Bound By
Blissett Bookbinders
020 8992 3965
www.blissetts.com

EVALUATION OF THE FINANCIAL AND TECHNICAL IMPACTS OF CHANGING COMMERCIAL-SCALE PHARMACEUTICAL MANUFACTURING PROCESSES

A thesis submitted to University College London in fulfilment of the requirements for the degree of:

Doctor of Engineering (EngD)

by

Sunil Chhatre

The Advanced Centre for Biochemical Engineering
University College London
Gower Street
London
WC1E 7JE

May 2008

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

Growing pressures in the pharmaceutical industry are driving the need to optimise processes used for the manufacture of drugs at commercial-scale, in order to improve cost of goods, product throughput and production times. Evaluating the impacts of process optimisation upon these metrics presents a challenge due to complexities and trade-offs that are often encountered when developing a typical bioprocess. Such factors have resulted in a range of novel simulation- and experimental-based techniques being developed which enable rapid, accurate and cost effective assessment of manufacturing options for commercial-scale production. This thesis proposes a combination of modelling and experimental methods for evaluating the business- and process-related impacts of implementing changes to pre-existing commercial-scale pharmaceutical manufacturing processes. The approaches are illustrated through an industrial case study, focusing upon a process operated by Protherics U.K. Limited for the manufacture of the FDA-approved rattlesnake anti-venom CroFab™ (Crotalidae Polyvalent Immune Fab (Ovine)).

The novel methods developed and illustrated in this thesis include:

- Investigating the effects of process changes upon calculated yields and processing times within the production framework for a pre-existing FDA-approved bio-manufacturing process
- Evaluating the impacts of both developing and implementing process changes, combining output metrics into a single value to simplify the assessment
- Developing a multi-layered simulation methodology for the rapid and efficient evaluation of bio-manufacturing process options
- Applying advanced sensitivity analysis techniques to identify the most critical factors that influence product yield and throughput
- Evaluating a novel synthetic Protein A matrix for the recovery and purification of polyclonal antibodies from hyperimmunised ovine serum
- Developing decision-support software to aid the design of chromatography steps for antibody purification at industrial scale
- Demonstrating the utility of such models by application to data and constraints derived from a full-scale industrial facility

ACKNOWLEDGEMENTS

University College London:

- Eli Keshavarz-Moore
- Nigel Titchener-Hooker
- Josh King
- Yuhong Zhou
- Suzanne Farid

Protherics U.K Limited (industrial sponsors):

- Anthony R. Newcombe
- Richard Francis
- Carl Jones
- Aline Denton
- Chrissie Cresswell
- Kieran O'Donovan

Millipore:

- Fred Mann (prototype adsorbent)

Funding:

- Engineering and Physical Sciences Research Council
- Royal Academy of Engineering
- Society of Chemical Industry
- University College London Graduate School

CONTENTS

LIST OF FIGURES

LIST OF TABLES

.

NOMENCLATURE AND ABBREVIATIONS

Nomenclature:

A_j	j^{th} Fourier sine coefficient
B_j	j^{th} Fourier cosine coefficient
B_{CIP}	Base case buffer volume (L)
b_1	Centrifugal caulk width (m)
C	Flowthrough concentration from a chromatography column (mol/m^3)
C_0	Feed concentration to a chromatography column (mol/m^3)
C_M	Concentration of mobile phase at equilibrium in batch mode
C_N	Number of operational cycles in a chromatography column (–)
C_S	Concentration of bound solute at equilibrium with the mobile phase in batch mode
c	Flowthrough concentration (mol/L) – used in the Beer-Lambert Law
D	Diafiltration ration (v/v)
$D_{i/wi}$	Variance due to a specific input variable or combination of variables
$D_{T/FAST}$	Total model variance
d	Particle diameter (µm)
d_c	Critical particle diameter – the smallest size of particle sedimented in the centrifuge (µm)
E	Extinction coefficient (M^{-1} cm^{-1}) – used in the Beer-Lambert Law
E_S	Efficiency (expressed as a fraction between 0 and 1) of process step S (-)
F_{no}	Number of filter housing (-)
f_1	Centrifugal correction factor
f_{sed}	Fraction of solids in the centrifugal input stream (-)
G	Dimensionless mobile phase flowthrough velocity (Thomas model)
$G(d)$	Grade efficiency i.e. the fraction of particles of size (d) sedimented in the centrifuge (–)
G_i	Transformation function used in FAST
g	Gravitational acceleration (m/s^2)
H	Chromatography bed height
$I_0(x)$	Zero order–modified Bessel function of the first kind
J	J function used in the Thomas model
k	Term in the centrifugal grade efficiency expression (–)
k_1	Forward rate constant – Thomas model (m^3/mol/s)
K	Term in the centrifugal grade efficiency expression (–)
K_d	Equilibrium dissociation constant – Thomas model (mol/m^3)
J	*J* function used in the Thomas model
L	Length of a chromatography column (m)
L_{IEX}	Unit loading of IgG to ion exchange per cycle (g/L)

l	Path length (cm) – used in the Beer-Lambert Law
M_A	Actual product mass (g F_{AB})
M_0	Lowest product mass set to represent the zero bound in normalisation (g F_{AB})
M_1	Highest product mass set to represent the one bound in normalisation (g F_{AB})
$M_{Fab,in}$	Inlet mass of F_{AB} to a process step (g)
$M_{Fab,out}$	Outlet mass of F_{AB} from a process step (g)
$M_{IgG,in,IEX}$	Mass of IgG fed to ion exchange (g)
$M_{IgG,in,papain}$	Mass of IgG fed into the digestion vessels (g)
$M_{VS.Fab,initial}$	Mass of venom-specific Fab present at the start of the process (g)
$M_{VS.Fab,final}$	Mass of venom-specific Fab present at the end of the process (g)
N	Dimensionless number of transfer units (Thomas model)
N_{cent}	Number of centrifugal cycles (-)
N_{IEX}	Number of ion exchange cycles (-)
N_R	Centrifugal revolution speed (rps)
N_i	Normalised value of the i^{th} performance metric (–)
N_m	Normalised product mass (–)
n	Number of disks in the centrifuge (-)
n_{LF}	Langmuir-Freundlich coefficient
P	Maximum harmonic considered in the FAST calculations (-)
q_m	Equilibrium saturation capacity of solid chromatographic resin (mol/m^3)
q_{ms}	Equilibrium saturation capacity of settled chromatographic resin (mol/m^3)
Q	Centrifugal volumetric flowrate (m^3/s)
Q_L	Linear column flowrate (m/hr)
$Q_{V,cent}$	Centrifugal volumetric flowrate (L/hr)
$Q_{v,filter}$	Volumetric flowrate of material in filtration (L/hr)
$Q_{V,UF}$	Volumetric flowrate in ultrafiltration (L/hr)
R_i	Inner disk radius (m)
R_o	Outer disk radius (m)
S_i	Sensitivity index
S_{T12}	Total Sensitivity Index for variables one and two i.e. $S_1 + S_2 + S_{12}$
T	Dimensionless residence time (Thomas model)
T_{conc}	Time for concentration in ultrafiltration (hr)
$T_{diafilter}$	Time for diafiltration in ultrafiltration (hr)
T_{flush}	Time for flushing depth filters (hr)
T_{prime}	Time for priming depth filters (hr)
T_{setup}	Time for setting up depth filters (hr)
t	Time (s or hr)
$t_{equilibrate}$	Duration of column equilibration (hr)

$t_{c\,/\,load}$	Duration of applying feed to a chromatography column (s/hr)
t_{elute}	Duration of elution (hr)
$t_{permeate}$	Duration of filtration (s)
$t_{sanitise}$	Duration of sanitisation (hr)
t_{wash}	Duration of bed washing (hr)
v	Interstitial velocity in a chromatography column (m/s)
$V_{activator}$	Volume of activator for enzymatic digestion (L)
v_{base}	Base case loading flowrate (m/s)
V_{Bed}	Packed bed volume (mL)
$V_{concentrate}$	Concentrate volume in ultrafiltration (L)
$V_{diafilter}$	Diafiltration buffer volume used in ultrafiltration (L)
V_F	Feed volume during ultrafiltration (L)
V_{flush}	Flush volume in ultrafiltration (L)
$V_{in,1\text{-}pass}$	Volume at the centrifuge inlet per cycle (L)
$V_{in,cent}$	Volume of centrifugal input stream (L)
$V_{in,filter}$	Feed volume into a filtration step (L)
$V_{in,IEX}$	Volume of ion exchange feed (L)
$V_{in,papain}$	Inlet feed volume to papain digestion (L)
$V_{in,pool}$	Inlet pool volume (L)
$V_{in,UF}$	Inlet volume to ultrafiltration (L)
V_L	Volume loaded up to 10% breakthrough (mL) – used in calculating DBC values
V_{load}	Volume of feed applied to a chromatography column per cycle (L)
V_M	Volume of chromatography matrix (L)
V_{matrix}	Packed bed volume (L)
$V_{matrix,a}$	Actual volume of chromatography matrix bed (L)
$V_{out,cent}$	Volume of centrifugal sedimented product (L)
$V_{out,UF}$	Outlet volume from ultrafiltration (L)
V_{papain}	Volume of papain (L)
$V_{permeate}$	Filtrate volume (m^3)
$V_{precip,1\text{-}pass}$	Precipitant volume used per centrifuge cycle for precipitate washing (L)
$V_{retentate}$	Retentate volume during ultrafiltration (L)
$V_{terminator}$	Volume of terminator for enzymatic digestion (L)
V_0	Void volume of chromatography bed (mL)
w_i	Weighting of the i^{th} performance metric (–)
X	$=C/C_0$ (normalised breakthrough fraction)
Y	Dimensionless separation factor (Thomas model)
y_{max}	Highest possible yield achieved with a fresh batch of matrix

Greek symbols:

