties, capacity and selectivity (94). Agarose based matrix is, for example, a highly porous material and is present in MabSelect Sure and rProtein A Sepharose^{TM} Fast Flow from GE Healthcare. Sorbents based on rigid ceramic, such as HyperD® F from Pall Life Sciences, or porous glass, such as ProSep®-A from Merck Millipore, can withstand high flow rates and pressures during regeneration phases.

The cost of Protein A is a major concern as it contributes to a quarter of the overall costs in mAb downstream processes. Protein A can therefore not be used as a disposable component. However, Protein A can sustain a relatively high number of cycles, and can be reused more than 200 times without noticeable loss in performance. Purification platforms excluding the use of Protein A have successfully been used in the past for purifying mAbs. However, such purification schemes require extensive process development on a case by case basis and therefore cannot be used for multi-product processing platforms.

1.2.2.3 Ion exchange chromatography

Ion exchange chromatography encompasses two techniques: cation and anion exchange chromatography. These techniques are generally used as intermediate and polishing purification platforms as, when combined, they allow effective clearance of contaminants such as HCP, host cell DNA and product aggregates (95). Proteins are molecules formed by polymerizing amino acids. Amino acids are structurally characterised by at least one acidic carboxylic, and one basic amide functional group (Figure 1.10). Each amino acid also possesses a side chain or R group, that can exhibit acidic or basic properties with respect to the nature of the titratable groups it contains. In fact, R group variability alone explains the difference in pKa between amino acids. Proteins, which are built of numerous amino acids, therefore exhibit numerous ionisable groups, and possess an overall net charge. The net charge varies with respect to the pH of the solution, the charge's density and distribution within, and at the surface of the protein. When the pH of the solution is low, proteins will tend to be positively charged and are called cations. At high pH, on the contrary, negative charges will be predominant and the proteins will become anions. Because the amino acid sequence and tri-dimensional structure are protein specific, at a given pH, each protein has its own unique net charge. Ion exchange chromatography takes advantages of these properties. The technique is based on the electrostatic interactions between charged amino acids chains and the stationary surface charge of a ion exchange sorbent. Electrostatic interaction strength depends on both the ionic charge and ionic radius, which vary locally in proteins. Theoretically, cations of high charge and small ionic radius have high electrostatic interaction strength, and are more likely to interact strongly to a negatively charged sorbent surface than cations of low charge and large radius. More importantly, proteins are "amphoteric", *i.e.* their net surface charge will change gradually as the pH of the environment changes.

In ion exchange chromatography, the mobile phase is a solution containing the molecules to be separated. The solution is acid or basic in cation or anion exchange mode respectively. The stationary phase usually consists of a sorbent coated with an organic layer presenting positively charged molecules at its surface, in anion ex-

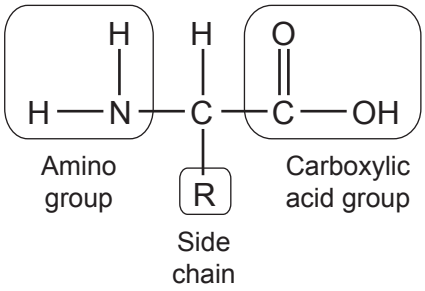


Figure 1.10: Amino acid structure

change mode, or negatively charged molecules in a cation exchange mode. When passed through the stationary phase, electrostatic interactions form between the charged groups of the side chains of the molecules present in the mobile phase, and the oppositely charged groups immobilised on the sorbent. This phenomenon of adsorption is controllable by varying the mobile phase pH and ionic strength. For example, in cation exchange mode, proteins exhibiting positives charges will adsorb on the sorbent. A slight increase in pH will result in lower charged cations and weaker electrostatic interaction energies and will promote the desorption of lightly bound cations from the stationary phase. Desorption can also be promoted by increasing the mobile phase ionic strength. By introducing counterions in the mobile phase, competition for both the immobilised sorbent and cations charged sites result in stoichiometric exchange with the bound charged groups of proteins (Figure 1.11). Sodium ion is the most commonly used counterion in cation exchange chromatography but other cations have been successfully used previously. Arginine, for example, proved to help prevent protein aggregation that sometimes occurs during elution (96). Guanidine, tetra-n-butylammonium, or calcium, have also shown to improve the selectivity of the method (97, 98).

Since ion exchange chromatography is really selective, it is possible to isolate a product of interest from relatively low quantities of contaminants. Therefore, even, if IEX represents, in some cases, a viable alternative to Protein A for the capture of mAb, the technique is often used for intermediate and polishing purification steps where HCP, protein aggregates and DNA amounts need to be reduced to clinically acceptable trace levels. In fact it is common to sequentially use a cation, then anion

exchange chromatography (99).

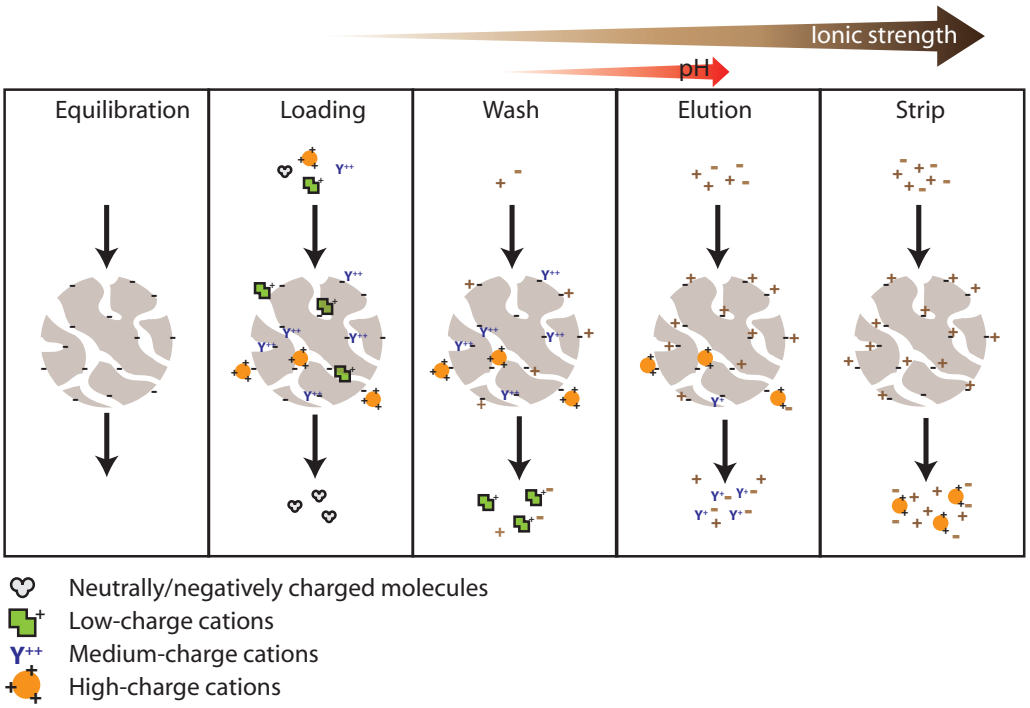


Figure 1.11: Cation exchange chromatography principle. Prior to loading, the sorbent is equilibrated with a buffer at relatively low pH and conductivity. During the loading, the negatively/neutrally charged molecules are flowing through the sorbent. A buffer with increasing ionic strength and sometimes pH is pumped through the sorbent allowing the bound molecules to sequentially elute from the sorbent. By collecting separate fractions, the product of interest can be isolated from some of its contaminants.

Ion exchange sorbents present interesting properties: sorbents are relatively cheap, have a high loading capacity and separate proteins under near physiological and non-denaturing conditions. Maximal dynamic binding capacity results from a combination of fast mass transport, large binding surface to volume ratio and quick and robust protein ligand binding. Sorbent design parameters include bead and pores sizes, ligand density as well as the chemistries of the backbone, spacers and ligand. In fact numerous sorbents are available on the market (99). Each option presents several advantages and disadvantages. The paragraph below only describes the most popular.

The nature of the sorbent backbone strongly depends on the manufacturer: Pall Life Sciences provides Q and S HyperCel^{TM} sorbents that are based on a rigid cellulose matrix that generates low back pressures. GE Healthcare products include

cross linked agarose backbones adsorbents. Polymethacrylate, used by Fractogel® for example, presents greater selectivity for removal of aggregates due to non-specific hydrophobic interactions with the backbone. Polystyrenedivinylbenzene is usually present in Poros®from Life Technologies. Weak ion exchangers, that stay ionised within a relatively narrow pH range, are usually reserved for specific applications and are not suited for a multi-product platform. Strong anion ligands include quaternary amino ethyl, triethylammonium ethyl, or quaternary aminoethyl immobilised to the sorbent. Among strong cation exchangers, it seems that sulphopropyl, sulphoethyl and sulphoisobutyl groups dominate the market (100). In order to improve performance, traditional materials have recently been coupled with polymers such as dextran to which the functional group is attached. This extended structure generally allows for better mass transfer and selectivity (101). Another modification is the filling of the bead pores with a functional polyacrilamide gel to form the "gel-in-a-shell" technology, available for both anion or cation exchange mode using the Pall Life Sciences HyperD® F sorbents.

1.2.2.4 Tangential ultrafiltration/diafiltration (UF/DF)

Tangential ultrafiltration is widely used for product concentration. The principle is the same as for microfiltration, previously detailed, except that the product of interest is retained by the membrane, while volumes of buffer are gradually eliminated from the feed, resulting in a concentration of the product. As opposed to the concentration step, diafiltration, or buffer exchange, consists of adding the buffer of the final desired composition to the retentate system at the same rate at which the permeate is removed, thus maintaining a constant volume upstream of the membrane. In mAb processing, UF/DF usually represents the last step before filling. At this stage the feed is relatively pure of contaminants and requires little to no optimisation. In fact, the process is mostly limited by the viscosity of the retentate. Ultrafiltration membranes are made from different polymers including polysulphone, polyethersulphone, polyvinylidene and regenerated cellulose. Cellulose membranes are, however, mostly used in biopharmaceutical industry due to their low protein

binding properties. Special care is, however, required as damage may occur when using harsh cleaning methods (102).

1.2.2.5 Other processing options

A variety of other technologies are, or have been developed in the past to accommodate mAb processing requirements. Among them, the charged micro and ultrafiltration membranes as alternatives to sorbent bead chromatography, and mixed-mode chromatography are particularly attractive.

The costs associated with the use of Protein A, the concern of immunogenic leached Protein A in purified product, as well as the cumbersome "3-stages purification scheme"paradigm have promoted the development of alternatives for more efficient methods of purification. Mixed-mode chromatography combines different mode for the separation of biomolecules. For example, MEP HyperCel sorbent from Pall Life Sciences takes advantage of hydrophobic charge-induction chromatography. The sorbent allows for the selective capture of mAbs by hydrophobic interaction under near-physiological conditions, then desorption, by reducing the pH of the mobile phase to promote charge repulsion between the ligand and the product. It is now possible to use mixed mode chromatography as an alternative to Protein A for capture, and to develop an efficient two-step purification scheme (103).

The efficacy of sorbent bead chromatography is mainly limited by slow purification flow-rates, due to the diffusion of the solutes into the pores of the beads. Moreover, sorbent bead chromatography requires high pressure, which necessitates robust pumps, columns, connectors and other hardware that can withstand high pressure. Chromatographic membranes are composed of porous polymer membranes. The pores are significantly larger than the pores of sorbent beads. Therefore, convective mass transfer is dominant over diffusion resulting in higher throughputs and shorter processing time without compromising product recovery (Figure 1.12). Membranes consist of polymeric substrate to which a functional ligand is chemically coupled. The ligands used are generally identical to those used in resin chromatography. Recent advances in membrane chromatography technology have yielded products suit-

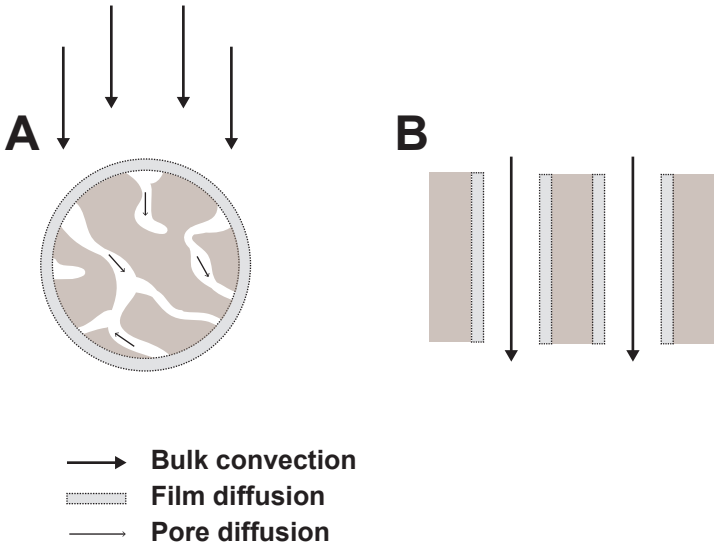


Figure 1.12: Comparison of the mass transfer mechanisms involved in resin chromatography (A) and membrane chromatography (B). As opposed to resin chromatography, convective diffusion is dominant over diffusive diffusion, resulting generally in relatively shorter processing time.

able for almost every kind of chromatographic operations including product capture, intermediate purification and polishing. Membrane chromatography is now routinely used as polishing step. At slightly basic pH and low conductivities contaminants such as DNA, leached Protein A and host cell proteins will bind to the ligands while the mAb will flow through the membrane matrix without being bound (104). Future work on this technology should be conducted to overcome current drawbacks including ligand breakthrough from membranes, low capacity and uniform flow distribution.

1.2.3 Novel manufacturing platforms

Biopharmaceutical companies face challenges to modify their manufacturing platforms, especially for small to medium scale early productions of therapeutic candidates. There is a strong need for easy, cost effective production platforms that would allow the production of milligram to gram quantities of biopharmaceuticals with minimal to no process development. Conventional bioprocessing systems are not adapted. Mostly made of stainless steel, they have been specifically designed to support the production of only one to a few molecules. A quick turn-around of

molecules requires flexibility and adaptable production platforms.

1.2.3.1 The adoption of disposable technologies in biopharma industries

The past decade has seen the emergence of single-use (SU) systems over more conventional stainless steel systems. First reserved to sterile liquid handling, connections and buffer preparation, disposables have successfully been implemented in complex bioprocessing steps. The best examples are probably the success of the WAVE Bioreactor® and WAVE Bioreactor™. In 1998, Wave Biotech was the first company to commercialise a completely disposable cell culture system (105). Wave bags are today widely used for applications at scales of up to 500L and cell expansion systems to feed stirred tank bioreactors (106). Wave bags are also used to carry out complex culture processes such as perfusion (107). Compared to stainless steel systems, SU options represent significant savings in costs, labour and time. SU also allows for maximum flexibility as challenging and time consuming cleaning and validation steps can be avoided. Therefore it is not surprising to have the SU market growing annually at a steady 15-18% rate. The advantages of SU over stainless steel are detailed in Table 1.4.

There is now a multitude of disposable bioreactors available on the market. Cylindrical or cubic bioreactor bags mimic conventional bioreactors. They are designed to fit in a stainless steel outer vessel including a heating jacket (108). Gas transfer is promoted by disposable mixing systems while air is injected into the air phase of the bag. Mechanical mixers are efficient but are a source of shear damage for the cells. Other options involve moving platforms supporting a bag. Air is sparged continuously into the headspace of the bag. The rocking motion of the platform creates waves at the liquid-air interface enhancing the aeration of the medium. Recently, two-dimensional wave motion resulted in high mass transfer capacities which are able to support the oxygen demand of high density cell cultures. Orbital shake bioreactors also represent an attractive alternative to other bioreactors as orbital mixing results in low sheer stress and the large surface of gas exchange allows efficient oxygen transfer (109).

Upstream technologies are way ahead of downstream technologies in terms of disposable solutions. In fact, according to Langer, the expansion of the SU market is slowed down by the lack of disposable options for the downstream bioprocessing of biopharmaceuticals (110), cost and scalability being the main issues. For example, the cost of Protein A sorbent, or chromatography membranes are not compatible with a SU application. Future innovations rely on a plug-and-play approach towards managing product design and innovation. This would help to provide a framework to identify where and how disposable solutions will be of real benefit for easier and cheaper processes. This could be driven by the development of processing platforms for the processing of mAbs.

1.2.3.2 Process integration

A challenge for the next decade will be to develop processing solutions adapted to the production of a multitude of products. Innovation will be driven by (i) a modular and "plug and play"approach to manufacturing and (ii) the development of well characterised and flexible platforms. In fact, the modular approach to biomanufacturing already exists in some areas. At small scale, QIAGEN provides processing disposable solutions for the purification of plasmid DNA or the preparation of genomic DNA. The solutions fit in a box that includes all the necessary components to quickly and easily perform a sequential set of tasks (112). At large scale, and in the area of protein production, Merck Millipore already offers its Mobius® FlexReady solutions. Indeed, the main drawback of stainless steel installations is the rigidity of the pipelines and connections between processing units as well as the impossibility to switch between units if necessary. Mobius product's innovation lies in the fact that each processing unit is preassembled and features all the sterile tubing, connectors and required handling systems on a moving cart (113). Therefore, units can easily connect to each other in a modular way. These process options are, however, more adapted to medium to large scale bioprocessing. In fact, there is currently a gap for a completely integrated disposable solution for the flexible, cost effective and easy production of milligram to gram of biopharmaceuticals. An approach here could be

Table 1.4: Advantages/Disadvantages of single-use systems over stainless steel systems (111).

Critical Factor	Stainless Steel System	Driver Direction and Strength	Single-Use System
Capital investment	High	>>>>	Low
Extractables/ leachables	Low	<<	High
Area footprint	High	>>	Small
Sterilisation methods	Steam, dry heat	>>>>	Gamma irradiation
Utilities supply	High use of water, steam and cleaning agent	>>>>	Low
Chemical compatibility	Good	<	Limitation due to plastic components
Physical compatibility	Good	<<	Temperature and pressure constraints
Solid waste diposal	Low	<<<	High
Liquid waste disposal	High	>>>>	Low
Energy use	High for steam and WFI water	>>>	Low
Labour requirements	High	>>>	Pre-assembled and pre-sterilised
Scale-up	Fully scalable	<<	Limitations for some process units
Process flexibility	Low	>>>>	High
Sterility assurance	Assurance by SOPs, validation and system design	>>>>	Assurance with sealed and pre-sterilised systems
Available technologies	High	<<<<	Some components not fully developed

to combine both the concept of "a process in a box" used by QIAGEN, and the modular approach of bioprocessing, each box corresponding to one processing unit. The whole boxed package would allow the realisation of all the steps resulting in the production and purification of biopharmaceuticals "on the bench". For the concept to succeed, each processing unit should be a platform, *i.e.* able to accommodate many biopharmaceuticals. Each platform should also be interfaceable with each other not only from a physical , but also biological point of view.

Platform development and process integration are complex tasks, especially with interdependent, multi-stage processes involved with the production of biopharmaceuticals. Process integration methodology typically involves different activities that have been described in the literature (114).

The first one is called "Task Identification" which aims to explicitly express the goal to achieve, and describes it as an actionable task. For example: speed-up the production of therapeutic candidates to support pre-clinical tests. This could be achieved by developing an integrated platform for the production of therapeutic proteins using disposable processing technologies. The second, called "Targeting", refers to the identification of performance benchmarks ahead of detailed design. In this project the overall objective is to produce a quantity of 1g of a purified reporter protein using a transient upstream expression system coupled with a specifically developed purification platform. Targeting is a structure-independent step. However, within each stage of the process, targets can be identified and used in the following stages. For example, in a two stage protein purification process with an overall target of 99% of purity, the targeting of 95% purity to achieve in the first stage would lead the scientist to identify the second stage targeting as the removal of 80% of the contaminants. The third one consists in "Generating Alternatives" or covering the range of possible solutions to reach the target. For example, reusing Protein A sorbent can be an alternative to decreasing the amount of sorbent needed to purify 1g of mAb. This activity is also aimed to identify the integer and continuous variables: integer variables correspond to the existence or absence of certain technologies and pieces of equipment in the solution. Continuous variables correspond to non discrete

design and operating factors such as flowrates, pressures, units sizes and assembling. The fourth activity is aimed at selecting the best alternatives, extracting the optimum and/or the most promising solution(s). Finally, the selected alternatives are tested and analysed by experimentation. Every activity in the methodology goes through a process of questioning. Some are reported in Table 1.5.

Table 1.5: Process of questioning in process integration. Adapted from (115).

1. Goal		What solution to achieve the goal?
2.Targeting	Inputs	What are the inputs? What standards must they meet?
	Outputs	What are the outputs? What standards must they meet?
3. Generate Alternatives	Inputs	How is their adequacy assessed? Do they need to be stocked, formulated?
	Outputs	How is their adequacy assessed? Do they need to be stocked, formulated?
	Operations	What operations to converts inputs in outputs? What tools, equipment and resources are needed to complete those operations? What resources are generated? What initiates the operation? What terminates it?
	Connections	What kind of connections to connect operations between each other? What tools, equipment are needed to connect operations between each other?
	Operator	Who or what is the operator? What capabilities does the operator require?
	Controller	Who or what serves as a controller? What tools, equipment are needed to control the operations? What tools, equipment control the information about Inputs and outputs?
4. Select Alternatives	Inputs	Do they fulfil the standards?
	Outputs	Do they fulfil the standards?
	Operations	What are the best tools, equipment to complete the necessary operations? Are they available? Can the operations be integrated? What are the advantages of each operation in terms of : cost? processing time? robustness? output's nature and quality? complexity?
	Controllers	Can the number of controllers be minimised?
5. Analyse of the alternatives	Operations	Can the selected operations reach the target(s)? Do the operations need optimisation?
	Controllers	Is the information accurate?

1.3 Process development using design of experiment approach

Experimentation is at the core of any process development. With respect to the new challenges that biopharmaceutical companies are facing, there is a considerable need to maximize the benefits of experimentation *i.e.*, increase the amount of meaningful data and perform more thorough process characterisation while minimizing the number of experiments for quick and cost effective development.

1.3.1 DOE methodology

Any process can be visualised as a black box in which inputs are transformed into outputs. Input variables, or *factors*, regroup discrete or continuous controllable, uncontrollable and unknown variables. The interaction of factors with the process results in the formation of outputs, also called *responses*. Some of these responses are quantifiable and can be used to develop a process.

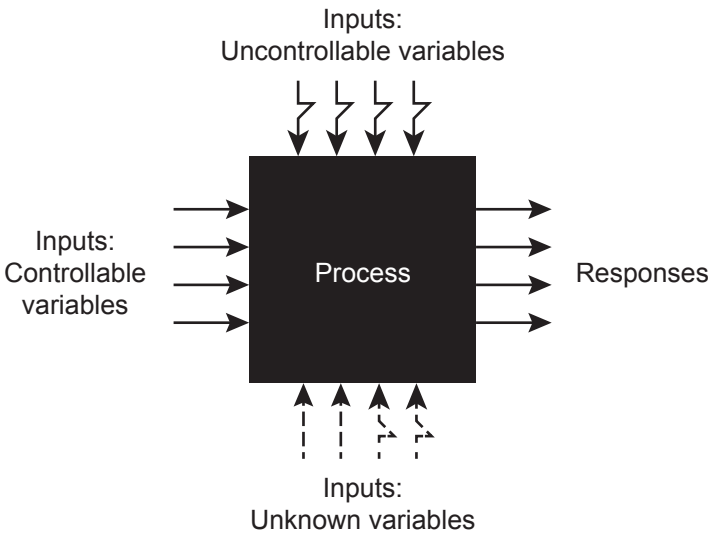


Figure 1.13: The process box. Unknown variables regroup both controllable and uncontrollable variables that may affect the process response independently, or via interaction with the known input variables.

Compared to chemical processes, biotechnology processes are significantly more complex, involving numerous reactions inside a host cell, most of them being unknown. This has two consequences: first, product and processes are highly interde-

pendent and only a slight variation in input variables may significantly affect process responses. The second is the difficulty of characterising intermediates and narrowing down the evolution of one response to one particular reaction. In this context, process development represents a critical step prior to transfer at manufacturing scale. The main difficulty being the numerous inputs variables, known or unknown, that may affect process responses in some way and complexify the characterisation and optimisation of the process. With regards to this complexity, experimentation should rely on an organised scientific approach, from the planning through conduction and analysis of experiments. Moreover, due to the cost and practical concerns, most experiments are generally conducted at small scale, laboratory experiments being a small scale representation of what the future manufacturing process could be. Because the environment greatly differs from the early research experimentation phases in the laboratory and the manufacturing environment (different experimenter, different lots of chemicals, different equipment, ...) it is critical to assess the robustness of the process before the transfer to larger scale. It is also of primary importance to screen for factors that, not varied at small scale, could potentially affect the process at large scale. These reasons support the idea that experimentation should rely on an organised scientific approach, from the planning, throughout the conduction and analysis of experiments.

The full benefits of DOE appear when the comparison is made with one factor at a time (OFAT) methodology. Conventional experimentation consists in varying one factor at a time while keeping the others constant and observes the impact on one or several responses. This operation is repeated for all factors. As described by Anderson and Whitcomb , this approach presents several disadvantages (116) :

- OFAT fails at estimating the effect of interactions of factors on the response(s).
- OFAT fails at identifying the true optimal conditions (Figure 1.14).
- OFAT requires a large number of experiments to get to conclusions, especially when continuous factors are involved.
- Process optimisation with OFAT becomes difficult when multiple responses

evolve differently with a change in factors.

To this list could be added that OFAT does not allow the scientist to know exactly to what extent the process is robust, as well as how the process responses evolve from a given distance from the optimum.

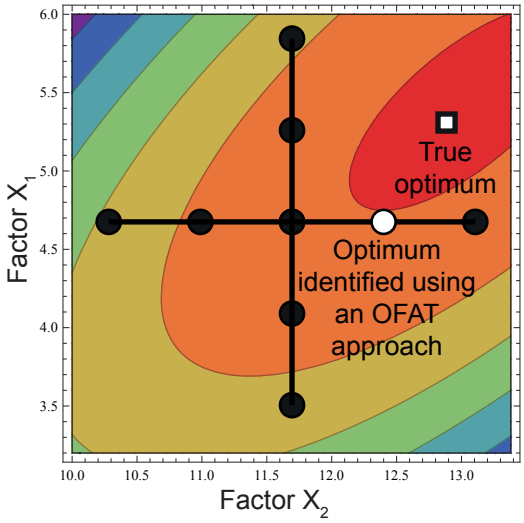


Figure 1.14: OFAT fails at identifying optimal condition. While keeping the factor X_1 constant, five experiments varying factor X_2 are performed. Then starting from the best condition (the centre point in this figure), factor X_1 is maintained constant while another set of five experiments varying factor X_2 is carried out. This OFAT approach leads to the identification of an optimum that is not the true optimum as factor X_1 and X_2 are interdependent (Contour plot was generated from a DOE-optimised PEI transient transfection process in CHO M cells (73)).

DOE mainly differs from OFAT in that all factors studied are varied simultaneously, following a previously constructed plan of experiments. Data is statistically analysed afterwards, thus allowing the scientist to get more information from fewer experiments. The DOE approach to process development is a multi-stage process. Firstly, the objective of the experiment dictates the type of DOE design to use. Following this choice, the experimenter identifies the factors and responses that needs investigating and accordingly sets up an experimental plan. DOE general methods also apply to experimentation with three basic principles for later meaningful statistical analysis: *randomisation*, *blocking* and *replication* (117). Indeed randomisation of experiments is a prerequisite to any DOE analysis involving the estimation of factors significance. Randomisation eliminates the lurking effect of unknown variables

e.g. time. Blocking, on the other hand, consists of grouping experiments in different blocks relative to a known influential variable, which the experimenter does not want to take into account during the analysis of data. In other words, blocking minimises the variation between experiments that are not accountable to the factors under study. Finally, replication consists in repeating the full experiments to estimate the experimental error but also obtain a more precise estimate of a factor effect.

An ANOVA test is then performed to reveal factors and factor interactions that significantly affect the process response. As opposed to OFAT method, which tends to gets closer from the optimum by experimentation only, the identification of the optimum is based on the generation of a mathematical model from the data collected, and the use of this model to statistically "map" a predicted response across the design space. This empirical model links process factors (X_1, X_2, ...,X_n) to one process response (Y). The same empirical model can be further analysed to map and identify the extent of process robustness. An infinite number of empirical models can be generated. But only a few are useful, the rest being misleading. Model validation is critical for meaningful analysis and often overlooked by experimenters due to its relative complexity. This will be detailed further in this chapter.

1.3.2 The DOE designs involved in process development

Several types of DOE design exist, each one being tailored for a specific purpose. Below are presented the designs that are most generally used in process development *i.e.* factorial and response surface designs.

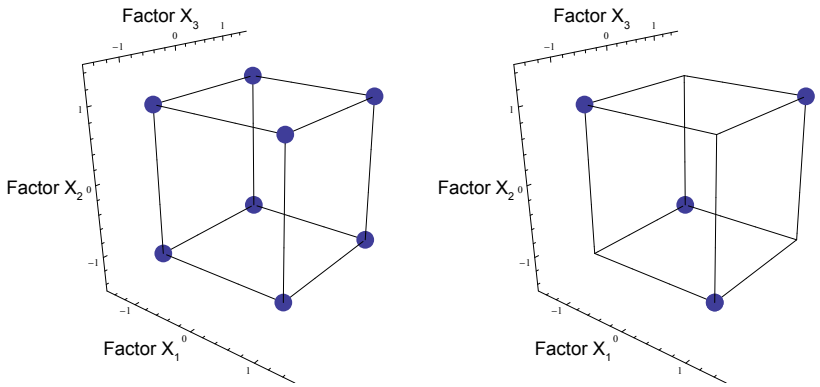
Factorial designs are used for screening the factors affecting the process responses. In this type of design, each factor is balanced at two or more *levels*, the experimental values taking the possible combination of levels across the range of factors screened. When all the possible combinations of factor levels are experimentally tested the design is called a *full factorial design*. The number of experiments to perform is relative to the number of factors, as well as the number of levels for each factor. Factorial designs are effective tools for screening of factors, especially when the factors are balanced only at two levels. Indeed, with more than two levels

per factor, the number of experiments to perform become excessive with respect to the amount of knowledge gained. A viable approach consists of performing several 2-levels factorial designs. A large number of factors to screen also leads to an excessive number of experiments. In this case a viable option is to go for a fractional factorial design in which only a fraction of the full factorial design will be experimentally tested. The fraction is chosen in the manner that high-order interactions (large number of factors interacting altogether to affect a process response) are aliased with low-order interactions. The viability of this design relies on the principle that the effect of high order interactions on process responses is negligible and that only effects due to main factors or low-order interactions are significant (118; 119) (Figure 1.15). The number of experiments is as follows:

$$Number\ of\ experiments = levels^{factors-p} \tag{1.1}$$

in which p is a real number as a $\frac{1}{2^p}$ fraction of the factorial design actually experimentally tested. The effects of different input variables on a process response can only be assessed and compared if these variables are transformed into normal, dimensionless, coded variables. To do so the upper and lower levels of factors are coded +1 and -1 respectively. Factorial designs are then really useful for the screening of input process variables, assuming the non-existence of high order-interactions.

Increasing p results, therefore, in smaller experimental design to the expense of aliasing lower order interactions with main effects. Every experimental design can be defined by a *resolution* number that varies according to the number of factors screened and the number of experiments to perform. The resolution of a design defines the alias relations between factors and their interactions (Table 1.6). Resolution II designs are useless as the effect of main factors are aliased with each other. Resolution V designs are generally considered safe for most screening steps as the effect of main factors and low order factor interactions are aliased with high order factors interactions only (120). The resolution R of the design is sometimes included in the design notation as an indication. For example a 2_V^{5-1} is a 2 levels fractional factorial experimental design of resolution V, including 5 factors and 16



(a) 2^3 full factorial design **(b)** 2^{3-1} fractional factorial design

Figure 1.15: Types of factorial designs. In the example of a 2^3 full factorial design, each of the three factors screened is balanced at two levels, bringing the total of the possible experiments to 8. Reliable information about the process can still be obtained by performing only a fraction of the total possible combinations. However, the experiments to perform must be chosen carefully and the effect of some interactions of factors on the response not identifiable.

Table 1.6: Alias structures associated with different design resolution

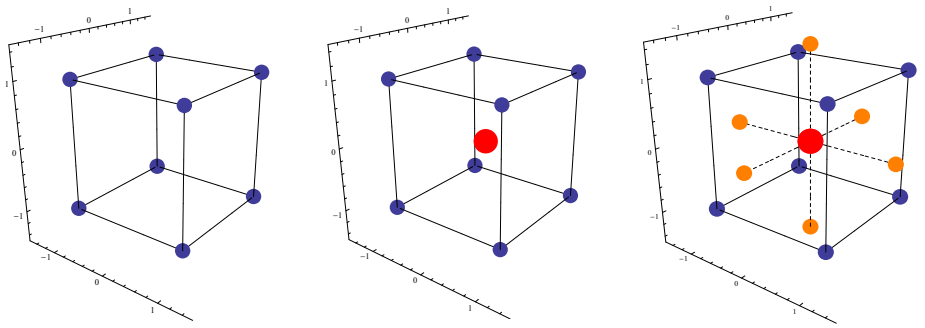
Resolution	Alias structure
II	Main factors are aliased with each other
III	Main factors are aliased with two factors interaction
IV	Main factors are aliased with three factors interactions. Two factors interaction are aliased with each other
V	Main factors are aliased with four factors interactions. Two factors interaction are aliased with three factors interactions

experiments.

Once the few vital factors as well as their interactions have been identified, the process can be further optimised by looking at optimal combination of factors as well as their optimal levels that enhance one or several responses in a desired way. Response surface methods are useful for that matter. During experimentation and for practical reasons, it is likely that the scientist will not test the exact combination of factors levels that will lead to the discovery of the true response optimum. Response surface methods (RSM) are specific DOE methods aimed at mapping process responses across a given range of levels of factors. They can provide the scientist with three and two-dimensional representations of process responses as well as the possibility to predict the response at a given, but not experimentally tested, co-ordinate

of the *characterisation space*. This is at the core of experimentation in QbD. As the experimenter gets closer from the response peak, two level factorial designs do not provide enough information to adequately model the response. Indeed, close from the optimum the likelihood to encounter a curvature in the response is relatively high. To adequately estimate this curvature, more complex designs, such as CCD or Box-Behnken designs, involving more experimental points are required (121). The most practical feature of CCD is their ability to be built blocks by blocks. In fact CCD are augmented factorial designs, with a centre point and several axial points. The centre point provides information for the estimation of curvature and, when replicated, for estimation of experimental error. Axial points allow for the accurate modelling of the process response. (Figure 1.16). The distance of axial points from the centre point can be varied by the experimenter under four constraints: *rotability*, *orthogonality blocking*, operational constraint and the number of factors. A design is rotatable if the variance of the predicted response at any point depends only on the distance of the point from the centre point. In other words, the standard error of all the points at the same distance from the centre point is equal. A design should also be orthogonally blocked, which means that there is no correlation among the factors included in the model (122). This property is important for good estimation of quadratic terms of a model. In general rotability and orthogonality can both be satisfied by positioning the axial points outside the initial design space (*circumscribed CCD*) or by positioning the factorial points inside the design space, the axial points being at the extreme of the design space (*inscribed CCD*). The latter becomes useful when the values that a factor can take are constrained within the range of the possibility that can physically be investigated. In some cases however, it is impossible to test the process with factors taking values outside the design space, whereas accuracy of the model estimates at the extremes of the design space is still required. In this case, the axial points can be positioned at the extreme of the design space instead of outside to form a cubic type of design called a *face-centred CCD*, or CCF. With a CCF, however, the distance of the axial points and the factorial points from the centre differ. The design is therefore not rotatable and estimation

of the quadratic terms of the model is relatively poor (123).



(a) Full factorial design

(b) Augmented full factorial design with a centre point

(c) Central composite design (CCD)

Figure 1.16: The construction of a central composite design (CCD) starts with a 2^3 factorial design. CCD are augmented factorial design with the presence of a centre point, that provides for the estimation of curvature and pure error and axial points, that allow accurate modelling of the process response. The distance of the axial points from the centre point can vary and specify the nature of the CCD used.

Despite increasing popularity, the use of DOE in our industry is not fully adopted due to several reasons. Firstly, "pure" biologists are not well familiarised with the technique. Secondly, DOE methodology involves complex statistics which, without the help of statistical packages, is not practical to adopt. Finally, DOE methodology reveals its true power when the whole process, from early parameter screening to final identification of an optimum in response, is developed using a rationally integrated set of DOE tools. To our knowledge, only one study, published recently, describes a global integrated DOE approach towards drug and process development (124).

1.4 Thesis outline

The work presented in this thesis describes the development of a platform for the production of milligrams to gram of recombinant protein. The platform should accommodate with the flexibility required by the quick and cost effective production of multi-products. Innovative technologies such as transient expression, and the use of statistically driven tools and methods were developed to achieve this objective.

Chapter 2 details the development of statistical tools for the creation of a computational algorithm to considerably simplify, secure and speed up the analysis of DOE-generated response surface experimental designs. In addition to the algorithm's unique properties, its performances were compared to three commercially available statistical packages.

Chapter 3 presents the development of a method for the optimisation of multivariate transient gene expression processes in CHO cells. Recombinant mAb titres could be enhanced 200 fold while maintaining cells viable using non specifically designed DNA vector, cell-line and culture medium.

The subsequent two chapters elaborates on the development of a purification train for a reporter mAb protein, and the integration of the processing units. The main aim was to achieve a reduction of the protein's contaminants to an acceptable level for preclinical studies, while maintaining its native conformation. Purification relies on an integrated three stage chromatography. In Chapter 4, a cation exchange resin chromatography step was characterised and developed using a Quality by Design approach to make it a mAb robust intermediate purification platform. In Chapter 5, this step was finally integrated with a pre-existing Protein A based capture step, as well as a developed anion membrane chromatography step as a final polishing chromatography step. The whole platform was then scaled-up to allow the production, then purification of 1g of antibody. This Chapter also introduces the economical aspect of the developed production platform.

Chapter 2

Development of an algorithm for the automated analysis of DOE-Response Surface Methods designs

This work attempts to address the difficulties associated with the building of empirical regression models for DOE-RSM designs. After presenting the theory behind the construction of regression models, the Chapter focus on the development of statistical tools, then an algorithm. Finally, the performances of the algorithm were compared to those of available statistical softwares through a case-study.

2.1 Introduction

Response surface modelling (RSM) is part of DOE methodology and regroups several mathematical and statistical techniques for empirical model building. Empirical models can be used to estimate the evolution of a process response across a continuous design space, and identify a set of process parameters leading to a minimum or maximum in response. RSM is, therefore, extensively used in process development and is systematically used in QbD methodology. RSM is a sequential methodology

that consists in (Figure 2.1):

1. Formulating the objective of the experiment
2. Choosing the appropriate RSM design
3. Identifying a suitable design space
4. Designing a plan of experiment
5. Performing the experiments
6. Modeling data
7. Analyzing the model
8. Drawing conclusions

Recent reports showed that the FDA strongly encourages biopharmaceutical companies to adopt a Design of experiment (DOE) driven methodology towards process development. Whilst it is possible to conduct DOE with general statistical softwares, a recent survey showed that considerable progress still has to be made. Indeed, 29% of the respondents judged that DOE software packages can lead to "illogical biological recommendations", whereas 26% of them found the packages not being user friendly and 21% too complex (125). This study clearly points out that DOE softwares should eliminate the need for statistical expertise on the part of the user. If currently available softwares now guide the user through the DOE driven experimentation by performing all the calculations, none to our knowledge tackled the problem of empirical model selection. Yet, model selection represents the biggest challenge in RSM. Indeed, the success of a RSM design can be judged by its ability to generate a model that properly fits experimental data and makes accurate predictions across the operational space. Generating a model is relatively easy thanks to the available statistical softwares. However, model validation, prerequisite to the use of the model, is so far a non-automated task and is the responsibility of the experimenter. Moreover, several assumptions relative to the model analysis, as well as model characteristics need to be validated or checked prior to using the model. A large number

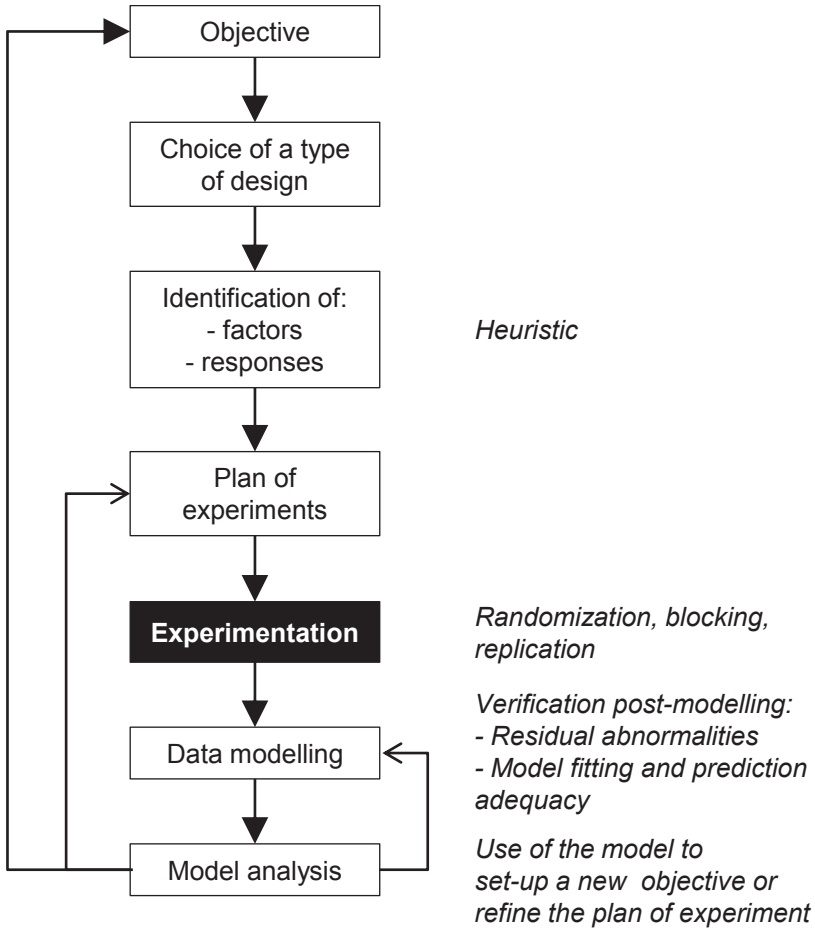


Figure 2.1: DOE-RSM approach to experimentation. RSM is a sequential methodology aims at modelling experimental data to estimate with confidence a process response on a continuous scale, as well as identifying a set of process parameters that optimise the response while minimizing the amount of experiments required to do so.

of models can be generated: 64 for a 3 factors Central Composite Design (CCD) or a Box-Behnken Design (BBD), 1024 for a 4 factors CCD, and this by limiting the construction to linear and quadratic models and ignoring three factors and higher order factors interactions. Selecting the right model trough all these possibilities is a challenging and time consuming task. Algorithms are frequently used in economy and medicine to automatically construct empirical regression model. However, as described by Babyak, they tend to produce overfitted models leading to spurious conclusions (126). They have also not yet been integrated with DOE methodology. Moreover, in an industry where development times and costs need to be reduced, it becomes critical to select, then work, with the most accurate model. In this context,

the development of statistical tools for the automation of model selection would be of real value.

In this study we describe the development of an algorithm in Mathematica software for the automated selection then analysis of the most accurate model(s) for a DOE-RSM (BBD, CCD, CCF or CCI with 2,3 or 4 factors) generated set of data. The algorithm integrates several independent statistical modules to test all the potential models (instead of constructing it step by step) and gradually rules out the inappropriate ones.

2.2 Theory

2.2.1 Data modelling

RSM is a specific DOE methodology that produces visual representation of a process response based on an empirical model. All empirical models present the same properties (127):

- The process response, usually noted y
- The mathematical function noted $f(\vec{x};\vec{\beta})$
- The error ϵ, $\epsilon \in \mathbb{R}$

The mathematical function regroups two coordinate vectors. The dimension of the vector $\vec{x}$ is equal to the number of *model variables* (sometimes called regressors), and varies from one model to another. For example the list of variables of a three factors full quadratic model will contain the process factors (x_1, x_2, x_3) , but also the full list of factors interaction $(x_1.x_2, x_1.x_3, x_2.x_3)$ and quadratic terms (x_1^2, x_2^2, x_3^2). The coordinates of the vector $\vec{\beta}$ are the regression coefficients or equation *parameters*. The term ϵ stands for an observation *error i.e.* the sources of variability not accounted for in the model, such as measurement error on the response, background noise, or the effect of other factors than the one studied. The general form of the model is generally written:

$$y = f(\vec{x};\vec{\beta}) + \epsilon \tag{2.1}$$

In DOE, the model is a multiple linear regression model. Low order polynomial models, such as linear models, are adapted to approximate process responses in a relatively small design space, or when a curvature of response is not expected across the design space. In other cases, second order polynomial approximations must be used. Third or higher order polynomials are, however, generally not needed if the experimental plan is appropriately chosen at the start (128), *i.e.* not too wide. Moreover, the relatively low number of experiments of the two most commonly used RSM designs, CCD and BBD, do not allow for third-order response approximations

without aliasing. For a series of k process variables $(x_1, x_2, ...x_k)$ the model equation can be written:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_i x_i^2 + \sum_{i<j} \sum_{j=2}^{k} \beta_{ij} x_i x_j + \epsilon \tag{2.2}$$

Generally the structure of the relationship between the process response and the model variables is unknown. In other words, the exact model equation is unknown. The first step in RSM is, therefore, to find a suitable approximation of the true relationship. This is done by estimating the coordinates of the vector $\vec{\hat{\beta}}$ whose coordinates $(\hat{\beta}_1, \hat{\beta}_2, ...\hat{\beta}_k)$ are the estimates of $(\beta_1, \beta_2, ...\beta_k)$, and choosing the right vector $\vec{x}$. The estimated regression equation, or *fitted model*, taking the form:

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i + \sum_{i=1}^{k} \hat{\beta}_i x_i^2 + \sum_{i<j} \sum_{j=2}^{k} \hat{\beta}_{ij} x_i x_j \tag{2.3}$$

The difference between the observation y_i and the model fitted value $\hat{y}_i$ is the *residual* $e_i = y_i - \hat{y}_i$. In other words the residuals are the estimation of the errors. The method of least squares regression used for the construction of linear regression models consists in finding the coordinates of the vector $\vec{\hat{\beta}}$ that minimises the sum of squares of the residuals:

$$SS_E = \sum_{i=1}^{k} {e_i}^2 \tag{2.4}$$

The fitted model can be used to graphically map the predicted response across the design space as well as identifying the values of the process factors that maximise or minimise the response. However, before proceeding, a critical step in DOE is to validate the model.

2.2.2 Model validation

Once the model has been constructed it is critical to verify that (i) the model accurately fits experimental data, and (ii) is an accurate representation of the studied process response across the design space.

2.2.2.1 Quantifying the model ability to fit experimental data

With the computational power of today's available statistical packages, constructing complex model is relatively easy. However, complex does not mean appropriate. An overly complex model will describe noise instead of the actual relationship between process parameters and process response. This phenomenon, called *overfitting*, results in models that overestimate the variations due to experimental error, in other words, explain error that exists in sample data but do not really exists in the population, and hence will not replicate. The number of model variables is too high in overfitted models. In other words, overfitted models include one or several variables that do not contribute to explain true variation in a set of data. Several statistical tools can be used to screen for overfitted models. The first one is called analysis of variance (ANOVA). ANOVA consists of simultaneous hypothesis tests to determine if any of the effects attributed to the model variables is significant or not. ANOVA relies on computing several statistics:

- Sum of Squares (SS): sum of all the squared effects
- Degrees of freedom (df): number of free units of information
- Mean square (MS): SS divided by df. Computed only for the model and the model variables.
- F-statistic: MS divided by MS of pure error.

