Process Understanding and Design Methodology for Industrial Biotechnology

Kirsty Jane McLachlan

A thesis presented for the degree of Engineering Doctorate in Biopharmaceutical Process Development

Biopharmaceutical & Bioprocessing Technology Centre,

School of Chemical Engineering and Advanced Materials,

Newcastle University

June 2017

Acknowledgements

I would like to acknowledge the support I have received from Britest, my industrial sponsor, in particular Dr Charles Gordon for his sustained input over the duration of the project. In addition I would like to thank the Britest member companies who were involved in the research for their time and knowledge.

I would like to thank my initial academic supervisor Dr Lucy Foley for all the coffee through first year, and my subsequent academic supervisor Professor Jarka Glassey for her invaluable and unwavering help and support over the last 4 years.

Thanks go to my fellow EngDs, friends and family, who have tolerated me through the many grumpy days, and to Shaky and Dudley for keeping me sane.

Finally a huge thank you to my colleagues at CPI for the guidance, encouragement and motivation in the final writing up stages.

Preface

This thesis describes research that was undertaken as part of an Engineering Doctorate (EngD) in Biopharmaceutical Process Development which was carried out in collaboration with Britest Ltd. and sponsored by the Engineering and Physical Sciences Research Council (EPSRC) (Grant number EP/G037620/1).

Being an industry sponsored Engineering Doctorate, the project reflects the research requirements of Britest Ltd., and was conducted with an industrial focus.

The work considers the toolkit with the aim of developing tools to allow Britest to move into the bioprocessing sector. Tools suitable for application to bioprocessing are required before companies from the bioprocessing sector can be confident that membership of Britest Ltd. will be beneficial, and the associated cost justified.

The thesis sets out recommendations for tools to Britest Ltd. that have been made based on the outcomes of the research.

Abstract

Many types of knowledge exist within a bioprocess, but the utilisation of this knowledge is not always as straightforward as collecting and analysing data. The Quality by Design initiative (ICH Guideline, 2009) has increased the need for thorough process understanding within bioprocessing. Fundamental process understanding is imperative to adequately implement a QbD approach to a bioprocess. Formalised knowledge capture techniques have been developed previously (West, 1992; Ranjan et al., 2002; Stowell, 2013), but these tend to be designed only to capture information rather than increase understanding. Equally, modelling techniques can be utilised to predict process behaviour and therefore increase understanding, but these rely on the user to have an understanding of the underlying science. This can be problematic in interdisciplinary industries such as bioprocessing, as there are many factors to build into a model. With this in mind, this research considers the Britest tools with respect specifically to biotechnological applications, and formulates a whole bioprocess development methodology. The Britest tools are a suite of qualitative tools and methodologies which were designed to highlight the knowledge gaps within chemical and physical processes, and to promote innovative process design solutions. The tools can help to identify areas where optimisation may be possible, and also increase the understanding of the process as a whole across a range of disciplines.

The Britest tools were first considered with respect to four bioprocesses (Monoclonal Antibody production, Insulin production, Waste Water Treatment and Penicillin production), simulated within SuperPro Designer. The range of processes gave an indication of breadth of application, while the depth of information available in the simulations allowed the research to be unhindered by data availability. From here, several gaps within the toolkit were identified, including the potential for variability and the interactions between multiple parameters.

Variability is inherent within a bioprocess, and the reduction of this variability is a key driver for the implementation for QbD. The Reaction/Reagent Transformation Tracker (R2T2) was designed to capture this variability, and allow the user to evaluate the potential for various scenarios to arise. The tool facilitates a whole process view, without the information becoming overwhelming and confusing for the users.

Understanding the interactions between Critical Quality Attributes (CQAs) and Critical Process Parameters (CPPs) is essential to the successful implementation of QbD, and was not covered by the original Britest toolkit. To combat this the Interaction Analysis Table (IAT) was created. The tool was designed to be applied in the early stages of process development, to guide the application of Design of Experiments (DoE) approaches when data is in short supply but process knowledge is available. Finally, the IAT was evaluated for sensitivity, to investigate the potential influence of uncertainty/human error on the outcome. The work identified a parameter and a threshold value enabling the user to assess the confidence in the proposed process analysis outcome.

This work sought to develop novel knowledge management tools which had been designed specifically for application to bioprocessing. It aimed to establish the applicability of the Britest toolkit for this purpose, as Britest tools have only previously been applied to chemical and physical processes. A Britest toolkit for bioprocessing could be utilised to aid in the adoption of a QbD approach, through tools specifically designed to capture the knowledge of the process. This knowledge would be difficult to adequately represent in statistical models and could be lost between disciplines without a structured methodology to apply. The toolkit can be used to facilitate better communication in an interdisciplinary environment, and provide key information to enable better process design from an early stage.

Table of Contents

List of Figures

List of Tables

List of Appendices

This thesis contains supplementary electronic appendices intended to support the reader in understanding work included within chapters of the research. The contents of these are detailed below.

Appendix A

Appendix A contains the full Britest studies for the SuperPro Designer research discussed in chapter 3 of this thesis. Each folder is named by process (e.g. Monoclonal Antibody Production), and within these folders are the relevant completed Britest tools for the process as it exists within the SuperPro Designer example bioprocess files. The Insulin production process was the focus of chapter 3, the other processes are included to demonstrate the breadth of tool application.

Appendix B

Appendix B contains the data which was employed to construct IATs within chapter 6 of this thesis, where different methods of organising the data were trialled to ascertain the impact of this on the resulting constructed IAT. The file "Approach 1" contains the data sorted into the highest and lowest results for each output. The file "Approach 2" contains the data in line graph format, which was used to identify trends by visual inspection.

Appendix C

Appendix C contains the sensitivity analysis files which were used to perform analysis in chapter 7 of this thesis. IAT files labelled without a letter (e.g. IAT 4) contained 5 outcomes, whereas IAT files ending in the letter a (e.g. IAT4a) contained ten outcomes. IAT files ending in b investigate the IAT weightings to ±3, those ending in c investigate the 1, 5, 10 scoring system and those ending in d investigate the effect on the threshold value of having 7 outcomes.

List of Abbreviations

Abbreviation	Full Name
ADH	Alcohol Dehydrogenase
AE	Adverse Effects
ATA	Anti-Therapeutic Antibody
Britest	Best Route Innovative Technology Evaluation and Selection Techniques
CMO	Contract Manufacturing Organisation
CPI	Centre for Process Innovation
CPP	Critical Process Parameter
CQA	Critical Quality Attribute
DE9	Design Expert 9 ™
DFA	Driving Force Analysis
DMADV	Define Measure Analyse Design Verify
DMAIC	Define Measure Analyse Improve Control
DNA	Deoxyribo Nucleic Acid
DoE	Design of Experiments
EDTA	Ethylenediaminetetraacetic acid
EngD	Engineering Doctorate
EPSRC	Engineering and Physical Sciences Research Council
ESBES	European Society of Biochemical Engineering Sciences
FDA	Food and Drug Administration
FFIC	FujiFilm Imaging Colourants
FIAT	Fermentation Interaction Analysis Table
FMEA	Failure Modes and Effects Analysis

GM	Granular Media
HIC	Hydrophobic Interaction Chromatography
IAT	Interaction Analysis Table
IB	Inclusion Bodies
ICES	Institute of Chemical and Engineering Sciences
IEX	Ion Exchange Chromatography
IMAC	Immobilised Metal Ion Chromatography
ISA	Initial Screening Analysis
KATKit	Knowledge Acquisition Technique Kit
KM	Knowledge Management
mAb	Monoclonal Antibody
MCS	Monte Carlo Simulations
MMC	Mixed Mode Chromatography
MT	Mass Transfer
OD	Optical Density
PAT	Process Analytical Technology
PD	Pharmacodynamics
PDD	Process Definition Diagram
PenV	Penicillin V
PK	Pharmacokinetics
PrISM	Process Information Summary Map
QA	Quality Assurance
QbD	Quality By Design
R	Reactions

R&D	Research and Development
R2T2	Reaction/Reagent Transformation Tracker
RP-HPLC	Reverse Phase High Performance Liquid Chromatography
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
TE3PO	Transformation, Entities, Properties, Physics, Parameters, Order of Magnitude
TM	Transformation Map
TUB	Technical University Berlin

Chapter 1 Introduction

1.1 Research Objectives

This Engineering Doctorate (EngD) thesis has presented work undertaken in collaboration with Britest Ltd to develop the Britest tools for application to bioprocessing. This research aimed to:

1. Develop novel knowledge management tools designed specifically for bioprocessing
2. Test these tools on a range of industrially relevant datasets
3. Identify the stage of process development at which the tools would add the most value
4. Compare these to alternative methods of enhancing process understanding
5. Investigate whether the Britest tools could be applied to bioprocessing to fill the gaps identified in objectives 1-4

The following chapter gives an insight into the background of the subject areas which relate closely to the topic of this research, and an overview of the research structure.

1.2 Bioprocessing

Bioprocessing is generally the method of choice for the manufacture of biological molecules, as recreating the same chemical structure using chemical synthesis methods can be difficult and expensive. A typical bioprocess is split into two sections, upstream processing where the cell line is grown and the product synthesised, and downstream processing where the cell mass and other contaminants are removed and the product is captured in a pure form. Often downstream processing units are those which have previously been developed and employed in chemical processing, such as

chromatography, filtration and centrifugation, whereas upstream processing is more specialised. The cell line must be selected, and then manipulated to produce the desired product correctly. This in itself can be complicated, as variation in cell line can cause significant variation within the product.

Genome manipulation for this purpose has been applied to many organisms including whole plants, whole insects, whole animals and a range of cell culture types (Gordon *et al.*, 1980; Shinmyo *et al.*, 2004; Van Der Vossen *et al.*, 2005). Within cell culture there are 4 main expression system options which are widely used: mammalian, insect, yeast and bacteria. Each of these has its own merits and drawbacks (Table 1.1), and all have their place in both research and industrial systems. In general, micro-organisms are the favoured host due to the rapid generation time, higher reliability and ease of handling. They have been used for many years and so a range of well characterized expression systems are available. However for some large molecules, in particular monoclonal antibodies, mammalian expression systems would be the host of choice, due to their enhanced ability to produce complex proteins.

Table 1.1 - Characteristics of production systems used within bioprocessing. Taken from Fernandez and Hoeffler (1998).

CHARACTERISTICS	***E. COLI***	**YEAST**	**INSECT CELLS**	**MAMMALIAN CELLS**
CELL GROWTH	rapid (30 min)	rapid (90 min)	slow (18-24 h)	slow (24 h)
COMPLEXITY OF GROWTH MEDIUM	minimum	minimum	complex	complex
COST OF GROWTH MEDIUM	low	low	high	high
EXPRESSION LEVEL	high	low - high	low - high	low - moderate
EXTRACELLULAR EXPRESSION	secretion to periplasm	secretion to medium	secretion to medium	secretion to medium
POSTTRANSLATIONAL MODIFICATIONS				
PROTEIN FOLDING	refolding usually required	refolding may be required	proper folding	proper folding
N-LINKED GLYCOSYLATION	none	high mannose	simple, no sialic acid	complex
O-LINKED GLYCOSYLATION	no	yes	yes	yes
PHOSPHORYLATION	no	yes	yes	yes
ACETYLATION	no	yes	yes	yes
ACYLATION	no	yes	yes	yes
GAMMA-CARBOXYLATION	no	no	no	yes

1.3 Upstream processing

This research begins by considering a multitude of bioprocesses, however the focus of the later stages of research was on bioprocesses employing microbial expression hosts. This was due to the availability of microbial upstream and downstream datasets for tool testing (Chapters 5 and 6). Microbial expression systems are typically used for proteins with no or simple post translational modifications, or those which can be modified post translation chemically after cell fermentation. A microbial cultivation, or indeed a cultivation of any cell type, will involve four stages of growth: the lag, log, stationary and death phases (Figure 1.1). In the lag phase the bacteria will be starting to double in number with each generation, causing a slow rise in cell number as the micro-organism adapts to the conditions for the cultivation. In the log phase exponential growth is observed as the cell doubling causes this sharp rise in number. At the stationary phase, the rate of cells being produced is equal to the rate at which cells are dying. It is during the log and stationary phases of cell culture where the cells are the most productive in terms of product generation. During the stationary phase toxic metabolites will start to accumulate, and the nutrient supply will be depleted, which will cause a shift into the death phase, where cells are being produced more slowly than they are dying.

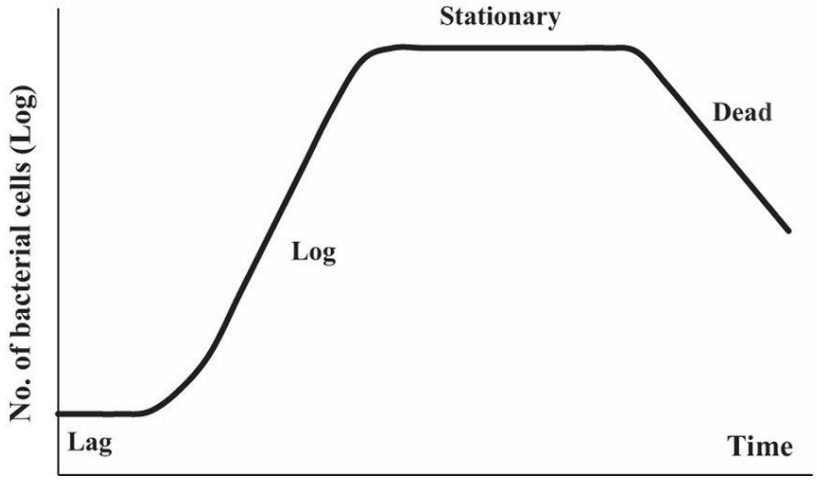


Figure 1.1 - Stages of growth within a cultivation. Taken from Wang *et al.* (2015).

Microbial cultivations generally use simple media, and have a rapid generation time. This combined with a long history of use and regulatory approval makes them an attractive host system for simple products such as peptides, or precursor molecules where post cultivation processing can configure the correct product. The nutrient requirements of a cultivation will differ depending on the strain being used and the fermentation conditions being implemented, however in general the bacteria will require a carbon source, a nitrogen source and trace minerals. Microbial cultivations have been used for a variety of different purposes, over an extended period of history. These purposes include, but are not limited to, waste treatment, food and drink production and recombinant protein production.

There are a range of reasons for employing a recombinant protein expression system in industry or research. Research may use this approach to understand a protein in greater detail or for reverse genetic engineering, where the gene encoding a protein is available but the protein itself is not, and to investigate Structure-Activity Relationships (SAR) (Stewart *et al.*, 1986; De Lalla *et al.*, 1996; Chapman *et al.*, 1998). It may also facilitate development of novel proteins (Zoller, 1992). Industrial processes use recombinant expression systems to produce large quantities of a desired protein which may have a range of applications, including therapeutic. The protein may only be available from natural sources in small quantities, making extraction from the natural source economically unviable. Alternatively the natural source may be toxic or difficult to handle. One instance of this would be the chlorotoxin protein, which is scorpion derived. In this case, although the protein has clear therapeutic potential (Xiang *et al.*, 2011; Graf *et al.*, 2012), the associated handling of a large number of scorpions would make the industrial process both logistically complex and dangerous. The alternative to this situation is to transform a cell line to express the chlorotoxin, making large volumes easy to obtain and simplifying the extraction and purification processes. This holds the additional benefit of reducing the ethical concerns, and

makes any product suitable for vegetarians. The production of insulin from animals including pigs had not only made it unsuitable for vegetarians, but the differences in structure made it unreliable and often unpredictable. The advent of homologous expression systems has eliminated this problem, as the human insulin protein can be produced in large volumes using cell culture (Bell *et al.*, 1984; Chen *et al.*, 1995). While there are a variety of reasons for employing recombinant protein production, the most lucrative market is undoubtedly for production of therapeutic proteins, an important part of the growing pharmaceutical market.

1.4 Downstream Processing

Downstream processing within bioprocessing is generally comprised of centrifugation, chromatography and filtration, in various combinations. Cell lysis will be included if the product is intracellular, and can employ mechanical or chemical mode of actions. Past the initial purification stage downstream processes can vary widely depending on the product and host, and any further chemical processing required. However monoclonal antibody production has become well understood, with generalised platform processes being found to be broadly applicable (Birch and Racher, 2006; Kelley, 2007; Shukla *et al.*, 2007; Hogwood *et al.*, 2013). Within these platforms Protein A purification is generally the most expensive stage of the process. The purification of monoclonal antibodies, and associated challenges, has been discussed at length by Sommerfeld and Strube (2005), Shukla *et al.* (2007), and by Shukla and Thömmes (2010). Organisations have been active in the pursuit of an alternative technology, discussed in detail by Ghose *et al.* (2006), but the high efficiency of Protein A chromatography, combined with high levels of understanding and a well-documented history of use, mean that it remains an attractive process choice, despite the associated cost. An increase in titre, with claims of titres in excess of 10g/L (Kelley, 2009), has increased the potential for profit from each batch; however it has also increased the burden on downstream processing to be able to purify such concentrated solutions.

Alternative modes of chromatography are detailed in Table 1.2, with associated references for more detailed reviews of mode of action and applicability.

Table 1.2 - Types of chromatography available with accompanying references.

MODE OF ACTION	OVERVIEW	
AFFINITY CHROMATOGRAPHY	Exploits interactions between molecules to separate impurities from the desired product (e.g. Protein a chromatography)	*(HOBER ET AL., 2007)*
IMMOBILISED METAL ION CHROMATOGRAPHY (IMAC)	The product displays a tag which binds selectively to the metal ions within the column (e.g. His-tags)	*(BLOCK ET AL., 2009)*
ION EXCHANGE CHROMATOGRAPHY (IEX)	Separates molecules based on their isoelectric points (e.g. Anion exchange/cation exchange)	*(STANTON, 2004)*
SIZE EXCLUSION CHROMATOGRAPHY (SEC)	Separates molecules based on their size	*(CALIBRATION ET AL., 1994)*
HYDROPHOBIC INTERACTION CHROMATOGRAPHY (HIC)	Uses hydrophobicity properties of the product to separate from impurities	*(OCHOA, 1978)*
MIXED-MODE CHROMATOGRAPHY (MMC)	Incorporates multiple modes of chromatography on a single resin	*(MCLAUGHLIN, 1989)*

Recent trends point to improving abilities to obtain higher titres in upstream processing (Kamachi, 2016; Chen *et al.*, 2017), and in light of this there has been a shift within the bioprocessing sector from considering the upstream product production to be the limiting factor for final product yield to the downstream capacity becoming the limiting factor (Gronemeyer *et al.*, 2014; Pinto *et al.*, 2015). In light of this, it is important that any tools

developed within this research are applicable to both upstream and downstream production to ensure the bottleneck can be addressed regardless of where in the process it is occurring.

1.5 The Biopharmaceutical Industry

The term bioprocessing can cover a range of sectors, including waste water treatment, biological therapeutic production, biofuel production, and even food production e.g. marmite (Hassan and Heath, 1986; Grady Jr et al., 2011; Bornscheuer et al., 2012; Cheng et al., 2012; Liu et al., 2014; Marmite Museum, 2015). The tools developed within this research were designed with broad applicability in mind, particularly within Chapter 3. However, the focus of the research has been on biopharmaceutical processing, due to the highly competitive nature of the market creating a clear need for streamlined process development, which effective knowledge management has been shown to support (Pan and Scarbrough, 1999).

Therapeutic pharmaceutical developments, and advances in diagnostics, have been a major contributor to not only the increase in life expectancy, but also the rise in quality of life. When the sponsor of this research, Britest, was established in 2001, the average life expectancy globally was 66.7 years. By 2015 when this research was in progress, this had extended to 71.4 years. In the UK alone the rise was from 78 to 81.2 years (WHO, 2017). The pharmaceutical drug market is worth billions of pounds each year, and this is increasing year on year as new drugs are discovered, new processes for production developed and new diseases emerge. Altogether the top ten pharmaceutical companies had revenue values in 2016 in excess of US$440 billion (Datta (2016), Table 1.3), and sales values are set to rise over the course of the next ten years.

Table 1.3 - Top ten pharmaceutical companies by revenue in 2016. Sourced from Datta (2016).

COMPANY	TOTAL REVENUE IN 2016 (US$bn)
JOHNSON & JOHNSON	$70.1
BAYER	$51.4
NOVARTIS	$49.4
PFIZER	$48.9
ROCHE	$48.1
MERCK & CO.	$39.5
SANOFI GENZYME	$34.5
GILEAD	$32.6
ASTRAZENECA	$24.7
GLAXOSMITHKLINE	$23.9

To generate revenues of this magnitude, pharmaceutical companies have multiple products, often for multiple indications, and they invest significant amounts of their money into research and development for drug discovery and development. Pipelines can be extensive (Citeline (2014), Table 1.4), and billions of dollars are spent on Research & Development (R&D) each year to maintain market share (Carroll (2016), Table 1.5/Figure 1.2).

Table 1.4 - Top ten pharmaceutical companies by number of pipeline drugs in 2014. Sourced from Citeline (2014).

COMPANY	DRUGS IN PIPELINE
GLAXOSMITHKLINE	261
ROCHE	248
NOVARTIS	223
PFIZER	205
ASTRAZENECA	197
MERCK&CO	186
SANOFI	180
JOHNSON & JOHNSON	164
BRISTOL-MYERS SQUIBB	133
TAKEDA	132

Table 1.5 - Top ten pharmaceutical companies by R&D spend in 2015. Sourced from Carroll (2016).

COMPANY	R&D SPEND IN 2015 (US$bn)
ROCHE	$9.7
JOHNSON & JOHNSON	$9.0
NOVARTIS	$8.9
PFIZER	$7.7
MERCK & CO.	$6.7
BRISTOL-MYERS SQUIBB	$5.9
ASTRAZENECA	$5.6
SANOFI	$5.6
ELI LILLY	$4.8
GLAXOSMITHKLINE	$4.4

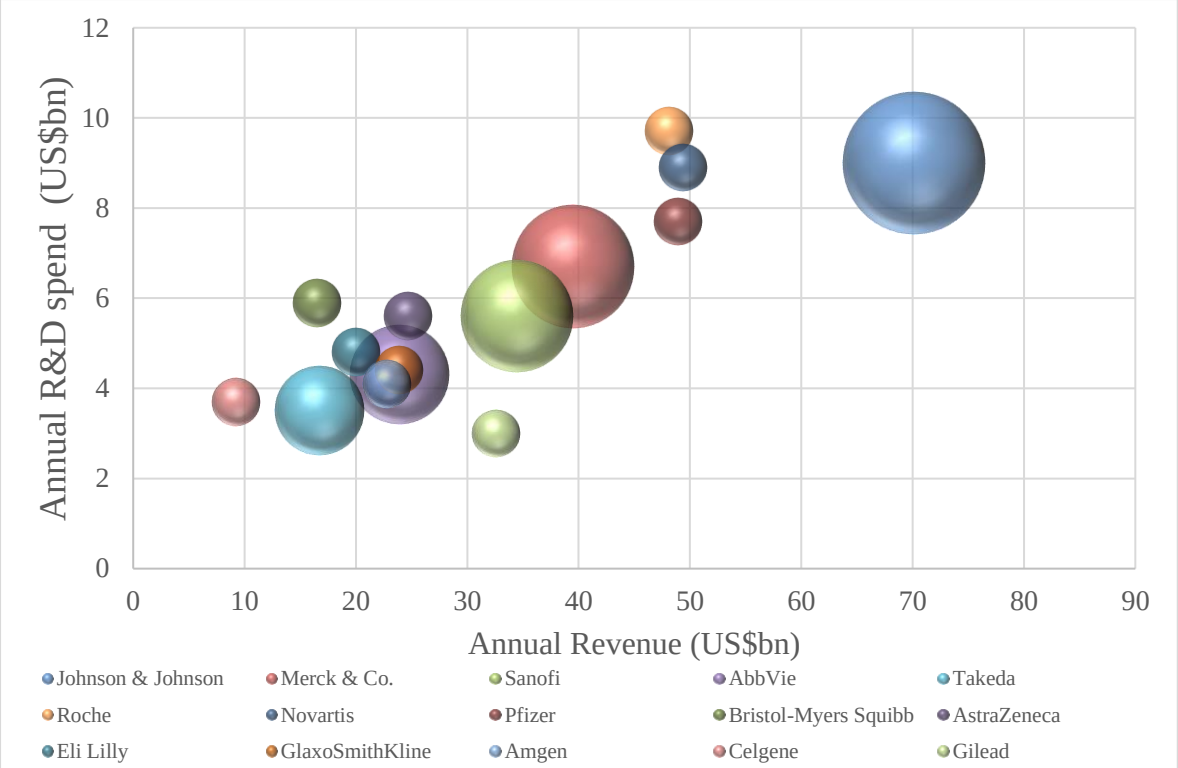


Figure 1.2 – Bubble chart showing the top 15 pharmaceutical companies total revenue compared to R&D spend in 2016. The size of the bubbles corresponds to the size of the company.

These R&D programmes give rise to multiple drug candidates, for a range of indications, however the risky business of pharmaceutical production often results in promising lead molecules being rejected after significant sums of money have been invested in development. At the time of writing, there are 622 drugs in Phase I clinical trials, 597 in Phase II and 285 in Phase III (DataMonitor, 2016). Studies have shown that 10,000 drug candidates must be investigated to give rise to a single patented molecule (Figure 1.3, Guilfoyle (2016)). This is generally due to adverse effects encountered during trials, or the drug showing a lack of efficacy. As a result, pharmaceutical companies must have the R&D costs for 10,000 candidate drugs to be covered by a single successful drug product. This has led to a high value market, where the ability to predict a drugs performance or manufacturability comes with a high value.

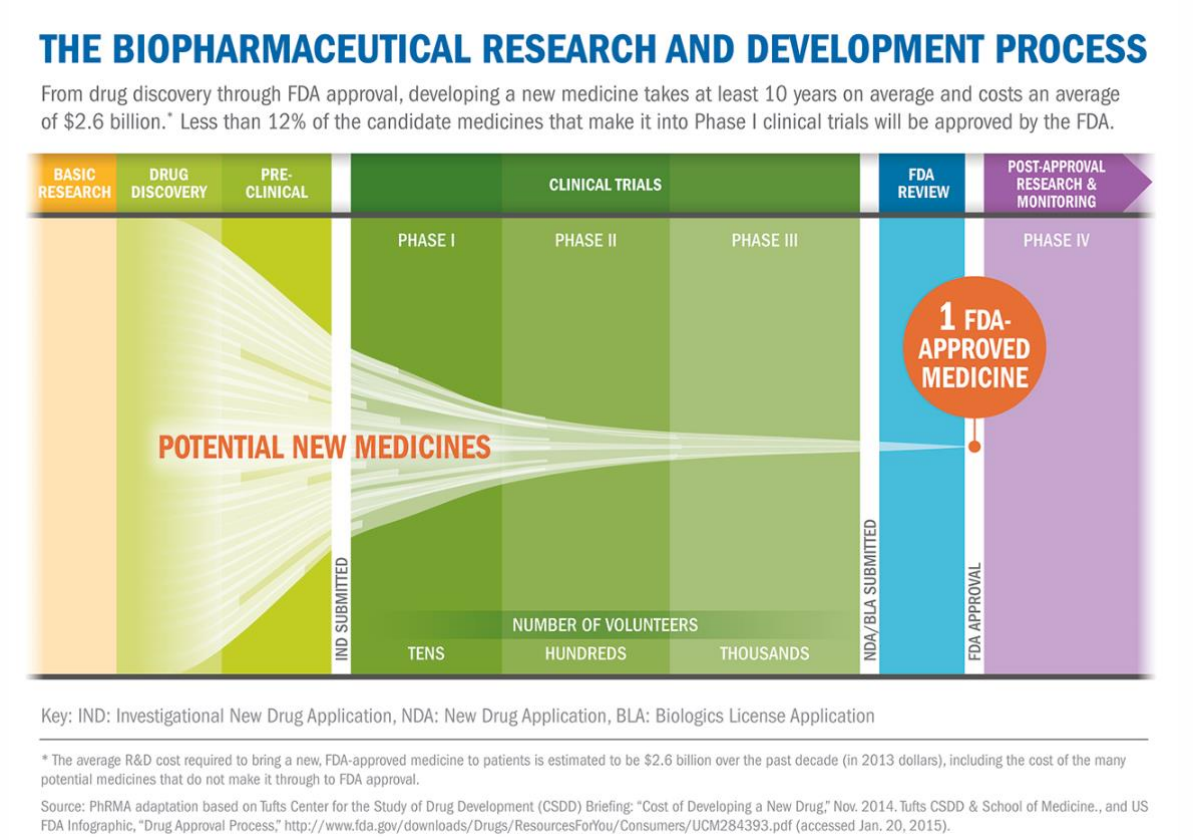


Figure 1.3 - The number of drug candidates progressing at each stage of a pharmaceutical development pipeline. Reproduced from Guilfoyle (2016).

The pharmaceutical landscape is changing every year, however in 2016 the majority of R&D pipeline outputs were anticipated to be in oncology (Figure 1.4). Oncology is a large market ($107 Billion in 2015, IMSHealth (2016)), which covers a variety of diseases, each of which have associated variations. Spanning the breadth of the pharmaceutical landscape, chemical compounds have previously been at the heart of the development pipeline, but advances in recent years have made biologics serious contenders as treatment options for a range of conditions. In 2015, just under 3,000 biological products were either marketed or approved for market (DataMonitor, 2016), treating a range of conditions from diabetes to Multiple Sclerosis to wrinkles. Though the benefits are clear, the production of biological products can be problematic due to the uncertainty and variability associated with live biological systems, and the range of expertise required to design a successful bioprocess.

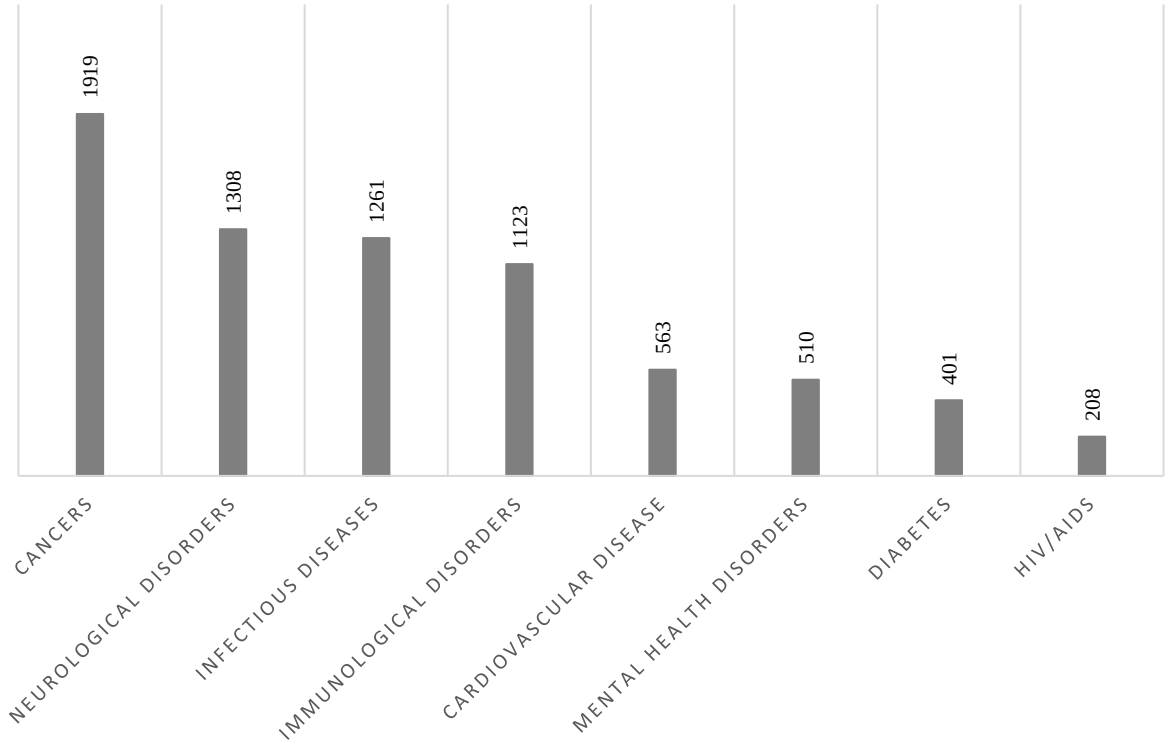


Figure 1.4 - Biopharmaceutical pipeline drugs by indication in 2016. Sourced from Guilfoyle (2016). The numbers indicate the exact number of drugs in development for each indication at the time of writing.

1.6 Knowledge Management

The competitive nature of the pharmaceutical industry is clear, and the potential for a company to make significant profit in the biologics sector is considerable. This being said, the high failure rate of potential drug candidates, combined with long development times and a reduced patent lifetime (due to the length of time candidates take to develop), means that companies operating in this space must maximise their efficiency to successfully tap into this market potential. In an interdisciplinary sector such as bioprocessing, the successful management of the different types of knowledge is vital for efficient process design, and so knowledge management (KM) could be a useful technique to maximise potential value within a business.

Knowledge can exist in many forms within an organisation, broadly being split into tacit and explicit (Nonaka, 1994; Polanyi and Sen, 2009). Explicit knowledge is easily communicated

and captured (Duffy, 2000), either through technical documentation, operating procedures or data. This knowledge can be transferred between individuals or departments with minimal requirement for formal transfer activities. Conversely, tacit knowledge could be the beliefs or viewpoint of an individual, or the application of ability (Scott, 1998). This is more difficult to communicate and transfer, and as a result formal KM techniques have been created to attempt to simplify tacit knowledge transfer and capture. It is common that the two cannot easily be separated, and that some tacit knowledge can be required to successfully apply or understand explicit knowledge (Wakefield, 2005).

There are many management techniques which can be employed in a multitude of sectors to aid in the application of knowledge management. Knowledge Management was defined by Bassi (1997) as the "creation, acquisition, sharing and utilisation of knowledge for the promotion of organisational performance", and within Quality by Design (QbD) as "a systematic approach to acquiring, analysing, storing and disseminating information related to products, manufacturing processes and components" (I.C.H Guideline, 2008). Many types of knowledge exist within a business, but the presence of knowledge does not always mean that the knowledge is fully utilised. Knowledge can be used to achieve a desired outcome, or indeed to avoid a negative outcome. Knowledge was said to be only part of a larger relationship within a successful business (Andersen, 1999); this relationship is shown in Figure 1.5.

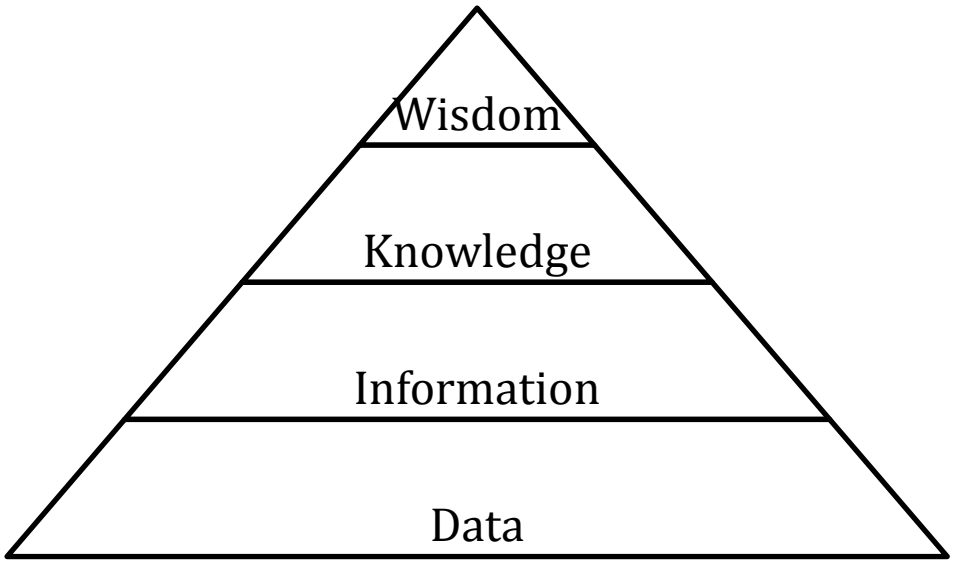


Figure 1.5 - Showing the relationships between data, information, knowledge and wisdom. Adapted from Andersen (1999).

It is clear from Figure 1.5, that while knowledge is important to improve a process, it is not in itself a way of determining best practices. Data is required to generate information, and from this knowledge can be assembled. Armed with this knowledge, a company can seek wisdom, the use of this knowledge to change the company reaction to a situation. This was devised with respect to the business model; however parallels could be drawn between this hierarchy and the Quality by Design initiative in bioprocessing. The data, in the case of QbD would be the readings from probes and results from analytical methods. The information would be features or characteristics that could be inferred from these results or readings. The knowledge would be the understanding of whether these readings and results were conforming to predetermined quality standards. The wisdom to be able to act on this information within a pre-defined parameter space is the underlying principle of QbD.

Capturing and using knowledge can be a powerful ethos within a business; a central store of information can be invaluable not only for troubleshooting purposes, but also to allow the best decisions to be reached first time. Making information available to employees can aid their understanding of a process or business, and its effective sharing through the business can facilitate communication between departments (De Vries *et al.*, 2006). The most effective

knowledge stores employ a structured approach, to ensure straightforward navigation and full utilisation of the resource within the business (Wen, 2009).

The high volume of data generated by companies has meant a trend towards using software and databases for this capture and storing of knowledge. While this can be invaluable for raw data storage, such as readings from probes, it can make navigation and analysis of this information difficult for the individual. The programmes used can be complex to implement and run (Liao, 2003). In addition, it is not unimaginable for a company to employ the program as the solution to knowledge sharing, rather than as part of a larger company ethos. In fact if the information is not used then it is of little benefit to the company. There is additionally the ongoing battle with maintaining the database, not only to ensure the information within it is up to date, but also to ensure it is running effectively (Liao, 2003). However, each of these is transcended by the difficulty in obtaining tacit knowledge from employees. While there is much to be gained from readings and measurements, the experience of operators of a process can be as valuable, if not more so.

Many knowledge elicitation techniques are available and have been used in bioprocessing. The KATKit was one such system developed previously (Ranjan *et al.*, 2002), which focused on how to best draw out the relevant knowledge from process experts. The early stages of the KATkit system involved knowledge elicitation using a unique exception logic, which was used to create rules for the various fermentations running at an industrial partner site (Eli Lilly). The knowledge elicitation technique relied on an independent elicitation facilitator running the sessions, and documenting the outcomes. These were then coded into a software based control system to be implemented on the site. While this gave a significant benefit to the company, the requirements in terms of time were significant (many person months), and the requirement for an independent elicitation expert trained in the KATkit approach made it

unsuitable for large scale integration into Eli Lilly. However, the approach itself was shown to be a valid knowledge elicitation technique which could add value to a bioprocess.

Different knowledge management strategies suit different organisations (Kim *et al.*, 2014), and there is not a one size fits all approach. With that in mind, this research sought to ascertain the potential value of the Britest approach for bioprocessing. Bioprocessing is an industry which relies on efficient interdisciplinary working, and the effective management of the different areas of knowledge within a plant or process can be key to the success of a bioprocessing business. Experience in working on a plant or process is invaluable to process development, and this is demonstrated through the expanding Contract Manufacturing (CMO) market within bioprocessing. The experience a CMO derives from working on a variety of products is invaluable, and is the reason that the CMO market in bioprocessing is expanding (Stanton, 2015).

However, even in the established CMOs, the sharing of this knowledge relies on effective communication skills. Technology transfer and process design are core capabilities for organisations within contract manufacturing in any sector, however for these to be effectively employed communication skills are key (Santoro and Gopalakrishnan, 2001). However in companies not specialising in this, effective communication can be problematic. This becomes increasingly difficult when bioprocesses are involved, as the range of skills required is broad. Effective knowledge capture and management techniques have previously been examined in relation to technology transfer (Salazar Alvarez, 2003; Wakefield, 2005), and it has been shown that by employing KM techniques to streamline communication channels organisations can reduce the number of mistakes made during development, and potentially therefore increase organisational effectiveness and reduce time to market (Pan and Scarbrough, 1999; Ofek and Sarvary, 2001).

1.7 Quality by Design

The ability to successfully manage the knowledge within a process is undoubtedly valuable, especially within a sector involving so many different disciplines such as bioprocessing. The extended time it takes to market a biopharmaceutical, combined with the tremendous associated costs, makes every potential saving of significant importance. The highly regulated environment that pharmaceutical companies operate in only adds to the pressure to perform in a maximally efficient manner. In light of this, any approach which can be adopted to give a competitive advantage could add significant value to a pharmaceutical company. The concept of incorporating Quality by Design into pharmaceutical production has therefore been met with great interest by companies operating in this space.

Quality by Design was originally defined as "A maximally efficient, agile, flexible pharmaceutical manufacturing sector that reliably produces high quality drug products without extensive regulatory oversight" (Woodcock, 2005). The traditional approach to producing pharmaceutical drug products was to follow a set protocol, with the aim of achieving a consistent result. However, this does not account for changes in raw material quality, environmental influences, and other uncontrollable factors. The QbD approach, in its simplest sense, allows for those variations to be taken into account, and the process changed within certain parameters to counteract the sources of variability. The range of conditions the process can operate within is termed the design space (FDA, 2006), and the ability to move the process around this design space to obtain a consistent product quality is the driver behind QbD. To achieve a QbD approach in a process, the various parameters making up the design space must be measured and controlled. In light of this, the Process Analytical Technology (PAT) guidelines followed from the QbD guidelines (FDA, 2004). PAT, in its broadest sense, covers the instrumentation and techniques used to ensure the process remains in its allocated design space.

The process would typically not operate within the whole of the design space. Generally the scheme outlined in Figure 1.6 would be followed, where the characterised space is large, the design space is a smaller part of the characterised space, and the control space, where the process operates, is smaller still.

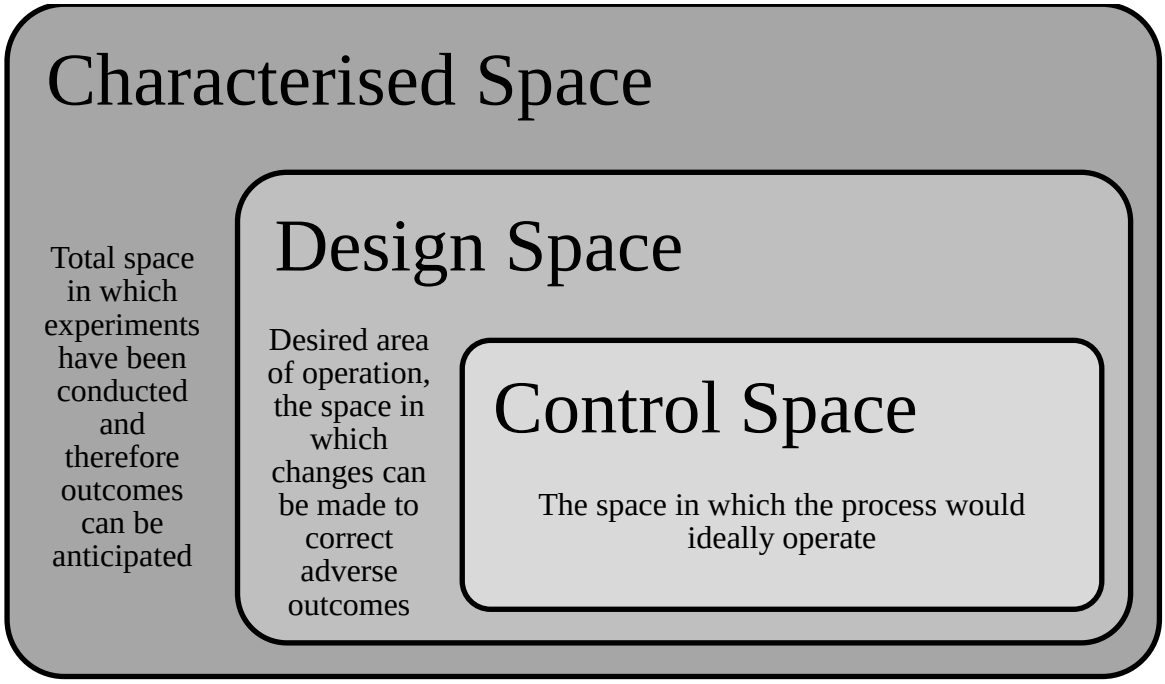


Figure 1.6 - Schematic of the relationships between characterised, design and control space.

The characterised space defines the whole area of knowledge. It encompasses not only the right result from the process, but also the potential wrong results. These may at times be too extreme to correct through a change in processing, and so the design space is a smaller subset in which corrections can be made to ensure product quality. Within the design space, some corrective measures may be too extreme or costly to feasibly implement, and so the control space becomes the smaller space within this in which making the changes would be a viable option, both economically and safely.

Each company will have an individual approach to defining the design and control space. The varying strategies will have varying degrees of robustness, but are generally based on a combination of process understanding and experimentation. There is not currently a standard approach which is recommended, and this means there can be no guarantee of the robustness of the design space identified. Harms *et al.* (2008) attempted to define the design space for fermentation of *Pichia pastoris*. To achieve this they first characterised the process risk using Failure Modes and Effects Analysis (FMEA), followed by the development of a scale down model. This was followed by characterising the process. The resulting design space was a combination of temperature, pH and dissolved oxygen, which were all defined as key process parameters. Three Optical Density (OD) readings, at the start, of the inoculum and at induction, were all shown not to impact the process performance, in addition to the feed rate. As a result these were not included in the design space. While this approach did create a design space for the fermentation, it would be difficult to replicate in processes where scale down alternatives were not available, or in whole process examples (Harms *et al.*, 2008). The temptation could be to create a design space for each unit operation, however as Zhou and Titchener‐Hooker (1999) have shown; adopting a Windows of Operation approach is more effective for optimising the process outcome. Performing the same level of characterisation and risk analysis for a whole process, particularly for mammalian cell culture based processes with their associated high complexity, would be challenging and may not generate a design space with an associated high level of confidence.

In addition to facilitating the QbD approach to processing, effective knowledge capture has been correlated with organisational effectiveness (Gold and Arvind Malhotra, 2001), and many ways of facilitating this capture are available. Knowledge management in its entirety has been identified as possibly the biggest challenge for QbD implementation. Indeed, it has been claimed that without effective knowledge management approaches, it is not feasible to

understand how the attributes of a product affect the safety and efficacy of the product (Herwig *et al.*, 2015), and by extension it is therefore difficult to see how effective QbD manufacturing processes could be implemented without these KM systems.

The knowledge required to implement a QbD approach is outlined in Figure 1.7, taken from (Herwig *et al.*, 2015). This clearly demonstrates the importance of effective KM strategies over the product lifecycle to the stage of manufacture. As the life cycle progresses and intellectual property protection such as patents expire the importance of understanding only increases, as efficiency must be improved to maintain the economic viability of the product.

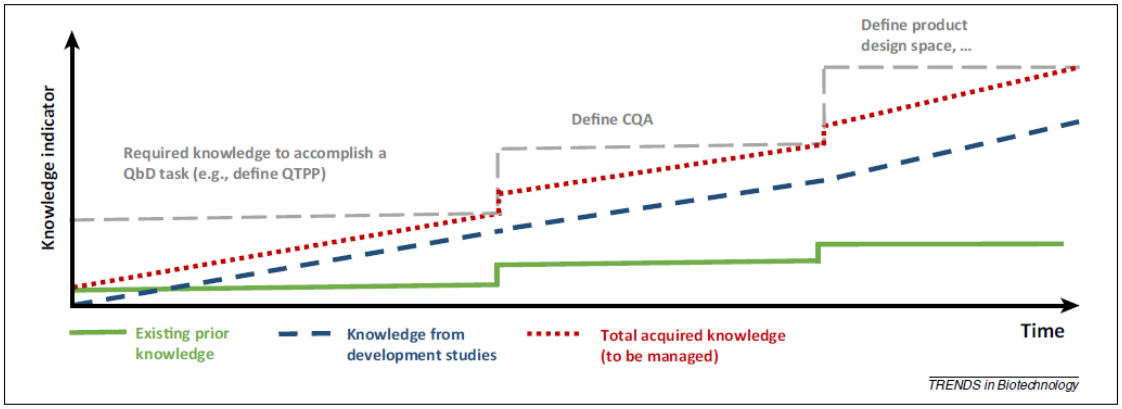


Figure 1.7 - Knowledge required at each stage of the bioprocess development timeline. Taken from Herwig *et al.* (2015). Abbreviations: QbD-Quality by Design, QTPP – Quality Target Product Profile, CQA – Critical Quality Attribute. Knowledge indicator is the total amount of required knowledge, shown here compared to the stage of development.