α	Argument of the *J* function in the Thomas model
β	Argument of the *J* function in the Thomas model
χ	Parameter used in enzyme progress curve
γ	Yield from a process step (specific unit operation or quantity identified in text by subscript)
ε	Voidage fraction of a chromatographic bed (–)
ε_i	Bead porosity of a chromatographic bed (–)
φ	Parameter used in enzyme progress curve
Λ	Parameter used in enzyme progress curve
λ	Term in the centrifugal grade efficiency expression (–)
λ_{GSA}	Input value in the zero to one range (-)
λ_L	Real lower bound
λ_{Scaled}	Scaled input value (-)
λ_U	Real upper bound
μ	Viscosity of feed to the centrifuge (NS/m^2)
ω	Centrifugal angular velocity = $2\pi N_R$
ω_i	Integer frequency used in the FAST method
ρ_{solid}	Solids phase density (kg/m^3)
ρ_{liquid}	Liquid phase density (kg/m^3)
$\Delta\rho$	Density difference between solids and liquid phases in the centrifuge (kg/m^3)
Σ	Centrifugal sedimentation area (m^2)
τ_{column}	Throughput in a chromatography column (g product/hr *or* g product/min)
$\tau_{process}$	Throughput from a whole process (g product/hr *or* g product/min)
θ	Half disk angle in the centrifuge (rad)
θ_{total}	Total duration of column operation
ξ	Integration parameter used in the $I_0(x)$ function

Abbreviations:

AU	Absorbance Units
BPS	BioProcess Simulator
CDR	Complementary Determining Region
CF	Concentration Factor in ultrafiltration
C_{H1-3}	Constant regions 1-3 of the heavy antibody chains
CIP	Cleaning in Place
C_L	Constant regions of the light antibody chains
COG	Cost of Goods
cGMP	Current Good Manufacturing Practice
DBC	Dynamic Binding Capacity

DNA	Deoxyribonucleic acid
DOE	Design of Experiments
ELISA	Enzyme Linked Immunosorbent Assay
F_{AB}	Antibody fraction for antigen binding
FAST	Fourier Amplitude Sensitivity Test
F_C	Antibody fraction that can be crystallised
FDA	Food and Drug Administration
GA	Genetic Algorithm
GP	Genetic Program
GSA	Global Sensitivity Analysis
IEX	Ion exchange chromatography
IgG	Immunoglobulin G
IOD	Integrated Optical Density
LSA	Local Sensitivity Analysis
mAb	Monoclonal antibody
MADM	Multi-Attribute Decision Making
MEP	Mercapto-Ethyl-Pyridine
OR	Overall Rank (–)
pAB	Polyclonal antibody
PAGE	Polyacrylamide Gel Electrophoresis
PAT	Process Analytical Technology
PBS	Phosphate Buffered Saline
pGH	Porcine Growth Hormone
QBD	Quality By Design
QC	Quality Control
RMS	Root Mean Square
SI	Sensitivity Index
SDS	Sodium Dodecyl Sulphate
SPD	SuperPro Designer
UCL	University College London
UF	Ultrafiltration step
USD	Ultra Scale-Down
VBA	Visual Basic for Applications
V_H	Variable regions of the heavy antibody chains
V_L	Variable regions of the light antibody chains

1: INTRODUCTION

1.1 Challenges faced by the pharmaceutical industry

The growth in the number and variety of pharmaceuticals being developed as a result of advances in genetic engineering and cell culture technologies has enabled the commercialisation of many novel therapeutics (Karri et al., 2001a; King et al., 2004), providing significant benefits to patients and pharmaceutical companies alike. From the instant that drug candidates are discovered, pharmaceutical companies have always strived to develop and market them as soon as possible (Levin and Papageorgiou, 2004; Hausner, 1998), because up to US $1 million income is lost for every day that a product is not on the market (Basu et al., 1998; Lim et al., 2004). To this end, companies historically spent little time in optimising manufacturing scale processes early in development, in order to minimise time to market and so maximise the returns made when selling under the exclusivity afforded by a patent. Any process optimisation that was attempted was often delayed until the latter stages of clinical trials, when the probabilities of drug success were sufficiently high to justify capital investment (Farid et al., 2000a). Due to high cash burn rates and with little time left before product launch, this typically meant that very few improvements were actually implemented (Mustafa et al., 2004), often resulting in technically and financially sub-optimal operation at large scale. Nevertheless, such strategies were often acceptable in the past because companies were able to command significant premiums for their drugs, resulting in revenues that far exceeded manufacturing costs (Pisano et al., 1995).

Recently, the pharmaceutical industry has experienced unprecedented commercial and regulatory pressures (Shah, 2004) that threaten the financial stability of the industry. The growing complexity of potential new therapeutics (e.g. genes or whole cells), the drive to apply current Good Manufacturing Practice (cGMP) to products early in development, longer times to market and more extensive clinical burdens have all driven up development costs and times (Drews and Ryser, 1997; Karri et al., 2001b; Rooney et al., 2001; Titchener-Hooker et al., 2001; Schmid and Smith, 2002). New therapeutics may cost US $500 - 800 million or more to develop and take upto 15 years to arrive in the marketplace (Bauer and Fischer, 2000; DiMasi et al., 2003), in part driven by the growing complexities of manufacturing technologies (Lim et al., 2004) and time required to complete clinical programs. For drugs that *are* successfully commercialised, shorter effective patent lives, greater worldwide competition, generic substitution and reimbursement pressures all jeopardise their profitability (Tranter, 1999; Terzi and Cavalieri, 2004; Deeds et al., 2000; Doblhoff-Dier, and Bliem, 1999; Karri et al., 2001 a/b). Such problems are compounded by the high attrition rate of drugs, with 75% of development costs associated with the expense of previous failures (Clewlow, 2004). All of these factors point towards a growing need to accelerate process development studies, maximise yields, improve resource utilisation and reduce cost of goods (Farid et al., 2000b).

1.2 Process development strategies

To achieve these goals, a number of manufacturing strategies are available which demonstrate potential to reduce manufacturing costs and times whilst boosting product throughput. The purpose of this thesis is to explore the challenges involved in adopting these changes for processes that are already used for commercial production. Examples of strategies being explored industrially include:

Techniques	Example	Potential considerations
Increasing initial product concentration	Fermentation and cell culture product titres have improved significantly resulting in far greater quantities of product being subjected to downstream purification	Although titre improvements potentially enable an increase in the amount of final product, it also increases the probability of bottlenecks appearing in recovery and purification
Consolidating multiple steps into a single step	Precipitation and centrifugation steps used for the initial capture of an antibody product could be replaced with a single chromatography column step	Any possible reduction in processing times and costs need to be balanced against the need for process re-validation
Replacing one unit operation with another	Replacing a centrifuge used for initial cell harvesting with microfiltration	The lower shear environment in filtration may provide advantages over centrifugation when processing delicate biological materials
Removing unit operations	Removing one or more purifying chromatography columns from a sequence of such steps	The cost and time advantages would need to be balanced against the added purification burden presented to operations further downstream

Table 1: Examples of strategies considered for their potential to achieve improvements to commercial-scale manufacturing processes

The challenge of evaluating these options has driven the development of methodologies for the rapid, inexpensive assessment of the likely financial and technical performance of production processes. These techniques include ultra scale down (USD), which facilitate rapid early stage evaluation of manufacturing options with millilitre quantities of process material (Varga et al., 1998; Boychyn et al., 2000) and bioprocess manufacturing simulations, which have attracted recent interest (Ungarala and Co, 2000; Lim et al., 2004) for their ability to evaluate and compare novel process routes (Petrides et al., 1989; Rouf et al., 2001; King et al., 2004; Farid et al., 2005; Chhatre et al., 2006). This thesis employs a combination of modelling and USD experimental techniques to evaluate the impacts of implementing changes such as those given in Table 1 to pre-existing

commercial manufacturing operations and illustrates the utility of these approaches by an industrial case study. A summary of these techniques is presented below.

1.3 Simulation-based approaches to evaluating process changes

1.3.1 Advantages of bioprocess modelling

Mathematical simulations provide a potentially valuable design tool for the bioprocess industry (Zhou et al., 1997) for reducing the times and costs of both process development and product manufacture (Titchener-Hooker et al., 2001; Karri et al., 2001a). Models which address both business *and* process aspects enable an engineer to answer technical and financial questions simultaneously (e.g. identifying the effects of changing column operating conditions on product yield as well as resource utilisation – Farid et al., 2000a). The benefits of employing such models are listed in Table 2 (Page 26); ultimately, these advantages enable a company to identify the best route(s) for optimising a process and to get it right the first time around (Venkatasubramanian, 2005). To put these advantages into context, the next section describes the development of bioprocess simulations to date and the current status of the field.

1.3.2 Current status of the bioprocess simulation field

The use of modelling software has increased rapidly within the manufacturing sector as a whole, driven by a growth in the size, cost and intricacy of projects and the need to achieve success the first time around (Giannasi et al., 2001). Despite the aggressive deployment of simulators for process design in sectors such as polymers or chemicals (Lim et al., 2004), its acceptance in the bioprocessing industry has been noticeably slower (Shanklin et al., 2001; Mustafa et al., 2006) and the technique remains underdeveloped for the description of industrial scale bio-manufacturing processes due to the following factors:

- Many unit operations are only poorly characterised and involve complex biophysical phenomena, meaning that the mathematical equations used to describe their behaviour only provide approximations of likely performance
- Products can include proteins, DNA or whole cells, the properties of which may be unknown or variable from batch to batch. Default values are often used in such cases, making it difficult to construct robust simulations
- The lack of available process data often results in simulations relying on, 'default,' input values that may vary significantly from reality
- Mainstream simulation packages generally model continuous processes, whereas bioprocesses predominantly operation in batch mode
- The impacts of interactions between process steps are often not incorporated into simulations

Focus laboratory and pilot scale development on the most worthwhile options and so allow a company to reduce the total number of studies required (Petrides et al., 2002a). Furthermore, options can be assessed more quickly and cheaply on a computer than by experimentation
Evaluate the likely capital investment and production costs of a manufacturing process *in silico* before committing resources to a project (Lim et al., 2004)
Compare different flowsheet choices to determine which one meets required goals, such as on grounds of cost effectiveness (King et al., 2004; Mustafa et al., 2004)
Quantify the impact of a process change, so providing justification for its implementation (Watson et al., 2000)
Provide a rationale for capital investment or process optimisation, ultimately facilitating fast and efficient process development (Ernst et al., 1997)
Determine the impacts of changes to one unit operation and the knock-on effects downstream on yields, costs, times and throughputs (Pisano, 1996; Groep et al., 2000). This is especially useful where the impacts of process change are unexpected or counter-intuitive (Williams, 1999)
Calculate resource utilisation (Workman, 2000) in order to eliminate bottlenecks and identify the best way to make more efficient use of limited resources (Petrides et al., 2002a)
Introduce a common language to facilitate communication between different groups such as primary recovery and purification (Petrides, 1994) or within research and development to achieve commercialization (Lim et al., 2005)
Allow investigation of an experimental design in order to shed light on the assumptions underlying a model (Holford et al., 2000)

Table 2: Advantages of bioprocess modelling

1.3.3 Commercially developed bio-simulation software

These restrictions have meant that (bio-) pharmaceutical simulations have only developed within the last two decades, with the majority of publications on the subject coming from the creators of commercially available software (Lim et al., 2004). There are two main packages in existence – BioProcess Simulator (AspenTech) and SuperPro Designer (Intelligen). BioProcess Simulator (BPS) contains an extensive bank of unit operations models which can be used to simulate a bioprocess flowsheet – from ancillary functions such as mixing and splitting modules, sterilising systems and heat exchanger systems, through to primary unit operations such as fermentation, centrifugation, ultrafiltration and chromatography. Petrides et al. (1989) used the package to model the manufacture of porcine growth hormone (pGH) and assess the financial viability of the proposed flowsheet with respect to capital and production costs. A critical drawback of the package is that the costing calculations are based on the chemical engineering simulation package Aspen Plus (Rouf et al., 2001), meaning that BPS fails to consider the higher expenses imposed by stringent regulatory requirements.