These statistics were calculated for the model, each model variable, the model residual and the pure error from experiment replicates. In ANOVA, the difference is made between the Model Sum of Squares and the Error Sum of Squares. The Model Sum of Squares represents the variability explained by independent variables. The Error Sum of Squares represents the variability not explained by the Model. These values are used to calculate the F-statistic. The F-statistic allows for the estimation of the significance, generally at 95%, of the model and the model variables. As a general rule, Type III ANOVA are used in RSM because compared to Type I/II ANOVA, it provides estimates that are not a function of the number of observations in a group.

Another statistical approach to assess if a model is overfitted consists in calculating then comparing two statistics: the R^2 and the $R^2_{adjusted}$. The R^2 measures the amount of variation around the mean explained by the model and can be defined as:

$$R^2 = 1 - \frac{SS_{residuals}}{SS_{residuals} + SS_{model}} \tag{2.5}$$

Usually, the better the regression model, the higher the R^2. However, R^2 counteracts the tendency for overfitting data when doing regression. However, as more variables come into the model, the R^2 cannot decrease, whether or not the effect of added variables is significant. As a result, the R^2 cannot be used to discriminate overfitted models. The $R^2_{adjusted}$ statistic, however, takes into account the number of variables in the model and will decrease with the addition of model variables that do not add value to the model.

$$R^2_{adjusted} = 1 - \frac{SS_{residuals}}{SS_{residuals} + SS_{model}} \times \frac{df_{residual}}{df_{residual} + df_{model}} \tag{2.6}$$

As a result, the $R^2_{adjusted}$ is always inferior or equal to the R^2. If $R^2_{adjusted}$ falls far below R^2, there is a good chance to have non significant terms in the model. This difference can be used to rule out the models presenting too much overfitting.

2.2.2.2 Quantifying the accuracy of model predictions

The purpose of data modeling is not only to accurately fit experimental data, but also use the constructed model to make predictions within the range of the experimental combinations tested. It is therefore critical to validate the ability of a model to make accurate predictions. The $R^2_{predicted}$ measures the amount of variation in new data explained by the model. This statistic measures the ability of the model to accurately predict a response from a given factors combination. The computation of this statistics is based on the calculation of the predicted residual error sum of squares (PRESS) (129).

$$PRESS = \sum_{i=1}^{n} \left(\frac{e_i}{1 - h_{ii}} \right)^2 \tag{2.7}$$

where h_{ii} represents the position ii in the diagonal of the hat matrix.

$$R^2_{predicted} = 1 - \frac{PRESS}{SS_{residuals} + SS_{model}} \tag{2.8}$$

The closer the $R^2_{predicted}$ will be from the $R^2_{adjusted}$, the more accurate the model will be on average. The difference between those two statistics can be used to rule out the models with poor prediction power. As a rule of thumb, a difference inferior to 0.2 is considered acceptable.

Prediction interval also represents a valid solution to assess the ability of a model to make accurate predictions. A prediction interval is an estimate of an interval in which future observations will fall with a certain probability. Boundaries from the prediction interval can be calculated using:

$$\hat{y} \pm t_{1-\alpha/2,\nu}\hat{\sigma}_p \tag{2.9}$$

where $\hat{y}$ denotes the estimated value of the regression function, $t_{1-\alpha/2,\nu}$ is the $1-\alpha/2$ critical value from the student-t distribution with ν degrees of freedom and

$$\hat{\sigma}_p = \sqrt{\hat{\sigma}^2 + \hat{\sigma}^2_f} \tag{2.10}$$

where $\hat{\sigma}^2$ is the estimate of the residual standard deviation and $\hat{\sigma}^2_f$ the estimate of the standard deviation of the predicted response for the experimental data.

Available softwares usually provide the user with an interval of confidence around the optimum in response. If this confidence interval is useful, it would also be of importance to know the confidence intervals around the factors value leading to this optimum. To our knowledge, commercial softwares only identify the set of factors values leading to an optimum in response, but do not calculate their associated confidence intervals. A way to calculate those intervals is to use the fact that, theoretically, an infinity of models with same fitting and prediction power can be constructed by varying only the model parameters values. Those models, however, will predict different response optimum and factors combinations. By assuming that

each model parameter follow a normal distribution, each parameter value can take a different value within a given confidence interval. Moreover, model parameters are correlated. Indeed, in order to yield a model with the same fitting and prediction power a change of value of one or several model parameter implies a change in other model parameters to adjust the new model to similar model characteristics. Using a general mathematical software, it is possible to use a multinormal distribution to generate as many models with same fitting and prediction power as wanted. From this list of models it is therefore possible to calculate a standard deviation around the studied process factors optimum leading to an optimum in response. Due to their relative complexity, and the fact that most mathematical software include multinormal distributions function, the detail of the calculations will not be presented in this chapter, but can be found in the code presented in appendix.

2.2.2.3 Least square assumptions verification

The use of least square regression is conditioned by fundamental assumptions between the studied process parameters and the process response. As described in (130), "Knowledge and understanding of the situations when violations of assumptions lead to serious biases, and when they are of little consequence, are essential to meaningful data analysis". From a practical point of view it is more convenient to check model assumptions by diagnosing the residual's population for:

- normal distribution (i)
- independence of residuals (ii)
- constant variance (homoscedasticity) (iii)
- zero mean (iv)

Generally these assumptions are checked using *studentised residuals* rather than raw residuals. Indeed, it is not unusual to have great variation in residual standard deviation, especially when measured observations vary by several orders of magnitude. Studentised residuals are raw residuals divided by their estimated stan-

dard deviation. Because residuals do not all have the same variance, studentising is helpful when it comes to compare residuals (131).

(i) Regression assumes normal distribution of model residuals. Non-normality may affect the Fisher test of significance used in ANOVA. There are several ways to check for normality of a population. The Shapiro-Wilk test tests the null hypothesis that a population is normally distributed. It is proved to be more efficient than numerous other tests such as the Kolmogorov-Smirnov test (132, 133). The test is based on the calculation of a statistic W.

$$W = \frac{\left(\sum_{i=1}^{n} a_i (x_{(n-i+1)} - x_{(i)})\right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{2.11}$$

with $x_{(i)}$ being the i$^{\text{th}}$ sample value, n the sample size, $\bar{x}$ the sample mean and a_i are constants generated from the means, variances and covariances of the order statistics (available in tables). The W statistic is then compared to a p-value. If the p-value is inferior to the chosen alpha level (0.05 or 0.01) the null hypothesis is rejected (at 95 or 99% respectively) and the sample distribution is considered normal.

Non-normality of a population can also be visually detected using normal probability plots. As described by Filiben *et al.*, a normal probability plot can be generated by plotting the quantiles of the observed sample as a function corresponding to the normal N(0,1) order statistics medians (134). A linear function can also be fitted to the studentised residuals to help assess the normality of the distribution, the further the residuals being from the line, the greater the departures from normality (Figure 2.2). However, the Fisher significance test is "quite" robust to small departures from normality. Therefore the experimenter should not worry if a slight non linear points pattern is visible (116). In case of significant departures from normality, data can be transformed using a Log, Inverse, Inverse Log, Square root or inverse square root transformation which can in some cases normalise the residual distribution.

(ii) Residuals should be independent from the tested factors' levels. In other words, plots showing residual distribution versus factors levels should present a random pattern. Experimenters should especially look for conical or curved patterns which indicate that a high order term is missing from the model equation or that

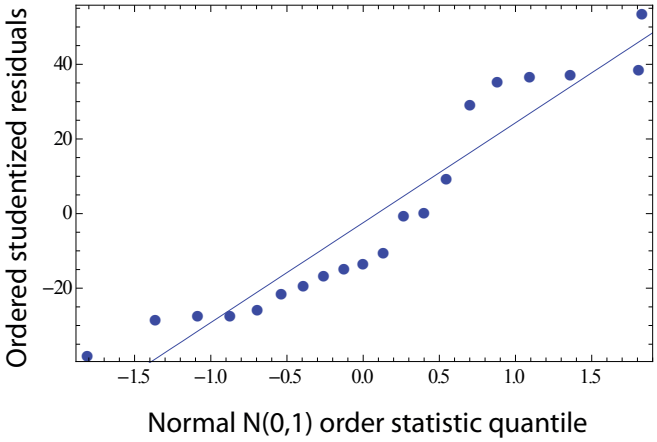


Figure 2.2: Normal probability plot. This type of representation is often used to visually assess the normality of the residual distribution. Several statistical tests are also available for this purpose like the Shapiro-Wilk test for example. In this example, points deviate from a straight line, so the normality assumption may not be satisfied.

data need a square root/inverse square root transformation (119).

(iii) and (iv) Homoscedasticity and residual zero mean can be checked at the same time. More importantly, by plotting the residuals versus the model fitted values, and looking at the distribution pattern, it is possible to assess which assumption is broken. The assumptions are validated if the residuals are distributed in a non-patterned fashion around 0. A curved pattern indicates that a quadratic term is missing from the model equation. A conical pattern (variation in residuals increasing or decreasing with an increase of fitted values) proves that the constant variance assumption is violated. Potential outliers, which usually have a much larger residual than the rest of the observations, can also be detected using this plot.

2.2.2.4 What to do if violation of assumptions occur? The Box-Cox transformation

In situations where one or several least square assumptions are violated, a power transformation can be applied to the data. The benefits from it depend entirely on the set of data but can include an improvement in the normality of the residual distribution and more stable variance. Power transformations are merely transformations that raise numbers to an exponent. Therefore, an infinity of different transformations can theoretically be applied to a set of data. However, the most

common includes adding a constant, or a conversion to square root $\left(y^{0.5}\right)$, inverse $\left(y^{-1}\right)$ or logarithmic $(log(y))$ scales. With the computational power available today, the difficulty does not lie in applying the transformation, but in identifying the transformation to apply. In 1964, Box and Cox solved this problem by publishing a method allowing to identify the power transformation λ that will minimise the residual sum of squares (135). For normal distribution of residuals, the Box-Cox distribution is written:

$$y^{\lambda} = \frac{y^{\lambda} - 1}{\lambda \times \bar{y}^{\lambda-1}} \quad \text{for } \lambda \neq 0$$

$$y^{\lambda} = \bar{y} ln(y) \quad \text{for } \lambda = 0 \tag{2.12}$$

where $\bar{y}$ is the geometric mean of the data with n observations:

$$\bar{y} = Exp\left(\frac{\sum_{i=1}^{n} ln(y_i)}{n}\right) \tag{2.13}$$

Myers *et al.* went further by including a confidence interval for λ that can be found by computing (136) :

$$SS^* = SS_R(\lambda)\left(1 + \frac{t^2_{\alpha/2,\nu}}{\nu}\right) \tag{2.14}$$

and by locating the two coordinates where SS^* intercepts the curve $SS_R(\lambda)$ (Figure 2.3). If the confidence interval contains 1, there is no need for transformation.

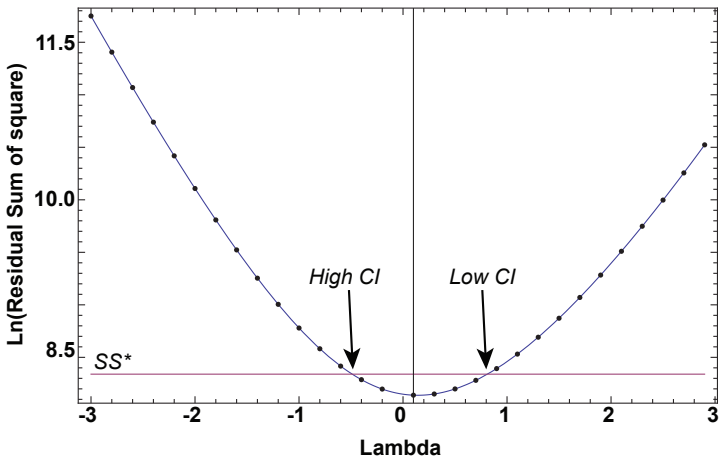


Figure 2.3: The Box-Cox procedure applied to a central composite design set of data In this example the optimal $\lambda = 0.12$ and the 95% confidence interval includes 0 and excludes 1. The use of a log transformation is therefore indicated.

2.3 Algorithm development

Because model selection and validation are relatively complex due to the multiple and interdependent statistics that need to be checked, it was decided to design a computational algorithm that will automatically perform these tasks for the experimenter. The algorithm was written in Mathematica (Wolfram Research, Champaign, IL) and is structured around a "mother file" that calls different "modules" in a given sequence to (i) generate all the possible models with respect to the type of DOE design and (ii) gradually rule out the models that do not satisfy a good data fitting and good prediction power. The first step of this work was to create a library of the necessary modules to write the algorithm. This library is presented in Table 2.1. The modules call for one or two inputs and generate an output. Under this format, the modules can easily be integrated in any order desired to provide the experimenter with a maximum of flexibility with respect to the development of an algorithm. Each module codes for a specific operation *i.e.* a Shapiro Wilk test, a Box-Cox transformation, the generation of a list of models from a set of data, *etc.*

The presence of a model variable in the model equation is conditioned by a test of significance, usually based on an ANOVA type III. ANOVA type III are generally used in DOE because of the treatment of unbalanced set of data and the potential significance of factors interactions. Algorithms for model construction have been developed in the past. They are based on a "forward" or "backward" procedure whether the model variables are sequentially added or removed from the model equation. This method is flawed as in type III ANOVA, the calculated results attributed to one model variable are correlated to the variation attributable to the effect after correcting for any other effects in the model. In other words, the significance of an effect of a model variable may depend on the presence or absence of other model variables in the model. Moreover, some developed algorithms took into consideration the overall model significance as only statistics to validate the model selection. However, with a type III ANOVA, the test for model significance can be positive while the effect of some model variables was not significant. This results is the selection of an overfitted model.

Table 2.1: Library of created modules.

Module name	Purpose	Input	Output
ModelsGenerator	Generate a list of all the regression models equation that can apply to a set of data, without aliased variables.	Data set	List of models equations
ModelsEqGenerator	Generates a list of all the regression models formal equation that can apply to a set of data, without aliased variables.	Data set	List of formal models equations
ShapiroWilkTest	Performs a Shapiro-Wilk test of normality and answer the question: is the residual distribution not normal?	Fitted model	"Yes" or "No"
BoxCoxTransformation	Performs a Box Cox transformation i.e. check if a data transformation is recommended.	Fitted model, Formal fitted model equation	A grid with the recommended power transformation, Lambda and lambda confidence interval limits.
DataTransformation	Generates a new data set applying a power transformation to the studied/observed data.	Data set, Power transformation	A new data set.
ModelSignificance	Test the model significance at 95% using an ANOVA.	Fitted Model	"Yes" or "No"
ModelOverfitting	Test if the model is overfitted	Fitted Model	"Overfitted" or "Not overfitted"
ModelStatistics	Calculates the model associated principal statistics: R-sq, Adj-R-sq, Pred-R-sq	Fitted model	A grod with the model associated principal statistics: R-sq, Adj-R-sq, Pred-R-sq
ModelOptimums	Generates a grid with the optimal predicted studied response as well as the factors leading to it and their associated confidence intervals	Formal model, Fitted model	A grid with the optimal predicted response, the optimal factors leading to it, and their associated confidence intervals.

ModelResAnalysis	Perform the model residual analysis (independence vs factors, linearity and homoscedasticity.	Fitted model	A table with the charts needed for the validation of the model residuals behavior.
ModelContourPlots	Generates contourplots associated with a model.	Fitted model	Generates contourplots associated with a model. The number of contourpolots depends on the nature of the design.

Each module performs simple operations and requires one or several inputs to generate one output. The code behind each module can be visualised in Appendix.

The algorithm developed in this study was based on the construction of all possible models with respect to any CCD (2,3,4 factors) and BBD generated set of data. The first step performed by the algorithm was to import the experimenter data from an excel file. The algorithm then constructs all the possible models: simple linear to full quadratic with all the intermediates. The algorithm then performed several loops in sequence aimed at testing a specific criteria, one model after the other. The first loop included a Shapiro-Wilk test that tested for the model's residual's normal distribution. If the distribution was estimated normal, the model was saved into the list "TableResNormalModels". If not, a Box-Cox- transformation was applied. If a power transformation other than 1 was indicated by the Box-Cox transformation, the data was transformed and a new model with the same variables was fitted to the data. Whether or not the distribution of the residuals of the new model was normal, the model was saved into the list "TableResNormalModels" (Figure 2.4).

For each model in "TableResNormalModels", a type III ANOVA was then performed. Significant models were saved into a newlist: "TableSignificantModels". For each model in this list, if the p-value associated to at least one model variable was superior to 0.1, the model was rejected. This step ensured to reject all the overfitted models. If the p-values associated to the model variables were inferior or equal to 0.1, the model was compiled into the list "GoodFittingModels". For each model of this list, if the difference between the $R^2_{adjusted}$ and $R^2_{predicted}$ was superior to 0.2, the model was compiled into the list "Potential Models". This step ensured to select only the models that presented a good prediction power (Figure 2.5).

Finally, for each model in the "PotentialModels" list, a full analysis was performed. Because, the validation of homoscedasticity and independence assumptions on residuals can only be checked visually, the algorithm provided the user with the necessary plots to select the most appropriate model. At this stage generally, only a few models (less than 10) are present in the list. Therefore, this visual check can be performed quickly.

At the different steps of the algorithm, if any list was found empty, the user was warned of the presence of one or several outliers in the data provided.

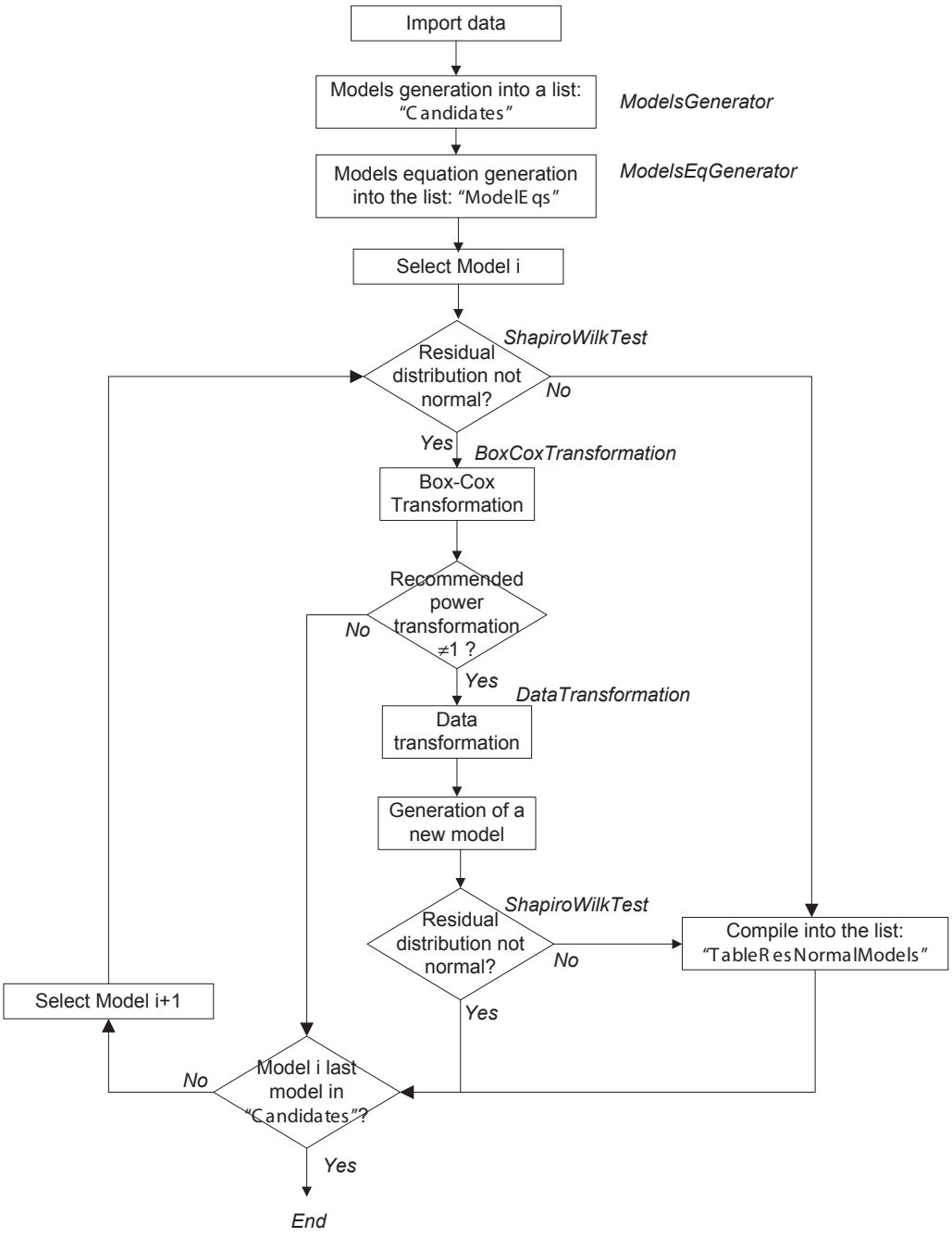
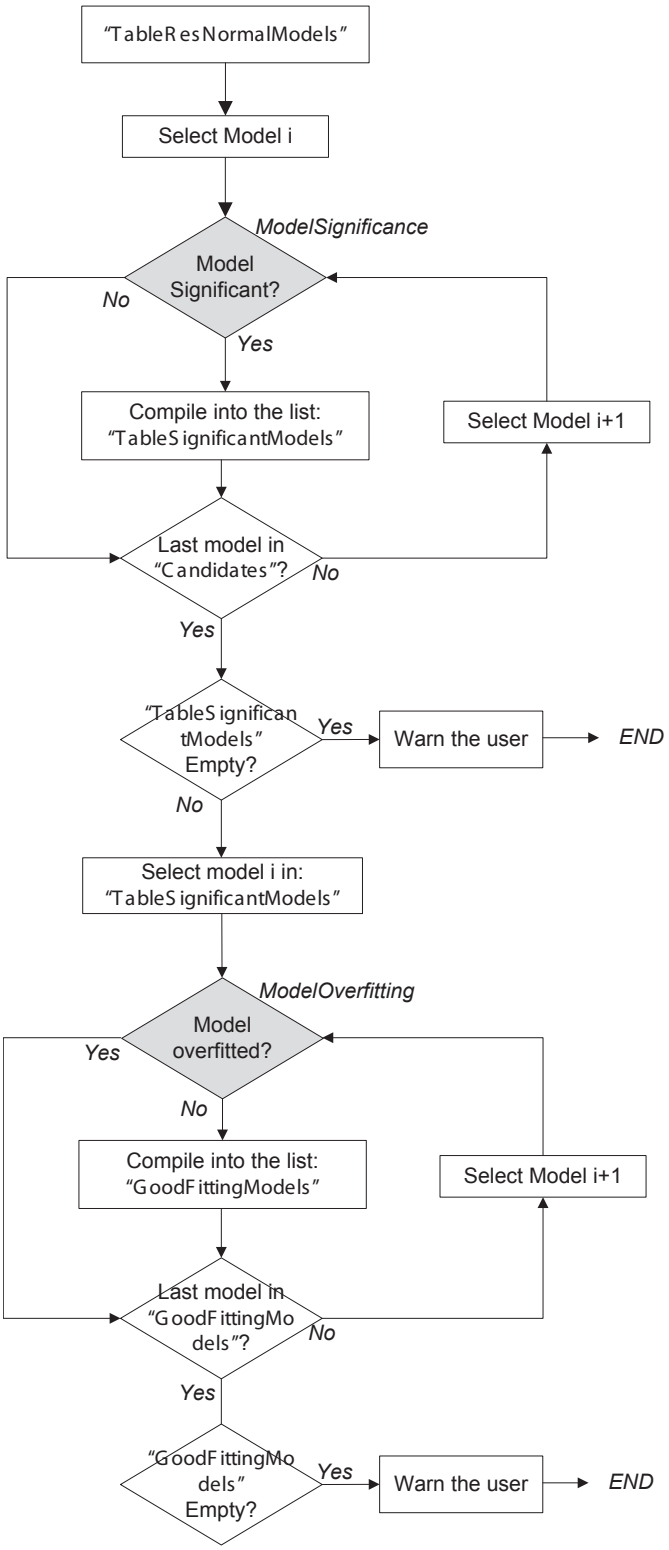


Figure 2.4: First algorithm loop: ruling out the models that present a non normal model's residuals distribution. Several modules were used in this algorithm loop. They are indicated in italic.


"TableResNormalModels"
Select Model i
ModelSignificance
Model Significant?
No
Yes
Compile into the list: "TableSignificantModels"
Select Model i+1
Last model in "Candidates"?
No
Yes
"TableSignificantModels" Empty?
Yes
Warn the user
END
No
Select model i in: "TableSignificantModels"
ModelOverfitting
Model overfitted?
Yes
No
Compile into the list: "GoodFittingModels"
Select Model i+1
Last model in "GoodFittingModels"?
No
Yes
"GoodFittingModels" Empty?
Yes
Warn the user
END

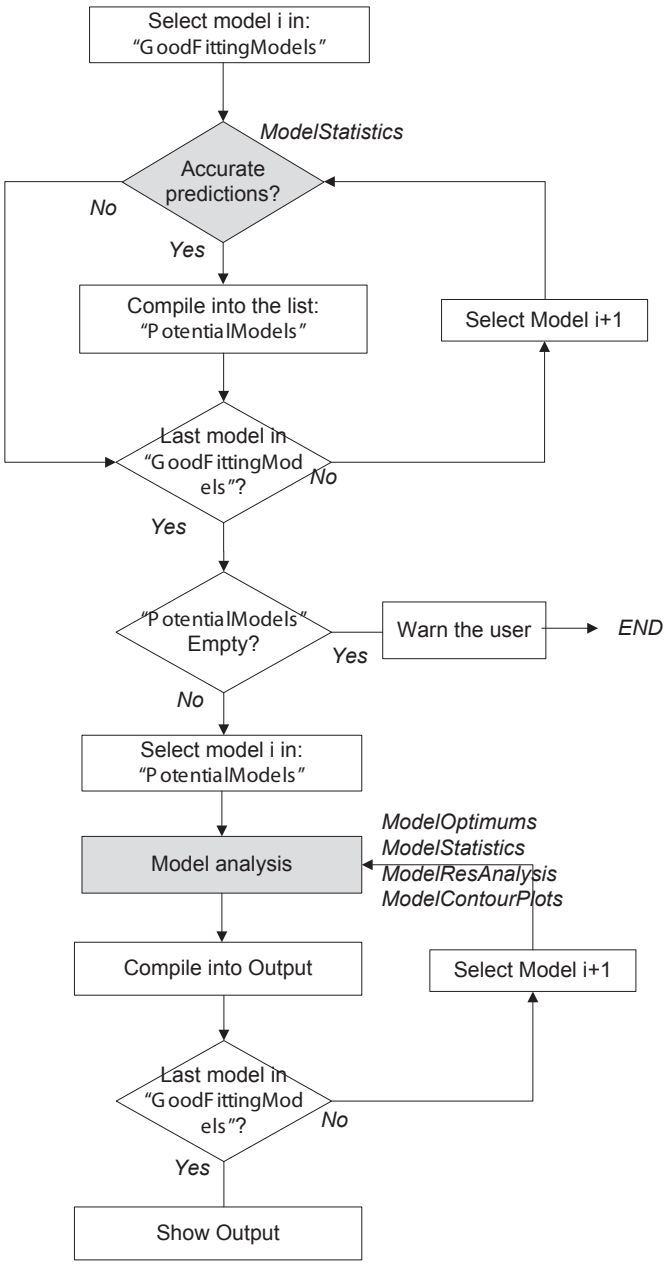


Figure 2.5: Algorithm for model selection. The modules used in the algorithm are indicated in italic on the figure.

2.4 A case-study: the optimisation of a PEI mediated transfection process in CHO cells

The algorithm was validated using twenty four DOE generated set of data including BBD (six designs), two, three and four factors CCD (three, twelve and three designs tested respectively). Data sets were found in books or generated from experiments in house (116, 121, 136). The algorithm output was compared to the ouputs of other statistical packages including Design-Expert® 7.0 (Stat-ease, Minneapolis, MN), Minitab® 16 (MinitabState, State College, PA) and JMP® (SAS Institute Inc., Cary, NC). A case study is presented below, in which a PEI-mediated transfection process was optimised using a three factors CCF, then a BBD.

2.4.1 Experimental method

The aim of the experiment was to optimise a PEI-mediated transfection process in CHO cells. More specifically, identify a combination of PEI, DNA and cell concentrations at transfection that would maximise protein titres 5 days post transfection. CHO-S cells were routinely cultured in vented 250mL shake flasks in a 5% CO_2 orbital shaker (125rpm, OT of 25mm) in CD-CHO (Life Technologies, Paisley, UK) supplemented with 8mM of glutamine (Lonza, Slough, UK) and passaged when the the culture reached mid exponential phase. Prior to transfection, cells were diluted to the desired concentrations in 50mL CultiFlasks (Sartorius AG, Goettingen, Germany) in CD-CHO supplemented with 8mM of glutamine in a total volume of 5mL. The reporter plasmid used in this study coded for the secreted alkaline phosphatase protein (SEAP). The plasmid construction details can be found in (73). PEI chemical (25 kDa linear, Polysciences, Warrington, USA) was dissolved into water at 1mg mL^{-1} and stored at -20°C. On the day of transfection, in separate micro-centrifuge tubes, the desired quantities of PEI and DNA were diluted with an equal volume of NaCl 300mM. PEI and DNA were then mixed together and allowed to complex for 1min in a total volume of 333μL made up with 150mM NaCl. The PEI/DNA solution was then added to the cells. CultiFlasks were immediately orbitally shaken

by hand and directly placed in an incubator at 37°C in 5% CO_2 with orbital shaking at 170rpm and an orbital throw of 50mm. Cultures were harvested after 5 days. Cell concentration and viability were assessed using trypan blue exclusion method on a Vi-Cell (Beckman-Coulter, High Wycombe, UK). Samples were taken and stored at -20°C for SEAP quantification using the SensoLyte® pNPP Secreted Alkaline Phosphatase colorimetric reporter gene assay kit (Cambridge Biosciences, Cambridge, UK) according to manufacturers instructions.

First experiments were performed following a three factors CCF design. PEI concentrations were varied from 7.5 to 20μg mL^{-1}, DNA from 1 to 10μg mL^{-1} and cells from 0.50E+06 to 2.50E+06 cells mL^{-1}. Then the process was re-optimised using a BBD design for which PEI concentrations ranging from 11 to 19μg mL^{-1}, DNA from 7 to 11μg mL^{-1} and cell counts from 1.75E+06 to 2.75E+06cells mL^{-1}.

2.4.2 Algorithm performance for the CCF

The algorithm output consisted of a list of two models. Out of the 64 models initially generated, nine did not meet the requirements of normal model's residuals distribution. Out of the 55 remaining, nine were found not significant at 95%. Among the 46 remaining a vast majority (40) were overfitted. In the end, four of the good fitting models were not making accurate predictions and were discarded and the algorithm output consisted of a list of two models. For both models, the data was transformed using a 0.5 power transformation. The models did not present any violations of residual homoscedasticity or independence. The predicted optimal concentrations of DNA, PEI and cell at transfection were very similar for both models. However one model presented better statistics in terms of fitting power (higher $R^2_{adjusted}$ and low difference between R^2 and $R^2_{adjusted}$) as well as prediction power (higher $R^2_{predicted}$ and low difference between $R^2_{adjusted}$ and $R^2_{predicted}$) (Figure 2.6). Model 2 was therefore selected and used to identify the combination of DNA, PEI and cell concentration for which an optimum in reporter protein yield could be achieved.

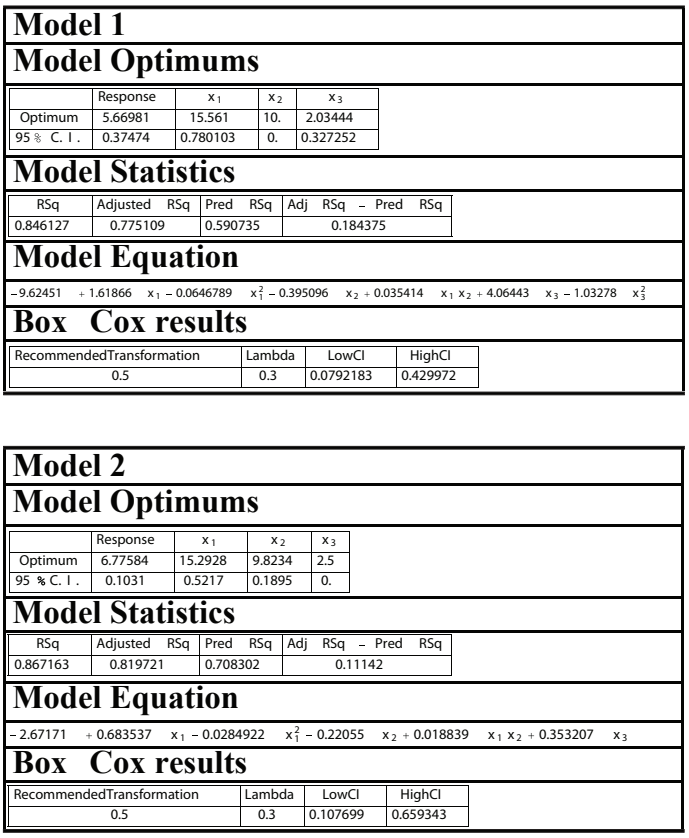
In the available statistical softwares, the experimenter can generally enter experimental data and automatically generate a model constructed with a list of model

variables of their choice (linear, linear with interactions, quadratic, *etc*). Using the same list of model variables of the model selected by the algorithm, models were constructed in Design Expert, Minitab and JMP, and analyzed. Model statistics were then compared (Table 2.2). The model fitting and prediction statistics (R^2, $R^2_{adjusted}$ and $R^2_{predicted}$) were similar across the different platforms. However, the algorithm is the only one to integrate the use of a Shapiro-Wilk test to check the normality of the model's residuals distribution. Indeed, a Shapiro-Wilk test can be performed using Minitab and JMP but requires the experimenter to first create the model, then isolate the residuals as a list and independently perform the test in a new data table. Shapiro-Wilk statistics were calculated in both Minitab, JMP and GraphPad Prism® (GraphPad, San Diego, CA) and shown to be equal to the one calculated by the algorithm (data not shown), proving that the test was executed correctly by the "ShapiroWilk" module.

All programs are able to perform a Box-Cox transformation. However, slight differences in the lambda values, as well as confidence intervals, could be observed. As Design Expert and the algorithm use the same calculations and Box-Cox transformation, the differences between the algorithm and Design Expert were due to calculation approximations within both programs (137). The differences observed with Minitab and JMP are probably due to a different Box-Cox transformation used.

2.4.3 Algorithm performances for the BBD

The algorithm output consisted of a list with 6 models. Out of the 64 models initially generated, 14 did not meet the requirements of normal model's residuals distribution. The total 50 remaining were significant at 95% according to the ANOVA tests. However, 43 of them were overfitted. Out of the seven remaining, only one was found not acceptable with a $R^2_{adjusted} - R^2_{predicted}$ equal to 0.2136. None of the six remaining models ("potential models") were constructed from a transformed set of data, mainly because transforming the data did not lead to significantly lower $SS_{residuals}$. Out of the six models proposed four presented significant violation of residual assumptions of independence and/or homoscedasticity. Therefore, from this



Model 1

Model Optimums

	Response	x_1	x_2	x_3
Optimum	5.66981	15.561	10.	2.03444
95 % C. I .	0.37474	0.780103	0.	0.327252

Model Statistics

RSq	Adjusted RSq	Pred RSq	Adj RSq – Pred RSq
0.846127	0.775109	0.590735	0.184375

Model Equation

$-9.62451 + 1.61866\ x_1 - 0.0646789\ x_1^2 - 0.395096\ x_2 + 0.035414\ x_1 x_2 + 4.06443\ x_3 - 1.03278\ x_3^2$

Box Cox results

RecommendedTransformation	Lambda	LowCI	HighCI
0.5	0.3	0.0792183	0.429972

Model 2

Model Optimums

	Response	x_1	x_2	x_3
Optimum	6.77584	15.2928	9.8234	2.5
95 % C. I .	0.1031	0.5217	0.1895	0.

Model Statistics

RSq	Adjusted RSq	Pred RSq	Adj RSq – Pred RSq
0.867163	0.819721	0.708302	0.11142

Model Equation

$-2.67171 + 0.683537\ x_1 - 0.0284922\ x_1^2 - 0.22055\ x_2 + 0.018839\ x_1 x_2 + 0.353207\ x_3$

Box Cox results

RecommendedTransformation	Lambda	LowCI	HighCI
0.5	0.3	0.107699	0.659343

Figure 2.6: The algorithm output for the CCF design included two models. Out of the 64 models constructed, only two models were selected by the algorithm. The model 2 presented better data fitting (higher $R^2_{adjusted}$ and low difference between R^2 and $R^2_{adjusted}$) and ability to make accurate predictions (higher $R^2_{predicted}$ and low difference between $R^2_{adjusted}$ and $R^2_{predicted}$).

Table 2.2: The exactitude of the calculations performed by the algorithm, was validated by comparing the calculated model statistics between the algorithm, and other commercial statistical softwares.

	Design Expert	**Minitab**	**JMP**	**Algorithm**
R^2	0.8672	0.8672	0.8672	0.8672
Adjusted R^2	0.8197	0.8197	0.8197	0.8197
Predicted R^2	0.7083	0.7083	-	0.7083
Shapiro Wilk statistic	-	-*	-*	0.8769
BoxCox				
Box Cox LowCI	0.04	0.41	-	0.12
BoxCox HighCI	0.68	1.09	-	0.65
Lambda	0.36	0.72	0.6	0.3
RecommendedTransformation	0.5	0.5	-	0.5
Optimums				
Optimum predicted response	6.7454	7.1637	7.1639	6.7758
Response Confidence interval	0.1296	-	0.1520	0.1031
Optimum factor x1	14.9126	15.3153	15.3011	15.2928
Factor x1 confidence interval	-	-	-	0.5217
Optimum factor x2	9.6395	10.0000	10.0000	9.8234
Factor x2 confidence interval	-	-	-	0.1895
Optimum factor x3	2.3491	2.5000	2.5000	>2.5
Factor x3 confidence interval	-	-	-	-

For a CCD, using the same model variables of the model identified using the algorithm, new models were constructed in Design Expert, Minitab and JMP. Models statistics, as well as the combination of DNA, PEI and cell concentration leading to a maximum in response, were then compared. The model statistics across all the platforms are comparable. The developed algorithm however offers the advantage of computing confidence interval for the optimal factors value.

Table 2.3: Comparison of the calculated model statistics within different programs for the Box-Behnken design.

	Design Expert	**Minitab**	**JMP**	**Algorithm**
R^2	0.9753	0.9753	0.9753	0.9753
Adjusted R^2	0.9560	0.9560	0.9660	0.9560
Predicted R^2	0.8790	0.8790	-	0.8790
Shapiro Wilk statistic	-	-*	-*	0.9634
BoxCox				
Box Cox LowCI	0	-1.62	-	0.2984
BoxCox HighCI	3.49	2.92	-	3.2570
Lambda	1.85	0.43	1.8	1.9000
RecommendedTransformation	None	0.5	-	None
Optimums				
Optimum predicted response	7.3467	7.3469	7.3467	7.4172
Response Confidence interval	0.1509	-	0.3415	0.1324
Optimum factor x1	17.3506	17.3623	17.3609	17.6141
Factor x1 confidence interval	-	-	-	0.5014
Optimum factor x2	9.0039	9.0136	9.0064	9.0490
Factor x2 confidence interval	-	-	-	0.0439
Optimum factor x3	2.7500	2.75	2.7500	>2.75
Factor x3 confidence interval				

analysis, only two models only showed to be appropriate for use. The one presented in Figure 2.7 presented a lower difference between the $R^2_{adjusted}$ and the $R^2_{predicted}$ (0.076991 for the first one, 0.152971 for the second). A comparison of the calculated model statistics and factors optimal concentrations across the different statistical softwares and the algorithm did not reveal any significant differences (Table 2.3). Major differences were observed for the Box-Cox transformation associated calculations, most probably because of the different way the transformation is coded and applied within the softwares and the algorithm.

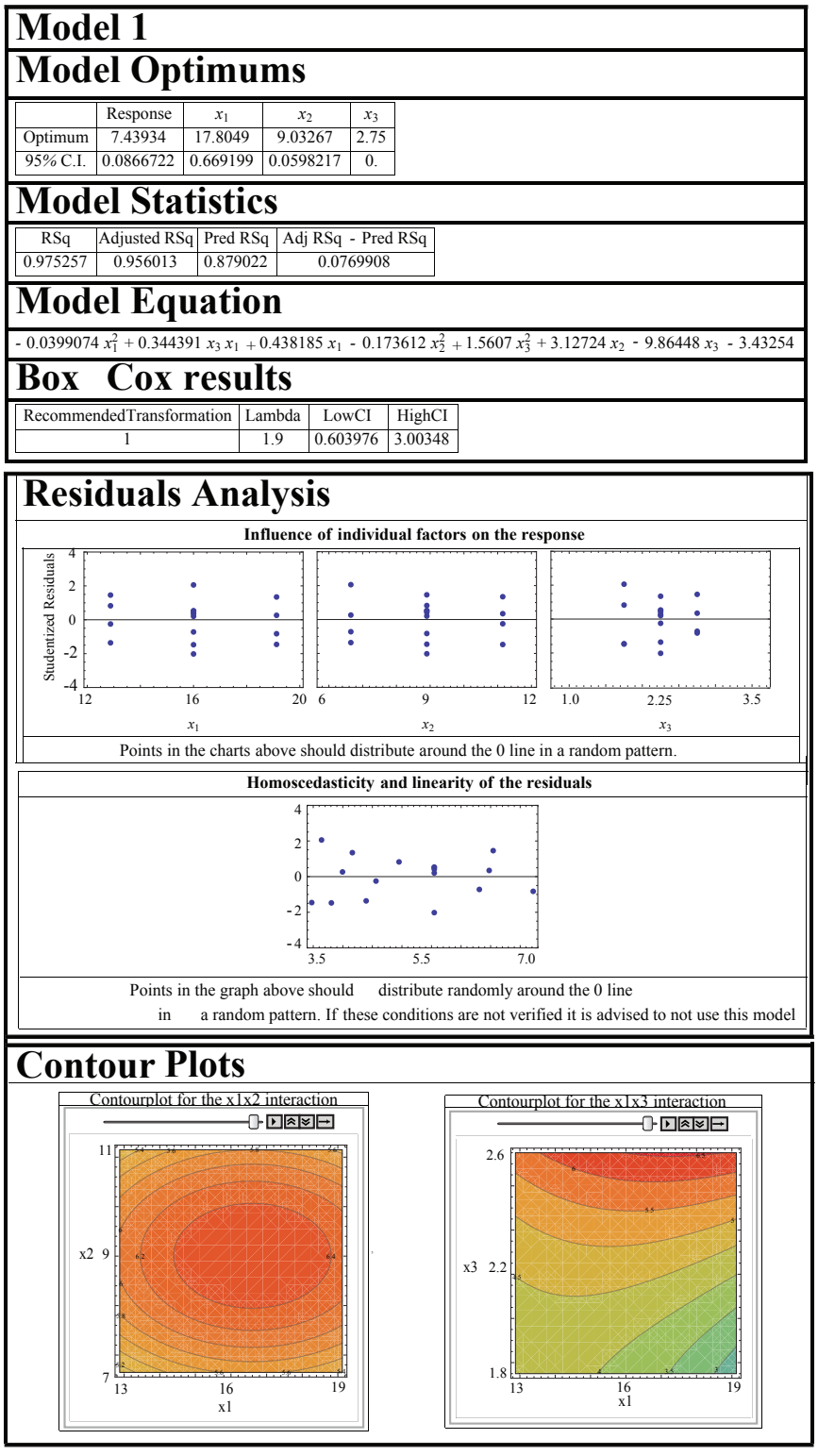


Figure 2.7: The algorithm output for the Box-Behnken design. The output provides the user with the analysed models that can be used. Here, only one model is shown. The output generates a report that includes the model optimums, equation and other fitting and prediction powers statistics, graphs to check for residual independence and homoscedasticity and two dynamic contour plots.

2.4.4 Summary

The algorithm and the statistical softwares were able to provide the user with a predicted maximum in response (protein titres) as well as the concentrations of PEI, DNA and cell leading to it (those optimal values are explicitly reported in Chapter 3). The optimal concentrations slightly varied with respect to the program used. However, the confidence intervals provided by the algorithm suggested that these differences were not significant and therefore most probably due to different combination of model parameters. Indeed, theoretically, an infinity of models with same fitting and prediction power can be constructed by varying only the model parameters values. The algorithm took advantage of this property to generate hundred of models and calculate the confidence intervals around the predicted optimal concentrations.

2.5 Discussion

The developed algorithm fills a gap left by commercial statistical softwares: the automated selection of an empirical model that both satisfies good fitting and prediction properties as well as the assumptions associated with linear regression (Table 2.4). The algorithm also proposes an accurate and more detailed model analysis with the calculation of confidence intervals around all predicted optimums (response and studied factors). Moreover, experimenters do not need to possess a broad knowledge of statistics to draw reliable conclusions from their set of data. Indeed, the only input required from the experimenter is an Excel file containing the experimental data. As a result, the use of algorithm considerably mitigates the risk of making mistakes during the model selection and analysis.

The main drawback is the need for the experimenter to possess the Mathematica software. The algorithm execution takes less than a minute for a three factors CCD and less than two on average for a four factors CCD. The algorithm presents another significant advantage over commercial statistical softwares: flexibility. Indeed, the algorithm is constructed in a modular way. Each module being independent they can be rearranged to create new algorithms. More importantly, recent surveys showed that DOE methodology implementation in industry suffers from the risks of misusing the methods, or available softwares. The risks of making mistakes during model selection and analysis is thought to be particularly high when using JMP or Minitab. Indeed, in these softwares the model selection and analysis fall to the experimenter. Moreover, the tools to assess the suitability of a model are either not present, or difficult to access without a strong statistical knowledge. JMP and Minitab main advantage remains in the user interface, made of icons on which the user can "point and click". Design Expert is thought to be more intuitive, providing the user with the tools and graphical representations to select an appropriate model. However, in this case as well, the model selection falls to the experimenter. This algorithm attempts to address these concerns as the automation of the model selection significantly mitigates those risks.

Table 2.4: Comparison of software and algorithm functionalities.

	Design Expert	**Minitab**	**JMP**	**Algorithm**
Purpose	Plan of experiment Model construction and analysis	Plan of experiment Model construction and analysis	Plan of experiment Model construction and analysis	Model construction, model selection and analysis.
Residual analysis				
Residual type	Native/ Studentized	Native/ Standardized	Native/ Studentized	Studentized
Normal distribution check	Half normal plot (visual)	Half normal plot (visual)/ Various normality tests	Half normal plot (visual)/ Various normality tests	Shapiro-Wilk test
Independence, Homoscedasticity, Linearity	Distribution pattern (visual)	Distribution pattern (visual)	Distribution pattern (visual)	Distribution pattern (visual)
Integrated Box Cox transformation	Yes	No	No	Yes
Model selection (Automated - Manual)	Manual	Manual	Manual	**Automated**
Risk of error in model selection (None-High-Very high)	High	Very high	Very high	**None**
Amount of knowledge required (None-Low-High)	Low	High	High	**None**
Flexibility (None-Low-High-Very high)	Low	High	High	**Very high** via the library of modules.