Process improvement and adoption of QbD through KM can be achieved through various tools, which can include but are not limited to data capture, text mining, visualisation tools, statistical analysers, and collaboration tools (Steinberg and Bursztyn, 2010; Schild and Fuchslueger, 2012; Turkay *et al.*, 2013; Giridhar *et al.*, 2014; Otasek *et al.*, 2014). The range of tools to suit an array of purposes is wide, and suitability will depend on a range of factors. One of the key reasons for employing KM strategies and QbD processes is process improvement. This thesis will first discuss two potential options for process improvement tools currently employed within bioprocessing, BioSolve and Six Sigma, before moving on to discuss the Britest methodologies, the focus of this research, in Chapter 2.

1.8 Process Understanding Tools

The following BioSolve case study was sourced from BioSolve directly, with the aim of illustrating the industrial applicability of tools of this nature. Results generated are the work of BioSolve, and have not been generated as a part of this research.

1.8.1 BioSolve

BioSolve is the core product from the company BioPharm Services, which was established in 1998 to create problem-solving software aimed at biologics, facilities and business strategy. BioSolve is designed to aid in the decision making process on a biological process. It aims to reduce manufacturing costs and aid the decision making process by incorporating the business perspective, rather than relying on the underlying science alone. The results are generated based on financial and process information, such as costs, timings, profit, materials and sales value.

Many approaches have been taken to improving the decision processes associated with monoclonal antibody production, and one of the most well-known examples would be that constructed by the C.M.C. Biotech Working Group (2009). This example is widely cited as an approach to implementing a QbD approach to a biological process. It relied on fundamental scientific understanding combined with scoring systems, designed to be used in conjunction with cost benefit analysis. An additional case study of using BioSolve on a Monoclonal Antibody (mAb) production process was constructed by BioPharm (2014), focusing more on the financial and numerical analysis than the underlying science of the process. It aims to address three main areas of concern in mAb production, outlined in Figure 1.8.

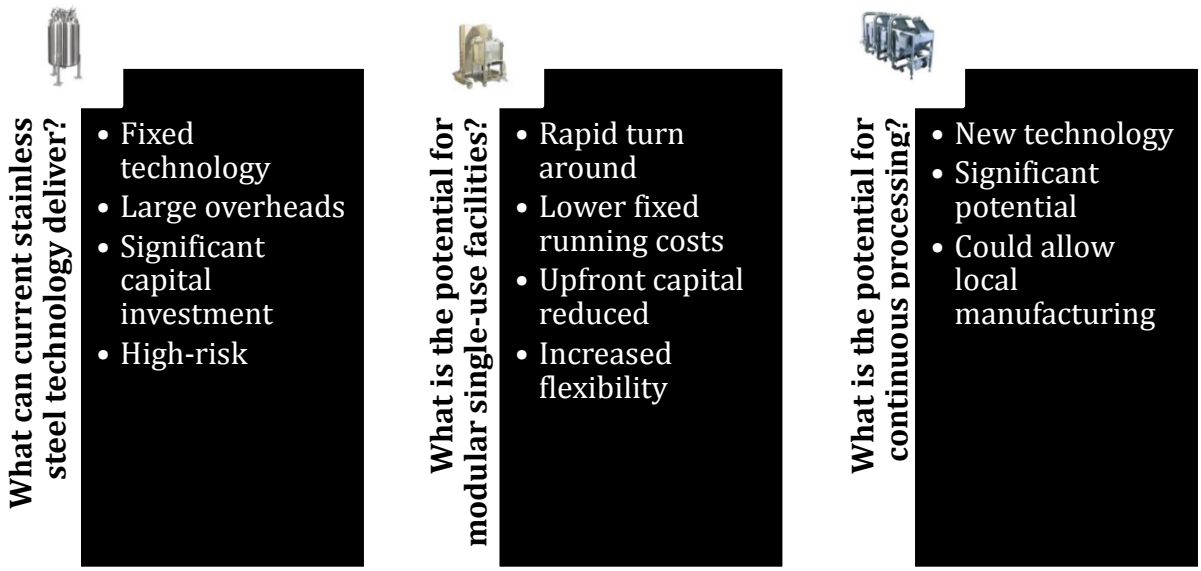


Figure 1.8 - Three areas of investigation and resulting conclusions from the case study of mAb production using BioSolve. Taken from BioPharm (2014).

The case study constructed by BioSolve considered each of these three questions in turn. Stainless steel is not a flexible technology, and as such the process must be robust and the market well established to make this a viable choice. This is a high risk approach, and the BioSolve software could be applied by a user when working on a process to investigate the impact of scale and titre on the capital investment required, allowing the user to make a decision based on both science and business case information.

In terms of single use facilities, the capital investment costs are lower, and there is a much greater degree of flexibility. For this case study, BioSolve was used to determine the harvest strategy which would give the highest yield without negating the increased productivity with the associated cost increase. The optimum option for pooling was also considered. The capital investment required was $250 million, a saving of over a million dollars when compared to the stainless steel version of the same process.

BioSolve was also used within the case study to investigate the potential of continuous processing. For this business case, a perfusion titre was set at 1g/L, lower than the fed batch titre of 5g/L. The process modelled was a 2000L bioreactor scale, which is smaller than the scales for the previous two business cases. However, the capital investment was found to be smaller, and the source of the biggest costs could be attributed to resin and media costs. When compared to the stainless steel option, a reduction was seen in cost of goods (down 10%), and upfront capital investment was reduced by 73%.

In summary, BioSolve allowed the user to make process decisions based not only on the scientific or engineering merit on an option, but also on the business case being presented. The models can be customised to a specific process or market, giving the user enhanced functionality and applicability. BioSolve can be a valuable asset to a company wishing to explore options for processing without expensive experimentation or building complex mathematical models. However, it relies on user information being correct, and the correct interpretation of the results to provide the full benefit. It also cannot analyse the fundamental science behind the process, or suggest alternatives which have not been input. The tool will improve in performance as more data is available, which is a limitation if large datasets are not available for a particular process, and conclusions are limited to the conditions in which there is data available. While it has clear potential benefits, in itself it will not increase interdisciplinary working, and like any tool should only be employed where suitable and not as a quick fix to a problem.

1.8.2 Six Sigma

Six Sigma is a methodology which is regularly employed across a multitude of sectors to enhance process efficiency. It was developed in 1986 (Motorola, 2009), and is currently used in a range of process sectors (McClusky, 2000; Buss and Ivey, 2001; Antony and Banuelas,

2002; De Feo and Bar-El, 2002). It is a set of tools for process improvement that has a base in statistical analysis and predictions. The Six Sigma approach is based on three assertions:

- Continuous improvement is vital to running a successful process
- Processes have both business and engineering aspects that can be measured, analysed, controlled and improved.
- To obtain the best process, all levels of the business must be committed to improving the process.

The Six Sigma process is outlined in ISO 13053:2011 (2011). Within Six Sigma there are two methodologies for a project, one for improving an existing process (DMAIC), and one for designing a new process (DMADV) (De Feo and Barnard, 2003). Within both of these are five phases, outlined in Figure 1.9.

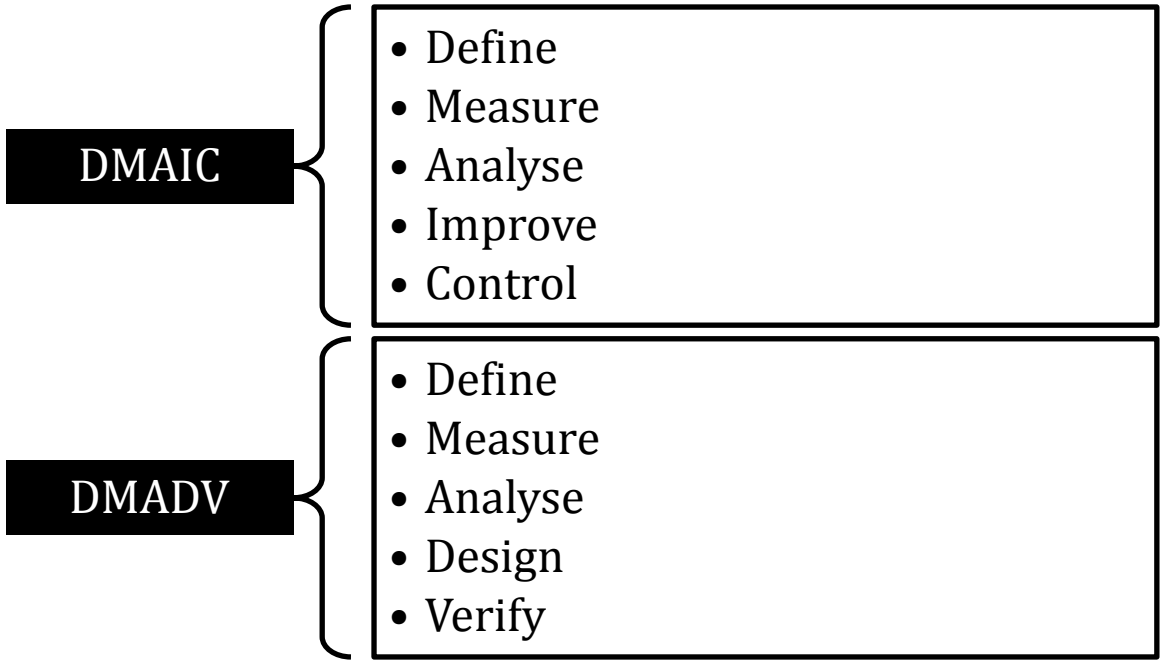


Figure 1.9 - Acronyms for two methodologies employed in Six Sigma.

Both the DMAIC and the DMADV start with the definition, measuring and analysis, for understanding of the system and alternative process designs respectively, of the system. From

here, the DMAIC methodology moves onto optimising the process using tools from within the toolkit. In contrast, the DMADV moves onto designing the new process using the results from the preceding analysis step. The DMAIC ends with the control stage, to correct deviations from the desired outcome before they result in a whole process failure. The DMADV methodology ends in the verification of the design, through experimentation and pilot runs, prior to running the process at full scale. A range of tools are encompassed within these methodologies, some based in statistical analysis, and some thought process tools.

With respect to bioprocessing, Dassau *et al.* (2006) employed the methodologies alongside process modelling techniques to consider a penicillin fermentation. After three cycles of Six Sigma evaluation of the process, the final conditions led to a 40% reduction in batch time, a 17% increase in throughput yield and a 33% reduction in impurities. The authors attribute the success to the adoption of a plant-wide approach to process improvement, previously discussed (Zhou and Titchener-Hooker, 1999), which would not have been adopted without the aid of the Six-Sigma methodologies. The adoption of a whole process view requires a shift in organisational culture, and the use of knowledge management tools to aid this shift was undoubtedly beneficial in the case presented by (Dassau *et al.*, 2006)

While the Six Sigma approach has many advantages, including wide applicability and a statistical basis for improvement, there are some drawbacks. The use of Six Sigma is within an organisation, and so any lessons that can be learned are only internal and not from other companies. Limited cross sector learning has been highlighted as a weakness within the pharmaceutical industry (Smith, 2014), and methods which promote looking only internally for improvement could potentially limit the improvement to the process and effective innovation. The tools follow a rigid structure, and while this could be considered a benefit, it could also limit the ability of the methods to diagnose a problem.

While both BioSolve and Six Sigma are viable options for enhancing process efficiency, neither will capture the fundamental science, nor explore process development options at an early stage of development. Both of these requirements are key to facilitate the adoption of a QbD approach in bioprocessing. BioSolve is a valuable tool for economic analysis, but cannot incorporate the underlying science behind the process design, and this could have a resounding impact of the process design. For example, if there was a technical feasibility impact for a processing option, this could not be incorporated other than as a cost. Conversely Six Sigma can be used to aid the capture of the fundamental science, but follows a set structure and does not necessarily lend itself to innovative process design.

1.9 Britest

This thesis is focused on the development of knowledge capture tools for application to bioprocessing specifically, starting from the Britest tools which were developed for chemical and physical processing. The research has been undertaken for an Engineering Doctorate, and is therefore sponsored industrially, in this case by Britest Ltd.

While tools such as BioSolve and Six Sigma can be employed for the continuous improvement of processes, and to investigate the impact of changes, Britest operates at a more fundamental level. The Britest tools aim to capture and explore the underlying science of the process, facilitating interdisciplinary communication and capturing the specialist knowledge of each discipline in a structured manner. The tools were initially developed as part of an EPSRC funded collaboration between The University of Manchester Institute of Science and Technology (UMIST), Imperial College London and University of Leeds, and in 2001 Britest was formally established as a company to maintain and develop the tools and methodologies. The name Britest was created from the acronym:

Best **R**oute **I**nnovative **T**echnology **E**valuation and **S**election **T**echniques

It is a not-for-profit membership-based organisation, currently encompassing twenty industrial and academic partners. The industrial members and associates are drawn from the pharmaceutical, fine and speciality chemical sectors. The inclusion of academic members, including Newcastle University, is intended to bridge the gap between academia and industry, thus ensuring that academic developments are applied to real-world problems. The open innovation model promoted by Britest gives members access not only to the tools, methodologies and enablers, but additionally to the knowledge and experience of other members. This provides an avenue for open discussion of processing problems between organisations, allowing each member to draw on the expertise of others without worries of confidentiality breaches. Often the answer to a process problem may come from a different industry sector, which would be less likely to be generated outside of the Britest consortium.

This research arose from a need identified within the consortium for application of the Britest tools to bioprocessing. At the time, Britest identified a growing interest within the consortium around bioprocessing. The project was approached from an academic perspective with an academic consortium member (Newcastle University). The aim of this research was to test the tools on a range of bioprocesses, and to investigate the potential for application to bioprocessing as a whole. The industrially based nature of an EngD means that the research must not only advance an area of knowledge, but also provide a business benefit for the sponsoring company. This thesis advances knowledge management within bioprocessing, and the resulting toolkit allows Britest to pursue recruiting new members from the bioprocessing community.

The research presented within this thesis sought to answer the following research questions:

Can the Britest tools which have been developed for Chemical and Physical processing be applied to Bioprocessing? Do they add value? Are adaptations/modifications required?

The work developed novel knowledge management tools specifically designed for capture and transfer of knowledge generated within bioprocessing. The ability of these tools to capture relevant and useful bioprocessing knowledge was assessed by understanding the most important factors within bioprocessing from both a technological and economic perspective. These were designed specifically with the adoption of a QbD approach to processing in mind, as this was an area identified as being a current major challenge within bioprocessing where tools such as those contained within Britest could add significant value.

1.10 Thesis Structure

The chapters in this thesis present the work carried out over the course of the Engineering Doctorate study. This thesis begins by discussing the current Britest tools in detail (Chapter 2), to ensure the reader has a clear understanding of how the tools are intended to work and the form in which they existed prior to the commencement of this research. Chapter 3 moves on to discuss applying the tools to four virtual bioprocesses using SuperPro Designer, to test the potential applicability on a whole bioprocess without the constraints of a real industrial process. From here, a gap was identified to drive the development of the Interaction Analysis Table (IAT), and development and testing on upstream and downstream processes are discussed in Chapters 4, 5 and 6 respectively. The thesis concludes with Chapter 7 which investigates the sensitivity of the weighting system within the IAT, to better understand the potential limitations or drawbacks of the tool. Chapter 8 presents the research conclusions, and a summary of the impact the research has had on the industrial sponsor.

Chapter 2-The Britest Tools and Methodology

This chapter will present the reader with the information on the Britest tools, how each of them works and an example of when they were used. This shows how the tools would be used on a chemical or physical process to add value to a process, which is necessary to

understand prior to considering whether they would be applicable to a bioprocess. It also covers the background to the Britest tools and methodology.

Chapter 3-Virtual Bioprocessing

Chapter 3 will consider the application of the Britest tools to four separate bioprocesses, using virtual processes simulated in SuperPro Designer. These cover four main sectors within bioprocessing: high value low volume (Monoclonal Antibody production), low value high volume (Insulin production), secondary metabolite production (Penicillin V) and waste water treatment. The chapter will focus on adaptations made to the tools with respect to the insulin production process, as the following chapters focus on microbial processing. This chapter presents the Reaction/Reagent/Transformation Tracker (R2T2), a new tool which was developed in response to limitations identified by the simulated Britest study. It also identifies a need for a tool to facilitate linking Critical Quality Attributes (CQAs) to Critical Process Parameters (CPPs), in keeping with the QbD initiative.

Chapter 4-Interaction Analysis Table Development

This chapter focuses on the requirement identified in Chapter 3 for a tool to facilitate the linking of CQAs and CPPs, and develops the Interaction Analysis Table (IAT) for this purpose. A range of options for tool development are discussed, with the final tool being presented at the conclusion of the chapter.

Chapter 5-IAT Upstream Testing

This chapter tests the newly developed IAT tool on an upstream dataset from early stage process development on a microbial process. The dataset is a publicly available academic dataset from Technical University Berlin (TUB), where a range of processing conditions were tested with respect to production of Alcohol Dehydrogenase (ADH).

Chapter 6-IAT Downstream Testing

This chapter tests the newly developed IAT tool on a downstream dataset from early stage process development on a microbial process. The dataset is a publicly available academic dataset from Technical University Berlin (TUB), where a range of reagents were tested with respect to cell lysis.

Chapter 7-Sensitivity Analysis

Chapter 7 concludes the thesis by investigating the effect of variability in the weightings of an IAT on the outcome. Two sets of IATs were simulated using Microsoft Excel (2010), one set with five outcomes and the other with ten outcomes. Each IAT consisted of ten parameter rows, with relationships and weightings simulated using random number generators. Weightings were investigated to ±1, to ascertain the impact of the inherent variability on the outcome of the tool. The work identified factors which could be used reliably to infer sensitivity and confidence in the result without the need for complex simulations, allowing the Britest consortium to use the tool and to have an indication of the reliability of the outcome through using a simple calculation which can be performed by hand on an IAT of any size. This chapter concludes the work presented in this thesis, along with making recommendations for the implementation of the new Britest tools on bioprocesses, and suggestions for future developments to the toolkit.

Chapter 8 – Research Conclusion and Industrial Impact

Chapter 8 concludes the research, and the impact the research presented in this thesis has had on Britest, the industrial sponsor. It includes statements from Britest members around both the R2T2 and the IAT, the membership increase to Britest as a result of the work, and the John Borland award which was presented to the authors in 2016 in recognition of the innovative approaches used in the research.

Chapter 2 The Britest Tools and Methodology

2.1 Introduction

Chapter 2 will discuss the application of the Britest tools to the simple process of making a cup of coffee. This aims to give the reader a working understanding of the toolkit, and the potential benefits each tool brings to a process, in a format which can be related to a broad audience. It will introduce the company background and structure, before moving onto considering each tool in turn, explaining the features of each tool along with the benefits it could bring.

Britest began in 1998 as a joint industry/academic collaborative research project funded by the Engineering and Physical Sciences Research Council (EPSRC). The project, which included academic and industrial partners, established collaborative thinking on radical new process design methodologies that could lead to greater understanding and drive significant improvements in sustainable manufacturing. Output from this collaborative project generated a set of innovative tools and methodologies which allow the analysis of product development and manufacturing processes to demonstrate where and how major improvements could be made.

In 2001, Britest was formally established as a company to maintain and develop the tools and methodologies. It is a not-for-profit membership-based organisation, currently encompassing 20 industrial and academic partners. The industrial members are drawn from the pharmaceutical, fine and speciality chemicals sectors. The inclusion of academic members, including Newcastle University, is intended to bridge the gap between academia and industry, thus ensuring that academic developments are applied in real-world problems. The open innovation model promoted by Britest gives members access not only to the tools, methodologies and enablers, but additionally to the

knowledge of other members. This provides an avenue for open discussion of processing problems between organisations, allowing each member to draw on the expertise of others without worries of confidentiality breaches. Often the answer to a process problem may come from a different industry sector, and the breadth of the Britest consortium facilitates collaborations of this nature.

The Britest tools can help to identify the best opportunities for process optimisation, and also increase understanding of the process as a whole, across a range of disciplines, considering the process as a whole and acting to highlight the 'unknown unknowns' within a process to identify areas where more understanding could prove beneficial. As reported by Britest, these tools have generated over £1 billion of value to member companies since they were first introduced (Britest, 2017), and it is anticipated that expansion into new areas such as bioprocessing will see this figure rise.

This research project was sponsored by Britest to aid the move into the bioprocessing sector, which would not be possible without a working toolkit to extend value to existing members, and also attract new member companies/institutions. This was in response to a need identified by the consortium members, some of whom are involved with bioprocessing already. Bioprocessing relies on the combination of a range of disciplines working collaboratively. The people involved in bioprocess development have a range of backgrounds, including biology, chemistry, engineering (chemical, biological and even mechanical), statistics and business management. The complexity of a biological process and the range of people involved can make effective communication problematic, and this, combined with the unpredictable nature of biological systems, creates a challenging environment in which to operate. As discussed in Chapter 1, the Quality by Design (QbD) initiative from the FDA (I.C.H Guideline, 2009) has shifted the focus of bioprocess development teams, from simply developing a fixed process which works most of the

time, to the specification of a design space in which the process can operate flexibly to ensure a consistent product regardless of variation between factors such as raw materials. The specification of an effective design space has made it even more important for bioprocessing professionals to communicate effectively and capture the basic process understanding which has led them to their desired design space. In light of the importance of effective communication, and emphasis on better process understanding, bioprocessing is an area where tools such as the Britest tools could add significant value. It is this challenge which this research aims to tackle.

This chapter will introduce the Britest toolkit as it currently exists, developed for application to chemical and physical processing.

2.2 Britest Toolkit

The current Britest toolkit consists of a number of tools and methodologies, designed to be applied to processes in different ways depending on the problem under consideration. Each will be suitable for different parts of the process, and could highlight different unknowns within the same process. The tools were designed to be applied by multidisciplinary teams, and therefore ease of application is imperative. A Britest study is generally supported by one or more facilitators, including people from a range of backgrounds. The following sections will give the reader an understanding of how each of the core tools is intended to be applied within a Britest study. For this purpose, two options were considered. Ideally a real process example would have been employed, to thoroughly demonstrate to the reader the Britest tools being applied within the context for which they were developed. However, the restrictions associated with process specific examples meant that this was not possible. Real process examples of the tools in use have been shared only within the Britest consortium, and so inclusion of these in this thesis

would not be possible due to the associated confidentiality agreements. Examples of using each tool on a single process were not available within the Britest consortium, and the use of a single process is the most effective method of showing the different benefits brought about by each individual tool.

In lieu of a process example for illustration purposes, the process of making a cup of coffee was utilised. The process involves phase changes and reactions much like a processing example, but does not have associated intellectual property. Additionally the process of making a cup of coffee is able to be understood by a reader of any background, unlike many chemical/physical processing examples.

Some tools have an inherent variability in application method, and so users will have their own preference as to what works best for their team or process. The description in this Chapter is not an exhaustive manual for tool application, but is intended to give the reader a basic understanding of the Britest toolkit in the state it existed prior to research commencement. The tools and their intended purpose/outcome are outlined in Table 2.1, along with their associated detail level, advantages and disadvantages when considered with respect to bioprocessing requirements. This is followed by more detailed application information in specific sections.

Table 2.1 - The Britest tools, purposes and relative strengths and drawbacks.

Tool	Purpose	Resulting Detail Level	Strengths	Drawbacks
Initial Screening Analysis (ISA)	Construct an overview of the process and inform subsequent tool use	Overview-Low	Consistent starting point for a Britest study to focus the people and give direction	Can generate large amounts of information
Process Information Summary Map (PrISM)	A high level overview of the key stages in a process, summarises process inputs and outputs, records key information [associated with each process stage, input and output]	Overview	Easy to understand, reduces process complexity, quick to apply, captures high level technoeconomic drivers	Can oversimplify, no intermediates captured
Process Definition Diagram (PDD)	Task-based whole process representation, showing where process materials are introduced and/or removed from the process, the phases present throughout each task, phase changes (e.g. dissolution, gas evolution, etc.), key energy balances	Medium	Independent of scale/equipment, cross-disciplinary, information rich	Time consuming to construct for long processes, less beneficial in single phase processes

Tool	Purpose	Resulting Detail Level	Strengths	Drawbacks
Rich Diagrams (Pictures/ Cartoons)	Rich Pictures/Cartoons are a way of visualising what is happening at a specific point within the process.	Dependant on the purpose, can range from Low to High.	Flexible, detail level defined by the user to give the desired benefit.	Lack of structure could lead to multiple versions being generated before relevant info is captured.
Transformation Map	A graphical portrayal of the network of transformations that convert raw materials into products within a process task. They should include both desired and undesired transformations, to support the use of other tools (e.g. Driving Force Analysis) to identify operating strategies favouring the desired transformations.	High	Forces user to consider all potential reactions, applicable across scales. Particularly useful for understanding multi-phase transformations	Time consuming if lots of detail required, multiple unknowns limits benefits. Can be confusing for large molecules.
Driving Force Analysis (DFA)	A qualitative model of the competing driving forces within a process to enable the identification of	High	Systematic application, helps understand impact of process changes, structured output	Requires completed Transformation Map, limited scope for inclusion of complex relationships

Tool	Purpose	Resulting Detail Level	Strengths	Drawbacks
	potential operating strategies.			
Transformation, Entities, Properties, Physics, Parameters, Order of Magnitude (TE3PO)	A tool used to record and analyse knowledge about transformations where the presence of parallel rate processes means that rates need to be balanced in order to deliver the optimum outcome	Medium	Information rich, breaks down process, macro/micro scale. Very useful for analysis of physical processes.	Difficult to interlink transformations, could be time consuming

2.3 Initial Screening Analysis

The Initial Screening Analysis (ISA) methodology is the starting point for Britest studies, allowing an overview of the process to be assembled. It can identify constraints on the process, either real or perceived, and is useful for noting key inputs and wastes.

The methodology consists of six steps (Figure 2.1). Through the application of this methodology, it should become clear where process improvement may be possible based on broader techno-economic drivers, for example through the increase of yield, reduction of waste, reduction in batch time or increase of throughput. The ISA is used to help the facilitator identify the additional tools which will be most beneficial, and the appropriate order for application.

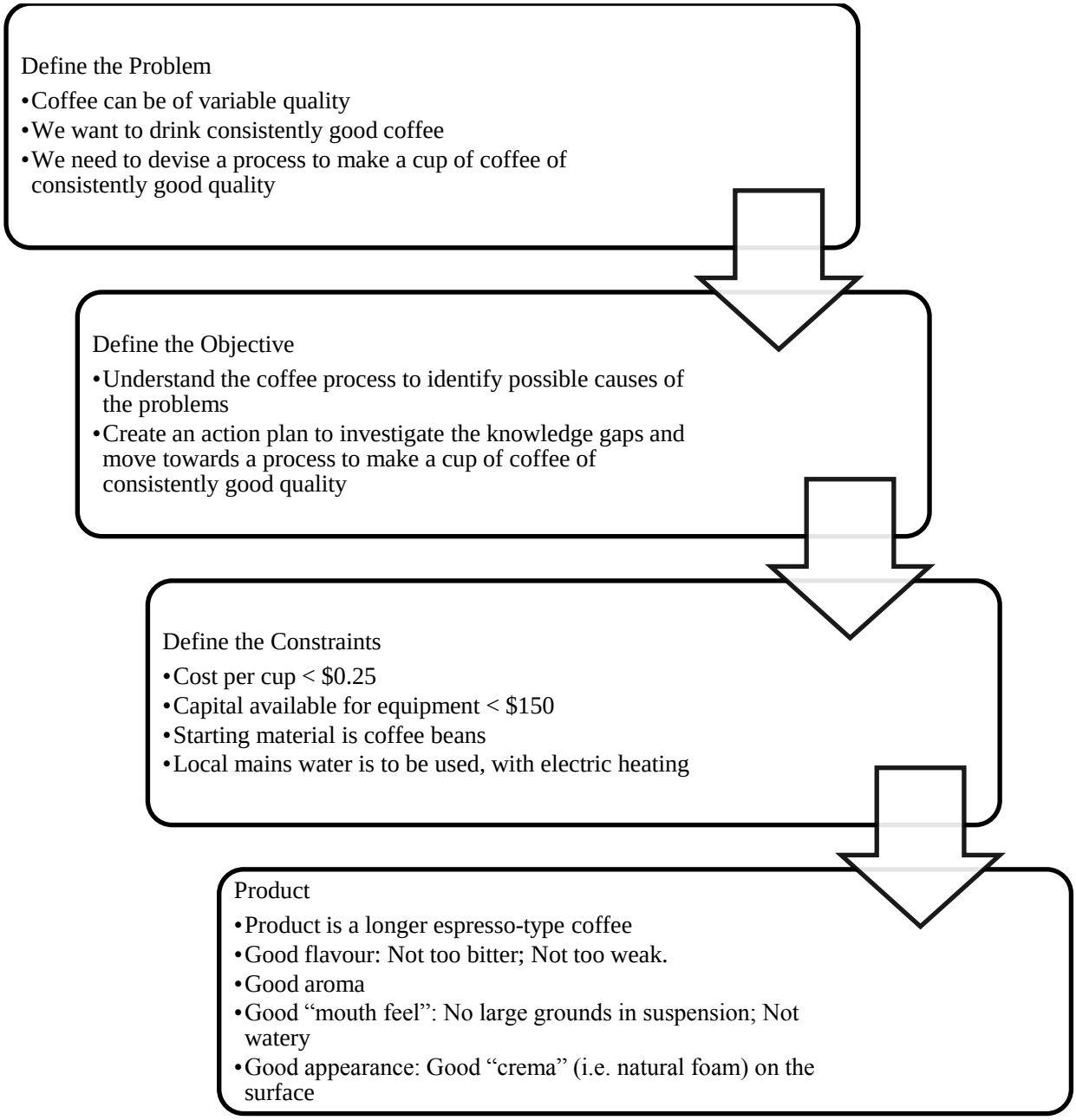


Figure 2.1 - ISA for coffee extraction process. Taken from Britest training materials.

2.4 Process Information Summary Map

During the ISA discussion, it is common to require an overview representation of the whole process. This could demonstrate how the actions performed in the process are relating to the product itself, giving an action to positively influence the final product quality, help to highlight the tasks which are likely to provide the largest benefits to the whole process, or where the cost/value of the process lies.

There are two tools aimed at representing whole processes, the Process Information Summary Map (PrISM) and the Process Definition Diagram (PDD). The PrISM is part of the ISA methodology, and so can be applied during early discussions to give a high level overview of the whole process. The PDD is constructed when more detailed process analysis is required.

In the PrISM (Figure 2.2), all inputs and outputs of each stage of the process are identified, including any waste. It is also useful to capture factors such as costs, step timescale, processing conditions and the yield of each step, which is analogous to value stream mapping within the Lean Toolkit. In this way the relative potential for process improvement of each step, for example based on cost, processing time or yield can be easily identified. Options for reducing cost, waste or increasing yield can be proposed. The output from this analysis is often surprising, as the stages which process technologists wish to consider may not be the ones with the greatest improvement potential. The PrISM can show more detail by including a table of inputs and outputs, for example to highlight potential quality assurance (QA) issues.

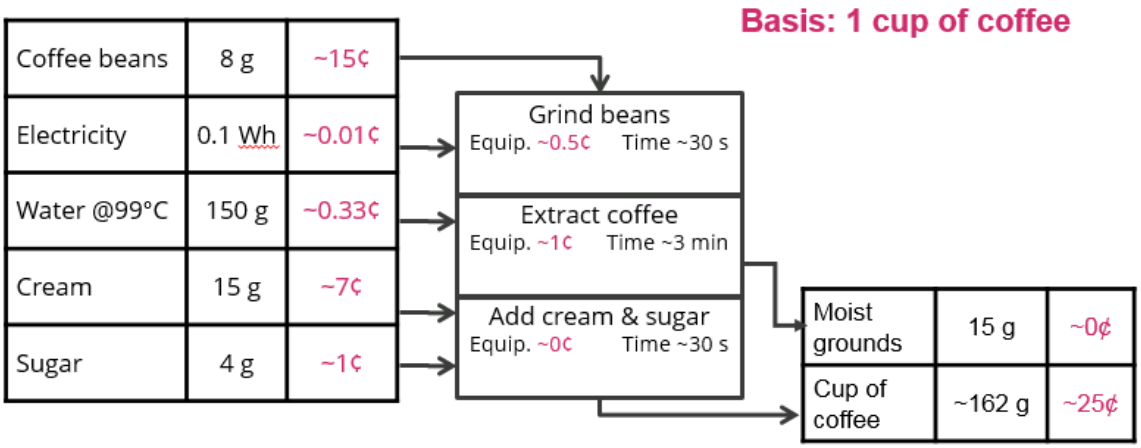


Figure 2.2- PrISM including inputs and outputs table for coffee extraction. Taken from the Britest training material.

2.5 Process Definition Diagram

In a complex process, or when a high level of detail is required, the PrISM can only provide a limited degree of insight into a process. A higher level of detail can be achieved by employing a PDD. The PDD was first constructed by Wall *et al.* (2001) , and though the style has changed over time, the principle remains widely applicable within processing (Teoh *et al.*, 2015). The PDD allows the process to be split into stages, without the restriction of unit operations. The detail generated is higher than that of a chemical equation. The pictorial representation makes it understandable to a team with a range of technical specialisations, facilitating effective communication.

A PDD is a form of State Task Network (Wall *et al.*, 2001). It consists of a series of boxes representing each of the tasks involved in a process, filled with a representation of the phases within a process. It encourages visualisation of the process as a set of tasks, not corresponding specifically to unit operations or particular types of equipment. It is a way of showing the physical changes occurring to the materials as they pass through the process, potentially influencing the CQA's of the product. Each box can show the change in phase ratios over time, capturing the accumulation or depletion of a phase. Energy streams are often included, to show where heating, cooling or mixing would be applied.

Each box is given a title to represent what is happening within that stage, for example Separation, Wetting, Mixing, and is numbered according to the order in which it occurs within the process, as shown in Figure 2.3. It is notable that stages will often not be named in terms of equipment, but rather in terms of the purpose of the step. For example solid/liquid separation could be used to represent a filter or a centrifuge. This allows alternative options to be considered, though it can be useful to include the current methodology as an annotation. The boxes are typically numbered in ascending multiple of

ten, to allow other stages to be added with ease if it becomes apparent that a stage would be better represented by splitting into several boxes.

For explanation purposes, Figure 2.3 shows a PDD of the process of making a cup of coffee. Initially the beans are ground and the water heated (Boxes 10 and 20). The beans then leach into the water and are filtered (Boxes 30 and 40). This generates grounds for disposal and the remaining coffee can have milk and sugar added according to taste (Box 50). While making a cup of coffee may appear simple, the PDD highlights the frequency and number of phase changes, and when applying to a complex chemical process the PDDs can give valuable information.

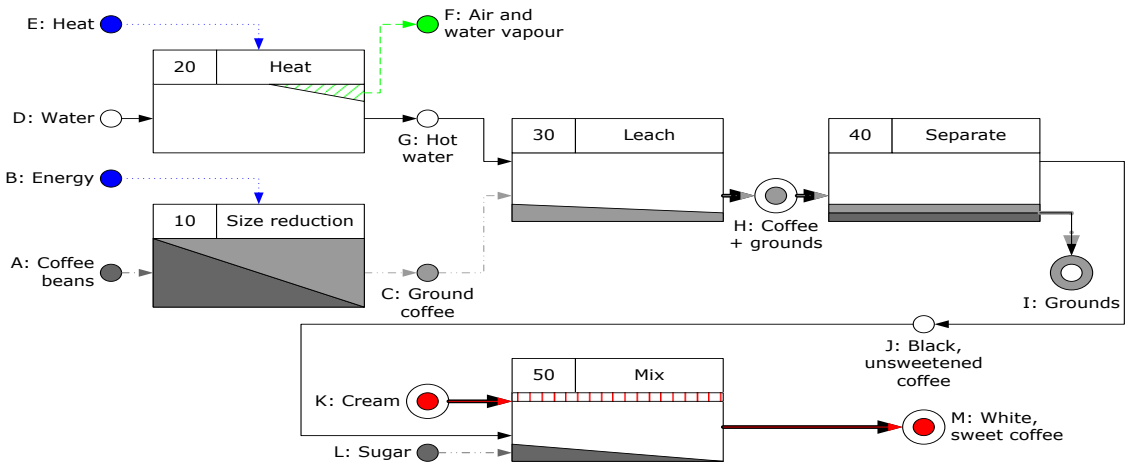


Figure 2.3 - An example of the PDD tool, representing coffee extraction. Taken from the Britest training materials. In this PDD white indicates a liquid, green indicates a gas phase, grey a solid phase, darker grey a denser solid and red an organic liquid. The circles between boxes indicate a multi-phase addition.

Annotations on boxes are useful, usually noted underneath the box in bullet point form. This captures additional important information to ensure the process as a whole is considered. PDDs have been used previously within technology transfer, process troubleshooting and to compare process options.

2.6 Rich Diagrams

2.6.1 Rich Pictures

Rich Pictures are a way of visualising what is happening at a specific point within the process. This could be mixing within a reactor, cleaning of pipes or any other part of the process. A typical Rich Picture will be a result of one stage of the PDD being identified as of particular interest, or the box not fully representing the reactions occurring. As an example, within a reactor it could identify issues such as inadequate mixing, the development of "hot spots", adhesion to walls, or settling.

Any scale can be used for rich pictures: either the whole unit can be drawn, or a smaller sub-section can be drawn. It is common to begin by drawing the whole unit but the result to be the need for further rich pictures to be drawn at a different scales, e.g. to focus on the macro (equipment), micro (solid/liquid structure) or molecular scale. Figure 2.4 shows a Rich Picture of a cup of coffee. This shows how solids and oils may be suspended in the aqueous phase, and provides understanding of the settled solids and the foam at the surface of the drink. These may seem trivial in the context of a cup of coffee, but in an industrial process inadequate mixing could be a serious hindrance. These are particularly useful in investigating localised effects and for troubleshooting.

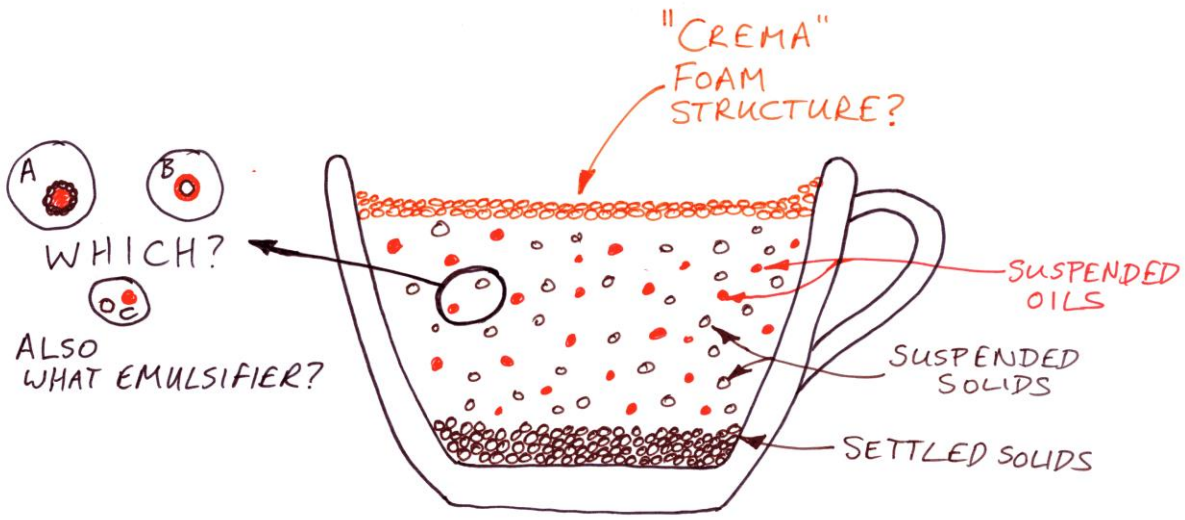


Figure 2.4 - An example Rich Picture showing coffee extraction. Taken from the Britest training materials.

2.6.2 Rich Cartoons

A rich cartoon is similar to a rich picture, but depicts the changes over a period of time rather than at one particular point in the process, much like a cartoon strip. This could aid in the visualisation of the process at a more in-depth level than the PDD.

2.7 Transformation Maps

The transformation map (TM) is intended as a method to explore all of the possible reactions or physical transformations that could take place within a single task, either desired or undesired. Used correctly, it can identify what causes these reactions to occur at a faster or slower rate. This can allow the selection of process conditions to push the reaction down the desired route, and to minimise undesired transformations.

It requires the equations for the reactions, and knowledge of all of the species potentially present. Initially a list of all species is constructed, prior to generation of the equations. The final step is to put these into a sequence for the reaction, and it is useful to colour code arrows to represent whether a reaction is desired or undesired. It is important to indicate whether a transformation is reversible or irreversible; note that mass transfer processes are by definition reversible. In the example process of making a cup of coffee, a

TM for the extraction of flavour from the beans may look like the one shown in Figure 2.5. Note that the different phases (solid, liquid and gas) are specifically highlighted to ensure that any mass transfer processes are captured.

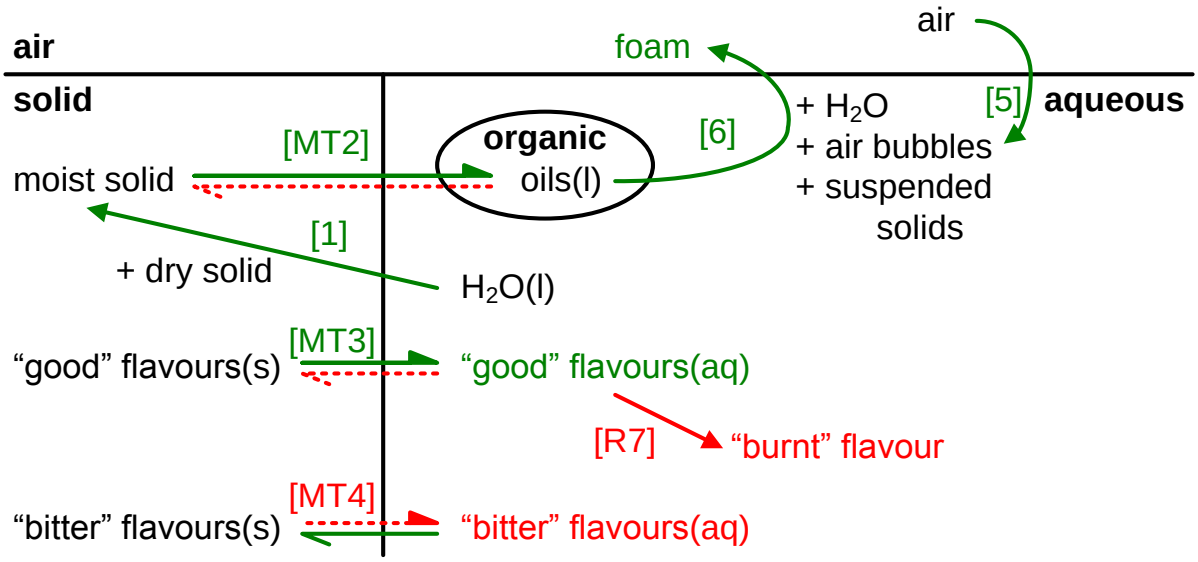


Figure 2.5 - Transformation map for the potential reactions within a coffee extraction. Taken from the Britest training materials.

2.8 Driving Force Analysis

In a typical Britest session, the Transformation Map is often followed by construction of a Driving Force Analysis (DFA) table (Sharratt *et al.*, 2003). The DFA table provides a structured approach to understanding the impact of each process driving force on the outcome of an individual transformation. Each column represents an individual transformation shown in the TM, and each row is a component or condition which may influence the transformation. It has been shown to be beneficial to colour code the columns according to whether a reaction is desired (usually green), or undesired (usually red). The table is completed by capturing the impact of the individual driving forces on each transformation using simple symbols such as plus and minus signs. It can also be useful to describe specific rates of reaction using words such as seconds or minutes.

To highlight the applicability of the symbols, the example of coffee extraction is outlined in Table 2.2. The headers in the DFA correlate to the reactions in the Transformation Map. The abbreviations MT and R are often used to differentiate mass transfer and reactions respectively.

Table 2.2 - An example DFA based on the process of coffee extraction. Taken from the Britest training materials. Columns correspond to the reactions in the Transformation Map (Figure 2.5).

Driving Force	[1]	[MT2]	[MT3]	[MT4]	[5]	[6]	[R7]
solid surface area	+	+	+	+			
"good" flavours(s)			+				+
"good" flavours(aq)			–/P				
"bitter" flavours(s)				+			
"bitter" flavours(aq)				–/P			
moist solid	P	0	0	0			
solid oil content		+					
suspended oil drops		–				+?	
suspended µ-solids						+?	
suspended air bubbles					P	++	
temperature	(+)?	(+)?	(+)?	(+)?	?	?	++
water hardness	?	?	?	U			?
shear & turbulence	+	+?	+	+	++	+	
rate	fast	variable secs – 10s secs	variable secs – 10s secs	variable < [MT3]	?	?	10s mins

Through filling in this table for each reaction, considering each influencing factor, it is possible to identify possible process operating strategies which may favour the desired reactions, and minimise undesired ones. In this example undesired reactions could be leaving the beans to brew for too long leading to bitter flavours, or the addition of too much milk or sugar. It would also demonstrate the addition of milk linking with cooling the temperature, which may or may not be desired.

2.9 Transformations Entities Properties Physics Parameters and Order of Magnitude (TE3PO) Table

The TE3PO table is a tool used to record and analyse knowledge about transformations. It is similar to a Driving Force Analysis table but was developed to capture information

about parallel rate processes where the rates need to be balanced in order to deliver the desired transformations such as physical processing operations and polymerisation chemistry.

The TE3PO draws upon information captured in Rich Pictures and Cartoons, and/or Transformation Maps. It structures and summarises process knowledge to aid in the troubleshooting of the process, identification of key parameters for process modelling and identification of knowledge gaps for planning experimental approach. An example TE3PO table is shown here in Figure 2.6.

Transformation	Entities	Properties	Physics	Parameters	Order of Magnitude
Liquid flow through bed	Liquid	Density Viscosity	Flow through packed bed (Ergun equation)	**Available pressure drop**	1 mm/s
	Bed of coffee	**Particle size** **Voidage** **Bed depth**			
Solids entrainment	Aqueous	Density Viscosity	Drag force on particle	Nothing controllable	?
	Fine solids	Particle size Density			
Extraction of oils	Coffee solids	***Oil content*** Affinity for oil	Desorption	**Temperature**	?
	Oils	-			
	Aqueous	Surfactant content?			

Figure 2.6 - TE3PO for the coffee extraction process. Taken from Britest training material.

Within the TE3PO table, each row corresponds to a single transformation, and the entities, properties, parameters and physics associated with the transformation are listed. There could be multiple entries in the subsequent columns, but transformations should always be considered independently. Through the completion of the TE3PO table the user can identify unknown influences requiring experimental clarification, highly influential parameters or help the user to identify the most important transformations to consider at an early stage of process development.