SuperPro Designer (SPD) works in a similar fashion, with a series of unit operations models represented graphically by icons that are cloned to a workspace and connected together in the required sequence to simulate the flow of materials between stages (Petrides et al., 2002b). Economic evaluation is achieved through a costing module that is specific to bioprocesses, making SPD superior to BPS in this respect (Rouf et al., 2001). For example, Fee (2001) used SPD to calculate the relative economic merits of different wash strategies during the expanded bed-based recovery of a protein product.

Despite their potential advantages, these commercial packages still present distinct problems which hamper their deployment for the routine design and development of bio-manufacturing processes:

- Such packages are static i.e. they do not account for the dynamic resource allocation over the time course of a simulation as dependent upon availability and demand by other parts of the process (Lim et al., 2004)
- The lack of rigorous unit operation mass balance models (with little or no capability to customise the in-built equations or enter user-defined equations and data) limits the ability of SPD or BPS to make accurate predictions (Shanklin et al., 2001). The resulting over-simplifications can often mean that little detailed process insight can be achieved
- These programs only tend to consider capital and manufacturing throughput and generally neglect costs and times involved in developing and implementing process improvements (Baker and Wheelwright, 2004)
- A review of the literature indicates that published studies have mainly focused on hypothetical case studies, thus neglecting the complexities and constraints imposed upon real industrial processes

1.3.4 Academic bio-simulation software

The current generation of bioprocess simulations produced by academic research groups have predominantly focused on individual unit operations alone (Pinto et al., 2001; Groep et al., 2000), often involving the solution of complex ordinary or partial differential equations to evaluate material balances, such as for fermentation (Gebicke et al., 1993; Jang and Barford, 2000; Chowdhury et al., 2003), homogenisation (Siddiqi et al., 1996; Wong et al., 1997), primary recovery such as filtration or centrifugation (Maybury et al., 1998; Shanklin et al., 2001; Chan et al., 2003; Reynolds et al., 2003) and chromatography (Li et al., 1998; Fee, 2001). By definition, models of individual unit operations fail to consider the likely interactions between process steps and although they constitute a useful first step toward the production of robust and predictive modelling packages, it is the simultaneous consideration of all operations and replication of the interactions between them that is needed for the assessment of whole process feasibility. The majority of models available only consider interactions in a qualitative, rather than quantitative fashion (Samsatli and Shah, 1996). Unit operations do not operate in isolation and even small changes at any given stage can dramatically affect operation there and further downstream. If a researcher purely optimises individual operations separately, with respect to the objectives defined for each stage alone, then the performance of the overall process is likely to be suboptimal (Groep et al., 2000). For the simulation to be beneficial, a completely interactive process flowsheet needs to be considered. Advances in this area so far have centred primarily on establishing modelling methodologies through the study of generic process flowsheets (Zhou and Titchener-Hooker, 1999a; 1999b) e.g. evaluation of the relative merits of expanded bed capture of a fermentation product as compared to the more conventional packed bed route (Mustafa et al., 2004) and the comparison between different process options for the manufacture of monoclonal antibodies (Lim et al., 2005). The methods developed in this thesis to overcome these problems are outlined in section 1.6 and are illustrated through an industrial case study described in the next section.

1.4 Simulations for assisting with regulatory concepts for design and manufacture

Growing emphasis is now being placed on using approaches such as Quality By Design (QBD) which employs DOE to achieve thorough process characterisation and thereby gain valuable understanding of which parameters are especially important in controlling a process and allowing it to satisfy product requirements. This information can be particularly useful in uncovering interactions between steps that can complicate the choice of the best operating conditions and utimately, this technique can help to achieve the most robust manufacturing strategy that can tolerate process variations without compromising on product quality. Nevertheless, although systematic application of the QBD philosophy provides useful data, time pressures within a typical drug development programme can make it difficult to carry out such a complete process characterisation. To circumvent this issue in the past, heuristics based on previous experience may have been used to identify an operating location, which may have met company and market goals but may not

necessarily have been optimal in a financial or scientific sense. In light of the growing commercial and regulatory pressures in this area indicated above, a more rigorous QBD-based evaluation of the design space is becoming increasingly necessary to allow production to match economic and technical expectations more closely.

Successful implementation of QBD will require methods for rapidly identifying the most critical process parameters and then characterising the behaviour of the process as a function of those parameters. Simulations can help to achieve these objectives by allowing a user to explore the design space rapidly and so gain insight into the likely shape of the response curve, determining where the flat regions are that enable robust operation and also where edges of failure exist in a process. The knowledge gained from these studies can then prompt fractional factorial design experiments that concentrate upon the most relevant variables alone in order to investigate robust process regions in depth and work out the most suitable way to achieve satisfactory product quality in a reproducible, controlled fashion.

The understanding provided by models can also help with the implementation of other regulatory initiatives such as Process Analytical Technology (PAT), which use real time process testing and manipulation of, 'critical quality and performance attributes,' (http://www.emea.europa.eu/) to counteract variations in feed or process properties. At present, protocols for operating processes are fixed towards the end of clinical trials (end of phase II or the start of phase III), meaning that it can be difficult to accommodate variability and leading to potentially unacceptable inconsistency in the quality and quantity of the product. Knowledge gained from modelling-driven DOE studies can help to suggest PAT strategies that can counteract the variation and so allow a process to recover acceptable product characteristics. For example, if the titres from a monoclonal antibody cell culture vary over a certain range, then DOE studies can be undertaken to investigate the how variables such as the flowrate in a subsequent Protein A column need to be adjusted across that titre range in order to prevent breakthrough on the column and thereby achieve a consistent antibody recovery.

1.5 Industrial case study

1.5.1 Introduction

To provide industrial validity to the methods proposed in this thesis, their utility is illustrated by application to a process operated by Protherics U.K. Limited (Blaenwaun, Ffostrasol, Llandysul, Wales, U.K.) for the manufacture of the FDA-approved rattlesnake anti-venom CroFab™ (Newcombe and Newcombe, 2006). Approximately 100,000 people are killed by snake bites every year worldwide and antibody-based therapeutics purified from animal sera offers the best treatment modality at present (Boyer et al., 2001; Neal et al., 2003), due to the high specificity of binding between the venom and antibody molecules and the complex number of toxins present within snake

venom. For example, rattlesnakes cause the largest number of snakebites and resultant fatalities in the United States of America (Juckett and Hancox, 2002) and hence there is a clear need for an effective treatment for both the local and systemic effects of rattlesnake envenomation. CroFab™ is an antibody fragment purified from ovine serum obtained from sheep that have been immunised with venom extracted from rattlesnakes. The product binds venom molecules and enables their distribution away from target tissues and hence their subsequent elimination (Neal et al., 2003). To put the characteristics and behaviour of this product into context, the next section discusses antibody generation through the immune response and the methods employed industrially for their purification.

1.5.2 The immune response and natural antibody manufacture

The immune system in humans and species employed to defend against attack by foreign particles can be divided into two groups: (1) the innate element, which consists of those parts which an individual has at birth and (2) the acquired element, which is more specialised and results from contact with the invader (immunisation), activating cells that manufacture bio-molecules to destroy the foreign agent (Abbas and Lichtman, 2004). Characteristics of acquired immunity include:

- Tailored: the response is specific to each foreign agent
- Adaptive: a response can be mounted against previously unknown molecules
- Discriminatory: the system distinguishes between 'self' and 'non-self' cells to target the response
- Anamnestic: the system retains knowledge of previous contact, generating an ability to resist subsequent exposure and responding more rapidly and extensively

The cells involved in acquiring immunity are lymphocytes and fall into two groups – B and T cells and the specificity of their immune response is explained by the clonal selection theory (Skerra, 2003):

- Self reactive lymphocytes that respond against self cells are eliminated during early development, allowing all reactivity to be focused on invaders
- T/B lymphocytes exhibit a very wide repertoire of specificities prior to association with the foreign agent, enabling a response that is specific to potentially unseen non-self entities
- Foreign bodies possess characteristic structures called antigens which are used by the immune system to recognise them as non-self. Lymphocytes bind to the antigen and hence the cells become activated and releases bio-active products
- Each lymphocyte is specific to one binding site on the antigen

The bio-active products manufactured by B cells are antibodies, which are glycosylated proteins that circulate in the bloodstream and bind to the antigen of non-self particles, inducing their neutralisation and ultimate agglutination and disintegration (Bailey and Ollis, 1986). Antibody binding occurs through a complementary 'lock and key' fit (Roque et al., 2004) with a portion of an antigen called an

epitope, by a range of non-covalent interactions such as hydrogen, ionic and hydrophobic bonds as well as London Dispersion and dipole-dipole forces. The epitope is typically a few amino acids long and there may be several on each antigen. Each foreign agent induces antibody production against each epitope of each antigen, resulting in a mixed antibody population termed a polyclonal response. The structure of an antibody facilitates its action (adapted from Coico et al., 2003):

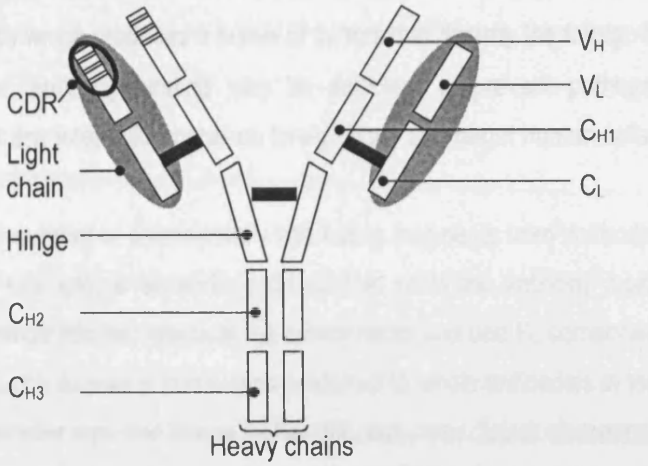


Figure 1: Antibody structure
V = variable; C = constant; H = heavy; L = light. A complementarity determining region (CDR) is shown in the V_H part of the molecule. The thick black bars represent interchain disulphide bonds.