While the algorithm is limited to the data analysis only, the algorithm is fully automatic which prevents the source of error, check the assumptions relative to the construction of empirical model, and ensure that the empirical model selected is the most appropriate

Chapter 3

Multivariate optimisation of transient production processes using an integrated set of "Design Of Experiment" tools

This Chapter presents a holistic approach to the development of transient transfection processes. The aim was to validate that, using an integrated set of DOE tools, it was possible to characterise, then optimise a process in a more efficient manner than by using a traditional one-factor-at-a-time experimental approach.

3.1 Introduction

Transient gene expression (TGE) allows for the quick conversion of a recombinant gene into protein product, without the need of long and labour intensive screening and amplification methodologies that underpin the development of a stable cell line. Among the basal process variables involved in transient gene expression processes, the concentrations of DNA, DNA vehicle and cells at transfection are almost systematically subject to optimisation in the studies attempting to develop a TGE

process. However, a comparative analysis of the results published highlights that the optimal combination of these variables is case-specific (Table 1.2). This suggests that numerous other factors such as the transfection protocol, the cell line or the culture medium, may critically influence the process efficiency as well. Moreover, if an efficient transfection is a prerequisite to a high yielding process, the expression of the transgene also represents a major obstacle. Isolated studies report the introduction of isolated process design innovations such as mild-hypothermia which consists in lowering the temperature of cultivation, or the addition of histone deacetylase inhibitors such as valproic acid or sodium butyrate, could significantly enhance transgene expression by preventing the condensation of transgene DNA. In other words, despite its apparent simplicity, TGE is a complex process whose efficiency is underpinned by numerous interdependent factors. In this context, a rational approach such as the one presented in Figure 3.1 for a holistic development of a TGE process, combined with the use of design tools to reduce the complexity and time taken to explore and generate a viable production process would be of real value.

Under the acronym DOE lies a set of statistical and mathematical techniques that present a preferable alternative to the one-factor-at-a-time approach to experimentation. Firstly, DOE is a rational method that focuses on characterizing the process by identifying the major influencing process variables, then optimising these variables simultaneously to develop a high yielding process. In other words, DOE takes into account the potential effect of process variable's interactions on a process response. Secondly, DOE derived plan of experimentation are constructed in such a way that, combined with a thorough statistical analysis, meaningful information can be obtained in a minimum amount of time and experiments. In a recent study, a PEI-mediated TGE process was successfully optimised using a DOE approach (73). However this study limited itself to the optimisation of the basal variables involved in a PEI-mediated transfection process. Here we describe a rational and integrated method for the flexible, holistic and rapid development of a transient gene expression platform in CHO cells. Using this approach it was possible to enhance the titres of

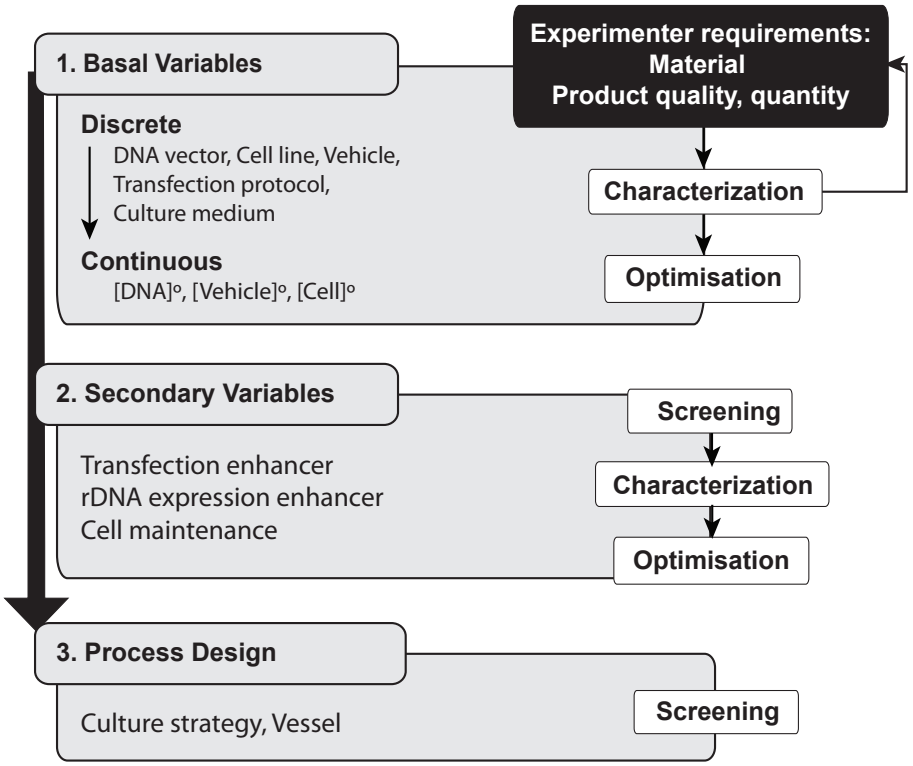


Figure 3.1: Rational and holistic development of a transient expression process. The variables underpinning a transient gene expression process can be classified in three categories. The basal known variables, without which the process does not exist, regroup both discrete and continuous variables. The nature of the combination of basal variables is chosen according to the experimenter needs. The second category regroups all the variables aimed at improving either the transfection or the protein expression efficiency. The effect of most of these variables is case specific and unknown prior to experimentation. As a result, a screening step is generally necessary to keep the vital few over the trivial many variables. The last category regroups the process design variables such as the mode of culture or the bioreactor type.

a recombinant mAb by more than 200 fold in 45 days.

3.2 Material and methods

Cell culture: CHO-S cells (Life Technologies, Paisley, UK) were routinely cultured in vented Erlenmeyer shake flasks (Corning, Surrey, UK) in CD-CHO medium (Life Technologies, Paisley, UK) supplemented with L-glutamine (Sigma-Aldrich, St. Louis, USA) at a concentration of 8mM, at 37°C in 5% (v/v) CO_2 with orbital shaking at 140 rpm. Cells were re-suspended in fresh medium every 3-4 days at a concentration of 2.00E+05 cells mL^{-1}. Cell concentration and viability were routinely measured by an automated Trypan Blue exclusion assay using a Vi-CELL® counter (Beckman-Coulter, High Wycombe, UK) according to manufacturer's instructions.

Reporter plasmids: Two different plasmids were used in this study. The first one, available in-house, was based on pSEAP2-Control (Clontech, Mountain View, CA) backbone. The SV40 enhancer of pSEAP2-Control was deleted by partial digestion with HpaI and BamHI. After blunting the ends with Klenow enzyme (Roche, Penzberg, Germany), the DNA was self-ligated. CMV promoter was amplified by PCR from pcDNA3.1/V5-His-TOPO/lacZ (Life Technologies) with primers: CMV-a, 5-GATCA GATC TCGA TGTA CGGG CCAG ATAT ACG-3 and CMV-bN, 5-GATC GAAT TCGA TCTG ACGG TTCA CTAA ACCA GCTC TGCT TATA TAGA CCTC CCAC-3 and cloned into the BglII and EcoRI sites of pSEAP2-Control. The sequence of the construct was finally verified by PCR (The synthesis of this vector had been performed by Andrew Tait). The second plasmid encoding the recombinant chimeric IgG4 mAb cB72.3 was provided by Lonza Biologics (Slough, UK). Plasmids DNA were purified using plasmid maxi purification kits (QIAGEN, Crawley, UK), re-suspended in a 10mM Tris-HCl buffer, pH 8.5 and stored at -20°C.

Transfection of CHO cells: CHO-S cells were cultured in shake flask to mid exponential phase prior to transfection. One hour prior to transfection, cells were diluted to the desired concentrations in 50mL CultiFlasks (Sartorius AG, Goettingen, Germany) in CD-CHO supplemented with 8mM of glutamine in a total volume of 5 mL. For PEI mediated transfections, in separate micro-centrifuge tubes, the

desired quantities of 25kD linear PEI (Polysciences, Warrington, USA) and DNA were diluted with an equal volume of 300mM NaCl. PEI and DNA were then mixed together and allow to complex for 1 min in a total volume of 333μL made up with 150mM NaCl. The PEI/DNA solution was then added to the cells. This procedure was noted *CF* in the text. When specified in the text, cell culture were spined down (200g, 5min), the conditioned media was discarded and cells resuspended in fresh media (CD-CHO supplemented with 8mM of glutamine, total volume of 5mL) to the desired cell concentration. Finally, DNA and PEI were directly added to the cells. This procedure was noted *DA* in the text. When specified in the text or the figures, cells post transfection were cultured in various culture media including ProCHO 4 and ProCHO 5 (Life Technologies). When specified in the text, cells were transfected with the FreeStyle^{TM} Reagent (Life Technologies). The procedure was carried out according to manufacturer's instructions. CultiFlasks were immediately orbitally shaken by hand and directly placed in an incubator at 37°C in 5% CO_2 with orbital shaking at 170rpm and an orbital throw of 50mm. Cultures were harvested after 5 days. Cell concentration and viability were assessed 5 days post transfection. Samples in separate micro-centrifuge tubes were stored at -20°C for reporter protein quantification. When specified in the text, the cultures were supplemented one hour post transfection with various chemicals: valproic acid, DMSO, LR3IGF and EDTA (all four chemicals were supplied by Sigma-Aldrich) and/or placed at 32°C in 5% CO_2 with orbital shaking at 170rpm and an orbital throw of 50mm. When specified in the text, cultures were fed at different days with FeedA and/or FeedB (Life Technologies).

Reporter protein quantification: SEAP was quantified using the SensoLyte pNPP Secreted Alkaline Phosphatase colorimetric reporter gene assay kit (Cambridge Biosciences, Cambridge, UK) according to manufacturer's instructions. IgG was quantified using the FastELYSA® Human IgG quantification Kit (R&D Biotech, Besancon, France).

3.3 Results

3.3.1 The effect of basal TGE variables on culture viability and IgG titres was response specific

Initially an augmented factorial design was constructed to study the effect of three factors, DNA, PEI and cell concentrations, on two process responses: cell culture viability and IgG titres, 5 days post transfection. This type of design is characterised by a relatively low number of experiments and the presence of a centre point (Figure 3.2). Each combination of factor levels is experimentally performed once, with the exception of the "centre point" located at the mid-level of each factor which is replicated four times. Replicating the centre point allows for an estimation of experimental error but also for a curvature test. This test investigates whether a linear model can accurately represent the process response. A significant curvature indicates that a linear model is not adapted, and that the design should be augmented to a CCD by the addition of other experiments before making any conclusions on the evolution of the response. However, if no curvature is detected, the linear model generated can be used to describe the evolution of the process response. The design space *i.e.* the ranges of factors to test were chosen following the method described by Thomson *et al.*: DNA concentration was varied from 1 to 10μg mL^{-1}, PEI from 1.5 to 20μg mL^{-1} and viable cell counts from 1.50E+06 to 2.50E+06 cells mL^{-1} (73). The extreme points of the concentration ranges were chosen as low and high factor levels to construct the design.For each response, a linear model was generated and analysed using the algorithm described in Chapter 2. A test for curvature was performed using Mathematica. Curvature was estimated significant when the p-value associated to the ratio of the mean square was associated to the centre point replicates, and the mean square of the model error was inferior or equal to 0.05.

With respect to culture viability response, the test for curvature came back negative at 95% confidence ($p - value = 0.0646$). The results from the centre point are represented in Figure 3.2, A and are located between the two lines, which is in accordance with the fact than an increase in PEI linearly affected cell culture

viability. However, using the model equation an intercept of 80.2 could be calculated when all the factors were at their mid-range level. Centre points results fell slightly below this value.

$$\begin{aligned} Ln(Cultureviability) = \\ 4.61545 \\ -3.45182E-003*[PEI] \\ -0074201*[DNA] \\ +2.29659E-003*[PEI]*[DNA] \end{aligned} \tag{3.1}$$

Therefore, a slight curvature in response may have been present, but did not significantly affect the model accuracy. As a result, the linear model presented here could be confidently used to analyse experimental culture viability data, and make predictions across the design space. The effect of cell concentration at transfection on culture viability was not found significant ($p-value = 0.12$). However, PEI and DNA concentrations, as well as their interaction, were found to significantly affect this response. High concentrations of PEI resulted in lower culture viability after 5 days. The effect of DNA, however, varied with respect to the concentration of PEI in culture. Indeed, at high PEI concentrations, variations in DNA concentrations did not influence the response. At low PEI concentrations, a high DNA concentration resulted in higher culture viabilities. This suggests that PEI was less cytotoxic when complexed with DNA than in its free form. The evolution of culture viability response with respect to DNA and PEI concentrations is represented in Figure 3.2, B. If a linear model was acceptable to describe the culture viability response, it clearly was not adapted to the IgG titre response. Indeed, IgG titre response exhibited a significant curvature ($p-value < 0.001$). It is therefore not surprising to observe the centre point results falling well above the predicted response lines in Figure 3.2, C.

While the design space construction method used by Thompson *et al.* (73) was valid when PEI was used as a DNA carrier, it did not hold true when it was adapted to the FreeStyle reagent. Indeed, the reagent was much less cytotoxic than PEI

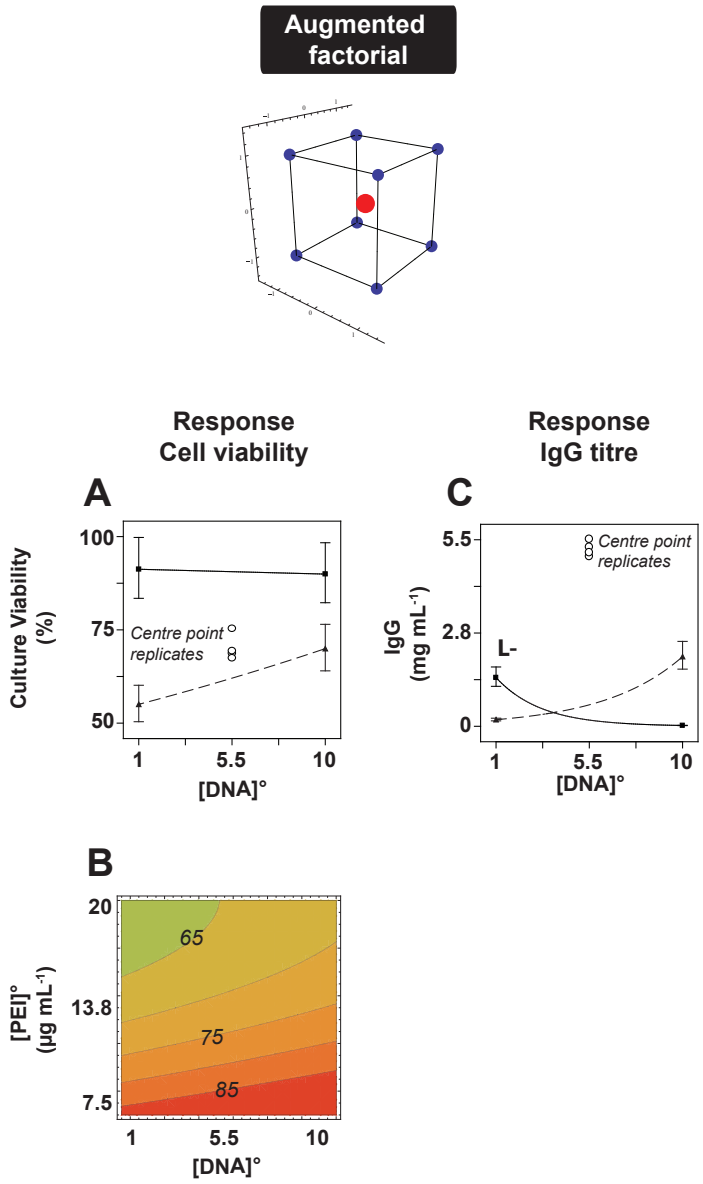


Figure 3.2: Augmented factorial design revealed curvature in IgG titre response but not in culture viability. This type of design is characterised by a relatively low number of experiment and the presence of a centre point. Each combination of factors levels is experimentally performed once, with the exception of the "centre point", located at the mid-level of each factor, which is replicated four times. The interaction plots display the mean in response (A: culture viability, B: IgG titres) for the different combination of the PEI and DNA factor levels tested. The centre point replicates can also be visualised on the plots. The presence of curvature in the response can be generally detected by the ex-centred location of the centre point replicates. When curvature is not detected in the response, contour plots can be used to estimate the evolution of the response across the design space (C), without requiring further experimentation.

with culture viability 5 days post transfection superior to 80% in all cases. In that case, despite a highly significant and good fitting linear model, the initial operating factors range tested was not adapted as a maximum in response was located outside the design space. It was however possible to "move" gradually towards the response maximum by adjusting the design space.

3.3.2 Augmenting the factorial design to a central composite design led to the identification of a combination of basal continuous factors that maximised protein titres with minimal experimental effort.

The previous augmented factorial design was augmented to a face-centred composite design (CCF) by adding axial points. The new axial combinations of factors were tested once, while the centre point was replicated twice. Compared to the previous augmented factorial design, the increased number of experiments allowed for the estimation of quadratic factors, and therefore, accurate estimation of the curvature in the response. Data in Figure 3.3, shows that the quadratic model identified was accurate at modelling the response curvature. Indeed, the intercept of the generated mathematical model equation at the centre point equalled 4.9mg L^{-1}. This value was reasonably close from the experimental values collected for the centre points replicates (5.34, 5.44, 4.52, 5.52, 5.16, 5.22 mg L^{-1}). The interaction plot pointed out the significance of the effect of the interaction of DNA and PEI on the IgG titres. The fact that protein yields declined when both factors were set at the same level revealed an antagonistic interaction. In fact, a simple maximum in response prediction could be identified within the design space for a given combination of basal continuous factors (Figure 3.3,B). Cell concentration at transfection proved to significantly affect IgG titres. Indeed, in the range of concentrations tested, the higher the concentration, the higher were the titres. An optimum of 6.7 mg L^{-1} could be achieved which represent a 550% increase compared to the titres obtained using the protocol described in (77).

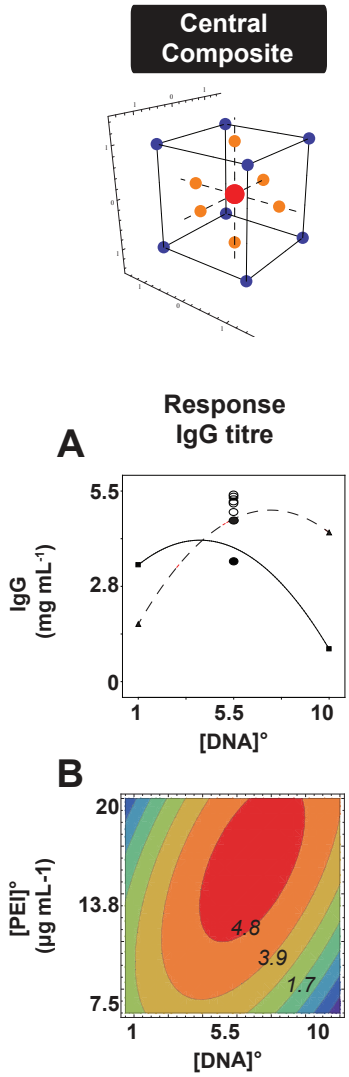


Figure 3.3: A central composite design led to the identification of a combination of basal continuous factors that maximised IgG titres. The previous augmented factorial design was augmented to a face-centred composite design (CCF) by adding axial points. The new axial combinations of factors were tested once, while the centre point was replicated twice. Compared to the previous augmented factorial design, the increased number of experiments allowed for the estimation of quadratic factors, and therefore, accurate estimation of the curvature in the response (A). As a result, the empirical model can now be used to graphically represent the evolution of the response across the design space.
Legend: ———: Low PEI concentration; – – –: High PEI concentration

3.3.3 The optimal combination of basal continuous variables depended on the early choice of discrete basal variables.

The experimental strategy described above was then reproduced for different combinations of basal discrete variables. The concentrations of basal continuous variables that maximised protein titres are reported in Table 3.1. For those conditions the culture viability at the time the cultures were stopped (5 days post transfection) was also reported. Results showed that the optimal concentration of the continuous factors greatly varied with respect to the discrete combination of factors chosen. Optimal combination of basal continuous factors differed with respect to the media utilised. While being similar for cultures in CD-CHO and ProCHO^{TM}4, optimal PEI concentration was significantly lower for cultures in ProCHO5. To reach an optimal protein titre, less DNA was required when ProCHO4 was used (5.7μg mL-1) compared to CD-CHO and ProCHO5 (9 and 9.9μg mL^{-1} respectively). Culture medium also proved to significantly affect maximal protein titres as only 5.13mg L^{-1} could be achieved using ProCHO4 whereas titres were 29% and 33% higher using CD-CHO and ProCHO5 respectively. At the conditions that maximised protein titres, culture viability after 5 days was higher when cells were transfected and cultivated in CD-CHO (71.4% after 5 days) while being the lowest in ProCHO4 (46.8%). Overall, transfecting and cultivating cells in CD-CHO proved to be most desirable option. Another transfection protocol was also investigated. Called "DA" for "direct addition protocol", this protocol was characterised by the fact that the plasmid DNA and DNA vehicle were added directly to a concentrated pool of cells. Prior to transfection, cells were spun down (200g, 10min) and re-suspended at a concentration of 20.00E+06 cells mL^{-1} in CD-CHO supplemented with 8mM of L-glutamine in CultiFlasks. Desired quantities of DNA, then DNA vehicle, were directly added to the cells and the CultiFlasks placed at 37°C in the incubator. One hour post-transfection, transfected cultures were diluted down to 5.00E+06 cells mL^{-1} with fresh CD-CHO supplemented with 8mM Glutamine and the cultures allowed to grow for 5 days. Transfecting cells using the DA protocol resulted in 57% higher titre than using the CF protocol. However cell viability was significantly lower reach-

ing 33.4% after 5 days in culture despite a lower optimal PEI concentration (12.4 rather than 17.54 μg mL^{-1}). The process also seemed more robust to variation in concentrations of basal continuous factors as shown by the relatively high 95% confidence interval around predictions. Using FreeStyle reagent resulted in lower protein titres than with PEI at the time the cultures were stopped. However culture viabilities were higher suggesting that cultures could be maintained for a longer period of time and potentially lead to better yields. Moreover, optimal DNA concentration at transfection was three times lower: around 3μg mL^{-1} using FreeStyle reagent rather than 9μg mL^{-1} using PEI. With respect to these results, it appears that transient production process performances are intrinsically linked to the combination of the basal discrete variables chosen. Introducing or changing a basal discrete variable justifies a complete process re-optimisation.

Table 3.1: Basal discrete and continuous factors combinations that maximised reporter protein titres.

Medium	*Carrier*	*Reporter protein*	*Protocol*	Protein (mg L^{-1})	Culture Viability (%)	Carrier (mg L^{-1})	Cell (10^{-6} mL^{-1})	DNA (mg L^{-1})
CD-CHO	*PEI*	*SEAP*	*CF*	7.26±0.68	71.4±3.6	17.54±0.82	2.5	9.01±1.06
ProCHO4	*PEI*	*SEAP*	*CF*	5.13±0.21	46.8±1.9	17.22±2.03	2.26±0.43	5.7±0.98
ProCHO5	*PEI*	*SEAP*	*CF*	7.62±0.54	62.1±3.1	12.09±0.36	2.1±0.49	9.94±0.6
CD-CHO	*PEI*	*SEAP*	*DA*	12.68±0.73	33.4±3.8	12.4±1.75	20±1	7.96±1.72
CD-CHO	*PEI*	*IgG*	*CF*	6.71±0.83	55.4±9.8	16.82±1.34	2.02±0.45	9.5±1.23
CD-CHO	*PEI*	*IgG*	*DA*	13.06±0.66	26±6.5	13.5±1.92	NA	8±1.55
CD-CHO	*FR*	*IgG*	*CF*	3.08±0.24	87.2±3.5	3.43±2	1±0.5	2.82±0.41
CD-CHO	*FR*	*IgG*	*DA*	6.30±0.46	84.5±2.1	4.59±1.15	1.5±0.7	3±2

CF: Complex formation pre transfection; DA: Direct addition; FR: FreeStyle reagent

3.3.4 Process performance could be further enhanced by controlling the culture environment post transfection

Despite resulting in higher protein yields, the DA transfection protocol induced a premature cell death and therefore significantly reduced the protein expression phase to a maximum of 5 days. Moreover the lysis of cells leads to the release of hydrolytic enzymes in the culture and can affect product quantity and quality as well as complicating the subsequent purification steps. By controlling the culture environment it should be possible to maintain the culture alive for longer, as well as protecting the DNA transcription machinery, to significantly increase protein titres. Data in literature showed that the addition of specific active factors in the medium such as growth factors or histone deacetylase inhibitors, as well as lowering the culture temperature (mild-hypothermia), could result in higher culture viability and protein titres (77, 80). Using a 2^{5-1} fractional factorial design, the effect of mildhypothermia (32°C culture temperature) as well as the addition of Valproic acid (VPA), LONG R3 Insulin-like growth factor I (LR3IGF), Dimethyl Sulfoxide (DMSO) and Ethylenediaminetetraacetic acid (EDTA) on the process performances was evaluated. Compared to a full 2^5 factorial design, only a half of the total possible factor combination was performed. This resulted in significantly shorter development time to the expense of aliasing. In other words, it was assumed that the effect of high order factors interactions (3,4,5 factors) were not significant. Transfections were performed in CD-CHO using the DA protocol. One hour post transfection, and following the plan of experience, different combinations of chemicals were added to the cultures, and for some of them, the incubation temperature was decreased to to 32°C. The titres of a reporter IgG, as well as the culture's viability were assessed 5 days post transfection. Results were analysed using half normal probability plots. These plots show the magnitude of the effect of the tested factors, and their interactions, ordered in an increasing magnitude. The Standardised Effect for a factor corresponds to the difference of the average process response over "high" factor levels, minus the average response over the "low" factor levels. The values on the y-axis are given by the idealised expected values for this number of effects, ranked

by increasing value. A significant effect is characterised graphically as an outlier with respect to the "line of chance" (Figure 3.4).

In Figure 3.4 A, the effect of factors D and E, EDTA and 37°C temperature respectively, fall in the top right hand corner, away from the "line of chance". Therefore, both these factors had a significant effect on cell viability. Individual factor plots revealed that EDTA was highly cytotoxic whereas mild-hypothermia resulted in higher cell viabilities compared to cultures placed at 37°C. Figure 3.4 C shows that factors C (DMSO) and D (EDTA), as well as numerous two factor interactions significantly affected IgG titres. While DMSO had a positive effect, the presence of EDTA in culture always resulted in lower IgG titres, most probably because EDTA proved to be cytotoxic, leading to a lower amount of viable producing cells, and lower product titres by the end of the culture. While factor A (VPA) was found to not have a significant effect alone (p-value of 0.13), it positively counteracted the negative effect of EDTA, suggesting that under stressed conditions VPA somehow enhanced transgene expression levels. In other words, it is probable that VPA could have a significant positive effect on transgene expression near the end of the culture. Overall, incubating the culture at 32°C instead of 37°C resulted in 32% higher culture viabilities 5 days post transfection. Interestingly, despite a higher proportion of living cells, mild hypothermia alone did not result in significantly higher IgG titres in culture. However, combined with DMSO the final IgG titres were 25% higher than when the cultures were incubated at 37°C. This suggest a synergistic positive effect of DMSO and mild-hypothermia on the culture performances.

To confirm the positive effect of VPA, DMSO on IgG titres and culture viability, two CHO cell cultures were transfected using the DA protocol. One hour post transfection, one culture was supplemented with VPA and DMSO at 0.75mM and 1% (v/v) respectively, while the other served as control. The culture vessels were placed at 32°C. IgG titres and culture viability were assessed every two days. The results presented in Figure 3.5 confirmed the benefit of supplementing the culture with VPA and DMSO. Indeed, the cells were viable for a longer period of time. At day 6 post transfection, the culture viability was maintained above 65%, while in

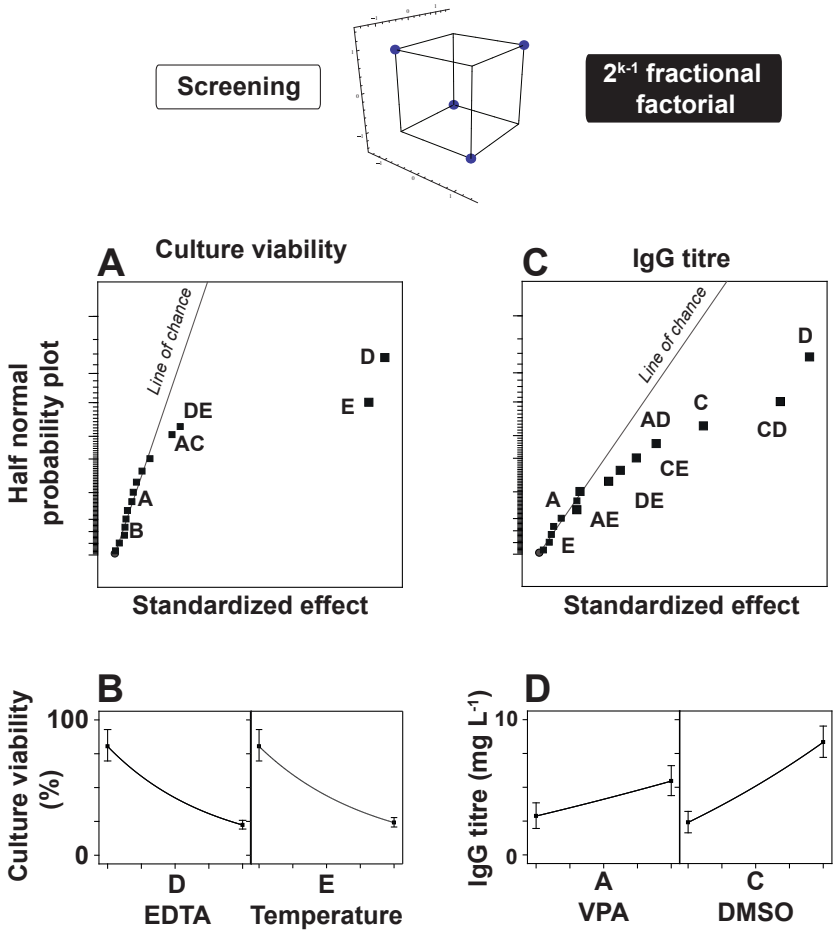


Figure 3.4: Identification of the variables that increased process performances using a fractional factorial design. Half normal probability plots show the magnitude of the effect of the tested factors, and their interactions, ordered in an increasing magnitude. The Standardised Effect for a factor corresponds to the difference of the average process response over "high" factor levels, minus the average response over the "low" factor levels. The values on the y-axis are given by the idealised, expected values for this number of effects, ranked by increasing value. Here, half normal probability plots were used to screen for the variables significantly affecting Culture viability (A) or IgG titres (C). The (B) and (D) figures show the evolution of the mean in culture viability and IgG titres respectively, with respect to a variation of one single variable.

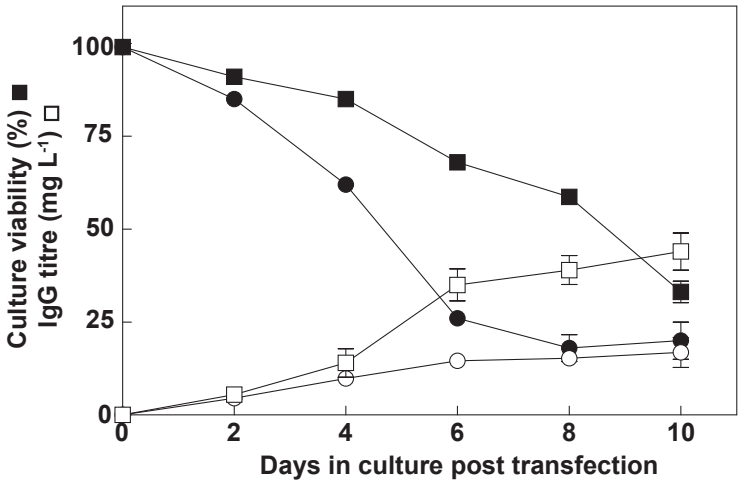


Figure 3.5: Identification of the variables that increased process performances using a fractional factorial design. CHO-S cultures were transfected at basal continuous variables optima using the direct addition protocol. Half of the cultures were supplemented with DMSO and VPA one hour post transfection and the culture vessels placed at 32°C (square symbols). The other half served as controls (circle symbols). Cell viability (closed symbols) and IgG titres (open symbols) were assessed every two days. Cultures were run in triplicates.

comparison, the viability of the control culture reached 25%.

In the presence of DMSO and VPA, protein titres reached 48mg L^{-1} 10 days post transfection, which represented an increase of 161% over the control. In the presence of chemical supplements, volumetric productivity gradually increased from day 0 to day 6 reaching a maximum of 10.5mg L^{-1} d^{-1} between day 4 and day 6 compared to 1.75L^{-1} d^{-1} for the control culture over the same period. More importantly, protein titres of 26mg L^{-1} after 5 days were higher than any of the titres associated with the experiments of the fractional factorial design, confirming the predicted beneficial synergistic effect of the addition of VPA and DMSO agents and mild hypothermia on volumetric productivity.

To further investigate the effect of concentrations of VPA and DMSO on IgG production, an augmented factorial design was performed. VPA concentration was varied from 0.5 to 10mM and DMSO concentration from 0.5 to 1.5%(v/v final culture). IgG titres were assessed 10 days post transfection. However, these factors did not significantly affect the IgG titres in culture (flat response surface, not shown). As a result, the effect of VPA and DMSO was concentration independent within the ranges tested.

3.3.5 Production phase could be increased further using a simple fed-batch strategy

The amount of recombinant protein achievable *in vitro* was limited by the ability of the cells to remain viable over an extended period of time. This was probably caused by the exhaustion of a key component in the culture medium and by the accumulation of toxic products that result in lower cell metabolic activity (138). Moreover, the high quantities of PEI required for transfection showed to promote cell death and subsequently shorten production phase duration. To overcome those issues a simple fed-batch strategy was developed. It was thought that the addition of medium in the culture would dilute toxic components while providing additional nutrients to sustain metabolic activity for longer. The effect of three different feed solutions (Feed A, Feed B and a 50/50 (v/v) mix of FeedA/Feed B) as well as 4 different feeding strategies (single addition at day 0 or day 5, or semi continuous addition starting at day 0 or day 2 post transfection) on culture viability and product titres was investigated. To do so, cultures of CHO-S cells were transfected using the DA protocol. One hour post transfection, cultures were supplemented with DMSO and VPA and incubated at 32°C. Feeding was started one hour post transfection following a general factorial plan of experiment (13 individual experiments replicated two times). A total of four feeding solutions were investigated. For half of the cultures, a unique addition of feed media (50% of the initial culture volume) was performed either immediately after transfection, or at day 5 post transfection. For the other half, the feeding was done semi-continuously, with four additions (12.5% of the initial culture volume each time) every two days. The semi continuous feeding was started either immediately after transfection or at day 2 post transfection. Culture viability and total cell concentration were measured 10 days after transfection by trypan blue exclusion, and samples from the culture were taken for IgG quantification. Results are presented in Figure 3.6.

The type of feed used (A, B or a mix of A and B) did not significantly affect process performance (Fisher-Snedecor test). Data for a 50/50 (v/v) mix of FeedA/Feed B are presented in Figure 3.6. In all cases feeding cultures enhanced process perfor-

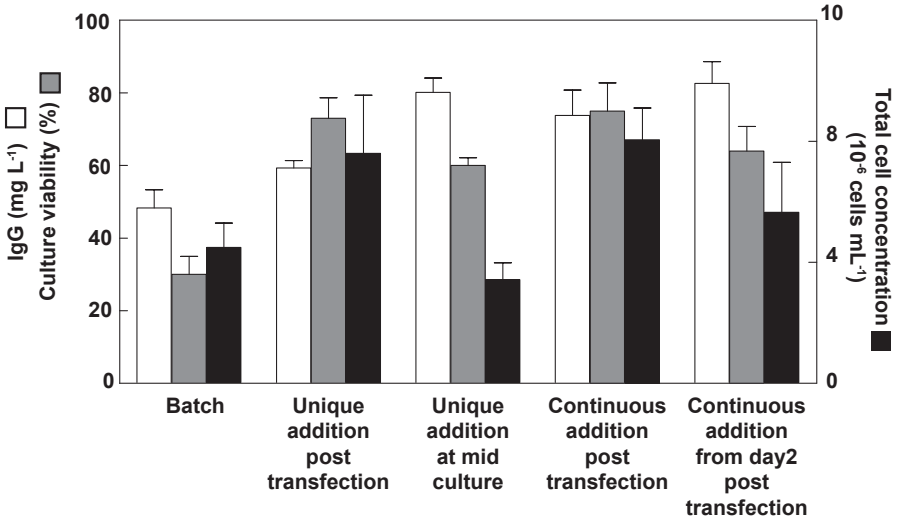


Figure 3.6: Screening of the effect of feed additions conditions on protein titres and culture viability 10 days post transfection. CHO-S cells were transfected using the DA protocol then supplemented with DMSO and VPA and placed at 32°C in 5mL volume. Cultures were subsequently fed with a 50/50 (v/v) mix of FeedA/Feed B. Four different feeding strategies were screened with a unique addition of 2.5mL at day 0 or day 5, or a semi-continuous fed-batch strategy with four additions of 0.625mL starting at day 0 or day 2. Cultures were run in duplicate.

mance. When fed, culture viabilities ranged from 59% (unique addition at day 5) to 72% (continuous addition from d0) after 10 days. This represented an improvement of 96.7% and 140% respectively over the batch culture that served as a control. Interestingly, culture viabilities were higher when cultures were supplemented immediately after transfection compared to later during the process, supporting previous observations that PEI cytotoxicity is concentration dependent (73), and that by diluting it early in the culture, higher culture viabilities could be obtained. Compared to the control, higher cell densities were obtained in fed cultures with the exception of the culture fed at day 5 post transfection. This result showed that when realised early, the supplementation post transfection also promoted cell growth. Compared to a batch process, a unique addition at mid-culture resulted in a doubling of titres by the end of the culture. As the final total cell densities are not significantly different, this effect can be solely attributed to the extended maintenance of living cells in culture and/or higher expression activity. A continuous addition starting at day 0 resulted in similar IgG titres, but also contributed to cell proliferation and lower cell specific productivity. However, the culture viability of 78% at day 10 post

transfection suggested that the culture could be maintained for a few more days and that higher titres were achievable.

3.4 Discussion

This data demonstrates the potential of an integrated DOE approach for the quick and cost-effective development of transient production processes in CHO cells. The basis of the approach is to only perform the experiments that will contribute to the understanding of the process. The fact that central composite designs can be constructed in a modular way (axial points added to an augmented factorial design), proved to be decisive for the quick characterisation, then optimisation of the process. The experimental design can therefore be constructed step by step, by increasing its complexity, and therefore the number of experiments to perform, with respect to the type of information required. Simple factor screening could be performed using fractional factorial design, while augmented factorial design was required to model the linear effect of factors on one or several process responses. Central composite designs allowed for data modelling where curvature was present in the response. (Figure 3.7). Using augmented factorial designs, it was possible to test early for the presence of curvature in the process response, and sequentially augment the design to a central composite design when necessary, which saved both time and number of experiments to perform. Using central composite designs, it was possible to identify a simple maximum in process response and the combinations of basal continuous factors that led to it. More importantly, by identification of ranges of confidence on basal factors, as well as combining process response contour plots, it is possible to assess how a variation in process factors will affect the process as a whole, and define a zone of control for guaranteed process performance (Figure 3.8). This is of critical importance in the development of production platforms to generate consistent material in terms of both quantity and quality. For example, it is known that the lysis of cells can affect protein titres and quality by the release of proteases and glycosidases as well as seriously complicating the purification of the product (139). Therefore, maintaining culture viability above a given threshold can be desirable in some applications. This approach to process development tends to become a standard in a "Quality by Design" and it is believed that it will be generalised in the future.

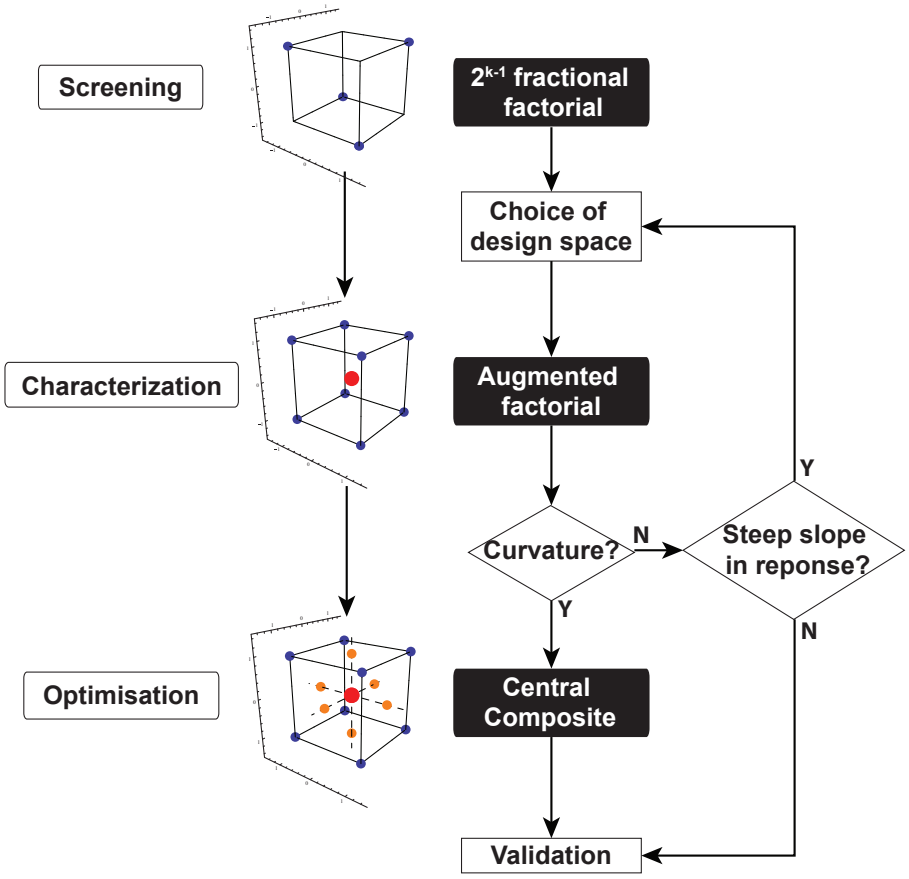


Figure 3.7: Integrated DOE approach to experimentation. The level of complexity of the experimental design is adjusted with respect to the purpose of the experiment. Fractional factorial designs are ideal for the screening of numerous factors but lack the ability to accurately model a process response. Augmented factorial designs are ideal for process characterisation: the presence of the centre point allows for the test of curvature in the response. Additional experiments allow for the linear modelling of process responses. However those designs are inadequate to estimate quadratic factors. By adding axial points, the design can be augmented to a composite design with which the curvature in response can be accurately modelled.

Using fractional factorial design it was possible to screen for the effect of 5 different factors, and their interactions, on protein titres and culture viability. By adding the highly cytotoxic EDTA chemical to the cultures, it was possible to generate a stressed culture environment, with a high proportion of cells dying rapidly. This enabled identification of the factors that beneficially impacted cell viability or protein titres after only 5 days in culture. Using this approach, the time dedicated to experimentation could be significantly reduced and it was possible to identify the factors that would not appear significant in non-stressed cultures.

By distributing the variables that impacted process performance into different

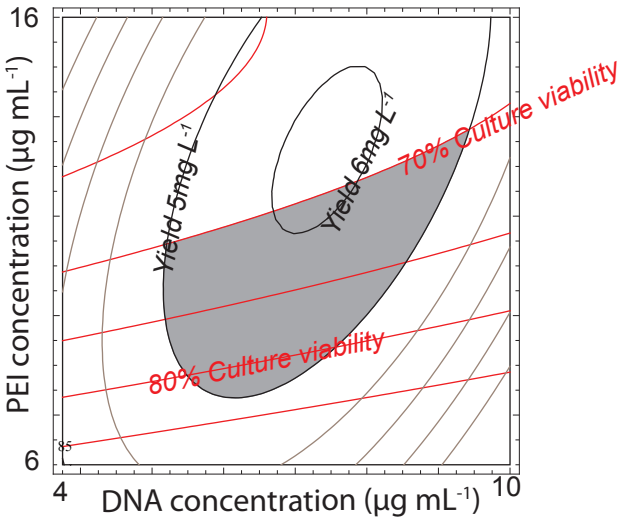


Figure 3.8: Operational design space for desirable process performance. Culture viability and IgG titres response contour plots were combined to identify a zone (represented by the grey area) for which IgG titres were above 5mg mL^{-1} and culture viability above 70%.

layers, it was possible to rationally and quickly develop a PEI-mediated transfection process. Compared to a non-optimised process, the optimisation of basal continuous variables resulted in a significantly higher yield (13.1mg L^{-1} of a reporter IgG 5 days in culture post transfection) to the expense of a drop in culture viability. The DA transfection protocol proved to promote cell death more quickly than the original protocol. Mechanical damage to the cells during the concentration step pre transfection, the culture medium exchange, as well as the transfection of a highly concentrated pool of cells with higher concentrations of PEI are reasons that may have contributed to this observation. It has been proved in the past that conditioned medium could prevent cell apoptosis and therefore result in higher culture viability (140). However, transfecting cells in conditioned medium proved to limit protein expression yields by somehow reducing transfection efficiency (141). In our protocol however, cells are re-suspended one hour post transfection. In this case, re-suspending the cells after transfection in a medium containing a defined percentage of conditioned medium could potentially limit the early death of cells post transfection, without inhibiting the transfection process itself.

Following the optimisation of the basal continuous factors it was decided to

further enhance protein titres by maintaining biomass in a productive state for longer. In other words, (i) maintain transcriptional and translational activity of the trangene and (ii) maintain cells viable. In this study, mild hypothermia, DMSO and VPA proved to synergistically enhance culture survivability and/or protein titres by 3-fold (Figure 3.9). Mild hypothermia alone correlated with higher culture viabili-

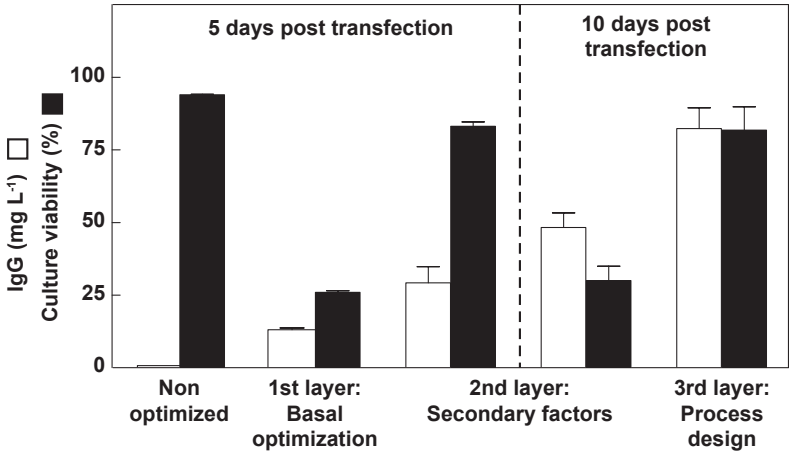


Figure 3.9: Evolution of process performances throughout the optimisation. The optimisation of the first layer consisted in identifying the optimal combination of basal continuous factors that maximised protein titres. Mild hypothermia coupled with the addition of VPA and DMSO to the culture medium, resulted in higher culture viabilities that allowed for the extension of the culture to 10 days. IgG titres and culture viability could be doubled by the end of the culture by adopting a simple fed-batch solution.

ties. The effect of mild hypothermia on CHO cell culture is today well documented (78, 142, 143). Reduced culture temperature would arrest cells in G_1 phase of the cell cycle as well as increasing the stability of cellular messenger RNAs. As a result, the lifespan of RNAs in the cell is extended, yielding to an extension of the cell lifetime but also higher cell productivity. In this study however, mild hypothermia induced lower specific cell productivity, probably by reducing protein synthetic rates. This result contrasts with previous published studies on transient expression, for which an increase in specific productivity (qP) was observed (77). The effect of mild hypothermia is therefore more than likely to be cell line specific and further optimisations with respect to the culture temperature should be conducted.