2.10 Summary

Chapter 2 of this thesis discussed the Britest toolkit in detail, to allow the reader to gain a working knowledge and appreciation of how the tools work. This gives the reader the appropriate understanding of the Britest tools to comprehend the research presented in this thesis. The tools were demonstrated on a simple process of making a cup of coffee, to allow a reader of any background to appreciate the methods involved. This highlighted the salient features of each tool, how they could be applied to a process, and the benefits each could bring. Chapter 3 will move on to consider the application of the Britest tools to bioprocessing specifically, using virtual processes in SuperPro Designer, before the remainder of the research presents developments and investigations within the toolkit which are required to adapt the Britest tools for effective use within bioprocessing.

Chapter 3 Virtual Bioprocessing

3.1 Introduction

The previous chapters have discussed the background to the work (Chapter 1) and introduced the reader to the Britest toolkit in its original form (Chapter 2). This thesis aims to develop the Britest tools for bioprocessing, and this chapter discusses the application of the toolkit to a range of simulated processes using SuperPro Designer (Petrides *et al.*, 2002b), to act as "best case scenario" examples of processes where variability is not influencing the outcome and data is available for each component for the duration of the process.

The design and development of sustainable and innovative processes is a challenge across a broad range of manufacturing sectors, especially in the high value sectors. Key difficulties include: pressure on development lead times to reduce time to market; complex systems where chemical, physical and/or biological properties are not fully understood; poor communication of critical process information between different technical disciplines; lack of detailed understanding of whole process challenges within a process made up of a number of separate unit operations; identification of viable process flowsheet concepts, and rapid identification of the most viable options.

In recent years, there has been great progress in the development of tools to support the design and development of chemical and biological processes (Zhou and Titchener-Hooker, 1999; Kalil *et al.*, 2000; Petrides *et al.*, 2002b; Posch *et al.*, 2013; Petrides *et al.*, 2014). Many of these are based on computational simulation of the different unit operations, and the integration of these operations into whole process flowsheets. In general, however, such approaches require large amounts of quantitative data about the different process steps. While some individual steps can be modelled based solely on theoretical data, the development of a

whole process model during the early stages of process design can be extremely challenging as a result of limited quantitative data availability. Computational simulation approaches are also often highly complex, requiring an expert user and significant periods of time to deliver a robust model. Furthermore, multidisciplinary communication of input and output from these models is often difficult for non-expert users.

The challenges posed by the complexity of the products/processes and highly regulated character of the industry exacerbate these issues within the bioprocessing/biopharmaceutical industry sector. Whilst the introduction of Quality by Design (QbD) and Process Analytical Technologies (PAT) (FDA, 2004; I.C.H Guideline, 2009) has contributed to the generation of much richer datasets through the bioprocess design and development process, it also raises additional challenges. The identification of Critical Quality Attributes (CQAs), Critical Process Parameters (CPPs) and the definition of the design and control space are frequently not straightforward, although fundamental to the process understanding and the ability to effectively control the process (Harms *et al.*, 2008; Rathore, 2009; Abu-Absi *et al.*, 2010; Glassey *et al.*, 2011; Kumar *et al.*, 2014).

Different approaches to defining the design and control space have varying degrees of robustness, but are generally based on a combination of process understanding and experimentation (Rathore, 2009). There is not currently a standard approach which is recommended, and this means there can be no guarantee of the robustness of the design space generated.

In addition to facilitating the QbD approach in processing, effective knowledge capture has been correlated with organisational effectiveness (Gold and Arvind Malhotra, 2001). In order to be useful, however, it is important that any knowledge capture approach used is able to

organise the information in a manner that enables its effective future use and supports process understanding.

One approach to the challenges of knowledge capture in scientific and engineering based companies, developed by Britest Ltd., has found broad use across the chemical-using sectors such as pharmaceuticals, and fine, speciality and consumer chemicals industries (Wall *et al.*, 2001). As previously discussed in Chapter 2, the Britest approach uses a set of qualitative and semi-quantitative tools and methodologies to enable cross-disciplinary understanding of industrial processes, therefore supporting innovative whole process design. The tools are deliberately designed to be complementary to more quantitative approaches such as computational process modelling, economic modelling or fluid dynamics calculations. This approach is not an expert system, and it is intended to be usable by technologists of all disciplines.

In this work, process simulations were used to provide a range of virtual biological processes on which to test the Britest toolkit. The virtual processes were available within SuperPro Designer (Petrides *et al.*, 1998; Petrides *et al.*, 2002a; Harrison *et al.*, 2015), and provided a "best case scenario" where all process units had significant information available. This level of detail would likely be unavailable on an industrial process in early stages of development, and so the simulations allowed testing of the tools where practical constraints and data availability were not a concern. It was anticipated that applicability to bioprocesses could be established, and required developments identified to enable the next stage of the research to test the developed toolkit on a process which better represents the level of detail available on an industrial process. Four types of bioprocess, spanning four markets, were selected to demonstrate broad applicability across a range of bioprocesses (Figure 3.1). These were monoclonal antibody (mAb) production, insulin production using *E. coli*, wastewater

treatment and penicillin V production. For the purpose of this chapter of the thesis, the focus will be on insulin production through an *E. coli* host expression system. The completed Britest tools for the remaining three bioprocesses are included in Supplementary Material as Appendix A.

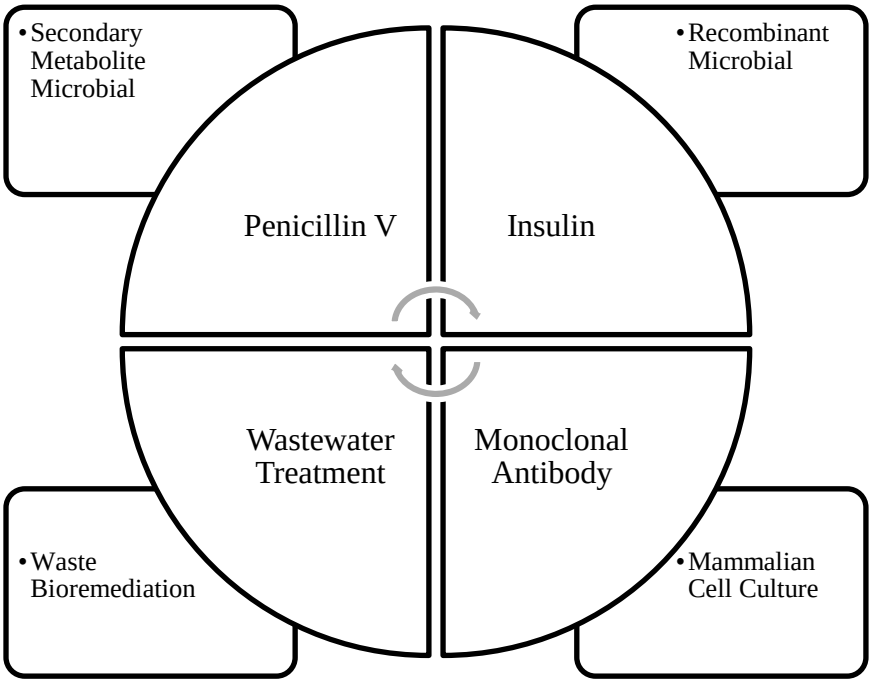


Figure 3.1 - The four types of bioprocess, and their associated markets, being considered for this research.

3.2 Methods

3.2.1 Process Simulation

The model process selected for detailed discussion in this chapter, the production of insulin from *E. coli*, is a complex process, which can be carried out using two methods (Kamionka, 2011). Either the chains could be synthesised separately and mixed, reduced and reoxidised after purification (Goeddel *et al.*, 1979). Alternatively, the bacterial culture produces proinsulin, which then undergoes extensive downstream processing to give biologically active insulin (Zündorf and Dingermann, 2001).

In this case, the proinsulin method was simulated using SuperPro Designer. This simulation of insulin expression in *E. coli* has been presented previously as part of Chapter 12 in Bioseparations Science and Engineering (Harrison *et al.*, 2015). The process scheme is summarised in Figure 3.2. The fermentation, producing Trp-LE-MET-proinsulin precursor, is performed in bioreactors using transformed *E. coli* cells. The fermentation duration is 18h and it is performed at 37^{o}C. The product is formed as inclusion bodies and a total yield of 30g/L is obtained. The primary recovery consists of cell lysis and purification of inclusion bodies, through centrifugation for cell separation, homogenisation to lyse the cells and then further centrifugation to separate the inclusion bodies from cellular debris. A detergent (Triton-X-100) is then added prior to the final centrifugation step, to aid further separation of the inclusion bodies. The reaction section of the downstream process starts with solubilising the inclusion bodies using urea and 2-mercaptoethanol to break the disulphide bonds prior to concentration through diafiltration. The solubilised inclusion bodies are then cleaved with cyanogen bromide to remove the signal sequence, and evaporated before sulfitolysis results in protein unfolding. The next stage is S-seraphose chromatography, followed by refolding and the final step, again using 2-mercaptoethanol. The resulting protein is purified with Hydrophobic Interaction Chromatography (HIC) before being cleaved enzymatically with trypsin to remove the C-terminal peptide. The final purification consists of four chromatography stages, followed by crystallisation of the insulin. Centrifugation is used to recover the crystals for freeze drying.

3.2.2 Qualitative Process Understanding Tools

The Britest tools were applied according to a framework developed for a chemical processing study. The main objectives of applying the tools in this case study were:

- To capture the purpose of each stage of the process and how it works
- To identify the potential for improvement within the process

- To outline experiments required to further understand and optimise the process

While the purpose of the work presented within this case study was to identify gaps within the toolkit in relation to bioprocessing, the study was designed to mirror the typical aims of a study supported by the Britest tools. Were the process not simulated, the study would be used to capture process understanding in each stage, in addition to exploring the underlying science of the process and identifying potential opportunities for process improvement. They could also be used for whole process analysis/design, to determine the impact of changes in one stage on others. The Britest tools are also particularly useful for facilitating interdisciplinary knowledge transfer, by providing a visual approach to knowledge capture, which is nonetheless based on the fundamental science under investigation. Such an approach is particularly pertinent to the bioprocessing sector, where many different disciplines can be involved in a single process, and effective communication of information between different disciplines can be extremely challenging.

The key tools are outlined previously as part of Chapter 2. Each tool was considered in turn, and relevant advantages and disadvantages used to determine which tools would be most appropriate for application to this particular bioprocess to achieve the intended knowledge outcomes. This study focussed on the Process Information Summary Map (PrISM), the Process Definition Diagram (PDD) and the TE3PO. The Transformation Map and Driving Force Analysis (DFA) are targeted at developing understanding of the chemical reactions occurring within a single process task, which was deemed too complex to consider for the fermentation step. The tools could be used within downstream processing steps, but this wasn't carried out within this study as the downstream processing units used in chemical processes do not differ significantly between chemical and biological processes. In the course

of this work, a new tool was developed (the Reaction/Reagent Transformation Tracker (R2T2)) and it was employed to further enhance process understanding.

The PrISM captures key data on all stages within a process, along with the inputs and outputs for each stage. This tool helps the team to focus their activities on the most appropriate parts of the process by providing an overview of the most critical material, time and energy dependencies.

The Process Definition Diagram (Wall *et al.*, 2001) is a tool that enables process technologists to describe a process independently of scale and equipment. It is a form of State Task Network, describing the process as a sequence of tasks that are performed to transform starting materials into products. The PDD provides an information rich summary of part or all of a process, which has been used for purposes such as cross-disciplinary knowledge sharing, whole process design, process technology transfer, and troubleshooting. The PDD uses a pre-defined set of symbols to denote the number and type of phases present in each process task as the presence of multiple phases can add significant complexity and risk to the scale-up of chemical and biochemical processes.

The TE3PO table is used to better understand the conversions and reactions when a driving force analysis is not possible. The tool was developed for physical processes, where clearly defined intermediates and reactions are not available or not fully understood. It is particularly useful when seeking to understand and balance reaction rates.

3.2.3 Simulated Bioprocesses

As mentioned earlier, this thesis chapter will focus on the production of insulin from *E. coli.* The insulin process starts with the fermentation, and then moves into primary recovery using a combination of centrifugation, blending and homogenisation to fully lyse the cells. This is followed by the solubilisation of the inclusion bodies, and a range of reactions to obtain the correct folding of the protein. The process ends with several filtration and purification steps to ensure the correct purity is obtained, excluding incorrectly folded proteins.

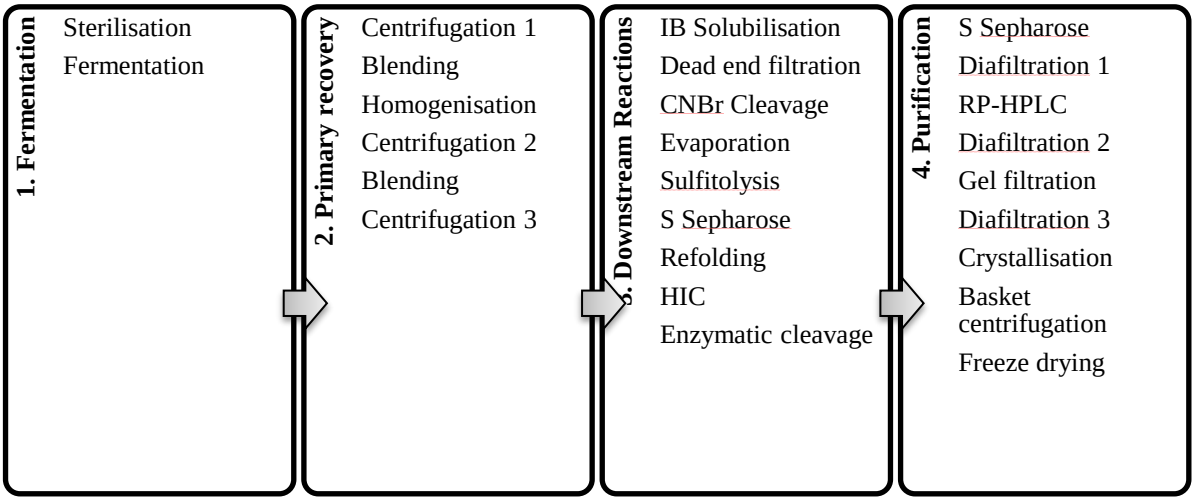


Figure 3.2 - Process outline for insulin production within SuperPro Designer. This was the process on which the Britest study was conducted.

The Penicillin V process also starts with fermentation, and is followed by primary recovery where the biomass is removed. The resulting broth is cooled and acidified to ensure the correct form of penicillin is produced. This is then purified using solvent washes, and centrifuged to purify the solids. These are subjected to fluid bed drying to remove any remaining solvent before leaving the process, and to ensure that the final product does not contain more than 0.05% water.

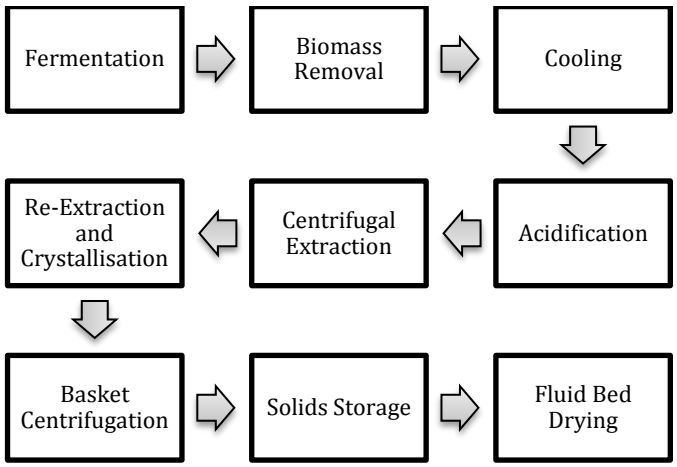


Figure 3.3 - Process outline for the Penicillin V production process.

The waste water treatment process starts with the mixing of influent, which is treated with two aerobic bio-oxidation steps prior to polymer addition at the second mixing stage. The polymer addition is designed to encourage the growth of flocs, increasing treatment effectiveness. The resulting water is clarified, and the process can then follow two branches. The first of these (Figure 3.4) is Granular Media filtration, after which the water is discharged into the main sewer system. The second (Figure 3.5) is belt filtration followed by sludge drying, to remove old biomass from the system. Recycle loops also operate to recycle the sludge and maintain a consistent population within the bio-oxidation stages.

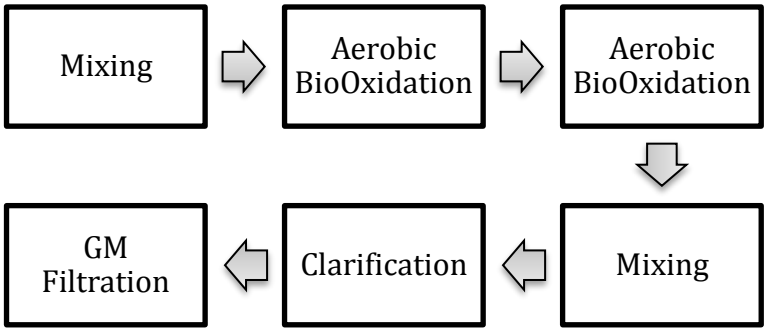


Figure 3.4 - Process outline for one of the branches of the Industrial Wastewater Treatment process.

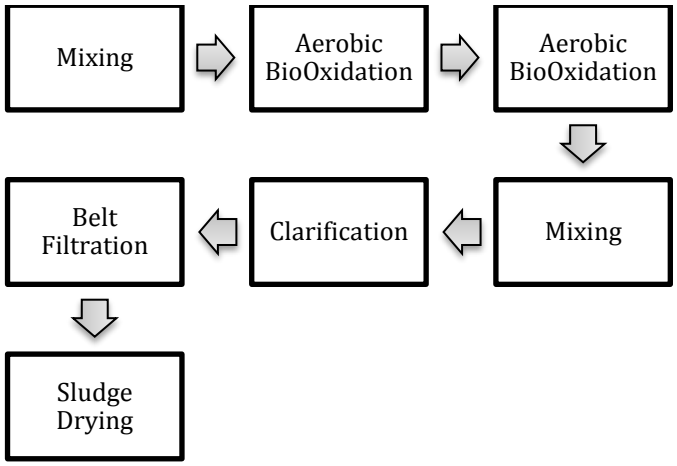


Figure 3.5 - Process outline for the second branch of the Industrial Wastewater Treatment process.

The monoclonal antibody production process follows the generalised scheme outlined in Figure 3.6. It begins with inoculum preparation and then the production bioreactor is run. This is followed by primary recovery in the form of centrifugation, and then Protein A chromatography. The next steps are virus inactivation, and two chromatography steps (Ion Exchange Chromatography (IEX) and Hydrophobic Interaction Chromatography (HIC) respectively). Finally viral exclusion precedes the final filtration step. The final product is frozen and leaves the plant in plastic packaging, remaining frozen in transport.

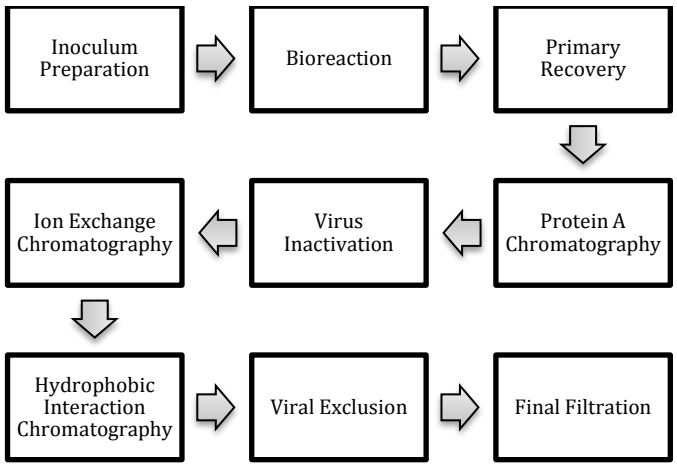


Figure 3.6 - Process outline for the monoclonal antibody production process.

Additional bioprocesses included in this work are detailed for information and their associated Britest studies are included in Appendix A.

3.3 Results

The PrISM for the insulin model process considered in this research is shown in part in Figure 3.7. In this representation, the process has been split into four high-level stages: fermentation, primary recovery, reactions and final purification. To complete the PrISM tool first the central column representing the various stages of the process were considered. Each central box was sized according to the length of that section of the process. For example, the reactions box was bigger than both the fermentation and primary recovery stages, as it takes 106h vs 34h and 30h respectively. This would give the user an indication of where the most time is being spent during the process, and this could be a factor worth investigating in further detail later in the Britest study, as time savings can often lead to cost savings.

Once the central column was completed for the four overarching stages, each stage had its associated inputs and waste captured in tables on the left and right hand side on the corresponding box. For example, in the fermentation Ammonia, Glucose, Nitrogen, Oxygen, Salts and Water are added. The waste produced from this stage consists of Ammonia, Carbon

Dioxide, Nitrogen and Oxygen. Anything else produced in the fermentation (e.g. biomass, insulin) is taken forwards through the process, and so is not captured in the outputs table. The amount of each reagent used is captured, as is the amount of each waste component produced. This could help the user to identify reagents which are used in excessively large amounts, which could indicate a process inefficiency.

The final stage of tool completion is to consider the costs associated with each reagent. Initially raw costs were used, but the high number of reagents (particularly in the reactions phase of the process) made it difficult for the user to discern the difference between each cost. To alleviate this, cost categories were introduced. The cut off points for each category would vary between processes, the cut off values applied for this study are shown in Table 3.1.

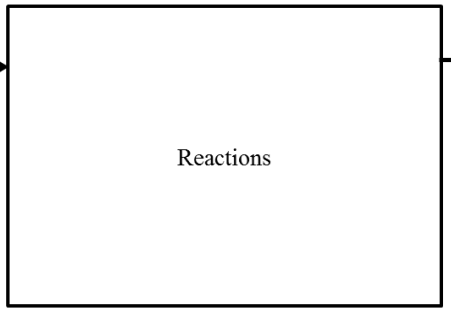
Table 3.1 - Cost category assignation based on US$ cost per unit.

COST CATEGORY	COST PER UNIT (US$)
1	≤1
2	≤10
3	≤20
4	≤100
5	≤1000
6	≤5,000
7	≤20,000
8	≤100,000
9	≤500,000
10	>500,000

The introduction of these cost categories enables the user to quickly discern the most expensive reagents being used, which could be used to focus the Britest study direction if there were alternatives to the expensive reagent available. This would be especially useful if an expensive reagent was being used in large amounts, and would allow this to be quickly identified for further investigation. In some cases this may be unavoidable (e.g. in the case of using a Protein A chromatography stage in monoclonal antibody production (Shukla *et al.*, 2007; Ayyar *et al.*, 2012; Bolton and Mehta, 2016)) , but in many situations a process could be altered to reduce the requirement for the expensive reagent, or indeed a cheaper alternative could be identified.

Within the insulin production process the most expensive reagents were the enzymes, and the main waste was generated at the reaction stage within the downstream processing (stage 3). This was also the longest stage of the process and additionally generated the highest contribution to the product cost (Figure 3.7).

Compound	kg/batch	Cost Category
Acetic Acid	34.7	2
CNBr	95.4	3
Enzymes	0.02	9
Formic acid	10947	2
Guanidine HCl	5035	2
MrEtOH	616.6	2
$Na_2O_6S_4$	151	1
NH4HCO3	34.7	1
Sodium Chloride	3593.3	1
Sodium Hydroxide	821.9	2
Sodium sulfite	302	1
Urea	19087.9	2
WFI	200409.4	1



Compound	kg/batch
Acetic-Acid	27.8
Biomass	0.42
Cell Debris	1.95
CNBr	95.4
Cont Proteins	154.2
Denatured Proin	2.1
EDTA	1.4
Formic acid	10947
Glucose	3.7
Guanidine HCl	5035
IBs	10.6
MrEtOH	300
$Na_2O_6S_4$	149.6
NH_4HCO_3	34.7
Nitrogen	73.4
Oxygen	22.3
Proinsulin	3.4
Proinsulin-SSO_3	9.9
Salts	0.54
Sodium Chloride	3515
Sodium Hydroxide	821.9
Sodium sulfite	300.6
TRIS Base	5.7
Triton-X-100	5.7
Trp-Proinsulin	8.7
Urea	17139
Water	168.7
WFI	96554

Figure 3.7 - Extract from the PrISM for the Insulin production process covering the reactions stage. The central box is sized relative to the duration of each step. The box on the left identifies additions to the process at each stage, the box on the right identifies additions to the process at each stage, the box on the right identifies waste leaving the process.

In a traditional Britest study, the next step would be to complete a PDD for the reactions section of the process, as this is where the PrISM has identified the most potential for cost and time reductions to be made. However, the use of the PDD is already well established within the Britest consortium for downstream processing units from chemical processes. In light of this, the PDD was constructed for the upstream processing (fermentation) stage, to investigate its applicability to biochemical transformations, rather than chemical or physical transformations as has been its primary application to date.

The PDD (Wall *et al.*, 2001) provides a task-based process overview, which also includes a notation that captures the states present during the course of a process (Figure 3.8). Completing the PDD the user begins by constructing the first task box, in this case solution

prep and sterilise. The task boxes are labelled in multiples of ten, to allow the user to retrospectively add boxes if tasks are missed without having to change the entire PDD numbering system. The task box is filled with the relevant patterns/shapes to represent the different phases present within a process task, and the additions to the task are captured using circles and arrows showing where the component is added or removed. For example in this case the first task box is has a small amount of grey denoted by a triangle, where the media powder is added to liquid and dissolved (therefore the powder amount decreases). The rest of the box is left white to denote an aqueous liquid state. Water is shown to be added by the white circle, and heat is applied which is denoted by the blue circle at the top of the task box. Process conditions will always be denoted at the top of the task box in this manner, to avoid confusing them with material inputs. The colour coding system used for this process (Figure 14) is outlined in Table 3.2. Once the first task box is completed, the user would create the next task box (in this case 20, prep for fermentation) and continue to complete based on the components of this task. This would continue until the process or section had been fully captured.

Table 3.2 - Colours used within the PDD and what these represent within the PDD.

Colour	*Represents*
White	Liquid
Green	Gas
Grey	Solid
Blue	Heat

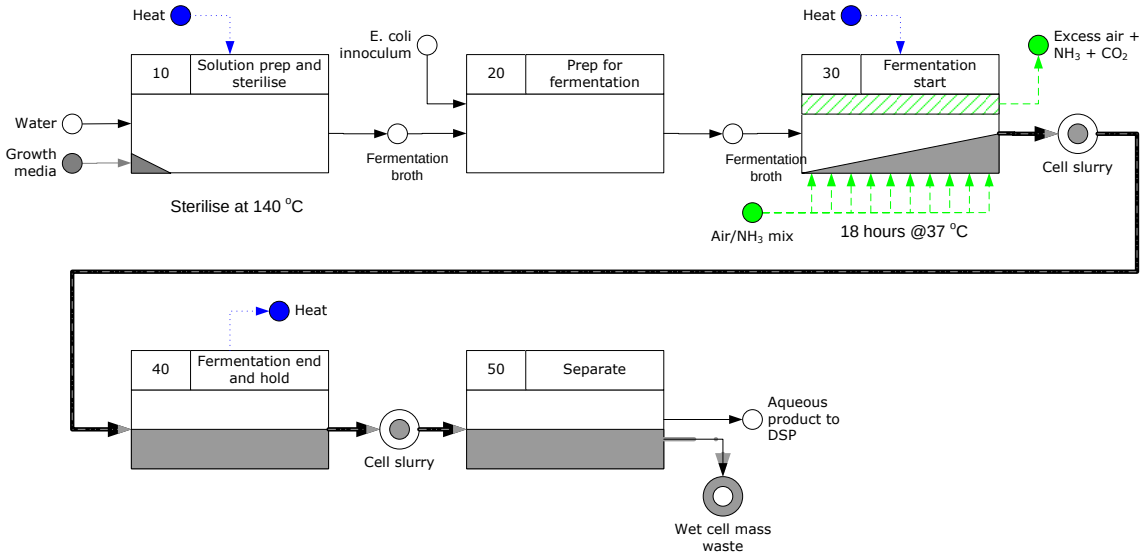


Figure 3.8 - Process Definition Diagram for the upstream stages of the Insulin production process. The different colours present in each box represent a different phase, as outlined in Table 3.2 (white-liquid, green-gas, grey-solid and blue-heat). In this PDD the cells are represented as a solid.

As noted in the previous section, the focus is not on equipment but rather process tasks, allowing changes to be considered independent of the "unit operation" thinking. The second level of detail is the capture of the phases present in each task, which can be critical in determining the complexity of many chemical and physical processes but can under-represent the complexity of many bioprocesses, owing to the presence of multiple components within both solid and aqueous phases. Annotations can be added, which could include operating parameters, observations, common issues etc.

After tool redevelopment, the tool was constructed in a similar manner, but using additional colouring to represent the different components present in each stage rather than simply the stage, to ascertain whether the addition of this information would add more value than the traditional PDD. While the act of tracking the components was useful for the purpose of better understanding what is happening at each stage of the process (vs Figure 3.8 where limited information is shown), the resulting PDD contained such a diverse range of colours and patterns that the user required a key to remember what each colour/pattern combination represented. It is worth noting that the same colour was used more than once with different

patterns, as otherwise there were not sufficient different colours to capture the number of components in the process.

As this tool was constructed using a single user, the necessitation of a key to understand the output highlighted the unsuitability of the tool for detailed analysis of bioprocesses where understanding of individual components is critical, though the ability to track reagents showed the potential to add value. In addition, the time taken to create the PDD with separate components was significant and would not be realistic for inclusion in the Britest study unless only a highly restricted section of the process was selected for very detailed investigation (e.g. a single unit operation). Based on this analysis, there was a clear need for an alternative tool that allowed the components of a process to be tracked, thus giving scope for understanding potential for process variability and improvement.

A new tool called the Reaction/Reagent Transformation Tracker (R2T2) was conceived to fill this gap. This tool aims to show how the amount of each process component changes through the course of the process, to provide a high-level view of the whole process. Colour coding is employed to capture the inherent variability when considering a biological system, allowing for understanding of the challenges involved in development of a process that delivers a consistent output. Incorporation of the variability in this manner helps to tackle the second aim of understanding the potential for improvement in the process. Each of the process stages, and the whole process, can be viewed in relation to the best and the worst-case scenarios, akin to a cost benefit analysis.

To construct the R2T2, the user begins by identifying the process sections (e.g. fermentation, primary recovery, reactions, formulation etc.) to list along the top of the tool, with each section corresponding to a column. There is a column on the left for each reagent to be listed, and the next column allows the user to note the purpose of the reagent (e.g. buffer

component/growth media/promote inclusion body refolding). This would ensure that everyone within a Britest study understood why each reagent was included, promoting effective communication between technologists involved in up- and down-stream, along with business stakeholders. A column on the right-hand side of the table is left to allow the capture of the final concentration of the formulation, either in % or in absolute amount. It would also allow the indication of any limits for purity of the final product, where the amount of an impurity has an upper limit at which it can be present and still acceptable.

From here, the R2T2 is ready to complete. The user can either list each reagent one at a time, and then track across the stages with a line the levels at which it is present at each stage, or could list all reagents first and then draw the tracking line after the list has been compiled. In this case, the list of reagents was generated in whole before tracking was captured, but this would likely be more difficult on a process which was not simulated. SuperPro Designer allowed the list of reagents to be exported into the R2T2 directly, streamlining the application process.

As this was a simulated process, there was no variability to be captured in the R2T2. However, different coloured lines could be used on the R2T2 to represent different scenarios. This could include red lines for a poor process, or green lines for a successful process. This would allow the user to identify where the most critical discrepancies occur. This could influence process monitoring options via the application of Process Analytical Technology (PAT), or could identify where experimentation would be required to reduce variability by changing the process in some way. The final stage of the R2T2 is to colour any cells in a solid colour where a reagent is not present within a process stage. In this case the colour orange was used to represent when a reagent was not present.

In this case study, the R2T2 generated the process overview shown in Figure 3.9. From this, it is evident that the biomass is eliminated completely during the primary recovery stages of the process. It is also clear that the insulin is only produced within the final stage of the process, and the requirement for the production of precursors is more apparent. The extent of reagents required to produce the insulin is easier to comprehend, and this highlights the required focus on downstream processing for process improvement. When considering the process using conventional methods, it may be tempting to focus on improving the yield from the fermentation, however the output from R2T2 makes it clear that the process improvement effort would be better expended on improving the downstream conversion reactions and purification scheme. The R2T2 took less time to complete than the PDD, and provided a whole process view that was more appropriate than the PDD for a bioprocess of this type. Additionally, the tool is simple to understand and apply, which are key criteria for delivering a new tool that will find broader application. The R2T2 fills a performance gap that cannot easily be addressed using the PDD tool. These tools are very complementary in nature, and the decision on whether to use PDD, R2T2, or both will depend on the problem being considered, the timelines, and the data available to the team.

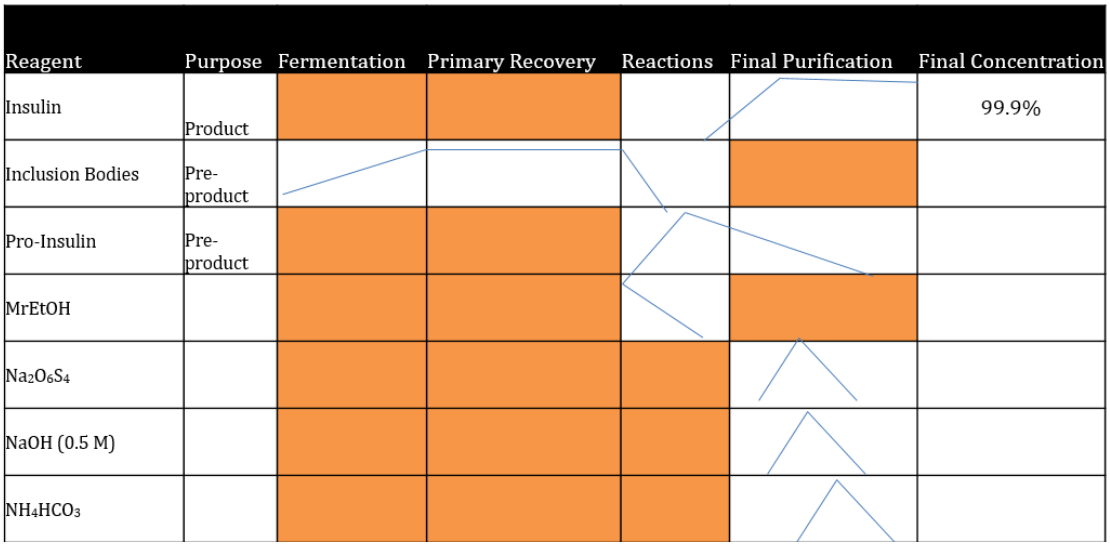


Figure 3.9 - Extract from the R2T2 of the process. Each reagent and its purpose is captured in the column on the left. The process stages make up the remaining columns. The reagent's presence is then tracked through the process with the blue line. Orange boxes indicate the absence of the reagent.

The Transformation, Entities, Properties, Physics, Parameters and Order of Magnitude (TE3PO) tool was employed in an attempt to link the process parameters with the outcomes for specific process tasks. This tool has been used to support understanding of complex physical processes such as milling, where balancing of input parameters related to both the input material and the equipment is necessary to deliver a desired outcome. For this study fermentation was selected for testing tool applicability, for the same reasons as the PDD above. Previous work illustrates the applicability to downstream operations, but application to fermentation is as yet unproven. The cellular growth aspect of fermentation was anticipated to be the aspect of the process which the tool had not already been tested on. There are many metabolic pathways within fermentation, and these are too numerous to be captured in a tool such as the TE3PO. Therefore, rather than considering each metabolic reaction as a separate reaction, a higher level overview approach was adopted. The aim of applying the TE3PO was to be able to link the process parameters and their associated impact on the fermentation outcome. The higher-level approach included reactions such as cell

growth, rather than individual pathways, to capture the relevant effect of each process parameter instead of all possible changes at a cellular level.

The tool was used to assess the cell growth within the fermenter shown in Table 3.3. Although it does not provide a direct means of optimisation for the cell growth, the tool can help in defining which parameters or properties could have an impact on a particular transformation. This is valuable information in helping to define which of these are fixed, and which can potentially be varied and to what extent.

One challenge identified is the fact that the metabolic pathways involved in bioprocesses are generally interlinked, whereas this tool considers each of the transformations separately, at least in its current form. While there was some benefit in using the tool to understand how particular parameters could influence the output of multiple transformations, the practicality of applying it to deliver deeper understanding of a fermentation process was more problematic. In metabolic pathways many of the reactions are interlinked, and not all are identified or understood. An alternative approach could be to capture all of the known reactions using a table of this type, and try to use the information collated to identify trends in the impact of input parameters and material properties on the overall output. However, to construct this tool in this level of detail would take a great deal of time, and once constructed the resulting table would contain such a high volume of information and conditions that it would be impossible to draw conclusions from the information. In addition, when the TE3PO was used to consider high level transformations (e.g. cell growth), the volume of information was too high to be suitable to draw conclusions from upstream processing because there were too many factors involved to draw meaningful conclusions. Based on this analysis, there is a requirement for a tool capable of linking the process parameters of a fermentation to the outputs. However, the TE3PO tool cannot deliver this requirement in its current form.

Table 3.3 - TE3PO for fermentation.

Transformation	Entities	Properties	Physics	Parameters	Order of Magnitude
Bioreaction-growth and production	Ammonia Glucose Nitrogen Oxygen Salts Water Inoculum	Liquid phase, grow to high cell density, productivity, ease of lysis, morphology, product structure	K_{La}	Reactor geometry, Oxygen transfer, mass transfer, agitation, temperature, feedstock composition, starting inoculum concentration, pressure, osmotic pressure	

3.4 Alternative Bioprocesses

This chapter has focussed specifically on the production of insulin by an *E. coli* expression system. However, the tools were evaluated with respect to multiple bioprocesses (Appendix A). For the purpose of this thesis, the outcomes regarding applicability to the broader bioprocessing industry have been summarised for the reader in Table 3.4. Each tool was evaluated with respect to the various processes and anticipated shortcomings documented, including the tools developed within the course of this work (R2T2 and TACO).

Table 3.4 - The challenges associated with applying Britest tools to the different types of simulated process being considered within Chapter 3.

Britest Tool	Aims	Changes so far	Specific Process	Specific Process Challenges	Resulting proposed changes	Completion Time
PrISM	Whole process overview, identify highest waste contributors, where product is lost, most expensive reagents, most time consuming steps	Addition of a cost category, colour coding of reagent amounts	Waste water treatment	Simple process-process sections could be too high level. Process splits - no definitive backbone	Use unit operations for simple processes. Use branches to allow the split.	30mins - 1 hour
			PenV	Simple process-process sections could be too high level	Use unit operations for simple processes	30mins - 1 hour
			MAb	Platform processes generally used so limited benefit	Templates could be generated and edited	30mins - 1 hour
PDD	Whole process knowledge capture, at a higher level of detail than the PrISM. Facilitate communication in interdisciplinary teams. Show the states present within a process, identify where multiple states are present, show the experience the materials have through the process	Trialled breaking down into reagents, the use of high level sections, both deemed unsuccessful	Waste water treatment	Capture of different species information	None: tool deemed not appropriate for this information	½-1 day
			PenV	Highly complex liquids	None: other tools better suited	½-1 day
			MAb	Highly complex liquids	None: other tools better suited	½-1 day
Rich Pictures	Detailed capture of a specific part/sequence of the process	None-applicable without changes	Waste water treatment	Highly variable process-will depend on feed, several may be needed	Use on a specific problem, not on all potential situations	30mins-1 hour
			PenV	None	None	30mins-1 hour
			MAb	None	None	1 hour

Britest Tool	Aims	Changes so far	Specific Process	Specific Process Challenges	Resulting proposed changes	Completion Time
TE3PO	Link reactions with controlling parameters	Targeted to specific reactions, not suitable for whole process use	Waste water treatment	Species interactions are highly complex	This tool will be unable to capture this information- potentially better suited to a Transformation Map. Limited applicability- TM probably better suited	1 hour
			PenV	Secondary metabolite production, therefore production will be more complex	Take care when targeting, applicable but must be used with caution	1-2 hours
			MAb	Eukaryotic expression systems more complex still, combined with a complex molecule. Production influences often not well understood	Metabolic pathways too complex and not well enough understood. Some potential for application if large amounts of data are present.	1-2 hours

Britest Tool	Aims	Changes so far	Specific Process	Specific Process Challenges	Resulting proposed changes	Completion Time
R2T2	Track reagents through the process, capture process variability, identify unknowns	NA	Waste water treatment	Variability is associated to different types of feed-not always the same process. Species dynamics not captured.	Construct more than one R2T2 for various commonly treated waste streams. Alternatively focus on high, medium and low toxicity waste, using colour coding to distinguish.	1 hour per stream. 60-90 mins for one with all info on one.
			PenV	None	None	1 hour
			MAb	None	None	1 hour

Table 3.4 shows that while all the processes are biological in basis, the applicability of the tools remains variable and dependent on the process itself. This is not dissimilar to the application of tools to chemical and physical processes, where the application of different tools to particular processes can vary greatly, depending on the problem being addressed. The variability reinforces the importance of developing the frameworks for tool application. Appropriate frameworks for application would help users to streamline the application of the tools, and facilitate development of an appropriate level of process understanding.

3.5 Discussion

This qualitative study of the insulin production process found results at each stage of the study. Initially, the PrISM was employed. Within the completion of this tool, the highest waste stream was identified, along with the most time-consuming stage of the process. The most expensive reagents were the enzymes. The tool gives a basic overview of the process in a clear and efficient manner, thus demonstrating its applicability to bioprocessing. The underlying concept of the tool is beneficial to a bioprocess, and the simple format in which it is employed is not so simplistic as to reduce the value of the contained information.

Within a process running using a QbD approach, the ability to demonstrate clearly process understanding is invaluable when applying for regulatory approval for a product (I.C.H Guideline, 2009; Zelenetz *et al.*, 2011; Wang and Chow, 2012). The PrISM tool has been demonstrated as an efficient way to summarise a process into a succinct format without losing crucial information about how the process operates. The PrISM could be used as a means to identify the section of the process with the most potential for improvement; from here efforts to decrease waste or enhance reaction efficiency can be investigated, either experimentally or theoretically through further tool application. The

clear explanation of why a change to a process could be required and where the efforts for change would be focussed could be crucial in justifying the changes. Additionally, if a PrISM was constructed for multiple scenarios it could be used to support the varying action required within the QbD approach to facilitate the same end result. Quality Attributes with respect to cost could be identified, but these could not be related to the CQAs of the product. The value of cost modelling within a process has been demonstrated previously within bioprocessing (Sinclair and Monge, 2002; Farid, 2007; Jiménez-González and Woodley, 2010), however these models are often complex and difficult to interpret. BioSolve is one tool which currently works on the cost modelling basis within bioprocessing, and while the benefit of detailed costs understanding is clear, the capturing of the cost information in a format which all employees could understand (such as a PrISM) could be useful for ensuring a broad understanding across the whole plant.

Following this, the PDD was tested on the simulated process. While this tool can be useful as a means of reviewing a process as a whole, facilitating communication between interdisciplinary teams and knowledge capture, in the case of the fermentation stage it proved difficult to achieve a balance between too much and insufficient detail. When used in its conventional form, where states present within each task are captured, the prevalence of a dominant liquid phase meant limited information could be gained from this aspect of the tool. However, when the liquid was split into components, the content of the liquids meant that the resulting PDD was highly complex and therefore could be difficult to understand. Knowledge transfer tools are most effective when easy completion and understanding enable effective knowledge capture (Gupta *et al.*, 2000; Goh, 2002; Tamer Cavusgil *et al.*, 2003). In the case of the PDD, the changes which were predicted to add benefit to process understanding negated this through the added complexity. It was

concluded that within a biological process, the ability to track individual reagents would provide greater benefit than representing the phases present.

The R2T2 is a novel knowledge management tool which was developed as a direct result of this study. The ability to view a snapshot of how each process component changes over the course of the process is envisaged to be beneficial in both knowledge capture and process improvement. The resulting process snapshot aims to provide a method for the capture of reagent purpose, gain/loss and final concentration. With respect to this process, those aims were met through the R2T2 in a manner which was found to be both user friendly and information rich. The ability to use colour coding to capture potential variability within a process was found to be of particular interest to biologically based processes, where reducing variability can be a key concern.

The ability to pinpoint the source of variability within a process, and consider the options available for reduction would be highly beneficial in a QbD process. In this tool, criticality of process components could be ascertained, but like with the PrISM tool, this could not be related to the CQAs through the R2T2 tool alone. The identification of variability and the attribution of this to a cause is the first step an organisation could take in effective process control. Without knowing why the resulting product from a process varies, it is impossible for the process owner to attempt to control this. In this case the process was simulated, and so no robust assessment of variability could be made. It is hypothesised that one important source would be the fermentation, as has been shown to be the case in previous studies (Neves *et al.*, 2001; Defernez *et al.*, 2007; Montague *et al.*, 2008; Delvigne *et al.*, 2014). If this was found to be the case, the process owner could increase monitoring efforts in the reactor to more tightly control the resulting broth, and therefore reduce the variability for the primary recovery. If the variability could not be controlled within the reactor, then it is possible that the conditions for the biomass

removal could be altered to accommodate the output from the fermentation and obtain the optimum results regardless. This is the underlying principle of QbD, and the R2T2 has been shown in this example to be of benefit in the early phases of implementing this approach.

The weakness of this study was the inability to correlate the CQAs with their controlling CPPs, facilitating the successful application of the QbD approach. A new tool would be required to fill this knowledge management gap in a simpler format. Whilst a new tool was not developed as part of this study, subsequent chapters will discuss how this was achieved.

The techniques employed for this qualitative understanding study originated from the Britest toolkit, which was developed for enhancing process understanding of chemical and physical processing. The study aimed to investigate the applicability to bioprocesses, and to overcome any potential gaps within the toolkit. It was clear from the PDD that the increased complexity within a biologically based process was the most significant barrier to application. The development of the R2T2 from this shows that the implementation can be critical to the capture of knowledge. The PDD could be used to capture the same information but was difficult to interpret. This demonstrates clearly the requirement for structured knowledge capture and management, rather than reliance on regulatory or internal documentation.

This study established the possibility of applying the current Britest tools to bioprocessing to enhance process understanding. While not all of the tools were directly transferable, it is envisaged that through further tool development, to allow for the complexity of a biological process to be captured, a user friendly qualitative toolkit for bioprocess understanding could be constructed. The value of such a toolkit is challenging to quantify.

However, the requirement for enhanced process understanding underlies the QbD initiative, a growing driver in industrial bioprocess development (Chhatre *et al.*, 2011; Neubauer *et al.*, 2013; Rathore, 2014).

The final stage in the study was the more detailed understanding of the fermentation, which was undertaken through the TE3PO table. The TE3PO is the tool within this study which showed the most potential to be able to correlate the CQAs with their controlling CPPs, facilitating the application of the QbD approach. Linking of process parameters and outcomes is generally performed using Design of Experiments (DoE) (Bade *et al.*, 2012; Zhang and Mao, 2016), however in a data lean environment a tool such as the TE3PO to link the anticipated effects could be useful for influencing the experimental approach. However, while the information was captured, again this was not in a user-friendly format and it was clear that a new tool would be required to fill this knowledge management gap in a simpler format.

The techniques employed for this qualitative understanding study originated from the Britest toolkit, which was developed for enhancing process understanding of chemical and physical processing. The study aimed to investigate the applicability to bioprocesses, and to overcome any potential gaps within the toolkit. It was clear from both the PDD and the TE3PO that the increased complexity within a biologically based process was the most significant barrier to application. The development of the R2T2 from this shows that the implementation can be critical to the capture of knowledge. The PDD could be used to capture the same information but was difficult to interpret. This demonstrates clearly the requirement for structured knowledge capture and management, rather than reliance on regulatory or internal documentation. The TE3PO table is another clear example of where the knowledge is successfully captured but the use of the knowledge is potentially limited through the representation. A tool to enable these effects to be captured visually for

clearer understanding would be highly beneficial to achieving the aim of the tool while maximising usability.