Antibodies consist of two heavy chains and two light chains, with interchain disulphide bonds providing stability to the molecule. There are also intrachain disulphide bonds, around which the chains fold into compact globular domains possessing homologous sequences; light chains have two domains (one in the variable region, one in the constant region), whilst heavy chains have one in the variable region and three in the constant region (Farid, 2001). Hydrophobic interactions between certain domains also further stabilise the molecule (V_H-V_L; C_{H1}-C_L; C_{H2}-C_{H2}; C_{H3}-C_{H3}). The greatest variability between different antibodies is found in three regions of the N-terminal 110 amino acid sequence of both light and heavy chain of the first domain (V_L and V_H) – Coico et al. (2003). These areas are said to be hypervariable and since they are complementary to the structure of the epitope, are called complementary determining regions (CDR). Differences in CDRs between different antibodies provide for their antigen binding specificities (Brekke and Sandlie, 2003). It should also be noted that the hinge region between the first and second CH provides flexibility in opening up the two diagonal arms to bind to epitopes or antigens that may be some distance away from each other.

In addition to hypervariability in the N-terminal region, there are also a number of different types (isotypes) of antibodies, which are normally classified by the type of heavy chain (γ, α, μ, δ, ε) and light chain (κ, λ). The different heavy chain antibodies are called immunoglobulin isotypes – each with the same epitope specificity but a different biological function – and are referred to IgG, IgA, IgM, IgD and IgE, with IgG being the most prevalent (~80%). These differ according to the structure

in the constant domains of the heavy chains, such as with respect to their glycosylation patterns (Watt et al., 2003). It is this part of the molecule that mediates biological functions such as:

- Neutralisation of toxins
- Immobilisation and agglutination of organisms, which result in formation of precipitates that are readily eliminated by phagocytosis (Mackean and Jones, 1985; Davies, 1997)
- Triggering the complement cascade – a system of more than thirty proteins linked through an enzymatic sequence which produces a series of factors that destroy the foreign body
- For large particles, antibody binding may be sufficient to prevent pathogenic action e.g. adhesion can block any interaction between foreign body and target human cells

It is possible to obtain a number of commercially interesting fragments from antibodies by enzymatic digestion. Papain, for example, is an endopeptidase that splits the antibody molecule above the disulphide bond at the hinge into two identical F_{AB} components and one F_C component (Porter, 1959; Aubrey et al., 2004). F_{AB} on its own is sometimes preferred to whole antibodies in terms of biological activity because of its smaller size and hence higher diffusion rates, faster clearance from circulation and reduced immunogenicity (Coleman and Mahler, 2003). In the case of CroFab™, the therapeutic activity of the anti-venom derives entirely from the Fab region; removing the Fc component reduces potential adverse immunogenicity and enables the smaller Fab fragment to leave the blood stream more quickly than intact IgG (Neal et al., 2003). Hence, IgG digestion by papain is used by Protherics for CroFab™ manufacture.

1.5.3 Therapeutic antibodies

Monoclonal antibody-based (mAb) therapies, containing immunoglobulins with only one specificity, have made many valuable clinical contributions (Farid, 2007), for conditions ranging from cancer and infectious diseases through to the treatment of autoimmune responses (Torphy, 2002). A number of monoclonal antibodies have received regulatory approval and have been commercialised including Rituxan (non-Hodgkins lymphoma), Herceptin (breast cancer) or Zenapax (kidney transplant rejection) with many more in the pipeline (Shukla et al., 2007). As such, antibodies constitute one of the most important classes of biotherapeutics in development and production (Flatman et al., 2007) – a trend that is likely to continue for the foreseeable future.

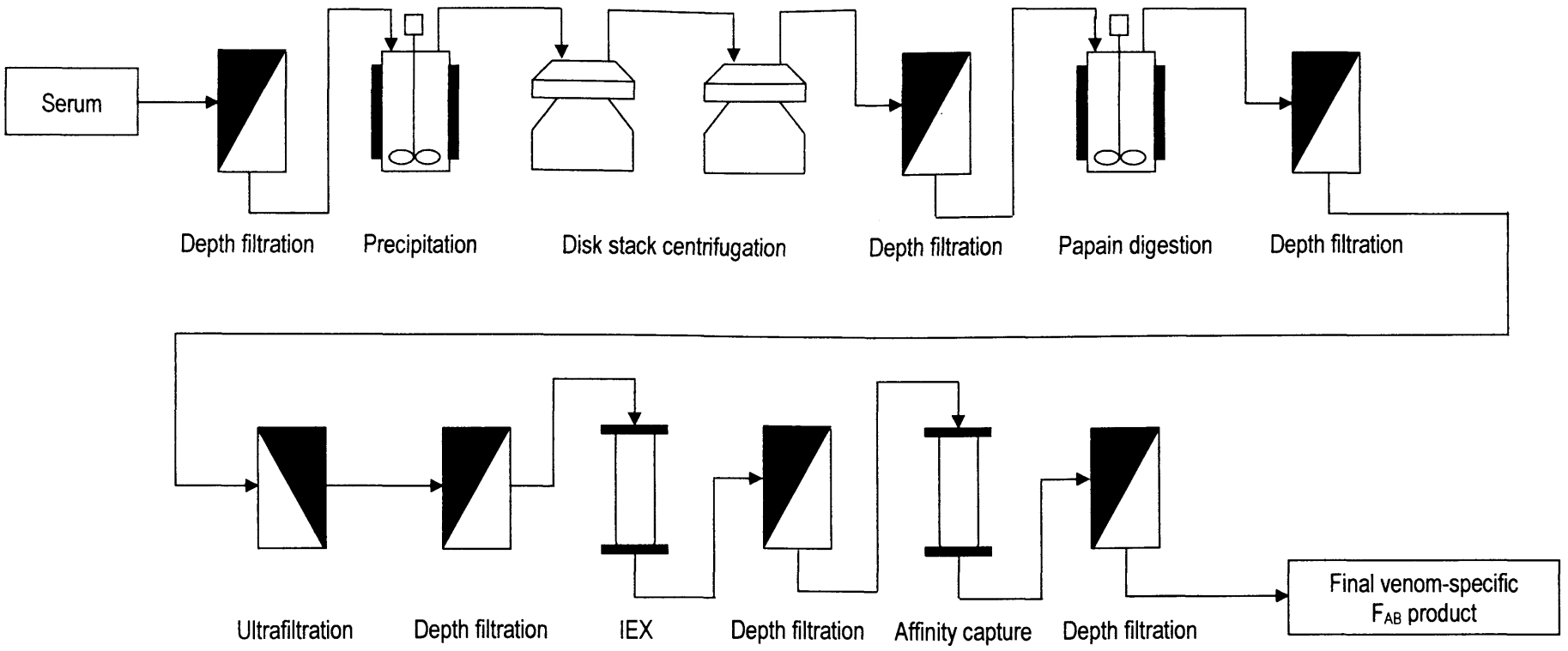
Although the majority of antibodies of commercial interest are monoclonal, polyclonals (pAbs) nevertheless also constitute an important class of therapy for neutralising toxins and overcoming immunoglobulin deficiencies (Newcombe and Newcombe, 2007). As outlined above, a pAb population is the natural result in animals or humans following infection or immunisation resulting from the simultaneous activation of the clones from many different B lymphocytes. A pAb mixture therefore consists of many different antibodies that have been raised against a range of epitopes on each antigen on a foreign molecule. Although it is technically possible to manufacture all of these

antibodies as separate monoclonals and mix them together, this would require multiple parallel culture systems for each mAb, which would be time consuming to set-up, difficult to control and highly capitally intensive. Furthermore, each separate antibody would require its own quality control, quality assurance and validation (Newcombe and Newcombe, 2007). These regulatory and financial obstacles makes polyclonal antibodies purified from animal sera more attractive and since there are many different antibody types in a pAb mixture, this also potentially offers greater efficacy than a mAb that is specific to one epitope alone, therefore enhancing the therapeutic value of a product such as CroFab™.

1.5.4 Details of the commercial manufacturing process

The process flowsheet used for CroFab™ manufacture is shown in Figure 2. Preparation of the antivenom product starts with hyperimmunisation of sheep. All flocks are located in Australia so as to comply with the regulatory requirement that serum-derived products can be certified as free of transmissible spongiform encephalopathies. Four different rattlesnake (Crotalidae) venoms are purified – C. atrox (CCT), C. adamanteus (CCD), C. scutulatus (CCS) and Agkistrodon piscivorus (CAP). After approximately 26 weeks, when sufficient venom-specific IgG levels have built up, blood is collected by cannulation of the jugular vein, clotted, depth filtered and frozen prior to shipping to the U.K. Following thawing and pooling of the ovine serum, the material is depth filtered to remove any cryoprecipitates formed during freezing and thawing the bulk. Fractional precipitation by sodium sulphate is then used to sediment the antibodies (typical IgG purities at this stage are greater than 85% – Newcombe et al., 2005) and the flocs are separated from the contaminating soluble phase by disk stack centrifugation. The sediment is washed in phosphate buffered saline and sodium sulphate and subjected to centrifugation, resulting in antibodies recovery in the solids phase again. Depth filtration clarifies the process stream and thus prepares it for enzymic digestion by papain in which the immunoglobulin molecule is broken at its hinge region into its F_{AB} and F_C components. Following depth filtration, ultrafiltration is used in concentration and diafiltration modes to reduce the volume of the in process material and reduce the salt concentration respectively in preparation for the next step. Ion exchange chromatography removes the F_C fragment and other minor impurities such as residual albumin. Finally, the product fraction is purified by an affinity capture step, in which the resin consists of snake venom molecules immobilised to a glass base matrix. This is used to recover the efficacious F_{AB} fragments and separate them from the non-specific F_{AB} present at the start of the process. The process shown in Figure 2 is run four times to isolate the different anti-venom F_{AB} species, which are then blended together in equal amounts to produce the final material, concentrated and filtered to produce the final pre-formulation drug substance. Pooling samples from the four different antibody types creates a broad spectrum treatment applicable to a range of different bites (Neal et al., 2003). After viral filtration of this solution, diafiltration into phosphate buffer saline, depth filtration and concentration, the material is then shipped backed to USA for lyophilisation and final distribution.