The presence of DMSO in the culture medium significantly increased protein titres. The mechanisms of action remains unclear. DMSO is well known for per-

meabilizing cell membranes and has therefore been used in the past to increase transfection efficiency using electroporation or calcium phosphate (144, 145). However, recent studies showed than a HSPG mediated endocytosis was primarily involved in the transportation of transgenes through the cell membrane of CHO-S cells (73). Moreover, the passage through the cell membrane is thought to happen within minutes after introduction of DNA and PEI in the medium (65). In this study, DMSO has been added one hour post transfection. Therefore, the effect of DMSO on transfection efficiency is more than likely to be negligible here. DMSO has also been found to increase CHO cell-specific productivity. However its mode of action remains unclear. Ye *et al.*. proved that maintenance of transcriptionally active DNA was critical in obtaining high protein expression levels and they suggest that DMSO could help maintain DNA integrity within the cell (146). The effect of valproic acid on cell culture is also well documented and it is used routinely by some companies to increase their transient protein titres (147). Acting as a histone deacetylases inhibitor, VPA prevents DNA condensation, extending the lifespan of viable cells (DNA condensation is known to be involved in apoptotic mechanisms) but also maintaining the accessibility of transgene to RNA polymerases for longer. VPA is also known to block cell proliferation, preventing the gradual loss of transgene through cell divisions. This effect however could not be specifically identified during this study. Interestingly, the addition of LR3IGF to the cultures did not result in higher titres or culture viability which is in disagreement with the results reported by Galbraith *et al.* (77). As Galbraith utilised the CHO K1SV cell line, a possible explanation to this observation would be that the effect of LR3IGF is cell-line specific. Feeding the cultures post transfection contributed to further extend the culture lifespan and increase protein titres. Since the composition of the feed tested was unknown it is difficult to explain in details the mechanisms responsible for these observations. However our results suggest that higher culture viabilities were due to the dilution of toxic-free PEI molecules, but also a possible nutrient depletion during the original batch culture. It is also possible that some components in the feed such as amino acids, may have enhanced protection against apoptosis (148).

In conclusion, a rational development of a PEI-mediated transient production platform in CHO-S cells could be developed in less than 5 weeks with minimal experimentation, yielding up to 90mg L^{-1} and 82mg L^{-1} of a reporter IgG2 and IgG1 respectively, in 10 days of culture. It is important to note that commercially available chemically defined culture media, as well as a non optimised DNA vector were used in this study. Screening for more culture medium at the early stage of development could enable much higher titres to be reached. Modified DNA vectors, with the introduction of specific intron sequences, also showed promising results (149). More importantly, the method presented here is the first example to our knowledge of a rational and holistic approach towards transient transfection process optimisation. The method relies on the sequential screening and/or optimisation of process variables at different levels: the transfection process, the protein expression modulation via the control of the cell environment, and the cell culture process. It allows for greater flexibility as the development can be extended to different process variables and responses. The use of an integrated set of DOE tools minimise the number of irrelevant experiments that generally extend development times, and considerably strengthens the characterisation of the process. Indeed, the effect of process factor variations on process performances can be explored and then rationally controlled for the production of product of high consistency.

Chapter 4

Development of a cation exchange chromatography platform using a Quality by Design approach.

This Chapter presents the development of a cation exchange chromatography step as an intermediate purification platform for the processing of monoclonal antibodies. Using a rational Quality by Design approach, it was possible to evaluate the effect of a variation of operational process parameters on different process responses, and compare the performances of three different sorbents over a relatively wide design space.

4.1 Introduction

Since no single chromatography step can achieve the product purity required for biopharmaceuticals, a two or three process step is generally adopted. Due to its ability to remove host cell proteins (HCP), DNA and protein aggregates, cation exchange chromatography (CEX) is today routinely used as an intermediate or polishing purification step (95, 150). As opposed to protein A, several cation exchange ligands are available on the market. This, coupled with a diversity of base matrix

and manufacturer's specific grafting techniques, contribute to create a large library of sorbents, exhibiting different characteristics. Ghose et al. (99) cite no less than twelve different sorbents commonly used for antibody late purification stages. The efficacy of a CEX step is generally measured by three criteria: (i) the level of clearance of contaminants such as HCP for example, (ii) the dynamic binding capacity as binding a large quantity of protein in one go is generally desirable and (iii) the process yield and duration. The performance of a ion exchange chromatography sorbent are highly dependent on the mobile phases pH and ionic strength. As a result, the implementation of a CEX step requires some optimization. Moreover, the production of monoclonal antibodies, introduced new considerations into the development. The formation of product variants is frequent for this class of molecule. As those variants can alter the product safety and efficacy, it became of primary importance to monitor them throughout the process. The integration of the CEX step within the overall manufacturing process should also be taken into account in the development, the ideal being to limit the number of operations required to adjust the feed material between different processing steps.

The work presented in this Chapter describes the development of a CEX platform for the intermediate purification of a monoclonal antibodies using a Quality by Design (QbD) approach. This work also provides a comparative analysis on the performance of three sorbents: S HyperCel (Pall Life Sciences), Capto^{TM} S and SP HP (GE Healthcare), and a methodology framework for future development with other sorbents or other molecules. In this study, process performance was established in terms of integration with pre and post CEX step: Protein A capture step and anion exchange (AEX) polishing step (Table 4.1) as well as product quality attributes (Table 4.2) and process attributes (maximise yield, minimise processing time).

Table 4.1: Process linkages

Integration criteria	Elution from Protein A	CEX loading	CEX elution	AEX Loading
mAb solution	6mg mL^{-1} in 20mM citrate buffer	No buffer exchange Processing time <4hours	Maximise yield Elute in 20mM phosphate buffer	5-20mg mL^{-1}
pH	3.75±0.1	Identical to ProA elution pool if possible	Identical to AEX loading elution pool if possible	6.5-8.5
Conductivity (mS cm^{-1})	1.6±0.1	Identical to ProA elution pool if possible	Identical to AEX loading if possible	<10

Table 4.2: Product quality attributes involved in CEX

Product variants	Aggregates Charged variants	Less than 5% Equal or inferior to input feed
Purity	HCP DNA	Reduction Reduction
Product attributes	pH	0.5 unit lower than product pI

4.2 Material and methods

Feed material The feed material consisted of 10g of a recombinant IgG1 purified from a CHO-S cell culture supernatant (stable cell clone) on an AxiChrom^{TM} 70/300 column containing 900mL of MabSelect Sure and connected to an ÄKTA Pilot^{TM} system (GE Healthcare, Uppsala, Sweden). The material was aliquoted and stored at-20°C in Protein A elution buffer (20mM citrate, pH 3.75±0.1, conductivity 1.6±0.1mS cm^{-1}) at a concentration of 6mg mL^{-1}. This particular mAb has a pI of 8.8.

Chromatography sorbents and buffers Lab scale experiments were performed on 1mL prepacked columns with three different types of sorbent: S HyperCel (Pall Life Sciences, Portsmouth, UK), HiTrap^{TM} Capto S 1mL and HiTrap SP Sepharose HP (GE Healthcare). Columns were connected to an ÄKTA Explorer 100 chromatography system (GE Healthcare). Operating buffers consisted of 10mM citrate for the equilibration and wash steps, 10mM citrate/ 1M NaCl for the strip step, 20mM sodium phosphate for the elution step and 0.5M NaOH for sanitizing the sorbents. Buffers were titrated by the addition of 1M hydrochloric acid/ 1M sodium hydroxide (Sigma-Aldrich, St Louis, MI) as required. Buffers ionic strength were adjusted by adjusting the conductivity in solution with buffers of the same composition and pH, but supplemented with 1M NaCl. At large scale, a LRC column 10/80-200 (Pall) packed with 11mL of S HyperCel sorbent (Pall), also connected to an Äkta Explorer 100, was used.

Quality control Desired samples were analysed for their mAb concentration and aggregates content by protein A and size exclusion chromatography (SEC) respectively using an HPLC (Shimadzu, Milton Keynes, UK). Quantification of HCP was performed using the CHO Host Cell proteins 3rd generation ELISA by Cygnus Technologies (Southport, NC). Quantification of DNA was performed using a qPCR CHO residual DNA quantification assay (Invitrogen, Paysley, UK). All the quality controls assays were performed by the QC team, Pall Life Sciences, Portsmouth. DOE

generated experimental designs were analysed using Design-Expert (Stat-ease, Minneapolis, MN) and the algorithm described in Chapter 2.

4.3 Results

4.3.1 A risk management approach guided the characterization studies to perform

An initial literature review, coupled with in-house experience, served to establish a list of the critical process parameters *i.e.* the parameters whose variability may have an impact on a product quality attribute, affect a process attributes or process integration. This was done through a risk assessment in which the likelihood of an existing effect was ranked from high (known to affect) to low (no data, or rationale, supporting the idea of the existence of the effect assessed) through medium (may affect). The likelihood of a product or process attribute to be affected by a combination of process parameters guided the nature of the method of experimentation to perform. Indeed, the process parameters known to interact with each other to affect a process response were incorporated into a DOE-guided characterization study. The impact of independent parameters on the process were measured using conventional one factor at a time experimentation. The risk matrix for this CEX step is presented in Table 4.3.

Three product quality attributes (Aggregation, HCP, and charged variants contamination in the eluted product), two process attributes (yield and processing time) as well as process integration with subsequent AEX step were considered. Aggregation is a major concern during manufacturing as a relatively high percentage of aggregates can potentially enhance immunogenicity and/or weaker binding to the molecular target than mAb monomers (151). Moreover, as aggregation can be promoted by numerous factors such as the contact with chromatography sorbents, a change in the chemistry of the solution, or the concentration of the mAb in solution, it was necessary to monitor it through the purification process (152). CEX is a commonly used method for clearance of HCP. Indeed most of HCP being charged it is theoretically possible to isolate them from the product during throughout the purification steps by adjusting the mobile phase pH and conductivity. As a result, mobile phase pH and conductivity during loading, the wash and elution were

likely to affect the proportion of HCPs in the eluate. MAbs, like many proteins, have naturally charge heterogeneity that optimises the balance of gaining favorable electrostatic interactions and determines their structure, stability, binding affinity, chemical properties and therefore their biological reactivity. Charge heterogeneity can however be promoted by process of production and purification by sudden changes in mobile phase chemistry, electrostatic contact with different charged support and temperature among others (153, 154). As a result it was estimated that a CEX purification scheme, that encompasses numerous changes in mobile phase pH could promote the formation of charged variants. Maximizing the binding as well as minimizing loss during following wash and elution step was considered of primary importance to yield an economic chromatography process. The binding/elution of the charged antibodies depending mainly on mobile phase chemistry, it was of primal importance to control mobile phase pH and conductivity to maximise the binding during loading, minimise the loss during the wash step, and maximise the recovery during elution. Processing time was also chosen a marker of process performance. Indeed, it was critical for this project to be able to complete the purification run in less than a working day *i.e.* 8 hours. Yet, processing time could be affected by several operational parameters such as the flowrate and residence time during the loading, wash and elution steps, or even the nature of the sorbent of chromatography. Indeed, different sorbents are prone to exhibit different mass transfer properties. The mobile phase chemistry can potentially affect the rate of desorption of the product from the column, in a sorbent specific manner, and therefore shorten or lengthen the required processing time to achieve an acceptable product recovery or contaminants removal. Finally, with respect to integration with the subsequent AEX process, the pH and conductivity of the mobile phase during elution should, if possible, be adjusted to be compatible with a direct load on an AEX system.

Table 4.3: Risk assessment matrix for CEX step

Step	Process parameters	Aggregates	HCP	Charged variants	Yield	Processing time	Integration with AEX	Risk mitigation
Loading	sorbent	**High**	Medium				Low	Univariate
	pH							DOE
	Conductivity							DOE
	Residence time							DOE
Wash	sorbent							Univariate
	pH							DOE
	Conductivity							DOE
	Flowrate							DOE
Elution	sorbent							Univariate
	pH							DOE
	Conductivity							DOE
	Flowrate							DOE
	Elution stop							Univariate

4.3.2 Characterization and optimization of the product binding

4.3.2.1 Optimal binding capacity conditions were sorbent specific.

Optimal binding capacity was detrmined using the dynamic binding capacity at 10% indicator. The dynamic binding capacity of a media is the amount of protein that can bind to the media before significant breakthrough of unbound proteins occurs. The dynamic binding capacity at 10% breakthrough was determined for the three sorbents tested in this study at various pH, conductivity and residence time combinations. The pH and conductivity of the feed material were adjusted prior to loading on the column by addition of acid/base and salt. Sorbents were equilibrated prior to loading with 10 column volumes (CV) of 10mM citrate buffer at pH and conductivity values identical to the feed material. Experiments were carried out in a random order following a DOE designed plan of experiment. The design assess the effects of various residence time, mobile phase pH and conductivity on the DBC. The design space boundaries were chosen according to data available in literature and experience in house (85, 155–157). The initial tested pH range for S HyperCel sorbent was lower (3.75 to 5.25) than the one tested for Capto S and SP HP sorbents (4.5 to 6). Tested ranges of conductivity (3 to 9mS cm^{-1}) and residence time (2 to 8min) were however identical for the three sorbents. Sample was loaded at the desired flowrate on the column until the UV280nm signal increased and reached a plateau. A 10% breakthrough corresponded at the quantity of mAb loaded on the column for which the UV signal was equal to 10% of the UV signal of the plateau. Dynamic binding capacity response was mapped across the design space as a two dimensional (contour plots) representation of the model equation that links significant variables (at 95%) to the response (Figure 4.1).These maps present a rising ridge type of surface for both S HyperCel and SP HP sorbents whereas a Saddle type can be observed for Capto S sorbent.

For the three sorbents, the effect of pH and conductivity on DBC was highly significant. Where low pH and conductivity seemed to result in high DBC for both S HyperCel and SP HP, a maximum in DBC was obtained at a significantly higher pH when using Capto S sorbent. The effect of the interaction between pH and

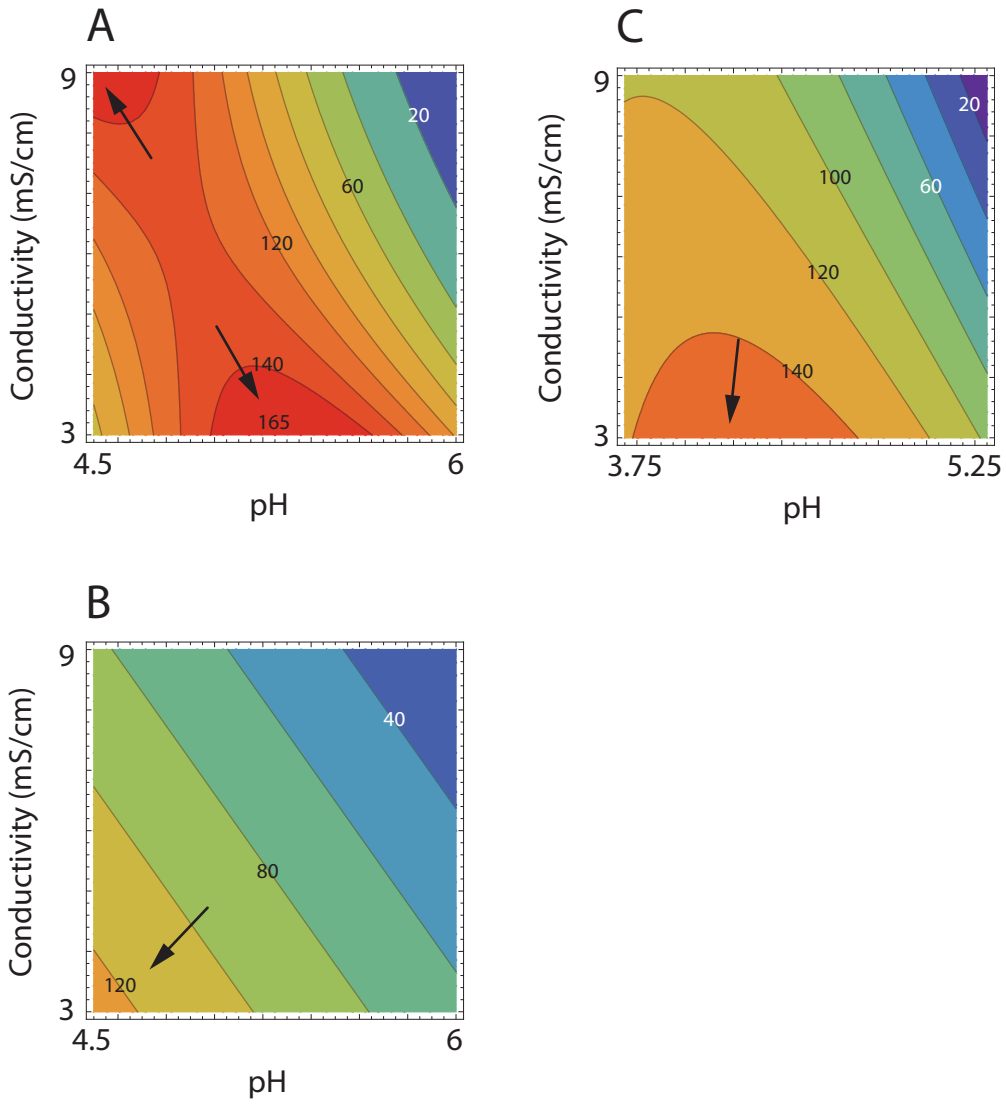


Figure 4.1: Dynamic binding capacity mapping on the initial design space. Recorded DBC at 10% breakthrough at various pH and conductivity for Capto S (A), SP HP (B) and S HyperCel (C) sorbents was mapped across the initial design space. Residence time was set at 5 minutes. The response surface took the form of a rising ridge for both S HyperCel and SPHP sorbent whereas a saddle form was detected for Capto S sorbent. Black arrows indicate the direction where binding conditions should result in higher DBC.

conductivity was also highly significant. When Using Capto S, a relatively high DBC could be achieved at various pH/conductivity ratios in what seems to be a kind of trade off between the two factors. For both S HyperCel and SPHP, DBC was not affected by the residence time of the feed material. However, for Capto S sorbent, increasing the residence time from 2 to 8 minutes resulted in higher DBC.

4.3.2.2 A peak in DBC was located outside of operational space

Figure 4.1 shows that higher DBC were predicted outside the initial design space indicating that experimental design space should be adjusted. In this case, the design space was simply extended by adding experiments. However, the range of experiments to perform was limited by the composition of the feed material. Indeed, the feed material used in this study had a pH of 3.75 and a conductivity of 1.6mS cm^{-1}. Titration of the feed material using acid/base automatically resulted in an increase of the feed solution conductivity. For example, it was practically impossible to get closer from the theoretical maximal DBC using Capto S sorbent as raising the pH of the material to 5.3 resulted in an increase of the feed conductivity from 1.6 to 3mS cm^{-1}. In other words, this co dependence between pH titration and solution conductivity resulted in constraints on the factors settings, defining an "impracticality zone" conflicting with the use of conventional RSM designs (BBK and CCD). This zone is represented by the hatched area in Figure 4.2. Extending the design space to the new irregular experimental region could be done using a specific type of DOE-RSM designs called D-optimal designs. D-optimal designs are computer generated design that establish the set of experimental runs to perform that will minimise the generalised variance of the estimates of the model parameter. Using the Design Expert software and the constraint equation specified by the impracticality zone, it was possible to generate a plan of experiment including the runs already performed. Theoretically, for a D-optimal design involving two factors A and B the constraint equation is:

$$1 \leq \frac{A - LL_A}{CPA - LL_A} + \frac{HL_B - B}{HL_B - CP_B} \tag{4.1}$$

where LL, HL and CP stand for "low level", "high level" and "centre point" respectively. In this study, The main difference with conventional RSM-DOE design lies in the fact that the nature of the model (linear, interactions, quadratic) should be specified before hand. In this particular study, the model chosen was a full quadratic for two factors.

D-optimal designs were run for both S HyperCel and SP HP sorbents only. Data showed that DBC could be significantly increased by loading a feed at low conductivity. Indeed, for SP HP sorbent, a 20% increase in DBC could be achieved, reaching 130mg mL^{-1} at pH 4.5, conductivity 1.6mS cm^{-1}. For S HyperCel sorbent a 10% increase in DBC could be achieved, reaching 160mg mL^{-1} by lowering the conductivity at 1.6mS cm^{-1} as well. Lowering the pH below 4 generally resulted in lower DBC. For both sorbents however the response surface took the form of a rising ridge, meaning that operating boundaries were hit before reaching a peak in response.

Overall, Capto S sorbent allows the binding of more mAb (165mg mL^{-1}) than SP HP (130mg mL^{-1}) or S HyperCel (160mg mL^{-1}) sorbents. However, obtaining high DBC on Capto S required fine adjustments of the feed pH and conductivity prior to loading whereas the protein A eluate could immediately be loaded onto a S HyperCel sorbent.

4.3.3 Optimization of the elution step

The influence of the pH, conductivity and buffer flow velocity on the process and product quality attributes was investigated using a DOE approach. Design spaces were chosen according to literature and previous experience in house. Therefore, despite testing the same range of pH (7 to 8.5) and flow velocity (61 to 306 cm h^{-1}) for the three sorbents, the conductivity range on Capto S sorbent was adjusted to 3 to 9mS cm^{-1}, instead of 9 to 18mS cm^{-1} for both SP HP and S HyperCel. Elution was performed over a number of CV for which, assuming 100% recovery, the final product concentration in the eluate would be 5mg mL^{-1}. For each condition, the concentration of monoclonal antibody in the eluate, the percentage of aggregates as

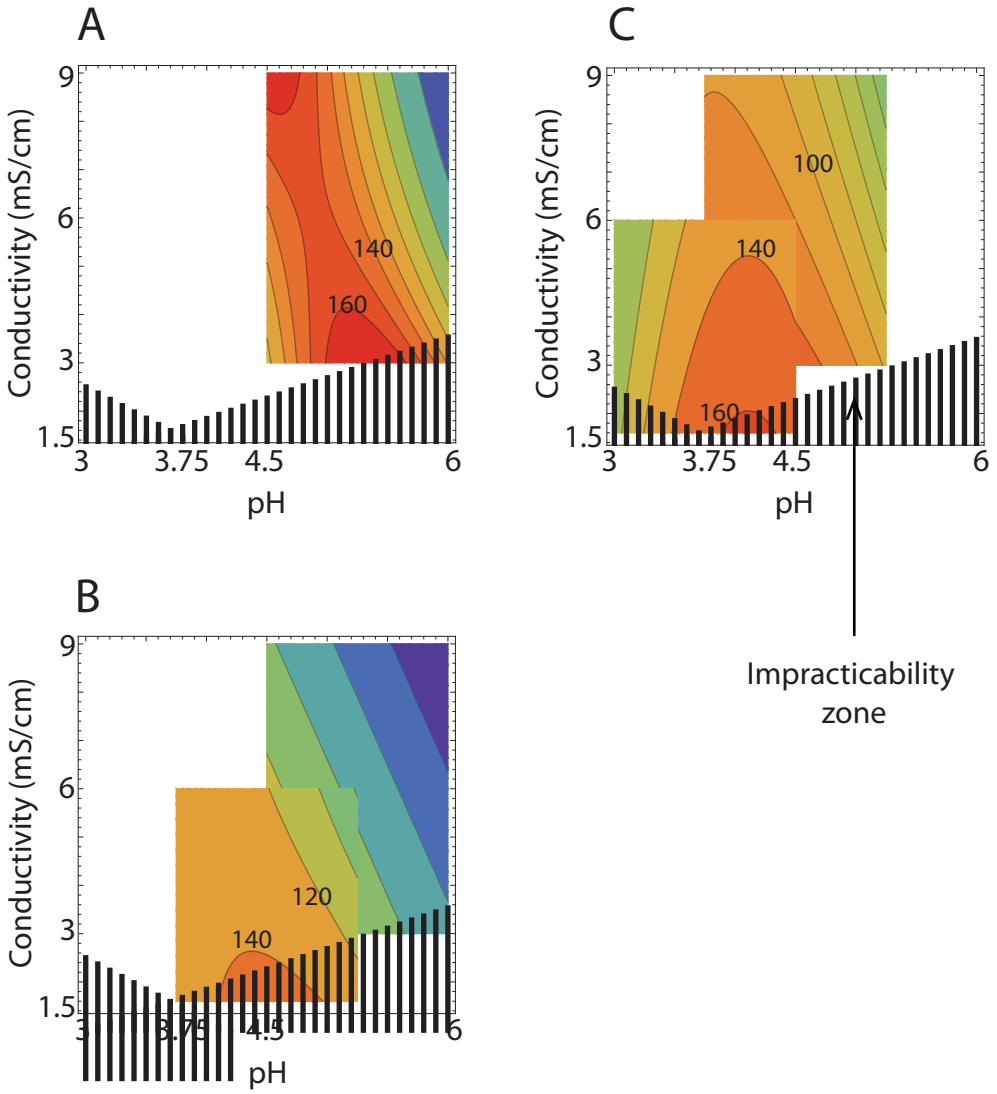


Figure 4.2: Extended dynamic binding capacity response mapping. The initial design space was extended to an irregular experimental design using D-optimal designs. The specificity of this type of design is to integrate operational constraints in its construction to design experiments at the edge of the impracticality zone. Experimental points are represented by the blue squares. Further characterization of DBC was performed for S HyperCel (C) and SP HP (B) only. As the theoretical maximum DBC when using Capto S resin was located in the impracticability zone (A), no further charcaterisation was performed. Residence time factor was fixed at 5 minutes.

well as concentration of HCP for every fraction was determined.

4.3.3.1 The conditions maximizing process yields were sorbent specific.

During elution experiments, fractions during sample load, wash, elution, and strip were collected. Those fractions were analysed for their content in mAb. The overall step recovery was then calculated using mass balance calculations. The step recovery gradually increased with increases in the elution buffer pH and conductivity (Figure 4.3). The effect of an increase in pH alone resulted in an increase in recovery up to a certain extent for S HyperCel and SP HP sorbents. Indeed, past a buffer conductivity of approximately 16mS cm^{-1}, the effect of an increase in pH did not result in an increase in recovery anymore. An increase in buffer conductivity however always resulted in an increase in recovery. However, the decreasing slope of the response surfaces seem to indicate that this effect was less pronounced as the conductivity increased, and that a plateau in response will be achieved by increasing the buffer conductivity further. The effect of pH/conductivity interaction was found highly significant for the Capto S and S HyperCel (associated p-values of 0.0001 and 0.0006 respectively), and significant at 94% for SP HP. In this study, the simultaneous effect of an increase in the elution buffer pH and conductivity resulted in a significant increase in recovery in a rising ridge type of response surface.

Reducing the elution buffer flow velocity had a significant positive effect on recovery for S HyperCel (p-value of 0.0034) (Figure 4.3, D). The effect of flow velocity was however not significant when the Capto S and SP HP sorbents were used. The pH/flow velocity and conductivity/flow velocity interactions effects were not significant, showing that the mass transfers variations due to the elution buffer flow velocity were not influenced by variations in the elution buffer chemical properties.

Within the design space explored, the maximum product recoveries at optimal elution conditions were sorbent specific, with the highest recovery being 99% for Capto S, the lowest being 83% for SP HP while 91% product recovery was achievable with S HyperCel type of sorbent. Data presented in Figure 4.3, A shows that a high recovery was possible with Capto S at relatively low conductivity (below 10mS cm^{-1}).

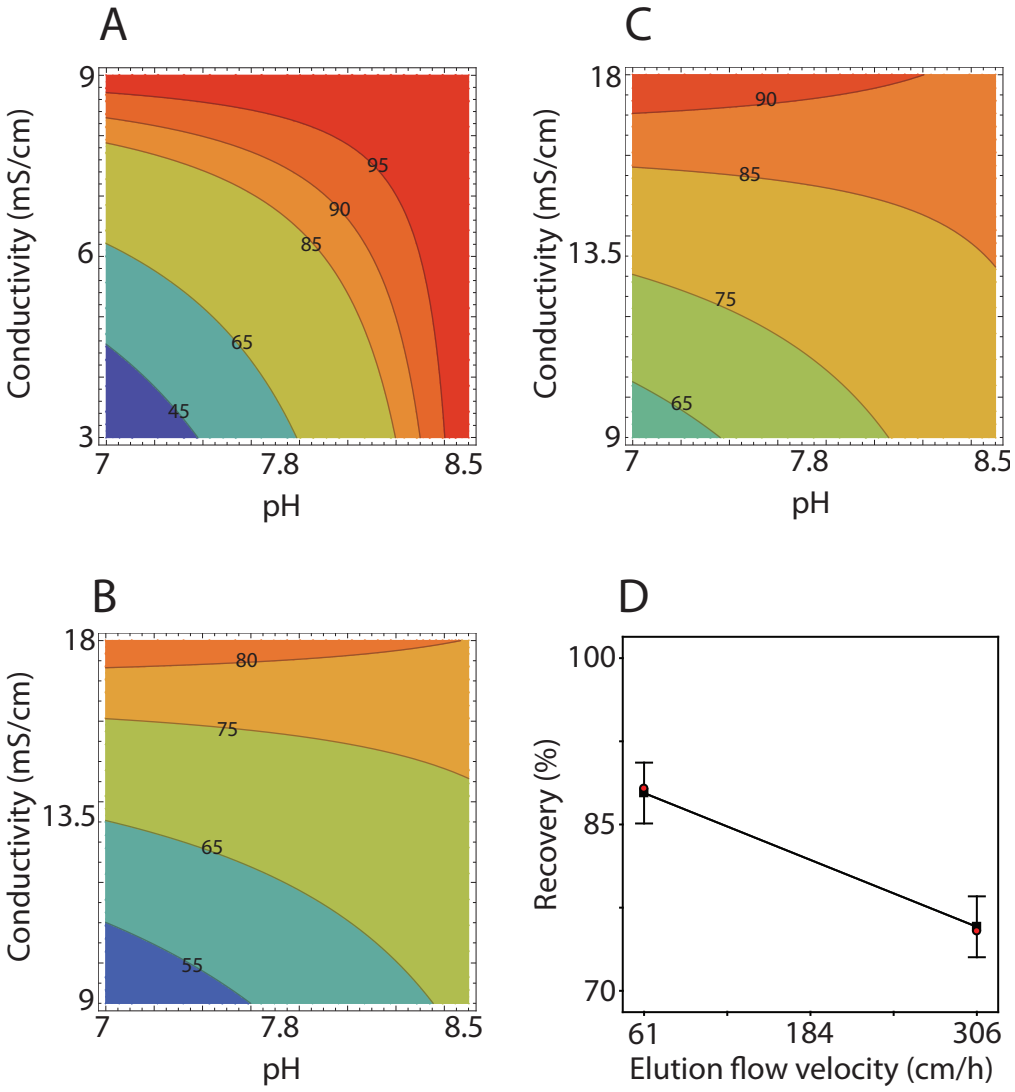


Figure 4.3: Product recovery for the three sorbents tested. All sorbents were loaded at 80% of the maximal DBC identified previously. sorbents were subsequently washed with 10CV of loading buffer, and then eluted with approximately 25CV of elution buffer (20mM sodium phosphate) at various pH, conductivity and flow velocity. Percentage recovery response was mapped for Capto S (A), SP HP (B) and S HyperCel (C) with respect to pH and conductivity factors (flow velocity of 61cm h^{-1}). Percentage recovery for S HyperCel with respect to elution flow velocity is presented in D.

Therefore material eluted from Capto S could be loaded directly on an AEX system without prior dilution. Data also suggest that eluting product from S HyperCel or SP HP sorbent with a buffer at higher conductivity than 18mS cm^{-1} would improve recovery further. However the slope of the response surface is relatively low which suggest that significantly higher conductivity will be required for a relatively low gain in recovery. Figure 4.3, D suggests that eluting the product at lower flowrate from S HyperCel sorbent could significantly improve the percentage recovery. However, because of the minimal flowrate achievable imposed by the AKTA operating system this option could not be validated at small scale.

The observation of differences in product recoveries can be correlated to the profiles of the chromatograms during the product elution and the strip and sanitization of the column. As shown in Figure 4.4, B and C, a significantly higher quantity of material was removed from the sorbent during the strip and sanitization steps when SP HP was used compared to Capto S or S HyperCel. Profiles presented in Figure 4.4, A show significant peak tailing during elution, the tailing being the less important with Capto S while being much broader for SP HP. Interestingly, in all cases, the strip test, performed with a buffer at high conductivity (1M NaCl, greater than 90mS cm^{-1} conductivity) did not suffice to remove all the material bound to the sorbents.

4.3.3.2 The CEX step promoted reversible aggregation in a sorbent specific manner

Preliminary aggregation study showed that samples post CEX exhibited higher percentage of aggregates than the feed material, and this for the three sorbents tested. Full ANOVA analysis showed that none of the three operating factors tested in the elution (pH, conductivity and elution flow velocity) significantly affected aggregation. Therefore, aggregation that occurred during the process was therefore not caused by the variations in mobile phases during elution. The means and standard deviations of monomer percentages recorded for each sorbent were then computed. Results are presented in Table 4.4.

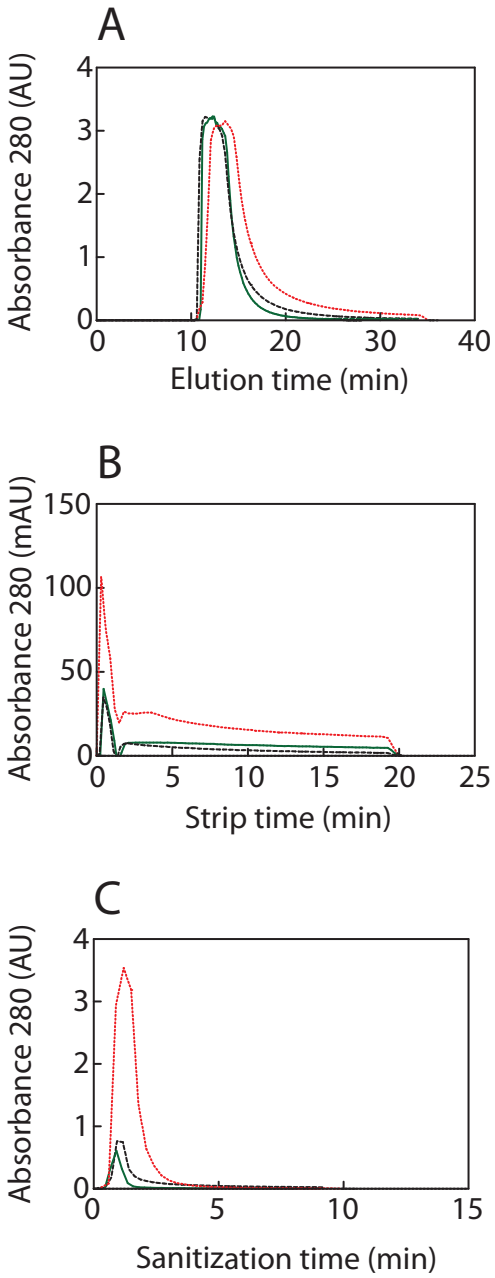


Figure 4.4: UV chromatograms for the elution, strip and sanitization steps for the three sorbents tested. Sorbents were loaded at 80% of the maximal DBC identified previously then washed with 10CV of loading buffer. UV 280nm signal was recorded during the three steps following the wash: elution, strip and sanitization. Elution: the product was eluted with 25CV of 20mM phosphate buffers at the pH and conductivity that maximised product recovery (A). Stripping: the sorbents were then stripped with 20CV of a 10mM citrate/1M NaCl buffer pH 7 (B). Sanitization: finally, all the three sorbents werecleaned of any residues still bound using 10CV of 0.5M sodium hydroxide at flowrates allowing 10min of residence time (C).
Legend: Capto S, S-HyperCel, SP HP.

Table 4.4: Comparison of the percentage of aggregates in eluate for the three sorbents tested.

Sample source	Mean of percentage of aggregates
Feed material	1.37±0.02
Capto S	3.68±1.48
SP HP	18.30±8.72
S HyperCel	10.88±5.56

The percentage of aggregates in the eluate for every single experiment conducted was quantified by HPLC-SEC. The mean and standard deviation of this percentage across the different sorbents tested was calculated and compared with the feed material.

A Tukeys multiple comparison test identified significant difference in means between all pairs of sorbents: SP HP and S HyperCel, SP HP and Capto S as well as Capto S and S HyperCel. As slow loading flow velocity significantly increased the processing time, and therefore the time the feed material was left at room temperature before being loaded on the column, a further assay aimed at assessing the stability of the feed material at room temperature was performed. It consisted in measuring the aggregate content in the feed material right after thawing, and after 2 days left at room temperature. No significant differences were observed with 1.38% and 1.36% of aggregates respectively. As all experiments were conducted with the same feed material, and special care was taken to freeze down the samples down to -20°C right after elution it was concluded that the CEX step promoted aggregation in a sorbent specific manner. Samples purified from SP HP exhibited the highest percentage of aggregates while samples purified on Capto S exhibited the least.

However, these observations could not be repeated afterwards. The percentage of aggregates in the eluate using S HyperCel sorbent was constantly maintained below 4% during further screening experiments. After investigations, it appeared that the percentage of aggregates in solution was proportional to the time the eluate was left at room temperature before performing the quantification of aggregates assay. In other words, aggregation that occurred during the purification on the column was mostly reversible, and that a relatively high percentage of monomers could be recovered when the eluate was incubated at room temperature for 3 hours before performing the quantification assay.

4.3.3.3 HCP clearance efficiency was sorbent specific

The HCP concentration in the column eluates was then quantified. The effect of the pH, conductivity and the elution buffer flow velocity on the final relative HCP quantities in the eluate (ng HCP per mg mAb) were then statistically quantified for the Capto S and S HyperCel sorbent. Using Capto S, none of the experimental parameters were found to affect the relative amount of HCP in the eluate meaning that elution of HCP follows the same dynamic than the mAb product. Using S HyperCel however, a significant cross over effect could be observed for the interaction of the buffer pH and conductivity (p value¡0.001). The interaction plot is presented in Figure 4.5. When conductivity was maintained to its low level (9mS cm^{-1}, increasing

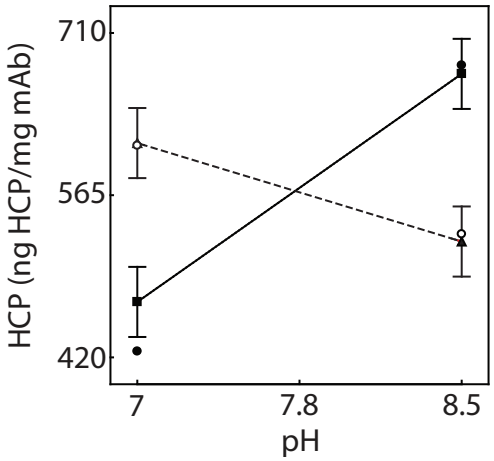


Figure 4.5: The effect of the elution buffer pH-conductivity interaction on the relative amount of HCP eluted from S HyperCel sorbent.
Legend: (———) low conductivity; (– – –): high conductivity. Error bars represent 95% predicted confidence intervals.

the pH of the elution buffer resulted in a higher proportion of HCP in the eluate. When elution buffer conductivity was at its high level however, increasing the pH resulted in a lower relative HCP amount. These observations underlined the fact that a significant amount of bound HCP had a lower pI than the mAb and eluted from the sorbent at lower pH/conductivity combinations than the mAb product. At high pH and conductivity the elution pool was enriched with mAb product, driving the HCP relative amount down. Interestingly the effect of flow velocity on the relative ratio was not estimated significant. In other words, the amount of HCP in the eluate pool followed the same dynamic that the mAb product with respect to

the flow velocity.

HCP clearance (ratio between loaded HCP amount on the column and amount of HCP in the eluate) was more effective with Capto S with which a 84.7% reduction was possible whereas only a 36.8% and 8.9% reduction were achievable with S HyperCel and SP HP respectively. A Tukeys multiple comparison test across the three sorbents tested, revealed significant differences in the means of relative HCP amounts (ppm) in the eluate. Compared to Capto S, HCP relative amount were 4.1 and 6 times greater when eluted from S HyperCel and SP HP respectively (Figure 4.6).

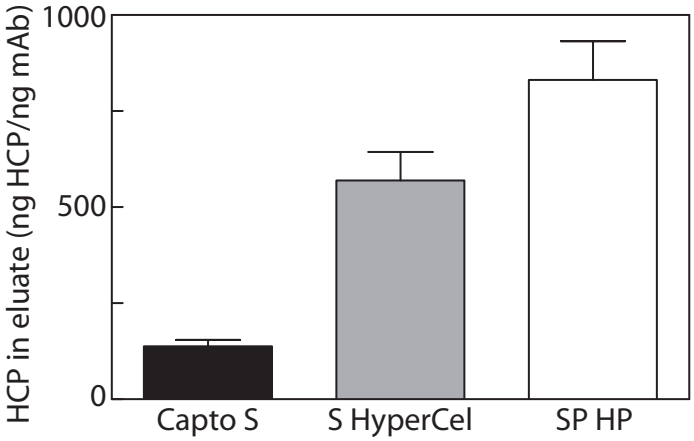


Figure 4.6: Comparison of the HCP amounts (ppm) in eluate across the three sorbents tested. The histograms represent the mean of the HCP amounts of all the experiments performed for a given sorbent. Error bar is the representation of a 95% confidence interval calculated on all the experiments performed for a given sorbent.

4.3.3.4 HCP clearance with S HyperCel could be increased by optimizing the wash step

Conductivity factor was thought being critical for an efficient washing step. Indeed, theoretically high conductivity will weaken the electrostatic interaction between molecules and the resin binding sites, allowing the clearance of weakly bound HCP. However, increasing the buffer conductivity can also result in a loss of product and therefore a lower recovery. Theoretically, the optimal wash buffer conductivity should be as close as possible as the one at which the product of interest will start to elute, allowing all the molecules with lower pI than the product to desorb from

the resin. To identify the optimal wash buffer conductivity, Capto S, SP HP and S HyperCel sorbents were loaded at 80% of the maximal DBC and at the optimal pH/conductivity conditions identified previously. The sorbents were washed with 10CV of 10mM citrate buffer at a pH and conductivity identical to the feed. A 50% gradient elution over 20CV using 10mM citrate/1M NaCl buffer was then performed and the UV280 signal monitored. The conductivities at which the signal started to increase, and at which the signal reached 10% of the maximum were qualified as "minimal wash conductivity" and "maximal wash conductivity" respectively. Experiments were run in duplicate. Results are reported in Table 4.5.

Table 4.5: Wash buffer conductivities boundaries for the three sorbents tested

sorbent	Conductivity Loading (mS cm^{-1})	Minimal wash conductivity (mS cm^{-1})	Maximal wash conductivity (mS cm^{-1})
Capto S	3±0.1	4.49±0.26	7.71±0.2
S HyperCel	1.6±0.1	3.02±1.27	12.62±1
SP HP	1.6±0.1	3.49±0.08	-

Capto S, SP HP and S HyperCel sorbents were loaded at 80% of the maximal DBC and at the optimal pH/conductivity conditions identified previously. The sorbents were washed with 10CV of 10mM citrate buffer at the same pH and conductivity than the feed material. A 50% gradient elution over 20CV using 10mM citrate/1M NaCl buffer at a pH identical to the feed material was then performed and the UV280 signal monitored. The conductivities at which the signal started to increase, and at which the signal reached 10% of the maximum were qualified as "minimal wash conductivity" and "maximal wash conductivity" respectively. Experiments were run in duplicate.

The minimal conductivities of the wash buffers were relatively close from the loading conductivity for the three sorbents tested. The maximal washing conductivities were significantly higher than the minimal conductivities. S HyperCel is more robust to variations in buffer conductivity than Capto S as shown by the smaller difference between the minimal and maximal conductivity identified for Capto S. Maximal conductivities for SP HP could not be established as the UV signal varied with respect to a gradual increase in buffer taking the form of three confounded peaks.

In order to minimise the amount of HCP in the eluate using S HyperCel, the sorbent was washed for 10CV with a buffer at pH 1.6 and conductivity of 10mS

cm^{-1} (80% of the maximal washing conductivity) and a flow velocity of 61cm/h. The mAb product was eluted at the optimal elution conditions identified previously and the final amount of HCP quantified by ELISA. Results showed a reduction of 36.8% of the relative HCP amount (397.36ppm) bringing the HCP clearance from 36.8% to 55.9%. The step recovery was 90.3%.

4.3.3.5 The percentage of charged variants was not affected by the purification step

The analysis of HPLC-IEX assay revealed that the percentage of charged variants in the eluates where not significantly different between all the elution conditions tested. Therefore the pH and conductivity of the mobile phase, as well as the flow velocity during elution, did not promote charge modifications of the product. Moreover, the percentage of charged variants was highly comparable to the feed material (56.23% and 55.1% respectively) with a comparable percentage of acidic and basic variants. As a result, it is believed the whole CEX purification step did not affect the amount of product charged variants.

4.3.4 Ten fold scale-up

The chromatography process developed on S HyperCel sorbent was then scaled-up to allow the binding of around 1g of antibody. It was decided to scale up the process by keeping constant the residence time during loading and the column aspect ratio between width and length. The elution flow velocity was however decrease from 61 to 52cm h^{-1} as, assuming a continuity in prediction accuracy outside the design space, the small scale model predicted an increase in yield with a lower elution flow velocity. This hypothesis was confirmed. Indeed the UV chromatograms during elution, presented in Figure 4.7, shows that less CV were required to elute more than 90% of the bound product. A recovery of 96% could be achieved at large scale.

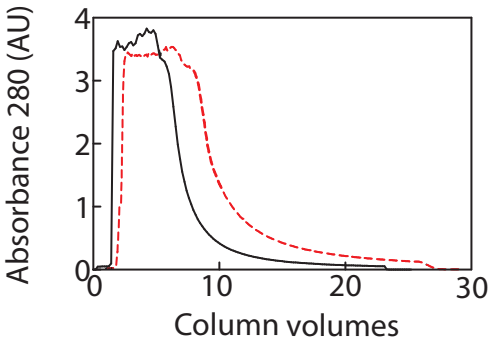


Figure 4.7: Comparison of the elution related UV chromatograms between the small and large scale run using S HyperCel sorbent. The process developed with S HyperCel was scaled-up 10 fold. Flow velocity during elution was decreased from 61 to 52cm h^{-1} to improve the step recovery. Legend: (———): large scale; (– – –): small scale.