This study established the possibility of applying the current Britest tools to bioprocessing to enhance process understanding. While not all of the tools were directly transferable, it is envisaged that through further tool development, to allow for the complexity of a biological process to be captured, a user friendly qualitative toolkit for bioprocess understanding could be constructed. The value of such a toolkit is challenging to quantify. However, the requirement for enhanced process understanding underlies the QbD initiative, a growing driver in industrial bioprocess development.

3.6 Summary

This chapter considered the application of the Britest qualitative knowledge capture tools outlined in Chapter 2 to a simulated bioprocess to ascertain the potential for employing the tools within the bioprocessing sector. It is anticipated that the requirement for methods such as those presented within this research will increase as the QbD approach becomes more widespread within bioprocessing. Some of the Britest tools were found to be directly transferable, particularly the Process Information Summary Map, while the Process Definition Diagram has a clear gap in capturing the complexity of bioprocesses. More specifically, this relates to effective capture of the complexity of homogeneous phases containing multiple components. In light of this challenge, a novel knowledge capture tool (the Reaction/Reagent Transformation Tracker) was developed to provide a means of tracking multiple components through a whole process.

Overall, this highlights the value of using qualitative tools such as those developed by Britest to support whole process understanding and knowledge transfer for complex biological processes. However, it also flags the limitations of the existing tools, and

demonstrates the requirement for new or amended tools to be developed to fill the current gaps, in particular the linking of CQAs to CPPs. With the increasing pressures to improve process understanding (I.C.H Guideline, 2009) to comply with the Quality by Design initiative, tools such as these can play an important role in enhancing cross-disciplinary process understanding in complex biological systems. Qualitative tools of this type can also provide an invaluable means of identifying the depth of knowledge and understanding of a process, and thus support targeting of more detailed experimental and/or modelling studies.

From here, the next stage of the research was to investigate Britest tool application to more realistic datasets where information was incomplete. However, the work on virtual processes highlighted the need for a tool to link the process parameters and outcomes. Chapter 4 discusses the development of a tool for this purpose (the Interaction Analysis Table-IAT) before Chapters 5 and 6 move onto considering this tool with respect to Upstream and Downstream processing datasets from academic studies, before Chapter 7 investigates the tools sensitivity.

Chapter 4 Interaction Analysis Table Development

4.1 Introduction

The previous Chapters of this thesis discussed the background to the research (Chapter 1), and the Britest tools both in their original format (Chapter 2) and adapted for bioprocessing through simulated processes in SuperPro Designer (Chapter 3). SuperPro Designer allowed the toolkit to be tested on ideal processes, where data was available for each stage of the process and values were fixed. At this stage in the research a basic Britest toolkit was constructed for bioprocessing, and the next stage was to test this toolkit on processes which better represents a real-world scenario. When this section of the research was commenced, a further gap was identified and a new tool required to fill the gap was developed. To this end, this Chapter will focus on the development of the Interaction Analysis Table (IAT).

Quality by Design (QbD), as previously discussed in Chapter 1, is becoming increasingly important within the bioprocessing sector (I.C.H Guideline, 2009; Jiang *et al.*, 2010; Chhatre *et al.*, 2011; Rathore, 2014). The implementation of QbD relies on a high level of process understanding, accompanied by a high level of knowledge of the product. Thorough process understanding will enable a flexible process to be adopted, where changes can be made throughout process operation, to rectify problems as they arise, to reduce lost batches, increase confidence in product quality, and reduce the environmental footprint of processes (Junker, 2010; Koch, 2011; Neubauer *et al.*, 2013). This clear added value means more companies are investing time early in the process development stages to ensure this level of control could be achieved on the final process through high levels of process understanding. High throughput automated systems such as the ambr250™ (Sartorius Stedim/TAP Biosystems, (Ngibuini, 2017)) and MicroMatrix™ (Applikon, (Bareither and Pollard, 2011)) are making it easier than ever before to generate this process understanding at the early stages of bioprocess development, to increase confidence before scale up. These techniques are capable of generating large amounts of data, however uncertainty can then arise in the analysis and interpretation of these large amounts of

data due to wide range of data analysis techniques being available, and at times a lack of mechanistic understanding (Charaniya *et al.*, 2008; Ündey *et al.*, 2010; Mercier *et al.*, 2014).

One of the main features of a successful QbD approach is being able to accurately predict the effect a change in process parameters will have on the quality attributes of the product. This can be achieved through first principles models, or data analysis when large datasets are available, but can be problematic in early stage process development, where available data may be limited. Design of Experiments (DoE) is often employed to test the effects of multiple parameters on process outcomes; however, the successful application of DoE requires the identification of the appropriate parameters and their associated ranges (Harms *et al.*, 2008; Streefland *et al.*, 2009; Abu-Absi *et al.*, 2010). Currently the success of the approach relies on the expertise of the user, which is not always applied in a structured and reproducible manner. The IAT was developed to support the application of a DoE approach to optimisation, with the aim of generating a structured approach to DoE implementation. Other uses have included the supporting of scientific rationale behind effects of parameters on process outcomes, and to bridge the gap between data based and first principles models.

However, while DoE is a powerful technique when applied correctly, if applied incorrectly the time and money required to develop a process can increase drastically. It is not uncommon for companies to employ a screening round of DoE prior to a full DoE experimental set up, to ensure the design covers the optimum design space prior to investing large amounts of money and time (Mandenius and Brundin, 2008; Shivhare and McCreath, 2010; Kumar *et al.*, 2014). While this approach is undoubtedly valuable, its successful application relies on the user's process knowledge. This is the point of development at which the Britest tools could provide value to the bioprocessing industry. The Britest tools have previously generated significant benefits within other industries (Infineum, 2011; Johnson-Matthey, 2014), and would be applied to guide an experimental approach to process development. The tools at this stage in the research were able to consider various aspects of a bioprocess, and showed they could be used successfully to capture relevant information specific to bioprocessing.

In spite of this successful application to a range of bioprocesses, the tools were unable to satisfactorily link the process parameters to the process outcomes, a key consideration during bioprocess development. In light of this, a new tool was sought to facilitate this linking, to ensure the Britest tools were able to provide value to a bioprocess development study and therefore be attractive to the bioprocessing industry. The IAT originated from an industrial collaboration between Britest, AbbVie and Pfizer, to meet a requirement for a tool to support whole process understanding of antibiotic fermentation processes. The IAT is not currently part of the core Britest toolkit, due to issues arising in early stage testing. The tool has been reported to be difficult to implement, problematic to interpret and requires an experienced facilitator as it is not as visual or intuitive as the rest of the Britest toolkit. The aim of this work was to redevelop the tool, resulting in an IAT which was easier to apply to a process, and gave a more easily interpreted output.

The original IAT was named the (F)IAT (Fermentation IAT), and was developed as a direct result of a wish to link the fermentation process parameters to the properties of the final fermentation broth. The tool was based on the Driving Force Analysis (DFA), in which symbols such as + and – represent the effect a condition or component has on each outcome. The DFA was developed for chemical processing, and so requires the user to consider each transformation (chemical or physical) possible in the process stage being investigated. These include both desired and undesired reactions. For a chemical process, this list of potential transformations can vary. However, in *E. coli* the list of metabolic reactions possible is in excess of 700 (Karp *et al.*, 1996; Ouzounis and Karp, 2000). This is only the core metabolism, and does not include those pathways involved in recombinant expression of a protein. For a mammalian cell line, the potential metabolic reactions would be even greater, and so the DFA would not be a suitable tool for this purpose. A higher-level approach was sought in the original (F)IAT development. The developed tool was thought to be applicable and useable by those involved in its development, however at the beta-testing stage other Britest users found it difficult to understand and so it has not been included within the broader Britest toolkit to date. Rather than develop a new tool from scratch,

this work focussed on considering whether the (F)IAT could add value if some further development work was carried out to overcome the issues raised with applicability and added value in the original beta-test.

For this work, a dataset originating at the Technical University of Berlin (TUB) was employed. A small set of experiments were performed to produce Alcohol Dehydrogenase (ADH) using *E. coli* as an expression system (Knepper *et al.*, 2014). The rest of the Britest toolkit provided little support in interpreting this dataset, and the benefit of a working IAT was clear. The high complexity of fermentation was a large driver for the development of the IAT; most of the current Britest tools were developed to be applied to relatively simple chemical processes where the complexity is significantly lower. Additionally, the variation of input parameters across the process can have a significant effect on process outcomes that occur further downstream, and the current Britest toolkit is not capable of tracking these effects. To this end, the IAT was redeveloped to allow datasets from biopharmaceutical development of this nature to be analysed using the Britest approach. The implementation of the redeveloped IAT for application to upstream data is discussed in detail in Chapter 5. This Chapter will focus on the development of the tool from the original version of the IAT into a new tool which overcomes many of the issues reported in beta-testing previously, with the aim of developing an IAT which is user friendly, adds value to a Britest study, and is able to link the CPPs of a fermentation to the CQAs. The ability to utilise the tool in other areas in which Britest are active was considered throughout development, in addition to the ability of the tool to be applied to other process units rather than being restricted to fermentation.

4.2 Methods

4.2.1 Assessment of Current Tool

A subsection of the original (F)IAT for illustration purposes is shown in Figure 4.1. This subsection describes the main aspects of the tool.

	Key Outcomes (Measurements)				
	Outcome 1	Outcome 2	Outcome 3	Outcome 4	Outcome 5
Potential Constraints					
Constraint 1	+, S	+, S	+, S	+, S	+, S
Constraint 2					
Constraint 3					
Constraint 4			?	?	
Constraint 5			-, L		-
Media/Operating Conditions					
Parameter 1	+	+		+	
Parameter 2	+	+	+	+	+
Parameter 3	+	+			
Parameter 4			+, S		+, S
Parameter 5					
Rating	**3**	**5**	**5**	**7**	**10**

Figure 4.1 -Section of an IAT produced by AbbVie which has been anonymised for the purpose of this thesis. Down the left hand side the constraints and parameters are listed, and the outcomes are listed along the top. The rating is decided with the business and process benefit of the outcome in mind.

The tool comprises of two tables, one focussing on effects of process parameters on process outcomes, and the other on the constraints of the process. These are displayed one under the other, and while a column by column approach is recommended, the two are completed in parallel. The first step the user would follow would be to construct these tables, with the constraints and parameters as rows, and the outcomes as columns. These could be identified from previous tools, for example in the original (F)IAT work these were derived from a PDD. To illustrate, constraints could be factors such as "fermenter must be kept below 150ºC". Parameters could be media components, temperature, dO_2, pH etc. Outcomes would be anything which could be measured or observed about the outcome, which in a fermentation could be cell mass, viable cell count, lactate, product yield, product stability etc.

After the table is constructed and the constraints, parameters and outcomes noted, the outcomes are then designated weightings. These weightings indicate their importance to the process from both a technical and business perspective, as assessed using the Britest Initial Screening Analysis

(ISA) methodology. For example, an outcome which makes the process more efficient but does not lower costs could be attributed a score of 5 or 6, showing there is some benefit to improving the outcome but it is not perhaps as substantial as another factor which both improves the process and reduces cost (which could be designated a 9). Scores can range from 1-10, with 1 being least influential and 10 being most influential. A score can be replicated; the system is to score outcomes rather than to rank them in order. User judgement is required for designating these weightings, and it is anticipated that this could spark discussion between individuals.

Once the outcomes have all been attributed weighting scores, the table is filled in. Each cell in the table is used to examine the impact of the factor in the row on the outcome in the column. This can be represented by several symbols, which were based on those used for the DFA tool. A single plus sign shows a positive relationship, i.e. that an increase in the parameter will cause an increase in the associated outcome. A double plus sign indicates that this effect is more substantial than in those attributed a single plus. An example of this could be the effect of increasing a key nutrient on biomass level. Following this trend, a minus sign would indicate that an increase in the parameter causes a decrease in the associated outcome, and a double minus would be used to indicate a stronger inhibition.

Not all relationships are linear, and an asymptotic relationship is particularly common in bioprocessing. Where live cells are being handled in upstream processing this asymptotic relationship could be expected from almost every parameter within the process. To reflect this, the plus or minus sign is followed by an S. An L would be used to indicate that this constraint reached a limiting value. The final symbol, a question mark, would be used to indicate an unknown effect.

The intended outcomes of the IAT were two-fold. Firstly, the tool would be used to interrogate the process in a structured manner, to allow the degree of knowledge and understanding to be captured. Secondly the completed tool could act as a visual log of the key interactions within the fermentation process. In its original state the IAT could be used to serve this purpose, however the knowledge captured and visual log were only able to be understood by those involved in the

original Britest study where the IAT was constructed. Individual cells could be annotated with additional observations using Microsoft Excel, but this did not make interpretation easier. Tests where other users were asked to interpret an IAT they had not been involved in constructing showed that they found it challenging to understand exactly what had been noted. As such, the tool could be used by an individual, but the knowledge was less easily transferable across individuals or teams, and it was this limitation that this study aimed to address.

The preliminary consideration of the tool for the linking of process parameters and outcomes showed several areas in which redevelopment would be required. The first of these was the high level of complexity of the tool. The inclusion of two tables, with differences between symbols being difficult to ascertain at a glance, each contributed to making the tool complicated to implement on a process. This complexity was the overarching problem with the tool in its original form, but additional concerns included the ambiguity of the weightings and output, the time to complete and the limited scope of information obtained and captured within the tool. Each concern about the tool was considered in a structured manner, with various options for redevelopment being trialled until an optimal solution was obtained. The flow of redevelopment is shown as Figure 4.2. Each development to the tool had at least one alternative considered.

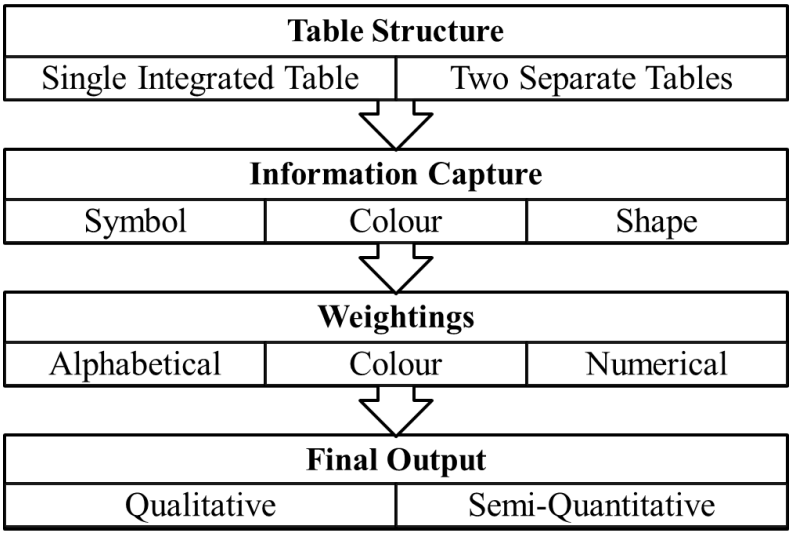


Figure 4.2 - The work flow for the tool development. The over-arching sections show the four areas for tool improvement considered. The sub-sections of each show the options which were considered.

4.2.2 Table Structure

Two options were considered for the table structure. The original approach had two tables constructed in parallel, and examined together when considering potential changes to a process. Both aspects of the table are important within bioprocessing, but simultaneous examination was a main contributing factor to the high complexity of the tool. The options considered for overcoming this limitation were to amalgamate the tables into one, or to consider them independently in sequence rather than simultaneously.

4.2.3 Capture System

Three methods for capturing the information were tested: a numerical system, a colour-based system or a shape-based system. Numbers could be employed to give a scoring system, resulting in a quantitative tool. Positive numbers could indicate the strength of the positive interaction, and negative indicate the strength of the negative interaction. Colours could be used in a system similar to heat mapping where green spots indicated the best results, and red indicated adverse results. Alternatively, shapes could be used to generate a "reaction profile" to show the relationships between the parameters and outcomes.

4.2.4 Weightings

A numerical system was originally applied for the weighting of outcomes. The weighting system was tested with a numerical system, an alphabetical system and a colour based system. The various weighting systems were tested to ascertain which would be the most user-friendly. In addition to testing different methods for implementation, the categories were more clearly defined, and a system to incorporate these weightings into a final output was sought.

4.2.5 Final Output

The original tool did not have a clear output to identify opportunities for process improvement. While the completion of the tool is of value, and a conclusive output is not always sought from a

Britest tool, when linking the process parameters to outcomes the anticipated benefit would be to make a change to the process to positively influence the outcome. The decision to change would be better supported by the clear output of a tool than by subjective interpretation where results could be ambiguous. Outputs of a qualitative and quantitative nature were tested.

4.3 Results and Discussion

4.3.1 Table Structure

Amalgamation of two tables was the first option considered, with the intention to integrate the table of limitations into the table of reactant relationships. This had already been partially achieved through the use of the "L" symbol to indicate limiting values for constraints. However, this increases the amount of information contained within the single table, which could increase the complexity. One of the overriding concerns about applying the IAT successfully is the inability to draw a clear conclusion from the information contained within the cells, and the addition of further symbols is unlikely to reduce complexity. Instead, the two components of the table were split to be utilised separately.

In Chapter 3 of this thesis, the application of structured methodology for tool employment within a study has shown promise. Splitting of the table into two separate parts would enable a more flexible methodology for bioprocess analysis and design. The work in Chapter 3 showed the bioprocess type (e.g. antibiotic production, monoclonal antibody production, etc.) influenced the manner in which tools were applied. It is possible that the IAT may require adaptation for use on different bioprocess types. However, currently it is suggested that at the start of the Britest study, the ISA methodology would be employed to guide which table would be most beneficial to complete first. It is anticipated that rather than bioprocess type, the scenario in which the tool is being employed will be used to determine the order in which the two tables are completed. If a process is already running within a regulated environment, for example, then it may be more suitable to complete the constraints table before the effects table, as the constraints will greatly limit the changes which could be made. Conversely, if a process is in the early design stages and

has a large amount of scope for changes to be made, then examining the interactions before the constraints would be more appropriate.

4.3.2 Capture System

One of the primary barriers to effective tool employment on process was the ambiguous interpretation of results. While this was partially overcome through the introduction of a scoring system (discussed in the next section), the use of symbols such as +. – etc. for information capture was seen as a key factor in the complex nature of the tool. The symbol-based system had two main issues: the difficulty in distinguishing the symbols from each other, and the lack of depth to the information captured except through the use of hidden annotations. Considering these in turn, the difficulty in distinguishing symbols led to the decision to use a visual system that could be interpreted at a glance. This approach has been employed in previous tool development work, in particular on the R2T2 and Process Definition Diagram (PDD), and has been shown to increase information accessibility when conclusions need to be drawn with limited time and resources for detailed analysis. The lack of depth was due to the lack of variation within the symbols, meaning only basic relationships could be captured in the tool. This approach is successful in the DFA for chemical processes but the increased complexity of bioprocesses makes this approach less suitable. Asymptotic relationships could be shown using the original symbol-based system. However, there are a huge range of reaction profiles for biological reactions. Reactions can have an optimum value, especially when enzymes are involved, or may have a sudden plateau rather than an asymptote. Neither of these options could be captured satisfactorily in the original IAT. Multiple signs could be used to show particularly strong or weak interactions, but this was not a strongly visual way of showing how the increased outputs compared to each other. The IAT output was a classic case of needing to meet two opposing criteria. More information was required within the IAT but in increasing the detail level the tool then becomes more difficult to understand. Satisfying both of these criteria simultaneously required major changes to the tool. The methods employed for this were colour and shape, eventually surpassed by a combination of the two approaches.

Initially it was envisaged that colour could be used to create a "heat-map" approach to finding the optimum performance conditions. Each cell was defined in terms of high and low parameter and outcome combination (shown in Figure 4.3), and the colours red, yellow green were used to show the effect of the parameter on the outcome at a particular spot. This was effective at showing the best space in which to operate the reaction, but was not as simple to construct on a computer. An illustration was carried out by hand, which has been the standard approach for all Britest tools; however with the recent introduction of the Facilitator Support Tool (FST), which allows capture of the output from a Britest study in electronic fashion, it would be more beneficial for the Britest members if tools could be constructed easily in electronic form. To combat this, row was split into three in Microsoft Excel (Figure 4.3). The splitting into three allowed a universally familiar traffic light system to be applied to indicate which levels of each parameter would lead to the optimal response. The ideal outcome for the tool in this form would be a green colour across a single level of a parameter. For example, in the case of Parameter 1 in Figure 4.3, a medium level would be the optimal solution, as green (good) results are seen for outcomes 1 and 3, and a yellow (moderate) result seen for outcome 2. This is not the perfect result, but is better than at high levels of parameter 1 (where a red-poor-result is seen for outcome 2), or at low levels of parameter 1 where outcome 2 shows a good result, but outcomes 1 and 3 show poor and moderate results respectively. In this case there would be a compromise made on outcome 2, to preserve the positive results for the other 2 outcomes.

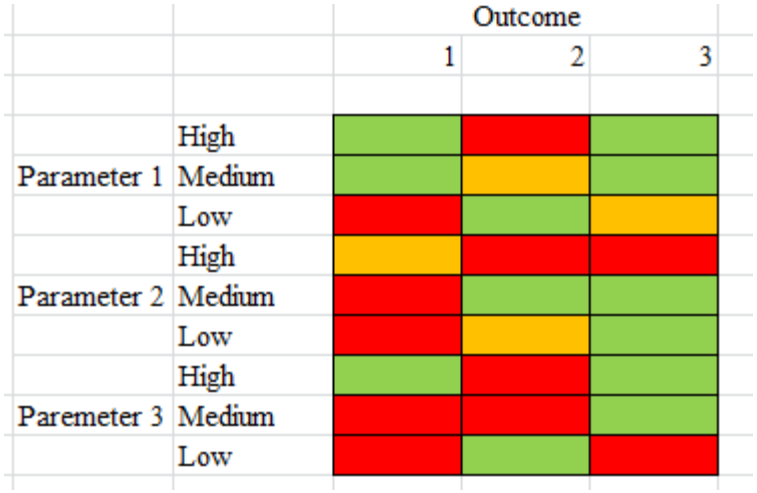


Figure 4.3 - Heat map IAT showing how the colour coding system could work if employed in the IAT.

This approach overcame the issues around application on an electronic platform, but was only capable of showing high, medium and low levels of each parameter/outcome. More cells would give better resolution, but would also make it more difficult to interpret at a glance. When this approach was enlisted for the original AbbVie version of the IAT, it was difficult to ascertain where each parameter and outcome started and ended. While this showed the lack of depth of information with the symbol-based system had been overcome, it did not satisfy the ease of understanding criterion for tool development.

The initial colour coding scheme appeared to generate a tool which satisfied both criteria to a limited extent, and so this approach was examined in greater detail. The ability to easily visualise the process parameter levels which corresponded to the desired process outcomes was a strength of the colour-based system. Nevertheless, the colours did not capture the response in its entirety. If three parameters had a positive linear effect on a particular outcome, the differences in impact were not captured. The complex nature of bioprocesses, especially fermentation, means that the relationships between parameters and outcomes do not always fall within set categories. For example, a linear increase may suddenly cease when a parameter reaches a certain level, or a slight decrease may increase drastically below a certain threshold value. These intricacies could be better captured by the colour-coding system than the previous symbol-based system. However, a method utilising shapes was devised which was fully capable of capturing the detail of these

relationships, in a comparable way to the outputs from DoE approaches (Figure 4.4). The shapes system could be completed using Microsoft Excel using the Shapes function. The resulting tool was effectively a grid of "reaction profiles" showing where optima existed, which reactions were positively/negatively linear (and to what degree), and allowed asymptotic curves to be captured in more detail than the original "S". The location of the asymptote was now shown relative to the high and low levels of each parameter/outcome combination, rather than simply showing that an asymptote existed.

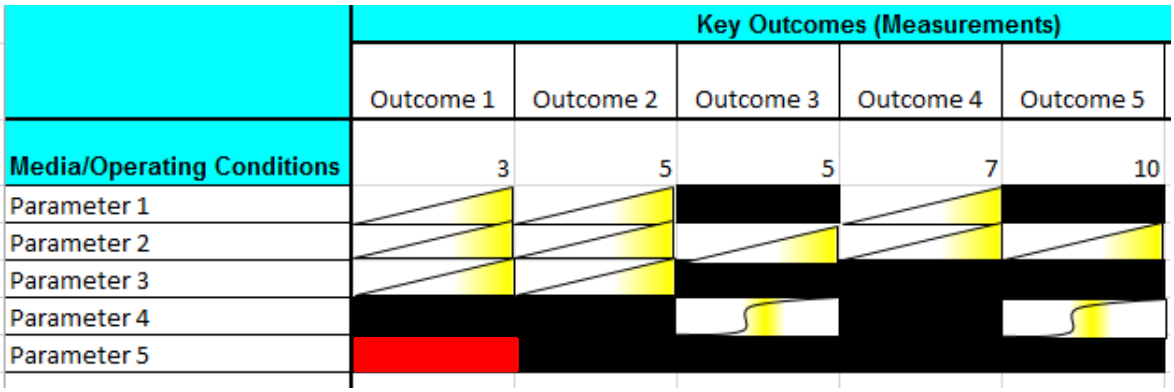


Figure 4.4 - IAT constructed using a combination of shapes and colour to represent the interactions between process parameters and outcomes. Cells filled in red are unknown relationships, cells filled in black are where no relationship exists.

The shape system worked well for capturing the reactions within the process, and was a marked improvement for implementation from the colour system. The lack of colour made the tool difficult to interpret at a glance, and so some colour was incorporated to show where the process would need to operate to obtain the optimum result for each outcome. Outcomes were split into desirable (yellow) and undesirable (blue), and the areas in the cells corresponding to the optimum value for each outcome (maximised for desirable, minimised for undesirable) was highlighted with the corresponding colour. The colours yellow and blue were selected as a result of previous feedback around red/green colour-blind users, but these could be changed based on user preference.

The resulting tool, at this stage of development, consisted of a structured approach for implementation, and also showed in an information-rich manner where the process should be

operated for optimum output. The next stage of development was to consider the action required when optimum conditions for a process parameter were different for multiple outcomes.

4.3.3 Weightings

The original tool contained weightings for the outcomes relating to the impact on the process and the business. These were displayed at the bottom of the tool as the final row. This removed the focus on the weightings, and so the first stage of development was to move this to become the row underneath the outcomes. In this way, it was more obvious which outcome corresponded with which weighting.

Weightings were displayed using a numerical system, balancing both the business and process benefits. The incorporation of these into a single value meant that it could be unclear when revisiting the tool at a later date on what basis the weighting was assigned. To rectify this, the weightings were split into two. Each outcome was assigned a weighting with a value between 1 and 5 for the potential business benefit, and then a second weighting between 1 and 5 for the potential process benefit. In the initial development version, the weightings remained separate, but when the tool was redeveloped further a row for combining the scores was added. This is discussed further in the next section, and the weighting system robustness (along with alternative systems) is discussed in detail in Chapter 7.

The final development was the added level of clarity when considering the weighting assignment. The original system comprised of only two defined levels (1 being "Improvement in outcome is unnecessary for business and process needs" and 10 being "Improvement in outcome would have significant process and business benefit"), meaning that some variability could be introduced between users. While systems of this nature are common in QbD assessments (C.M.C Biotech Working Group, 2009; Patil and Pethe, 2013; Kepert *et al.*, 2016), they do not fully capture the process as a whole, only the most critical parameters/outcomes. Qualitative tools are notoriously difficult to reproduce in a consistent manner (Konstantinov and Yoshida, 1992; De Ruyter and Scholl, 1998; Patton, 1999; Glassey *et al.*, 2000) and as this is a long known weakness of the

approach, a more robust weighting assigning system was devised. Weightings for the outcomes had a value of between 1 and 5, and the associated importance of the outcome for the improvement of the business/process is outlined in Table 4.1. The same weighting categories are used for both business and process impacts.

Table 4.1 - The weighting scale and the corresponding definition for use within the IAT

Weighting	Business Case Definition	Process Case Definition
1	Unimportant to business case	Unimportant to process
2	Slightly Important to business case	Slightly Important to process
3	Important to business case	Important to process
4	Very Important to business case	Very Important to process
5	Critical to business case	Critical to process

The more clear definition of the weightings aims to reduce the potential for ambiguity across different users, however ambiguity and alternative systems are discussed in more detail in chapter 7.

4.4.4 Final Output

The final output of the tool was an important consideration for tool development. As with many of the Britest tools, the original output for the IAT was the completed tool, with decisions about further work or process changes being based on the discussion generated through completing the

tool. While qualitative output of this nature is valuable, the IAT as a tool was shown to lend itself to an output of a semi-quantitative nature. The revised tool, at the stage of development discussed, consisted of coloured "profiles" as an indicator of the interactions between each parameter and outcome, weightings to show how important the improvement of each outcome would be to the business and process respectively. Additionally, in practice, the focus of the ISA analysis and output should be guided by the ISA to ensure maximum benefit to the stakeholders. While the potential for improvement can be seen from analysis of the "profiles", there was no output to show how contradicting optima could be best handled. For example, a parameter may positively influence one outcome, but negatively influence another, and while the weightings may make it obvious which outcome would be more important to improve, this is not as clear when high numbers of outputs and parameters are being considered. To rectify this, a scoring system was devised to allow the balancing of outcomes to be shown in relation to the parameters.

The scoring system is not complex, allowing it to be calculated by hand rather than using complex computer software. Three columns were added to the end of the IAT, labelled "Drive to Increase", "Optimum" and "Drive to Decrease". They could either be calculated separately for the business and process (using the two individual weighting values), or the weightings could be added (or multiplied if the user felt appropriate) together to give an overall weighting from 2-10 which could be used. This would be determined by the user, and could be the subject of further tool development as case studies are constructed.

The score is then generated using the coloured areas of each cell in the table. If the coloured area is to the right of the cell then it is a drive to increase (as this corresponds with high levels of the parameter), and the weighting for that combination is added to the "Drive to Increase" column as a positive integer (blue circles on Figure 4.5). These are added up to give a final score for the "Drive to Increase". This score will always be positive and represents the overall incentive to increase each parameter in turn. If the coloured area is to the left of the cell then it represents a "Drive to Decrease", and so the weightings from these columns are converted to be negative, and

are subtracted from each other in this column. This gives an overall negative value in the “Drive to Decrease” column for each parameter in turn (red circles in Figure 4.5).

The ‘Optimum’ column is employed if the coloured area is in the centre of the cell, usually when a bell shaped curve has been required. It is calculated in the same manner as the “Drive to Increase/Decrease” columns, and is positive. A diagram showing the calculation of two of the three score columns is shown as Figure 4.5, to illustrate the generation of the final output in a simplified form.

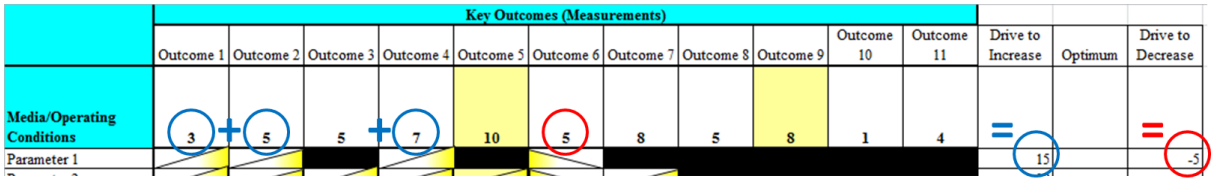


Figure 4.5 - Calculation of the "Drive to Increase/Decrease" columns in the IAT. The blue circles represent scores to be added to the "Drive to Increase" column, and the red circles represent scores to be added to the "Drive to Decrease column".

Once the values for these final columns are calculated, they can be used to determine a future plan for optimisation. The parameters which have a large difference between one column and the other two show a clear incentive to make a change to the process. Those which have very little difference show no incentive to change the process from current operating conditions. This system does not make the tool fully quantitative, but the ability to compare outcomes using numerical measures means it can be classed as semi-quantitative. This is not as powerful as a large dataset where multiple studies have been employed and effects can be statistically analysed, but is powerful enough to inform the design of such studies. Potential improvements are not in absolute values, but in relation to the other parameters which it is proposed would be of great value in early stage studies where data is not available. The final IAT is shown in Figure 4.6.

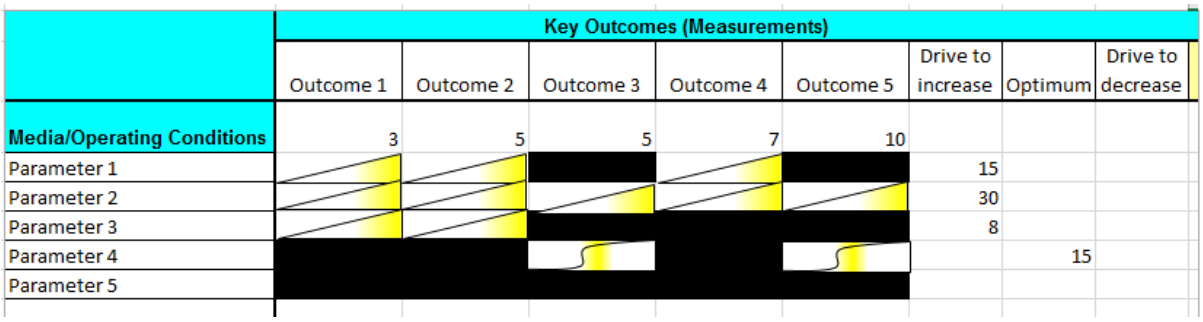


Figure 4.6 - Final version of the IAT. This is a much more visual tool that that originally devised (Figure 4.1), and gives the user a clearer course of action following tool utilisation.

4.4.5 IAT completion

In chapter 2 of this work, the simple process of making a cup of coffee was used to illustrate how the Britest toolkit would be used on a process, without a requirement for any detailed scientific knowledge. For the same reason the IAT was constructed for the same simple process, to demonstrate how the tool would be constructed and applied.

The first stage of the IAT is to complete the outcomes list, where each column is attributed an outcome associated with the process (Figure 4.7), in this case desired sweetness, temperature, bitter flavour and calorie content. The parameters are then completed down the rows, in this case amount of coffee beans, sugar, milk and water temperature.

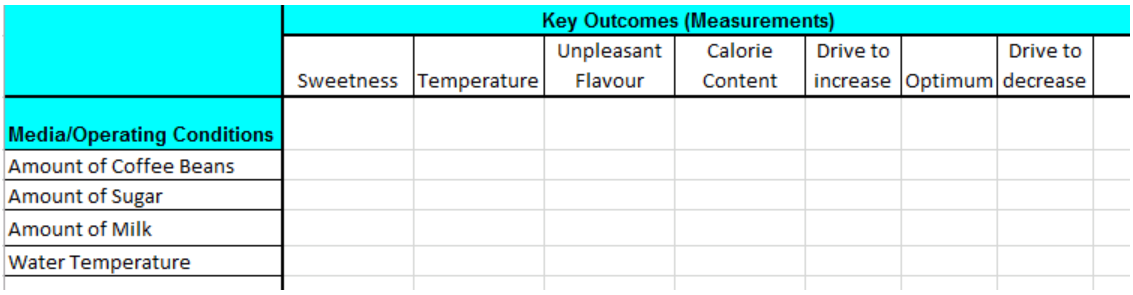


Figure 4.7 - The first 2 stages of building the IAT, in this case for the process of making a cup of coffee.

From this stage, the weightings are then completed, where the business case and potential for process improvement are considered using the guide set out in Table 4.1, shown here in Figure 4.8. In this example of making a cup of coffee, sweetness was deemed a 3 with respect to business case (as it could incur an extra business cost if provided free to the customer), and a 3 with respect to process improvement (as it is not the primary attribute of a cup of coffee). Assuming the coffee

was a standard filter coffee, temperature would be an important attribute to increase, giving it a 4 with respect to process improvement, and a 4 with respect to business case as a cold cup of coffee could prevent repeat business. An unpleasant and bitter flavour would be a 5 with respect to process improvement, as it is critical to providing a quality product, however was only designated a 3 with respect to business case as often other factors such as convenience can maintain customer levels. Calorie content would vary depending on the customer, and so was attributed a 1 with respect to business case and a 1 with respect to process improvement. In a real-world example this weighting assignation would be carried out by an interdisciplinary team.

	Key Outcomes (Measurements)						
	Sweetness	Temperature	Unpleasant Flavour	Calorie Content	Drive to increase	Optimum	Drive to decrease
Media/Operating Conditions	6	8	8	2			
Amount of Coffee Beans							
Amount of Sugar							
Amount of Milk							
Water Temperature							

Figure 4.8 - IAT for making a cup of coffee once the weightings have been attributed according to the system in Table 4.1.

The next stage of the IAT completion process is to show where relationships do not exist (in black), and if relationships do exist what the relationship looks like. In this case (Figure 4.9) for example there is no relationship between the amount of beans and both temperature and calorie content, and the amount of sugar does not influence the temperature of the coffee. However the more milk that is added to the coffee, the lower the temperature will be. In addition, the more sugar that is added the higher the calorie content. At this stage desirability of a quality is not indicated by colour, the relationship is therefore shown with white shapes.

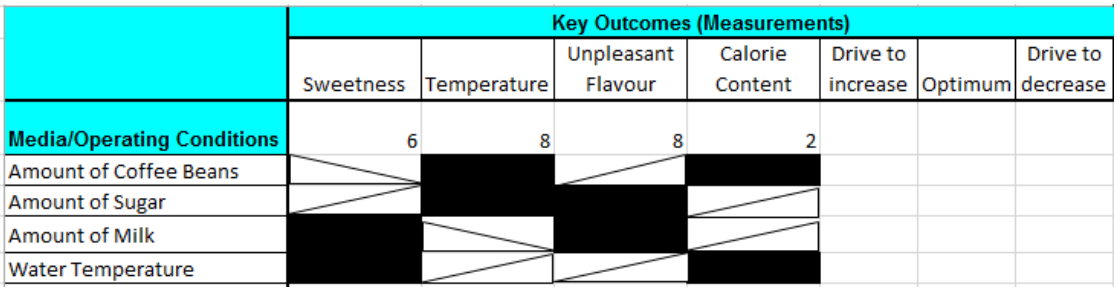


Figure 4.9 - Relationships between the parameters and outcomes for making a cup of coffee, shown in the IAT.

The next stage of completing the IAT is to consider the desired area of operation using the shading system. Yellow is used for features which are desirable, blue for those which are undesirable. This is due to the relative rarity of blue/yellow colour blindness (also known as Tritanopia) compared to red/green colour blindness (Simunovic, 2010), and ensures the broad majority of potential users would be able to distinguish between the colours used. The area of the shape which represents the desired outcome is the area which is shaded. In this example (Figure 4.10), the assumed desire was a standard filter coffee with milk and one sugar. This is why the middle of the amount of sugar vs sweetener (shown in the red oval) is shaded in the centre. If the drinker preferred 3 or 4 sugars, the area towards the right-hand side would be shaded instead to indicate this preference. Conversely the unpleasant flavour associated with using water which is too hot was sought to be avoided, and so the area towards the left for the shape was shaded blue (shown in the green oval).

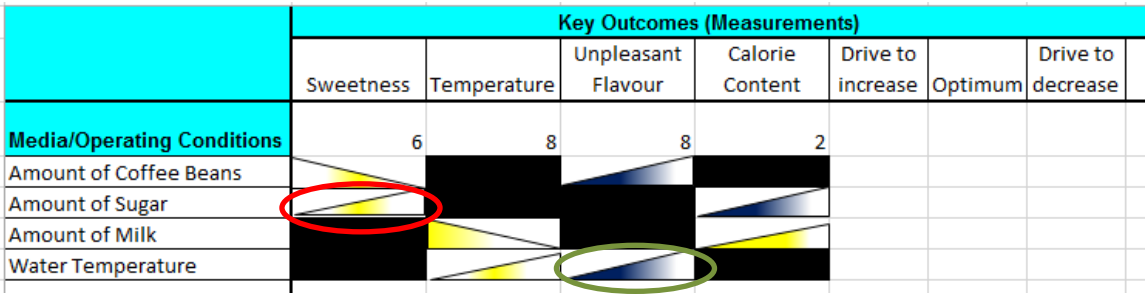
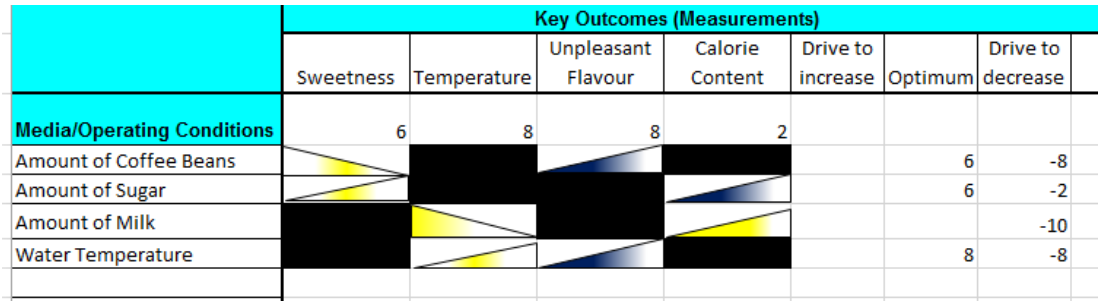


Figure 4.10 - The IAT for making a coffee with milk and one sugar when shading is completed. Yellow shading indicates desirable characteristics, blue indicates undesirable.

The final stage is to complete the calculation to give the drive to increase and decrease each parameter, illustrated previously in Figure 4.5. In this case (Figure 4.11) there is a drive to limit the amount of coffee beans and sugar used, a drive to decrease the amount of milk in the coffee, and a drive to maximise the temperature of the water but only to a defined limit.



	Key Outcomes (Measurements)						
	Sweetness	Temperature	Unpleasant Flavour	Calorie Content	Drive to increase	Optimum	Drive to decrease
Media/Operating Conditions	6	8	8	2			
Amount of Coffee Beans						6	-8
Amount of Sugar						6	-2
Amount of Milk							-10
Water Temperature						8	-8

Figure 4.11 - The completed IAT for making a cup of coffee where the drinker wishes for milk and one sugar.

4.5 Summary

This chapter showed the redevelopment process for the IAT tool. The tool was originally developed through industrial collaboration, but had several associated challenges which prevented Britest integrating it into the standard the toolkit. The tool was redeveloped through structured changes to the table layout, order of application, information capture system, weightings, and the addition of a scoring function to generate a conclusive final output. The new IAT offers a user-friendly approach to systematically analyse the potential impact of each process parameter on each process outcome. The effective linking of parameters and outcomes is imperative to the adoption of a QbD approach to bioprocessing. In order to be able to make changes to a process during operation to positively influence the outcome, the operator must fully understand the link between process parameters and product attributes. Without this understanding, the impact of changing the process cannot be fully understood, and so changes cannot be performed with confidence. The IAT would be used to encourage the linking of the parameters and outcomes from an early stage of process design/optimisation, supporting effective experimental design. The IAT tool would be particularly well suited to early stage studies where there is a major drive to "do more with less" on a short timescale. This tool could be used to help minimise the

experimental burden in these development stages, and ensure that experimental plans are based on clear process understanding. The semi-quantitative nature can aid the user in prioritising the further sets of experimentation to be performed.

The subsequent Chapters will test the IAT on upstream and downstream unit operations (Chapters 5 and 6 respectively) to ascertain the ability of the tool to handle the range of data types which can be generated within bioprocessing. Chapters 5 and 6 discuss the application of the tool to datasets accessed through an academic collaboration, due to limitations in suitable data availability from within the Britest consortium. Finally, Chapter 7 explores the sensitivity of the weightings within the IAT.

Chapter 5 Upstream Testing of the Interaction Analysis Table (IAT)

5.1 Introduction

The previous chapters of this thesis have set out the bioprocessing background (Chapter 1), the Britest tools in their original form (Chapter 2), the basic redeveloped toolkit using "best case scenario" simulated processes (Chapter 3) and the redevelopment of the Interaction Analysis Table (IAT) tool (Chapter 4). This Chapter will focus on the application of this tool to an upstream processing dataset which was originally generated as part of a research study at Technical University Berlin (TUB), exploring the optimisation of heterologous protein production in *E. coli*. .

Heterologous protein production is the manipulation of an organism to produce a protein which would not be produced in the untransformed "wild-type" organism. It employs the recombinant gene sequence for a polypeptide to produce the protein. This manipulation has been applied to many organisms including whole plants, whole insects, whole animals and a range of cell culture types (Gordon *et al.*, 1980; Shinmyo *et al.*, 2004; Van Der Vossen *et al.*, 2005). As discussed in Chapter 1, there are four main options for host expression system which are widely used: mammalian, insect, yeast and bacteria. Each of these has its own merits and drawbacks, and all have their place in both research and industrial systems. In general, micro-organisms are the favoured host due to the rapid generation time, higher reliability and ease of handling (Sadava *et al.*, 2009; Edwards, 2011). They have been used for many years and so a range of well characterized expression systems are available.

Escherichia coli (*E. coli*) is a common host for the expression of proteins which do not require complex post-translational modifications to be applied, usually proteins of prokaryotic origin. For this study, *E. coli* was employed to recombinantly express alcohol dehydrogenase (ADH). The ADH produced within this study is derived from *Lactobacillus*. Alcohol dehydrogenase is an enzyme often used as a biocatalyst (Leuchs and Greiner, 2011) to catalyse the reduction of carbonyl compounds to enantioenriched (r)-alcohols in an enantioselective manner (Müller *et al.*,

2005). It is known for its versatility, making it a valuable product within the biotechnology market. Other attractive properties include the stereo-selectivity, producing almost exclusively (R)-alcohols (Leuchs and Greiner, 2011), and the substrate specificity (Wolberg *et al.*, 2001; Ernst *et al.*, 2005). It is effective in a range of atypical conditions including in the presence of organic solvents and gaseous reactants (Leuchs and Greiner, 2011). It has been suggested that its activity could be further enhanced by supplementing with co-factors (Machielsen *et al.*, 2009). The main drawback to any biocatalyst is the high purification costs incurred when compared to chemical catalysts (Faber, 2011). Host cell proteins must be fully eliminated from the enzymes to ensure additional unwanted reactions are avoided (Bommarius and Riebel-Bommarius, 2007). Whole cell systems can be employed to reduce purification costs for the biocatalyst producer, but with this comes the increased risk of undesired reactions. In addition, both purified enzymes and whole cell systems may have a narrow range of process conditions under which the desired catalytic reaction occurs. Enzymes which arise naturally from a biological origin tend to be sensitive to temperature and pH, which is not always beneficial depending on the desired reaction conditions (Zhao, 2006). Within industry, *Lactobacillus* derived ADH is used to produce chiral alcohols, which may be used as building blocks within fine chemical and pharmaceutical production (Schmid *et al.*, 2001; Schoemaker *et al.*, 2003; Panke *et al.*, 2004).

The Bioprocess Engineering group at TUB performed a series of optimisation experiments for the production of ADH from *E. coli*, investigating the effect of three media components on the outputs of the process (Knepper, 2014). The outputs were measured at specific time-points, which were not always equidistant. This led to the generation of a dataset containing 24 experiments, which did not contain evenly spaced sampling points, and did not have readings for all outputs at all time-points. While 24 experiments are not a large dataset, the original work involved visual analysis of the results using line graphs, with no structured approach to data analysis being employed. For a single factor this approach would not be impractical, however with 24 experiments to compare consisting of three factors and five outputs the comparison was time

consuming. Additionally, there was the possibility that different people analysing the results would come to different conclusions, due to the lack of structure in the analysis.

The IAT was identified as being the tool most suited to a study of this nature. This was due to the upstream focus of the study, as previous work has shown the high level of complexity within fermentation makes the Britest tools difficult to apply. The IAT was designed to link process parameters to outcomes, which is consistent with the dataset generated through these experiments. Once the tool had been redeveloped to overcome the problematic application process (Chapter 4), the TUB dataset was used as an alpha-test of the tool in its redeveloped form.