Figure 2: CroFab™ process flowsheet



1.6 Unit operation theory

1.6.1 Introduction

The following describes the theory underlying the unit operations used for this process. Their mathematical basis and the equations used for the models are specific to each chapter and are therefore presented later.

1.6.2 Precipitation

Precipitation is an operation involving simple process equipment whereby a previously soluble material comes out of solution as a solid and is commonly used at the start of a purification process to achieve inexpensive volume reduction and a potential increase in product purity (Harrison et al., 2003). In its native form, the protein assumes a structure that locks the apolar hydrophobic residues in the centre to maximise their Van der Waals interactions and minimises the unfavourable interactions with water, leaving the polar and ionic groups on the surface to hydrogen bond with water. This achieves some degree of solubility, which is enhanced by the double hydration layer that typically forms around most proteins that force them apart from each other and hence encourage them to remain solvated. Nevertheless, this structure only has a relative small Gibbs free energy (around 10-20 kcal/mol), meaning that the structure may only have marginal stability (Harrison et al., 2003) and can therefore be disrupted by modest alterations to the environment surrounding the proteins. This can be achieved by adding salts such as sodium sulphate (Neal, 2005) and depending upon the ionic strength, the protein may either remain soluble (a process called salting in) or precipitated (salting out). Solubility generally peaks at a relatively low salt concentration and any significant increase cause solubility to fall and precipitation begins to dominate. When using aqueous solvents, salting-out effects pre-dominate, although certain proteins such as albumins can remain highly soluble at high ionic strengths. This property is exploited in by Protherics in the CroFab™ process by fractionally precipitating the protein of interest (Neal et al., 2003, 2004) and leaving the majority of the remaining contaminating serum proteins in the mother liquid. This can achieve significant purification – for CroFab™, 75-80% of the albumin in the feed is removed in this step.

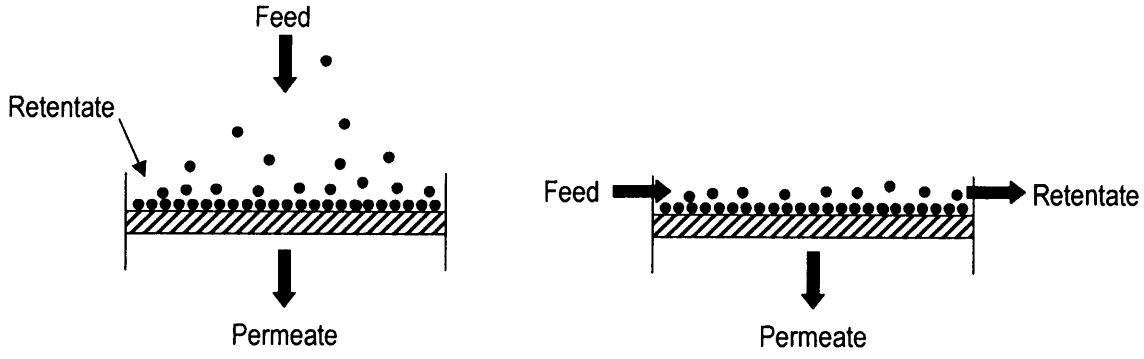
1.6.3 Filtration

Filtration membranes separate particulates from soluble components in a fluid suspension or solution according to their size, by flowing the liquid phase across a porous medium across which a pressure gradient is established to force the bulk fluid flow through (Harrison et al., 2003). Filter separations rely on the use of a membrane with pores of a certain size which permit particles of that diameter or below to pass through into the permeate (also called the filtrate), whilst particles that are too large are rejected (or retained) in the retentate phase. Two main types of filtration exist – (a) conventional/dead end filtration and (b) cross flow/tangential flow. In the first, the fluid flow

containing solids is normal to a membrane surface, with the particulates accumulating as a cake on the surface of the membrane, whilst in cross flow mode, liquid is pumped in parallel across the membrane surface, as shown in Figure 3. For dead end filtration, the solids thickness continues to build up and permeate flux reduces with time until it reaches zero, whereas with cross flow filtration, where the scouring action of the flow removes some of the cake, the feed with either a solute or solid becomes concentrated at the surface and eventually, the permeate flux reaches a constant steady state value. For the CroFab™ process, dead end filtration using depth filters (which rely on the tortuosity of the channels through the membrane and statistical probability to remove particulates) are used throughout the process in order to remove colloidal material to reduce turbidity that results from either the ovine feed or the precipitation step. The ultrafiltration membrane used for CroFab™ has pore sizes which with a cut off size of 30 KDa, allowing retention of F_{AB} from the previous digestion step, and is used in two modes:

- Concentration mode, in which the product becomes concentrated on the retenate side, whilst smaller molecular weight species pass through with bulk permeate flow
- Diafiltration mode, in which a buffer is added to the concentrate tank at the same rate as the permeate is formed when passing the feed across the membrane, meaning that the retentate phase retains the same volume. For the CroFab™ process, diafiltration reduces the salt concentration, conditioning the process stream for the subsequent ion exchange step, which requires a low ionic strength feed to encourage binding of the F_C component to the matrix

Membrane steps are also used after the main process in order to concentrate and depth filter the material and also to achieve viral elimination.



(a) Conventional filtration *(b) Crossflow filtration*

Figure 3: Modes of filtration
(Adapted from Harrison et al., 2003)

1.6.4 Centrifugation

Centrifugation is commonly used at large scale for separating solids from a liquid feed stream and relies on the density difference between the two phases to achieve partitioning. It is regularly employed for operations such for cell harvesting following fermentation or removing debris following cellular disruption by homogenisation. The rotational force achieved in a centrifuge provides enhanced settlement to a feed stream relative to that experienced purely under gravity and generates two discharge streams – one consisting of heavy sediment (containing the majority of the heavy solid particles) and the other of the lighter, predominantly particle-free liquid (supernatant). Large scale centrifuges can achieve accelerations of 20000 times that of gravity (20000g). Many different types of production-scale centrifuge exist, including tubular bowl, scroll and basket. Protherics employ a disk stack, which is a system of rotating inverted conical disks stacked upon each other and separated by spacer caulks of about 1 mm in height, thus minimising the duration needed for recovering dense solids while also allowing continuous liquid feed and discharge (Harrison et al., 2003). The feed is applied through the central shaft at the top of the bowl and under pressure, the flow path moves to the bottom of the bowl and then back upwards between the disks, as indicated by the arrows in Figure 4. The net force on the solids causes them to emerge from the liquid stream and accumulate on the underside of the disks, where they slide down and are held in a holding area at the side of the bowl. There, they can either be collected continuously through nozzles, or be discharged at the end of a batch (intermittent discharge). Meanwhile, the supernatant flows up between the disks and is discharged from the top of the bowl.

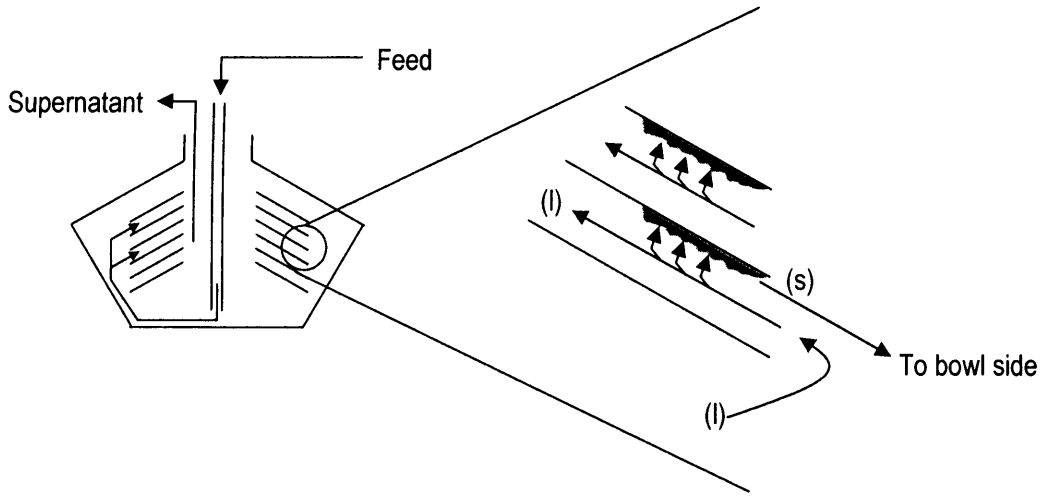


Figure 4: Simplified cross section of a disk-stack centrifuge and mode of separation

1.6.5 Ion exchange and affinity chromatography

Chromatography is a method of purifying a feed stream by its application to a column containing a stationary matrix (resin). Due to their molecular characteristics, different feed solutes possess different affinities for the matrix, enabling their separation. Many different types of chromatography matrix exist, two of which are employed in CroFab™ manufacture: ion exchange and affinity capture.

The ion exchanger uses an anionic matrix which therefore enables it to bind negatively charged impurities such as the F_C component remaining after digestion, as well as digestion excipients and other residual proteins such as albumin. The F_{AB} exhibits very little interaction with the resin and the majority passes straight through the column. The feed is applied at low ionic strength using a weak phosphate buffer (hence the diafiltration step in the ultrafiltration), whilst elution of the adsorbed impurities occurs at a 100 fold higher salt concentration.

The affinity step uses venom molecules that are immobilised to a custom manufactured matrix. Four different custom matrices are used to capture each of the different F_{AB} components. The matrix is equilibrated with a phosphate buffer, after which the feed is loaded with a saline solution and then eluted by a low pH glycine buffer to produce the final venom-specific product.