4.4 Discussion

Data presented in this chapter demonstrate that the three sorbents tested differed substantially in their abilities to bind and elute product, as well as separate HCP. Despite similar operating mode, the factors combination for optimal binding and elution proved to be sorbent specific. Those differences can partially be explained by differences in (i) ion exchange equilibrium capacities and (ii) mass transfer both in and outside the resin beads.

Maximal DBC strongly depended on the combination of mobile phase pH and conductivity during loading. Theoretically, at low pH and conductivity, electrostatic bond energies between molecules and binding sites of any CEX sorbent are stronger. However, for Capto S sorbent, DBC could be maximised at relatively high mobile phase pH, at which cations have a global lower positive charge, and are supposed to establish weaker bonds with the sorbent. As a result, it seems that maximal DBC did not only depend on the ability of cation molecules to bind to the sorbent, but also on the accessibility for those sites within the sorbent pores. Indeed, at low pH, accessibility could have been limited (i) steric hindrance *i.e.* strongly bound molecules at the pore surface prevent passage of cations inside the bead and (ii) local charge repulsion between the bound and free cation at the surface of the beads. The smaller the pores, and/or the bigger the molecules, the stronger the effects. At higher pH however, mass transfer could be improved and resulted in higher DBC. It is interesting to note that, in the case of Capto S, the detrimental effect of low pH on DBC can be counteracted by a high mobile phase conductivity. This parameter interaction is not a surprise as both pH and conductivity of the mobile phase impact sorbent capacity, and cation overall positive charge.

The differences in product recovery observed, as well as peak broadening during elution are thought to be principally due to differences in mass transfer throughout the different sorbents. Resistance to mass transfer can be caused by two mechanisms. The first one concerns narrow and deep beads pores that can slow down the diffusion of large molecules inside the beads. The SP HP sorbent main characteristic is precisely its relatively small beads and pores diameters. Despite increased resolu-

tion, this characteristic may have resulted in peak broadening. In other words, the use of SP HP sorbent over Capto S or S HyperCel would potentially be justified in cases where a high resolution is required, *i.e.* the product of interest needs to be separated from contaminants having a close net global charge. The second mechanism concerns the kinetics of adsorption/desorption of the molecules to the binding sites. Indeed it takes a certain amount of time for a cation molecule to equilibrate between stationary and mobile phase. In other words, molecules in the mobile phase move ahead of the stationary phase. At constant desorption rate, as the displacement of molecules within the bead pores relies only on a diffusive mode, higher flow velocity will therefore result in accentuated band broadening (158). This effect is highly significant with S HyperCel as shown by the increase in yield when the velocity of the mobile phase during elution was decreased. Removal of contaminants during sorbent washes and sanitisation steps was also logically affected by sorbent mass transfer steps. Those steps needed to be performed at low velocity when using S HyperCel which considerably lengthen the overall process. To save buffer, an alternative could consist in marking a pause during the elution/wash/sanitisation to allow complete desorption of product/contaminants within the sorbent beads pores, before restarting the flow again. The two mechanisms presented could also explain the differences observed in HCP clearance. Because of better mass transfers, HCP clearance is more effective using Capto S than S HyperCel or SP HP sorbents. However, the improved clearance of HCP after optimizing the conductivity of the wash buffer when using the S HyperCel sorbent, showed that a significant amount of HCP with a lower pI than the mAb product, also bound to the resin. Because the loading must be performed at higher pH when using Capto S, it is possible that HCP that bound to the S HyperCel sorbent, did not to a Capto S sorbent.

For S HyperCel and SP HP, the first round of small scale experiments revealed a high percentage of aggregates in the eluate, regardless of the column or the feedstock lot used. This result however could not be replicated, the large scale run giving a reasonably low amount of aggregates. Stability of the sample post elution was therefore investigated by performing aggregate assays right after elution up to 30 days

post elution. Results showed that 10 to 25% of aggregates was present in the eluates regardless of the pH and conductivity of the solution. However, these percentages dropped to less than 5% after few hours only (data generated by John Welsh, Pall Life Sciences, not shown). Therefore, the aggregation that occurred during the purification was mostly reversible. As aggregation can take several forms and depends on numerous factors, it remains a phenomenon hard to foreseen (159) and it is difficult to know with exactitude at what stage of the phenomenon happened. One possible explanation is that the maintained low pH during the loading and washing steps, as well as the close proximity of bound molecules on the sorbent could have both promoted oligomerisation. Aggregation being product specific, the processing of other mAbs using this developed platform should be carefully controlled to ensure less than 5% of aggregates in the final product.

The integration of the CEX step with the previous Protein A and post AEX runs was considered key in this development. S HyperCel sorbent presents here an advantage over Capto S as the feed material from Protein A could directly be loaded on the column without any prior titration. In other words the Protein A elution buffer and CEX equilibration buffer were identical. However, due to the relatively high concentration in salts, the eluate will need to be diluted before being loaded onto an AEX system. Finally, as opposed to Capto S, the amount of protein binding on S HyperCel sorbent was less dependent on the mobile phase pH and conductivity. In other words, a process using S HyperCel sorbent will be more robust to variations in mobile phase chemistry. This will represent a clear advantage for a process designed to be run in a lab with simple, non automated equipment. It is believed that the use of a rational approach towards the development of this CEX platform significantly reduced development times. The outcome for this thesis project was to develop a process based on S Hypercel resin. However, the work presented in this Chapter could also be used as a methodology guide for future characterization of chromatography resins and/or the development of chromatographic step.

Table 4.6: Tested conditions and performance summary

		Capto S	SP HP	S HyperCel
DBC				
Initial screening	pH range	4.5-6	4.5-6	3.75-5.25
	Conductivity range (mS cm^{-1})	3-9	3-9	3-9
	Residence time (min)	2-8	2-8	2-8
Extended screening	pH range	NA	3-4.5	3-4.5
	Conductivity range (mS cm^{-1})	NA	1.6-6	1.6-6
	Residence time (min)	NA	NA	NA
Optimal conditions	Max DBC	165	130	160
	pH	5.3	4.1	4.3
	Residence time (min)	5	2	2
	Conductivity (mS cm^{-1})	3	1.6	1.6
Elution				
Initial screening	pH range	7-8.5	7-8.5	7-8.5
	Conductivity range (mS cm^{-1})	3-9	9-18	9-18
	Flow velocity (cm h^{-1})	61-306	61-306	61-306
Optimal conditions	pH	7	7	7
	Conductivity (mS cm^{-1})	9	18	18
	Flow velocity (cm h^{-1})	184	61	61
Wash				
Optimal conditions	Conductivity (mS cm^{-1})	7.71±0.1	NA	12.62±1
	Flow velocity (cm h^{-1})	184	61	61
Performance				
	Recovery (%)	99	81	96
	HCP clearance (%)	84.7	8.9	55.9

Chapter 5

Development of an integrated processing platform

This Chapter focuses on the scale-up and integration of the developed processing steps to yield a global production platform. The economic aspects of the production of one gram of mAb using this platform are presented, and compared with those associated with the development of a platform using a stable producing cell line.

5.1 Introduction

The adoption of templated production-purification schemes by industry would considerably shorten development time-lines and raw material inventory through the utilisation of common components. The use of process platforms, early in the development of a biopharmaceutical, would also result in streamlined documentation and facilitated transfer to manufacturing. Taking advantage of the biochemical properties shared among all the variety of mAbs, such platforms would be of real value as theoretically, any mAb could be processed through it. While being more frequently used at large scale, such platforms are not yet generalised to the production then purification of relatively low quantities of proteins (100 to 1000mg). Yet, recent studies underlined the fact that production method should be revisited to develop high-throughput platforms able to deliver the quantities of a multitude of therapeu-

tic candidates to support preclinical trials (160). In this context, such platforms would be of real value.

The main challenge of platform development lies in having a holistic approach to process development (161). In other words, the development of a platform should focus on developing a succession of processing steps, able altogether to comply with the wanted product quality attributes, while taking into account factors such as capacity and connectivity as well as cost of production. Currently, the long and tedious development of stable producing cell lines is still a prerequisite to the production of enough quantities of a therapeutic candidates. This development leads to a process dedicated to the expression of only one molecule, and is therefore not compatible with the notion of platform of production. Transient gene expression (TGE) technology however allows for the production of milligrams of recombinant proteins in few weeks only, regardless of the nature of the protein. Moreover, if the development of a TGE platform is often a prerequisite to the expression of a candidate in enough quantities (see Chapter 3), the platform could then be used generically to express protein candidates of the same class, such as different mAbs for example. The following purification platform should therefore accommodate with low protein titres in cell supernatant, characteristic of current transient gene expression processes. Using Protein A sorbent in the initial capture step represents an attractive solution due to its high selectivity (162). Most mAbs have basic isoelectric points that facilitate the use of cation exchange media in a bind-and-elute mode as well as anion exchange in flow-through mode. Combined together, those technologies generally allow for the reduction of contaminants such as HCP and DNA to an acceptable level. Factors such as feed pH and conductivity are critical throughout the platform as they can considerably affect process performance. For example, maintaining mAb in low pH solutions is known to promote aggregation and should be avoided (163). However, low pH solutions is a prerequisite to efficient binding on CEX S HyperCel sorbent. It appears clearly that these factors need to be adjusted throughout the purification process (Figure 5.1).

This chapter describes the development of an integrated platform for the produc-

tion of mAbs in CHO-S cells to sustain preclinical assays. The platform specifications include the production of 0.5 to 1g of protein, the reduction of contaminants such as HCP to hundreds of parts per million (ppm) and DNA to less than 100ppm. Maintaining the percentage of aggregation to less than 5% also represented an objective. The platform was based on an upstream transient expression process coupled to a three stages purification scheme and a final ultrafiltration/diafiltration process. The overall production using this platform allowed for significant time reduction and savings compared to conventional production system.

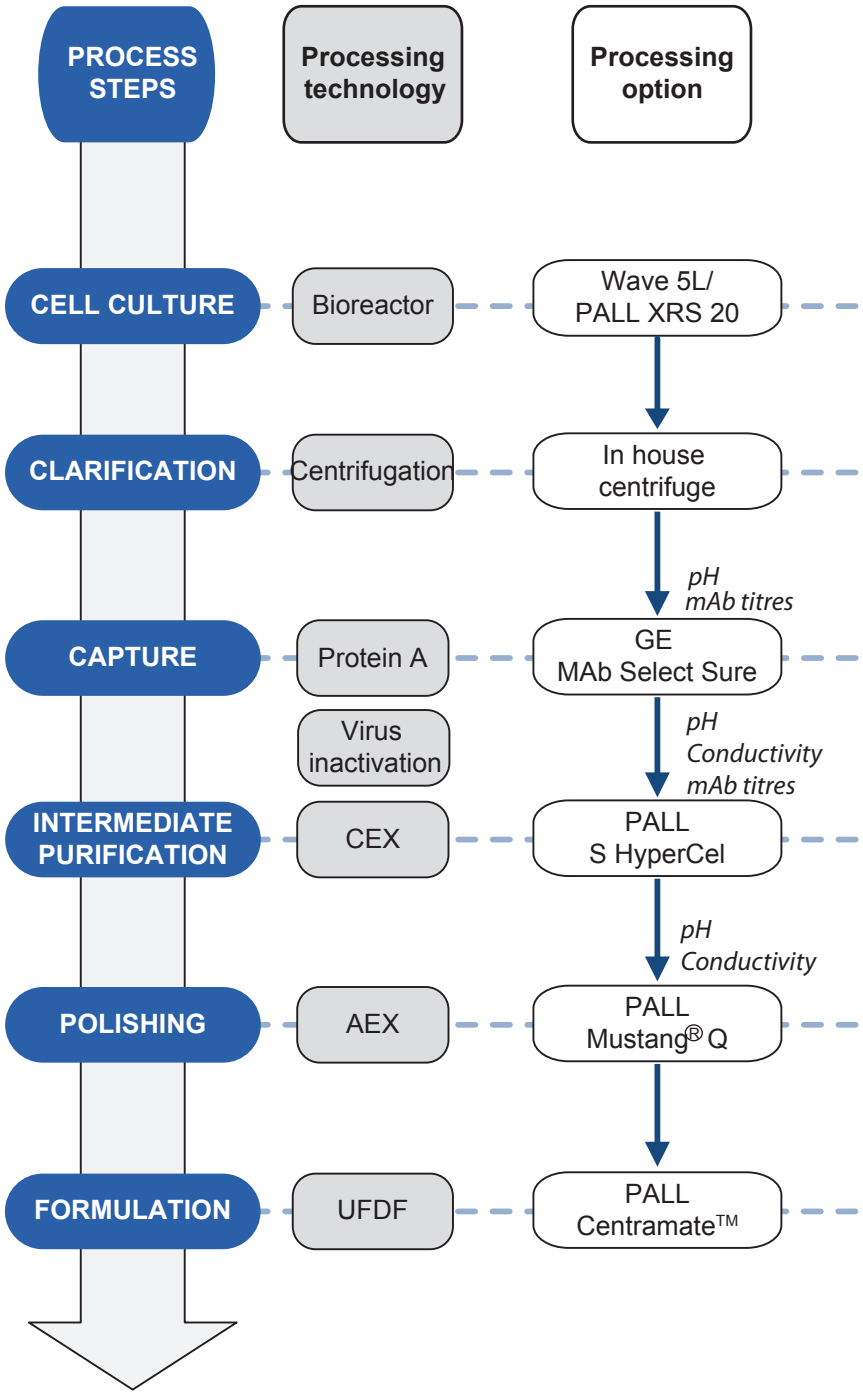


Figure 5.1: Integrated platform for the production of recombinant mAbs. The column on the right presents the upstream processing options retained for the platform. The parameters taken into account for the integration of each step are shown in italic. The upstream cell culture process is performed at 5L (pilot scale) and 20L (production scale).

5.2 Material and methods

DNA production Two different plasmids were used in this study. The first one encoded the recombinant chimeric IgG4 mAb cB72.3 and was provided by Lonza Biologics (Slough, UK). The second one, encoding the recombinant humanised Herceptin IgG1 Her mAb, was provided by Cobra Biologics (Keele, UK). Plasmids were transformed in DH5α^{TM} competent cells from Life Technologies (Paisley, UK). The seed cultures were started from glycerol stocks inoculated into LB medium (Sigma-Aldrich, St-Louis, MI) plus 100μg mL^{-1} ampicillin (Sigma-Aldrich) and grown in 1L baffled vented shake flasks (Sigma-Aldrich) to an OD_{600} comprised between 0.8 and 1.2. The seed cultures were used to provide 1% inoculums for the fermentations. All fermentations occurred in New Brunswick Scientific BioFlo® 310 bioreactors (New Brunswick, Edison, NJ) containing 10L of LB media supplemented with ampicillin at a concentration of 100μg mL^{-1}. Cultures were maintained at 37°C. The pH was controlled at 7.0±0.1 using a 2M Sulphuric acid and 20% Ammonium hydroxide solutions. The dissolved oxygen probe was calibrated to 0% by disconnecting it from the system and 100% with air saturation. The vessel was aerated at one volume of gas per volume of medium per minute and dissolved oxygen was maintained at 40% by proportional-integral control of agitation. Culture were stopped 16 hours post inoculation. DNA plasmids were purified using Giga DNA purification kits either from Sigma-Aldrich or QIAGEN (Crawley, UK). Purified plasmids DNA were re-suspended in endotoxin free water, and stored at -20°C. A fraction of the purified DNA plasmid was plated on an agar plate incubated at 37°C for quality control.

Pilot scale transient production - 5L scale in Wave CHO-S cells (Life Technologies) were routinely cultured in roller bottles (Corning, Surrey, UK) in CD-CHO medium (Life Technologies) supplemented with L-glutamine (Lonza) at a concentration of 8mM at 37°C in 5% (v/v) CO_2 and rotated at 2rpm. Cells were re-suspended in fresh medium every 3-4 days at a concentration of 2.00E+05 cells mL^{-1}. Cell concentration and viability were routinely measured using an hematocytometer and the Trypan Blue exclusion assay. The passage no. 5 was used to start the preculture by

inoculating an AppliFlex 10L wave-motioned bioreactor (Applikon Biotechnology, Tewkesbury, UK) containing 5L of CD-CHO and 8mM of L-glutamine, at a concentration of 2.00E+05 cells mL^{-1}. Cells were grown until mid-exponential phase then harvested in 1L centrifuge bottles. Cells were spun down at 200g for 20min, the spent media discarded, and the cells re-suspended at 20.00E+06 cells mL^{-1} with 0.83L of fresh medium. The re-suspended culture was transferred into two 2L roller bottles. Re-suspended cells were transfected by adding 15mg of pB72.3 DNA plasmid (Lonza,) then 20mL of a 1mg L^{-1} solution of PEI in each bottle (Polysciences, Warrington, USA). The transfected cultures were incubated for one hour at 32°C in 5% (v/v) CO_2 in roller bottles rotated at 2rpm. The cultures were then supplemented with 0.75mM of valproic acid and 1% v/v of liquid DMSO (Sigma-Aldrich). In the mean time, a volume of 2.5L of fresh medium was pumped back into the bioreactor and warmed up to 32°C. After incubation, the transfected cultures were pumped back into the bioreactor and agitated at 25rpm and a rocking angle of 8°. The culture was fed with a 0.42L mix of 50% (v/v) of FeedA/FeedB (Life Technologies) every two days, starting at day 2 post transfection (Figure 5.2).

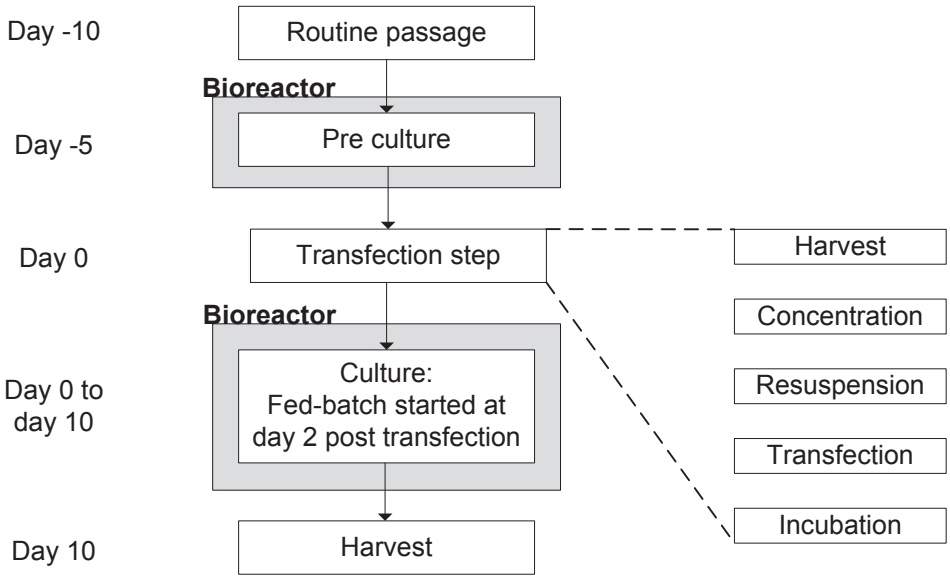


Figure 5.2: Pilot scale transient upstream production process flowchart. The whole process duration lasted for 20 days. The preculture and the culture post transfection were carried out in the same disposable bioreactor. Prior to transfecting cells, the culture was harvested then concentrated down using a centrifugation operation. The cells were incubated for one hour post transfection in an off line incubator before being transferred back into the bioreactor.

Large scale transient production - 20L scale in XRS The procedure was identical to the one used at the pilot scale with the exception of some minor modifications detailed below:

- A plasmid coding for a humanised IgG1 Her was used.
- CHO-S cells were routinely cultivated in shake flasks instead of roller bottles.
- The concentration/re-suspension step took 1hr instead of 30min
- The preculture was re-supended in 6 bottles. A total of 6 individual identical transfections were performed within those bottles.
- All the quantities were multiplied by 4.

Capture by Protein A Two pre-packed HiTrap MabSelect Sure 5mL columns (GE Healthcare, Upssala, Sweden) were connected in series to an ÄKTA Explorer 100 (GE Healthcare). The columns were equilibrated with 10CV of 20mM sodium phosphate buffer pH 6.8 at 5mL min^{-1}. After loading the sample, columns were washed with 5CV citrate buffer pH 5 at 5mL min^{-1}. The product was then eluted with 3CV of 20mM citrate buffer pH3.5 at a flowrate of 5mL min^{-1}. Columns were stripped with 20mM citrate buffer pH2.5 then re-equilibrated. Columns were sanitised with 10CV of 0.1M sodium hydroxide at 5mL min^{-1} every three runs.

Cation exchange purification A LRC column 10/80-200 (Pall Life Sciences, Portsmouth, UK) was packed wth 11mL of S HyperCel sorbent (Pall Life Sciences) and connected to an ÄKTA explorer 100 (GE). The column was equilibrated with 10CV of 10mM citrate buffer pH 3.75, conductivity 1.6mS cm^{-1}. Protein A eluate was then loaded on the column in one go at a velocity of 14.3cm min^{-1}. The column was washed with 5CV of 10mM citrate buffer, pH 3.75, conductivity 10mS cm^{-1}. Bound mAb was eluted with 6CV of 20mM phosphate buffer pH7, conductivity 18mS cm^{-1}. The wash and elution buffers were pumped at a flow velocity of 52cm h^{-1}. The column was then stripped with 10CV of 20mM sodium phosphate/1M sodium chloride pH 7 and sanitised with 5CV of 1M sodium hydroxide before being

re-equilibrated.

Anion exchange purification Small scale development experiments were conducted using the AcroPrep^{TM} Advance Mustang Q 96 well plates (Pall Life Sciences). The eluates collected during the development of the cation exchange chromatography step served as feed material (refer to Chapter 4). Those eluates were first diluted 15 times with 1400μL of 10mM sodium phosphate/sodium chloride buffer at the same pH and conductivity as the corresponding eluate solution. A volume of 250μL of each of the diluted eluate solutions was then pipetted into the wells of the plate. The volume inside each well was first reduced to 100μL before adding 150μL of sodium phosphate/sodium chloride buffer. This step was repeated three times. For each well, the mAb concentration as well as the percentage of aggregation were quantified using HPLC-Protein A and HPLC-Size exclusion chromatography respectively.

Production scale purification was performed using Mustang Q coin mounted into an appropriate stainless steel housing (Pall Life Sciences). The set-up was connected to an ÄKTA Explorer 100 (GE Healthcare).

Concentration/buffer exchange Concentration/buffer exchange was performed using a 30kDa T-series Centramate cassette, 0.02m^2 filtration area (Pall Life Sciences). The cassette was mounted in a specific holder and washed as specified by the manufacturer's instructions. The system was then equilibrated by pumping 300mL of a 2.4mM L-histidine/50mM D(+)-trehalose formulation buffer. The flowthrough of the anion exchange purification step was first reduced to reach a concentration of approximately 50mg mL^{-1} of mAb in solution (15 times concentration). Then the buffer was exchanged with the formulation buffer using diafiltration mode by maintaining constant volume throughout the system.

Quality control Desired samples were analysed for their mAb concentration and aggregates content by Protein A and SEC respectively using an HPLC (Shimadzu,

Milton Keynes, UK). Quantification of HCP was performed using the CHO Host Cell proteins 3rd generation ELISA by Cygnus Technologies (Southport, NC). Quantification of DNA was performed using a qPCR CHO residual DNA quantification assay (Life Technologies). All the quality controls assays were performed by the QC team, Pall Life Sciences, Portsmouth.

5.3 Results

5.3.1 The scale-up of the TGE process proved to be challenging

The optimised transient transfection process described in Chapter 3 was then scaled-up from 10mL culture in CultiFlasks to a 5L culture in a Wave bag. The seed culture grown in roller bottles was used to inoculate a 5L preculture in Wave bag. The cells proved to adapt with difficulty to the culture in Wave bag. Indeed, the culture viability at the end of the preculture was only 90.2%. As a comparison, the cell viability of the preculture in the shake flasks used for the small scale experiment could be maintained above 98% prior to transfection.

The 5L preculture was then harvested, the cells concentrated, transfected and finally transferred back in the Wave bag as described in the Material and Methods section. IgG titres and culture viability were assessed throughout the culture and the results compared with data from small scale experiment (Figure 5.3). Both the small scale and pilot scale cultures were marked by an increase in productivity 4 days post transfection. This increase was correlated with a doubling in viable cell density between day 2 and day 4 for the small scale experiment, and an increase of 29% between day 2 and day 4 for the pilot scale culture. Indeed at day 4, viable cell density was estimated at 9.63E+06 cells mL^{-1} in small scale culture and 6.46E+06 cells mL^{-1} at pilot scale. As a result, it is most probable that the transfected plasmid DNA was diluted during the first cell divisions. As cell division resulted in an increase in overall productivity, it is probable that the transfected cells protein expression machinery was saturated by an excess of plasmid DNA. Compared to the small scale experiment, the culture in Wave bag was characterised by a lower cell productivity between day 4 and day 10 (5.8 mg L^{-1} d^{-1} compared to 11.2 mg L^{-1} d^{-1}). Moreover, cell productivity declined significantly after day 8 with the production of only 1.48mg L^{-1} d^{-1} over the last two days in culture. In comparison, productivity at small scale remained constant throughout the culture. Cell specific productivity were comparable with 17 and 14 pg $cell^{-1}$ day^{-1} between day 4 and day 8 for the small and pilot scale respectively. Because of the lower amount of viable cells, IgG

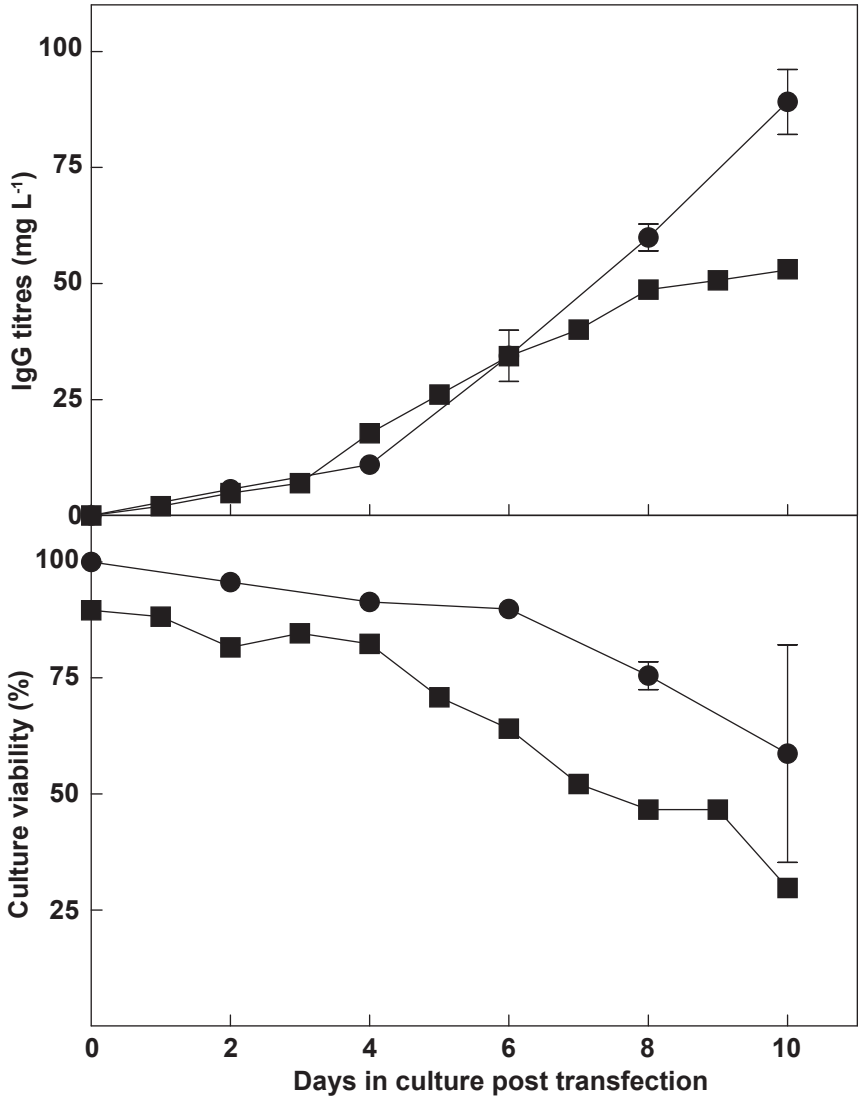


Figure 5.3: Evaluation of the transient expression process scalability. Square symbols: 5L Wave; circle symbols: 10mL culture in CultiFlasks. The TGE process developed in 50mL CultiFlasks was scaled-up to a 5L scale in Wave bag. A 10L Wave bag containing 5L of media was first inoculated at 2.00E+05 cells mL^{-1} with cells routinely grown in roller bottles. After 5 days, the culture was harvested in centrifuge bottles. Cells were pelleted down and the conditioned medium discarded. Cells were then re-suspended at 20.00E+06 cells mL^{-1} in 0.83L of fresh medium and transfected by adding DNA then PEI before being incubated at 32°C. In the mean time, 2.6L of fresh medium were pumped back in the Wave bag and warmed-up to 32°C. One hour post transfection, the Wave bag was inoculated with the transfected culture to reach a concentration of 5.00E+06 cells mL^{-1}. Culture was fed every two days, starting at day 2 post transfection, with 1.67L of Feed A/FeedB (50% v/v). Square symbols: 5L Wave; circle symbols: 10mL culture in CultiFlasks.

titres were significantly lower in the Wave bag after day 6, reaching a maximum of 53mg L^{-1} at day 10, 40% less than at small scale. Culture viability was lower in the Wave bag throughout the culture and started to drop significantly after day 4. This drop in viability was also observable at small scale, but only occurred at day 6. In other words, the culture started to die earlier at 5L scale. This effect may be attributed to the starting viability post transfection, which was significantly lower at 5L scale than at small scale (89.6% and 98% respectively), which in return, was the consequence of a relatively low viability at the end of the initial preculture carried out in the Wave bag. This effect could be attributed to the agitation in roller bottles which was limited by the maximal operating rotation speed of the incubator. This agitation proved to be insufficient to maintain the totality of the culture in suspension and prevent early cell clumping and cell death compared to a culture in shake flask.

The process was then scaled-up in a XRS bioreactor to a scale of 20L. Here, the cells routinely cultivated in shake flasks proved to adapt well in the bioreactor. Indeed, cells could be cultivated until mid exponential phase in 4.5 days while maintaining the preculture viability above 98%. The concentration step post preculture proved to be relatively complex due to the volume of culture. Indeed the culture had to be split up in 16 individual 1L centrifuge bottles. A total of three concentration steps by centrifugation and 16 individual re-suspensions were performed. Yet, the cells did not seem affected by the process as a culture viability of 98.5% could be measured in the bag post transfection could be measured. Data acquired until day 4 showed that the transfection process could be performed successfully. Indeed at day 3, the culture viability was at 95.8% and IgG titre in culture at 21.6mg L^{-1}. However, a bacterial growth occurred between day 3 and 4, that compromised the culture.

As the 20L XRS transient culture was compromised by the presence of bacteria, it was decided to generate a simulant, as close as possible as what the original transient culture (yields of 100mg L^{-1}) would have been. To do so, 20L of feed material from a stable CHO batch culture containing the IgG1 Her at 1.2g L^{-1} was

purified by Protein A using 900mL of MabSelect Sure sorbent, an ÄKTA pilot and an Axichrom 70/300 column (GE Healthcare). The Protein A sorbent captured the mAb in solution, and the process generated a flowthrough with levels of IgG under the detection threshold of 10mg L^{-1}. As the eluate of the column was relatively pure of contaminants, it is believed that most of these contaminants such as HCP and DNA were present in the flowthrough. Purified IgG was subsequently spiked in the flowthrough to generate a feed at a concentration of 100mg L^{-1}. The concentration was checked by HPLC-Protein A and estimated at 102mg L^{-1}. This simulant served as feed material to test the purification platform.

5.3.2 Development of a Protein A sorbent based capture step

Preliminary experiments consisted in evaluating the relation between the mAb concentration in solution, the residence time within the column of chromatography and the quantity of mAb that can be bound to 1mL of MabSelect Sure Protein A sorbent. Feed material containing the IgG1 Her at concentration ranging from 0.5 to 6.5mg mL^{-1} were generated by spiking the protein into CHO-S cell culture supernatants. Residence time within the column was varied from 1 to 5min. For each combination of residence time and IgG concentration, fractions of the flowthrough were taken during the loading, and mAb concentrations in each fractions quantified by HPLC-Protein A. The total mass of mAb which was bound before breakthrough occurred was calculated by multiplying the concentration of mAb in the feedstock by the total volume loaded up to a 10% breakthrough point. Results showed that the DBC was affected by the feed residence time. Indeed, the DBC got closer to the theoretical sorbent capacity with an increase in residence time (Figure 5.4). The effect of mAb concentration in solution on DBC however, was not found significant at 95%. In other words, the amount of mAb that can be bound to Protein A before breakthrough can be increased by increasing the residence time within the column, up to the point where the actual capacity of the column is achieved. The quantification of mAb quantities in the fraction collected during the DBC study showed that breakthrough occurred relatively suddenly after a given amount of mAb was bound

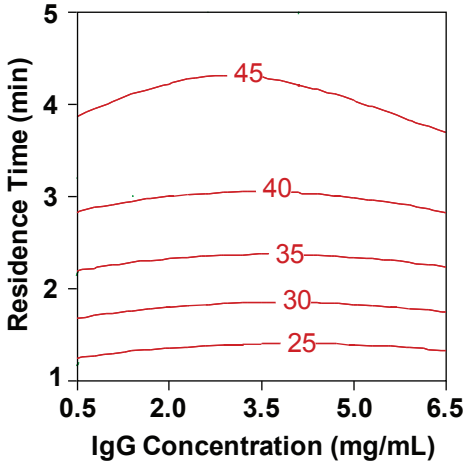


Figure 5.4: MabSelect Sure DBC study. Feed material containing the IgG1 Her at concentration ranging from 0.5 to 6.5mg mL^{-1} were generated by spiking the protein into CHO-S cell culture supernatants. Residence time within the column was varied from 1 to 5min. Fractions of the flowthrough were taken during the loading and mAb concentrations in each fractions quantified by HPLC-Protein A. The total mass of mAb which was bound before breakthrough occurred was calculated by multiplying the concentration of mAb in the feedstock by the total volume loaded up to a 10% breakthrough point. *Courtesy of Nigel Jackson, Pall Life Sciences.*

to the sorbent.

It is common practice to apply the sample to a final load of 80% of the DBC, 10% breakthrough (164). Therefore, at a 5min residence time, it could be established that 1mL of MabSelect Sure could bind 36.5mg of mAb. As a result, binding 1g of mAb in one attempt would require 27.4mL of sorbent. However, as transient feeds are characterised by low concentration of mAb, loading the feed on the column while respecting a 5min residence time would represent a rather long process. For example, with a feed at 100mg L^{-1}, the loading would take more than 30 hours on a single column. As culture feed was rich in nutrients and sterility was not maintained during cell harvest operation, the feed was prone to bacteria contamination and growth within few hours. Therefore, it was decided to reduce the time of loading to less than 4 hours and to scale out the capture process instead of scaling it up. Residence time was decreased to 2 minutes, and the volume of sorbent set to 10mL by connecting two 5mL prepacked columns in series. The column was loaded with only 115mg of mAb in one go (1150mL of feed). A series of 10 runs was performed by stripping and re-equilibrating the column between each run. The average yield was

95.9% ± 0.3% giving a total of 1103mg purified mAb. This yield was comparable to the yields obtained during the development of the capture step performed in house from a CHO supernatant containing a stably expressed mAb at a concentration of 1.2g L^{-1} (data provided by Nigel Jackson, Pall Life Sciences).

5.3.3 A membrane-based AEX process allowed for a fast and efficient polishing step.

The eluate of a cation exchange chromatography step was then used as feed material to develop an AEX process. The effect of the pH and conductivity of this feed material on AEX process performances (yield and product aggregation) was experimentally assessed at small scale using a 2 factors CCD design. The material pH values ranged from 6 to 8 while material conductivity ranged from 3 to 9mS cm^{-1}. Data showed that in the ranges tested, neither the pH or conductivity were impacting the process performances, with p-values of 0.18 and 0.33. The process yield was above 95% for all the experiments conducted while the percentage of monomers exceeded 97%. Analysis conducted on one sample (pH7, conductivity 6mS cm^{-1}) showed a Log reduction of 2.57 in DNA and a three times reduction in HCP.

The process was then scaled from the 96 well plate to a Mustang Q coin to process hundreds of millilitres. The feed material issued from a large scale cation exchange purification step was firstly diluted down using a 20mM sodium phosphate buffer pH7 to bring the conductivity down from 18 to 6mS cm^{-1}. The feed was loaded on the system at 3.5mL min^{-1} then washed for 5 minutes with 20mM Sodium Phosphate buffer pH7, conductivity 6mS cm^{-1}. The process showed to be scalable as a yield of 95.7% could be achieved while the aggregate content was maintained at its level pre-purification. The overall process could be completed in 125min with 100min for the sample loading only.

5.3.4 Step by step performance of the platform

Along the purification of the reporter mAb from the simulant feed, samples were taken at each processing step to determine the step yield, the percentage of monomers

Table 5.1: Platform performances step by step.

	IgG (g/L)	Monomers (%)	Charged variants (%)	Step yield (%)	mAb (mg)	Time	*HCP (ng mg^{-1})*	*DNA (ng mg^{-1})*
Small scale Transient	0.08	99.2	36.2	-	-	20d	-	-
Simulant	0.1	99.1	54.6	-	1150.00	-		
Capture	8.09	99.2	55	95.9	1102.85	5d	*813.6*	*32129.39*
Intermediate	19.35	97.8	55.1	95.8	1056.53	4hr	*307.6*	*112.4*
Polishing	3.02	98.4	54.2	95.7	1011.10	2hr	*102.4*	*0.3*
UF/DF	50.23	98.4	-	98.9	999.98	2hr	*80.7*	*0.44*

Samples were taken along the production and purification of the reporter mAb. The percentage of monomers were quantified by size exclusion chromatography, while the percentages of charged variants were obtained by HPLC-Ion exchange chromatography. MAb concentrations were determined by manual UV280nm spectrophotometry excepted for the cell culture small scale transient and simulant materials which were quantified using an HPLC-Protein A assay. Individual step yields were maintained above 95% leading to an overall efficiency of 86.9% from the expression of the mAb to the final UFDF step. The amount of HCP and DNA were quantified during the development of the platform at small scale with a feed material originated from a CHO-S supernantant containing a stably expressed IgG1 at 1.2g L^{-1}.

and correctly charged molecules. Results are reported in Table 5.1. Aggregation was mostly not affected by the different processing steps as the percentage of monomers was kept above 97.8% throughout the purification. Data on the cation exchange step showed however than significant aggregation could occur during the process (see Chapter 4). However the phenomenon proved to be reversible and therefore did not affect the quality of the final product. The percentage of charged variants was on average maintained at 55% throughout the purification. Therefore it is more than likely that molecules were altered during the cell culture step and not during the purification. The percentage of charged variants in a transient supernatant was 1.5 times lower than in a culture where the same mAb was stably transfected. In other words, transient gene expression led to better product quality than stable expression. The yields of all the processing steps were above 95% and the overall yield across the platform was 86.9%. While the upstream transient expression process took 20 days to perform, the purification of the mAb could be carried out relatively quickly.

The amount of HCP and DNA that figure in the table were quantified during the

development of the platform at small scale with an original supernatant containing a stably expressed mAb at 1.2g L^{-1}. Using this platform, it was possible to obtain a 10 times reduction of the relative amount of HCP (ng of HCP per mg of mAb) and a 4.9Log reduction in the relative amount of DNA.

5.4 Discussion

The scale-up of the transient gene expression process from 10mL scale to 20L proved to be particularly challenging. The main difficulty lied in maintaining sterility while harvesting the preculture in centrifuge bottles, centrifuging down litres of broth, then re-suspending the cells in fresh media. If these steps were successfully performed at the 5L scale, a contamination occurred at the 20L scale. It is difficult to identify the stage at which the contamination occurred. However, after a little investigation, the integrity of the seal in the lid of some centrifuge bottles showed to be compromised and it is more than likely that culture got contaminated when in contact with the lid of some bottles. It appears now of primal importance to be able to concentrate the cells in-line, *i.e.*, maintain the cells within the culture bag while removing culture media. This would significantly mitigate the risk of contamination but would also speed up the overall process. Concentration in-line could be performed by using a TFF micro filtration module connected to the culture bag. In house data showed that it was possible to concentrate a high density CHO cell broth (14E+06 cells mL^{-1} 25 times in 20 minutes while maintaining a high cell viability (Woodgate J., data not shown). Therefore this technology could be used to perform the concentration step prior to transfection. (Figure 5.5).

The costs of the platform were driven by the cost of consumables required for the DNA production and culture media for the transient gene expression step. Therefore, primary efforts in the future should focus on improving the transient titres up to 600mg L^{-1}. Preventing early cell death post transfection also represented a challenge. From a process design point of view, a perfusion culture system in which conditioned media is constantly replaced by fresh media would represent an interesting alternative to the fed-batch process presented here. Indeed, culture by products but also toxic molecules are known to infer with cell productivity. The key parameter for a conventional perfusion system is the retention of cells in the bioreactor. This is generally performed by the use of filters through which cells cannot pass. However in this particular case, because transient titres were relatively low compared to cultures of stable producing clone, it would be of primal importance to

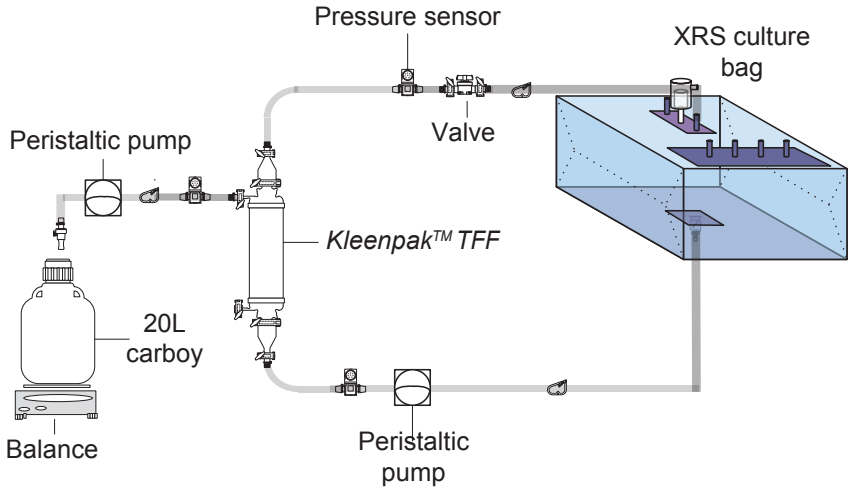


Figure 5.5: In line concentration of a cell culture using a Kleenpak TFF microfiltration module. By coupling a TFF module top the XRS bioreactor it should be possible to concentrate down the preculture relatively quickly while maintaining cells alive. The concentration factor could be calculated in real time by performing a mass balance between the volume of media in the bag, and the media collected on the balance. SciPres® pressure sensors (SciLog, Inc, Middleton, WI) can be sterilised by autoclaving and could easily be integrated to the system to apply a desired trans-membrane pressure. *XRS culture bag figure is a courtesy of John M. Woodgate.*

also retain the expressed product. In the case of mAbs, an ultrafiltration filter with a pore size of 30kDa could represent a viable solution. However, filters are subject to fouling. Wave Biotech published a few years ago the development of a floating perfusion filter at the surface of the culture liquid. The tangential movement of the filter during the rocking of the bioreactor platform would provide enough shearing to prevent filter fouling (107). Another solution could consist in designing a double bag bioreactor or a bag with an external layer. The inner face of the layer would be porous to allow nutrient exchange through the whole surface of the culture bag. Yet, by enhancing productivity titres to 0.5g L^{-1}, it would also become possible to perform the transient culture in smaller bioreactors such as shake flasks or even vented bottles. In this case, the bioreactor could be divided into two compartments separated by a porous layer. The bottom compartment would contain the cell culture, the top serving as manual exchange of fresh/conditioned media at regular intervals during the culture to mimic a semi continuous perfusion mode.

The purification of 1g of mAb could be performed relatively quickly in an efficient manner. The main difficulty being the capture of a feed lowly concentrated

in mAb. Working with an excess of Protein A sorbent allowed the experimenter to decrease the residence time and therefore speed up the process while maintaining a high step yield. Moreover, due to the cost of Protein A sorbent, significant savings could be made by scaling the process out, instead of scaling it up. The further DSP applications were relatively easy to implement and integrate as processing platforms. The Protein A eluate was normalised in terms of concentration, pH and conductivity. Therefore, the nature of the protein expressed, or the upstream culture process, would have a limited impact on the late purification steps performances. The transition between the Protein A capture step and the subsequent cation exchange step did not require any adjustment of the feed material. Indeed, the Protein A elution buffer and S HyperCel equilibration buffer were identical. Therefore, the eluate from Protein A could directly be loaded on the cation exchange chromatography column. This characteristic could in the future be employed to design a more continuous process with no interruption between the capture and the intermediate purification step. Using membrane based chromatography technology instead of sorbent beads proved to be advantageous when the volume of feed to process was relatively high. In this study, membrane chromatography was used for the late polishing purification step. At this stage, diluting the salts in the cation exchange eluate was a prerequisite before performing the final purification step. As a result, a relatively large volume of material needed to be processed. Because AEX chromatographic membranes can be loaded at high flowrate, the purification of 350mL could be performed in less than 3 hours. At that scale, the disposable membrane modules serving for the separation are relatively inexpensive. Therefore the stainless steel membrane housing system represented the main cost of the step. However, this cost represented only 4% of the overall production platform costs and can be amortised with the production of bigger quantities or different mAbs. Yet, a significant improvement to the step would be to design a disposable plastic capsule for the purification membrane. Chromatographic steps do not require any automation as simple peristaltic pumps could be used to load, wash and eluate the material at the desired flowrates.

Despite the contamination that occurred at the 20L scale transient culture, the

developed platform is thought to be capable of generating 1g of pure mAb. The integration of the different processing options simplified the overall process of production by limiting the intervention of the experimenter between the different processing steps. The platform was made of disposable/reusable components to minimise the use of stainless steel. Each processing step can, however, be presented as a disposable solution to prevent product cross contamination or eliminate a non negligible volume of buffers. Yet, some components of the platform such as the sorbents of chromatography or the cassette of filtration, could be reused to minimise the cost of the platform.

Chapter 6

General discussion and future directions

This Chapter summarises the contribution of my research, and attempt to inscribe it within the current state and needs of biopharma industry towards the development of new biopharmaceuticals. Directions for future work are also considered.

The core problem of today's biopharmaceutical companies is the lack of productivity in the early stages of drug development. The industry spends far more on R&D and produces far fewer new molecules than it did 20 years ago. It is not the modes of production that changed, but the market (165). If blockbuster medicines have helped large populations to face generic health disorders or illnesses, their efficacy varies from one patient to another. Yet, the demand for more effective medicines is rising, and as the population ages, new medical needs emerge. In other words, the market moves towards personalised medicine *i.e.* more effective drugs designed for much smaller numbers of patients. In this context, it becomes necessary to amplify the relevance of early molecules testing by bringing more molecules into testing while decreasing the costs of early production. At the moment, production systems rely on long, tedious and product specific developments. If those are a prerequisite to a large scale production, they are incompatible with the quick and cost effective production of numerous molecules to support preclinical development.