This Chapter reports the application of IAT for analysis of the dataset generated by TUB, and compares the outcome of this analysis to the conclusions drawn in the original study, and the outcome of a statistical analysis of the dataset using Design of Experiments (DoE).

5.2 Methods

5.2.1 Experimental

The production of ADH in *E. coli* was sought to be optimised in these experiments, where the ADH was recombinantly expressed. The ADH in this case was simply a protein which could be measured easily and is reliably produced, creating a model system, rather than a molecule which would be industrially beneficial. The dataset focussed on investigating the effects of changing three media components on the outcome of the fermentations.

The experimental methodology was performed as reported by Knepper et al. (2014). In summary, the *E. coli* was grown in 96-well microplates in an orbital shaker and attached to a liquid handling system to allow automated sampling. Three media components were used at differing levels to examine their effect on the output of the fermentation. These are detailed in Table 5.1. Reagent A is a glucose-releasing biocatalyst. Measurements taken were cell growth (using OD_{600}), concentrations of acetate and glucose, activity of ADH and the pH. The measurement methodology is detailed by Knepper (2014) and Ukkonen (2011).

Table 5.1 - Summary of the experiments performed for the TUB study. A tick indicates the level of a component used in an experiment; cells filled in black indicate this level of a component was not present in the experiment.

Experiment	Reagent A		Lactose		Glucose		
	0.6 U L^{-1}	1 U L^{-1}	0g L^{-1}	0.5 g L^{-1}	0g L^{-1}	0.5 g L^{-1}	1 g L^{-1}
A	✓		✓		✓		
B		✓	✓		✓		
C	✓			✓	✓		
D		✓		✓	✓		
E	✓			✓		✓	
F		✓		✓		✓	
G	✓			✓			✓
H		✓		✓			✓

5.2.2 Original Analysis

The original work sought to demonstrate the benefit of an automated system for optimisation work of this nature, and so no statistical analysis of the results was performed. Instead, the experiment which yielded the highest result for ADH activity was selected as representing the optimal conditions. Verbal discussions with the TUB researchers indicated that an attempt was made to understand the relationships between parameters and outcomes, but that this did not enhance process understanding due to the complex nature of the dataset.

5.2.3 IAT Analysis

The dataset generated by TUB contained process parameters and outputs, but had no structured methodology in place for analysis of the results. The conclusions drawn by the TUB researchers

were not a result of structured analysis, and did not consider each of the potential interactions present within the fermentation. The IAT was employed to test whether the tool could have identified interactions further to the original study. The IAT was completed from the study data, and was compared to the original output after completion. This ensured that conclusions from the tool were not being drawn with bias towards the original results. The IAT was completed with respect to the final values for each output. This was recorded at 44.3 hours for ADH activity, OD_{600} and pH, and 36.9 hours for acetate and glucose concentrations. The weightings within the IAT as defined in Chapter 4 are usually comprised of two individual weightings, related to the business benefit and the process benefit (discussed in Chapter 4). For this study they were generated only with respect to the process, as the academic nature of the study meant that no business benefit could be readily attributed.

5.2.4 Design Expert 9

Subsequent to qualitative analysis through the original study, and semi-quantitative analysis through the IAT, the results were analysed with Design Expert 9 ™, a software package intended to statistically evaluate Design of Experiments results. This study used a historic design to analyse the results from TUB. This ensured that the existing experimental setup and results available would be analysed, and extra experiments would not be required. The average values for each response were used for the analysis, meaning no replicates could be incorporated into the model. The p-value for significance was set at 0.05. Each response was considered individually, prior to optimisation to maximise cell specific productivity.

5.3 Results and Discussion

5.3.1 Original Analysis

The full results for each experiment are shown within Table 5.2. The TUB analysis concluded that the optimum conditions for the production of ADH from *E. coli* in the microwell plate system were as set for experiment G (shown as Figure 5.1-graph of outputs), which contained 0.6U/L of

Reagent A, 0.5g/L Lactose and 1g/L Glucose. As previously stated, this conclusion was based on the highest value of ADH activity, as attempts to understand the interactions between the parameters and outputs were time consuming and did not yield satisfactory results. Analysis of this nature with complex systems such as fermentation is unlikely to be predictive, as the understanding of the relationships between parameters and outputs is limited.

Table 5.2 - The results obtained from each experiment (Knepper, 2014)

Experiment	OD_{600}	pH	Glucose (g/L)	Acetate (g/L)	ADH Activity (U/L)
A	4.9	6.6	0.5	3.2	216.1
B	7.2	6.4	0.6	4.8	184.5
C	15.1	6.1	0.8	8.1	215.9
D	17.9	6.1	0.9	8.2	435.6
E	17.3	6.1	0.9	10.0	1096.3
F	20.0	6.1	0.9	8.7	854.0
G	15.8	6.1	0.7	8.0	1352.8
H	16.8	6.0	0.8	8.5	1255.8

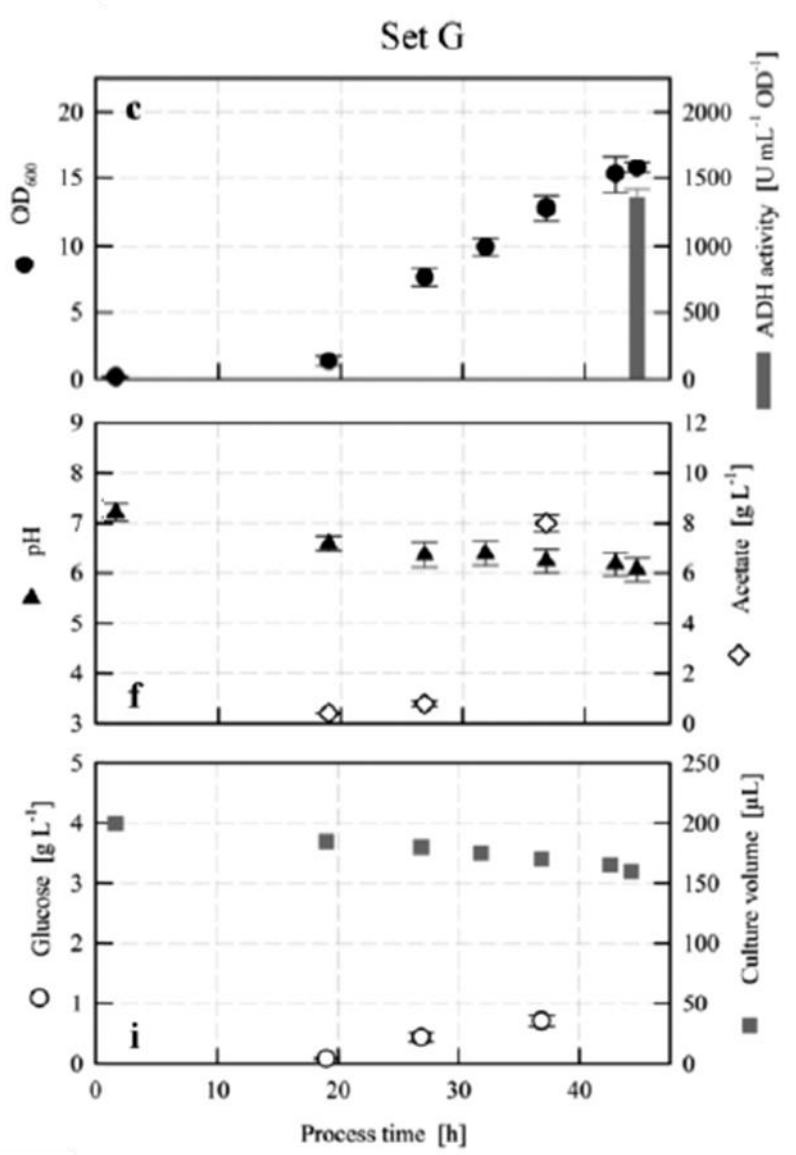


Figure 5.1 - The experiment determined to be the optimum conditions by the original research analysis. Reproduced from Knepper et al. (Knepper, 2014). The top graph (c) shows the OD_{600} measurements (solid circles) over the course of the fermentation and the final ADH activity levels (solid bar). The middle graph (f) shows the change in acetate (diamonds) and pH (solid triangles) over the course of the fermentation. The bottom graph (i) shows the change in glucose (circles) and culture volume (solid squares) over the course of the fermentation.

5.3.2 IAT Analysis

The IAT was constructed by considering each outcome individually, with ADH activity being the last to be completed. This ensured that the tool was not reverse engineered to fit the TUB researchers' optimum, and instead arrived at the conclusion in an independent, structured manner. The IAT constructed from the results is shown as Figure 5.2.

The weightings were assigned based on process benefits alone, consistent with the scale outlined in Chapter 2, due to the academic nature of the study. The overall growth was assigned a weighting of seven. The growth of the cells is imperative to the production of the ADH; however increasing cell specific productivity would be more beneficial than simply enhancing growth, justifying the value of seven. The acetate and pH deviations were both assigned weightings of four. Considering these in turn, an increasing acetate level will inhibit ADH production to an extent, but the cell growth and viability is inherently linked to the acetate production (Takahashi *et al.*, 1999). A weighting of four was therefore appropriate to demonstrate there is an adverse impact to acetate generation, but this is considered a necessary sacrifice within the system. Acetate is a waste product from the fermentation, when more cells are present more ADH will be produced, but more acetate will also be produced (Luli and Strohl, 1990; Han *et al.*, 1992). The pH deviations are similar in that the pH will always change within fermentation as waste products are produced. *E. coli* is known to produce acidic waste, therefore lowering the pH as cell growth occurs. This is exactly the effect seen in each of these experiments, making this acidic waste the likely reason for a reduction in pH. If no waste was produced the cells would not be growing and therefore no ADH would be produced. The final weighting was the ADH, attributed a value of ten. The ADH is the target product and maximisation of this was the primary aim of the experiments. Its value is critical in deciding the next round of optimisation, hence the highest weighting.

For simplicity and due to the sparse nature of the dataset, relationships between each parameter and outcome were assumed to be linear. The restriction of parameters to two or three levels meant that it was not possible to demonstrate asymptotic relationships, or those involving plateaus. It is recognised that the cell growth is likely to show an asymptotic relationship with all parameters if tested over a sufficiently wide range, which would require further experimentation to fully characterise.

For each parameter there was a compromise to be made between outcomes, showing the value of the weightings to the tool. This is common within bioprocessing, where waste will often inhibit

outcomes as previously discussed (Takahashi *et al.*, 1999). It is possible that other weightings for waste products would yield different results, and this sensitivity to weightings is investigated and discussed in Chapter 7. The IAT results indicate that the most significant factor for consideration in the next round of optimisation would be the increase of glucose concentration, followed by a reduction in the quantity of reagent A. As these are both linked to glucose levels, it would suggest that adding glucose as an individual component, rather than under the control of a biocatalyst, would better promote the production of ADH. If this strategy for optimisation was pursued, it would be notable that the oxygen in the fermentation would be depleted faster, and more waste would be produced. While the addition of glucose as an individual component would be possible within the fermentation, it does not consider the variation in availability of glucose over a period of time. Reagent A is employed to ensure sustained release of glucose through the fermentation, to enhance cell viability and production towards the end of the experiment. This is intended to simulate the fed-batch bioreactor environment. As with any data analysis tool, it is important, when using the IAT on a dataset, to incorporate the results with process knowledge, and in this case that could mean testing several combinations of glucose and reagent A both above and below the optimum indicated by the IAT. Lactose was the factor which showed the least potential to improve the process, having no discernible impact on the overall ADH activity. Irrespective of this, it was shown to have a positive influence on cell growth, and so would be worth including in future experimentation. *E. coli* has been shown to prefer glucose as a sugar source, and so the importance of lactose to cell growth could indicate that the cells do not have sufficient glucose for sustained growth (Donovan, 1996). In light of this, it would be proposed that the lactose could be investigated over a smaller range than the glucose, as it is anticipated its effects will be negated if sufficient glucose is present.

The consistent output between the IAT and the original investigation reinforces the usefulness of the tool. The IAT provided a structured methodology which could have aided the conclusion to be reached in a more robust manner. Additionally, a mechanistic justification was provided, rather than an empirical analysis based on evaluation of line graphs. It allowed the consideration of each

parameter's influence, which was not possible with the unstructured data analysis approach used in the original research. It also gave a clear direction for the subsequent optimisation experiments, and there were significant associated time savings.

When performing optimisation, many approaches are available. One Factor at a Time (OFAT) experiments can be useful for initial screening, but cannot identify interactions between factors. Design of Experiment approaches are designed to test for these interactions between factors to find the true optimum of a system. The IAT considers each factor individually, and so DoE was employed to compare the results from the IAT with an alternative approach which is quantitative and capable of predicting interactions between factors.

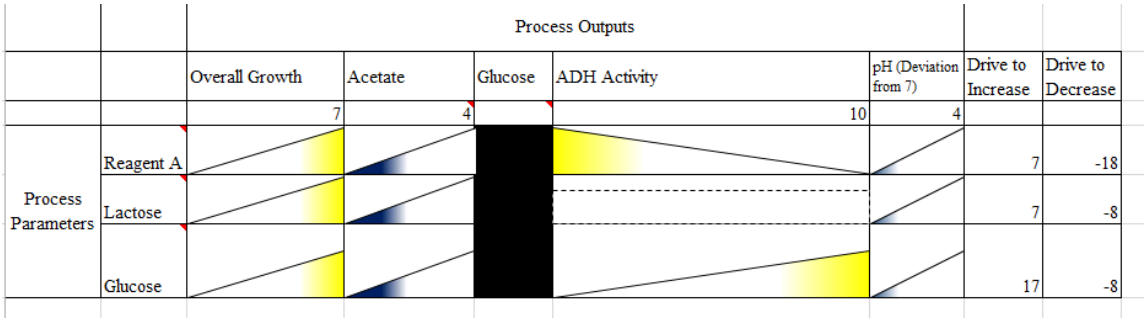


Figure 5.2 - IAT for the ADH fermentation. Each column corresponds to a measured output of the fermentation. Each row corresponds to the components being investigated. Shapes coloured yellow correspond to outputs sought to be maximised, Shapes coloured blue are sought to be minimised. Dotted shapes are used when no relationship could be discerned.

The IAT showed that the optimum conditions would be with low levels of Reagent A, and high levels of both lactose and glucose. This corresponded to experiment G, which was the experiment which the TUB researchers selected as the optimum. This consistency reinforced the tool accuracy; however, the additional information generated by the IAT would allow the TUB researchers to consider strategies for further optimisation, unlike the original analysis.

5.3.3 Design Expert 9

Experimental design analysis was performed using Design Expert 9 ™. A historic design was generated, and the results analysed with respect to each output individually before using the model generated for optimisation of the fermentation as a whole.

In light of the limitations of historic DoE analysis, it is important to highlight that the analysis was performed only to give an alternative analysis for comparison to both the original study and the IAT. It was not intended to generate a comprehensive model of the behaviours displayed within the fermentation and would certainly not be considered adequate for the purposes of optimisation and scale up. Each response was modelled separately prior to optimisation of the system using the five models generated.

Cell Growth (OD_{600})

Cell growth, measured through OD_{600}, was the first response considered. The normal plot of residuals (Figure 5.3) indicated that the significant term with respect to this response was lactose. The terms within the model all fell close to the line, indicating that the model was a good fit and noise levels were low. The ANOVA for this model attributed lactose a p-value of 0.0002, indicating a high level of statistical significance.

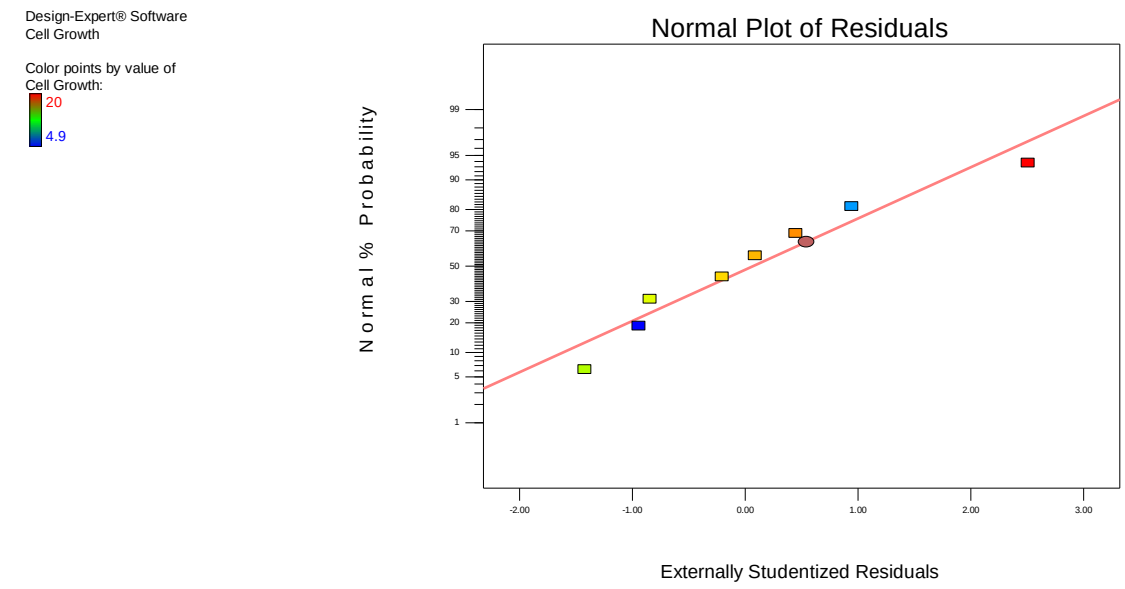


Figure 5.3 - Normal plot for the consideration of OD_{600} as a response. Lactose is shown to be significant through the distance from the normal effect line.

Considering the predicted vs actual plot (Figure 5.4) the results split into two clear sections, consisting of high and low values, but the trends within these groups could not be captured by the model. This is a result of the categorical nature of the analysis, and could be overcome if a response surface design was performed.

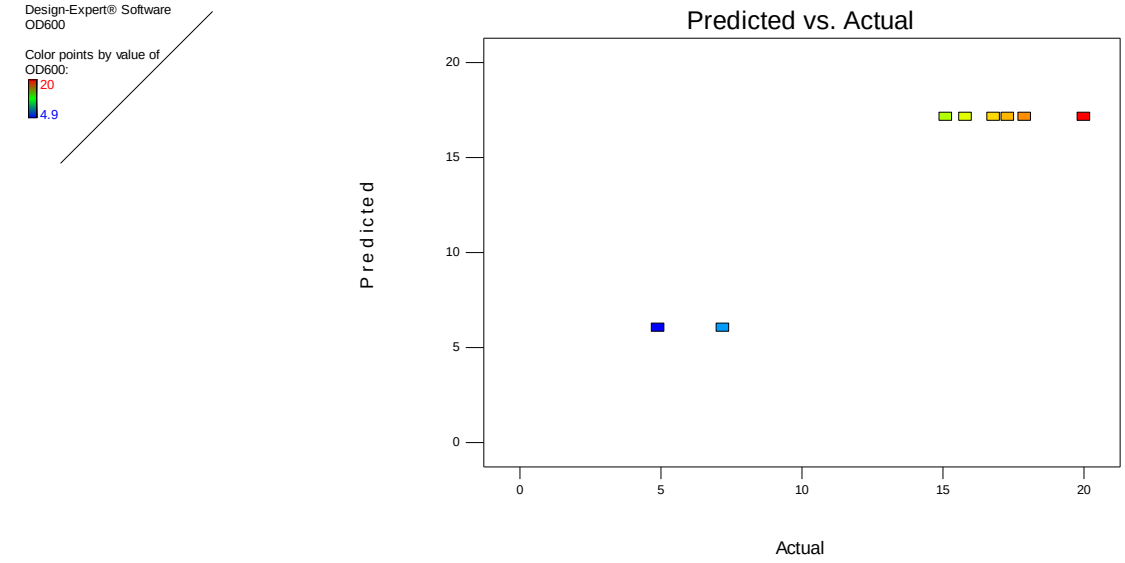


Figure 5.4 - Predicted vs actual plot for OD_{600}

The results showed the overall trend for the effect of lactose concentration on OD_{600} had been captured by the model, as points lie within a reasonable distance of the black solid line. All points would have fallen on the line if a perfect prediction had been achieved. The One Factor plot (Figure 5.5) did show a strong correlation between Lactose and OD_{600} response, increasing confidence in the model. When lactose was absent from the fermentation, cell growth was much lower than when lactose was present. The effects fall within a reasonably narrow 95% confidence limit, shown in Figure 5.5 as a dotted line, showing a high level of confidence in the inferred relationship.

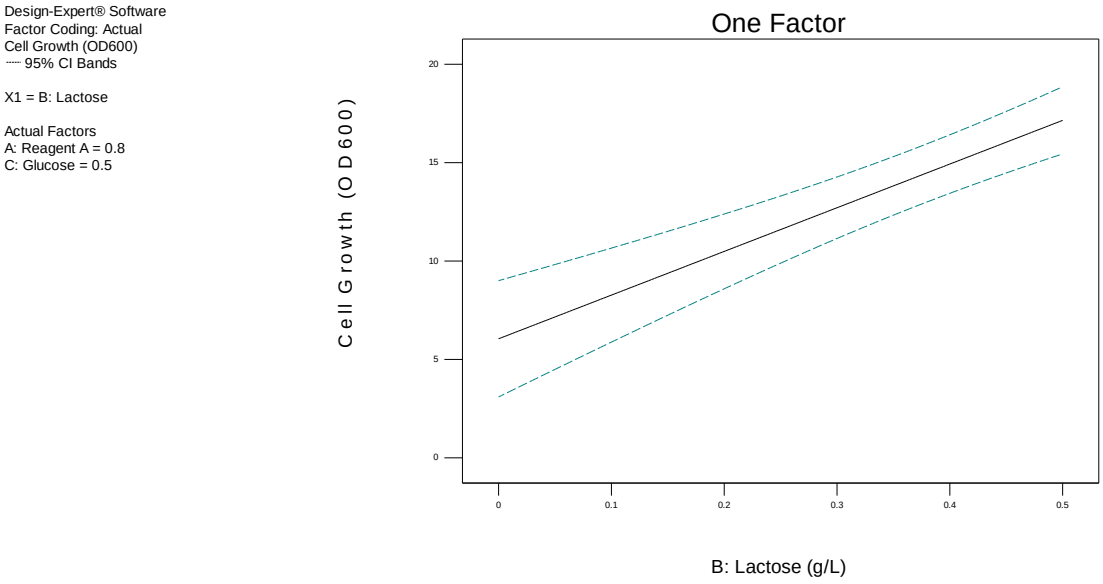


Figure 5.5 – One factor plot for OD_{600}

pH

Lactose was the most significant factor with respect to pH (p-value 0.0003), which was not unexpected. The pH of the fermentation will decrease as a result of waste product accumulation, which is directly linked to cell growth. Lactose is known to increase waste production (Donovan, 1996). High numbers of cells will mean a large amount of waste is generated, which will change the pH by a greater amount, as the pH was uncontrolled within the fermentation.

The groupings of high and low values were evident in this analysis, as with the analysis for OD_{600}. This is shown as Figure 5.6, the predicted vs actual plot, where two discrete groups were seen, one at low pH values, and another at higher pH values. The high variability within the group at higher pH further demonstrates the shortcomings of this type of analysis, but the general trend is captured. However, in this case there was increased variability within the categories, meaning the model was not as accurate as the previous analysis. The causes of this are likely not restricted to the categorical nature of the analysis. It would be anticipated that multiple factors are affecting the final pH which were not included within the model

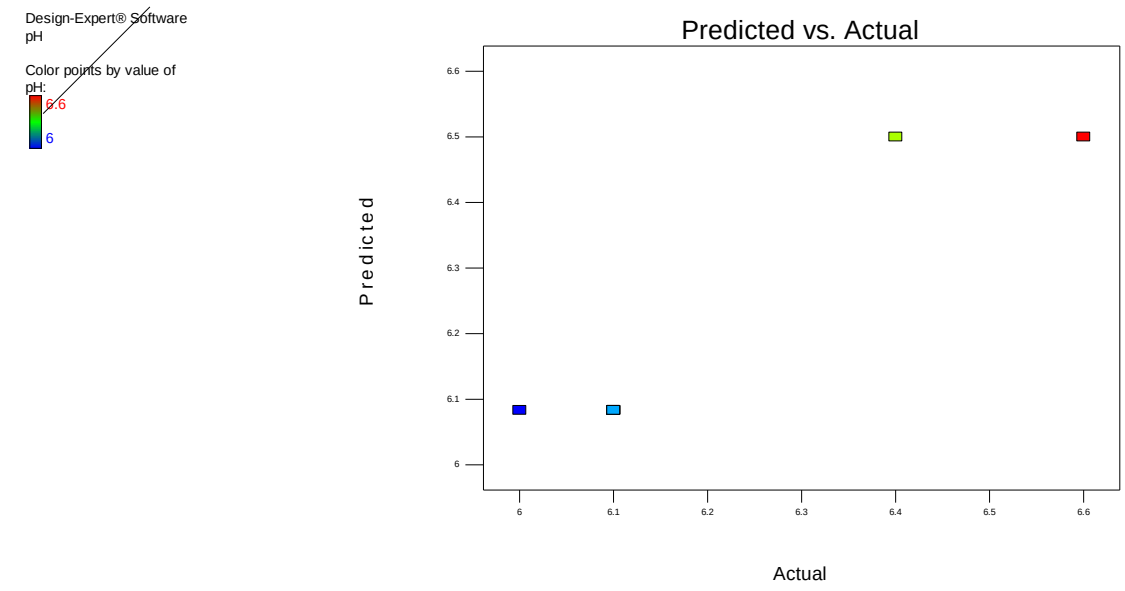


Figure 5.6 - Predicted vs Actual plot for pH

The results showed the overall trend for the effect of lactose concentration on pH had been captured by the model, as points lie within a reasonable distance of the black solid line. All points had fallen on the line if a perfect prediction had been achieved. As with the analysis for OD_{600} the One-Factor plot (Figure 5.7), shows the 95% confidence intervals were reasonably narrow, though broader than for the cell growth analysis, and at higher levels of lactose the pH was decreased.

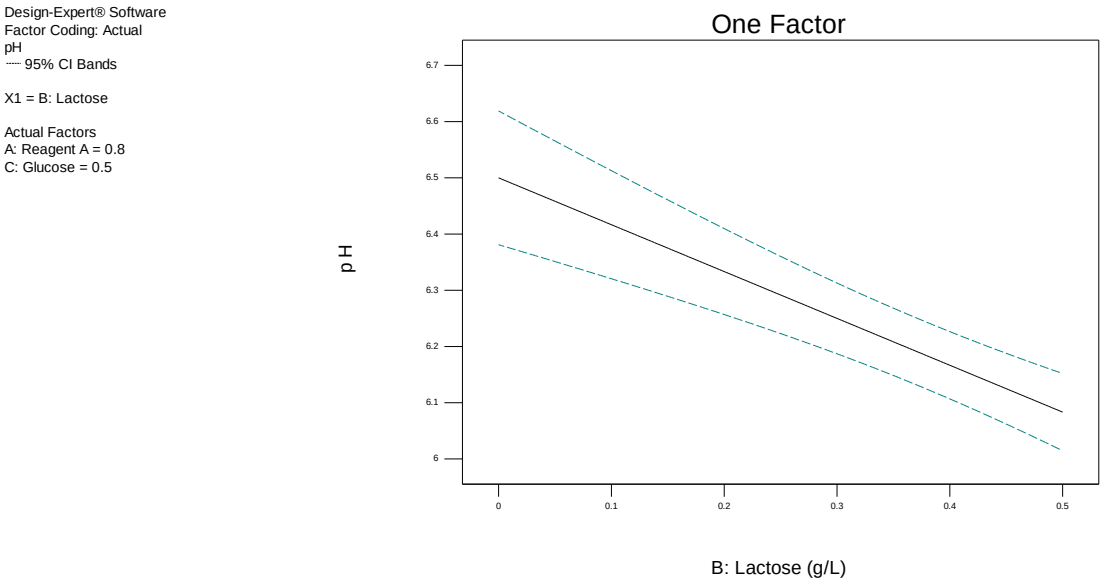


Figure 5.7 – One factor plot showing the interaction between lactose and pH

Acetate

As with pH, acetate concentration is related to cell growth and therefore OD_{600}. It was thus unsurprising that the results for this outcome were consistent with the pH and cell growth analyses in that lactose was the only significant factor with a p-value of 0.0005. The normal plot of residuals (Figure 5.8) showed noise within the model, potentially indicating a factor causing an effect which was not included within the model.

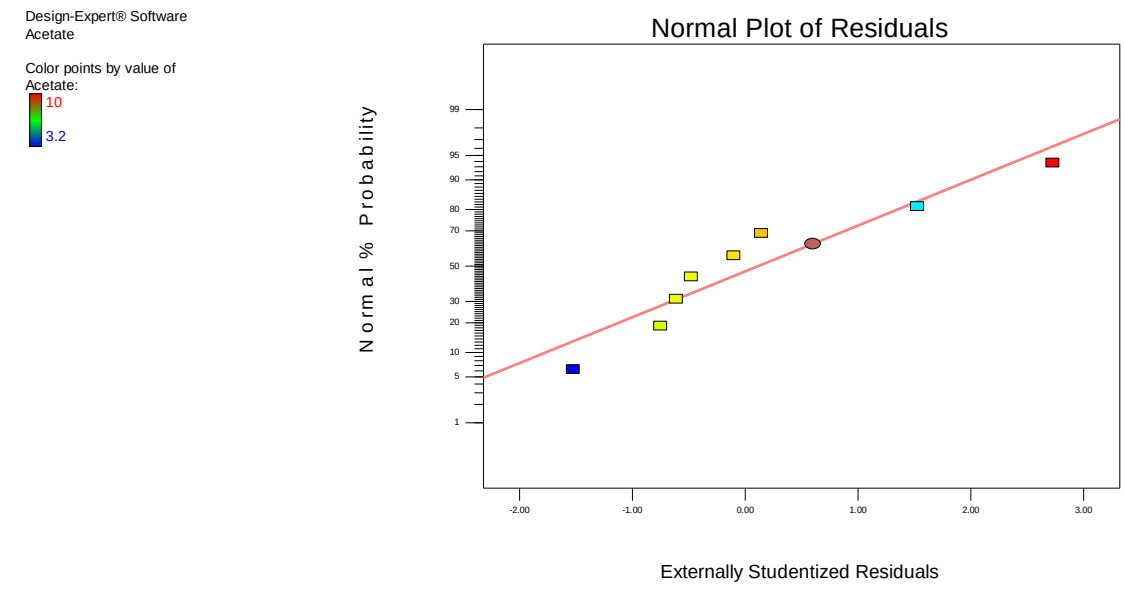


Figure 5.8 – Normal plot of residuals for the consideration of Acetate as a response

In keeping with previous analysis, the predicted outcome fell within reasonable 95% confidence interval on the one factor plot (Figure 5.9), indicating that the additional effect was not significant enough to undermine the relationship being inferred.

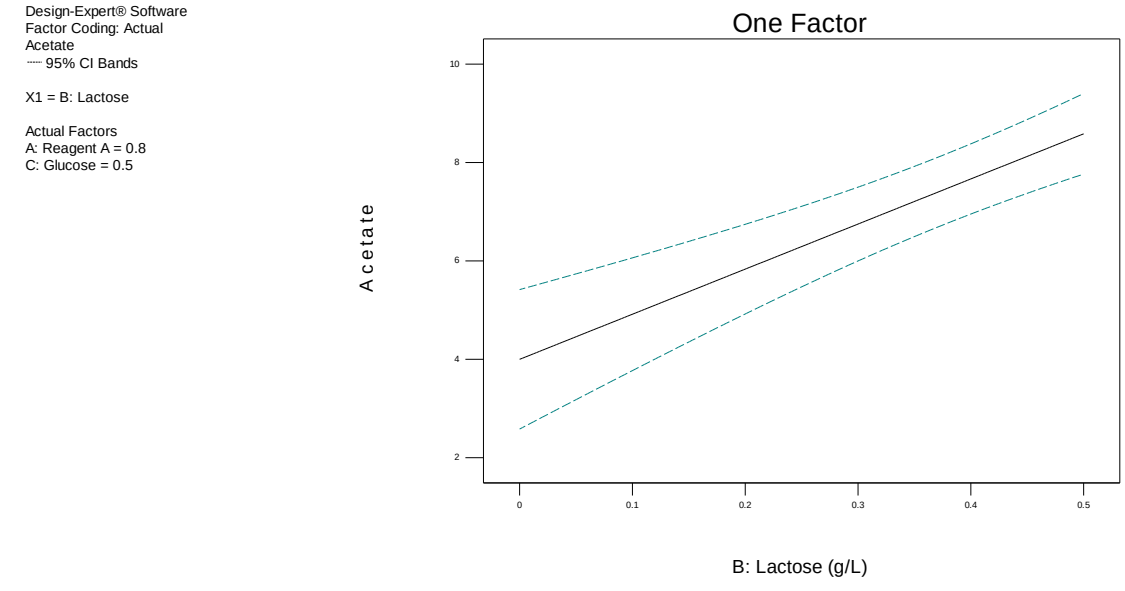


Figure 5.9 – One factor plot showing the effect of lactose concentration on acetate concentration

Glucose

Glucose was a medium component in two forms: active and inactive. The glucose measurements at the end of the fermentation would therefore be affected to three factors: the initial amounts of glucose and Reagent A respectively, and the amount of cell growth. When more cells were present, the glucose would be consumed faster, potentially leading to lower final values. It was therefore unsurprising that the significant term for the model was again lactose, the alternative source of carbon within the cultivation. This had a significance value of 0.0049, which was the highest significance value seen for lactose. This makes this the relationship most likely to be affected by noise, backed up by the normal residuals plot (Figure 5.10).

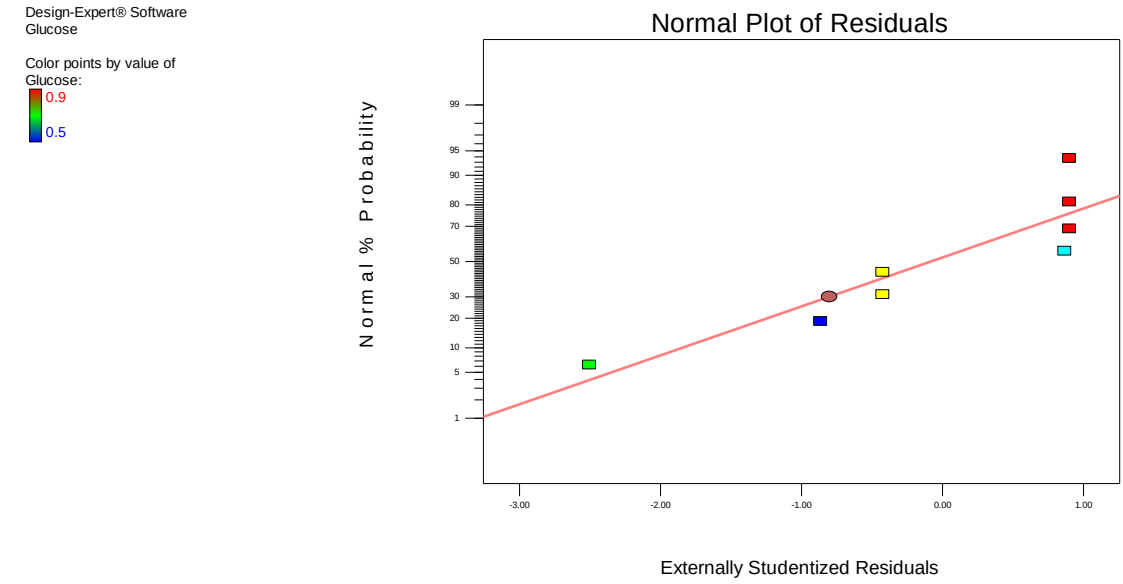


Figure 5.10 – Normal plot of residuals for glucose as a response

The One-Factor plot (Figure 5.11) shows this relationship, where high levels of lactose gave rise to higher levels of glucose. The 95% confidence intervals were fairly broad, consistent with the high levels of noise indicated in the Normal Plot of Residuals.

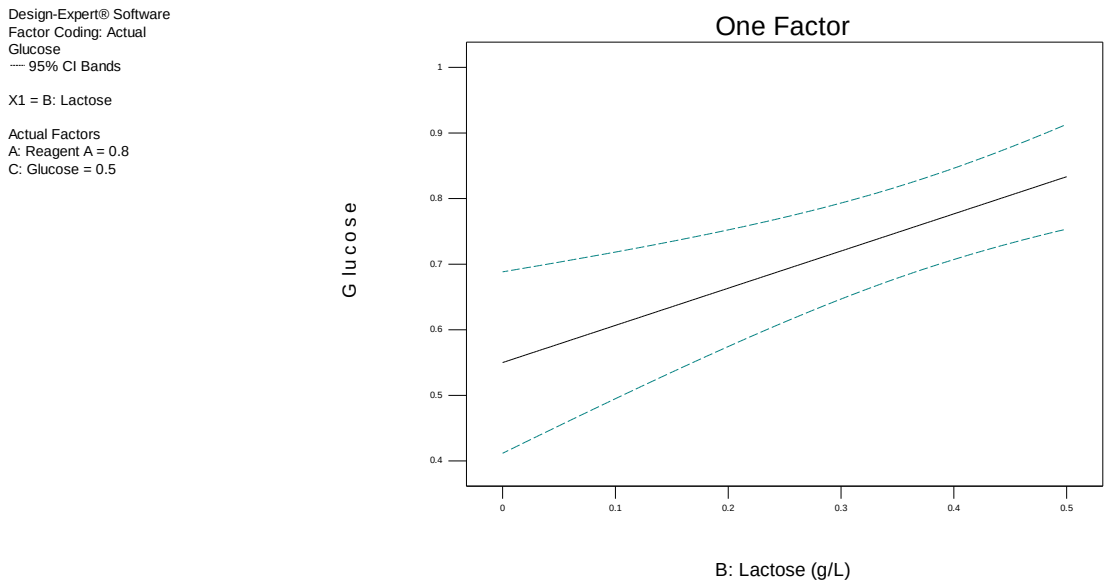


Figure 5.11 – One factor plot showing the effect of lactose concentration on glucose concentration

ADH Activity

All other factors were directly linked to cell growth, and so similar results were anticipated for ADH activity, as the most biomass present the likely more ADH being produced. The experiments generating higher cell numbers were anticipated to also generate high levels of ADH and ideally ADH activity, though activity is not always directly correlated to protein yield. It was therefore expected that lactose would again be the significant influencing factor on this attribute. It was therefore surprising that glucose was shown to be the only significant term for this model (Figure 5.12). The significance value of 0.0001 indicated this was a highly statistically significant effect.

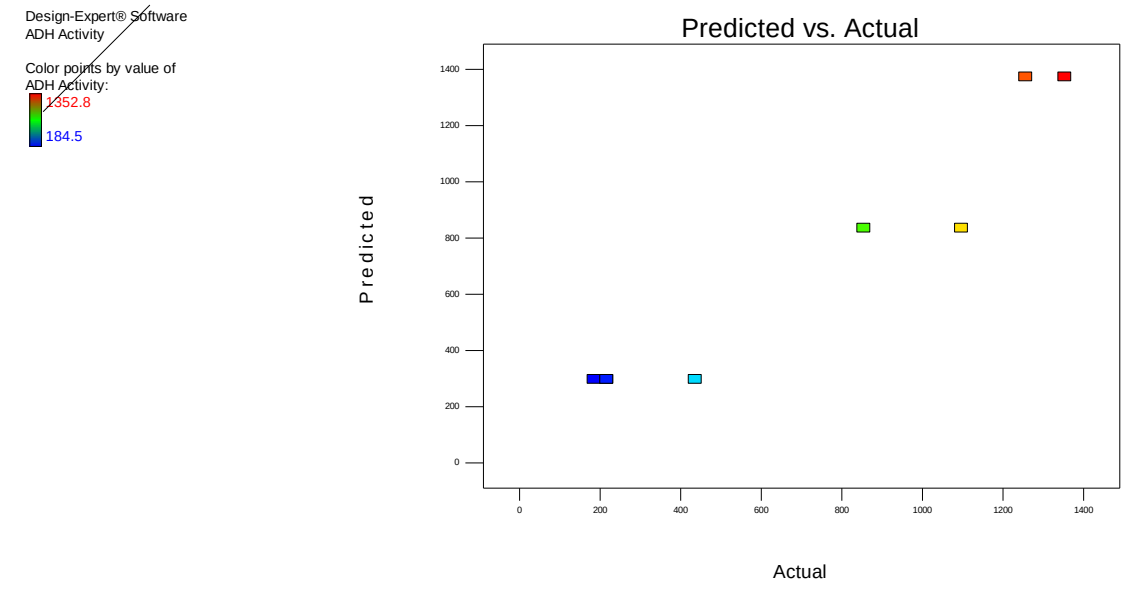


Figure 5.12 – Predicted vs actual plot for ADH activity

The predicted vs actual plot, Figure 5.13, showed a much-improved resolution within the model but this could be attributed to the higher number of levels within the glucose category when compared to lactose rather than an improvement in the model. The One-Factor plot shown here as Figure 5.13 did indicate that the model fit was improved, a much smaller 95% confidence interval.

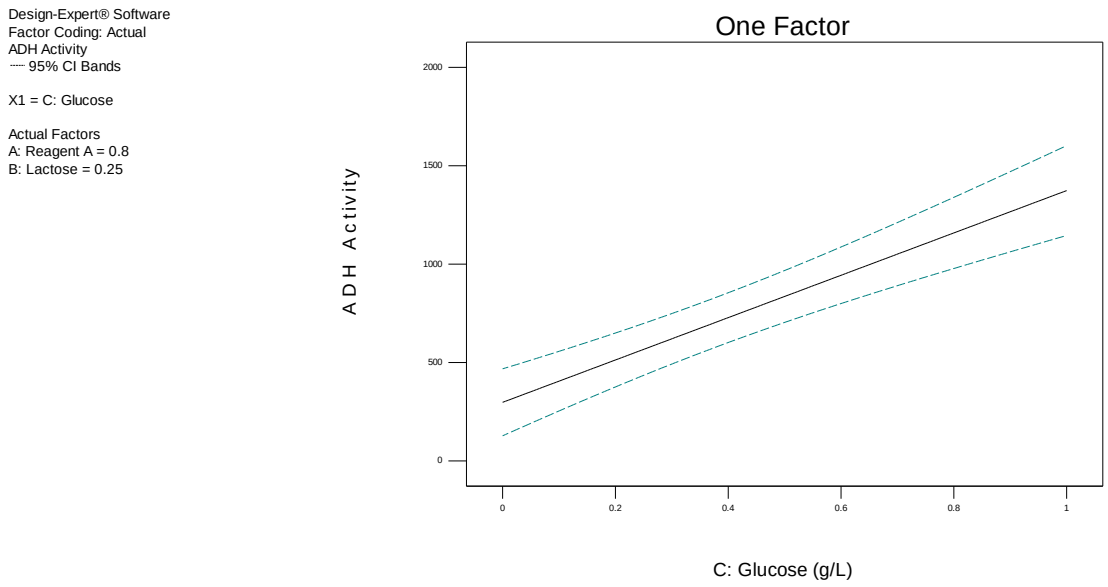


Figure 5.13 – One factor plot showing the effect of glucose concentration on ADH activity

Optimisation

Once the responses were modelled individually, Design Expert 9 ™ was employed to optimise the output. As there were no process specific criteria to meet, no threshold levels were set, and the objectives for optimisation were instead to minimise OD_{600} and maximise ADH activity. This would find the conditions which produced the most productive cells, rather than simply maximising cell number. If this process was scaled up this would allow the fermentation to be performed in a smaller bioreactor while maintaining product yield, and thus maximising profit as smaller bioreactors would be cheaper to operate and run. It would also minimise the size of the downstream processing capability required, as smaller fermentation broth volumes would be present. The optimal solution for this goal is shown in Figure 5.14, suggesting that lactose was to be minimised, and glucose was to be maximised. Only lactose and glucose levels could be optimised, as Reagent A was not shown to impact any of the responses measured within this study. This was a predictable output for optimisation, as lactose had been shown to positively influence cell growth related responses (OD_{600}, pH, Acetate and glucose consumption), and glucose was shown to maximise ADH activity (Figure 5.15).

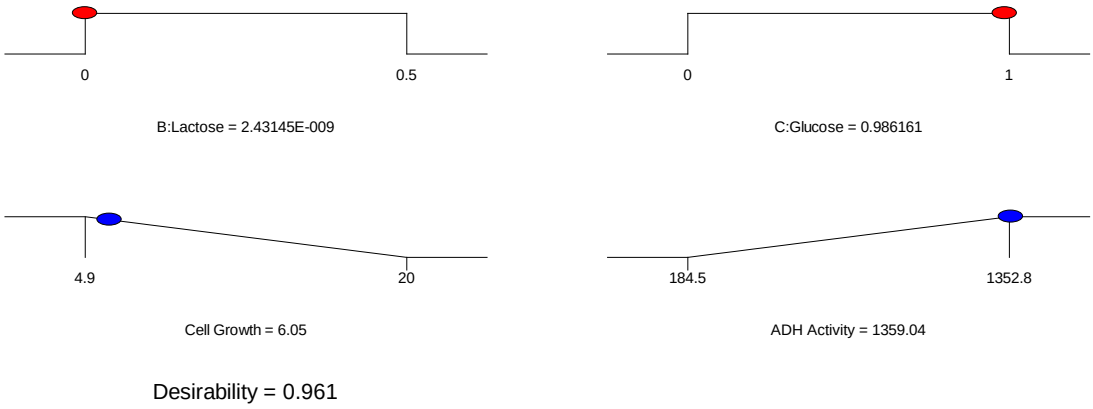


Figure 5.14 – Optimisation plot showing the optimal solution determined by DE9

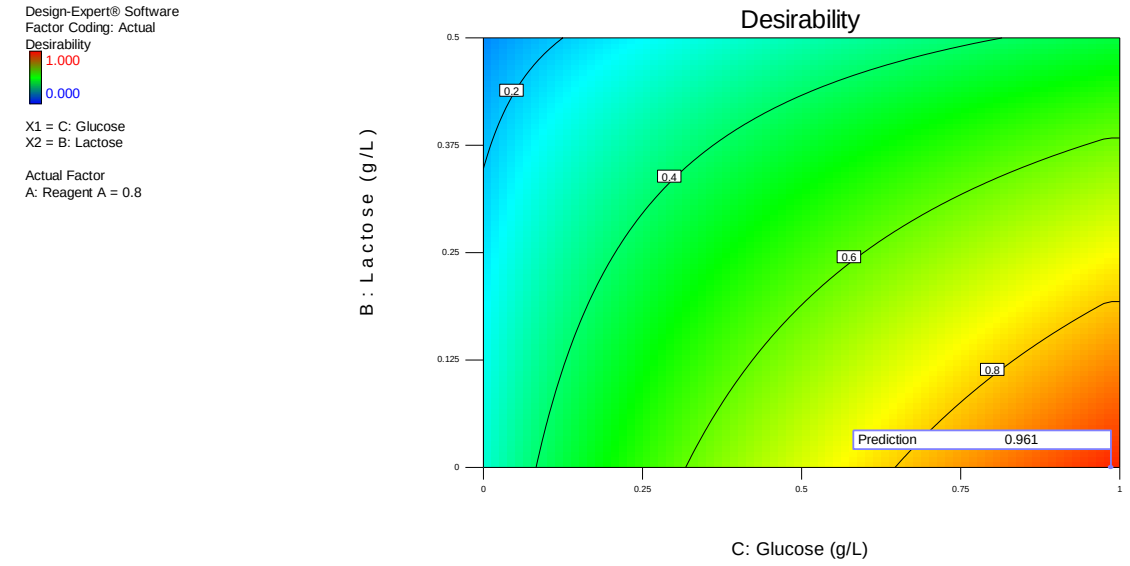


Figure 5.15 – Response surface for the optimisation work performed in DE9

5.3.4 Analysis Methodology Comparison

The original work selected experiment G as the optimum. The results were time consuming to analyse, reportedly taking hours of consideration compared to the single hour required to complete the IAT, and no structured approach was employed. The results from the original analysis were confirmed by the IAT when seeking to maximise ADH activity. However, when the

IAT was completed, lactose was shown to be linked only to cell growth rather than production, indicating the understanding of the results had been enhanced by employing the IAT. Donovan (1996) discussed the role of lactose in growth being linked to low levels of glucose, which would suggest that the glucose levels employed within the study are not sufficient. The induction system used in the study was auto-induction, which relies on the metabolic shift (Studier, 2005) from glucose to lactose metabolism. Once this shift occurs, the energy within the cell will be channelled into protein production, rather than cell growth (Studier, 2005). Therefore, low levels of glucose will cause this shift to happen earlier in the fermentation, resulting in fewer cells capable of producing ADH. This reduced number of cells is the likely cause of the reduced levels of ADH activity.