1.7 General aims of the thesis

Based on some of the key gaps in the field, the ultimate aim of the thesis was to produce a suite of development tools – primarily modelling based – for rapid day-to-day decision-making about the feasibility of options designed to achieve improvements to pre-existing processes. Since the majority of models available in the literature concentrate upon individual process steps, a key focus for this thesis was to simulate whole process sequences, populated by real industrial data and constraints rather than the more commonly used hypothetical input data. The models evaluate changes from both financial and process perspectives, examining both economic factors such as cost of goods alongside technical aspects such as product yield. This integrated approach provides an effective basis for rationalising process changes (Karri et al., 2001a).

1.8 Novel contributions made to the field by the thesis

1.8.1 Evaluating changes to pre-existing processes

The majority of bioprocess simulations focus upon modelling new processes for the manufacture of a brand new drug, covering all activities from initial drug discovery, through to clinical trials, process development and optimisation, plant installation, commissioning, validation and finally several years' worth of manufacture. This thesis differs in that it evaluates changes to an existing process and determines whether the effort involved in developing those strategies or the resulting improvements in production are large enough to make the change worthwhile. For example, if multiple process units are replaced by a single step in an attempt to shorten manufacturing schedules and reduce expenditure, then the models described in this thesis can determine whether the time and cost involved, such as for purchasing new columns or revalidating the process, achieves a large enough improvement to make the upgrade commercially viable.

1.8.2 Evaluating financial and technical performance of industrial manufacturing processes

In general, simulations are not as widely accepted for routine pharmaceutical process design compared to other sections of the process industries and in part, this stems from a lack of confidence that people have the robustness and validity of the calculated outputs. Many of the publicly-available simulations focus only on either the technical performance of an individual unit operation or the financial characteristics of a generic process flowsheet populated by industry-averaged data values. This thesis integrates both financial and technical factors together when evaluating potential improvements for whole processes and demonstrates the benefits of this approach using actual data and resource constraints gained from an FDA-approved production process. Use of company-derived data-sets gives the research a greater sense of commercial reality than it would otherwise have.

1.8.3 Novel techniques for assessing the feasibility of a process

The application of specific techniques to the evaluation of process changes demonstrates novelty e.g. the use of a multi-layered modelling approach which minimises the amount of time-consuming experimentation needed to obtain input data, or the application of global sensitivity analysis techniques to a sequence of steps in a bio-manufacturing processs.

1.9 Specific aims of the thesis

1.9.1 Software selection

In order to fulfil the aims in Section 1.6 successfully, the first task carried out in this thesis was to select the most suitable package in which to construct a bioprocess simulation. Due to the lack of suitable commercial or academic software for modelling bio-manufacturing operations as described above, the simulation environment was chosen by comparing a wide range of different packages on the market and quantitatively ranking them (see Chapter 2: Materials and Methods).

1.9.2 Framework for evaluating the manufacturing impacts of changing large scale processes

Having selected the most suitable software, it was then used to construct an initial modelling framework for the evaluation of the manufacturing impacts of implementing moderate changes to large scale processes in terms of product yields and process times. Changes examined included changing feed volumes or bed volumes used in purifying chromatography columns. This work provides the first application in the public domain of bioprocess modelling techniques to a real industrial process with its own data/resource constraints and serves as a reference point for future work reported in the thesis – see Chapter 3.

1.9.3 A multi-layered software methodology for the evaluation of bio-manufacturing options

Although the previous chapter illustrates the utility of biomanufacturing simulations, their wider industrial acceptance would be further enhanced by development of other approaches. Models often require large quantities of input data, which may need to be gathered from time-consuming experimental studies. Where many flowsheet options are evaluated and several turn out to be unfavourable, the laboratory time spent acquiring the relevant data will ultimately represent a waste of resources. Chapter 4 presents a multi-layered methodology designed to overcome this problem, enabling the efficient use of models and the underlying experimental studies. In each layer, inferior options are screened out, while more promising candidates are evaluated further in the subsequent, more refined layer, which uses more rigorous models that require more data from time-consuming experimentation. Screening ensures laboratory studies are focused only on options showing the greatest potential. To demonstrate the value of the methodology, it was used to investigate the attractiveness of options such as increasing feed volumes or replacing process steps.

1.9.4 Evaluation of the development and manufacturing impacts of changing processes

Whereas the first two research chapters concentrated upon assessing the manufacturing effects of changing a process, they did not consider the costs and timescales involved in doing so. Chapter 5 therefore describes a two-part decision-support model which is used to assess both the development- and manufacturing- related impacts of implementing a series of major changes to large scale processes. A multi-attribute decision-making technique is described which integrates the outputs of the two models into a single value to simplify the comparison between strategies. Changes that were examined included increasing feed titres, replacing multiple steps with a single operation and removing unit operations altogether. These options were assessed in terms of both developmental costs and times as well as manufacturing production levels, COG and batch times.

1.9.5 Global Sensitivity Analysis for determining parameter importance in bioprocesses

In previous chapters, the process changes that were evaluated were selected as being those of typical industrial relevance; however, it is also possible that a company may not know which input parameters are the most important to a process. Knowledge of which changes are likely to have the greatest influence before running a model would enhance its utility by identifying those process attributes which offer the greatest potential for achieving manufacturing improvements. This issue is addressed in Chapter 6, which describes the application of a technique called Global Sensitivity Analysis (GSA) to mathematical models of bioprocess flowsheets in order to identify the most critical variables that influence product yield and throughput.

1.9.6 Development of methods for the industrial-scale chromatographic purification of antibodies

Although simulations of whole processes are desirable in order to create methods that have industrial relevance, individual unit operations also require robust models in order to make accurate predictions of the best operating conditions to choose that satisfy production goals. This is particularly true of purification based on affinity chromatography, where expensive, custom-made resins may be employed for the purification of antibodies that are specific to particular antigens of interest (Thillaivinayagalingam et al., 2007). The high cost of such matrices necessitates efficient process design in order to maximise their economic potential and thus techniques identifying the most suitable conditions for achieving desired manufacturing performance provide significant value. Hence, the first part of Chapter 7 describes a computer model which incorporates the effects of capacity changes over consecutive chromatographic cycles in order to identify combinations of protein load and loading flowrate which satisfy preset constraints for product yield and throughput. The method is demonstrated by its application to an affinity column operated by Protherics for the manufacture of another polyclonal F_{AB} product called DigiFab™, for the treatment of digoxin toxicity.

Another strategy considered industrially involves column chromatography for initial product recovery and purification, instead of more conventional operations such as precipitation or centrifugation (Chhatre et al., 2007a). Although resins such as Protein A are commonly employed to capture IgG in monoclonal antibody production, several problems arise, including their high purchase cost of the matrix and the need for expensive cleaning agents. Furthermore, low capture efficiencies and binding affinities observed with crude, highly viscous feeds (Huse et al., 2002) and other questions of resin shelf life and ligand leakage (Yavuz et al., 2006) have driven the development of novel matrices which employ low molecular weight synthetic ligands. The second part of Chapter 7 focuses upon the evaluation of one such prototype agarose-based synthetic ligand adsorbent (Millipore, Consett, Co. Durham, U.K.) for recovering an antibody product from a crude industrial feedstock. The prototype is compared to commercially available matrices by static and dynamic capacity measurements and by inverted confocal microscopy to visualise product diffusion inside particles of the matrix.

2: MATERIALS AND METHODS

2.1 Introduction

This chapter describes the techniques required to complete the modelling and experimental studies outlined in the first chapter. The challenges encountered when modelling bio-manufacturing activities are summarised, followed by selection of the appropriate dynamic software needed in order to meet these challenges and generate the required simulations. A generic hierarchical structure that was applied to the different dynamic models is then put forward, explaining how the simulations were constructed in order to capture all of the aspects associated with bioprocess manufacture. Finally, the details of the synthetic chromatography experiments are described.

2.2 Challenges of bioprocess modelling

The successful modelling of bio-manufacturing activities presents three main challenges:

- Data relating to process costs, material balances and durations of unit operation activities need to be captured, entered into a suitable database and connected to the main modelling software
- Resource pools for labour, chromatography columns, process vessels etc. need to be created from which activities can draw resources and then replenish after use, if appropriate (this does not occur for disposable/'single-use' resources such as buffers)
- The principal activities underlying the operation of individual stages need to be generated, incorporating functions to fetch and return resources and time-delay functions to represent the durations of those activities

Other requirements of the model include the need for an intuitive user interface for housing and manipulating the input data and also a logical structure for organising the model which ensures that all of the aspects relevant to biomanufacturing are incorporated.

2.3 Software selection

2.3.1 Software criteria

- Overcoming these challenges requires a judicious selection of the most suitable modelling environment. Of these, there are many types available commercially and these can be categorised into two different groups – (a) Static or dynamic and (b) continuous or discrete

Spreadsheets are classical examples of a static environment and traditionally, custom in-house process models were frequently constructed in spreadsheets (Lim et al., 2004). Although their use is still widespread for decision making (McGill and Klobas, 2004), their static nature makes them unable to account for typical bioprocess dynamics. Static programs are unable to account for the dimension of time during calculations. Hence, for example, if a formula representing an output as a

function of time is entered into a cell in Microsoft® Excel (Microsoft Corporation, Washington, U.S.A.), the model will only be able to evaluate that equation at the values of the inputs, therefore only providing a single model output under the defined set of input conditions. Unless code is specifically introduced to iterate the calculation over different values of inputs, or the inputs are changed manually, the model will be unable to calculate the output function at any other input value. Dynamic simulations solve this problem by automatically calculating model values at each time step and permitting these to influence other parts of the model as needed. For example, dynamic modelling allows the presence of a pool from which resources are abstracted and replaced as necessary. Thus, if a particular resource is used up in one part of the manufacturing model, a call to use this resource from elsewhere will result in that part of the simulation halting and waiting until the resource is replenished in adequate quantities to cover the latest demand.

Models can alternatively be classified as either continuous or discrete event. In the former, there is a continuous flow of data values through the model, with data values re-calculated at regular intervals in time, enabling the model output to be evaluated if the input time is known. By their nature, manufacturing operations in a pharmaceutical facility occur irregularly – it is impossible to say that a particular event will occur at any given time, as it depends on the state of the model at that time. For example, if insufficient manufacturing staff is unavailable to carry out a specific task in that facility, then the model will not allow a given event to occur unless and until enough labour becomes available. This situation can be modelled as a discrete event simulation, where product batches moving through a process are represented as, 'items.' The status of an item only changes when an event occurs and not as a direct function of time e.g. a product batch (item) only moves into a unit operation model once a resource constraint is lifted.