6.1 The development of a transient production platform

This thesis aimed to develop a platform for the production of milligrams to gram of recombinant protein. The idea was that the platform should accommodate the flexibility required for the quick and cost effective production of multi-products. In theory, transient expression technology represented the ideal expression platform as it allows for the conversion of recombinant gene into protein products in days. However, transient expression processes utilises functionally heterogeneous parental cell populations, whose intrinsic genetic heterogeneity has not been exploited to derive a host cell clone intrinsically suited to the expression of the product. As a result transient production processes are generally low yielding (73). Moreover, the processes are relatively complex with numerous interdependent variables underpinning the overall process efficiency. A sub-project within this thesis was therefore to take into account the intrinsic complexity associated with transient production processes to develop a transient expression platform in CHO-S cells. This development was done rationally by sequentially optimizing three groups of variables involved either in the early transfection process, the transgene expression and cell maintenance, and/or culture process design. Using an integrated set of DOE tools, the whole platform could be developed in 45 days and the process yield be improved by more than 200 fold, yielding approximately 88 and 82mg L^{-1} of an IgG2 and an IgG1 respectively in 10 days. Product quality attributes such as charged variants and percentage of monomers were also checked. Interestingly, compared to a stably expressed product, a transiently expressed mAb showed a larger proportion of correctly charged molecules.

The process developed at small scale was scaled up to 5, then 20L. The scale-up proved to be particularly challenging as simple operations at lab scale were difficult to reproduce at large scale. Furthermore, mammalian cell culture can be characterised as very demanding processes where a single mistake can rapidly be associated with disastrous consequences on the process at large scale. At 5L, the non adapted routine passage method resulted in significantly lower yields than the ones achieved at lab scale. At the 20L scale, the limited capacity of the centrifuge used for the cell

concentration step resulted in the multiplication of the operations and the eventual contamination of the cultures. First days post transfection showed that the transfection process succesfully occurred and that the cells were expressing the reporter protein in the ranges observed at lab scale. This event stressed the need of including scalability as a parameter during the development of a process at lab scale, and the need for testing the protocol with a simulant at large scale prior to running the definitive process.

6.2 The development and integration of a purification platform

It was necessary at this stage to also evaluate the potential of integration of an appropriate purification platform with the upstream transient cell culture. The major challenge brought by transient culture is the relatively low concentration of product within the cell culture supernatant. If the high selectivity of Protein A guaranteed the capture of the majority of the product on a chromatography column, loading the whole supernatant in one go could not be accomplished without significantly long loading time, and eventually the contamination of the feed. Moreover, the efficiency of the capture on Protein A strongly depended on the residence time within the chromatography column, the capacity of the overall system tending to the maximal capacity with longer residence and therefore loading times. A solution to the problem required (i) increasing the volume of Protein A sorbent so the column could be loaded at high flowrates without losing product and (ii) scaling out the process. An advantage of this solution is its flexibility. Protein A sorbent based capture process are relatively robust for cell culture supernatant variations. Although providing excellent purification performance, some host cell proteins, as well as nucleic acids can co-elute with the product. By combining a cation exchange, followed by an anion exchange step, it was possible to considerably reduce these contaminants to an acceptable level. The cation exchange step was performed using S HyperCel sorbent as it was possible to load the Protein A eluate straight onto the column without the

need to titrate the Protein A eluate first. In this work, the Protein A capture and the CEX steps were performed in series. However, an improved version could be to develop a continuous process by eluting the product from Protein A directly on the S HyperCel sorbent. Final polishing purification and ultrafiltration/diafiltration steps using membrane chromatography and TFF respectively could be performed easily and relatively quickly.

The purification platform presented here was developed for monoclonal antibodies as they represent the current major class of biopharmaceuticals. It is clear that some changes should be introduced for the purification of other classes of recombinant proteins as the capture by Protein A sorbent is specific to mAbs. However, this work also provides a methodology framework that could be used in the future to develop new capture processes and subsequently adapt the platform as needed.

6.3 Economical considerations related to the use of the platform

For each processing step, a cost of goods analysis was carried out. Parameters such as consumables, the hours of labour (estimated at £10 hr^{-1}) and the specific investment in equipment required to generate 1g of purified mAb were taken into account. General equipment routinely used in cell line development such as bioreactor units, or cell culture incubators however, were not taken into account in the analysis. Capital charges such as energy costs and footprint area in the building were not taken into account as they vary strongly across different companies and are usually not communicated to the public. Results are presented in Figure 6.1.

DNA production costs were driven at 60% by the costs of the single-use DNA purification kits. This step also represented a labour intensive process as 43 hours of work in total was required to generate the amount of DNA required. The development of the transient gene expression process was the most labour-intensive step in the overall process with a total of 50 hours of work required. The transient gene expression process costs were driven by the cost of basal media (48%) as well as the

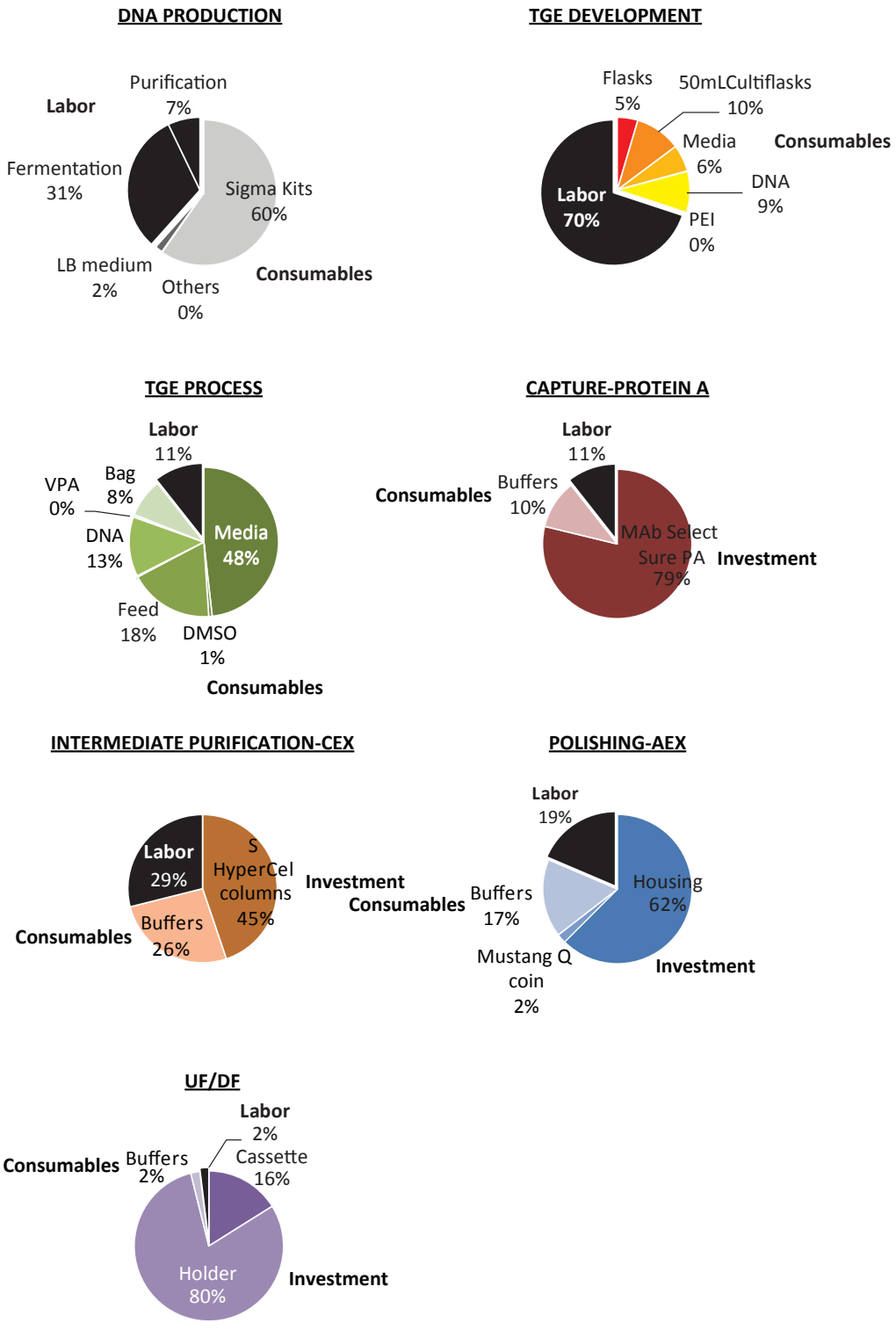


Figure 6.1: Breakdown of COG for performing each processing step once. The relative costs of consumables, specific materials and hours of labour for each processing step are represented. While consumables and labour were driving the costs of upstream processing steps, investment in specific equipment.material represented the main cost of the downstream processing applications.

feed media (18%). The cost of DNA also represented a significant part in the overall cost of this step. The costs of the downstream applications were on the contrary driven by the specific investment in reusable equipment. S HyperCel sorbent being relatively inexpensive, the relative costs of buffers and labour required for the intermediate purification steps were higher compared to the capture step. Indeed the chromatographic sorbents MabSelect Sure Protein A and S HyperCel accounted for 79% and 45% of the overall costs of the capture and intermediate purification steps respectively. The polishing purification and UF/DF steps costs were driven by the investment in the specific coin holder, and cassette housing, respectively.

The breakdown of the COG for the production of 1g of one mAb is detailed in Figure 6.2, A. The transient gene expression process represented the most costly operation at 33% of the overall costs. Despite the relatively high cost of Protein A sorbent, the capture step only represented 10% of the costs. The UF/DF step however accounted for more than a quarter of the overall costs. This is primarily due to the investment in a cassette holder. In fact, the investment in equipment and other reusable components represented 39.9% of the overall costs for the production of one mAb. The relative cost of consumables, labour and investment required to produce 1g of 10 different mAbs, or 10g of one mAb, were then summed-up with respect to each processing step, and compared with the costs associated with the production of one gram of only one mAb. As shown in Figure 6.2, the steps that required significant initial investment in equipment and/or labour now only represent a relatively low part of the overall costs (14%). In fact it appeared clearly that the costs of production were now mostly driven by the consumables necessary to perform the DNA production (21%) and the transient gene expression processes (57%). In other words, transient expression titres represented a major bottleneck in the platform and increasing titres would result in significant savings in the amount of consumables required to generate 1g of protein.

A simulation of what the cost of the production will be with respect to an increase in titres has been performed and the results are presented in Figure 6.3. Doubling transient titres would result in 22.4% reduction of the overall production

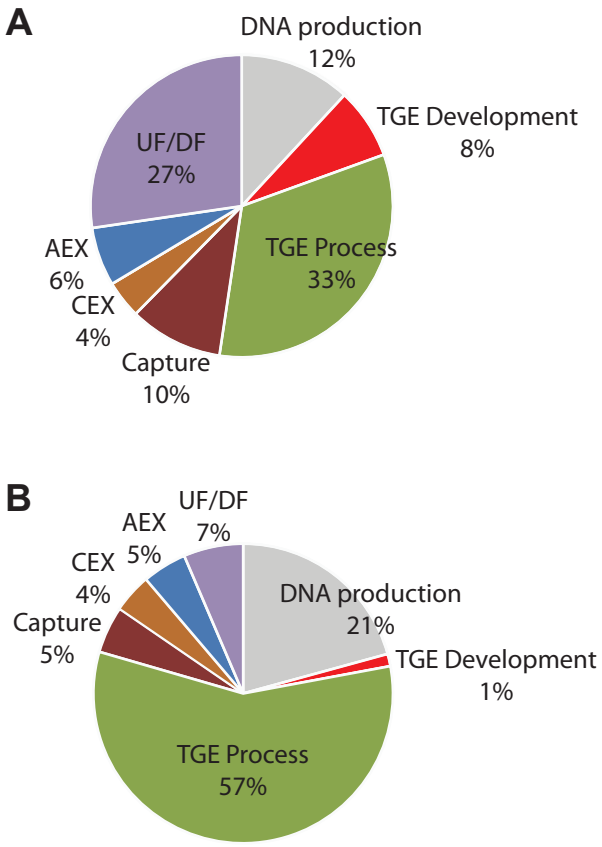


Figure 6.2: Breakdown of COG for the production of one mAb and 10 mAbs. The relative cost of consumables, labour and investment required to produce 1g of 10 different mAbs, or 10g of one mAb (B), were then summed-up with respect to each processing step, and compared with the costs associated with the production of one gram of only one mAb (A).

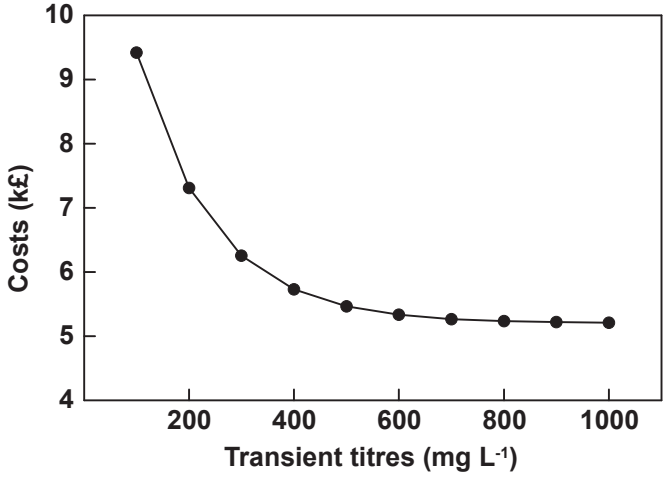


Figure 6.3: Cost benefits provided by an increase in transient expression titres. Consumables required in the DNA production step and transient gene expression process represented 36.7% of the overall production costs. By increasing transient titres, the amount of consumables, and therefore the costs to generate 1g of mAb could be significantly reduced.

cost. Increasing titres to 600mg L^{-1} would result in 43.4% savings. However increasing transient titres above 600mg L^{-1} would not result in significant savings anymore. At this stage, production costs become incompressible and the cost of production is mainly driven by downstream applications.

Finally the resources required to generate 1g of protein using the development of a stable producing clone, or the transient expression process, were compared (Figure 6.4). Genmab B.V. company evaluates at 13 months and 1.4 full time equivalent salary (28,000£) the resources required to manually develop a stable cell clone (166). Following the detailed protocol presented by Lundgren et. al, it was possible to estimate the resources required to develop a producing stable cell clone to £30,737 and 12 months (167). The details of this estimation are presented in Appendices. In comparison, developing a TGE platform and generating 1g of mAb was 5.7 times cheaper and could be achieved 4.24 times quicker (Figure 6.4). Moreover, previous data showed that once developed, a TGE platform can be reused to express other mAbs with no, to minor tweaking of the process. Therefore, it becomes possible to generate the required quantities of mAb in 34 days which reduces the overall production time by 59.5%.

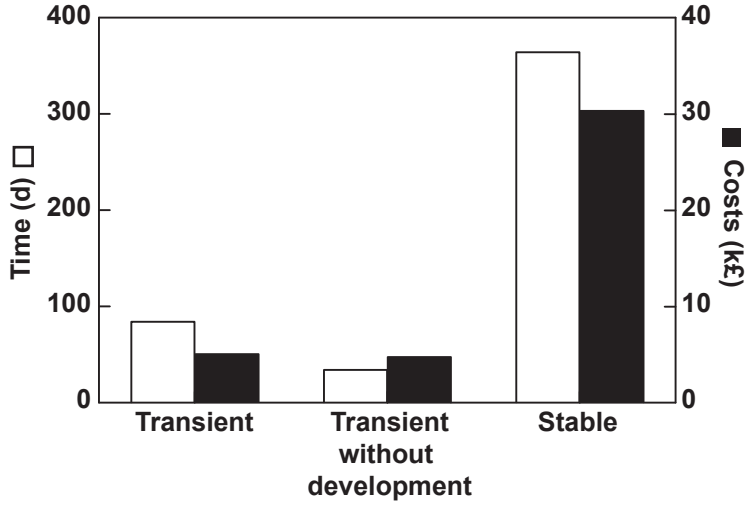


Figure 6.4: Comparison of the resources required to generate 1g of mAb using the developed transient production platform or a stable producing clone. For each step of a protocol detailing the development of a stable producing clone, an estimation of the required consumables and labour was performed (167). The associated costs and time of development were summed up and are presented here. Those numbers correlated with the numbers presented by Gerritsen *et al.* (166). The production of 1g of mAb once the production platform has been developed could be reduced from 84 to 34 days

6.4 Future improvement should focus on the upstream transient expression process

Cost analysis revealed that the obtaining 1g of protein using the developed TGE expression platform was much cheaper than having to generate a stable expressing clone. However the breakdown of COGs also revealed that the costs of the overall platform were driven by the price of consumables required at large scale. The costs associated with the DNA production are hardly compressible as the step is limited not by the bacteria host productivity, but the capacity of the disposable purification kits limited to 15mg. Therefore, further direct improvement should focus on increasing transient titres up to 600mg L^{-1}. This could be done by several means. Firstly, the screening of initial basal discrete variables, such as the culture medium, should be intensified. Indeed, a recent study showed that the nature of culture medium used at transfection could impact transient production titres by more than 400 fold (168). Secondly, the development of a DNA vector with a specific sequence aimed at (i) improving the half life of the vector within the cell, (ii) promoting

the vector transportation into the cell nucleus with, for example, a greater affinity for nuclear import sequences, (iii) introducing a stronger promoter than the actual CMV, (iv) including the genes coding for growth factors favouring the long term cell survival such as p21 or p28 and (v) including sequences coding for regulatory elements. Lastly, the development of new DNA vehicles less cytotoxic than PEI would aid production of higher titres.

At this stage of the development, the platform is characterised by a shift in scale between the upstream transient expression system, and the purification platform. Indeed, transient cultures are relatively low yielding. As a result, litres of culture need to be run to achieve 1g of protein. However this work shows that 1g of mAb could efficiently be purified using 10mL chromatography columns. Therefore, compared to the costs of consumables required in cell culture, the downstream processing material is relatively inexpensive. Moreover, initial equipment investment in cassette or AEX membrane holder can be reused to such a large extent (stainless steel) that their cost can be amortised.

6.5 Towards a fully disposable platform?

The current trend towards disposable manufacturing reflects the pressure on companies to increase their flexibility without exposure to excessive investment risk. Using disposable processing material allows companies to shift resources from upfront capital investment into variables costs of consumables spread over time. However, if the advantages of disposable over stainless steel and in cell culture processes are today widely accepted, the benefits of disposable manufacturing over reusable equipment for downstream applications still remains to be proven. The main advantage of disposable is to eliminate the need for cleaning in place then validation steps that are required with reusable equipment. However, the cost of some material is so high that using it as a disposable component is not economically viable. Moreover, the single-use-then-disposal of equipment capable of being reused highlights some ethical problems. With respect to mAb processing, the cost of downstream processing are generally driven by the cost of the Protein A sorbent. In this case, it is clear

that cleaning and validating the system between uses is cheaper than buying more sorbent. Interestingly, from both a practical and economical point of view, a solution would be to use both properties and introduce continuous processes. Using the Protein A example, performing several cycles of capture and elution throughout the culture would require a limited amount of sorbent while ensuring to use it as its maximum potential before disposal.

6.6 The place of DOE driven experimentation in data and knowledge management

With respect to the complexity of current pharmaceutical processes, managing experimentation and the treatment of data are of primary importance to avoid an information overload and gain maximum knowledge on the product and/or the process from less data. The difficulty shifts from not being able to produce enough material to quickly developing a process able to perform the task. According to Davenport and Pruzak, knowledge is located at the apex of a three levels pyramid (169). The first level consists of the generation of data by experimentation. Efforts conducted during experimentation can be converted to information after proper data analysis. Finally, the information gained allows the scientist to get knowledge on the product and/or process. Current regulatory context shows that it becomes of utmost importance to characterise a production process and the influence of basic operational parameters on the product and process quality attributes. In other words, providing a methodology framework for experimentation and development is now of primary importance. The traditional paradigm of optimizing processes by varying one factor at a time can only result in long, challenging development. Moreover, the limited gain in information and knowledge on the product/process leads to the biased estimation of the potential of a molecule to reach the market later on. Finally, this approach leads to the development of "locked" processes, unadapted to the inherent variability associated with the current biological systems in use. QbD is clearly aimed at addressing this situation by introducing a rational, cost effective

method of development. The approach is to develop as early as possible an understanding of the relationships between the input material, the process parameters and the resultant products quality attributes. This method leads to a more stringent screening of the potential of a molecule to reach the market. By using QbD, that encompasses the use of DOE, the development of a process can be conducted relatively quickly. More importantly, the gain in information is far superior than with traditional methods of experimentation thanks to a thorough and rational statistical analysis of data. The knowledge gained during the development can then be used during the manufacturing to adjust in real time the process parameters to constantly meet the product specifications (Figure 6.5).

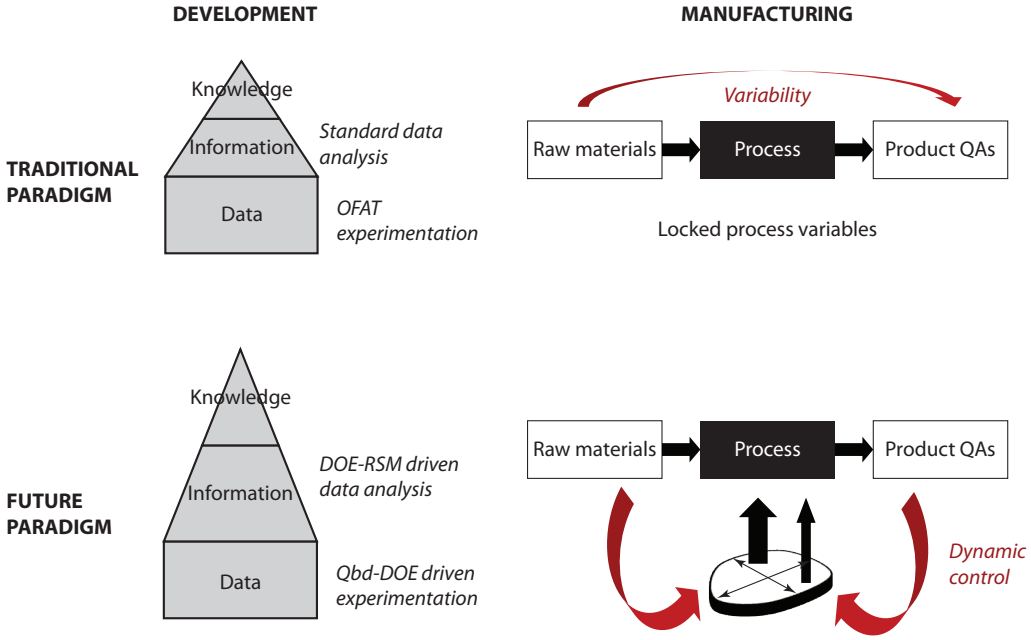


Figure 6.5: Traditional versus future paradigm for the development of biopharmaceutical processes. The traditional paradigm for process development leads to the development of locked processes for which any variability in input material will be transferred to the product. Compared to traditional development, the use of DOE driven experimentation allows the scientist to increase the amount of knowledge construct a design space within which the effect of the variations of input material and critical process parameters on the process/product is known. It is believed that in a near future, it will be possible to continuously adjust the process parameters to compensate for input material variability and constantly produce high quality product. *Adapted from (88).*

It appears that the democratisation of DOE in process development is one of the first steps to address the current challenges the biopharm industry is facing. However, the method is relatively complex and the knowledge of statistics required

for the use of DOE represents a major barrier to its generalisation within the industry. Available statistical packages are too generic and unintuitive. The industry will certainly gain from a dedicated software, or more integration with the current platform of development. It is believed that the method could be almost entirely generalised and automated. The integrated DOE set of tools used in Chapter 3 to develop a TGE platform represents an example of what a generic DOE approach to process optimisation could be. The algorithm, presented in Chapter 2 proves that even complex tasks such as the selection and validation of an empirical model from a set of data could be automated. Moreover, the past few years have seen the arrival on the market of small scale platforms for the expanded throughput of early screening of process conditions such as the Micro-24 MicroBioReactor from Pall Life Sciences. This is the perfect opportunity to democratise the use of DOE during process development. It is believed that the integration of a DOE module to the control software could be of real value.

References

[1] D. J. Adams, "The valley of death in anticancer drug development: a reassessment," *Trends in Pharmacological Science*, vol. 33, no. 4, pp. 173–180, 2012.

[2] B. S. Coller and R. M. Califf, "Traversing the valley of death: a guide to assessing prospects for translational success," *Science Translational Medicine*, vol. 1, no. 10, p. 10cm9, 2009.

[3] R. M. Frederickson, "Escaping the valley of death," *Molecular Therapy*, vol. 20, no. 3, pp. 476–478, 2012.

[4] "Pharmaceutical industry profile 2011," tech. rep., Pharmaceutical Research and Manufacturers of America, 2011. `www.phrma.org/sites/default/files/159/phrma_profile_2011_final.pdf`.

[5] "Innovation or stagnation," tech. rep., Food and Drug Administration, 2004. `www.fda.org/downloads/ScienceResearch/SpecialTopics/CriticalPathOpportunitiesReports/ucm113411.pdf`.

[6] E. S. Langer, "Biomanufacturing outsourcing outlook. industry optimism is on the rise for 2012.," *BioPharm International*, vol. 25, no. 2, pp. 15–16, 2012.

[7] S. L. Morrison, M. J. Johnsom, L. A. Herzenber, and V. T. OI, "Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 81, no. 21, pp. 6851–6855, 1984.

[8] G. Winter and C. Milstein, "Man-made antibodies," *Nature*, vol. 24, no. 349, pp. 293–299, 1991.

[9] R. R. Beerli and C. Rader, "Mining human antibody repertoires," *mAbs*, vol. 2, no. 4, pp. 365–378, 2011.

[10] T. Igawa, H. Tsunoda, T. Kuramochi, Z. Sampei, S. Ishii, and K. Hattori, "Engineering the variable region of therapeutic igg antibodies," *mAbs*, vol. 3, no. 3, pp. 243–252, 2011.

[11] Y. Durocher and M. Butler, "Expression systems for therapeutic glycoprotein production," *Current Opinion in Biotechnology*, vol. 20, no. 6, pp. 700–707, 2009.

[12] M. Matasci, D. L. Hacker, L. Baldi, and F. M. Wurm, "Recombinant therapeutic protein production in cultivated mammalian cells: current status and future prospects," *Drug Discovery Today*, vol. 5, no. 2-3, pp. e37–e42, 2008. electronic article.

[13] S. Soukharev, D. Hammond, N. M. Ananyeva, J. A. Anderson, C. A. Hauser, S. Pipe, and E. L. Saenki, "Expression of factor viii in recombinant and transgenic systems," *Blood cells, Molecules and Diseases*, vol. 28, no. 2, pp. 234–248, 2002.

[14] P. Thomas and T. G. Smart, "Hek293 cell line: a vehicle for the expression of recombinant proteins," *Journal of Pharmacological and Toxicological Methods*, vol. 51, no. 3, pp. 187–200, 2005.

[15] C. Rossman, N. Sharp, G. Allen, and D. Gewert, "Expression and purification of recombinant, glycosylated human interferon alpha 2b in murine myeloma NSO cells," *Protein Expression and Purification*, vol. 7, no. 4, pp. 335–342, 1996.

[16] M. J. Havenga, L. Holterman, I. Melis, S. Smits, J. Kaspers, E. Heemskerk, R. van der Vlugt, M. Koldjijk, G. J. Schouten, G. Hateboer, K. Brouwer,

R. Vogels, and J. Goudsmit, "Serum free transient production system based on adenoviral vector and PER.C6 technology: high yield and preserved bioactivity," *Biotechnology and Bioengineering*, vol. 100, no. 2, pp. 273–283, 2008.

[17] B. Mei, Y. Chen, J. Chen, C. Q. Pan, and J. E. Murphy, "Expression of human coagulation factor viii in a human hybrid cell line, HBK11," *Molecular Biotechnology*, vol. 34, no. 2, pp. 165–178, 2006.

[18] F. M. Wurm and M. Jordan, "Gene transfer and gene amplification in mammalian cells," in *Gene transfer and expression in mammalian cells* (S. Makrides, ed.), vol. 38 of *New Comprehensive Biochemistry*, pp. 309–335, Amsterdam: Elsevier Sciences B.V., 2003.

[19] L. Chu and D. K. Robinson, "Industrial choices for protein production by large-scale cell culture," *Current Opinion in Biotechnology*, vol. 12, no. 2, pp. 180–187, 2001.

[20] M. Jerums and X. Yang, "Optimization of cell culture media," *BioProcess International*, vol. 3, no. 4, pp. 38–44, 2005. Supplement.

[21] J. H. Nunberg, R. J. Kaufman, R. T. Schimke, G. Urlaub, and L. A. Chasin, "Amplified dihydrofolate reductase genes are localized to a homogeneously staining region of a single chromosome in a methotrexate-resistant chinese hamster ovary cell line," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 78, no. 10, pp. 6043–6047, 1978.

[22] M. E. Brown, G. Renner, R. P. Field, and T. Hassel, "Process development for the production of recombinant antibodies using the glutamine synthetase (GS) system," *Cytotechnology*, vol. 9, no. 1-3, pp. 231–236, 1992.

[23] W. Hu, *Methods and strategies in cell line development.* University of Minnesota Department of Chemical Engineering and Material Sciences, Lecture course, 2010. `http://hugroup.cems.umn.edu/NTU/NTU%202010/Lecture%20notes/Cell%20Culture%20Processes/Cell%20Line%20Development.pdf`.

[24] S. M. Browne and M. Al-Rubei, "Selection methods for high-producing mammalian cell lines," *Trends in Biotechnology*, vol. 25, no. 9, pp. 425–432, 2007.

[25] M. Kim, P. M. O'Callaghan, K. A. Droms, and D. C. James, "A mechanistic understanding of production instability in cho cell lines expressing recombinant monoclonal antibodies," *Biotechnology and Bioengineering*, vol. 108, no. 10, pp. 2434–3446, 2011.

[26] F. M. Wurm and A. Bernard, "Large scale transient expression in mammalian cells for recombinant protein production," *Current Opinion in Biotechnology*, vol. 10, no. 2, pp. 156–159, 1999.

[27] F. M. Wurm, "Transient gene expression: an emergent technology for large scale preclinical and clinical production?," in *Cell line development and engineering*, (Prague), 2010.

[28] T. Azzam and A. J. Domb, "Current developments in gene transfection agents," *Current Drug Delivery*, vol. 1, no. 2, pp. 165–193, 2004.

[29] L. H. Li, R. Shivakumar, S. Feller, C. Allen, J. M. Weiss, S. Dzekunov, V. Singh, J. Holaday, J. Fratantoni, and L. N. Liu, "Highly efficient, large volume flow electroporation," *Technology in Cancer Research and Treatment*, vol. 1, no. 5, pp. 341–350, 2002.

[30] A. Loyter, G. A. Scangos, and F. H. Ruddle, "Mechanisms of dna uptake by mammalian cells: fate of exogenously added DNA monitored by the use of fluorescent dyes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 79, no. 2, pp. 422–426, 1982.

[31] M. Jordan, A. Schallhorn, and F. M. Wurm, "Transfecting mammalian cells: optimization of critical parameters affecting calcium-phosphate precipitate formation," *Nucleic Acids Research*, vol. 24, no. 4, pp. 596–601, 1996.

[32] J. Lindell, P. Girard, N. Muller, M. Jordan, and F. M. Wurm, "Calfection: a novel gene transfer method for suspension cells," *Biochimica et Biophysica acta*, vol. 1676, no. 2, pp. 155–161, 2004.

[33] L. Baldi, N. Muller, S. Picasso, R. Jacquet, P. Girard, H. P. Thanh, E. Derow, and F. M. Wurm, "Transient gene expression in suspension HEK-293 cells: application to large-scale protein production," *Biotechnology Progress*, vol. 21, no. 1, pp. 148–153, 2005.

[34] S. Chenuet, D. Martinet, N. Besuchet-Schmutz, M. Wicht, N. Jaccard, A. C. Bon, M. Derouazi, D. L. Hacker, J. S. Beckmann, and F. M. Wurm, "Calcium phosphate transfection generates mammalian recombinant cell lines with higher specific productivity than polyfection," *Biotechnology and Bioengineering*, vol. 101, no. 5, pp. 937–945, 2008.

[35] N. Muller, *Transient gene expression for rapid protein production: studies and optimizations under serum-free conditions.* PhD thesis, Ecole Polytechnique Federale de Lausanne, 2005.

[36] T. Kost and J. Condreay, "Recombinant baculoviruses as mammalian cell gene-delivery vectors," *Trends in Biotechnology*, vol. 20, no. 4, pp. 173–180, 2002.

[37] C. Liu, B. Dalby, W. Chen, J. M. Kilzer, and H. C. Chiou, "Transient transfection factors for high-level recombinant protein production in suspension cultured mammalian cells," *Molecular Biotechnology*, vol. 39, no. 2, pp. 141–153, 2008.

[38] W. C. Tseng, F. R. Haselton, and T. D. Giorgio, "Transfection by cationic liposomes using simultaneous single cell measurements of plasmid delivery and transgene expression," *The Journal of Biological Chemistry*, vol. 272, no. 41, pp. 25641–25647, 1997.

[39] P. L. Felgner, T. R. Gadek, M. Holm, R. Roman, H. W. Chan, M. Wenz, J. P. Northrop, G. M. Ringold, and M. Danielsen, "Lipofection: a highly efficient, lipid-mediated dna-transfection procedure," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 84, no. 21, pp. 7413–7417, 1987.

[40] I. S. Zuhorn, R. Kalicharan, and D. Hoekstra, "Lipoplex-mediated transfection of mammalian cells occurs through the cholesterol-dependent clathrin-mediated pathway of endocytosis," *Journal of Biological Chemistry*, vol. 277, no. 20, pp. 18021–18028, 2002.

[41] E. J. Schlaeger and K. Christensen, "Transient gene expression in mammalian cells grown in serum-free suspension culture," *Cytotechnology*, vol. 30, no. 1-3, pp. 71–83, 1999.

[42] M. Derouazi, P. Girard, F. Van Tilborgh, K. Iglesias, N. Muller, M. Bertschinger, and F. M. Wurm, "Serum-free large-scale transient transfection of CHO cells," *Biotechnology and Bioengineering*, vol. 87, no. 4, pp. 537–545, 2004.

[43] P. L. Pham, A. Kamen, and Y. Durocher, "Large-scale transfection of mammalian cells for the fast production of recombinant protein," *Molecular Biotechnology*, vol. 34, no. 2, pp. 225–237, 2006.

[44] C. Raymond, R. Tom, S. Perret, P. Moussouami, D. L'Abbe, G. St-Laurent, and Y. Durocher, "A simplified polyethylenimine-mediated transfection process for large-scale and high throughput applications," *Methods*, vol. 55, no. 1, pp. 44–51, 2011.

[45] J. Rejman, M. Conese, and D. Hoekstra, "Gene transfer by means of lipo- and polyplexes: role of clathrin and caveolae-mediated endocytosis," *Journal of Liposome Research*, vol. 16, no. 3, pp. 237–247, 2006.

[46] K. von Gersdorff, N. Sanders, R. Vandenbroucke, W. E. De Smedt, S.C, and M. Ogris, "The internalization route resulting in successful gene expression depends on both cell line and polyethylenimine polyplex type," *Molecular Therapy*, vol. 14, no. 5, pp. 745–753, 2006.

[47] C. K. Payne, S. A. Jones, C. Chen, and X. Zhuang, "Internalization and trafficking of cell surface proteoglycans and proteoglycan-binding glycans," *Traffic*, vol. 8, no. 4, pp. 389–401, 2007.

[48] I. Kopatz, J. S. Remy, and J. P. Behr, "A model for non-viral gene delivery: through syndecan adhesion molecules and powered by actin," *The Journal of Gene Medecine*, vol. 6, no. 7, pp. 769–776, 2004.

[49] C. L. Grigsby and K. W. Leong, "Balancing protection and release of dna: tools to address a bottleneck of non-viral gene delivery," *Journal of the Royal Society*, vol. 7, no. Suppl 1, pp. 67–82, 2010.

[50] J. Suh, D. Wirtz, and J. Hanes, "Efficient active transport of gene nanocarriers to the cell nucleus," *Applied Biological Sciences*, vol. 100, no. 7, pp. 3878–3882, 2003.

[51] W. T. Godbey, K. K. Wu, and A. G. Mikos, "Tracking the intracellular path of poly(ethylenimine)/DNA complexes for gene delivery," *Proceedings of the National Academy of Sciences of United States of America*, vol. 96, no. 9, pp. 5177–5181, 1999.

[52] A. Remy-Kristensen, J. P. Clamme, C. Vuilleumier, J. G. Kuhry, and Y. Mely, "Role of endocytosis in the transfection of l929 fibroblasts by polyethylenimine/dna complexes," *Biochimica et Biophysica Acta*, vol. 1514, no. 1, pp. 21–32, 2001.

[53] A. S. Tait, C. J. Brown, D. J. Galbraith, M. J. Hines, M. Hoare, J. R. Birch, and D. C. James, "Transient production of recombinant proteins by chinese hamster ovary cells using polyethyleneimine/DNA complexes in combination with microtubule disrupting anti-mitotic agents," *Biotechnology and Bioengineering*, vol. 88, no. 6, pp. 707–721, 2004.

[54] E. Carpentier, S. Paris, A. A. Kamen, and Y. Durocher, "Limiting factors governing protein expression following polyethylenimine-mediated gene transfer in HEK293-EBNA1 cells," *Journal of Biotechnology*, vol. 128, no. 2, pp. 268–280, 2007.

[55] S. Brunner, E. Furtbauer, T. Sauer, M. Kursa, and E. Wagner, "Overcoming the nuclear barrier: cell cycle independent nonviral gene transfer with linear

polyethylenimine or electroporation," *Molecular Therapy*, vol. 5, no. 1, pp. 80–86, 2002.

[56] I. Mortimer, P. Tam, I. MacLachlan, R. W. Graham, E. G. Saravolac, and P. B. Joshi, "Cationic lipid-mediated transfection of cells in culture requires mitotic activity," *Gene Therapy*, vol. 6, no. 3, pp. 403–411, 1999.

[57] X. Han, Q. Fang, F. Yao, X. Wang, J. Wang, S. Yang, and B. Q. Shen, "The heterogeneous nature of polyethylenimine-DNA complex formation affects transient gene expression," *Cytotechnology*, vol. 60, no. 1-3, pp. 63–75, 2009.

[58] H. Pollard, J. S. Remy, G. Loussouarn, S. Demolombe, J. P. Behr, and D. Escande, "Polyethylenimine but not cationic lipids promotes transgene delivery to the nucleus in mammalian cells," *The Journal of Biological Chemistry*, vol. 273, no. 13, pp. 7507–7511, 1998.

[59] D. A. Dean, "Import of plasmid DNA into the nucleus is sequence specific," *Experimental Cell Research*, vol. 230, no. 2, pp. 293–302, 1997.

[60] G. L. Wilson, B. S. Dean, G. Wang, and D. A. Dean, "Nuclear import of plasmid dna in digitonin-permeabilized cells requires both cytoplasmic factors and specific dna sequences," *Journal of Biological Chemistry*, vol. 274, no. 31, pp. 22025–22032, 1999.

[61] D. V. Schaffer, N. A. Fidelman, N. Dan, and D. A. Lauffenburger, "Vector unpacking as a potential barrier for receptor-mediated polyplex gene delivery," *Biotechnology and Bioengineering*, vol. 67, no. 5, pp. 598–606, 2000.

[62] M. Thomas and A. M. Kilbanov, "Non-viral gene therapy: polycation-mediated DNA delivery.," *Applied Microbiology and Biotechnology*, vol. 62, no. 1, pp. 27–34, 2003.

[63] Z. Dai and C. Wu, "How does DNA complex with polyethylenimine with different chain lengths and topologies in their aqueous solution mixtures," *Macromolecules*, vol. 45, no. 10, pp. 4346–4353, 2012.

[64] A. E. Wu, *Examination of transient transfection as a potential means of recombinant protein production.* PhD thesis, 2007.

[65] M. Bertschinger, A. Scherteleib, J. Cevey, D. L. Hacker, and F. M. Wurm, "The kinetics of polyethylenimine-mediated transfection in suspension cultures of chinese hamster ovary cells," *Molecular Biotechnology*, vol. 40, no. 2, pp. 136–143, 2008.

[66] G. Backliwal, M. Hildinger, V. Hasija, and F. M. Wurm, "High-denity transfection with hek-293 cells allows doubling of transient titers and removes need for a priori DNA complex formation with pei," *Biotechnology and Bioengineering*, vol. 99, no. 3, pp. 721–727, 2007.

[67] A. Alshamsan, A. Haddadi, V. Incani, J. Samuel, A. Lavasanifar, and H. Uludaq, "Formulation and delivery of sirna by oleic acid and stearic acid modified polyethylenimine," *Molecular Pharmaceutics*, vol. 6, no. 1, pp. 121–133, 2009.

[68] J. Xiao, X. Duan, Q. Yin, L. Chen, Z. Zhang, and Y. Li, "Low molecular weight polyethylenimine-graft-tween 85 for effective gene delivery: synthesis and in vitro charcateristics," *Bioconjugate Chemistry*, vol. 23, no. 2, 2012.

[69] M. Zheng, Y. Zhong, F. Meng, R. Peng, and Z. Zhong, "Lipoic acid modified low molecular weight polyethylenimine mediates nontoxic and highly potent gene transfection," *Molecular Pharmaceutics*, vol. 8, no. 6, pp. 2434–2443, 2011.

[70] Y. Rajendra, D. Kiseljak, L. Baldi, D. L. Hacker, and F. M. Wurm, "A simple high-yielding process for transient gene expression in CHO cells," *Journal of Biotechnology*, vol. 153, no. 1-2, pp. 22–26, 2011.

[71] R. Haldankar, D. Li, Z. Saremi, C. Baikalov, and R. Deshpande, "Serum-free suspension large-scale transient transfection of CHO cells in WAVE bioreactors," *Molecular Biotechnology*, vol. 34, no. 2, pp. 191–199, 2006.

[72] B. S. Majors, M. J. Betenbaugh, N. E. Pederson, and G. G. Chiang, "Enhancement of transient gene expression and culture viability using chinese

hamster ovary cells overexpressing Bcl-x(L)," *Biotechnology and Bioengineering*, vol. 101, no. 3, pp. 567–578, 2008.

[73] B. C. Thompson, C. R. Segarra, O. L. Mozley, O. Daramola, R. P. Field, P. R. Levison, and D. C. James, "Cell line specific control of polyethylenimine-mediated transient transfection optimized with "Design of experiments" methodology," *Biotechnology Progress*, vol. 28, no. 1, pp. 179–187, 2012.

[74] M. Stettler, *Bioreactors processes based on disposable materials for the production of recombinant proteins from mammalian cells.* PhD thesis, 2007.

[75] X. Sun, H. C. Hia, P. E. Goh, and M. G. Yap, "High-density transient gene expression in suspension-adapted 293 ebna1 cells," *Biotechnology and Bioengineering*, vol. 99, no. 1, pp. 108–116, 2008.

[76] P. L. Pham, S. Perret, B. Cass, E. Carpentier, G. St-Laurent, L. Bisson, A. Kamen, and Y. Durocher, "Transient gene expression in hek293 cells: peptone addition posttransfection improves recombinant protein synthesis," *Biotechnology and Bioengineering*, vol. 90, no. 3, pp. 332–344, 2005.

[77] D. J. Galbraith, A. S. Tait, A. J. Racher, J. R. Birch, and D. C. James, "Control of culture environment for improved polyethylenimine-mediated transient production of recombinant monoclonal antibodies by cho cells," *Biotechnology Progress*, vol. 22, no. 3, pp. 753–762, 2006.

[78] S. Wulhfard, S. Tissot, S. Bouchet, J. Cevey, M. De Jesus, D. L. Hacker, and F. M. Wurm, "Mild hypothermia improves transient gene expression yields several fold in chinese hamster ovary cells," *Biotechnology Progress*, vol. 24, no. 2, pp. 458–465, 2008.

[79] E. Riu, Z. Y. Chen, H. Xu, C. Y. He, and M. A. Kay, "Histone modifications are associated with the persistence or silencing of vector-mediated transgene expression in vivo," *Molecular Therapy*, vol. 15, no. 7, pp. 1348–1355, 2007.

[80] G. Backliwal, M. Hildinger, I. Kuettel, F. Delegrange, D. L. Hacker, and F. M. Wurm, "Valproic acid: a viable alternative to sodium butyrate for enhancing

protein expression in mammalian cell cultures," *Biotechnology and Bioengineering*, vol. 101, no. 1, pp. 182–189, 2008.

[81] "Final report in pharmaceutical cGMPs for the 21st century - a risk based approach," tech. rep., Food and Drug Administration, 2004. www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/Manufacturing/QuestionsandAnswersonCurrentGoodManufacturingPracticescGMPforDrugs/UCM176374.pdf.

[82] J. Avellanet, "Why quality by design?," tech. rep., Cerulean associates LLC, 2008. www.ceruleanllc.com/wp-content/articles/eReport_QbD_Executive_Guide_CERULEAN.pdf.

[83] B. Leishman and F. Hoffman, "Quality by design in clinical projects: if you keep on doing what you always did...," in *Chiltern Seminar*, 2011. Roche.

[84] D. Kenett, "QbD in large molecules: The A-Mab case study by the CMC Biotech Working Group," in *The 2nd QbD conference*, 2010. TEVA pharmaceuticals industries Ltd.

[85] "A-Mab: a case study in bioprocess development, version 2.1," tech. rep., The CMC Biotech Working Group, 2009. www.ispe.org/pqli/a-mab-case-study-version-2.1.

[86] "Q8(R2) pharmaceutical development," tech. rep., International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, 2009. www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q8_R1/Step4/Q8_R2_Guideline.pdf.

[87] R. Mhatre and A. S. Rathore, "Quality by design: an overview of the basic concepts," in *Quality by Design for biopharmaceuticals* (R. Mhatre and A. Rathore, eds.), Wiley series in biotechnology and bioengineering, pp. 1–7, Hoboken: John Wiley & Sons, Inc., 2009.

[88] S. Kozlowski and P. Swann, "Considerations for biotechnology product quality by design," in *Quality by Design for biopharmaceuticals* (R. Mhatre and

A. Rathore, eds.), Wiley series in biotechnology and bioengineering, pp. 9–27, Hoboken: John Wiley & Sons, Inc., 2009.

[89] "Viral safety evaluation of biotechnology products derived from cell lines of human or animal origin Q5A(R1), institution = International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, note = `www.ich.org/fileadmin/public_web_site/ich_products/guidelines/quality/q5a_r1/step4/q5_r1_guideline.pdf`, year = 1999," tech. rep.

[90] J. Coffman, R. Shpritzer, and S. Vicik, "Flocculation of antibody-producing mammalian cells with precipitating solutions of soluble cations and anions," in *Recovery of biological products XII*, (Litchfield, AZ), 2006.

[91] B. Huang, F. F. Liu, X. Y. Dong, and Y. Sun, "Molecular mechanism of the affinity interactions between protein a and human immunoglobulin g1 revealed by molecular simulations," *Journal of Physical Chemistry B*, vol. 115, no. 4, pp. 4168–4176, 2011.

[92] S. Ghose, B. Hubbard, and S. M. Cramer, "Binding capacity differences for antibodies and fc-fusion proteins on protein a affinity chromatographic materials," *Biotechnology and Bioengineering*, vol. 96, no. 4, pp. 768–779, 2007.

[93] J. N. Carter-Franklin, C. Victa, P. McDonald, and R. Fahrner, "Fragments of protein a eluted during protein a affinity chromatography," *Journal of Chromatograpy A*, vol. 1163, no. 1-2, pp. 105–111, 2007.