The IAT tool has been shown through this study to be applicable to a fermentation to enhance process understanding and offer a robust structured manner for data analysis. The visual nature of the tool make it more user friendly, and especially useful in a team including non-experts. It is important to note that the tool is not intended to replace alternative analysis methods. The comparison to Design Expert 9 ™ results shows the importance of such analyses, and that the two can be complementary to each other. The IAT attributed the importance of lactose to cell growth, which had not been previously identified by the original study. The quantitative analysis in Design Expert 9 ™ identified glucose as being key to maximising ADH activity, reinforcing the previous qualitative analysis. This would have made experiments G and H the optima if this was the only goal for optimisation. When optimisation was performed from a whole process perspective, i.e. minimising downstream recovery while maximising product, the models predicted that a low level of lactose and high level of glucose would be optimal. These experiments were not performed and so it is not possible to infer the model accuracy from the data available. This was the same conclusion as the IAT would have generated under the same optimisation criteria, further reinforcing the value of the tool.

5.4 Summary

This chapter examined the value of three analysis methods for a set of experiments: unstructured/qualitative, structured/semi-quantitative and Design of Experiments. The qualitative analysis was limited in that it could select the best output of the experiments performed, but was unable to infer relationships or attribute the results to an individual factor. It was time consuming to perform and was not structured. The semi-quantitative analysis using the IAT could not only select the higher output for the experiments performed, but could additionally infer which parameter was controlling which aspect of the output. This led to the conclusion that further experimentation would be required to optimise the fermentation in a whole-process manner as low levels of lactose were identified as minimising the cell growth but high levels of glucose would maximise the ADH activity. This is the same conclusion derived by the DoE, but with one minor difference. The IAT results indicated that Reagent A did influence the fermentation, and needed to be minimised to enhance the ADH activity. However this relationship was not identified through Design Expert 9 ™. This could either indicate that the IAT was attributing significance to non-significant differences, an inherent risk of using methods with a qualitative nature. However, it could also indicate that the use of process knowledge employed when completing an IAT has identified a relationship that, while statistically insignificant at the range selected, could show greater influence over the fermentation if investigated across a larger range.

This work demonstrated the clear benefit of employing a semi-quantitative analysis method such as the IAT. The tool drew conclusions superior to the original qualitative analysis as it could attribute the impact of lactose being primarily on growth rather than production which was not discussed within the original research. However as previously discussed there would be further work required to verify the findings of the tool experimentally. From a tool development perspective, this study showed that the IAT in this form is able to be applied to a fermentation dataset and draw valuable conclusions. The next stage of tool development would be to test the IAT on a dataset incorporating downstream data, which could either be purely downstream data or ideally incorporating outputs from both upstream and downstream unit operations.

The next stage of this work (Chapter 6) moves on to consider the application of the IAT to a downstream dataset, following from the successful application to upstream data in this chapter. From here, the effect of the weightings was investigated using sensitivity analysis (Chapter 7) to ascertain the reliability of the tool outputs.

Chapter 6 Downstream Testing of the Interaction Analysis Table (IAT)

6.1 Introduction

The previous chapter discussed the application of the IAT tool to a fermentation, as the original purpose that the tool was developed for and which is the key difference between chemical and biochemical processes. The successful application to upstream processing is a good indicator that application to bioprocessing in general will be successful. However, to conclusively demonstrate applicability within bioprocessing, a downstream dataset was used to apply the IAT, to establish the applicability to a bioprocess as a whole rather than only certain parts of the process.

Downstream processing encompasses the cell lysis and subsequent purification processes to separate the product from impurities generated during the upstream phase of the process. This chapter will focus on Britest tool application to the cell lysis stage of the process.

As in previous chapters, public datasets were used for the testing, in this case from work performed by Glauche *et al.* (2016). This overcomes any confidentiality concerns, and also ensures that the dataset is from an early stage of bioprocess development, which would be difficult to source from within the Britest membership. The data relates to a. *E. coli* cell lysis experiment, where the group sought to optimise the lysis buffer used for the process. The buffer system was selected as the group work on developing high-throughput platforms, and the buffer system for cell lysis would enable automation to be employed, increasing throughput and reducing labour requirements.

The lysis buffer components were Lysozyme, Polymyxin B, Triton-X and EDTA. Lysozyme is an enzyme which breaks peptidoglycan bonds within the bacterial cell wall.

Derived originally from hen egg white, the mode of action is well documented (Weibull, 1953; Chassy and Giuffrida, 1980), and it has been used for cell lysis processes for many years. It is a staple component of "off the shelf" kits for cell lysis, and so is widely applied in academic research and small scale industrial research. Polymyxin B also attacks the cell wall of bacteria to cause lysis, however it is only effective in gram-negative bacteria such as *E. coli,* as the thicker membrane in gram-positive bacteria inhibits activity (Newton, 1956). In gram-negative bacteria it binds to a negatively charged site in the lipopolysaccharide layer, destabilising the outer membrane (Zavascki *et al.*, 2007). Triton-X is a detergent which has a range of protein denaturing applications, most commonly in SDS-PAGE gels and as part of the extraction buffer in DNA extraction kits (Van Tongeren *et al.*, 2011; Lever *et al.*, 2015). Ethylenediaminetetraacetic acid (EDTA) is well known for its ability to sequester metal ions. Within the context of protein extraction it is generally included to inhibit cation-dependant proteases (Wu and Tai, 2004), and so in this study it was included to preserve the protein and activity, rather than enhance cell lysis. Within the original study, three of the four reagents in the buffer were included to directly lyse the *E. coli* cells, and EDTA was included to ensure the extracted protein was not degraded. It is expected that the three lysis reagents would work synergistically, and that the maximum amount of lysis would be seen by including these multiple lysis agents with differing modes of action. This would ensure the highest amount of protein were released from the cells, and the EDTA would preserve the protein for analysis.

The data originates from two DoE campaigns, performed in sequence. The initial DoE identified that the boundaries set for each component were sub-optimal and would not give the highest yields of protein possible from the *E. coli.* The second DoE did find the optimal conditions, with the experimental boundaries set from the information gained

from the first set of experiments. This study aimed to ascertain whether the first experiments could have been reduced in number, or indeed avoided completely, if the IAT had been employed by the group prior to commencing the experimental work. The group are well established in working on *E. coli* based processes, and so a certain level of process knowledge and understanding is assumed. While the IAT has been investigated in this research project for its potential to add value during early process development, it would not be intended to be applied by people with no knowledge of the process whatsoever. This is consistent with the other Britest tools, where the quality of knowledge extracted and process understanding gained will depend on the experience and knowledge of the participants. This is explored further in relation to the IAT specifically in Chapter 7 of this thesis through sensitivity analysis.

This study focussed on the IAT application to the dataset, assuming a basic level of process understanding within the original research group. As the experimental work had already been performed, and the results published, it was not possible to ascertain the level of process understanding within the group at the time when the IAT would have been applied. In light of this, the study focussed on investigating the number of experiments required to complete the IAT using the limits set within the first DoE. The design would have been constructed by the group anticipating that further DoE were unlikely to be required, as they expected the optimum to exist between the original limits. The screening design that the IAT offers relies on an understanding of the interactions between each parameter and outcome, although identifying a lack of understanding can be beneficial. It is likely that within a group of process experts, the process knowledge assumed to be present within the group would likely eliminate some of these experiments. The reduction of experimental burden during early process development stages would streamline the development process with respect to both cost and time, which are the two

most important factors in industrial process development. The potential for significant cost and time savings through the reduction of experimental burden would provide a robust business case for bioprocessing companies to adopt the Britest toolkit, including the IAT.

6.2 Methods

6.2.1 IAT

Completing the IAT is performed first by setting up a table with the components (EDTA, Lysozyme, Polymyxin-B and Triton-X) listed in the first column. The next 2 columns correspond to the outcomes. In this case the outcomes were soluble protein and β-galactosidase. Soluble protein is the amount of soluble protein measured in the samples. The β-galactosidase represents enzyme activity. Weightings were not included within this IAT. As in previous work, the study was academic in context and it was not known what the associated business benefit would be if scale up work was performed. With respect to the technical benefit, the outcomes were approximately equal, therefore the weightings would have been identical and would not have provided extra information to the analysis. The cells within the IAT are filled with a shape indicating the nature of the relationship between the component and outcome. For the purpose of this study all relationships were assumed to be linear, though it is recognised that this is not always the case, and a non-linear relationship is more common with an optimum value between the upper and lower bounds. The colour yellow is then used to highlight the level at which the outcome is highest, to allow for ease of understanding. For example, if a high amount of EDTA resulted in a high amount of soluble protein, the area on the right hand side on the cell would be shaded yellow (as shown in the generic example in Figure 6.1). In this case both outcomes were sought to be maximised, so yellow was the only colour required. Blue can

be used to shade optimal operating areas in the case of undesirable outcomes such as impurities.

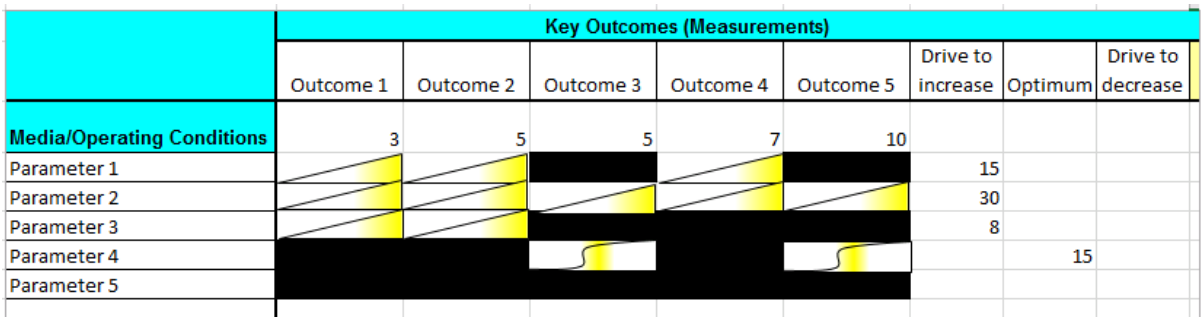


Figure 6.1 - Generic IAT to show how the tool would usually appear.

Within this research three IATs were constructed. These represented three stages of process development, and are outlined in Figure 6.2. In the first two instances the results from the two DoE datasets were used to complete the IAT, simulating a later stage process development study where data is abundant. In the third IAT, a more data lean environment was simulated, to investigate the potential for the IAT to add value in an early stage of process development.

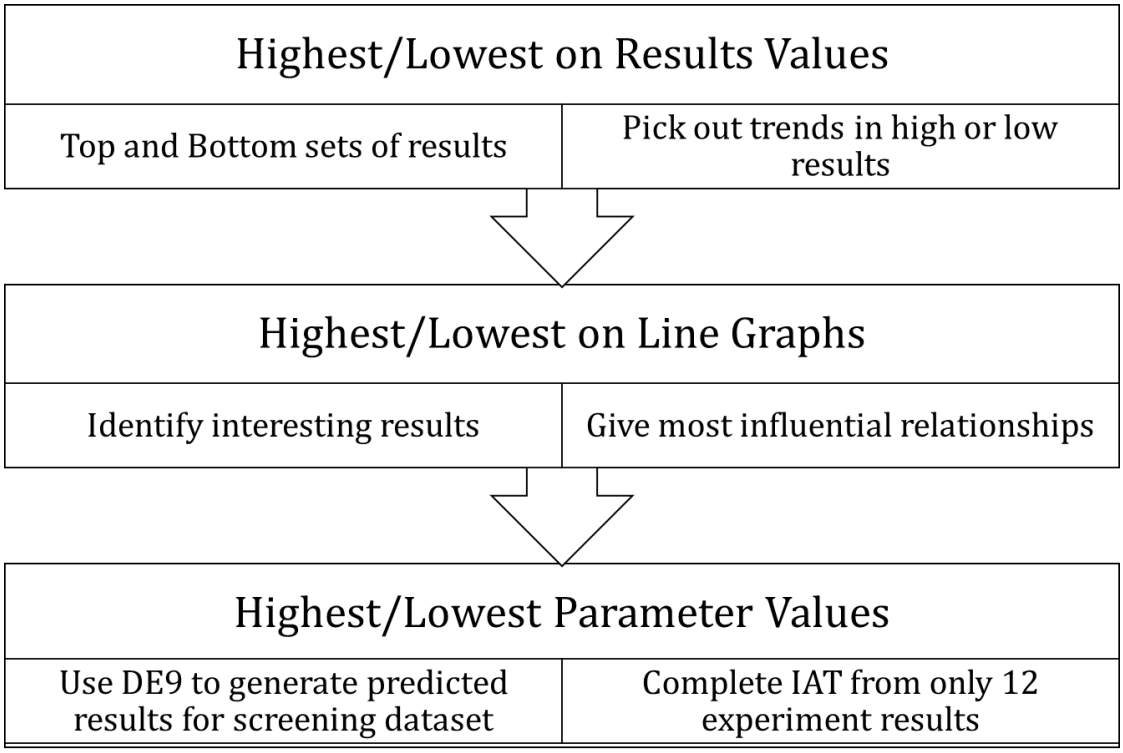


Figure 6.2 - Showing the flow of work for this section of the research.

The first IAT sorted the results by high to low, for each outcome individually. Cut off points were used to define which experiments yielded a "high" or "low" result. The experiments yielding the highest and lowest results for each outcome were examined visually for any obvious patterns or consistencies (e.g. all low EDTA values/all high lysozyme values). With respect to soluble protein, results over 1.4 were selected as being "high", and results lower than 0.5 were selected as being "low". With respect to β-galactosidase, results over 0.3 were selected as being "high", and results below 0.01 were selected as being "low". The conditions which gave rise to these high and low results were then examined to ascertain whether there were any trends within these sets of results. This reflected a situation in which the IAT was less likely to be employed in isolation, as statistical analysis techniques would likely be used to complement the process understanding generated by qualitative tools such as the IAT.

The second IAT method again used the results from both complete datasets, amalgamated together. Both sets of results were represented graphically using line graphs for each individual parameter, and obviously high or low results examined and their associated conditions used to complete an IAT. This meant fewer results were used to construct the IAT than in the first IAT, but the potential to conclude a trend with an anomalous result was increased.

In the third IAT method a data lean environment was simulated. This reflected the situation in which the IAT would likely provide maximum benefit, where process knowledge may exist but a complete DoE dataset may not yet be available. The IAT was initially constructed in a blank format, and a list of desired experiments which would be performed as the next stage of the investigation were constructed. No prior knowledge of the relationships between parameters and outcomes was assumed. The list of desired experiments is included below as Table 6.1. Each parameter would be investigated at a high and low point, and at a mid-point when not at an extreme value. The exception to this was lysozyme, as this is well understood and it was not unreasonable to assume this relationship would be well known by the experts involved in a Britest study.

Table 6.1 - The screening design used in this research to generate the IAT. Results were simulated using DE9 software, as none of these experiments were performed in the original design.

EDTA (mM)	Polymyxin B (µM)	Triton X (%)	Lysozyme (U/ml)
0.5	25	1.367	9000
10	25	1.367	9000
5.25	0.1	1.367	9000
5.25	50	1.367	9000
5.25	25	0.1	9000
5.25	25	2	9000
0.5	25	1.367	300
10	25	1.367	300
5.25	0.1	1.367	300
5.25	50	1.367	300
5.25	25	0.1	300
5.25	25	2	300

In the case of each of these experiments, the results were not already available as part of the datasets generated by Glauche *et al.* (2016). To overcome this limitation in results, Design Expert 9^{TM} (DE9) was employed as a tool to construct a model of the data and predict what the results would likely have been if the experiments had been performed.

6.2.2 Design of Experiments

Design of Experiment analysis was performed in Design Expert 9 ^{TM}, to facilitate completion of the IAT under data lean conditions as described above. Additionally, the analysis was used to determine optimal conditions for the process to compare to the three IATs generated from the datasets. This would ascertain whether the completion of the

minimal screening experiments required to complete the IAT could have led the authors to the same conclusion as the more complex and time/labour intensive initial DoE study.

A "Historical" Design of Experiments (DoE) analysis approach was employed for this study, due to the experimental design and results already being available from previous analyses. DE9 was used to generate a model which could predict the values for both soluble protein and β-galactosidase under experimental conditions which were not original experimental design points. A response surface was used, as all factors were Numeric rather than Categoric. This analysis was restricted to the first DoE dataset, which contained 91 experiments, and not the subsequent second DoE dataset where the experimental boundaries were revised based on the output of the initial 91 experiments.

6.3 Results and Discussion

For the initial IAT the outputs from the original study were used, combining the results from both datasets. This was to ascertain whether the IAT could add value to a study if data availability was not restricted, unlike in the above example where only 12 experimental points were selected. Working in a data lean environment is an important potential benefit of the IAT, and this would be a key phase of application. However it was recognised that an organisation may have extensive data from previous experiments or process runs which they want to use to derive process understanding, and may not have the in-house statistical experience to generate a meaningful analysis. Equally the statistical analysis may not deliver a meaningful output, if the system under investigation has not been adequately analysed. In this case, the IAT could be combined with plant knowledge from the experts on site to derive some value, though value can be added through complementary statistical analysis. In this instance, the IAT would not replace

statistical analysis, but could aid process understanding in a resource restricted environment.

The results from both DoE campaigns were sorted into order according to the results for both β-galactosidase and amount of soluble protein. The trends from the highest and lowest results for each criteria were determined by eye, and this was used to complete an IAT (Figure 6.3).

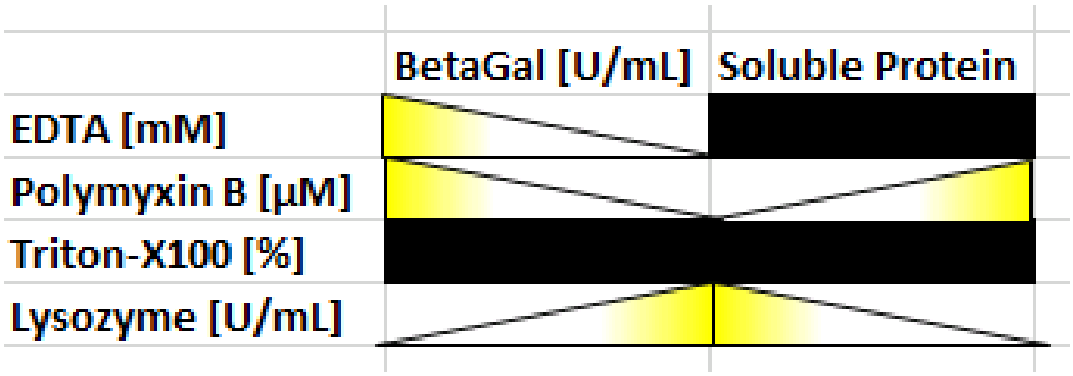


Figure 6.3- IAT using the best and worst results obtained within the datasets. The result of stage 1 of the research.

This analysis concluded different optima to the original authors, and the screening IAT approach outlined above. This analysis suggested that for a high β-galactosidase output EDTA would need to be minimised, Polymyxin B would need to be minimised, and lysozyme would need to be maximised. For maximum soluble protein production Polymyxin B would need to be maximised and lysozyme minimised. Triton-X was not shown to be impacting the results. While there was some overlap, particularly in the desire to minimise EDTA and maximise lysozyme, the lack of complete agreement would give a cause for concern if the IAT was intended to be used on large datasets. It would appear that the complexity of the dataset, combined with the various interactions, would make IAT application benefit limited, and in fact if employed on a process dataset could lead to adverse process impacts.

The second method devised to employ the IAT on the large dataset was to use line graphs (included in electronic Appendix B) to show the results for each parameter, and any pattern which was evident was used to complete the IAT. For example if a result was particularly high the experimental conditions associated with that point would be highlighted as being advantageous in the IAT. The resulting IAT is shown in Figure 6.5. The various graphs produced from the dataset are included as Appendix B, Figure 6.4 comparing the EDTA concentration to the soluble protein yield is included as an illustrative example.

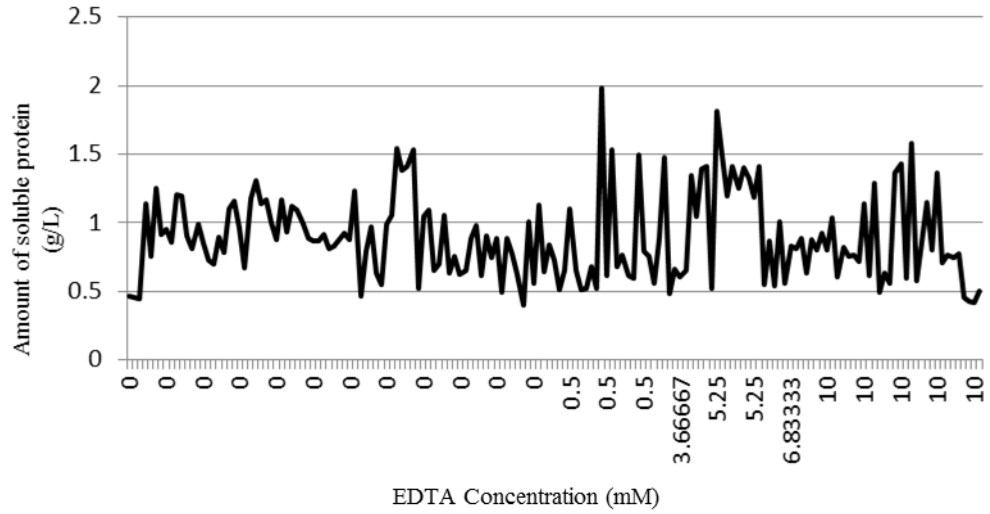


Figure 6.4 - Graph showing the EDTA concentration and the associated amounts of soluble protein.

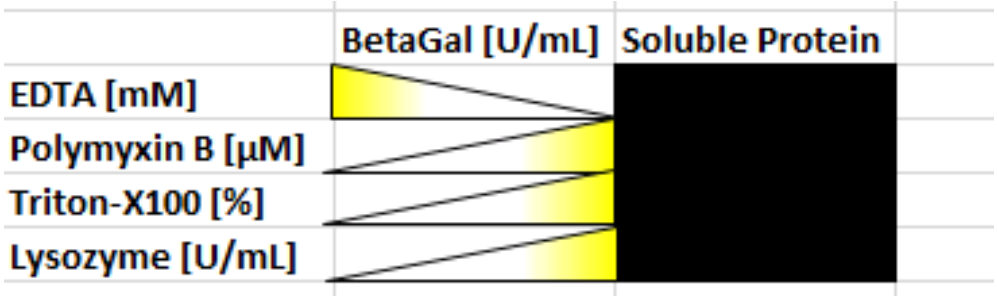


Figure 6.5 - IAT using the obviously best and worst results from line graphs of the datasets. The result of stage 2 of the research.

In this case, the results showed there was no correlation between any of the factors and soluble protein output, which the detailed data analysis shows to be incorrect. The

minimisation of EDTA remains consistent, with all other factors to be maximised. This is more consistent with the detailed statistical analysis, but the lack of any effect shown between the reagents and soluble protein levels would still suggest a limitation in the IAT application to large datasets.

To this point the IAT has been discussed in relation to scenarios where data is available. This is analogous to the use of Six Sigma tools (Harry, 1998; Pande *et al.*, 2000; Pyzdek and Keller, 2014) for continuous process improvement. Studies of this nature are used widely in industry to examine established processes. This could be to reduce process time, reduce raw material costs, improve safety, or to aid the decision making process (Harry, 1998; Eldridge *et al.*, 2006; Kumar and Sosnoski, 2009; Yang and Hsieh, 2009). Evaluation of process options is also a frequent occurrence in Contract Manufacturing/Research Organisations (CMOs), where technology transfer is a regular activity. From the work performed to this point in the research, the IAT has been shown to be applicable to scenarios where data and process knowledge are available, however the above discussed limitations would suggest that combining tools of this type with complementary statistical analysis could give an added benefit.

Following this established process investigation, the IAT was examined for applicability to a process where data availability was limited. This mirrored more closely the early stage of process development, where the IAT was anticipated to add significant value. From the initial list of screening experiments which would be desired, those for which results were not available within the two DoE campaign datasets were simulated using DE9. Only the results from the first DoE campaign dataset were used for the DE9 analysis. It is worth noting that in the course of the original work the authors found ETDA to be irrelevant, and so the second DoE study was performed without this as a factor, in addition to moving the design space based on the original analysis. The aim of

this study was to determine whether the IAT could have been used to create a better design from the beginning of the research, and so it was more appropriate to consider the first design in isolation as this was created with the same information that the IAT would have been created with.

It was anticipated that the lysozyme would have a significant effect on the outcome of the lysis. In light of this, the results were considered with respect to high and low levels of lysozyme initially, and it was the effect of the other 3 reagents which were investigated. Lysozyme is well understood and has a well characterised mode of action, meaning it was deemed reasonable to assume this positive linear correlation would be generally accepted within a Britest study.

The IAT would require the effect of EDTA, Polymyxin B and Triton-X to be understood independent of each other. The interactions between reagents are not insignificant, but this is where the IAT is distinguished from more complex data analysis methods. It has been shown to be able to be applied when little data is available, and given the significant amounts of data required for interaction analysis, it was considered that a basic screening dataset would be suitable for completing the IAT in this instance.

The results from the predictive experiments are shown in Table 6.2 and Table 6.3, with an indication of the results seen with respect to both β-galactosidase and Soluble Protein (activity and solubility). The resulting IAT is shown as Figure 6.6.

Table 6.2 - Results from the IAT screening experiments at high lysozyme concentrations.

High Lysozyme					
EDTA (mM)	Polymyxin B (µM)	Triton X (%)	Lysozyme (U/ml)	**β-galactosidase (U/ml)**	**Soluble Protein (g/L)**
0.5	25	1.367	9000	Med/High	Med
10	25	1.367	9000	Med	Med/Low
5.25	0.1	1.367	9000	Low	Low
5.25	50	1.367	9000	Med	Med
5.25	25	0.1	9000	Low	Low
5.25	25	2	9000	High	High

Table 6.3 - Results from the IAT screening experiments at low lysozyme concentrations.

Low Lysozyme					
EDTA (mM)	Polymyxin B (µM)	Triton X (%)	Lysozyme (U/ml)	**β-galactosidase (U/ml)**	**Soluble Protein (g/L)**
0.5	25	1.367	300	Med	High
10	25	1.367	300	Low	Med
5.25	0.1	1.367	300	Low	Low
5.25	50	1.367	300	Low	Low
5.25	25	0.1	300	Low	Low
5.25	25	2	300	Med	High

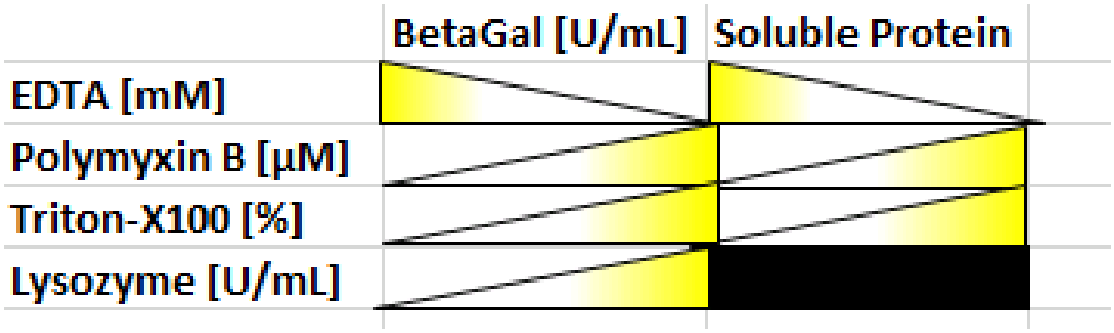


Figure 6.6 - IAT generated using the results from the IAT screening experiments. The result of stage 3 of the research.

As predicted by the IAT, the optimal results were found at high levels of lysozyme, when the cells would be lysed most effectively (e.g. Figure 6.7 and Figure 6.8). The presence of a set of results where high levels of activity and soluble protein are present would indicate that an optimal solution can be obtained.

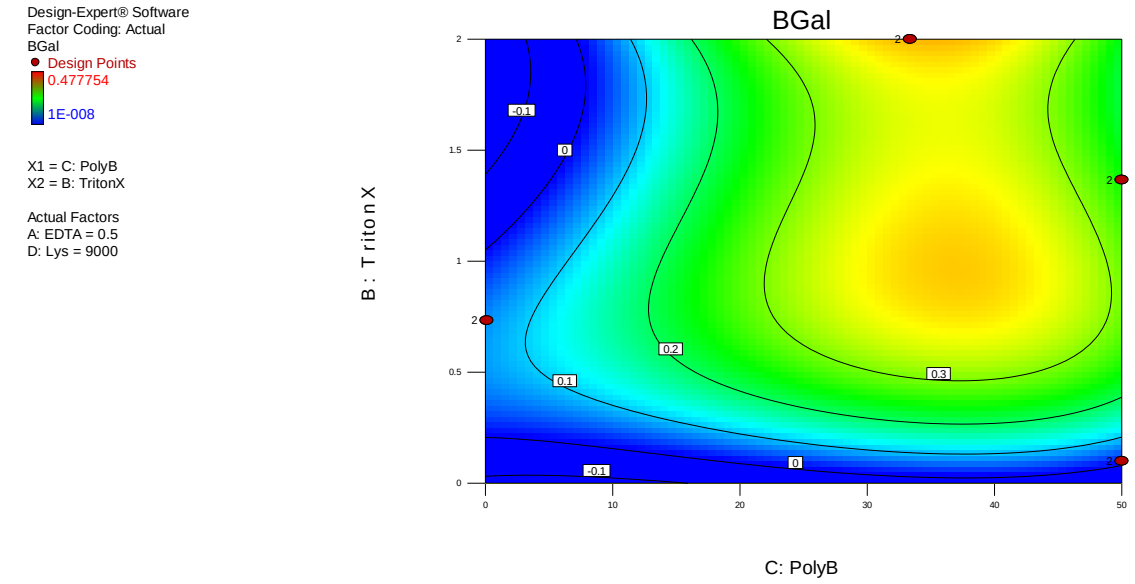


Figure 6.7 - Beta-galactosidase assay result at low levels of EDTA and high levels of lysozyme in relation to Triton-X and Polymyxin B concentrations. Generated using DE9.

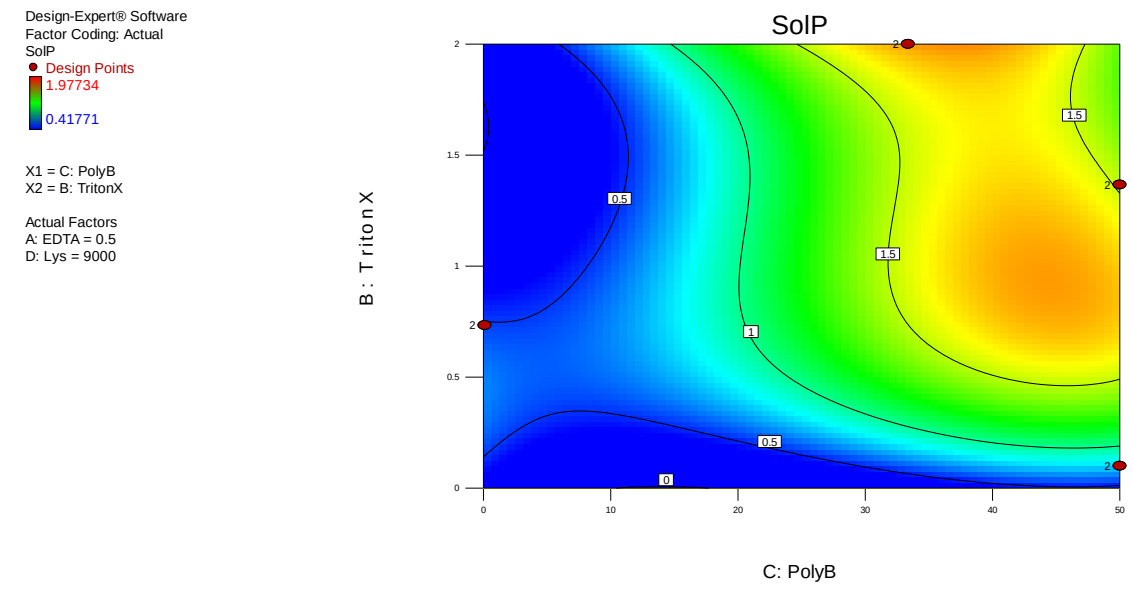


Figure 6.8 - Soluble Protein result at low levels of EDTA and high levels of lysozyme in relation to Triton-X and Polymyxin B concentrations. Generated using DE9.

Using this scheme for completing the IAT would have required 12 experiments. This would be the minimum required, and it is likely that a mid-point experiment for each variable would be desirable. Even with the inclusion of a mid-point, the same conclusion regarding the range at which to perform the experiments would have been obtained. With a mid-point the number of experiments would rise from 12 to 18, which is still a dramatic reduction from the 91 experiment dataset that was used for the original analysis.

As DE9 was being employed as a tool for this research to simulate the results required to complete the set of IAT screening experiments, optimisation using this model was also carried out to ascertain whether this would give a different result to both the IAT's generated, and the original author conclusions. When the output was optimised for maximum soluble protein and β-galactosidase (Figure 6.9 and Figure 6.10) there was a solution which gave higher levels of both outputs than the solution presented previously. This would suggest that the DoE was likely focussed on a sub-optimal design space. If the experiments which were suggested by the IAT had been performed, the simulated results

would indicate that higher values for the reagents would have been selected. This was exactly what Glauche *et al.* (2016) decided to pursue for the second DoE they performed on this process, although they omitted EDTA as a factor.

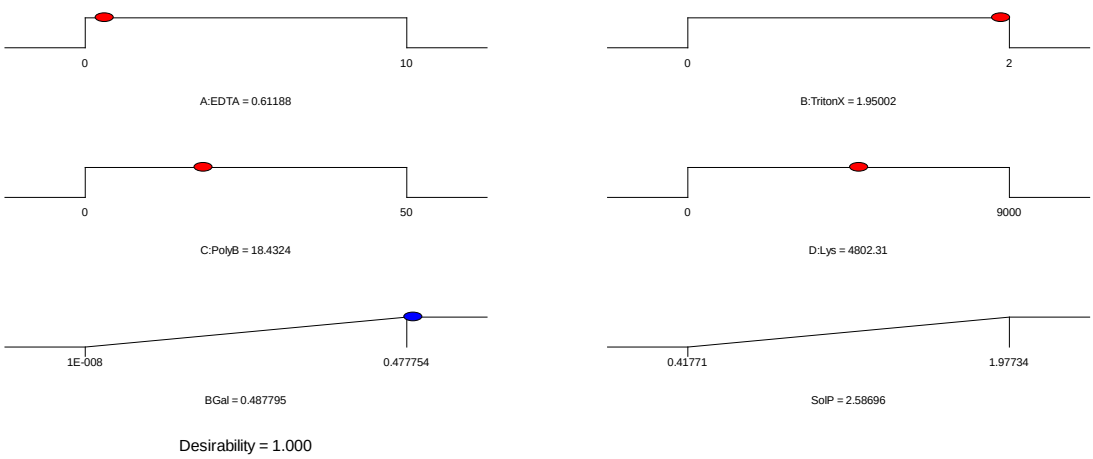


Figure 6.9 - Optimal solution for maximising soluble protein and beta-galactosidase using DE9.

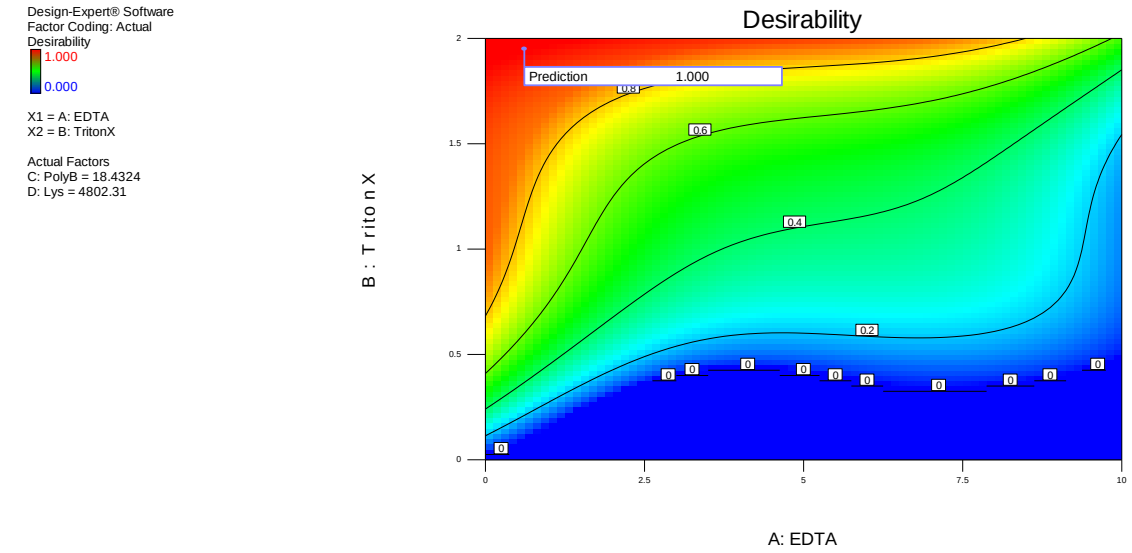


Figure 6.10 - Response surface showing the relationship between EDTA and Triton X at mid-levels of Polymyxin B and Lysozyme. Created using DE9.

The original research presented by Glauche *et al.* (2016) concluded that the optimum values for cell lysis were obtained at high levels of Triton-X and lysozyme, with medium

levels of Polymyxin B and low levels of EDTA. This was relatively consistent with the IAT suggested optima, with the exception of Polymyxin B. However the conclusions were drawn using both DoE campaign results, rather than the initial campaign as in the IAT. Therefore it is possible than an optimum value of Polymyxin B exists above the limits for the first DoE campaign but at the medium point in the second campaign.

EDTA was not shown in the original research to be influencing the output of the experiments, leading to the removal of this factor for the second DoE campaign. In the DoE model generated for this research EDTA was shown to be adversely impacting the β-galactosidase results. In the optimisation from this model it was suggested that the levels of EDTA should be minimised, which is consistent with the original approach. Polymyxin B was shown in this work to be required at a high level, which was consistent with the original research. The second DoE included in the original work increased the amount of Polymyxin B used, which would have been the decision taken had the authors used the proposed screening approach to build an IAT. Triton X had a significant influence on the output of the experiments, and again this was consistent with the original research. As with Polymyxin B this was increased in the second DoE, which would have been the case had the IAT been implemented in the original work. Lysozyme is a well understood enzyme used broadly to lyse *E. coli* cells at a range of scales and for a range of purposes. In light of this, it was unsurprising that better results were seen at higher concentrations of lysozyme. It is anticipated that experiments to demonstrate this may have been carried out for completeness had the original authors opted for the proposed screening IAT approach, however given the extensive history of application this may have been considered unnecessary for an early stage study.

Glauche *et al.* (2016) completed 91 experiments to obtain this information, to allow the authors to determine the optimal design space required for a more detailed second DoE.

The proposed screening IAT approach outlined within this work would suggest that the authors could have completed only 12 screening experiments and combined the output with an IAT focussed Britest study to conclude the same design space was required. This would have saved significant resource and time, not only in the experimental set up and clean up, but also in the data analysis. One author spent a significant amount of time (>1 day) analysing the results, before presenting their analysis to the group for discussion. Had an IAT been employed it is likely that the analysis and redesigned experimental space could have been achieved within half a day. The reduction in number of experiments required would also have saved significant time. While the equipment employed is a high-throughput system and reductions in time may be incremental, they are accompanied by a reduction in the associated costs (e.g. set up, analysis and materials) which can be a significant factor when conducting research.

It is anticipated that the proposed screening IAT approach discussed above would mirror the approach taken by the original authors, had they been aware of the IAT at the point of conducting their original research. However, this work also considered whether the IAT could have been applied by the original authors if they had applied the IAT after their first DoE. The Britest approach is commonly used once development has started, especially when problems are encountered, and so it was important to understand whether the IAT had the potential to bring benefit in later stage studies if needed.

It is likely that in reality, the screening dataset could be replaced (at least partially, if not in whole) by knowledge from the team. For the purpose of this study, as the experiments had already been performed, it would not have been possible to perform a Britest study to create an IAT with the team's knowledge before the experimental data had been understood. The team had already analysed significant amounts of data from the experiments and so their knowledge level would be greatly enhanced compared to that of

when the study was first developed. This would have made the Britest study not a true reflection of the knowledge level at an early stage of development, and so the DE9 analysis was employed to fill this gap.

Tools for process understanding where data is already available are generally statistical in nature, and while these have been shown to add significant value to a process the implications if the analysis is incorrect can be significant. Combining these statistical tools with qualitative knowledge tools, such as those offered by Britest, to enhance the value derived from statistical tools can only aid in delivering value to organisations. At earlier stages in process development where data on a process is scarce, qualitative tools could been employed to allow options to be explored without significant resource requirements. This is one area where the IAT would be expected to add value to an organisation, as it can be used with minimal data, and can be tailored to the users' requirements.

Six Sigma provides value to a variety of industries (Koning *et al.*, 2006; Saleh *et al.*, 2007; Junker *et al.*, 2011; Siddh *et al.*, 2014; Antony *et al.*, 2016), in addition to a range of other knowledge management techniques discussed in Chapter 1. Design of Experiments is also engaged as an approach in a variety of specialities, ranging from drug discovery to motor manufacturing (Tye, 2004; Franceschini and Macchietto, 2008; Sakkas *et al.*, 2010; Ford, 2011; Kumar *et al.*, 2014), and is particularly useful in early process design. While the value of an analysis approach which is statistical in nature which is capable of modelling complex system behaviour is clear, the ability of users to successfully apply the approach can be less successful. There are a broad range of design options available for studies (Montgomery, 1991), and the deciding on the design to select is not the only aspect which can cause difficulty. The selection of the appropriate boundaries within which to base the design is critical to the success of the experimental

campaign, but relies on a level of expertise which can be lacking, especially within early process design stages.

In cases where data is not available to support a process decision, especially in commercial environments where costs can be tightly controlled, tools such as the IAT can add value to users, particularly when combined with other complementary approaches. Not only can the IAT facilitate effective communication and provide a system for knowledge capture, but also requires the user to justify design parameter selection. This can be key for avoiding poor experimental design, which can be an expensive mistake for organisations in terms of wasted resource and costly errors. The combination of qualitative tools, such as some of those contained within the Six Sigma toolkit, with Design of Experiments approaches can ensure organisations have the highest chance of success when implementing a DoE (Conklin, 2004; Raisinghani *et al.*, 2005). Six Sigma tools can be employed to aid a user in applying DoE designs (García-López *et al.*, 2015; Gupta *et al.*, 2016), and the IAT has been shown through this research to be a viable alternative, which was specifically designed with bioprocessing in mind.

6.4 Summary

This study considered the application of the IAT tool to a downstream process, cell lysis of *E. coli.* Chapter 5 discussed the successful application of the IAT to an upstream process, and this chapter followed on to ascertain whether the tool could be applied to other stages of the process, and potentially to the whole process if sufficient process knowledge existed within a team.

The IAT was applied using two methodologies, to the whole dataset available in the original study. The application within these examples was less successful, though it is recognised that this could be aided by using process knowledge combined with the data

available rather than relying on data alone as in this study. Previously the IAT was developed for an antibiotic fermentation where large amounts of data were available, and it aided in focussing the users on the important interactions within the fermentation. However, it is clear that the IAT can add significant value in early process development, where knowledge is relied on for effective experimental design.

Following this, the study considered how the IAT would be applied within a Britest study. It is anticipated that this would be at an early stage of investigation, where little to no data existed. Britest tools have added value in early process design (Sharratt *et al.*, 2003), and so the IAT was applied to construct a desired screening dataset to reflect a common data limited scenario. Some of the experimental points required for this were not directly available, and so a historic DoE approach was employed to predict what the values for these data points would have been. This approach is not without its limitations, but for the purposes of this research it was assumed that the approach would be sufficient to reflect a level of assumed process knowledge present within a Britest study.

In this IAT approach, the IAT was able to ascertain the same conclusions about optimal design space as the original research, and had it been applied the experimental requirements could have been dropped from 91 separate experiments to only 12, with the associated time and cost savings. This is clear evidence for the value of Britest application to process development in the early, data limited, stages. While the IAT analysis would lack the statistical element of DoE analysis, the same conclusion would have been drawn, and the statistical approach used from there to further investigate the design space would have ensured robustness within the process development. This is without considering the additional benefits of more efficient teamwork and communication, or the impact if multiple process units were considered, or multiple Britest tools employed.

Further work has been done to assess the IAT application to chromatography datasets. This utilised datasets originating from a biopharmaceutical company, and due to confidentiality restrictions the work may not be included as part of this thesis. The IAT was able to draw conclusions from the dataset, but separate IATs were required for each resin being investigated. This would suggest the IAT would be better suited to optimisation experiments rather than selecting resins/membranes, unless comparisons of running conditions were required.

In conclusion, the IAT is a powerful process understanding tool, which can be applied to both upstream and downstream processes. Deriving the most value from the IAT would be through application in very early process design/optimisation stages, when data is not available and process knowledge and understanding is the main driver for the work. This is consistent with the intention of the IAT, and also with the other Britest tools. Where complex statistical analysis is possible this would be the preferable option provided sufficient underpinning mechanistic understanding exists to ensure that the statistical analysis is properly set up. However at early stages where resources may be limited qualitative process understanding tools such as those presented in this thesis add significant value while minimising the resource required to add value. The application of the IAT to both upstream and downstream units is promising, and it is not inconceivable that the IAT could be applied to multiple unit operations or indeed a whole process. This would add significant value to a company, not only in terms of better process design but also in aiding in the development of a QbD approach to process design.

This chapter, combined with Chapter 5 focussing on upstream application, clearly demonstrates the value of applying the IAT in a data-lean, early process development stage of a bioprocess. The applicability to both upstream and downstream units means the tool shows promise for whole-process application, which would be a significant benefit

from a user perspective. It would provide a structured framework for considering the process design of a bioprocess, to aid intelligent experimental design. In addition to the benefit to the user, this provides a notable benefit for Britest with respect to attracting new members from within bioprocessing. However, the IAT is limited by the knowledge within the user group. If the users have little to no process understanding, the resulting IAT would be of no benefit. However, the process of identifying where process understanding is limited can itself be useful for an organisation, and this is a purpose for which the Britest tools have found broad applicability. The limitations around a lack of process understanding are particularly evident in relation to the weighting attribution system, and so having established applicability to both up and downstream processing units, investigation into this sensitivity was the next stage of the research (Chapter 7).