On this basis, the first task involved choosing appropriate dynamic and discrete event software, capable of calculating process times within the available resource constraints, as well as evaluating material balances and cost of goods. To this end, a review of available software was carried out on the World Wide Web, examining software reviews, images of the interfaces and downloaded sample simulations to identify their advantages and disadvantages. The following criteria required for each program were used in their appraisal, specifying that the software should...

1. Enable rapid model development, with the ability to add and manipulate parameters or properties to represent bioprocesses accurately
2. Have a graphical and intuitive interface
3. Use a hierarchical structure to enable inheritance of properties, reducing time involved in repeatedly configuring new objects
4. Not be excessively processor or memory intensive, as this could produce a slow simulation
5. Provide methods for the automatic export of output data from the simulation

6. Use a programming language already known by the author (namely Visual Basic or C++)
7. Be relatively easy to learn
8. Have the capacity to correct model errors during the course of a simulation run (facilitates faster debugging)

Eighteen simulation environments were identified and scored between 0 and 10 for each of these features, zero representing a feature that is not supported at all and ten indicating a fully supported feature. Scoring was based upon the numerical representation of the qualitative impressions gained from each program. Where little information was available about a feature other than its existence, a default value of five was used. The details of each individual environment are given in Appendix 1.

2.3.2 Software evaluation

A more detailed summary of each piece of software, including the manufacturer, are in the appendix.

	Criteria								
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Average (%)**
Analytica	5	8	5	9	8	5	5	5	63
Anylogic	5	8	5	2	8	5	5	5	54
Arena	8	7	5	7	3	5	5	5	56
Crystal Ball	4	3	5	8	8	5	4	0	46
Decision Pro	5	8	5	8	8	5	5	5	61
emPower	5	7	10	3	8	5	5	5	60
Enterprise Dynamics	7	5	10	8	4	5	5	5	61
Extend™	9	8	10	7	5	7	7	5	73
Flexsim	9	8	10	5	6	7	5	5	69
GPSS World	6	8	10	8	4	4	4	5	61
Pasion Simulation System	7	7	10	8	7	3	4	0	58
PIMSS	4	7	5	3	4	5	4	5	46
ProVision	5	6	10	5	8	5	5	0	55
Resource Manager	5	4	5	4	5	5	5	5	48
SansGUI	5	7	10	8	7	7	2	5	64
ShowFlow	7	8	10	8	7	5	5	5	69
VisSim	8	8	5	6	4	7	5	5	60
Visual Simulation Environ.	5	7	10	5	6	5	6	5	61

Table 3: Scores of eighteen different modelling environments

Evaluation was carried out using a 1.4 GHz Pentium computer using 256 MB and operating on the UCL network with Windows® 2000 Professional. Based on the qualitative assessments and the scores above, 'Extend™,' version 6 (Imagine That, Inc., San Jose, U.S.A.) was selected. This is a powerful dynamic and discrete event simulation environment which accurately reflects the impacts of resource constraints upon manufacturing throughput. Additionally, Extend™ provides an intuitive modelling and data reporting system and consequently, has been used by other researchers to represent bio-manufacturing activities (e.g. Lim et al., 2004; Mustafa et al., 2004) and the drug development pathway (Rajapakse et al., 2004; Rajapakse et al., 2006). This experience was also a key motivating factor in choosing the package. All Extend models were developed and executed on a 1.4 GHz Pentium M 256 MB RAM computer running Microsoft® Windows® XP (Microsoft Corporation, Washington, U.S.A.).

2.4 Model construction in Extend™

Extend™ consists of many libraries which containing pre-coded modules (called blocks) which can be customised to represent the standard functions required to model a pharmaceutical production facility. Properties of the block can be manipulated by double clicking on them to access their dialog boxes. Data values can also be entered using data inlet ports which are attached to each block. Models of individual unit operations are constructed by cloning such blocks onto a workspace, adjusting their properties and linking them together with stream lines. These individual unit operation sub-models are then connected together to produce the dynamic model for the entire production process. The main block types required to achieve this are given in Table 4.

Extend™ automatically tracks the passage of time through a simulation in order to calculate campaign lengths. It should also be noted that model usage is aided by the animation facilities in Extend™, which enable visualisation of activities as they occur in the simulation. This feature is used by both the developer to eliminate coding errors and also by the end-user, such as for identifying bottlenecks. In the models developed in this thesis, outputs of total process duration, costs and material composition were exported from Extend™ back to a spreadsheet in Microsoft® Excel XP for further analysis.

Block type	Description
Program and Exit	Program blocks generate items for use in the simulation, with its dialog box providing a means to enter a schedule for when batches start. Exit blocks take items out of simulations and provide a defined means for batches of product to leave the model
Activity delay	These represent the duration of every step in every unit operation, with a data inlet port allowing time delay data to be inserted into that block and the item to be delayed for that duration
Attributes ('Set', 'Get' and 'Change')	Attributes are stream properties that are loaded on to items to represent their physical state e.g. material balance data. Set attribute blocks are used to assign initial values to each attribute as an item commences its progress through the simulation. As an item moves through each unit operation model, attribute values can be retrieved by the Get Attribute block and updated by the Change Attribute block – such as for mass balance calculations
Queues	These can be used to store an in-process batch, representing a holding point
Resource pool ('Get' and 'Release')	Resource pools are special types of queues. Pools are established for each resource (labour, holding vessels or disposable bags, specialised equipment (e.g. centrifuges/chromatography columns), chromatographic matrices, filter membranes and buffers). Get and release resource pool blocks are positioned either side of activity delay blocks in order to request and replenish the resource pools (the latter is not used for single use resources i.e. buffers/disposable bags). Items go into each Get Resource block and if sufficient resources are available to cover the demand, the items carries on – otherwise, the item is forced to wait until resources are freed from other tasks
Data communication	Extend™ allows data to be imported and exported from sources such as spreadsheets. This enables input data to be stored externally and manipulated through an external user interface, such as one set-up in Visual Basic for Applications within Microsoft® Excel. This approach provides a more user friendly way of entering information into the model than can be achieved in Extend™ and also allows programming code to be entered in order to ensure that inserted data have an acceptable format or quantity (e.g., disallowing the entry of insufficient vessel numbers to hold process volumes)
Batching	Batching is used to combine items together e.g. for the CroFab™ process, where items representing each of the four venom types are modelled, these need to be combined when the affinity purified F_{AB}s are blended together to produce one process stream
Hierarchical blocks	Extend™ provides an ability to group parts of a model together and hide their details within a 'hierarchical' block. Double clicking on the hierarchical block opens up a separate window including all of the selected blocks that were previously on the main workspace. This process can be repeated as many times as are needed, enabling hierarchical structures of blocks to be built up that provide a logical and ordered way of structuring a model. This also enables suitable groups of blocks to be re-used by simply copying and pasting their hierarchical 'container'. The structure of the model according to a hierarchical decomposition is outlined in Section 2.6
Throw and catch	Extend™ allows different parts of a model to be segregated e.g. one part of the main workspace may contain a hierarchical block containing the mass balance models, whilst another may contain the resource blocks, a third set to represent each process step etc. As items go through each process step, they will 'call' the other relevant hierarchical blocks e.g. requesting resources or calculating a material balance. Throw and catch blocks are used to divert items from one hierarchical block to another and back again in order to achieve the desired flow of items through the simulation
Decision blocks	These are another set of blocks which are used to determine the flow of items through a process – depending on the conditions entered into a Decision block, this allows items to be routed into different parts of the model as required
Operators /equations	Extend™ includes the full set of usual mathematical operators and there are also equation blocks which perform mathematical manipulations
Plotters	These can be used to create resource utilisation profiles
Cost accumulator	Resources can be assigned a cost as an amount per unit e.g. £/L. Usage is automatically tracked and a cost report is produced at the end of the simulation

Table 4: Blocks used in Extend™

2.5 Example of model construction

The figure below illustrates part of a model (from the elution sub-task of an ion exchange step) in the industrial case study, showing how blocks are connected together and how items pass between them. As outlined above, items mimic the flow of batches through a modelled process, passing through the simulation along stream lines (the double lines between blocks A-D in Figure 5), requesting and releasing resources or being delayed in the appropriate blocks to represent times spent in individual parts of the process.

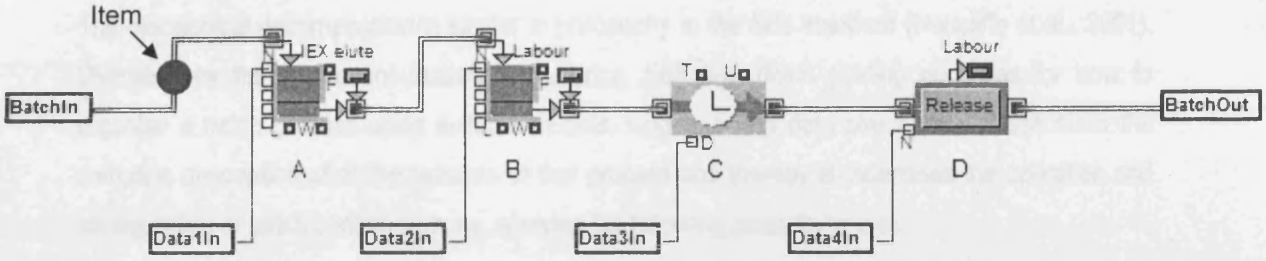


Figure 5: Sample screenshot of the elution sub-task in the ion exchange step
The circular item represents a product batch; Key: (A) 'Queue, Resource Pool' block, which requests elution buffer; (B) 'Queue, Resource Pool' block, which requests labour to perform the task; (C) 'Activity, Delay' block, which represents the duration of the elution; (D) 'Release Resource Pool' block, returning labour to its pool

Figure 5 shows an item (represented by the darkened circle on the left hand side) progressing along a stream line and about to enter a 'Queue, Resource Pool' block (A) which requests the elution buffer. Subsequently, the item enters the next block which requests labour to perform the elution task (B), after which it passes into an 'Activity, Delay' block (C) which causes the item to wait for a set period of time,. Finally, it enters the 'Release Resource Pool' block (D) which returns the labour resource to its pool. If at any point insufficient resources are available (e.g. if insufficient manufacturing labour is present within the pool to cover the demand), the item is forced to wait until an adequate quantity of resources become available. The boxes marked BatchIn and BatchOut represent the points at which the item moves into and out of the elution task sub-model from the previous and subsequent parts of the simulation respectively. Data values from the user interface which were saved in a spreadsheet were then linked to Extend™ through the data access points indicated as Data1–4In in Figure 5.