[94] D. Gao, S. W. Mia, H. B. Wang, S. J. Yao, and H. Bai, "Comparison of protein A affinity media for chromatographic performance and purification of monoclonal antibodies"," *Chinese Journal of Pharmaceutical Biotechnology*, vol. 19, no. 1, pp. 40–44, 2012.

[95] X. Han, A. Hewig, and G. Vedantham, "Recovery and purification of antibody," in *Antibody expression and production* (M. Al-Rubeai, ed.), pp. 305–340, London: Springer, 2011.

[96] T. Arakawa, Y. Kita, and D. Ejima, "Stress-free chromatography: IEC and HIC," *Current Pharmaceutical Biotechnology*, vol. 10, no. 4, pp. 461–463, 2009.

[97] G. V. Annathur, J. J. Buckley, K. Muthurania, and N. Ramasubramanyan, "Application of arginine as an efficient eluent in cation exchange chromatographic purification of a PEGylated peptide," *Journal of Chromatography A*, vol. 1217, no. 24, pp. 3783–3793, 2010.

[98] E. Perez Almodovar, B. Glatz, and G. Carta, "Counterion effects on protein adsorption equilibrium and kinetics in polymer-grafted cation exchangers," *Journal of Chromatography A*, 2012. In press, accepted manuscript.

[99] S. Ghose, M. Jin, J. Liu, and J. Hickey, "Integrated polishing steps for monoclonal antibody purification," in *Process scale purification of antibodies* (U. Gottschalk, ed.), pp. 145–167, Hoboken: John Wiley & Sons, Inc., 2009.

[100] H. F. Liu, M. Junfen, C. Winter, and R. Bayer, "Recovery and purification process development for monoclonal antibody purification," *mAbs*, vol. 2, no. 5, pp. 480–499, 2010.

[101] B. D. Bowes, H. Koku, K. J. Czymmek, and A. M. Lenhoff, "Protein adsorption and transport in dextran-modified ion-exchange media i. adsorption," *Journal of Chromatograpy A*, vol. 1216, no. 45, pp. 7774–7784, 2009.

[102] R. van Reis and A. L. Zydney, "Protein ultrafiltration," in *Encyclopedia of Bioprocess technology: fermentation, biocatalysis and bioseparation* (M. C. Flickinger and S. W. Drew, eds.), pp. 2197–2214, New York: John Wiley & Sons, Inc.

[103] M. Toueille, A. Uzel, J. F. Depoisier, and R. Gantier, "Designing new monoclonal antibody purification processes using mixed-mode chromatography sorbents," *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, vol. 879, no. 13-14, pp. 836–843, 2011.

[104] J. X. Zhou and T. Tressel, "Basic concepts in q membrane chromatography for large scale antibody production," *Biotechnology Progress*, vol. 22, no. 2, pp. 341–349, 2006.

[105] V. Singh, "Disposable bioreactor for cell culture using wave-induced agitation," *Cytotechnology*, vol. 30, no. 1-3, pp. 149–158, 1999.

[106] R. Brecht, "Disposable bioreactors: maturation into pharmaceutical glycoprotein," in *Disposable bioreactors* (R. Eibl and D. Eibl, eds.), vol. 115 of *Advances in Biochemical Engineering/biotechnology*, pp. 1–31, Heidelberg: Springer, 2009.

[107] V. Singh, "Disposable perfusion bioreactor for cell culture." Patent 6544788, USA, 2003.

[108] P. M. Galliher, G. L. Hodge, M. Fisher, P. Guertin, and D. Mardirosian, "Continuous perfusion bioreactor system." Patent 20090035856, USA, 2009. Xcellerex, Inc.

[109] M. De Jesus and F. M. Wurm, "Manufacturing recombinant proteins in kg-ton quantities using animal cells in bioreactors," *European Journal of Pharmaceutics and Bipharmaceuticals*, vol. 78, no. 2, pp. 184–188, 2011.

[110] E. S. Langer, "Improved downstream technologies are needed to boost single-use adoption," *BioProcess International*, vol. 9, no. S2, pp. 6–12, 2011.

[111] B. Rawlings and H. Pora, "Environmental impact of single-use and reusable bioprocess systems," *BioProcess International*, vol. 7, no. 2, pp. 18–26, 2009.

[112] QIAGEN, *QIAGEN Plasmid Plus purification handbook*, january 2009 ed., 2009. `www.qiagen.com/literature/render.aspx?id=104365`.

[113] Millipore, *Mobius® FlexReady solution for TFF*, 2011. `http://www.millipore.com/publications.nsf/a73664f9f981af8c852569b9005b4eee/3014311ba72692a0852575d1004a9ecf/$FILE/DS3125EN00.pdf`.

[114] M. El-Halwagi, "Introduction to sustainability, sustainable design, and process integration," in *Sustainable design through process integration*, pp. 1–14, Waltham: Elsevier, 2012.

[115] F. Nickols, "Inside the process box," tech. rep., BPTrends, 2008. `http://www.bptrends.com/publicationfiles/11-08-ART-Inside%20the%20Process%20Box-Nickols.doc-final.pdf`.

[116] M. Anderson and P. Whitcomb, *DOE Simplified*. New York: Productivity Press, 2nd ed., 2007.

[117] R. Fisher, *The Design of Experiments*. Edinburgh: Oliver and Boyd, 1935.

[118] G. Box and J. Hunter, "The 2^{k-p} fractional factorial designs part i," *Technometrics*, vol. 3, no. 3, pp. 311–351, 1961.

[119] D. Montgomery, *Design and analysis of experiments*. New York: John Wiley and sons, Inc., 4th ed., 1997.

[120] P. Mathews, "Fractional factorial experiments," in *Design of experiments with MINITAB*, Milwaukee: ASQ Quality Press, 2005.

[121] M. Anderson and P. Whitcomb, *RSM Simplified*. New York: Productivity Press, 2005.

[122] G. Box and K. Wilson, "On the experimental attainment of optimum conditions," *Journal of the royal statistical society. Series B (methodological)*, vol. 13, no. 1, pp. 1–45, 1951.

[123] L. Trutna, P. Spagon, E. del Castillo, T. Moore, S. Hartley, and A. Hurwitz, "Process improvement," in *NIST/SEMANTECH e-Handbook of statistical methods*, National Institute of standards and technology, 2006. `http://www.itl.nist.gov/div898/handbook/index.htm`.

[124] Z. Jiang, K. Droms, Z. Geng, S. Casnocha, Z. Xiao, S. Gorfien, and S. Jacobia, "Fed-batch cell culture process optimization," *BioProcess International*, vol. 10, no. 3, pp. 40–45, 2012.

[125] J. Comley, "Design of experiments: useful statistical tool in assay development or vendor disconnect!," *Drug Discovery World*, vol. Winter 2009/10, pp. 41–50, 2009. Electronic article, `http://www.ddw-online.com/media/32/2273/winter-2009-design-of-experiments.pdf`.

[126] M. A. Babyak, "What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models," *Psychosomatic Medicine*, vol. 66, no. 3, p. 411421, 2004.

[127] W. Guthrie, J. Filiben, and A. Heckert, "Process modeling," in *NIST/SEMANTECH e-Handbook of statistical methods*, National Institute of standards and technology, 2006. Electronic book, `http://www.itl.nist.gov/div898/handbook/index.htm`.

[128] G. Steenackers, F. Presezniak, and P. Guillaume, "Development of an adaptive response surface method for optimization of computation-intensive models," *Computers and Industrial Engineering*, vol. 57, no. 3, pp. 847–855, 2009.

[129] P. Sharkey and K. Warwick, "Automatic nonlinear predictive model-construction algorithm using forward regression and the press statistic," *Control Theory and Applications, IEE Proceedings*, vol. 150, no. 3, pp. 245–254, 2003.

[130] E. Pedhazur, *Multiple regression in behavorial research.* Crawfordsville: Wadsworth Publishing, 3rd ed., 1997.

[131] D. Behnken and N. Draper, "Residuals and their variance patterns," *Technometrics*, vol. 14, no. 1, pp. 101–111, 1972.

[132] S. Shapiro and M. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.

[133] S. Shapiro, M. Wilk, and H. Chen, "A comparative study of various tests for normality," *Journal of the American statistical association*, vol. 63, no. 324, pp. 1343–1372, 1968.

[134] J. Filiben, "Explanatory data analysis," in *NIST/SEMANTECH e-Handbook of statistical methods*, National Institute of standards and technology, 2006. Electronic book, `http://www.itl.nist.gov/div898/handbook/index.htm`.

[135] G. E. P. Box and C. D. R., "An analysis of transformations," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 26, no. 2, pp. 211–252, 1964.

[136] R. Myers, D. Montgomery, and C. Anderson-Cook, *Response surface methodology: process and product optimization using designed experiments*. Wiley series in probability and statistics, Hoboken: John Wiley and sons, Inc., 3rd ed., 2009.

[137] W. ADAMS, "Programmer at stat-ease, personal communication, e-mail," April 2010.

[138] S. O. Hwang and G. M. Lee, "Nutrient deprivation induces autophagy as well as apoptosis in chinese hamster ovary cell culture," *Biotechnology and Bioengineering*, vol. 99, no. 3, pp. 678–685, 2008.

[139] K. Hansen, M. Kjalke, R. P. B., L. Kongerslec, and M. Ezban, "Proteolytic cleavage of recombinant two-chain factor VIII during cell culture production is mediated by protease(s) from lysed cells. the use of pulse labelling directly in production medium," *Cytotechnology*, vol. 24, no. 3, pp. 227–234, 1997.

[140] I. Bramke and J. F. Burke, "Cell culture medium." Patent 20110104754A1, USA, 2011. Genetix Limited.

[141] J. Peirera, Y. Rajendra, L. Baldi, D. L. Hacker, and F. M. Wurm, "Transient gene expression with CHO cells in conditioned medium: a study using TubeSpin® bioreactors," in *22nd European Society for Animal Cell Technology (ESACT)*, vol. 5, BMC Proceedings, 2011.

[142] K. Sunley, T. Tharmalingam, and M. Buttler, "Cho cells adapted to hypothermic growth produce high yields of recombinant beta-interferon," *Biotechnology Progress*, vol. 24, no. 4, pp. 898–906, 2008.

[143] S. Becerra, J. Berrios, N. Osses, and C. Altamirano, "Exploring the effect of mild hypothermia on cho cell productivity," *Biochemical Engineering Journal*, vol. 60, pp. 1–8, 2012.

[144] H. Melkonyan, C. Sorg, and M. Klempt, "Electroporation efficiency in mammalian cells is increased by dimethyl sulfoxide (dmso)," *Nucleic Acides Research*, vol. 24, no. 21, pp. 4356–4357, 1996.

[145] M. Jordan and F. M. Wurm, "Transfection of adherent and suspended cells by calcium phosphate," *Methods*, vol. 33, no. 2, pp. 136–143, 2004.

[146] J. Ye, V. Kober, M. Tellers, Z. Naji, P. Salmon, and J. F. Markusen, "High-level protein expression in scalable CHO transient transfection," *Biotechnology and Bioengineering*, vol. 103, no. 3, pp. 542–551, 2009.

[147] M. Hildinger, G. Backliwal, and F. M. Wurm, "Use of valproic acid for enhancing production of recombinant proteins in mammalian cells." Patent 20090023186, 2009. ExcellGene SA.

[148] N. Arden and M. J. Betenbaugh, "Life and death in mammalian cell culture: strategies for apoptosis inhibition," *TRENDS in Biotechnology*, vol. 22, no. 4, pp. 174–180, 2004.

[149] R. Klein, B. Ruttkowski, E. Knapp, B. Salmons, W. H. Gnzburg, and C. Hohenadl, "Wpre-mediated enhancement of gene expression is promoter and cell line specific," *Gene*, vol. 10, no. 372, pp. 153–161, 2006.

[150] R. Fahrner, H. Knudsen, C. Basey, W. Galan, D. Feuerhelm, M. Vanderlaan, and G. Blank, "Industrial purification of pharmaceutical antibodies: development, operation and validation of chromatography processes.," *Biotechnology and genetic engineering reviews*, vol. 18, pp. 301–327, 2001.

[151] W. Wang, S. Singh, N. Li, M. Toler, K. King, and S. Nema, "Immunogenicity of protein aggregates - concern and realities," *International Journal of Pharmaceutics*, vol. 431, no. 1-2, pp. 1–11, 2012.

[152] Z. Xu, J. Li, and J. Zhou, "Iprocess development for robust removal of aggregates using cation exchange chromatography in monoclonal antibody purification with implementation of quality by design," *Preparative Chemistry and Biotechnology*, vol. 42, no. 2, pp. 183–202, 2012.

[153] B. B. Boonyaratanakornkit, C. B. Park, and D. S. Clark, "Pressure effects on intra- and intermolecular intercations within proteins.," *Biochimica et Biophysica Acta*, vol. 1595, no. 1-2, pp. 235–249, 2002.

[154] K. Diepold, K. Bomans, M. Wiedmann, B. Zimmermann, A. Petzold, T. Schlothauer, M. Molhoj, D. Treusch, and P. Bulau, "Simultaneous assessment of Asp isomerization and Asn deamidation in recombinant antibodies by LC-MS following incubation at elevated temperatures," *PLoS ONE*, vol. 7, no. 1, p. e30295, 2012.

[155] GE Healthcare, "Screening and optimization of the loading conditions on capto s." online, www.gelifesciences.com, 2006. Application note 28-4078-16 AA.

[156] GE Healthcare, "Capto s cation exchange exchanger for post protein a purification of monoclonal antibodies." online, www.gelifesciences.com, 2006. Application note 28-4078-17 AA.

[157] PALL, "Q and s hypercel sorbents." online, www.pall.com, 2009. Application overview USD2591(1).

[158] P. Haddad and P. Jackson, *Ion chromatography principles and applications*, vol. 46 of *Journal of chromatography library*. Amsterdam: Elsevier, 1990.

[159] J. S. Philo and T. Arakawa, "Mechanisms of protein aggregation," *Current Pharmaceutical Biotechnology*, vol. 10, no. 4, pp. 348–351, 2009.

[160] "Pharma 2020: Supplying the future," tech. rep., PwC, 2011. `www.pwc.com/gx/en/pharma-life-sciences/pharma-2020-supplying-the-future.jhtml`.

[161] T. Peuker and D. Eibl, "Biopharmaceutical manufacturing facilities integrating single-use systems," in *Single-use technology in biopharmaceutical manufacture* (R. Eibl and D. Eibl, eds.), ch. 12, Hoboken: John Wiley & Sons, Inc., 2010.

[162] Y. Li, D. W. Kahn, O. Galperina, E. Blatter, R. Luo, Y. Wu, and G. Zhang, "Development of a platform process for the purification of therapeutic monoclonal antibodies," in *Process scale purification of antibodies* (U. Gottshalk, ed.), ch. 9, Hoboken: John Wiley & Sons, Inc., 2009.

[163] S. B. Hari, H. Lau, V. I. Razinkov, S. Chen, and R. F. Latypov, "Acid-induced aggregation of human monoclonal IgG1 and IgG2: Molecular mechanism and the effect of solution composition," *Biochemistry*, vol. 49, no. 43, pp. 9328–9338, 2010.

[164] S. Vunnum, G. Vedantham, and B. Hubbard, "Protein a-based affinity chromatography," in *Process scale purification of antibodies* (U. Gottschalk, ed.), ch. 4, Hoboken: John Wiley & Sons, Inc., 2009.

[165] A. Busch, "Pharma's and biotech's future: R& d alliances," in *Biopharma Asia Convention*, (Singapore), 2012.

[166] J. Gerritsen, "Optimal selection of cell lines producing biopharmaceutical human IgGs." Optimizing mammalian cell lines, Conference Boston, 22-24 August 2011. Genmab B.V.

[167] K. Lindgren, A. Salmen, M. Lundgren, L. Bylund, A. Ebler, E. Faldt, L. Sorvik, C. Fenge, and U. Skoging-Nyberg, "Automation of cell line development," *Cytotechnology*, vol. 59, pp. 1–10, 2009.

[168] J. Pereira, Y. Rajendra, L. Baldi, D. Hacker, and F. Wurm, "Transient gene expression with cho cells in conditioned medium: a study using TubeSpin® bioreactors," *BMC Proceedings*, vol. 5, no. Suppl 8, p. p38, 2011.

[169] T. H. Davenport and L. Pruzak, *Working knowledge: how organizations manage what they know?* Boston: Haravard Business Review Press, 1998.

Appendix A

Statistical modules library

All the modules have been written in Mathematica as "packages" that the experimenter calls within the software.

A.1 Mother file - Algorithm

```
SetDirectory[
   Input["Please insert the path of the directory that
      contains the \
excel file of your data. The format of the path has to be \
''C:/...''"]] data =
 Import[Input[
    "Please insert the name of the file from which you would
       like to \
analyze the data. The format of the name has to be \
''filename.xls''"]] Needs["ModelsGenerator`"]
Candidates = ModelsGenerator[data];

Needs["ModelsEqGenerator`"]
ModelsEqGenerator[data];

Needs["ShapiroWilkTest`"]
Needs["BoxCoxTransformation`"]
Needs["DataTransformation`"]
TableFullNormalResModels = Table[
   NormalResModel = Candidates[[n]];
   NormalResModelEq = ModelsEqGenerator[data][[n]];
   If[ShapiroWilkTest[NormalResModel] == "Yes",
    (
     BoxCoxTest =
      BoxCoxTransformation[NormalResModel, NormalResModelEq
         ];
     NeedTransformation = If[BoxCoxTest[[1, 2, 1]] == 1, "No
        ", "Yes"];
      TModelResNormal = If[NeedTransformation == "Yes",
       (
```

```
        Tdata = DataTransformation[data, BoxCoxTest[[1, 2,
            1]]];
        l = NormalResModel["BestFitParameters"];
        ListParameters = Table[l[[n, 1]], {n, Length[l]}];
        ListFactors =
         Array[Subscript[
            Global`x, #] &, {Length[Global`data[[1, 1, 1
                ;;]]] - 1}];
        TModel =
         NonlinearModelFit[Tdata, NormalResModelEq,
            ListParameters,
          ListFactors];
        ShapiroWilkTest[TModel]
        ), "Null"
       ];
     If[TModelResNormal == "No", TModel]
     ),
   NormalResModel], {n, Length[Candidates]}];

TableNormalResModels = Cases[TableFullNormalResModels,
  Except[Null]];

Needs["ModelSignificance`"]
TableSignificantModels =
 Cases[Table[
   If[ModelSignificance[TableNormalResModels[[n]]] == "Yes
      ",
    TableNormalResModels[[n]]]
   , {n, Length[TableNormalResModels]}
   ],
  Except[Null]
  ];

Needs["ModelOverfitting`"]
GoodFittingModels =
 Cases[
  Table[If[
    ModelOverfitting[TableSignificantModels[[n]]] ==
     "Not overfitted", TableSignificantModels[[n]]], {n,
    Length[TableSignificantModels]}],
  Except[Null]
  ];


Needs["ModelStatistics`"]
PotentialModels =
 Cases[Table[
   If[ModelStatistics[GoodFittingModels[[n]]][[1, 2, 4]] <
     0.2,
```

```
      GoodFittingModels[[n]]],
     {n, Length[GoodFittingModels]}],
    Except[Null]];
PosModelEq =
  Flatten[Table[
    Position[TableFullNormalResModels, PotentialModels[[n
       ]]], {n,
     Length[PotentialModels]}]];
ModelFullEq =
  Table[ModelsEqGenerator[data][[PosModelEq[[n]]]], {n,
    Length[PotentialModels]}];

Needs["ModelOptimums`"]
Needs["ModelContourPlots`"]
Needs["ModelStatistics`"]
Needs["ModelResAnalysis`"]
Needs["BoxCoxTransformation`"]
Table[Grid[{{Style["Model" n, Bold, Large]}, {Style["
   ModelOptimums",
     Bold, Large]}, {ModelOptimums[ModelFullEq[[n]],
     PotentialModels[[n]]]}, {Style["Model Statistics", Bold
        ,
     Large]}, {ModelStatistics[PotentialModels[[n]]]}, {
        Style[
     "Model Equation", Bold, Large],
    "(caution, model may be transformed, see below the box-
       cox \
results)"}, {Normal[PotentialModels[[n]]]},
   {Style["Box-Cox results" , Bold, Large]}, {
      BoxCoxTransformation[
     PotentialModels[[n]], ModelFullEq[[n]]]}, {
        ModelResAnalysis[
     PotentialModels[[n]]]},
   {Style["Contour plots", Large, Bold]}, {ModelContourPlots
      [
     PotentialModels[[n]]]}}, Alignment -> Left, Frame ->
        All,
  FrameStyle -> Thick], {n, Length[PotentialModels]}]
```

A.2 ModelsGenerator

```
BeginPackage["ModelsGenerator`"]
ModelsGenerator::usage="ModelsGenerator generates a table of
    regression models."
Begin["Private`"]
ModelsGenerator[data_]:=(Length[data[[1,1,1;;]]];
lmf=Array[Subscript[Global`x,#]&,{Length[data
    [[1,1,1;;]]]-1}];
lofa=DeleteDuplicates[Flatten[Table[Subscript[Global`x,i]
    lmf,{i,Length[lmf]}]]];
lp=Take[{Global`a,Global`b,Global`c,Global`d,Global`e,Global
    `f,Global`g,Global`h,Global`i,Global`j,Global`k,Global`l,
    Global`m,Global`n,Global`o,Global`p,Global`q,Global`r,
    Global`s,Global`t,Global`u,Global`v,Global`w},Length[lmf
    ]+Length[lofa]+1];
lmp=Take[{Global`a,Global`b,Global`c,Global`d,Global`e,
    Global`f,Global`g,Global`h,Global`i,Global`j,Global`k,
    Global`l,Global`m,Global`n,Global`o,Global`p,Global`q,
    Global`r,Global`s,Global`t,Global`u,Global`v,Global`w},
    Length[lmf]+1];lsp=Drop[lp,Length[lmp]];
MfM=Total[lmp*Prepend[lmf,1]];
MatM=lofa*lsp*Transpose[Tuples[{1,0},Length[lofa]]];
ListeModele=Table[Total[MatM[[1;;,i]]]+MfM,{i,Length[Tuples
    [{1,0},Length[lofa]]]}];
PreTableModele=Table[NonlinearModelFit[data[[1,1;;]],
    ListeModele[[n]],lp,lmf],{n,Length[ListeModele]}]//
    MatrixForm;
Table[l=Chop[PreTableModele[[1,n]]["BestFitParameters"]];
Listp=Cases[Table[If[(n/.l)!=0,n],{n,Flatten[{lmp,lsp}]}],
    Except[Null]];
PModel=NonlinearModelFit[data[[1,1;;]],ListeModele[[n]],
    Listp[[All]],lmf],{n,Length[ListeModele]}]
)
End[]
EndPackage[]
```

A.3 ModelsEqGenerator

```
BeginPackage["ModelsEqGenerator`"]
ModelsEqGenerator::usage="ModelsEqGenerator generates a list
    of all the possible regression models formal equation
    for a set of data."
Begin["Private`"]
ModelsEqGenerator[data_]:=(Length[data[[1,1,1;;]]];
lmf=Array[Subscript[Global`x,#]&,{Length[data
    [[1,1,1;;]]]-1}];
lofa=DeleteDuplicates[Flatten[Table[Subscript[Global`x,i]
    lmf,{i,Length[lmf]}]]];
lp=Take[{Global`a,Global`b,Global`c,Global`d,Global`e,Global
    `f,Global`g,Global`h,Global`i,Global`j,Global`k,Global`l,
    Global`m,Global`n,Global`o,Global`p,Global`q,Global`r,
    Global`s,Global`t,Global`u,Global`v,Global`w},Length[lmf
    ]+Length[lofa]+1];
lmp=Take[{Global`a,Global`b,Global`c,Global`d,Global`e,
    Global`f,Global`g,Global`h,Global`i,Global`j,Global`k,
    Global`l,Global`m,Global`n,Global`o,Global`p,Global`q,
    Global`r,Global`s,Global`t,Global`u,Global`v,Global`w},
    Length[lmf]+1];lsp=Drop[lp,Length[lmp]];
MfM=Total[lmp*Prepend[lmf,1]];
MatM=lofa*lsp*Transpose[Tuples[{1,0},Length[lofa]]];
Table[Total[MatM[[1;;,i]]]+MfM,{i,Length[Tuples[{1,0},Length
    [lofa]]]}])
End[]
EndPackage[]
```

A.4 DataTransformation

```
BeginPackage["DataTransformation`"]
DataTransformation::usage="Transform a set of data with a
    power transformation"
Begin["Private`"]
DataTransformation[data_,RecommendedTransformation_]:=(
If[RecommendedTransformation==0, NewResponse=Log[Global`data
    [[1,1;;,Length[Global`data[[1,1,1;;]]]]]],NewResponse=
    Global`data[[1,1;;,Length[Global`data[[1,1,1;;]]]]]^
    RecommendedTransformation];
Newdata=Partition[Flatten[Riffle[Global`data[[1,1;;,1;;
    Length[Global`data[[1,1,1;;]]]-1]],NewResponse]],Length[
    Global`data[[1,1,1;;]]]]
)
End[]
EndPackage[]
```

A.5 BoxCoxTransformation

```
BeginPackage["BoxCoxTransformation`"]
BoxCoxTransformation::usage="Performs a BoxCox
    transformation i.e. test if data should be transformed"
Begin["Private`"]
BoxCoxTransformation[ModelFullEquation_,ModelFormalEquation_
    ]:=(
l=ModelFullEquation["BestFitParameters"];
Listparam=Table[l[[n,1]],{n,Length[l]}];
GM=Exp[(\!\(
\*UnderoverscriptBox[\(\[Sum]\), \(p = 1\), \(Length[Global`
    data[\([1, \(\(1\)\(;;\)\)]\)]]\)]\(Log[Global`data[\([1,
     p, Length[Global`data[\([1, 1]\)]]]\)]]\)\))/Length[
    Global`data[[1,1;;]]]];
s[x_,w_]:=((x^w)-1)/(w*GM^(w-1));
TTableneg=Table[s[Global`data[[1,1;;,Length[Global`data
    [[1,1]]]]],n],{n,-3,-0.1,0.2}];
TTablepos=Table[s[Global`data[[1,1;;,Length[Global`data
    [[1,1]]]]],n],{n,0.1,3,0.2}];
TTable=Join[TTableneg,TTablepos];
lmf=Array[Subscript[Global`x,#]&,{Length[Global`data
    [[1,1,1;;]]]-1}];
tdata=Partition[Flatten[Riffle[Global`data[[1,1;;,1;;Length[
    Global`data[[1,1]]]-1]],TTable[[2,;;]]]],Length[Global`
    data[[1,1]]]];
ListeLogSSRes=Table[tdata=Partition[Flatten[Riffle[Global`
    data[[1,1;;,1;;Length[Global`data[[1,1]]]-1]],TTable[[n
    ,;;]]]],Length[Global`data[[1,1]]]];
TModel=NonlinearModelFit[tdata,ModelFormalEquation,Listparam
    ,lmf];
Log[TModel["ANOVATable"][[1,1,3,3]]],{n,Length[TTable
    [[1;;,1]]]}];
ListeLambda=Join[Chop[Range[-3,-0.1,0.2]],Chop[Range
    [0.1,3,0.2]]];
Listeres=Partition[Riffle[ListeLambda,ListeLogSSRes],2];
Lambda=Cases[Table[If[Listeres[[n,2]]==Min[ListeLogSSRes],
    Listeres[[n,1]]],{n,Length[Listeres[[1;;,1]]]}],Except[
    Null]];
BestTransformedResp=s[Global`data[[1,1;;,Length[Global`data
    [[1,1]]]]],Lambda[[1]]];
TransformedData=Partition[Flatten[Riffle[Global`data
    [[1,1;;,1;;Length[Global`data[[1,1]]]-1]],
    BestTransformedResp]],Length[Global`data[[1,1]]]];
BestTransformedModel=NonlinearModelFit[TransformedData,
    ModelFormalEquation,Listparam,lmf];
AnovaBTM=BestTransformedModel["ANOVATable"];SSStar=AnovaBTM
    [[1,1,3,3]]*(1+(Quantile[StudentTDistribution[DofResidual
```

```
    ],0.95])^2/DofResidual);
SSStar=AnovaBTM[[1,1,3,3]]*(1+(Quantile[StudentTDistribution
    [DofResidual],0.95])^2/DofResidual);
DofModel=Length[Normal[ModelFullEquation]]-1;
DofResidual=ModelFullEquation["ANOVATable"][[1,1,5,2]]-
    DofModel;
IFun=Interpolation[Listeres];
HighCI=x/.Quiet[FindRoot[IFun[x]== Log[SSStar],{x,3}]];
LowCI=x/.Quiet[FindRoot[IFun[x]== Log[SSStar],{x,-3}]];
If[(LowCI<=1)==True&&(HighCI>=1)==True,
    RecommendedTransformation={1},RecommendedTransformation=
    Nearest[{-1,-0.5,0,0.5},Lambda]];
BoxCoxTransformationGrid=Grid[{{"RecommendedTransformation",
     "Lambda", "LowCI", "HighCI"},{RecommendedTransformation
    [[1]], Lambda[[1]], LowCI, HighCI}},Frame->All]
)
End[]
EndPackage[]
```

A.6 ModelSignificance

```
BeginPackage["ModelSignificance`"]
ModelSignificance::usage="Test if a model is significant at
    95% using an ANOVA test."
Begin["Private`"]
ModelSignificance[ModelFullEquation_]:=(test=
    ModelFullEquation["ANOVATable"];
DofModel=Length[Normal[ModelFullEquation]]-1;
SSModel=test[[1,1,5,3]]-test[[1,1,3,3]];MSModel=SSModel/
    DofModel;
DofResidual=test[[1,1,5,2]]-DofModel;
MSResidual=test[[1,1,3,3]]/DofResidual;
Fisherstatistic=SetPrecision[MSModel/MSResidual,4];Maxi=
    FindMaximum[{PDF[FRatioDistribution[DofModel,DofResidual
    ],x],x>0},{x}];Needs["HypothesisTesting`"];If[
    Fisherstatistic<x/.Maxi[[2]],valeur=CDF[
    FRatioDistribution[DofModel,DofResidual],Fisherstatistic
    ],valeur=1-CDF[FRatioDistribution[DofModel,DofResidual],
    Fisherstatistic]];
If[valeur<= 0.05, "Yes","No"])
End[]
EndPackage[]
```

A.7 ModelOverfitting

```
BeginPackage["ModelOverfitting`"]
ModelOverfitting::usage="Test if the model is overfitted or
    not by checking pvalues of model eq factors at 10%."
Begin["Private`"]
ModelOverfitting[ModelFullEquation_]:=(
If[Select[Drop[ModelFullEquation["ParameterPValues"],Length[
    Global`data[[1,1]]]],#> 0.10&]!={},"Overfitted","Not
    overfitted"])
End[]
EndPackage[]
```

A.8 ModelOptimums

```
BeginPackage["ModelOptimums`"]
ModelOptimums::usage="Generates a grid with the optimal
    sutied response as well as the factors leading to it and
    their associated confidence intervals"
Needs["MultivariateStatistics`"];
Begin["Private`"]
ModelOptimums[ModelFormalEquation_, ModelFullEquation_]:=(
CovM=Chop[ModelFullEquation["CovarianceMatrix"]];
CovM2=SetPrecision[CovM,5];
mnd=Quiet[MultinormalDistribution[ModelFullEquation["
    BestFitParameters"][[All,2]],CovM2]];
l=ModelFullEquation["BestFitParameters"];
Listparam=Table[l[[n,1]],{n,Length[l]}];
fgt=Function[##,Evaluate[ModelFormalEquation]]&@@ {Listparam
    };
fgt2=Apply[fgt,RandomReal[mnd]];
lmf=Array[Subscript[Global`x,#]&,{Length[Global`data
    [[1,1,1;;]]]-1}];
Stats=Table[Maximize[{Apply[fgt,RandomReal[mnd]],Table[Min[
    Global`data[[1,All,n]]]<=lmf[[n]]<=Max[Global`data[[1,All
    ,n]]]],{n,1,Length[lmf]}]},lmf],{10}];
MeanOptResp=Through[{Mean,StandardDeviation}[Stats[[All
    ,1]]]];
MeanOptFactors=Through[{Mean,StandardDeviation}[Stats[[All
    ,2,All,2]]]];
Zut=Table[lmf[[n]],{n,Length[lmf]}];
Zut2=Table[MeanOptFactors[[1,n]],{n,Length[lmf]}];
Zut3=Table[MeanOptFactors[[2,n]],{n,Length[lmf]}];
PL=Flatten[Join[{"  "},{"Response"},{Zut}]];
DL=Flatten[Join[{"Optimum"},{MeanOptResp[[1]]},Zut2]];
TL=Flatten[Join[{"95% C.I."},{MeanOptResp[[2]]},Zut3]];
Grid[{PL,DL,TL},Frame->All]
)
End[]
EndPackage[]
```

A.9 ModelStatistics

```
BeginPackage["ModelStatistics`"]
ModelStatistics::usage="Generate a table with the basic
    statistics associated with a non linear model fit."
Begin["Private`"]
ModelStatistics[ModelFullEquation_]:=(PRESS=Total[Table[(
    ModelFullEquation["FitResiduals"][[m]]/(1-
    ModelFullEquation["HatDiagonal"][[m]]))^2,{m,1,Length[
    Global`data[[1,1;;]]]}]]; test=ModelFullEquation["
    ANOVATable"]; SSModel=test[[1,1,5,3]]-test[[1,1,3,3]];
RSqM=SSModel/test[[1,1,5,3]];
samplesize=Length[Global`data[[1,All]]];
RSqAdjM=1-((samplesize-1)/(samplesize-Length[Normal[
    ModelFullEquation]]))*(1-RSqM); RSqPredM=1-(PRESS/test
    [[1,1,5,3]]); Diff=RSqAdjM-RSqPredM;
Grid[{{"RSq","Ajusted RSq","Pred RSq","Adj RSq - Pred RSq
    "},{RSqM,RSqAdjM,RSqPredM,RSqAdjM-RSqPredM}},Frame->All])
End[]
EndPackage[]
```

A.10 ModelResAnalysis

```
BeginPackage["ModelResAnalysis`"]
ModelResAnalysis::usage="Perform residual analysis on a
    specified model"
Begin["Private`"]
ModelResAnalysis[Model_]:=(
PR=Model["PredictedResponse"];
SR=Model["StudentizedResiduals"];
CD=Table[{PR[[n]],SR[[n]]},{n,Length[PR]}];
TableF=Table[Global`data[[1,1;;,n]],{n,Length[Global`data
    [[1,1]]]-1}];
Coordinates=Table[Partition[Riffle[TableF[[n]],SR],2],{n,
    Length[Global`data[[1,1]]]-1}];
lmf=Array[Subscript[Global`x,#]&,{Length[Global`data
    [[1,1,1;;]]]-1}];
HLPlot=ListPlot[CD,PlotStyle->{PointSize->Large},Frame->True
    ,PlotRange->{-4,4}];
FLPlots=Table[ListPlot[Coordinates[[i,1;;]],PlotRange->{{Min
    [Coordinates[[i,1;;,1]]]-1,Max[Coordinates[[i
    ,1;;,1]]]+1},{-4,4}},AxesOrigin->{Min[Coordinates[[i
    ,1;;,1]]]-1,0},Frame->True,PlotStyle->{PointSize->Large},
    FrameLabel->{lmf[[i]],"Studentized Residuals"}],{i,Length
    [TableF]}];
Grid[{{Style["Residuals Analysis",Bold,Large]},{Grid[{{Style
    ["Influence of individual factors on the response",Bold
    ]},{FLPlots},{"Points in the charts above should
    distribute around the 0 line in a random pattern."} },
    Frame->All]},{Grid[{{Style["Homoscedasticity and
    linearity of the residuals",Bold]},{HLPlot},{"Points in
    the graph above should (i) distribute randomly around the
     0 line in (ii) a random pattern. If these conditions are
     not verified it is advised to not use this model"}},
    Frame->All]}},Frame->All,Alignment->Left]
)
End[]
EndPackage[]
```

A.11 ModelContourPlots

```
BeginPackage["ModelContourPlots`"]
ModelContourPlots::usage"hh"
Begin["Private`"]
ModelContourPlots[ModelFullEquation_]:=(
res=Table[ContourPlot[ModelFullEquation[Subscript[Global`x,
    1],Subscript[Global`x, 2],n],{Subscript[Global`x, 1],Min[
    Global`data[[1,All,1]]],Max[Global`data[[1,All,1]]]},{
    Subscript[Global`x, 2],Min[Global`data[[1,All,2]]],Max[
    Global`data[[1,All,2]]]},ContourLabels->True,
    ColorFunction->"Rainbow", Frame->True,
    ColorFunctionScaling->{0,n/Max[Global`data[[1,All,3]]]},
    BoundaryStyle->Black],{n,Min[Global`data[[1,All,3]]],Max[
    Global`data[[1,All,3]]],Max[Global`data[[1,All,3]]]/10}];
lmf=Array[Subscript[Global`x,#]&,{Length[Global`data
    [[1,1,1;;]]]-1}];
IP1=Grid[{{"Contourplot for the x1x2 interaction"},{
    ListAnimate[res,ControlPlacement->Top,AnimationRunning->
    False]}},Frame->All];
If[Length[lmf]==3,
(
res2=Table[ContourPlot[ModelFullEquation[Subscript[Global`x,
     1],n,Subscript[Global`x, 3]],{Subscript[Global`x, 1],Min
    [Global`data[[1,All,1]]],Max[Global`data[[1,All,1]]]},{
    Subscript[Global`x, 3],Min[Global`data[[1,All,3]]],Max[
    Global`data[[1,All,3]]]},ContourLabels->True,
    ColorFunction->"Rainbow", Frame->True,
    ColorFunctionScaling->{0,n/Max[Global`data[[1,All,2]]]},
    BoundaryStyle->Black],{n,Min[Global`data[[1,All,2]]],Max[
    Global`data[[1,All,2]]],Max[Global`data[[1,All,2]]]/10}];
    IP2=
IP2=Grid[{{"Contourplot for the x1x3 interaction"},{
    ListAnimate[res2,ControlPlacement->Top,AnimationRunning->
    False]}},Frame->All])];
{IP1,IP2}
)
End[]
EndPackage[]
```

Appendix B

Cost analysis of the development of a stable producing clone

This cost analysis has been performed by estimating the cost of consumables and hours of work (10£ hr^{-1}) for each steps involved in the selection of a stable cell line from 2000 clones and including one sub cloning step.

Step	Consumables	Unit	Price	Supplier	Labor (hr)	Cost
CHOK1 culture	Flask 250mL	4	131.1	Sigma	4	40
	CD-CHO	400mL	17.2	Invitrogen		
	Glutamine 200mM	16mL	0.32	Lonza		
Electroporation	DNA	200ug	100	Experimenter	4	40
Culture in mSX	Flask 250mL	4	131.1	Sigma	8	80
	CD-CHO	400mL	17.2	Invitrogen		
	Glutamine 200mM	16mL	0.32	Lonza		
	MSX-3 hydrate	75uM	239.1	SIGMA		
Dilution in 96 well plates	96 well plates	20	2976	Cole-Parmer	8	80
	CD-CHO	192mL	8.6	Invitrogen		
	Glutamine 200mM	8mL	0.15	Lonza		
	MSX-3 hydrate	75uM	191.3	Sigma		
Screening for colonies and ELISA	ELISA plates	12	1134	RD Biotech	16	160
Expansion in 24 well plates	24 well plates	20	3288	Cole-Parmer	8	80
	CD-CHO	2400mL	106	Invitrogen		
	Glutamine 200mM	96mL	1.92	Lonza		
	MSX-3 hydrate	75uM	2293.6	Sigma		

Screening for colonies and ELISA	ELISA plates	5	472.5	RD Biotech	16	160
Expansion in shake flasks	Flasks 125mL	20	143.4	Sigma	6	60
	CD-CHO	600mL	25.8	Invitrogen		
	Glutamine 200mM	24mL	0.45	Lonza		
	MSX-3 hydrate	75uM	358.65	Sigma		
Screening for colonies and ELISA	ELISA plates	5	472.5	RD Biotech	16	160
Dilution in 96 well plates	96 well plates	20	2976	Cole-Parmer	8	80
	CD-CHO	192mL	8.6	Invitrogen		
	Glutamine 200mM	8mL	0.15	Lonza		
	MSX-3 hydrate	75uM	191.3	Sigma		
Screening for colonies and ELISA	ELISA plates	12	1134	RD Biotech	16	160
Expansion in 24 well plates	24 well plates	20	3288	Cole-Parmer	8	80
	CD-CHO	2400mL	106	Invitrogen		
	Glutamine 200mM	96mL	1.92	Lonza		
	MSX-3 hydrate	75uM	2293.6	Sigma		
Screening HPLC	Assay	320	1280	In house	16	160

Expansion in 6 well plates	6 well plates	16	4195.2	Cole-Parmer	8	80
	CD-CHO	480mL	21.2	Invitrogen		
	Glutamine 200mM	16mL	0.32	Lonza		
	MSX-3 hydrate	75uM	459.1	Sigma		
Screening HPLC	Assay	80	320		6	60
Expansion shake flasks	Flask 250mL	20	654	Sigma	6	60
	CD-CHO	5000mL	213.5	Invitrogen		
	Glutamine 200mM	180mL	3.6	Lonza		
Sum			29255.7		154	1540
Total						30736.7

Appendix C

Case study on a Trebuchet

In RSM Simplified, the authors presented the resolution of a problem involving a trebuchet (121). The aim of the study being to analyse the influence of the arm length, the counterweight, and the missile weight on the distance at which the missile could be thrown. To do so, they built up a model scale trebuchet firing racquetballs. Then performed experimentation following a Box Behnken type of RSM design. The results of their experimentation is presented in Table (Table C.1)

To analyse the data the authors goes through a whole procedure involving a lack of fit tests, the analysis of various statistics and finally an ANOVA. Despite identifying that one factor in their model is not significant, they decided to carry on.

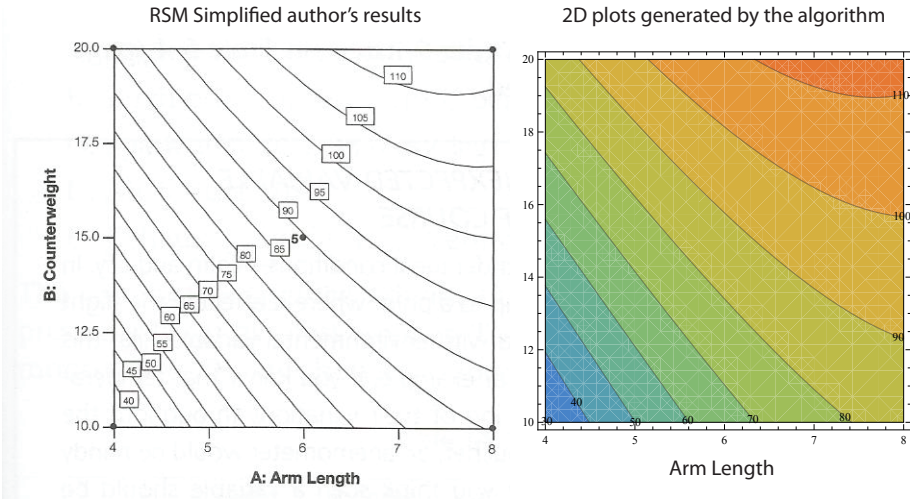
To assess if the algorithm developed in this thesis could be used to automatically conduct the study, the experimental data was fed into the algorithm. Results show that the algorithm successfully selected a quadratic model without the factor identified by the authors as being insignificant. 2D plot generated by the algorithm was very similar to the one generated by the authors. Using this plot is was possible to identify a combination of arm length, counterweight and missile weight to reach any desired thrown distance.

In summary, in this particular case, the algorithm allowed the user to achieve similar results than the authors, without, however, any manual analysis to perform, as the algorithm is automated.

Table C.1: Results table from trebuchet experiment

Arm length (inches)	Counterweight (pounds)	Missile weight (ounce)	Thrown distance (yards)
4	10	2.5	33
8	10	2.5	85
4	20	2.5	86
8	20	2.5	113
4	15	2	75
8	15	2	104
4	15	3	40
8	15	3	89
6	10	2	83
6	20	2	108
6	10	3	49
6	20	3	101
6	15	2.5	88
6	15	2.5	91
6	15	2.5	91
6	15	2.5	87
6	15	2.5	91

Figure C.1: Comparison of the 2D results plot generated by the authors of RSM Simplified book, and the algorithm developed in this thesis



Appendix D

Publication

Cell Line Specific Control of Polyethylenimine-Mediated Transient Transfection Optimized with ''Design of Experiments'' Methodology

Ben C. Thompson, Camille R. J. Segarra, and Olivia L. Mozley
Dept. of Chemical and Biological Engineering, University of Sheffield, Mappin St., Sheffield S1 3JD, U.K.

Olalekan Daramola and Ray Field
Dept. of Cell Sciences, MedImmune, Granta Park, Cambridge, CB21 6GH, U.K.

Peter R. Levison
Dept. of the Technology Development, Pall Europe Ltd, Walton Road, Portsmouth PO6 1TD, U.K.

David C. James
Dept. of Chemical and Biological Engineering, University of Sheffield, Mappin St., Sheffield S1 3JD, U.K.

DOI 10.1002/btpr.715
Published online October 14, 2011 in Wiley Online Library (wileyonlinelibrary.com).

We describe a design of experiments (DoE) response surface modeling strategy to optimize the concentration of basal variables underpinning polyethylenimine (PEI) mediated transfection of different CHO-K1 derived parental cell populations in a chemically defined medium, specifically the relative concentration of linear 25 kD PEI, host CHO cells and plasmid DNA. Using recombinant secreted alkaline phosphatase (SEAP) reporter activity as the modeled response, a discrete simple maximum was predicted for each CHO host cell population. Differences between the modeled optima derived from host cell specific differences in PEI cytotoxicity, such that the PEI:cell interaction effectively limited PEI-DNA polyplex load at a relatively constant PEI:DNA ratio. However, across the three CHO host cell populations, SEAP reporter production was not proportional to plasmid DNA input at the host cell specific predicted basal variable optima. A 10-fold variation in SEAP reporter output per mass of plasmid DNA delivered was observed. To determine the cellular basis of this difference in transient productivity, host CHO cells were transfected with fluorescently labeled polyplexes followed by flow cytometric analysis. Each CHO host cell population exhibited a distinct functional phenotype, varying in the extent of PEI-DNA polyplex binding to the cell surface and degree of polyplex internalization. SEAP production was directly proportional to the level of polyplex internalization and heparan sulfate proteoglycan level. Taken together, these data show that choice of host CHO cell line is a critical parameter, which should rationally precede cell line specific transient production platform design using DoE methodology. © 2011 American Institute of Chemical Engineers *Biotechnol. Prog.,* 28: 179–187, 2012
Keywords: Chinese hamster ovary cells, design of experiments, transient transfection, polyethylenimine

Introduction

Chinese hamster ovary (CHO) cells remain the most commonly used mammalian cell type for biopharmaceutical manufacture.[1,2] In all cases, to generate a CHO cell based production system capable of generating sufficient product for preliminary clinical trials and toxicology testing it has been necessary to use a cell line development process based on stable transfection of a parental host CHO cell population, followed by intensive cell clone isolation and screening operations.[3] This process identifies cell clones that not only have transcriptionally active recombinant genes, but a variety of other functional capabilities that permit high-level manufacture of the protein product.[4–6] The CHO cell clone factory background itself has to be "permissive" to the required function (e.g., Condon et al.[7]), where substantial variation in functional capability between individual CHO cell lines is frequently observed,[8] deriving from acquired genetic heterogeneity within the host cell population.[6] Despite this, cell line functional variability has not been harnessed in the development of alternative transient production processes, which have the potential advantage of a relatively rapid conversion of recombinant gene into protein product. In combination with new disposable processing technology for mammalian cell based production processes,[9] there is a real opportunity for transient production systems to routinely provide early stage product for the clinic and toxicology laboratory rapidly and inexpensively. Indeed, for early process development applications, the use of a transient expression

Correspondence concerning this article should be addressed to D. C. James at d.c.james@sheffield.ac.uk..