Chapter 7 Sensitivity Analysis within the Interaction Analysis Table (IAT)

7.1 Introduction

This research originated with an investigation into the applicability of the Britest toolkit to bioprocesses, simulated using SuperPro Designer (Chapter 3). This resulted in the development of the R2T2 tool, and a basic bio-suitable Britest toolkit ready for application to a range of industrial bioprocesses. From here, the tools were applied to academic datasets, and it was found that linking of the process inputs to process outputs would be a valuable addition to the toolkit. This took the form of the IAT tool, developed in industry in collaboration with AbbVie and Pfizer but not in a stage of development suitable for inclusion in the Britest toolkit. As previously discussed in detail (Chapter 4), the tool required significant redevelopment, before being applied successfully to both upstream and downstream data sets (Chapters 5 and 6).

The IAT was subsequently tested on academic datasets, in addition to industrial processes by Britest members, and was shown to enhance process understanding, while retaining its user friendly character. AbbVie gave positive feedback having employed the new IAT on their in-house bioprocesses, and feedback from the Britest Member's Day IAT workshop (2015) was overwhelmingly positive. This is discussed in more detail in Chapter 8. The designated system for attributing weightings was a feature many users found beneficial, however the question of incorrect weighting was raised. Weightings have a subjective element, which can be influenced by the experts included within the Britest study. The tables described in Chapter 4 for attributing weightings were designed with this in mind, as a method of limiting the potential for ambiguous weightings, but the qualitative nature

of a Britest study means that this potential shortcoming of the IAT cannot be fully eradicated. Therefore, different teams may attribute different weightings, generating different results. While the weighting definition tables go some way towards limiting the scope for incorrect weighting attribution, the sensitivity of these weightings remained unknown. For this reason, the study described in this Chapter was devised.

Multiple systems are available to aid in the adoption of quality by design (I.C.H Guideline, 2009). These range from basic flow charts, to more complex statistical methods such as Failure Mode and Effects Analysis (FMEA) or Monte Carlo Simulations (MCS). Failure Modes and Effects Analysis (FMEA) is a systematic method used to pre-empt causes of failure, to enable preventative action to be taken to minimise loss or disruption in the case of a failure occurring. This may be minimising with respect to loss of product, plant time or profit, depending on the process and value of the product. The standard form for this analysis comprises of identifying the potential problems, and then analysing the effectiveness of the remedial action which could be taken (Stamatis, 2003). This is shown in more detail in Figure 7.1:

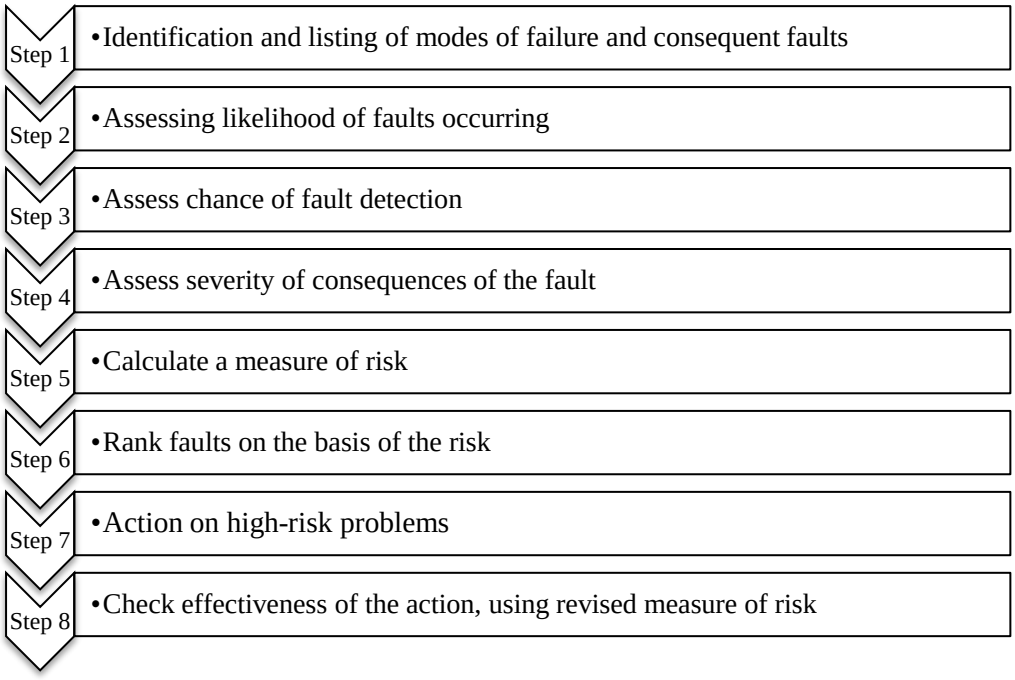


Figure 7.1 - Detailing the stages required for a FMEA analysis in the order they would be applied. Adapted from Gilchrist (1993).

FMEA follows a logical set of steps in an ordered fashion, with clear actions being generated at each step. In this sense it is a strong method for implementation in an industrial process as each stage has a purpose and the strong pattern should make it easy to follow with sufficient process understanding. However, it is not fully quantitative and so has limitations for applicability to complex risk scenarios.

Monte Carlo simulations (MCS) are the most widely applicable method for full risk quantification, rather than the semi-quantitative previously discussed. It allows weights and cost functions to be applied to variables, and combines this information with probability distributions to give a full risk analysis with a statistical basis to be constructed.

In its very basic form, the Monte Carlo method is similar to what if analysis in that it accounts for every possible outcome (Vose, 1996). The key difference is that it accounts for every possible value within a range, and uses the probability to weight how likely this

value is of occurring. In contrast, what if analysis is a crude methodology where sets of values are decided upon for each variable. The statistical basis for the Monte Carlo means that the models generated are of high quality and can be accurately used to describe the risk within a process (Vose, 1996). Both FMEA and MCS have been used to examine sensitivity and risk within a range of bioprocesses (Marchal *et al.*, 2001; Biwer *et al.*, 2005; Mollah, 2005; Farid, 2007; Witcher, 2014).

The ability to demonstrate process understanding is critical to regulatory approval for a product, and typically a range of techniques would be employed to examine the potential risks associated with a product/process. Within process understanding tools, weighting or scoring systems are not uncommon. They can be highly beneficial when seeking to understand the criticality of process conditions or outcomes. This is especially relevant to both the QbD approach, and risk assessment approaches. One example of using scores within process understanding tools is the A-Mab case study (C.M.C. Biotech Working Group, 2009). Several tools were developed within this work, and most employed a scoring system. The scoring systems employed varied with tools used, and each system was developed for the tool it is employed with, rather than utilising a single scoring system for all tools. As an example, Tool #1 encompassed an "Impact Score" and an "Uncertainty Score". Definitions for these are laid out in Tables 1 and 2. Within the A-Mab case study, these tools were used to investigate the risk for various characteristics of a product in terms of uncertainty (Tool #1,Table 7.1) and likelihood (Tool #2, Table 7.2) of occurrence. The Impact Score ranges from 2-20, and the Uncertainty Score ranged from 1 to 7. Using these scores, the Risk Score is calculated by multiplying the two together. The ability to use tools of this nature to critically consider the risk associated with a product, in this case a monoclonal antibody, is invaluable when considering the product in terms of regulatory approval. The pre-determined categories, similar to the

IAT, make the potential for ambiguity in assigning criticality minimal, but not unimaginable.

Table 7.1 - Tool #1 from the A-Mab Case Study C.M.C Biotech Working Group (2009). Abbreviations – PK=pharmacokinetics, PD=pharmacodynamics, ATA=anti-therapeutic antibody, AE=Adverse Effects.

Impact Score	Biological Activity or Efficacy	PK/PD[a]	Immunogenicity	Safety
Very High (20)	Very Significant Change	Significant Change on PK	ATA detected and confers limits on safety	Irreversible AEs
High (16)	Significant Change	Moderate Change with impact on PD	ATA detected and confers limits on efficacy	Reversible AEs
Moderate (12)	Moderate Change	Moderate Change with no impact on PD	ATA detected with *in vivo* effect that can be managed	Manageable AEs
Low (4)	Acceptable Change	Acceptable Change with no impact on PD	ATA detected with minimal *in vivo* effect	Minor, transient AEs
None(2)	No Change	No impact on PK or PD	ATA not detected or ATA detected with no relevant *in vivo* effect	No AEs

Table 7.2 - Tool #2 from the A-Mab Case Study C.M.C Biotech Working Group (2009).

Uncertainty Score	Description (Variants and Host Related Impurities)	Description (Process Raw Material)[a]
7 (Very High)	No Information (New Variant)	No Information (new impurity)
5 (High)	Published external literature for variant in related molecule	-
3 (Moderate)	Nonclinical or in vitro data with this molecule. Data (nonclinical, *in vitro* or clinical) from a similar class of molecule.	Component used in Previous Processes
2 (Low)	Variant has been present in material used in clinical trials.	-
1 (Very Low)	Impact of specific variant established in Clinical Studies with this molecule.	Generally Regarded as Safe (GRAS) or studied in clinical trials

While both tools gave the user an assessment of the risks involved, and the criticality of a characteristic to the success of the process, neither could definitively attribute criticality. Results were broadly consistent across both tools, with only minor inconsistencies in criticality. The main difference between these tools and the Britest tools is the limited ability of the A-Mab tools to increase process understanding. The A-Mab tools rely on the user understanding the quality attributes of the product prior to tool employment. However, the Britest tools would aim to enhance the process understanding through structured application requiring all known information relating to the process/product to be captured within the tool. A user could be unaware of a quality attribute when employing the A-Mab tools and could remain unaware after tool completion, whereas the Britest tools would aim to uncover the knowledge gaps, allowing the user(s) to investigate further.

Sensitivity analysis is the process of understanding how a change in designated conditions would affect the final output of a calculation or process. It can be performed in a multitude of software packages including, but not restricted to, MATLAB, Microsoft Excel and Minitab. Sensitivity can be tested either by changing multiple inputs simultaneously (similar to DoE experimental designs), or through changing each factor individually. In the case of investigations represented in this chapter, it was more appropriate to consider the changes simultaneously. It is possible to be incorrect on every single weighting, or on only one, and so by testing all weightings simultaneously, the impact of all possible scenarios could be investigated and understood.

The following chapter investigates the impact of a weighting within the IAT being incorrect by a value of ± 1. The IATs within this study have all been generated using random number generators unless otherwise stated. All possible combinations of weightings were simulated, and relationships generated at random. Ten parameters were simulated for each set of weightings. This totalled 25 IATs, or 250 parameters, for each number of outcomes (5 and 10) tested in this case.

The work aimed to ascertain at what point a parameter score could be considered reliable. A result was considered 100% reliable if 100% of the possible results generated the same indicated direction of change to the parameter. For example, if a parameter had a score of +5, for this to be considered 100% reliable all of the possible results would have to be positive in value. For the purpose of this work, it was assumed that the initial randomly generated result was the correct result, and the possible variations of ±1 were the permutations that could have been generated in a Britest study. It also tested whether this was possible to determine without running the extensive simulations. The study then moves on to consider the effect of a score variability of ±3, and the effect of employing an alternative weighting system. The IAT weighting system was compared to a system

employed within one of the Britest members, a multinational pharmaceutical company. The system preferred by this multinational utilises 1, 5 and 10 as weightings. For this comparison, a sample of IATs had their weightings converted to this system to ascertain the impact of this on the overall score for each parameter. The study concludes by considering the sensitivity of an IAT generated by AbbVie in the initial tool development work.

7.2 Methods

7.2.1 Simulations

The IAT has not currently been launched within Britest as a tool. In the absence of access to IATs generated as a result of Britest studies, simulated IATs were created for the purposes of sensitivity analysis. Microsoft Excel (USA, 2013) was used to create 50 IATs. Twenty five of the IATs had ten parameters and five outcomes. The remaining twenty five had ten parameters and ten outcomes. All random number generators were set to generate using a normal distribution and were required to be whole numbers.

Each simulation started with using a random number generator to give the weighting values (red box, Figure 7.2). The RANDBETWEEN function was used to create five or ten weightings between 2 and 10. The lower bound and upper bounds were set in line with the possible lowest and highest weightings when an IAT is created within a Britest study using the system first described in Chapter 4.

	Outcome 1	Outcome 2	Outcome 3	Outcome 4	Outcome 5	Outcome 6	Outcome 7	Outcome 8	Outcome 9	Outcome 10
Weighting →	10	9	7	2	3	8	4	7	9	8
Parameter 1	1	1	-1	0	1	0	-1	1	0	-1
Parameter 2	-1	1	-1	1	1	-1	-1	0	0	1
Parameter 3	0	0	0	0	-1	-1	1	1	1	0
Parameter 4	0	1	0	0	1	1	0	-1	-1	-1
Parameter 5	1	1	-1	0	0	1	0	-1	-1	0
Parameter 6	1	-1	1	0	0	0	0	0	0	0
Parameter 7	-1	-1	1	1	0	1	-1	0	0	-1
Parameter 8	0	1	-1	0	1	-1	1	0	-1	1
Parameter 9	1	1	-1	1	0	-1	-1	-1	0	1
Parameter 10	1	-1	1	-1	1	0	-1	-1	-1	0

Figure 7.2 - Example IAT where the red box highlights the outcomes and associated weightings, and the black box highlights the relationships between the parameters and outcomes (+1 for positive, 0 for none, -1 for negative).

The randomly generated weightings were then transferred into a Table shown in Figure 7.3. This was then completed to show the possible weightings if the weighting of respective outcomes was incorrect by ±1. In the example case shown in Figures 7.1 and 7.2 this gave 8, 9 and 10 as possible weightings for Outcome 2. If the weighting was assigned as 2 (Outcome 4), then only 2 and 3 would be considered, as it would not be possible to generate an IAT with a weighting of less than 2. Likewise if a weighting was assigned a 10 (Outcome 1) then only 9 and 10 would be considered as it would not be possible to generate an IAT with a weighting of greater than 10.

Outcome 1a	Outcome 2a	Outcome 3a	Outcome 4a	Outcome 5a	Outcome 6a	Outcome 7a	Outcome 8a	Outcome 9a	Outcome 10a
Outcome 1	Outcome 2	Outcome 3	Outcome 4	Outcome 5	Outcome 6	Outcome 7	Outcome 8	Outcome 9	Outcome 10
9	8	6		2	7	3	6	8	7
10	9	7	2	3	8	4	7	9	8
	10	8	3	4	9	5	8	10	9

Figure 7.3 - Example of the Table used for the query function to generate all possible combinations of the weightings ±1.

Tables such as Figure 7.3 were used to generate all possible combinations of each column. This was completed using the Microsoft Query Function. Figure 7.4 shows an example result which would be obtained. In this case, the result contained 26245 possible combinations.

Q26231

	A	B	C	D	E	F	G	H	I	J
1	Outcome 1	Outcome 2	Outcome 3	Outcome 4	Outcome 5	Outcome 6	Outcome 7	Outcome 8	Outcome 9	Outcome 10
2	9	8	6	2	2	7	3	6	8	7
3	9	8	6	3	2	7	3	6	8	7
4	10	8	6	2	2	7	3	6	8	7
5	10	8	6	3	2	7	3	6	8	7
6	9	8	6	2	2	7	3	6	8	8
7	9	8	6	3	2	7	3	6	8	8
8	10	8	6	2	2	7	3	6	8	8
9	10	8	6	3	2	7	3	6	8	8
10	9	8	6	2	2	7	3	6	8	9
11	9	8	6	3	2	7	3	6	8	9
12	10	8	6	2	2	7	3	6	8	9
13	10	8	6	3	2	7	3	6	8	9
14	9	9	6	2	2	7	3	6	8	7
15	9	9	6	3	2	7	3	6	8	7
16	10	9	6	2	2	7	3	6	8	7
17	10	9	6	3	2	7	3	6	8	7
18	9	9	6	2	2	7	3	6	8	8
19	9	9	6	3	2	7	3	6	8	8
20	10	9	6	2	2	7	3	6	8	8
21	10	9	6	3	2	7	3	6	8	8
22	9	9	6	2	2	7	3	6	8	9
23	9	9	6	3	2	7	3	6	8	9
24	10	9	6	2	2	7	3	6	8	9
25	10	9	6	3	2	7	3	6	8	9
26	9	10	6	2	2	7	3	6	8	7
27	9	10	6	3	2	7	3	6	8	7
28	10	10	6	2	2	7	3	6	8	7
29	10	10	6	3	2	7	3	6	8	7
30	9	10	6	2	2	7	3	6	8	8
31	9	10	6	3	2	7	3	6	8	8
32	10	10	6	2	2	7	3	6	8	8
33	10	10	6	3	2	7	3	6	8	8
34	9	10	6	2	2	7	3	6	8	9
35	9	10	6	3	2	7	3	6	8	9
36	10	10	6	2	2	7	3	6	8	9
37	10	10	6	3	2	7	3	6	8	9
38	9	8	7	2	2	7	3	6	8	7

Figure 7.4 - Example output from the Query function.

Once weightings were assigned and combinations had been generated, the relationships between each parameter and outcome were simulated, again using the RANDBETWEEN function. In this case the relationships could be -1, 0 or +1. A -1 represented a negative relationship, a 0 indicated no relationship, and +1 indicated a positive relationship. Across each parameter the relationships used for the simulation included both positive and negative relationships (Black square, Figure 7.2). If a parameter has an exclusively positive or negative influence on all outcomes then the overall drive direction will remain unchanged regardless of the weighting values.

The completion of both weightings and parameters then allowed the sensitivity of each result to be tested. The scores for drive to increase/decrease were generated separately.

Their formulas are shown below in word form, and in Microsoft Excel format, respectively:

Drive to Increase:

If the original weighting multiplied by the relationship is more than zero then show the value of the corresponding cell from the possible combinations, if it is less than or equal to 0 then show 0.

=IF((B10*B11)>0,Variations!A2,IF((B10*B11)<=0,0))

Drive to Decrease:

If the original weighting multiplied by the relationship is less than zero then show the value of the corresponding cell from the possible combinations multiplied by minus 1, if it is more than or equal to 0 then show 0.

=IF((B10*B11)<0,(Variations!A2*-1),IF((B10*B11)>=0,0))

Where "Variations!" refers to the sheet in which the possible combinations of the weightings is contained.

In these, cell B10 (red) in Figure 7.5 contained the original weighting, and cell B11 (blue) contained the parameter relationship to the outcome. "Variations!A2" linked to the first row in the sheet containing each possible combination. This calculation is shown in Figure 7.5.

Summary												
	Mean	9.5		DI	29							
	Mode	9		DD	-19							
	High	16		Score	10							
	Low	3										
	+	26244	100	26244	26259							
	0	0	0									
	-	0	0									
	10	**9**	**7**	**2**	**3**	**8**	**4**	**7**	**9**	**8**		
Parameter 1	1	1	-1	0	1	0	-1	1	0	-1		**Score**
Drive to increase	10	9	0	0	3	0	0	7	0	0	29	10
Drive to decrease	0	0	-7	0	0	0	-4	0	0	-8	-19	
	DI										**Total**	
	9	8	0	0	2	0	0	6	0	0	25	
	9	8	0	0	2	0	0	6	0	0	25	
	10	8	0	0	2	0	0	6	0	0	26	
	10	8	0	0	2	0	0	6	0	0	26	
	9	8	0	0	2	0	0	6	0	0	25	
	9	8	0	0	2	0	0	6	0	0	25	
	10	8	0	0	2	0	0	6	0	0	26	
	10	8	0	0	2	0	0	6	0	0	26	
	9	8	0	0	2	0	0	6	0	0	25	
	9	8	0	0	2	0	0	6	0	0	25	
	10	8	0	0	2	0	0	6	0	0	26	

DD										**Total**		**Score**
0	0	-6	0	0	0	-3	0	0	-7	-16		9
0	0	-6	0	0	0	-3	0	0	-7	-16		9
0	0	-6	0	0	0	-3	0	0	-7	-16		10
0	0	-6	0	0	0	-3	0	0	-7	-16		10
0	0	-6	0	0	0	-3	0	0	-8	-17		8
0	0	-6	0	0	0	-3	0	0	-8	-17		8
0	0	-6	0	0	0	-3	0	0	-8	-17		9
0	0	-6	0	0	0	-3	0	0	-8	-17		9
0	0	-6	0	0	0	-3	0	0	-9	-18		7
0	0	-6	0	0	0	-3	0	0	-9	-18		7
0	0	-6	0	0	0	-3	0	0	-9	-18		8

Figure 7.5 - Example output for each Parameter in the IAT.

This formula was used to generate the possible scores for each row in the variations table, resulting in an output similar to that shown as an example in Figure 7.5. The overall score was also calculated for each row, giving the whole range of possible scores for the parameter and weightings (green-Figure 7.5).

A range of statistics was produced for each parameter. The Mean, Mode, highest value and lowest value were all included, along with the % of overall scores which were positive, zero and negative. A graph of results such as that shown in Figure 7.6 was produced.

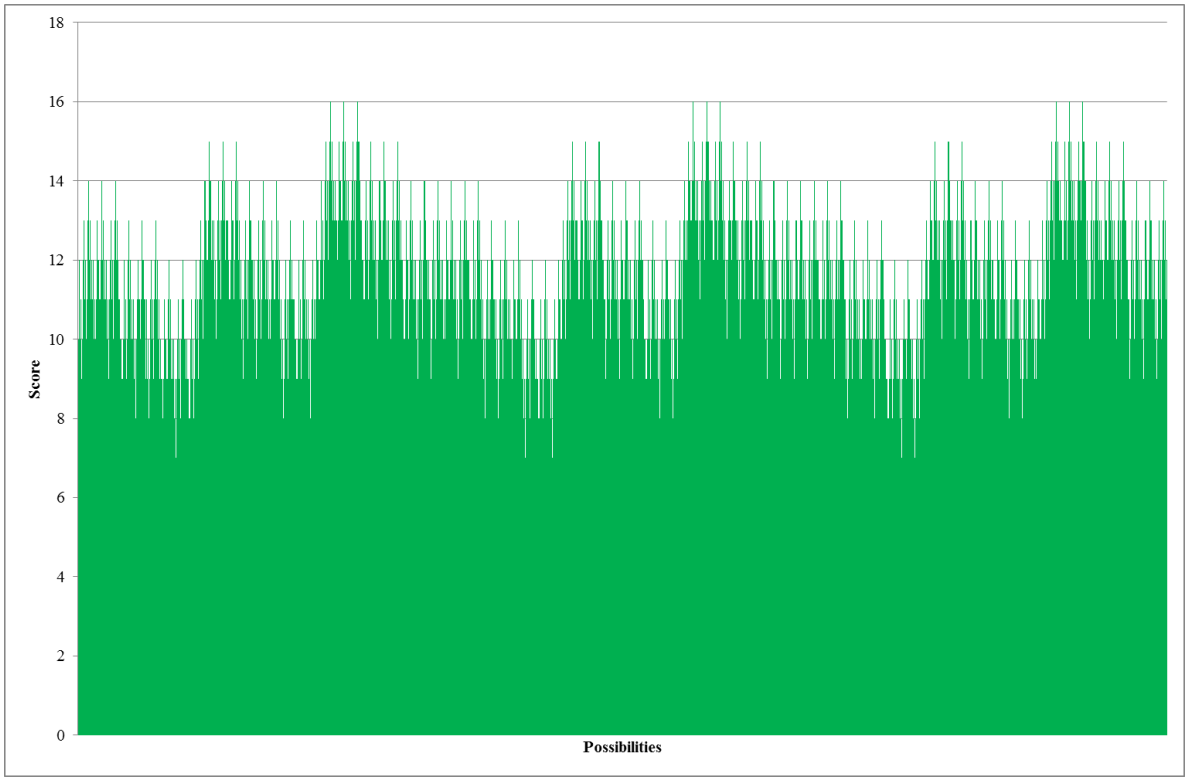


Figure 7.6 - Example graph of results showing the possible scores for each combination of weightings for a single parameter.

	Outcome	Outcome 2	Outcome 3	Outcome 4	Outcome 5	Outcome 6	Outcome 7	Outcome 8	Outcome 9	Outcome 10	Drive to Increase	Drive to Decrease	Score	Mean	Mode	High	Low	+%	0%	-%
Weighting ➔	10	9	7	2	3	8	4	7	9	8										
Parameter 1	1	1	-1	0	1	0	-1	1	0	-1	29	-19	10	9.5	9	16	3	100	0	0
Parameter 2	-1	1	-1	1	1	-1	-1	0	0	1	22	-29	-7	-6	-6	1	-13	0.03	0.27	99.7
Parameter 3	0	0	0	0	-1	-1	1	1	1	0	20	-11	9	9	9	14	4	100	0	0
Parameter 4	0	1	0	0	1	1	0	-1	-1	-1	20	-24	-4	-4	-4	2	-10	0.96	2.88	96.16
Parameter 5	1	1	-1	0	0	1	0	-1	-1	0	27	-23	4	3.5	4	9	-2	94.44	4.12	1.44
Parameter 6	1	-1	1	0	0	0	0	0	0	0	17	-9	8	7.5	7	10	5	100	0	0
Parameter 7	-1	-1	1	1	0	1	-1	0	0	-1	17	-31	-14	-13	-13	-7	-19	0	0	100
Parameter 8	0	1	-1	0	1	-1	1	0	-1	1	24	-24	0	0	0	7	-7	41.02	17.97	41.02
Parameter 9	1	1	-1	1	0	-1	-1	-1	0	1	29	-26	3	3	3	10	-4	87.93	7.24	4.84
Parameter 10	1	-1	1	-1	1	0	-1	-1	-1	0	20	-31	-11	-11	-12	-5	-19	0	0	100

Figure 7.7 - Final IAT including summary statistics (rounded to 2 decimal places)

This process was repeated for ten parameters for each IAT (Figure 7.7), resulting in ten graphs and ten sets of summary statistics. Fifty IATs in total were generated, twenty five with five outcomes and twenty five with ten outcomes, all with ten parameters.

7.2.2 Alternative weighting system

A subset of IATs were tested with the alternative weighing system. To convert from the original system, a limit of ±2 was applied. This meant any value of 2 was designated a 1, any values between 3 and 7 were designated 5, and any values 8 or over were designated a 10. The score for each parameter was compared to that of the original IAT.

7.2.3 Industrial Case Study

The IAT constructed by AbbVie for their antibiotic fermentation was tested using the sensitivity analysis method described above. The random numbers were replaced with the numbers attributed by AbbVie, but the generation of variations and the calculation methods remained the same as described above. It is notable that the weightings were attributed without using the system presented within Chapter 4. Additionally, the original IAT contained eleven outcomes, which was beyond the computing capability of Microsoft Excel. In light of this, one outcome was removed from the IAT to give ten outcomes which would make the sensitivity analysis possible. This outcome was selected as it had the lowest weighting value (1), and therefore AbbVie had deemed to the least influential of the possible outcomes. For the purpose of this thesis in light of confidentiality restrictions the outcomes and parameters for the fermentation have had to be anonymised.

7.3 Results and Discussion

This work aimed to test how sensitive to change the weightings of the IAT were with a small and large number of outcomes. The weightings within the IAT are attributed by experts working on a process, and each person may have a different view as to the importance of an outcome depending on their area of expertise. This ambiguity could be a shortcoming in the tool, and this study sought to ascertain how confident users of the tool could be in the resulting scores.

Initially Microsoft Excel (USA, 2013) was used to generate IATs using random number generators, to allow a test on a high number of completed tools without requiring information from a high number of industrial processes, which could be difficult to obtain. In addition to the results from the simulations, this gave rise to a spreadsheet which could be used on a real IAT to ascertain weighting sensitivity in a quick and efficient manner. This will allow organisations from the Britest consortium to make a business case for any arising work from Britest studies involving the IAT with confidence.

All IATs had ten parameters, with randomly generated relationships to the outcomes. All parameters contained a mix of positive and negative relationships, as anything with exclusively positive or negative relationships would give 100% confidence regardless of the variations in weightings. Initially the work focussed on simulated IATs with 5 outcomes, before moving onto consider 10 outcomes. The effect of the weightings of the outcomes being incorrect by ±1 number was investigated. For example, if an outcome was assigned a weighting of 6 the work would examine the effect of it being 5, 6 or 7.

For each IAT, the 5/10 outcomes were varied by ±1 and all possible combinations of these were generated in Excel. These were then tested with the various parameter

relationships to generate all possible scores for each parameter if the weighting was incorrect by ±1. The results were summarised with the number of parameters involved in the relationship, mean, mode, drive to increase, drive to decrease, score, range and confidence. The confidence was defined as the number of possible results which would result in the same action on the parameter (i.e. generated a positive or negative score) as a % of the total possible number of results (Equation 7.1). The results from the IATs are shown in full in Appendix C. Results were summarised as Table 7.3.

Equation 7.1

$$\frac{Number\ of\ results\ consistent\ with\ the\ original\ action}{Number\ of\ possible\ combinations}\ x\ 100$$

Table 7.3 - Format for results from the IATs generated.

IAT	Parameter	Parameters Involved			True Values					Distance from zero					Range	Confidence	Weighted Score
		Positive	Negative	Total	DI	DD	Score	Mean	Mode	DI	DD	Score	Mean	Mode			
1	1																
1	2	1	1	2	4	-3	1	1	1	4	3	7	1	1	-4	66.67	0.5
1	3	2	1	3	11	-10	1	1.5	1	11	10	21	1.5	1	-5	77.78	0.33
1	4	2	2	4	14	-11	3	2.5	3	14	11	25	2.5	3	-7	90.74	0.75
1	5	1	2	3	4	-11	-7	-7	-7	4	11	15	7	7	-6	100	-2.33
1	6	2	2	4	7	-19	-12	-11.5	-12	7	19	26	11.5	12	-7	100	-3
1	7	1	2	3	8	-7	1	1	1	8	7	15	1	1	-6	62.96	0.33
1	8	2	1	3	13	-3	10	10	10	13	3	16	10	10	-6	100	3.33
1	9	3	2	5	21	-13	8	7.5	7	21	13	34	7.5	7	-9	100	1.6
1	10	2	2	4	13	-18	-5	-4.5	-5	13	18	31	4.5	5	-7	100	-1.25
2	11	2	2	4	13	-11	2	2	2	13	11	24	2	2	-8	81.48	0.5
2	12																
⬇	⬇																
25	250																

Initially it was hypothesised that a correlation would be likely between confidence and score, with the resulting graphs from 5 and 10 outcomes respectively shown in Figure 7.8. The anticipated correlation was confirmed, with 100% confidence being more prevalent in results with higher scores. However, a single threshold value could not be ascertained. For example, in one case, a score of 9 did not give 100% confidence, but in another the score of 3 did give 100% confidence.

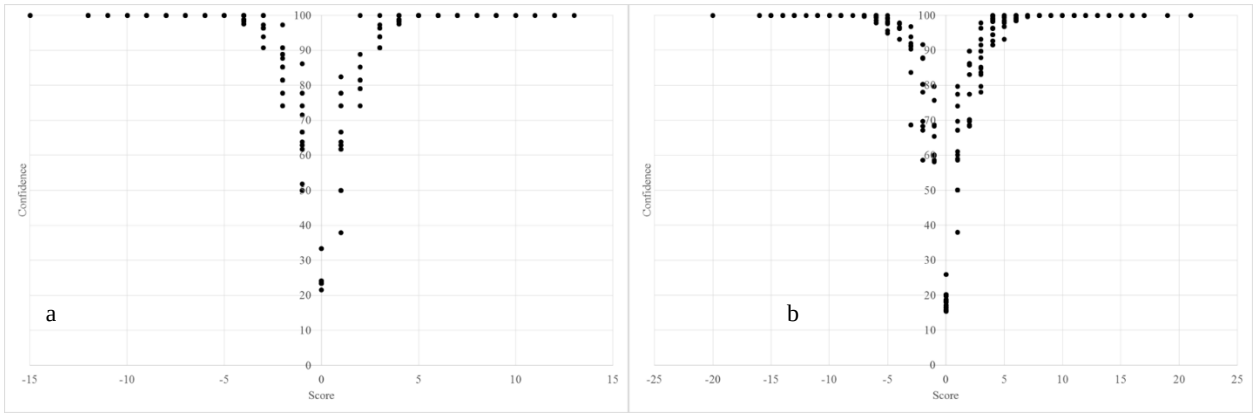


Figure 7.8 - Graphs comparing the score of each IAT to the confidence in the result. Results from the IATs with 5 outcomes as graph a, those with 10 outcomes as graph b.

Despite a clear trend, there was a lack of a defined threshold value, and so the overall "drive" was tested against confidence, i.e. the drive to increase and decrease values added together without the negative prefix for the drive to decrease. So a drive to increase of 10 and a drive to decrease of -12 would give a total of 22 rather than the original score of -2. This yielded poor results, with no correlation within the results. Next the difference between the mean or mode and the original score was examined, on the basis that this difference would be lower in the cases where confidence was high. This also yielded no discernible pattern, and so the range was compared to the confidence. It was assumed that cases with a high range would show a low confidence value, but this was also shown not to be the case. All results from these investigations are included in Appendix C.

Detailed examination of Figure 7.8 (weighted score vs confidence) showed that the number of parameters was influencing the results. In light of this, a weighted score was calculated by dividing the score by the number of parameters involved. The results are shown in Figure 7.9.

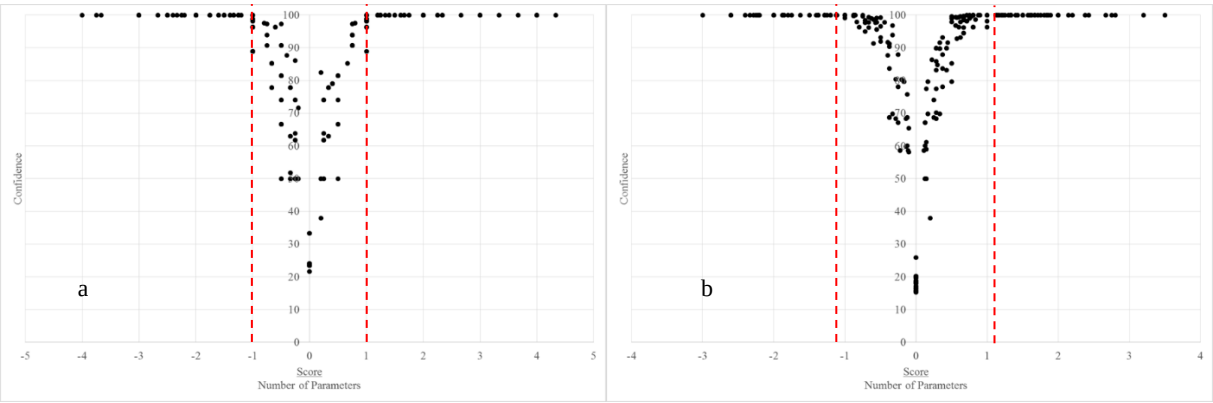


Figure 7.9 - Constant value (Score/Number of Parameters) against the confidence in the result. Results from the IATs with 5 outcomes are on the left, those with 10 outcomes on the right.

Starting with Figure 7.9a, where IATs with 5 outcomes were considered, the graph would indicate that results with a weighted score closer to zero are more likely to have low confidence. While the original hypothesis for the work proposed that a threshold score could be possible, past which point confidence would be high, that was shown not to be the case in Figure 7.8. However, the plotting of the weighted score against the confidence shows that a threshold value is present, when the number of parameters is accounted for. Figure 7.9a shows that for 5 outcomes, this threshold is 1.2, and the Figure 12b shows the same relationship is true for IATs with 10 outcomes, though the threshold in that case is 1 (indicated by the red dashed lines).

These threshold values suggest that to be confident in the results from an IAT the number of parameters involved needs to be low, or the score needs to be high. For example, a relationship where 3 parameters are involved and the weightings are all 8 or above would give a result with a high level of confidence. Conversely a relationship involving 7 parameters, all of which are weighted between 4 and 6, would be less certain. Following the

trend identified from the IATs with 5 and 10 outcomes, it would appear that as the number of parameters increases, the threshold value decreases. This trend held true for the AbbVie IAT (Table 7.4), where the only parameter showing less than 100% confidence (Parameter 3 level) had a constant value of 1, with a score of 3 and 3 parameters involved. The confidence level was 96%, showing the probability of being incorrect was low, though it was surprising that this showed any ambiguity and the weighted score should be used only as an indication of likely confidence and not a steadfast rule. From here several further research questions were raised.

Table 7.4 – IAT and associated summary statistics from AbbVie.

	Key Outcomes (Measurements) ➔																
	Outcome 1	Outcome 2	Outcome 3	Outcome 4	Outcome 5	Outcome 6	Outcome 7	Outcome 8	Outcome 9	Outcome 10	DI	DD	Score	+	0	-	Constant
Media/ Operating Conditions (Parameters) ➔	**3**	**5**	**5**	**7**	**10**	**5**	**8**	**5**	**8**	**4**							
Parameter 1	1	1	0	1	0	-1	0	0	0	0	15	-5	10	100	0	0	2.5
Parameter 2	1	1	1	1	1	-1	1	0	0	0	38	-5	33	100	0	0	4.71
Parameter 3	1	1	0	0	0	-1	0	0	0	0	8	-5	3	96.3	3.7	0	1
Parameter 4	0	0	1	0	1	0	0	0	0	0	15	0	15	100	0	0	7.5
Parameter 5	0	0	0	0	0	0	0	1	1	0	13	0	13	100	0	0	6.5
Parameter 6	0	0	1	1	1	0	1	-1	0	0	30	-5	25	100	0	0	5
Parameter 7	1	0	-1	1	-1	-1	0	0	0	-1	10	-24	-14	0	0	100	-2.33
Parameter 8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	
Parameter 9	1	1	0	0	0	1	0	0	0	0	13	0	13	100	0	0	4.33
Parameter 10	1	1	0	0	0	1	0	0	0	1	17	0	17	100	0	0	4.25
Parameter 11	0	0	-1	0	-1	0	1	0	-1	0	8	-23	-15	0	0	100	-3.75

It was not possible within the study to test all arising hypotheses, and so three research questions were selected for further investigation. Firstly, it was decided to investigate the effect of greater uncertainty on the threshold value. Rather than test all possible options, it was decided that change of outcome weightings by ±3, using just the extreme values would give an adequate indication of whether it was possible to reach a situation where 100% confidence is not possible. Secondly, the effect of the number of outcomes on threshold value was tested. Microsoft Excel was not capable of testing more than ten outcomes, however 5 IATs were generated with 7 outcomes to be used as indicators of threshold value effect. It was anticipated that the threshold value for these would lie around 1.1 as this was directly in the middle of the thresholds for 5 and 10 outcomes. Lastly, the results from the sensitivity of the current weighting system were compared to that of a QbD accepted system. The 1, 5, 10 system for importance is used, for example, in a major pharmaceutical Britest member company currently when carrying out risk analysis for QbD. Compared with the weighting system proposed for the IAT, this forces the user to separate attributes by greater margins, to ensure the most critical attributes are the most obvious. However, this enhanced separation is also expected to increase the error impact. Comparing the two systems aimed to allow the user to determine the most appropriate weighting system for the Britest study.

Considering first the increased error size, it was found that this increased the threshold value significantly, to 2.1 (5 outcomes). The high rate of error made it difficult to obtain any results with 100% confidence, but this did happen in some cases. The test was done with IATs 1-5, where the relationships remained the same to allow for direct comparison, and only the high/low values were changed. Most results did not have 100% confidence, though in a minority of cases this was achieved.

The threshold value for 7 outcomes was found to be between 1 and 1.2, as predicted (Figure 7.10). The smaller sample size made it difficult to determine a threshold as conclusively as in the IATs with 5 and 10 outcomes, but parameters associated with a weighted score of 1.2 or above resulted consistently in a 100% confidence value, with occasional incidences of 100% confidence at a weighted score of 1.0. If this trend continues beyond the number of outcomes tested then the higher the number of outcomes, the lower the threshold value for 100% confidence. This is highly significant when applying the IAT within bioprocessing, where the number of outcomes could be high, or when applying the IAT to multiple unit operations where many outcomes could be measured.

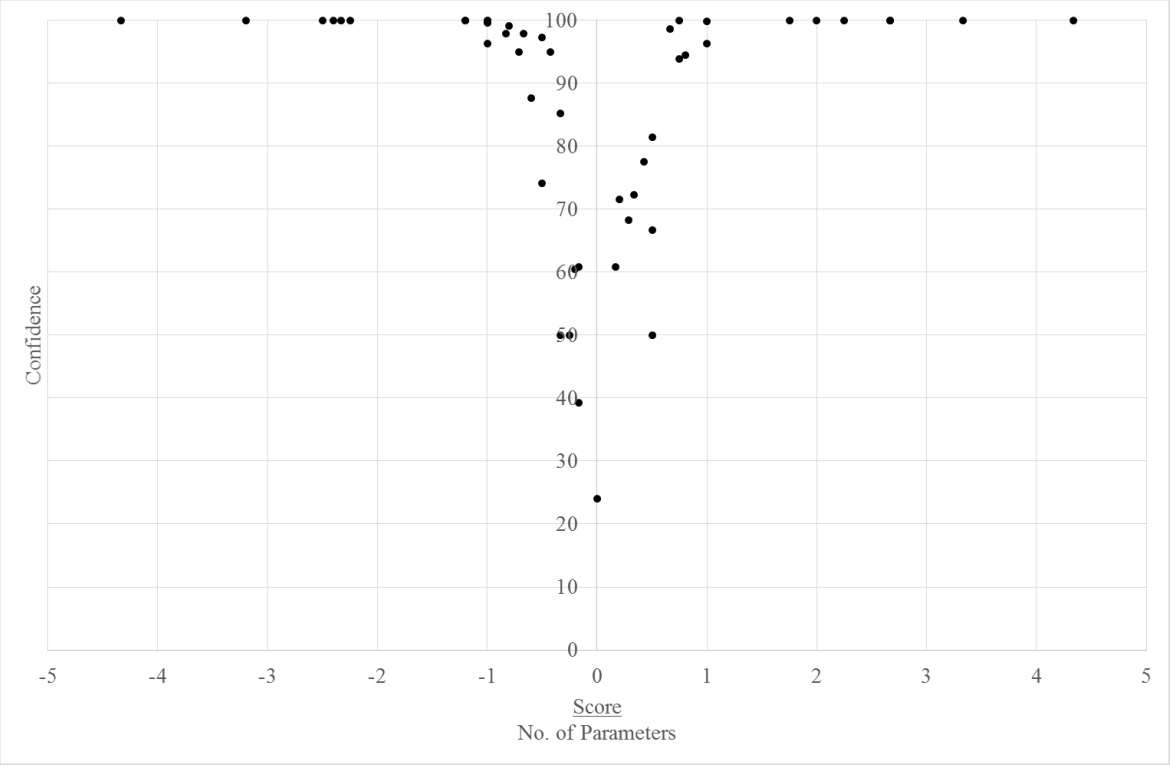


Figure 7.10 - Score vs constant for IATs with 7 outcomes.

Comparing the IAT weighting system to the QbD scoring system was achieved by using IATs 1-5, and converting the scores to the QbD system. In this system, any value of 2 or below was designated a 1, values between 3 and 7 were designated 5, and values 8 or over were assigned a 10. In most cases, the weighting system was shown not to affect the overall

outcome. Most parameters would have different scores, and so prioritisation of experiments may be altered, but in general scores remained positive or negative. However in some cases, the scores changed from positive to negative, or vice versa. The exaggeration of effects through using a 1, 5, 10 system is potentially powerful for minimising ambiguity when constructing an IAT, but this could also reduce confidence in the result and also potentially negatively impact the process by suggesting a drive to decrease a parameter when an increase is required. Therefore the original structured weighting system for the IAT which utilises two sets of scores out of 5 combining business and process benefit has been shown to be a superior system. The concern over a group being unable to agree on a weighting, or being reluctant to fully utilise the range of weightings available should be overcome through the clear categories provided for the business and process benefit. In addition to this, the ability to use the score and number of parameters involved to indicate the likely reliability of the weightings to within ±1 will ensure any possible ambiguity is understood prior to experimentation being carried out.

Within the AbbVie IAT case, the change in scoring system did not affect the way in which a parameter would be altered. Scores varied slightly in value, but the overall drive to increase or decrease a parameter was unaffected. However, the IAT contained only one parameter which did not generate 100% confidence (Parameter 3, Table 7.4), and 50% of the outcomes fitted with the 1, 5, 10 system without alteration. In a more varied IAT these results could have been different.

7.4 Summary

This study examined the impact of the weighting sensitivity on the outcomes of the IAT tool, to ascertain the impact of human error on the outcomes from the tool. While it was anticipated that the score alone could be used to infer the confidence in the results, it was

found that the number of parameters contributing to the score was required to be considered. Dividing the score by this gave a reliable indication of the confidence in the results, though the number of outcomes influenced the point past which the confidence would be 100%. If a larger error in outcome weighting was considered (±3 rather than ±1) then the threshold value increased, and it was rare to see a result which displayed 100% confidence. The implementation of the structured weighting attribution system should minimise error, and the calculation of the threshold value of the weighted score will allow users to determine not only which parameters could be negatively impacting the process the most but also which can be changed with the highest degree of confidence. Finally an alternative weighting system not designed specifically for the IAT was considered, which was shown to adversely affect results in rare cases, suggesting that the original structured approach would be better suited to the IAT tool.

Across a multitude of industries weighting systems are employed for a multitude of purposes (Burgess and Brennan, 2001; Kleiner *et al.*, 2005; Shaeri *et al.*, 2006; Kumschick *et al.*, 2015; Valtorta *et al.*, 2015; Nentwig *et al.*, 2016). A range of systems exist, and each will have associated merits and drawbacks for the situation it is applied within, and the system is often employed by organisations based on historical application rather than critical assessment of these. As pharmaceutical industry adopts a "Quality by Design" (I.C.H Guideline, 2009) approach, qualitative process understanding tools will become invaluable. While tools of this nature undoubtedly add value to a process, this work highlights the importance of considering the potential for error arising from such systems. This work clearly demonstrates the benefits of considering the possible ambiguity being introduced into a decision making process through these weighting systems, and the caution with which they should be applied within high value processes such as in the biopharmaceutical industry.

Chapter 8 Research Conclusion and Industrial Impact

8.1 Research Conclusion

This Engineering Doctorate (EngD) thesis has presented work undertaken in collaboration with Britest Ltd to develop the Britest tools for application to bioprocessing. The need for knowledge management tools within bioprocessing to support QbD adoption has been identified as a challenge (Herwig *et al.*, 2015), and this research sought to fill this gap through the creation of tools designed for this purpose.

This research aimed to:

1. Develop novel knowledge management tools designed specifically for bioprocessing
2. Investigate whether the Britest tools could be applied to bioprocessing to perform this function
3. Test these tools on a range of industrially relevant datasets
4. Identify the stage of process development at which the tools would add the most value
5. Compare these to alternative methods of enhancing process understanding

Considering these in turn, the work developed two new novel knowledge management tools for bioprocessing, the R2T2 and the IAT. The R2T2 was a redevelopment of the PDD, and the successful demonstration of application to virtual bioprocesses was deemed sufficient to add value to bioprocessing. The IAT was redeveloped significantly from an earlier Britest development, and as such was tested on upstream and downstream datasets to establish suitability for bioprocessing. In addition to these new tools, some tools from within the Britest toolkit such as the ISA/PrISM were shown to be directly applicable

(Chapter 3), whereas others such as the PDD did not address some key features of bioprocessing in their original format.

The focus of the work has been on developing the IAT (Chapter 4), and testing it on industrially relevant upstream and downstream datasets (Chapters 5 & 6), before testing the sensitivity of the weightings included within the tool (Chapter 7). The research has successfully demonstrated the benefits of applying tools of this kind within bioprocess development, and the limitations of such approaches. The resulting toolkit designed for application to bioprocesses is ready for Britest to deploy into their wider membership.

The work presented in this thesis has been shown to add significant value in the early stages of process development, where there is a requirement for tools designed to facilitate interdisciplinary, using qualitative approaches which can be applied in data lean environments following a structured format to give a consistent output, thus overcoming the various shortcomings of the identified alternatives.

Current tools to support QbD adoption and the early stages of process development work in a range of ways, each with their own associated shortcomings. Software implementation methods can use sophisticated systems to generate consistent outputs, but they struggle to overcome the challenges associated with interdisciplinary working (Liao, 2003). Frameworks which follow a structured application flow give support to interdisciplinary teams but do not have a consistent structured output for ease of knowledge transfer (Rathore, 2009). Mathematical modelling can capture highly detailed relationships but require a high level of data to generate a useful model. Finally more qualitative tools such as Six Sigma (Motorola, 2009) give a consistent output with a structured approach designed to facilitate interdisciplinary working, but these were designed for broad application, not specifically to support QbD and bioprocess

development. Within Chapters 5 and 6 the IAT outputs were compared with the original analysis, and reanalysis using a DoE simulation approach which assumed no experimental design information was present. The IAT gave consistent outputs to the original research, and could have significantly reduced the number of experiments required for optimisation had it been employed in the original experimental design.

8.2 Industrial Impact

The EngD is differentiated from a PhD through the industrial sponsorship, and so a successful project must not only advance knowledge in the project area, but a benefit for the industrial sponsor from the research should be demonstrated. In the course of this research several Britest member companies have been presented with the research outcomes, through regular Britest-wide bioprocessing teleconferences, individual interactions, Britest Members Day presentations/posters and through Britest studies. Feedback has been overwhelmingly positive on both the R2T2 and the IAT, both from companies directly involved in their development (in particular AbbVie), and those not directly involved in the research (in particular Johnson Matthey, Infineum, AstraZeneca and Shasun). The industrial impact of this research is shown best through the comments received after presenting the work as part of a workshop at a Britest Members day in October 2015. The session included an introduction to the project, followed by both the R2T2 and the IAT being explained to the audience. The audience was then invited to give feedback and discuss the potential for using both tools within their companies.

Question – Do you feel that the IAT could be applied to any of the process challenges you encounter in your own work? If so please indicate which ones.

AbbVie *– Yes, we have utilised and plan to continue to utilise the IAT for fermentation technology transfer and continuous improvement. We would specifically use it for:*

- *Initial review of technology transfer data*
- *Review of internal pilot plant data prior to scale up*
- *Review of production data for further process understanding and continuous improvement*

FFIC *– I think the tool may have potential in the area of physical processing-however it's difficult to judge until I have more experience of the tool.*

AstraZeneca *– Yes, I'm looking at some polymerisation and both the IAT and R2T2 could be applicable.*

ICES *– Partially. It might provide a link between the biologists and engineers to provide clear communication and mutual understanding.*

Question – Do you feel that the R2T2 could be applied to any of the process challenges you encounter in your own work? If so, please indicate which ones.

AbbVie *– Yes, this tool is useful for any downstream process where impurity clearance is important (should be broadly useful).*

FFIC *– I think the tool may have potential in both physical processing and chemical processing, in particular where multiple phases are present. As with the IAT it's difficult to judge until I have experience using the tool.*

AstraZeneca *– Yes, I'm looking at some polymerisation and both tools could be applicable.*

Shasun *– I can see it might be useful for chemical processes to track what's going on in the reaction mixture during processing operations – raw material consumption, product formation, impurity generation etc.*

ICES *– Yes, need to track progress through process.*

Question: What are your immediate impressions on the structure of the R2T2, and its potential ease of use?

AbbVie *– I think the tool structure is simple and effective. Would make sure to tie the process tasks to PDD number.*

FFIC *– difficult to comment as I have no experience.*

AstraZeneca *– I think a chemistry version would be a mixture between a PDD and DFA. This could be useful and is a current gap.*

Shasun *– Looks good and intuitive. A scale on each box may be useful if tracking.*

After the conclusion of this research, in January 2017, Biogen became Associates of Britest (Britest, 2017). The developing interest in bioprocessing demonstrated in this research was one of the key features that encouraged them to take this decision. In addition interactions between Britest and a mid-sized multinational pharmaceutical company have been opened as a direct consequence of presenting the research contained within Chapter 7 at the ESBES conference in Dublin in September 2016 (McLachlan, 2016).

Since the research project ended, the IAT and R2T2 have both been used by Britest in studies focussing on formulation. This clearly demonstrates the applicability across disciplines. While application to areas outside of bioprocessing was not a major driver in tool development, it is an additional benefit to Britest and is in keeping with the broad applicability of the rest of the toolkit.

This research project was awarded the John Borland Award in October 2016, for making a significant contribution to innovation within Britest. The award recipient is determined by the scientific advisory board from within Britest, and this project was chosen to be the inaugural recipient from a range of projects being carried both within academia and industry (CPI, 2016).

References

Abu-Absi, S.F., Yang, L., Thompson, P., Jiang, C., Kandula, S., Schilling, B. and Shukla, A.A. (2010) 'Defining process design space for monoclonal antibody cell culture', *Biotechnology and Bioengineering*, 106(6), pp. 894-905.

Andersen, A. (1999) 'Business Consulting', *Zukai Knowledge Management, Tokyo Keizar Inc., Japan*, 2001.

Antony, J., Antony, J., Gijo, E.V., Gijo, E.V., Kumar, V., Kumar, V., Ghadge, A. and Ghadge, A. (2016) 'A multiple case study analysis of Six Sigma practices in Indian manufacturing companies', *International Journal of Quality & Reliability Management*, 33(8), pp. 1138-1149.

Antony, J. and Banuelas, R. (2002) 'Key ingredients for the effective implementation of Six Sigma program', *Measuring Business Excellence*, 6(4), pp. 20-27.

Ayyar, B.V., Arora, S., Murphy, C. and O'Kennedy, R. (2012) 'Affinity chromatography as a tool for antibody purification', *Methods*, 56(2), pp. 116-129.

Bade, P.D., Kotu, S.P. and Rathore, A.S. (2012) 'Optimization of a refolding step for a therapeutic fusion protein in the quality by design (QbD) paradigm', *Journal of Separation Science*, 35(22), pp. 3160-3169.

Bareither, R. and Pollard, D. (2011) 'A review of advanced small-scale parallel bioreactor technology for accelerated process development: Current state and future need', *Biotechnology Progress*, 27(1), pp. 2-14.

Bassi, L.J. (1997) 'Harnessing the power of intellectual capital', *Training & Development*, 51(12), pp. 25-31.

Bell, G., Pictet, R., Goodman, H.M. and Rutter, W.J. (1984) 'DNA transfer vector and transformed microorganism containing human proinsulin and pre-proinsulin genes'. Google Patents.

BioPharm (2014) *MAb manufacturing today and tomorrow*. [Online]. Available at: http://biopharmservices.com/wp-content/uploads/2014/05/Biopharm_Mab_Manufacturing_White_Paper_FINAL.pdf (Accessed: 09/07/2014).

Birch, J.R. and Racher, A.J. (2006) 'Antibody production', *Advanced Drug Delivery Reviews*, 58(5), pp. 671-685.

Biwer, A., Griffith, S. and Cooney, C. (2005) 'Uncertainty analysis of penicillin V production using Monte Carlo simulation', *Biotechnology and Bioengineering*, 90(2), pp. 167-179.

Block, H., Maertens, B., Spriestersbach, A., Brinker, N., Kubicek, J., Fabis, R., Labahn, J. and Schäfer, F. (2009) 'Immobilized-metal affinity chromatography (IMAC): a review', *Methods in Enzymology*, 463, pp. 439-473.

Bolton, G.R. and Mehta, K.K. (2016) 'The role of more than 40 years of improvement in protein A chromatography in the growth of the therapeutic antibody industry', *Biotechnology Progress*, 32(5), pp. 1193-1202.

Bommarius, A.S. and Riebel-Bommarius, B.R. (2007) *Biocatalysis: fundamentals and applications*. John Wiley & Sons.

Bornscheuer, U.T., Huisman, G.W., Kazlauskas, R.J., Lutz, S., Moore, J.C. and Robins, K. (2012) 'Engineering the third wave of biocatalysis', *Nature*, 485(7397), pp. 185-194.

Britest (2017a) *Biogen becomes Britest Associate*. Available at: https://www.britest.co.uk/news/biogen_britest_associate/ (Accessed: 11/03/17).

Britest (2017b) *Britest Value*. Available at: https://www.britest.co.uk/about_us/britest_value/ (Accessed: 11/03/17).

Burgess, A.A. and Brennan, D.J. (2001) 'Application of life cycle assessment to chemical processes', *Chemical Engineering Science*, 56(8), pp. 2589-2604.

Buss, P. and Ivey, N. (2001) *Simulation Conference, 2001. Proceedings of the Winter*. IEEE.

Calibration, U., Packings, I.-B., Packings, P.-B., Microcolumn, S.E.C. and Preparative, S.E.C. (1994) 'Size exclusion chromatography', *Anal. Chem*, 66, pp. 595R-620R.

Carroll, J. (2016) *The 15 top R&D spenders in the global biopharma business*. Available at: https://endpts.com/top-pharma-biotech-research-development-budgets/ (Accessed: 25/03/17).

Chapman, M.D., Vailes, L.D., Arruda, L.K., Melén, E. and Pomés, A. (1998) 'Source characterization and molecular structure of cockroach allergens', *Revue française d'allergologie et d'immunologie clinique*, 38(10), pp. 842-845.

Charaniya, S., Hu, W.-S. and Karypis, G. (2008) 'Mining bioprocess data: opportunities and challenges', *Trends in Biotechnology*, 26(12), pp. 690-699.

Chassy, B.M. and Giuffrida, A. (1980) 'Method for the lysis of Gram-positive, asporogenous bacteria with lysozyme', *Applied and Environmental Microbiology*, 39(1), pp. 153-158.

Chen, A., Chitta, R., Chang, D. and Amanullah, A. (2009) 'Twenty-four well plate miniature bioreactor system as a scale-down model for cell culture process development', *Biotechnology and Bioengineering*, 102(1), pp. 148-160.

Chen, G., Su, Z., Li, F. and Liu, H.F. (2017) 'Application of calcium phosphate flocculation in high-density cell culture fluid with high product titer of monoclonal antibody', *Bioprocess and Biosystems Engineering*, pp. 1-12.

Chen, J.Q., Zhang, H.T., Hu, M.H. and Tang, J.G. (1995) 'Production of human insulin in an E. coli system with Met-Lys-human proinsulin as the expressed precursor', *Applied Biochemistry and Biotechnology*, 55(1), pp. 5-15.

Cheng, C.-L., Che, P.-Y., Chen, B.-Y., Lee, W.-J., Lin, C.-Y. and Chang, J.-S. (2012) 'Biobutanol production from agricultural waste by an acclimated mixed bacterial microflora', *Applied Energy*, 100, pp. 3-9.

Chhatre, S., Farid, S.S., Coffman, J., Bird, P., Newcombe, A.R. and Titchener-Hooker, N.J. (2011) 'How implementation of quality by design and advances in biochemical engineering are enabling efficient bioprocess development and manufacture', *Journal of Chemical Technology and Biotechnology*, 86(9), pp. 1125-1129.

Citeline (2014) *Citeline Pharma R&D Annual Review*. Bloomberg (Accessed: 25/03/17).

Conklin, J.D. (2004) 'DOE and six sigma', *Quality Progress*, 37(3), p. 66.

CPI (2016) *CPI employee receives award for innovation*. Available at: https://www.uk-cpi.com/news/cpi-employee-receives-award-innovation/ (Accessed: 11/03/17).

Dassau, E., Zadok, I. and Lewin, D.R. (2006) 'Combining six-sigma with integrated design and control for yield enhancement in bioprocessing', *Industrial & engineering chemistry research*, 45(25), pp. 8299-8309.

DataMonitor (2016) *Company Pipeline Analysis*.

Datta, A. (2016) *Top 10 pharmaceutical companies in 2016* Available at: http://blog.proclinical.com/top-10-pharmaceutical-companies-2016 (Accessed: 25/03/17).

De Feo, J. and Bar-El, Z. (2002) 'Creating strategic change more efficiently with a new design for six sigma process', *Journal of Change Management*, 3(1), pp. 60-80.

De Feo, J.A. and Barnard, W. (2003) *Juran Institute's six sigma: breakthrough and beyond: quality performance breakthrough methods*. McGraw-Hill Professional.

De Lalla, C., Tamborini, E., Longhi, R., Tresoldi, E., Manoni, M., Siccardi, A.G., Arosio, P. and Sidoli, A. (1996) 'Human recombinant antibody fragments specific for a rye-grass pollen allergen: Characterization and potential applications', *Molecular Immunology*, 33(13), pp. 1049-1058.

De Ruyter, K. and Scholl, N. (1998) 'Positioning qualitative market research: reflections from theory and practice', *Qualitative Market Research: An International Journal*, 1(1), pp. 7-14.

De Vries, R.E., Van den Hooff, B. and de Ridder, J.A. (2006) 'Explaining knowledge sharing: The role of team communication styles, job satisfaction, and performance beliefs', *Communication Research*, 33(2), pp. 115-135.

Defernez, M., Foxall, R.J., O'Malley, C.J., Montague, G., Ring, S.M. and Kemsley, E.K. (2007) 'Modelling beer fermentation variability', *Journal of Food Engineering*, 83(2), pp. 167-172.

Delvigne, F., Zune, Q., Lara, A.R., Al-Soud, W. and Sørensen, S.J. (2014) 'Metabolic variability in bioprocessing: implications of microbial phenotypic heterogeneity', *Trends in Biotechnology*, 32(12), pp. 608-616.

Donovan, R.S.R., CW; Glick, BR (1996) 'Review: optimizing inducer and culture conditions for expression of foreign proteins under the control of thelac promoter', *Journal of Industrial Microbiology*, 16(3), pp. 145-154.

Duffy, J. (2000) 'The KM technology infrastructure', *Information Management*, 34(2), p. 62.

Edwards, M. (2011) *Gene Technology*. Lecture. Newcastle University.

Eldridge, N.E., Woods, S.S., Bonello, R.S., Clutter, K., Ellingson, L., Harris, M.A., Livingston, B.K., Bagian, J.P., Danko, L.H. and Dunn, E.J. (2006) 'Using the six sigma process to implement the Centers for Disease Control and Prevention Guideline for Hand Hygiene in 4 intensive care units', *Journal of General Internal Medicine*, 21(S2), pp. S35-S42.

Enders, D. and Jaeger, K.-E. (2007) *Asymmetric synthesis with chemical and biological methods*. John Wiley & Sons.

Ernst, M., Kaup, B., Müller, M., Bringer-Meyer, S. and Sahm, H. (2005) 'Enantioselective reduction of carbonyl compounds by whole-cell biotransformation, combining a formate dehydrogenase and a (R)-specific alcohol dehydrogenase', *Applied Microbiology and Biotechnology*, 66(6), pp. 629-634.

Faber, K. (2011) *Biotransformations in organic chemistry: a textbook*. 6th Revised and corrected edn. Berlin: Springer Science & Business Media.

Farid, S.S. (2007) 'Process economics of industrial monoclonal antibody manufacture', *Journal of Chromatography B*, 848(1), pp. 8-18.

FDA (2004) 'Guidance for industry: PAT—A framework for innovative pharmaceutical development, manufacturing, and quality assurance', *DHHS, Rockville, MD*.

FDA (2006) 'Guidance for industry: Q8 pharmaceutical development', *Rockville, MD*.

Fernandez, J.M. and Hoeffler, J.P. (1998) *Gene expression systems: using nature for the art of expression*. Academic Press.

Ford (2011) *Launching a Flawless Fiesta: Ford Motor Company*. Available at: http://www.minitab.com/en-us/Case-Studies/Ford-Motor-Company-DOE/ (Accessed: 19/03/17).

Franceschini, G. and Macchietto, S. (2008) 'Model-based design of experiments for parameter precision: State of the art', *Chemical Engineering Science*, 63(19), pp. 4846-4872.
G. Bell, R.P., H. Goodman, W. Rutter (1984) *DNA Transfer vector and transformed microorganism containing human proinsulin and pre-proinsulin genes*. Agency, U.S.P.
García-López, E., Siller, H.R. and Rodríguez, C.A. (2015) *Industrial Engineering and Engineering Management (IEEM), 2015 IEEE International Conference on*. IEEE.
Ghose, S., Hubbard, B. and Cramer, S.M. (2006) 'Evaluation and comparison of alternatives to Protein A chromatography: Mimetic and hydrophobic charge induction chromatographic stationary phases', *Journal of Chromatography A*, 1122(1), pp. 144-152.
Gilchrist, W. (1993) 'Modelling failure modes and effects analysis', *International Journal of Quality & Reliability Management*, 10(5).
Giridhar, A., Gupta, A., Louvier, M., Joglekar, G., Nagy, Z.K. and Reklaitis, G.V. (2014) *24th European Symposium on Computer Aided Process Engineering. Elsevier Science*.
Glassey, J., Gernaey, K.V., Clemens, C., Schulz, T.W., Oliveira, R., Striedner, G. and Mandenius, C.F. (2011) 'Process analytical technology (PAT) for biopharmaceuticals', *Biotechnology Journal*, 6(4), pp. 369-377.
Glassey, J., Montague, G. and Mohan, P. (2000) 'Issues in the development of an industrial bioprocess advisory system', *Trends in Biotechnology*, 18(4), pp. 136-141.
Glauche, F., Pilarek, M., Bournazou, M.N.C., Grunzel, P. and Neubauer, P. (2016) 'Design of experiments-based high-throughput strategy for development and optimization of efficient cell disruption protocols', *Engineering in Life Sciences*.
Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K. and Riggs, A.D. (1979) 'Expression in *Escherichia coli* of chemically synthesized genes for human insulin', *Proceedings of the National Academy of Sciences*, 76(1), pp. 106-110.
Goh, S.C. (2002) 'Managing effective knowledge transfer: an integrative framework and some practice implications', *Journal of Knowledge Management*, 6(1), pp. 23-30.
Gold, A.H. and Arvind Malhotra, A.H.S. (2001) 'Knowledge management: An organizational capabilities perspective', *Journal of Management Information Systems*, 18(1), pp. 185-214.
Gordon, J.W., Scangos, G.A., Plotkin, D.J., Barbosa, J.A. and Ruddle, F.H. (1980) 'Genetic transformation of mouse embryos by microinjection of purified DNA', *Proceedings of the National Academy of Sciences of the United States of America*, 77(12 II), pp. 7380-7384.
Grady Jr, C.P.L., Daigger, G.T., Love, N.G. and Filipe, C.D.M. (2011) *Biological wastewater treatment*. CRC press.

Graf, N., Mokhtari, T.E., Papayannopoulos, I.A. and Lippard, S.J. (2012) 'Platinum(IV)-chlorotoxin (CTX) conjugates for targeting cancer cells', *Journal of Inorganic Biochemistry*, 110, pp. 58-63.

Gronemeyer, P., Ditz, R. and Strube, J. (2014) 'Trends in upstream and downstream process development for antibody manufacturing', *Bioengineering*, 1(4), pp. 188-212.

Group, C.M.C.B.W. (2009) 'A-Mab: a case study in bioprocess development', *Emeryville, CA: CASSS*.

Guilfoyle, S. (2016) *Big Money in Big Pharma*. Available at: http://analyteguru.com/big-money-in-big-pharma/ (Accessed: 18/03/17).

Gupta, B., Iyer, L.S. and Aronson, J.E. (2000) 'Knowledge management: practices and challenges', *Industrial Management & Data Systems*, 100(1), pp. 17-21.

Gupta, R., Verma, P.L., Manoria, A. and Bajpai, L. (2016) 'Simplifying Six Sigma Methodology Using Shainin DOE', *Development*, 3(3).

Han, K., Lim, H.C. and Hong, J. (1992) 'Acetic acid formation in Escherichia coli fermentation', *Biotechnology and Bioengineering*, 39(6), pp. 663-671.

Harms, J., Wang, X., Kim, T., Yang, X. and Rathore, A.S. (2008) 'Defining process design space for biotech products: case study of Pichia pastoris fermentation', *Biotechnology Progress*, 24(3), pp. 655-662.

Harrison, R.G., Todd, P., Todd, P.W., Petrides, D.P. and Rudge, S.R. (2015) *Bioseparations science and engineering*. Oxford University Press, USA.

Harry, M.J. (1998) 'Six Sigma: a breakthrough strategy for profitability', *Quality progress*, 31(5), p. 60.

Hassan, T.E. and Heath, J.L. (1986) 'Biological fermentation of fish waste for potential use in animal and poultry feeds', *Agricultural wastes*, 15(1), pp. 1-15.

Herwig, C., Garcia-Aponte, O.F., Golabgir, A. and Rathore, A.S. (2015) 'Knowledge management in the QbD paradigm: manufacturing of biotech therapeutics', *Trends in biotechnology*, 33(7), pp. 381-387.

Hober, S., Nord, K. and Linhult, M. (2007) 'Protein A chromatography for antibody purification', *Journal of Chromatography B*, 848(1), pp. 40-47.

Hogwood, C.E.M., Tait, A.S., Koloteva-Levine, N., Bracewell, D.G. and Smales, C.M. (2013) 'The dynamics of the CHO host cell protein profile during clarification and protein A capture in a platform antibody purification process', *Biotechnology and Bioengineering*, 110(1), pp. 240-251.

I.C.H Guideline, H.T. (2008) 'Pharmaceutical quality system q10', *Current Step*, 4.

I.C.H Guideline, H.T. (2009) 'Pharmaceutical Development Q8 (R2)', *Current step*, 4.
IMSHealth (2016) *IMS Health Study: Global Market for Cancer Treatments Grows to $107 Billion in 2015, Fueled by Record Level of Innovation*. Available at: http://www.imshealth.com/en/about-us/news/ims-health-study-global-market-for-cancer-treatments-grows-to-107-billion-in-2015-fueled-by-record-level-of-innovation (Accessed: 18/03/2017).
Infineum (2011) *Using the Britest Tools to Troubleshoot Reaction Issues at Infineum*. Available at: https://www.britest.co.uk/case_studies/case_study_h/ (Accessed: 12/03/17).
Isett, K., George, H., Herber, W. and Amanullah, A. (2007) 'Twenty-four-well plate miniature bioreactor high-throughput system: Assessment for microbial cultivations', *Biotechnology and Bioengineering*, 98(5), pp. 1017-1028.
Jiang, C., Flansburg, L., Ghose, S., Jorjorian, P. and Shukla, A.A. (2010) 'Defining process design space for a hydrophobic interaction chromatography (HIC) purification step: application of quality by design (QbD) principles', *Biotechnology and Bioengineering*, 107(6), pp. 985-997.
Jiménez-González, C. and Woodley, J.M. (2010) 'Bioprocesses: Modeling needs for process evaluation and sustainability assessment', *Computers & Chemical Engineering*, 34(7), pp. 1009-1017.
Johnson-Matthey (2014) *Using Britest Tools to Deliver Johnson Matthey's Core Values*. Available at: https://www.britest.co.uk/case_studies/case_study_i/ (Accessed: 12/03/17).
Junker, B. (2010) 'Minimizing the environmental footprint of bioprocesses', *BioProcess Int*, 8(9).
Junker, B., Maheshwari, G., Ranheim, T., Altaras, N., Stankevicz, M., Harmon, L., Rios, S. and D'Anjou, M. (2011) 'Design-for-six-sigma to develop a bioprocess knowledge management framework', *PDA Journal of Pharmaceutical Science and Technology*, 65(2), pp. 140-165.
Kalil, S.J., Maugeri, F. and Rodrigues, M.I. (2000) 'Response surface analysis and simulation as a tool for bioprocess design and optimization', *Process Biochemistry*, 35(6), pp. 539-550.
Kamachi, Y. (2016) 'Development of hyper osmotic resistant CHO host cells'.
Kamionka, M. (2011) 'Engineering of therapeutic proteins production in Escherichia coli', *Current Pharmaceutical Biotechnology*, 12(2), pp. 268-274.
Karp, P.D., Riley, M., Paley, S.M. and Pelligrini-Toole, A. (1996) 'EcoCyc: an encyclopedia of Escherichia coli genes and metabolism', *Nucleic Acids Research*, 24(1), pp. 32-39.

Kelley, B. (2007) 'Very large scale monoclonal antibody purification: the case for conventional unit operations', *Biotechnology Progress*, 23(5), pp. 995-1008.
Kelley, B. (2009) *MAbs*. Taylor & Francis.
Kepert, J.F., Cromwell, M., Engler, N., Finkler, C., Gellermann, G., Gennaro, L., Harris, R., Iverson, R., Kelley, B. and Krummen, L. (2016) 'Establishing a control system using QbD principles', *Biologicals*, 44(5), pp. 319-331.
Kim, T.H., Lee, J.-N., Chun, J.U. and Benbasat, I. (2014) 'Understanding the effect of knowledge management strategies on knowledge management performance: A contingency perspective', *Information & Management*, 51(4), pp. 398-416.
Kleiner, D.E., Brunt, E.M., Van Natta, M., Behling, C., Contos, M.J., Cummings, O.W., Ferrell, L.D., Liu, Y.C., Torbenson, M.S. and Unalp-Arida, A. (2005) 'Design and validation of a histological scoring system for nonalcoholic fatty liver disease', *Hepatology*, 41(6), pp. 1313-1321.
Knepper, A.H., Michael; Glauche, Florian; Neubauer, Peter (2014) 'Robotic Platform for Parallelized Cultivation and Monitoring of Microbial Growth Parameters in Microwell Plates', *Journal of Laboratory Automation*, 19(6), pp. 593-601.
Koch, M. (2011) *Pat applied in biopharmaceutical process development and manufacturing: an enabling tool for quality-by-design*. CRC Press.
Koning, H., Verver, J.P.S., Heuvel, J., Bisgaard, S. and Does, R.J.M.M. (2006) 'Lean six sigma in healthcare', *Journal for Healthcare Quality*, 28(2), pp. 4-11.
Konstantinov, K.B. and Yoshida, T. (1992) 'Real-time qualitative analysis of the temporal shapes of (bio) process variables', *AIChE Journal*, 38(11), pp. 1703-1715.
Kumar, S. and Sosnoski, M. (2009) 'Using DMAIC Six Sigma to systematically improve shopfloor production quality and costs', *International Journal of Productivity and Performance Management*, 58(3), pp. 254-273.
Kumar, V., Bhalla, A. and Rathore, A.S. (2014) 'Design of experiments applications in bioprocessing: concepts and approach', *Biotechnology Progress*, 30(1), pp. 86-99.
Kumschick, S., Bacher, S., Evans, T., Markova, Z., Pergl, J., Pyšek, P., Vaes-Petignat, S., Veer, G., Vilà, M. and Nentwig, W. (2015) 'Comparing impacts of alien plants and animals in Europe using a standard scoring system', *Journal of Applied Ecology*, 52(3), pp. 552-561.
Leuchs, S. and Greiner, L. (2011) 'Alcohol dehydrogenase from Lactobacillus brevis: a versatile robust catalyst for enantioselective transformations', *Chemical and Biochemical Engineering Quarterly*, 25(2), pp. 267-281.

Lever, M.A., Torti, A., Eickenbusch, P., Michaud, A.B., Šantl-Temkiv, T. and Jørgensen, B.B. (2015) 'A modular method for the extraction of DNA and RNA, and the separation of DNA pools from diverse environmental sample types', *Frontiers in Microbiology*, 6, p. 476.
Liao, S.-h. (2003) 'Knowledge management technologies and applications—literature review from 1995 to 2002', *Expert Systems with Applications*, 25(2), pp. 155-164.
Liu, D., Chen, Y., Ding, F.-Y., Zhao, T., Wu, J.-L., Guo, T., Ren, H.-F., Li, B.-B., Niu, H.-Q. and Cao, Z. (2014) 'Biobutanol production in a Clostridium acetobutylicum biofilm reactor integrated with simultaneous product recovery by adsorption', *Biotechnology for Biofuels*, 7(1), p. 5.
Luli, G.W. and Strohl, W.R. (1990) 'Comparison of growth, acetate production, and acetate inhibition of Escherichia coli strains in batch and fed-batch fermentations', *Applied and Environmental Microbiology*, 56(4), pp. 1004-1011.
Machielsen, R., Looger, L.L., Raedts, J., Dijkhuizen, S., Hummel, W., Hennemann, H.G., Daussmann, T. and Van der Oost, J. (2009) 'Cofactor engineering of Lactobacillus brevis alcohol dehydrogenase by computational design', *Engineering in Life Sciences*, 9(1), pp. 38-44.
Mandenius, C.F. and Brundin, A. (2008) 'Bioprocess optimization using design-of-experiments methodology', *Biotechnology Progress*, 24(6), pp. 1191-1203.
Marchal, L., Zondervan, J., Bergsma, J., Beeftink, H. and Tramper, J. (2001) 'Monte Carlo simulation of the a-amylolysis of amylopectin potato starch', *Bioprocess and Biosystems Engineering*, 24(3), pp. 163-170.
McClusky, R. (2000) 'The rise, fall and revival of Six Sigma quality', *Measuring Business Excellence*, 4(2), p. 6.
McLachlan, K.J. (2016) *ESBES*. Dublin. Presentation.
McLaughlin, L.W. (1989) 'Mixed-mode chromatography of nucleic acids', *Chemical Reviews*, 89(2), pp. 309-319.
Mercier, S.M., Diepenbroek, B., Wijffels, R.H. and Streefland, M. (2014) 'Multivariate PAT solutions for biopharmaceutical cultivation: current progress and limitations', *Trends in Biotechnology*, 32(6), pp. 329-336.
Mollah, A.H. (2005) 'Application of failure mode and effect analysis (FMEA) for process risk assessment', *BioProcess International*, 3(10), pp. 12-20.
Montague, G.A., Martin, E.B. and O'Malley, C.J. (2008) 'Forecasting for fermentation operational decision making', *Biotechnology Progress*, 24(5), pp. 1033-1041.
Montgomery, D.C. (1991) 'Design and analysis of experiments'.

Motorola (2009) *About Motorola University:the inventors of Six Sigma*. Available at: http://web.archive.org/web/20051106025733/http://www.motorola.com/content/0,,3079,00.html (Accessed: 09/07/2014).

Müller, M., Wolberg, M., Schubert, T. and Hummel, W. (2005) 'Enzyme-catalyzed regio-and enantioselective ketone reductions', in *Technology Transfer in Biotechnology*. Springer, pp. 261-287.

Museum, M. (2015) *How is Marmite made?* Available at: http://www.marmitemuseum.co.uk/tag/production/ (Accessed: 25/03/17).

Nentwig, W., Bacher, S., Pyšek, P., Vilà, M. and Kumschick, S. (2016) 'The generic impact scoring system (GISS): a standardized tool to quantify the impacts of alien species', *Environmental monitoring and assessment*, 188(5), pp. 1-13.

Neubauer, P., Cruz, N., Glauche, F., Junne, S., Knepper, A. and Raven, M. (2013) 'Consistent development of bioprocesses from microliter cultures to the industrial scale', *Engineering in Life Sciences*, 13(3), pp. 224-238.

Neves, A.A., Vieira, L.M. and Menezes, J.C. (2001) 'Effects of preculture variability on clavulanic acid fermentation', *Biotechnology and Bioengineering*, 72(6), pp. 628-633.

Newton, B.A. (1956) 'The properties and mode of action of the polymyxins', *Bacteriological Reviews*, 20(1), p. 14.

Ngibuini, M. (2017) *Accelerate Your Process Development With High-Throughput, Single-Use, Fully Automated Bioreactors*. Available at: http://www.tapbiosystems.com/tap/cell_culture/media/ambr250_media/ssb_media/SU%20Auto%20Bioreactors%20NEW.pdf (Accessed: 12/03/17).

Nonaka, I. (1994) 'A dynamic theory of organizational knowledge creation', *Organization science*, 5(1), pp. 14-37.

Ochoa, J.L. (1978) 'Hydrophobic (interaction) chromatography', *Biochimie*, 60(1), pp. 1-15.

Ofek, E. and Sarvary, M. (2001) 'Leveraging the customer base: Creating competitive advantage through knowledge management', *Management Science*, 47(11), pp. 1441-1456.

Otasek, D., Pastrello, C., Holzinger, A. and Jurisica, I. (2014) 'Visual data mining: effective exploration of the biological universe', in *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics*. Springer, pp. 19-33.

Ouzounis, C.A. and Karp, P.D. (2000) 'Global properties of the metabolic map of Escherichia coli', *Genome Research*, 10(4), pp. 568-576.

Pan, S.L. and Scarbrough, H. (1999) 'Knowledge management in practice: An exploratory case study', *Technology Analysis & Strategic Management*, 11(3), pp. 359-374.

Pande, P.S., Neuman, R.P. and Cavanagh, R.R. (2000) *The six sigma way*. McGraw-Hill.
Panke, S., Held, M. and Wubbolts, M. (2004) 'Trends and innovations in industrial biocatalysis for the production of fine chemicals', *Current Opinion in Biotechnology*, 15(4), pp. 272-279.
Patil, A.S. and Pethe, A.M. (2013) 'Quality by Design (QbD): A new concept for development of quality pharmaceuticals', *Int J Qual Assur*, 4, pp. 13-9.
Patton, M.Q. (1999) 'Enhancing the quality and credibility of qualitative analysis', *Health Services Research*, 34(5 Pt 2), p. 1189.
Petrides, D., Carmichael, D., Siletti, C. and Koulouris, A. (2014) 'Biopharmaceutical process optimization with simulation and scheduling tools', *Bioengineering*, 1(4), pp. 154-187.
Petrides, D., Cruz, R. and Calandranis, J. (1998) 'Optimization of wastewater treatment facilities using process simulation', *Computers & Chemical Engineering*, 22, pp. S339-S346.
Petrides, D., Koulouris, A. and Siletti, C. (2002a) 'Throughput analysis and debottlenecking of biomanufacturing facilities', *Chimica Oggi*, 20(5), pp. 10-17.
Petrides, D.P., Koulouris, A. and Lagonikos, P.T. (2002b) 'The role of process simulation in pharmaceutical process development and product commercialization', *Pharmaceutical Engineering*, 22(1), pp. 56-65.
Pinto, I.F., Aires-Barros, M.R. and Azevedo, A.M. (2015) 'Multimodal chromatography: debottlenecking the downstream processing of monoclonal antibodies', *Pharm. Bioprocess*, 3, pp. 263-279.
Polanyi, M. and Sen, A. (2009) *The tacit dimension*. University of Chicago press.
Posch, A.E., Herwig, C. and Spadiut, O. (2013) 'Science-based bioprocess design for filamentous fungi', *Trends in Biotechnology*, 31(1), pp. 37-44.
Pyzdek, T. and Keller, P.A. (2014) *The six sigma handbook*. McGraw-Hill Education.
Raisinghani, M.S., Ette, H., Pierce, R., Cannon, G. and Daripaly, P. (2005) 'Six Sigma: concepts, tools, and applications', *Industrial Management & Data Systems*, 105(4), pp. 491-505.
Ranjan, A., Glassey, J., Montague, G. and Mohan, P. (2002) 'From process experts to a real-time knowledge-based system', *Expert Systems*, 19(2), pp. 69-79.
Rathore, A.S. (2009) 'Roadmap for implementation of quality by design (QbD) for biotechnology products', *Trends in Biotechnology*, 27(9), pp. 546-553.
Rathore, A.S. (2014) 'QbD/PAT for bioprocessing: moving from theory to implementation', *Current Opinion in Chemical Engineering*, 6, pp. 1-8.

Sadava, D.E., Hillis, D.M., Heller, H.C. and Berenbaum, M. (2009) *Life: the science of biology*. Macmillan.

Sakkas, V.A., Islam, M.A., Stalikas, C. and Albanis, T.A. (2010) 'Photocatalytic degradation using design of experiments: a review and example of the Congo red degradation', *Journal of Hazardous Materials*, 175(1), pp. 33-44.

Salazar Alvarez, A.J. (2003) 'Managing knowledge flows between high-tech firms and universities: empirical evidence from the biotechnology industry in the UK', *Journal of Information & Knowledge Management*, 2(04), pp. 309-319.

Saleh, M.S., Alrabiah, A. and Bakry, S.H. (2007) 'Using ISO 17799: 2005 information security management: a STOPE view with six sigma approach', *International Journal of Network Management*, 17(1), pp. 85-97.

Santoro, M.D. and Gopalakrishnan, S. (2001) 'Relationship dynamics between university research centers and industrial firms: Their impact on technology transfer activities', *The Journal of Technology Transfer*, 26(1-2), pp. 163-171.

Schild, A. and Fuchslueger, U. (2012) 'Introducing a paperless lab: how optimization and simultaneous electronic support of actual laboratory processes can boost efficiency', *Sci. Comput*, 29, pp. 7-11.

Schmid, A., Dordick, J., Hauer, B., Kiener, A., Wubbolts, M. and Witholt, B. (2001) 'Industrial biocatalysis today and tomorrow', *Nature*, 409(6817), pp. 258-268.

Schoemaker, H.E., Mink, D. and Wubbolts, M.G. (2003) 'Dispelling the myths--biocatalysis in industrial synthesis', *Science*, 299(5613), pp. 1694-1697.

Scott, J.E. (1998) 'Organizational knowledge and the intranet', *Decision Support Systems*, 23(1), pp. 3-17.

Shaeri, J., Wohlgemuth, R. and Woodley, J.M. (2006) 'Semiquantitative process screening for the biocatalytic synthesis of D-xylulose-5-phosphate', *Organic Process Research & Development*, 10(3), pp. 605-610.

Sharratt, P.N., Wall, K. and Borland, J.N. (2003) 'Generating innovative process designs using limited data', *Journal of Chemical Technology and Biotechnology*, 78(2-3), pp. 156-160.

Shinmyo, Y., Mito, T., Matsushita, T., Sarashina, I., Miyawaki, K., Ohuchi, H. and Noji, S. (2004) 'piggyBac-mediated somatic transformation of the two-spotted cricket, Gryllus bimaculatus', *Development Growth and Differentiation*, 46(4), pp. 343-349.

Shivhare, M. and McCreath, G. (2010) 'Practical considerations for DoE implementation in quality by design', *BioProcess Int*, 8(6), pp. 22-30.

Shukla, A.A., Hubbard, B., Tressel, T., Guhan, S. and Low, D. (2007) 'Downstream processing of monoclonal antibodies—application of platform approaches', *Journal of Chromatography B*, 848(1), pp. 28-39.
Shukla, A.A. and Thömmes, J. (2010) 'Recent advances in large-scale production of monoclonal antibodies and related proteins', *Trends in Biotechnology*, 28(5), pp. 253-261.
Siddh, M.M., Soni, G., Gadekar, G. and Jain, R. (2014) *Business and Information Management (ICBIM), 2014 2nd International Conference on*. IEEE.
Simunovic, M.P. (2010) 'Colour vision deficiency', *Eye*, 24(5), pp. 747-755.
Sinclair, A. and Monge, M. (2002) 'Quantitative economic evaluation of single use disposables in bioprocessing', *Pharmaceutical Engineering*, 22(3), pp. 20-34.
Smith, B. (2014) 'Sex is a good trick', *Phamaceutical Market Europe*, April 2014, p. 22.
Sommerfeld, S. and Strube, J. (2005) 'Challenges in biotechnology production—generic processes and process optimization for monoclonal antibodies', *Chemical Engineering and Processing: Process Intensification*, 44(10), pp. 1123-1137.
Stamatis, D.H. (2003) *Failure mode and effect analysis: FMEA from theory to execution*. ASQ Quality Press.
Stanton, D. (2015) *Bio-CMO sector to grow over 40% on Big Biopharma's burgeoning pipelines*. BioPharma-Reporter. [Online]. Available at: http://www.biopharma-reporter.com/Upstream-Processing/Bio-CMO-sector-to-grow-40-on-Big-Biopharma-s-burgeoning-pipelines.
Stanton, P. (2004) 'Ion-exchange chromatography', *HPLC of Peptides and Proteins: Methods and Protocols*, pp. 23-43.
Steinberg, D.M. and Bursztyn, D. (2010) 'Response surface methodology in biotechnology', *Quality Engineering*, 22(2), pp. 78-87.
Stewart, F., Leigh, S.E. and Thomson, J.A. (1987) 'Application of recombinant DNA techniques to structure-function studies of equine protein hormones', *Journal of Reproduction and Fertility. Supplement*, 35, pp. 1-8.
Stowell, F. (2013) 'The Appreciative Inquiry Method—A Suitable Candidate for Action Research?', *Systems Research and Behavioral Science*, 30(1), pp. 15-30.
Streefland, M., Van Herpen, P.F.G., Van de Waterbeemd, B., Van der Pol, L.A., Beuvery, E.C., Tramper, J., Martens, D.E. and Toft, M. (2009) 'A practical approach for exploration and modeling of the design space of a bacterial vaccine cultivation process', *Biotechnology and Bioengineering*, 104(3), pp. 492-504.

Studier, F.W. (2005) 'Protein production by auto-induction in high-density shaking cultures', *Protein Expression and Purification*, 41(1), pp. 207-234.
Takahashi, C.M., Takahashi, D.F., Carvalhal, M.L. and Alterthum, F. (1999) 'Effects of acetate on the growth and fermentation performance of Escherichia coli KO11', *Applied Biochemistry and Biotechnology*, 81(3), pp. 193-203.
Tamer Cavusgil, S., Calantone, R.J. and Zhao, Y. (2003) 'Tacit knowledge transfer and firm innovation capability', *Journal of Business & Industrial Marketing*, 18(1), pp. 6-21.
Teoh, S.K., Rathi, C. and Sharratt, P. (2015) 'Practical assessment methodology for converting fine chemicals processes from batch to continuous', *Organic Process Research & Development*, 20(2), pp. 414-431.
Turkay, C., Lundervold, A., Lundervold, A.J. and Hauser, H. (2013) 'Hypothesis generation by interactive visual exploration of heterogeneous medical data', in *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data*. Springer, pp. 1-12.
Tye, H. (2004) 'Application of statistical 'design of experiments' methods in drug discovery', *Drug Discovery Today*, 9(11), pp. 485-491.
Ukkonen, K.V., Antti; Ojamo, Heikki; Neubauer, Peter (2011) 'High-yield production of biologically active recombinant protein in shake flask culture by combination of enzyme-based glucose delivery and increased oxygen transfer', *Microb Cell Fact*, 10, p. 107.
Ündey, C., Ertunç, S., Mistretta, T. and Looze, B. (2010) 'Applied advanced process analytics in biopharmaceutical manufacturing: Challenges and prospects in real-time monitoring and control', *Journal of Process Control*, 20(9), pp. 1009-1018.
Valtorta, E., Martino, C., Sartore-Bianchi, A., Penaullt-Llorca, F., Viale, G., Risio, M., Rugge, M., Grigioni, W., Bencardino, K. and Lonardi, S. (2015) 'Assessment of a HER2 scoring system for colorectal cancer: results from a validation study', *Modern Pathology*.
Van Der Vossen, E.A.G., Gros, J., Sikkema, A., Muskens, M., Wouters, D., Wolters, P., Pereira, A. and Allefs, S. (2005) 'The Rpi-blb2 gene from Solanum bulbocastanum is an Mi-1 gene homolog conferring broad-spectrum late blight resistance in potato', *Plant Journal*, 44(2), pp. 208-222.
Van Tongeren, S.P., Degener, J.E. and Harmsen, H.J.M. (2011) 'Comparison of three rapid and easy bacterial DNA extraction methods for use with quantitative real-time PCR', *European Journal of Clinical Microbiology & Infectious Diseases*, 30(9), pp. 1053-1061.
Vose, D. (1996) *Quantitative risk analysis: a guide to Monte Carlo simulation modelling*. John Wiley & Sons.

Wakefield, R.L. (2005) 'Identifying knowledge agents in a KM strategy: the use of the structural influence index', *Information & Management*, 42(7), pp. 935-945.
Wall, K., Sharratt, P.N., Sadr-Kazemi, N. and Borland, J.N. (2001) 'Plant-independent process representation', *Organic Process Research & Development*, 5(4), pp. 434-437.
Wang, J. and Chow, S.-C. (2012) 'On the regulatory approval pathway of biosimilar products', *Pharmaceuticals*, 5(4), pp. 353-368.
Wang, L., Fan, D., Chen, W. and Terentjev, E.M. (2015) 'Bacterial growth, detachment and cell size control on polyethylene terephthalate surfaces', *Scientific Reports*, 5.
Weibull, C. (1953) 'The isolation of protoplasts from Bacillus megaterium by controlled treatment with lysozyme', *Journal of Bacteriology*, 66(6), p. 688.
Wen, Y.-F. (2009) 'An effectiveness measurement model for knowledge management', *Knowledge-Based Systems*, 22(5), pp. 363-367.
West, D. (1992) 'Knowledge elicitation as an inquiring system: towards a 'subjective'knowledge elicitation methodology', *Information Systems Journal*, 2(1), pp. 31-44.
WHO (2017) *Life Expectancy Data by Country*. Available at: http://apps.who.int/gho/data/view.main.SDG2016LEXv?lang=en (Accessed: 18/03/17).
Witcher, M.F. (2014) 'Failure Mode and Effect Analysis (FMEA) as a Quality by Design (QbD) Tool for Managing Biopharmaceutical Product Development and Manufacturing Risks', *BioProcessing Journal*, 13(3).
Wolberg, M., Hummel, W. and Müller, M. (2001) 'Biocatalytic Reduction of β, δ-Diketo Esters: A Highly Stereoselective Approach to All Four Stereoisomers of a Chlorinated β, δ-Dihydroxy Hexanoate', *Chemistry-A European Journal*, 7(21), pp. 4562-4571.
Woodcock, J. (2005) *AAPS workshop: Pharmaceutical Quality Assessment-A Science and Risk-based CMC Approach in the 21st Century, North Bethesda, MD*.
Wu, K.-H. and Tai, P.C. (2004) 'Cys32 and His105 are the critical residues for the calcium-dependent cysteine proteolytic activity of CvaB, an ATP-binding cassette transporter', *Journal of Biological Chemistry*, 279(2), pp. 901-909.
Xiang, Y., Liang, L., Wang, X., Wang, J., Zhang, X. and Zhang, Q. (2011) 'Chloride channel-mediated brain glioma targeting of chlorotoxin-modified doxorubicine-loaded liposomes', *Journal of Controlled Release*, 152(3), pp. 402-410.
Yang, T. and Hsieh, C.-H. (2009) 'Six-Sigma project selection using national quality award criteria and Delphi fuzzy multiple criteria decision-making method', *Expert Systems with Applications*, 36(4), pp. 7594-7603.

Zavascki, A.P., Goldani, L.Z., Li, J. and Nation, R.L. (2007) 'Polymyxin B for the treatment of multidrug-resistant pathogens: a critical review', *Journal of Antimicrobial Chemotherapy*, 60(6), pp. 1206-1215.

Zelenetz, A.D., Ahmed, I., Braud, E.L., Cross, J.D., Davenport-Ennis, N., Dickinson, B.D., Goldberg, S.E., Gottlieb, S., Johnson, P.E. and Lyman, G.H. (2011) 'NCCN biosimilars white paper: regulatory, scientific, and patient safety perspectives', *Journal of the National Comprehensive Cancer Network*, 9(Suppl 4), pp. S-1-S-22.

Zhang, L. and Mao, S. (2016) 'Application of quality by design in the current drug development', *Asian Journal of Pharmaceutical Sciences*.

Zhao, H. (2006) 'Tyler Johannes Michael R. Simurdiak'.

Zhou, Y.H. and Titchener-Hooker, N.J. (1999) 'Visualizing integrated bioprocess designs through “windows of operation”', *Biotechnology and Bioengineering*, 65(5), pp. 550-557.

Zoller, M.J. (1992) 'New recombinant DNA methodology for protein engineering', *Current Opinion in Biotechnology*, 3(4), pp. 348-354.

Zündorf, I. and Dingermann, T. (2001) 'Vom Rinder-, Schweine-, Pferde-Insulin zum Humaninsulin: Die biotechnische und gentechnische Insulin-Herstellung: Bereitstellung ausreichender Mengen von Humaninsulin', *Pharmazie in unserer Zeit*, 30(1), pp. 27-32.