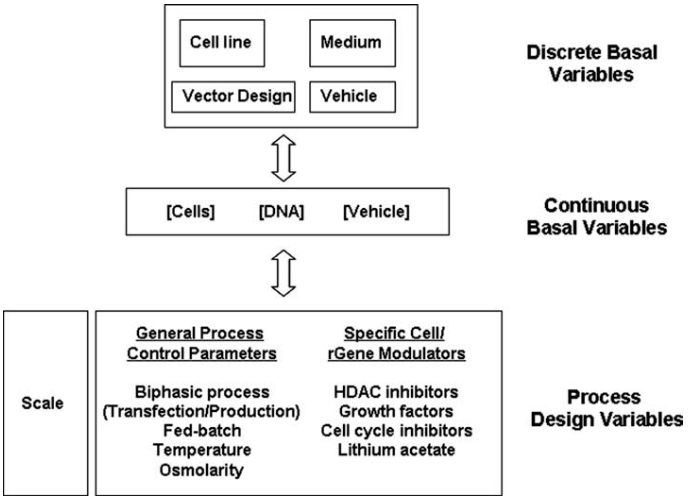


Figure 1. Interacting variables underpinning mammalian host cell based transient production process design.

system to predict reliably key manufacturability attributes (e.g., yield, post-translational modifications, aggregation, etc) is highly desirable in itself. In fact, the key difference between transient production systems and stable cell line development procedures is the general lack of cell clone functional screening/isolation technology (that underpins stable cell line generation) applied to the former. A typical transient production process uses functionally heterogeneous parental cell populations, whose intrinsic genetic/functional heterogeneity has not been exploited at all to derive a host cell line intrinsically suited to the task. Moreover, there are very few examples of gene expression technology specifically designed for transient production (e.g., Liao and Sunstrom, 2006[10]).

Therefore, scalable transient production processes based on CHO cells are generally low yielding, with maximum reported titers of recombinant proteins in the range 80–100 mg L^{-1},[11,12] and typically complex with respect to optimization of the numerous factors that have been shown to influence transient process yield (Figure 1). These include discrete, process-specific choices (e.g., base medium type and gene delivery vehicle) and a range of interacting process design variables or effectors that can potentially influence the quantity and quality of the recombinant product.[12–14] Development of a productive transient process therefore represents a significant parameter optimization problem, where the use of design tools to reduce the complexity and time taken to generate a production process would be of real value. To achieve this design of experiments (DoE) methodologies represent an attractive solution, although few reports describe the application of DoE methods for mammalian cell based bioprocess design. DoE methods are preferable to relatively slow and cumbersome one-factor-at-a-time (OFAT) approaches as they avoid experimental bias and quasi-optima with a reduced number of experiments required.[15] A key advantage of DoE methods is their ability to identify reliably dependency or interactions between variables affecting the process outputs, which is not possible via OFAT optimization.

In this study, we provide a DoE based strategy that uses response surface methodology to rapidly optimize the basal continuous variables underpinning polyethylenimine (PEI)-mediated transfection of different CHO host cell lines in a chemically defined medium. Based on this, we demonstrate for the first time that there is substantial inherent variability between CHO cell lines with respect to their ability to be transfected by this method, and that this variation is primarily a consequence of cell line specific differences in binding and internalization of PEI polyplexes.

Materials and Methods

Cell culture

CHO-K1 derived suspension adapted CHO host cell lines, CHO-S (Invitrogen, Paisley, UK), CHO-L (CHOK1SV^{TM}, Lonza Biologics, Slough, UK) and CHO-M (MedImmune, Cambridge, UK) were routinely cultured in vented Ehrlenmeyer shake flasks (Corning, Surrey, UK) in CD-CHO medium (Invitrogen, Paisley, UK) supplemented with L-glutamine (Sigma-Aldrich, St. Louis; 6 mM CHO-L and CHO-M, 8 mM CHO-S) at 37°C in 5% (v/v) CO_2 with orbital shaking at 140 rpm. Cells were resuspended in fresh medium every 3–4 days at a concentration of 2×10^5 cells mL^{-1}. Cell concentration and viability were routinely measured by an automated Trypan Blue exclusion assay using a Vi-Cell counter (Beckman-Coulter, High Wycombe, UK) according to manufacturer's instructions.

Reporter plasmid preparation

The plasmid used in this study was based on pSEAP2-Control (Clontech, Mountain View, CA) backbone. The SV40 enhancer of pSEAP2-Control was deleted by partial digestion with *Hpa*I and *BamH*I. After blunting the ends with Klenow enzyme (Roche, Penzberg, Germany), the DNA was self-ligated. CMV promoter was amplified by PCR from pcDNA3.1/V5-His-TOPO/*lac*Z (Invitrogen) with primers: CMV-a, 5′-GATCAGATCTCGATGTACGGGCCAGATATACG-3′ and CMV-bN, 5′-GATCGAATTCGATCTGACGGTTCACTAAACCAGCTCTGCTTATATAGACCTCCCAC-3′ and cloned into the *Bgl*II and *Eco*RI sites of pSEAP2-Control. The sequence of all constructs was verified. Plasmid DNA was purified using a plasmid midi purification kit (Qiagen, Crawley, UK), resuspended in a Tris-HCl buffer, pH 8.5 and stored at −20°C.

Transfection of CHO cells with PEI

Cells were cultured in shake flask to mid exponential phase prior to transfection. One hour prior to transfection, cells were diluted to the desired concentrations in 50 mL Cultiflasks (Sartorius AG, Goettingen, Germany) in CD-CHO supplemented with 8 mM (CHO S) or 6 mM (CHO M, CHO L) of glutamine in a total volume of 5 mL. In separate eppendorf tubes, the desired quantities of 25 kD linear PEI (Polysciences, Warrington) and DNA were diluted with an equal volume of NaCl 300 mM. PEI and DNA were then mixed together and allow to complex for 1 min in a total volume of 333 μL made up with 150 mM NaCl. The PEI/DNA solution was then added to the cells. Cultiflasks were immediately orbitally shaken by hand and directly placed in an incubator at 37°C in 5% CO_2 with orbital shaking at 170 rpm and an orbital throw of 50 mm. Cultures were harvested after five days. Cell concentration and viability were assessed. Samples were taken and stored at −20°C for secreted alkaline phosphatase (SEAP) quantification using the Sensolyte pNPP Secreted Alkaline Phosphatase colorimetric reporter gene assay kit (Cambridge Biosciences, Cambridge, UK) according to manufacturer's instructions.

Transfection of CHO cells by electroporation

At two days post subculturing, 1×10^7 cells were resuspended in 100 μL nucleofection solution and 2 μg of SEAP plasmid DNA was added to the suspension prior to cuvette electroporation in an Amaxa Nucleofector (Lonza, Slough, UK) using the standard CHO cell protocol. Cells were subsequently resuspended in 40 mL medium and cultured in suspension prior to analysis.

Flow cytometry

All flow cytometric analyses were performed on a FACSCalibur™ instrument (BD Biosciences, Oxford, UK). Plasmid DNA was labeled with fluorescein using a "Label IT" kit (Mirus, Madison, WI) at a ratio of 2:1 fluorophore to DNA. Prior to transfection this was mixed with unlabeled DNA at a 1:4 ratio (labeled:unlabeled) and transfections were carried out using standard procedures as stated. Cells were washed twice with PBS buffer prior to analysis. To measure polyplex uptake post-transfection, cells were washed in PBS, resuspended in "Cellscrub™" complex removal buffer (Genlantis, San Diego, CA), incubated for 10 min at RT and washed twice in PBS prior to flow cytometric analysis. For anti-heparan sulfate immunostaining, cells were washed in PBS, fixed in 4% (w/v) paraformaldehyde (PFA) on ice for 20 min then washed twice with ice-cold PBS and stored at 4°C at a concentration of 1×10^6 mL^{-1}. Prior to immunostaining, fixed cells were washed with PBSB buffer, [PBS, 1% (w/v) BSA] and incubated with anti-HS antibodies (Seikagaku 10E4 or HepSS1 murine monoclonals; AMSbio, Abingdon, UK) at a dilution of 1:100 for 30 min at 4°C. Cells were washed twice in PBSB and stained with anti-mouse IgM FITC-labeled secondary antibody (Invitrogen) at a dilution of 1:500 for 30 min at 4°C followed by three washes in PBSB prior to analysis.

Results

Empirical identification of the design space for RSM optimization of PEI mediated transfection

The three basal continuous variables that underpin the productivity of a non-viral transient production system are the relative concentration of host cells, plasmid DNA encoding the product and chemical gene delivery vehicle (Figure 1). In this study, we are concerned with the transfection of CHO cell lines with plasmid DNA condensed with the cationic polymer PEI. The latter is known to be cytotoxic.[16] Therefore, despite the known modulation of PEI cytotoxicity by DNA[17,18] we considered that only the discrete interaction between PEI and the host cell impacts significantly on the design space (upper) limit as it defines an inherently critical response factor, host cell viability. The other possible interactions, host cell/DNA and DNA/PEI do not directly inform on design space boundaries. Accordingly, we first measured the impact of increased PEI concentration on the proliferation and viability of three suspension-adapted parental CHO cell lines, CHO-S, CHO-L, and CHO-M, each maintained in CD-CHO medium. Each CHO cell line was obtained from a different commercial source: although all cell lines were derived originally from CHO-K1 cells. As shown in Figure 2, although each CHO cell line exhibited a similar specific growth rate over a five-day culture period in control conditions [average μ ranging from 0.3 d^{-1} (CHO-S) to 0.28 d^{-1} (CHO-L), at an initial seeding density of 1.5×10^6 cells mL^{-1} diluted from a 96 h mid-exponential cell culture in each case], their response to added PEI was extremely cell line specific. CHO-L was the most resistant to PEI, maintaining the highest viable cell concentration as PEI concentration increased. In contrast, CHO-M exhibited a precipitous decline in cell viability at PEI concentrations >10 μg mL^{-1}. Based on these simple empirical

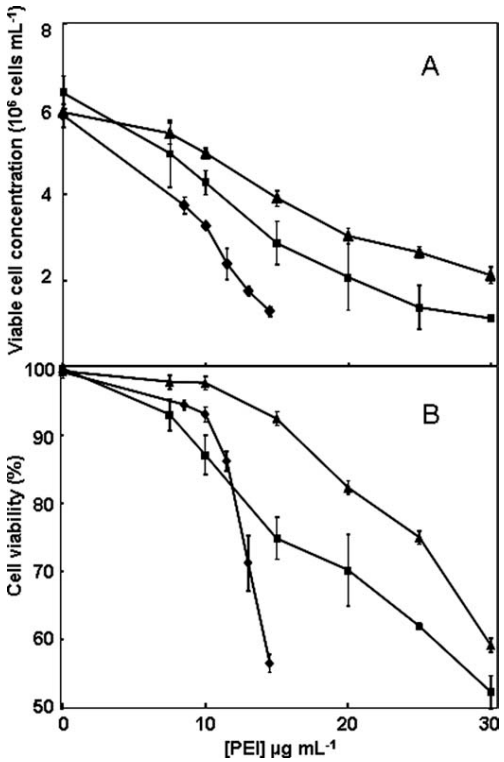


Figure 2. Empirical determination of CHO host cell responses to PEI to identify a DoE design space critical limit. Different CHO host cells (CHO-S ■, CHO-L ▲, and CHO-M ◆, each seeded at 1.5×10^6 cells mL^{-1}) were exposed to varying concentrations of PEI (25 kDa, linear).

Viable cell concentration (A) and percent cell viability (B) were measured after five days culture. $n = 3 \pm$ SD.

observations, the lower and upper bounds of PEI concentration employed were set to arbitrary, cell line specific limits, which corresponded to a final cell viability of 95% (upper bound) or 70% (lower bound) after five days growth at a given PEI concentration (Table 1). The plasmid DNA concentration range was set with reference to previous published literature.[19–21] For each cell line the DNA:PEI ratios used were the same, ranging between 1:2 and 1:7 (w/w). Therefore, for each cell line, the maximum DNA load employed was limited by the (cell line specific) maximum set PEI concentration. With respect to cell concentration range, an arbitrary lower bound of 0.5×10^6 cells mL^{-1} was employed uniformly, and the upper limit of initial cell concentration was also equivalent in each case, set to 2.5×10^6 cells mL^{-1}, the latter representing a concentration of mid-exponential cells from a diluted donor 96 h culture compatible with the maintenance of high cell viability over a subsequent five day transient production period under control conditions.

Sequential response surface models identify an optimal, cell line specific combination of basal variables for transient production

Initially, face-centered Box-Wilson (central composite) designs were constructed using the selected basal variable (PEI, DNA, and cells) concentration ranges as design space boundaries using Design Expert 7 software. A Box-Wilson design was employed initially because this method can theoretically generate accurate predictions of response variable output across a broad, untested design space as it uses basal variable combinations at the extremities of this space.[22] For each cell line SEAP reporter output after a five day production period was measured at 15 different coordinates within the cell line specific design space (three levels per variable), with the mid-point assay replicated six times to determine pure error. Calculated from SEAP reporter output, a response surface model was used to predict an optimal combination of basal variables for maximum SEAP reporter production for each cell line (Figure 3A). Each model mathematically described the relationship between SEAP response and basal variables, a second order function of individual, basal variable interaction and quadratic terms. Coefficients of determination (R^2 value) from the Box-Wilson design are 0.88, 0.93, and 0.76 for CHO S, L, and M, respectively, indicating a satisfactory model fit of experimental data. For example, the model derived from the CHO-L cell line Box-Wilson design analysis shown in Figure 3A is as follows:

$$\text{Relative SEAP} = -194.32 + 23.21 \times [\text{PEI}] - 4.00 \times [\text{DNA}] + 57.31 \times [\text{Cells}] + 0.54 \times [\text{PEI}][\text{DNA}] + 0.69 \times [\text{PEI}][\text{Cells}] + 0.47[\text{DNA}][\text{Cells}] - 0.70 \times [\text{PEI}]^2 - 0.35 \times [\text{DNA}]^2 - 18.3 \times [\text{Cells}]^2$$

For each cell line, a simple maximum response prediction (SEAP volumetric titer) was identified within the design space, associated with a cell line specific optimum combination of basal variable concentrations (Table 2). For each calculated basal variable optimum, a predicted 95% confidence interval (defined as the confidence that another model of the same order will predict an optimum within the given range) was calculated using an in-house program written within Mathematica software (Wolfram Research, Long Hanborough, U.K). These are shown in Table 2.

To validate the Box-Wilson modeled predictions of basal variable optima for each cell line and to produce a higher resolution prediction of these optima, we performed a Box-Behnken design using the Box-Wilson predicted optima as the mid-range level for each basal variable. The Box-Behnken design was used in sequence after the Box-Wilson design as the former method enables a prediction of response variable output using fewer combinations of basal variables than the Box-Wilson design within a design space that is not required to be tested at its extremities. For this analysis, the high to low range of basal variable concentrations employed was divided by two relative to the Box-Wilson design. For each cell line, SEAP reporter output after a five day production period was measured at 13 different coordinates within the cell line specific design space (three levels per variable), with the mid-point assay replicated five times to determine pure error. As described above for the Box-Wilson analysis, a response surface model was used to predict an optimal combination of basal variables for maximum SEAP reporter production for each cell line. An example of this analysis for the CHO-L host cell line is shown in Figure 3B, and the Box-Behnken predicted optimal basal variable concentrations for each host cell line are listed in Table 2. Coefficients of determination (R^2 values) from the Box-Behnken models were 0.98, 0.98, and 0.97 for CHO S, L, and M, respectively, indicating an excellent model fit of experimental data.

Table 1. Summary of Host CHO Cell Line Specific DoE Design Space Limits for Initial Box-Wilson Experimental Design

	Host cell		
Process variable	CHO S	CHO L	CHO M
[PEI] ($\mu g\ mL^{-1}$)	7.5–20	12–26	8.5–13
[DNA] ($\mu g\ mL^{-1}$)	1–10	1.7–13	1–6.5
[Cells] (10^6 cells mL^{-1})	0.5–2.5	0.5–2.5	0.5–2.5

Together, these data show clearly that the basal variable optima predicted by both methods are highly comparable, all Box-Behnken predicted optima lie within the 95% confidence intervals associated with the Box-Wilson predicted optima. The 95% confidence intervals associated with the Box-Behnken predicted optima are consistently lower than those associated with the Box-Wilson optima, likely deriving from the more restricted design space.

The predicted optimal basal variable concentrations were cell line specific. Most obviously, there was an apparent relationship between cell line resistance to PEI (Figure 2) and optimum predicted PEI concentration in the rank order CHO-L > CHO-S > CHO-M. As the DNA:PEI (w:w) ratio at each optimum was relatively constant, ranging from 1:1.8 (CHO-S cells) to 1:2.4 (CHO-M cells), cell cultures more resistant to PEI could be effectively "loaded" with more plasmid DNA, varying from 3.3 pg DNA $cell^{-1}$ (CHO-M) to 5.2 pg DNA $cell^{-1}$ (CHO-L).

These data suggest that at the predicted optima, whilst the PEI:DNA interaction is unlikely to be cell line specific, the PEI:cell interaction is cell line specific, and it is this that limits DNA input into the culture. Justifying the DoE approach taken here, at least one significant ($P < 0.05$, Fisher/Snedecor test) interaction between discrete basal variables was identified across the Box-Behnken design space for each cell line. However, the interactions identified as significant (either PEI-cell or cell-DNA or PEI-DNA) were cell line specific. Therefore, by comparison to the DoE method, necessarily cell line specific transient process optimization by the OFAT approach (e.g., Boussif et al., 1995[23]; Haldankar et al. 2006[24]) where optimal parameters are identified sequentially (with the inherently larger number of

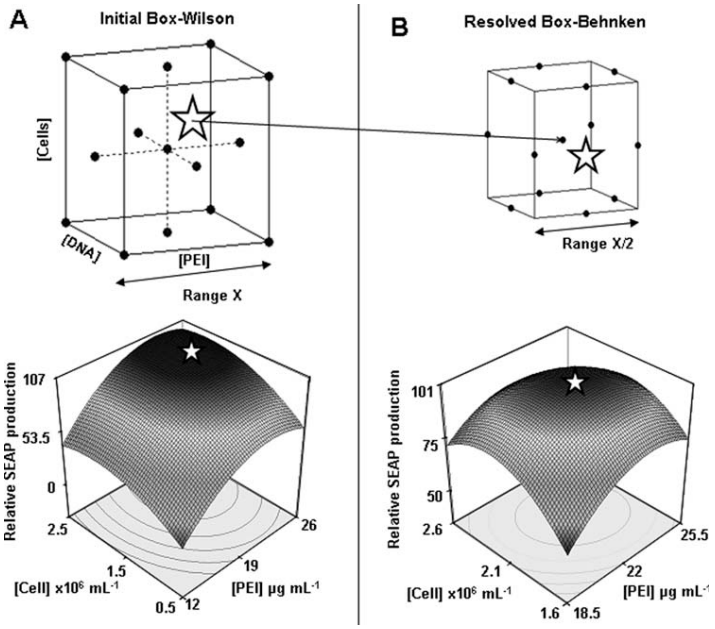


Figure 3. Optimization of continuous basal variables for transient transfection of CHO cells by sequential DoE response surface modeling.

Box-Wilson (A) followed by higher resolution Box-Behnken (B) response surface designs were used to predict the optimal concentration of continuous basal variables (PEI, cells, and DNA) for each CHO host cell line. The schematic boxes illustrate experimental design coordinates in each case, and an example of a modeled response surface (Cells-PEI interaction at the predicted optimal DNA concentration) for CHO-L cells is shown. The simple maximum predicted response for SEAP production is indicated by a star.

Table 2. Predicted CHO Host Cell Specific Basal Variable Optima for Maximum SEAP Production Identified by Sequential DoE Response Surface Modeling

Cell line		BW RSM	BB RSM
CHO S	[Cells] (10^6 cells mL^{-1})	2.4 (±0.3)	2.5 (*)
	[PEI] ($\mu g\ mL^{-1}$)	15 (±1.3)	16.3 (±0.9)
	[DNA] ($\mu g\ mL^{-1}$)	8.3 (±2.3)	9.0 (±0.3)
	Cell viability (%)	71.4 (±3.6)	69.3 (±6.4)
CHO L	[Cells] (10^6 cells mL^{-1})	2.1 (±0.1)	2.1 (±0.0)
	[PEI] ($\mu g\ mL^{-1}$)	22.4 (±0.5)	22.6 (±0.1)
	[DNA] ($\mu g\ mL^{-1}$)	11.6 (±1.2)	11.2 (±0.1)
	Cell viability (%)	97.5 (±2.4)	96.9 (+0.2)
CHO M	[Cells] (10^6 cells mL^{-1})	1.7 (±0.9)	1.6 (±0.0)
	[PEI] ($\mu g\ mL^{-1}$)	11.4 (±1.7)	13.0 (±1.0)
	[DNA] ($\mu g\ mL^{-1}$)	4.6 (±2.6)	5.4 (±0.9)
	Cell viability (%)	91.9 (±1.1)	94.3 (±0.4)

In each case, the predicted cell viability after a five day production period is shown.

2.5×10^6 cells mL^{-1} represents a practical optimum and not a prediction.

BW, Box-Wilson design; BB, Box-Behnken design.

experiments required) represents an inefficient and statistically limited means to identify basal variable optima.

Comparison of transient process outputs at cell line specific basal variable optima

The three CHO host cells clearly differed with respect to the overall level of recombinant protein produced during small-scale DoE optimization. To comparatively evaluate their relative performance at each discrete basal variable optimum, SEAP production and host cell growth were measured after a comparable five-day transient production period. These data are shown in Figure 4. CHO-S cells secreted approximately eight-fold more SEAP reporter than either CHO-L or CHO-M cells. With respect to the relative efficiency of product formation per mass of DNA substrate, CHO host cells differed substantially, where the relative ratio of SEAP produced per mass of DNA substrate was CHO-S (1):CHO-L (0.1):CHO-M (0.16). These data show clearly that despite rational optimization of the transient process, the intrinsic "transfectability" of each host cell varied substantially. However, these data also showed that the CHO host cell lines varied with respect to their relative rates of cell proliferation and death during the production process. This critical process design parameter cannot be easily inferred from an empirical determination of cell response to PEI alone (Figure 2) as DNA interaction with PEI may modulate its cytotoxicity. Cell death can be considered a critical output parameter as (i) release of hydrolytic enzymes from dead cells can potentially affect both product quality and quantity and (ii) potential extension of the culture process to increase product titer requires viable cells. Although exhibiting a higher productivity, the CHO-S based production process negatively affected host cell viability greater than the other host cells, with 2.5-fold more dead cells than either the CHO-L or CHO-M cultures. The specific rate of CHO-S cell death (0.17 d^{-1}) during transient production was two-fold higher than that of the other host cells (+/−0.02 d^{-1}), although the observed host cell viabilities generally concur with the DoE predicted cell viabilities (Table 2). Importantly, we note that, it would be possible to use both SEAP reporter output and cell viability as selection criteria for optimal process design using "Desirability" function within the DoE software. For example, for a CHO-S process, with a predicted cell viability, maintained in excess of 90%, predicted SEAP reporter output would be 80% of that observed at SEAP maximizing conditions.

Variation in cell surface polyplex binding and internalization underpins variation in CHO host cell transient process performance

As recent studies have shown that variation in CHO cell surface molecular composition can affect polyplex mediated transfection efficiency,[25–27] we tested the hypothesis that

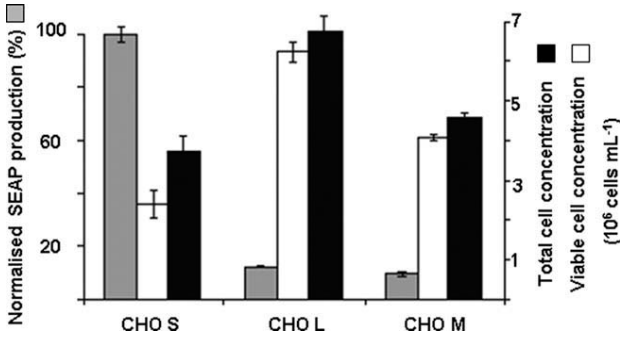


Figure 4. Host CHO cell line specific differences in PEI-mediated transient protein production.

Each CHO cell line was transfected at the predicted basal variable optimum identified by response surface modeling (Table 2). SEAP reporter production (gray bars), viable cell concentration (white bars), and total cell concentration (black bars) were measured after five days culture. $n = 3 \pm$ SD.

CHO host cells differed in their ability to bind and internalize polyplexes. For this experiment, all CHO host cells were transfected with fluorescently labeled plasmid DNA using identical conditions (i.e., at the same concentration of basal variables) to permit a direct comparison of polyplex particle binding and internalization by flow cytometry. The CHO-M basal variable optima (1.6×10^6 cells mL^{-1}, 13 μg PEI mL^{-1}, and 5.4 μg mL^{-1} DNA) were used for each CHO cell host as this modeled optimum (i) employed the least cytotoxic combination of basal variables at the common PEI:DNA (w/w) optimum (approx. 2:1) and (ii) was within the modeled design space for each host cell permitting prediction of SEAP output in each case. We distinguished between fluorescent polyplexes on the cell surface and those internalized using a proprietary reagent, CellScrub^{TM} designed to remove extracellular DNA complexes bound to the cell surface by electrostatic interactions.[28,29] As exemplified in Figure 5A, association of fluorescent polyplexes with cells was quantified by flow cytometry, and treatment of cells with CellScrub effectively removed cell surface polyplexes enabling quantification of internalized particles. Preliminary experiments with CHO-S cells analyzed up to 4 min post-transfection showed that >92% of total cell associated polyplex fluorescence was removed using the CellScrub procedure (data not shown). Comparative analysis of all CHO host cells after 30 min and 4 h transfection showed that each population differed markedly in the relative extent of polyplex binding and internalization (Figures 5B,C). These data show that (i) for each host cell polyplex binding to the cell surface was saturated by 30 min; CHO-M cells bound least polyplex at the cell surface, 75% less than CHO-S and 82% less than CHO-L, (ii) CHO-S cells exhibited the highest rate of polyplex internalization; >4-fold more polyplexes were internalized by CHO-S cells than either CHO-L or CHO-M cells, (iii) internalization of polyplexes was not proportional to cell surface polyplex binding; for CHO-L cells only 8.2% of total cell associated fluorescence was intracellular after 4 h transfection, whereas CHO-S and CHO-M cells internalized cell surface polyplexes more rapidly, with 41 and 34% of total cell fluorescence measured as intracellular after 4 h. We note that the observed host cell specific differences in polyplex binding were not related to cell surface area/volume ratio as each host cell population typically exhibited a similar size distribution prior to transfection (16–18 μm diameter).

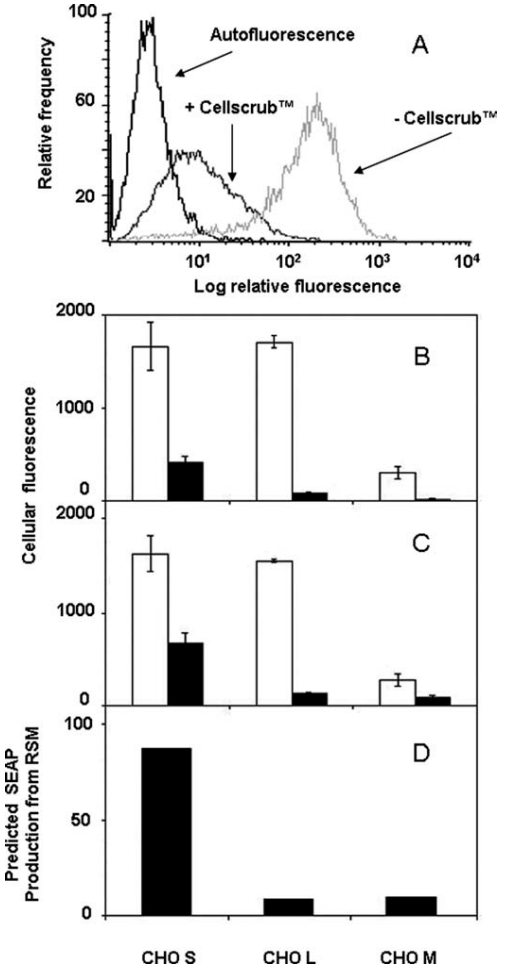


Figure 5. Host CHO cells vary in PEI-DNA polyplex binding and internalization.

Plasmid DNA was fluorescently labeled with fluorescein, and host CHO cells (1.6×10^6 cells mL^{-1}) were transfected with PEI-DNA polyplexes formed at a PEI:DNA (w/w) ratio of 2.4:1 with final concentrations in culture of 13 μg mL^{-1} PEI and 5.4 μg mL^{-1} DNA. For each CHO host cell, cell surface and internalized PEI-DNA polyplex fluorescence was measured by flow cytometry (A; CHO-L cells transfected for 30 min are shown as an example) either before (light gray line; total polyplex fluorescence) or after removal of extracellular polyplexes using CellScrub^{TM} reagent (dark gray line; internalized polyplex fluorescence only). Different CHO host cells were analyzed by flow cytometry either 30 min (B) or 4 h (C) after transfection, white bars represent total cell associated polyplex, black bars represent internalized polyplexes. $n = 2 \pm$ SD, geometric means are shown in each case. For each CHO host cell, the SEAP output predicted by response surface modeling at the basal variable concentrations employed is shown in Figure 5D (see Figure 3).

At the basal variable concentrations employed for this analysis the predicted relative SEAP production for each cell line from DoE response, surface modeling (Figure 5D) was clearly proportionate to the degree of polyplex internalization by each host cell population (Figure 5C). These data implied that molecular interactions between polyplexes and the host CHO cell surface and the associated endocytotic pathway(s) used were host cell line specific. As previous studies have shown that (i) CHO cell surface heparan sulfate proteoglycan (HSPG) level is related to polyplex internalization[25] and (ii) cell surface polyplexes colocalize with anti-HSPG antibodies on BS-C-1 cells,[26] we tested the hypothesis that differences in polyplex binding and/or internalization between host CHO cells is related to their cell surface HSPG

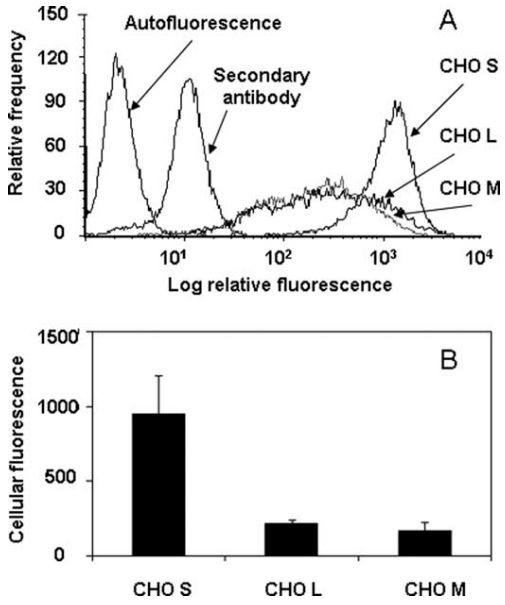


Figure 6. Host CHO cell heparan sulfate proteoglycan level covaries with PEI-DNA polyplex internalization.

Host CHO cells (1×10^6) were fixed in 4% (w/v) paraformaldehyde then labeled with murine anti-HSPG monoclonal antibodies followed by FITC-labeled goat anti-mouse secondary antibodies prior to analysis by flow cytometry. (A) Flow cytometric analysis of cell surface HSPG level. (B) Median CHO cell surface HSPG content. In each case, values were adjusted for the fluorescence of secondary antibody stained control cells. $n = 3 \pm$ SD.

content. As shown in Figure 6, CHO-S cells were more heavily stained (approx. 4–5 fold) with anti-HSPG monoclonal antibodies (10E4 and HepSS1) than either CHO-L or CHO-M cells. These data imply that CHO cell surface HSPG content is correlated with polyplex internalization rate (Figure 5C) rather than polyplex binding (Figure 5B), as previously described by Wong et al.,[25] and that this parameter is genetically variable between CHO cell lines.

Lastly, to confirm that the observed differences in SEAP production via PEI mediated transfection were a consequence of host cell specific variation in polyplex binding and internalization, each host cell population was also transfected by electroporation, a gene delivery mechanism that does not use the endocytotic pathway. Each host cell population was electroporated using identical conditions and SEAP production was measured after five days. These data are shown in Figure 7. Whilst host cell specific SEAP production was clearly less variable than for PEI mediated transfection, CHO-L cells demonstrated higher levels of SEAP production than either CHO-S or CHO-M cells. Taken together, these data indicate that host cell specific variation in PEI-mediated transfection is primarily a function of two discrete properties of the host cell surface, polyplex binding capacity and the rate of endocytotic internalization of bound particles.

Discussion

Our data demonstrate that host CHO cell lines differ substantially in their relative ability to be transfected via endocytosis. This is functional characteristic governs the intrinsic suitability of a given CHO cell as a vehicle for scalable transient production of recombinant proteins. This cell line specific variation impacts transient production process design in two fundamental and interdependent ways, (i) basic optimi-

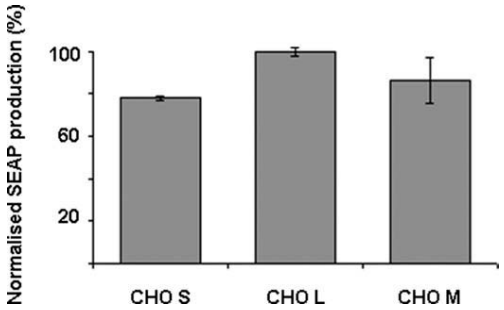


Figure 7. Host CHO cells exhibit, limited variation in transient production after gene delivery by electroporation.

CHO cells (1×10^7) were transfected with 2 μg of plasmid DNA then diluted to 2.5×10^5 cell mL^{-1} and cultured for five days prior to measurement of SEAP production. $n = 3 \pm$ SD.

zation of process variables has to be cell line specific and (ii) screening or engineering of host cell populations for improved transient production should primarily target cell surface binding and/or internalization of DNA nanoparticles. With respect to the former, we have shown clearly that DoE response-surface methodology (RSM) permits the rapid identification of an optimal combination of process variables. Although recombinant protein productivity is highlighted in the current study, host cell concentration and viability can also be considered important key process outputs. Using RSM modeling permits a "sweet spot" combination of responses to be achieved by manipulation of the input variables (i.e., DNA, PEI, and cell concentrations) using desirability functions.[30] For example, an accumulation of nonviable cells over the production process may adversely affect product titer or quality via the release of intracellular proteases or glycosidases.[31]

The main difficulty in response surface methodology lies in the choice of initial design space for the subsequent response surface optimization. Generally, using factorial designs and method of steepest ascent, the experimenter gradually moves towards the vicinity of the operating conditions where there is an optimal response and where RSM optimization should be conducted.[32] This sequential approach for design space identification can be time and labor consuming. In this study, we show that identifying the initial RSM design using empirical cell specific cytotoxicity to free PEI in the culture medium is a rapid and effective method.

Each CHO host cell population exhibited a distinct functional phenotype with respect to cell surface polyplex binding and internalization; either high binding/high internalization (CHO-S), high binding/low internalization (CHO-L), or low binding/high internalization (CHO-M). From these data, we infer that both the molecular composition of the cell surface and associated mechanisms of endocytosis are host cell population specific, yielding the observed differences in "permissivity" to polyplex mediated transfection. In general, this substantial phenotypic heterogeneity implies that these aspects of CHO cell function (binding to, and endocytosis of extracellular polymers), which may originally have been relevant to the function of the somatic CHO cell are now subject to neutral genetic drift in an unstable genetic background, irrelevant to the objective function (proliferation) of CHO cells during serial subculture in a synthetic environment. For cancer cell lines, it has been reported that the genome can evolve during subculture to an extent that caution should be exercised in using these cultures as models of human cancer.[33]

More specifically, we infer that host cell specific variation in polyplex binding may derive from differences in the relative abundance of cell surface proteins bearing polyanionic glycosaminoglycans (GAG) of varying molecular composition implicated in cell surface polycation binding and uptake.[34] For example, HSPG deficient CHO cells have been shown to exhibit reduced transfection efficiency mediated by polyplexes and lipoplexes[26] and more recently, FUT8 (fucosyltransferase 8) knockout CHO cells with a lower HSPG content than control cells exhibit a reduced uptake of PEI polyplexes.[25] In accordance with this, flow cytometric analysis of CHO-S cells with anti-HSPG monoclonal antibodies showed that this host cell line had a 4 to 5-fold higher cell surface HSPG content than either CHO-L or CHO-M cells, implying that cell surface HSPG level may be used to screen CHO cell populations for "transfectable" genetic variants. We also note in relation to this that variation in the cell surface content of different HSPGs may affect polyplex transfection as described by Paris et al.[35]

With respect to internalization of polyplexes, mammalian cells may use a variety of endocytotic pathways for internalization of molecular cargo.[36,37] Currently, there is no general consensus on the mechanistic link between nonspecific cell surface binding of polyplexes (or lipoplexes) by GAG molecules and the discrete mechanisms of endocytosis. A situation obscured by mammalian cell line specific observations made in a variety of studies,[38–40] which may relate to variations in the relative abundance of discrete HSPGs.[35] Where CHO cells have been used specifically, the available data suggests that polyplexes are likely to be internalized by nonclathrin, noncaveolae dependent[26] fluid-phase mechanisms[41] such as macropinocytosis.

In conclusion, this study shows that, as for stable production of recombinant proteins by CHO cells, discrete CHO host cell populations may be inherently more fit-for-purpose than others, where choice of CHO host cell is a key factor underpinning productivity.[4–6] We suggest that the optimization of a transient production process should rationally begin with cell clone screening operations to isolate transfection-competent host cells, followed by rapid DoE-based design of a clone-specific optimal production process.

Acknowledgments

The authors are grateful to Lonza Biologics for their provision of the CHOK1SV host cell line used in this study. This work was sponsored by both the Engineering and Physical Sciences Research Council and the Biotechnology and Biological Sciences Research Council via postgraduate studentships to BCT and CRJS, respectively.

Literature Cited

1. Birch JR, Racher AJ. Antibody production. *Adv Drug Delivery Rev*. 2006;58:671–685.
2. Walsh G. Biopharmaceutical benchmarks 2010. *Nat Biotechnol*. 2010;28:917–924.
3. Browne SM, Al-Rubeai M. Selection methods for high-producing mammalian cell lines. *Trends Biotechnol*. 2007;25:425–432.
4. Kennard ML, Goosney DL, Monteith D, Roe S, Fischer D, Mott J. Auditioning of CHO Host Cell Lines Using the Artificial Chromosome Expression (ACE) Technology. *Biotechnol Bioeng*. 2009;104:526–539.
5. O'Callaghan PM, James DC. Systems biotechnology of mammalian cell factories. *Brief Funct Genomic Proteomic*. 2008;7:95–110.
6. O'Callaghan PM, McLeod J, Pybus LP, Lovelady CS, Wilkinson SJ, Racher AJ, Porter A, James DC. Cell Line-Specific Control of Recombinant Monoclonal Antibody Production by CHO Cells. *Biotechnol Bioeng*. 2010;106:938–951.
7. Condon RGG, Schaefer EJ, Santoro M, Longley R, Tsao YS, Zurawski SM, Liu Z. Development of a Chinese hamster ovary cell line for recombinant adenovirus-mediated gene expression. *Biotechnol Prog*. 2003;19:137–143.
8. Barnes LM, Moy N, Dickson AJ. Phenotypic variation during cloning procedures: analysis of the growth behavior of clonal cell lines. *Biotechnol Bioeng*. 2006;94:530–537.
9. Singh V. Disposable bioreactor for cell culture using wave-induced agitation. *Cytotechnology*. 1999;30:149–158.
10. Liao MML, Sunstrom NA. A transient expression vector for recombinant protein production in Chinese hamster ovary cells. *J Chem Technol Biotechnol*. 2006;81:82–88.
11. Zhang X, Garcia I, Baldi L, Hacker D, Wurm F. Hyperosmolarity enhances transient recombinant protein yield in Chinese hamster ovary cells. *Biotechnol Lett*. 2010:1–6.
12. Ye JX, Kober V, Tellers M, Naji Z, Salmon P, Markusen JF. High-Level Protein Expression in Scalable CHO Transient Transfection. *Biotechnol Bioeng*. 2009;103:542–551.
13. Galbraith DJ, Tait AS, Racher AJ, Birch JR, James DC. Control of culture environment for improved polyethylenimine-mediated transient production of recombinant monoclonal antibodies by CHO cells. *Biotechnol Prog*. 2006;22:753–762.
14. Backliwal G, Hildinger M, Kuettel I, Delegrange F, Hacker DL, Wurm FM. Valproic acid: a viable alternative to sodium butyrate for enhancing protein expression in mammalian cell cultures. *Biotechnol Bioeng*. 2008;101:182–189.
15. Mandenius CF, Brundin A. Bioprocess optimization using design-of-experiments methodology. *Biotechnol Prog*. 2008;24:1191–1203.
16. Moghimi SM, Symonds P, Murray JC, Hunter AC, Debska G, Szewczyk A. A two-stage poly(ethylenimine)-mediated cytotoxicity: implications for gene transfer/therapy. *Mol Ther*. 2005;11:990–995.
17. Choosakoonkriang S, Lobo BA, Koe GS, Koe JG, Middaugh CR. Biophysical characterization of PEI/DNA complexes. *J Pharm Sci*. 2003;92:1710–1722.
18. Lee M. Apoptosis induced by polyethylenimine/DNA complex in polymer mediated gene delivery. *Bull Korean Chem Soc*. 2007;28:95–98.
19. Durocher Y, Perret S, Kamen A. High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. *Nucleic Acids Res*. 2002;30:9.
20. Chenuet S, Martinet D, Besuchet-Schirultz N, Wicht M, Jaccard N, Bon AC, Derouazi M, Hacker DL, Beckmann JS, Wurm FM. Calcium phosphate transfection generates mammalian recombinant cell lines with higher specific productivity than polyfection. *Biotechnol Bioeng*. 2008;101:937–945.
21. Reed SE, Staley EM, Mayginnes JP, Pintel DJ, Tullis GE. Transfection of mammalian cells using linear polyethylenimine is a simple and effective means of producing recombinant adeno-associated virus vectors. *J Virol Methods*. 2006;138:85–98.
22. Anderson MJ, Whitcomb PJ. *RSM simplified*. New York: Productivity press; 2005.
23. Boussif O, Lezoualch F, Zanta MA, Mergny MD, Scherman D, Demeneix B, Behr JP. A versatile vector for gene and oligonucleotide transfer into cells in culture and in-vivo–Polyethylenimine. *Proc Natl Acad Sci U S A*. 1995;92:7297–7301.
24. Haldankar R, Li DQ, Saremi Z, Baikalov C, Deshpande R. Serum-free suspension large-scale transient transfection of CHO cells in WAVE bioreactors. *Mol Biotechnol*. 2006;34:191–199.
25. Wong AW, Baginski TK, Reilly DE. Enhancement of DNA uptake in FUT8-deleted CHO cells for transient production of afucosylated antibodies. *Biotechnol Bioeng*. 2010;106:751–763.
26. Payne CK, Jones SA, Chen C, Zhuang XW. Internalization and trafficking of cell surface proteoglycans and proteoglycan-binding ligands. *Traffic*. 2007;8:389–401.
27. Poon GMK, Gariepy J. Cell-surface proteoglycans as molecular portals for cationic peptide and polymer entry into cells. *Biochem Soc Trans*. 2007;35:788–793.
28. Uzgun S, Akdemir O, Hasenpusch G, Maucksch C, Golas MM, Sander B, Stark H, Imker R, Lutz JF, Rudolph C.

Characterization of tailor-made copolymers of oligo(ethylene glycol) methyl ether methacrylate and N,N-dimethylaminoethyl methacrylate as nonviral gene transfer agents: influence of macromolecular structure on gene vector particle properties and transfection efficiency. *Biomacromolecules*. 2009;11:39–50.
29. Pun SH, Bellocq NC, Liu AJ, Jensen G, Machemer T, Quijano E, Schluep T, Wen SF, Engler H, Heidel J, Davis ME. Cyclodextrin-modified polyethylenimine polymers for gene delivery. *Bioconjugate Chem*. 2004;15:831–840.
30. Derringer GC. A balencing act-Optimizing a products properties. *Quality Prog*. 1994;27:51–58.
31. Cruz HJ, Dias EM, Peixoto CM, Moreira JL, Carrondo MJT. Product quality of a recombinant fusion protein expressed in immobilised baby hamster kidney cells grown in protein-free medium. *Biotechnol Lett*. 2000;22:677–682.
32. Zhuang YP, Chen B, Chu J, Zhang SL. Medium optimization for meilingmycin production by Streptomyces nanchangensis using response surface methodology. *Process Biochem.* 2006;41:405–409.
33. Hiorns LR, Bradshaw TD, Skelton LA, Yu Q, Kelland LR, Leyland-Jones B. Variation in RNA expression and genomic DNA content acquired during cell culture. *Br J Cancer*. 2004;90:476–482.
34. Mislick KA, Baldeschwieler JD. Evidence for the role of proteoglycans in cation-mediated gene transfer. *Proc Natl Acad Sci U S A*. 1996;93:12349–12354.
35. Paris S, Burlacu A, Durocher Y. Opposing roles of syndecan-1 and syndecan-2 in polyethyleneimine-mediated gene delivery. *J Biol Chem*. 2008;283:7697–7704.
36. Khalil IA, Kogure K, Akita H, Harashima H. Uptake pathways and subsequent intracellular trafficking in nonviral gene delivery. *Pharmacol Rev*. 2006;58:32–45.
37. Hansen CG, Nichols BJ. Molecular mechanisms of clathrin-independent endocytosis. *J Cell Sci*. 2009;122:1713–1721.
38. Rejman J, Bragonzi A, Conese M. Role of clathrin- and caveolae-mediated endocytosis in gene transfer mediated by lipo- and polyplexes. *Mol Ther*. 2005;12:468–474.
39. von Gersdorff K, Sanders NN, Vandenbroucke R, De Smedt SC, Wagner E, Ogris M. The internalization route resulting in successful gene expression depends on polyethylenimine both cell line and polyplex type. *Mol Ther*. 2006;14:745–753.
40. van der Aa M, Huth US, Hafele SY, Schubert R, Oosting RS, Mastrobattista E, Hennink WE, Peschka-Suss R, Koning GA, Crommelin DJA. Cellular uptake of cationic polymer-DNA complexes via caveolae plays a pivotal role in gene transfection in COS-7 cells. *Pharm Res*. 2007;24:1590–1598.
41. Hufnagel H, Hakim P, Lima A, Hollfelder F. Fluid phase endocytosis contributes to transfection of DNA by PEI-25. *Mol Ther*. 2009;17:1411–1417.

Manuscript received Jun. 8, 2011, and revision received July 29, 2011.