2.6 Hierarchical decomposition

As outlined above, the models were generally organised according to a hierarchical decomposition. These have been used several times to describe bioprocesses (e.g. Farid et al., 2000a) and form a logical structure on paper within which to create simulations, ensuring that all features of a manufacturing process are captured. Hierarchical representations of bio-manufacturing operations are described by Farid et al. (2000 a/b) as a series of layers, each of which contains blocks that represent a particular task and which act as containers for lower-level elements. The first layer of

blocks represents the entire process or manufacturing campaign and decomposes into stages such as product manufacture, materials preparation and equipment preparation. Each of these subdivides into specific manufacturing operations in the next layer – e.g. buffer preparation, CIP, up- and down-stream steps for product formation, recovery, dewatering, purification and concentration. Finally, individual steps in each unit operation can be detailed – e.g. charge, react, hold, mix, load, wash, discharge etc. Similar decompositions have been used by other authors (e.g. Lim et al., 2004; Mustafa et al., 2004).

The hierarchical decomposition is similar in philosophy to the S88 standard (Nortcliffe et al., 2001). Overseen by the Instrument Society of America, S88 lays down guiding principles for how to organise a batch process using suitable models, language and data structures. These allow the complete description of all the features of that process and thereby standardises the operation and configuration of batch control systems, allowing the following goals to be met:

- Defines a set way to determine how to organise and control batch operations
- Provides a common terminology to allow users to communicate their requirements effectively to each other
- Provides consistency between different vendors to allow integration of different equipment types/formats
- Enables the straightforward configuration of batch control

To achieve these goals, the S88 consists of a hierarchical structure for breaking down an enterprise such as a pharmaceutical company into its physical sites, which then reduce down to areas of those sites, followed by individual processes and then the component parts of that process. An example of the breakdown is shown below (Nortcliffe et al., 2001; http://www.isa.org):

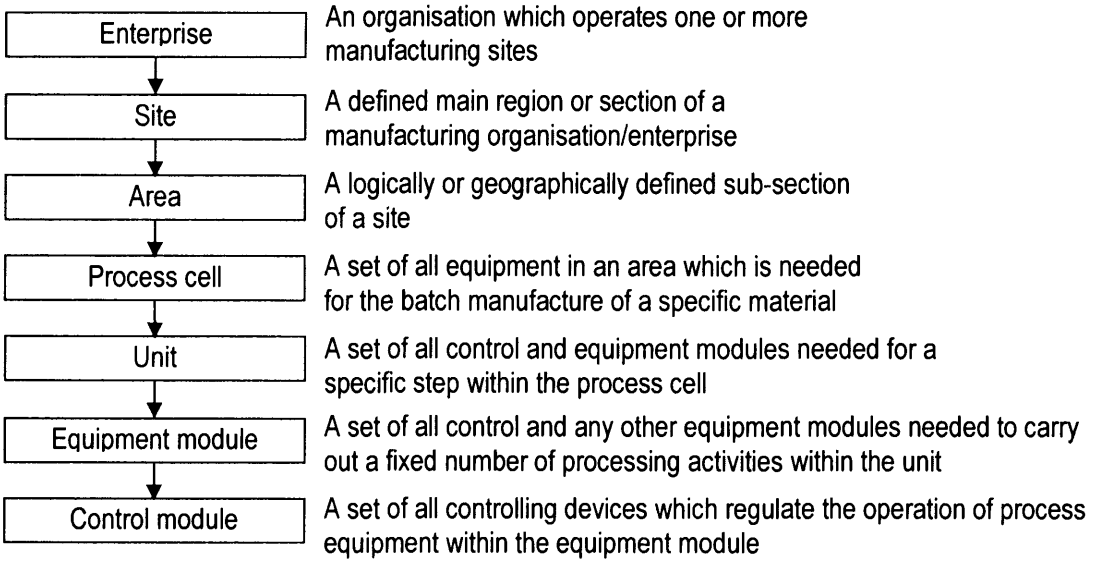


Figure 6: Outline of the S88 hierarchical standard for batch process control

This structure allows an end user to identify all of the equipment in a process as well as defining and implementing recipes which specify which equipment and control activities are carried out and the order in which these occur. The standard therefore provides several benefits, including bringing greater consistency between steps in processes/sites, enabling recipes to be re-used between different units where appropriate and allowing easier maintenance (http://www.rovisys.com).

A similar approach to that given in Figure 6 was used in this thesis, where the framework presented in Figure 7 was employed and which is specific to the requirements for modelling the CroFab™ process. Hence the manufacturing block at the top represents the highest level and decomposes into the operational characteristics level, which includes a group of all pre- and post- unit operation tasks for all unit operations as well as the main process steps and resources, such as labour, matrices, vessels, equipment and buffer resources. These feed into the individual procedures level blocks and as appropriate, resources are deployed and replenished. The operational characteristics level splits up again into the procedures level, which includes steps such as CIP, QC assays, desludging, buffer addition, product loading etc. It is at this level that calculations for those specific unit operations such as the mass balancing and cost evaluation are completed.

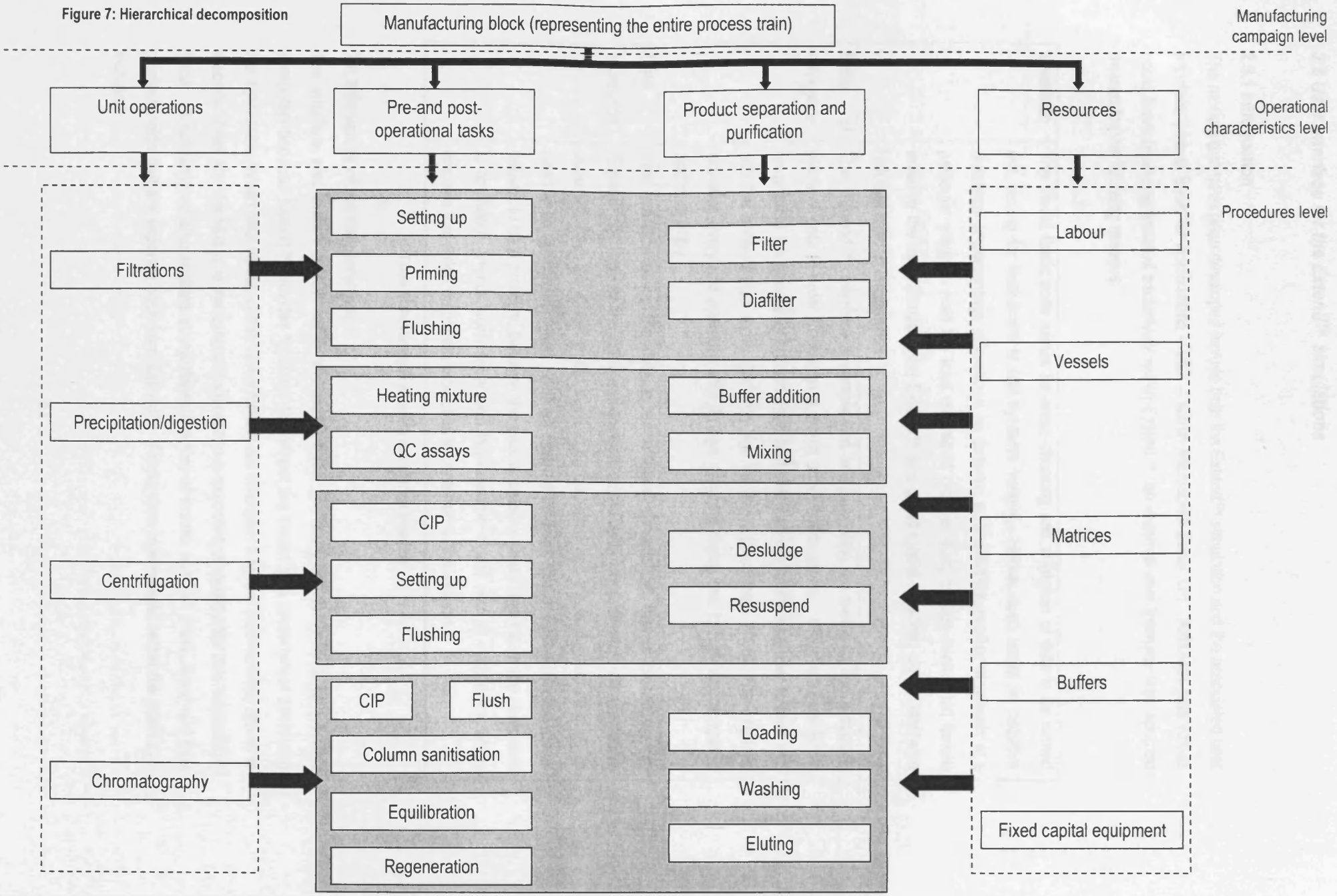
2.7 General model requirements

To facilitate the successful achievement of these goals, the following guidelines were drawn up for generic characteristics that all models must have:

Characteristic	Outcomes/comments
Ease of data manipulation	To facilitate data manipulation an external user interface will be set-up where needed for data manipulation
Model speed	Models must be created in such a way so as to run quickly and thus facilitate rapid flowsheet evaluation
Model outputs	The main outputs needed will be percentage yield, final product mass, total process times, throughput and capital/running costs and the models will be coded to export those values automatically

Table 5: General requirements of models

Figure 7: Hierarchical decomposition



2.8 User interface for the Extend™ simulations

2.8.1 Introduction

The models that have been developed include both the Extend™ simulation and the associated user interface (Visual Basic for Applications – VBA – within Microsoft® Excel XP). Although data values could have been manipulated exclusively within Extend™, an external user interface was adopted instead for the following reasons:

Validation	The Visual Basic code allows for cross-checking and validation of data to be carried out, testing for mathematical and syntactic validity. Hence, such errors as negative durations or percentage efficiencies (or entering an insufficient number of vessels at a particular stage to hold the total volume of process fluid) can be intercepted before entering the main simulation in Extend™ and avoid model runs that would otherwise halt due to such deficiencies
Ease of navigation	The, 'Extend™,' interface is complicated and navigating the model to find a specific piece of data in order to change it would prove cumbersome. Using a Visual Basic program in the structured fashion described below allows for the end user to read and edit the entered data more efficiently and intuitively (Workman, 2000). The interface provides every unit operation with its own separate dialog box to facilitate navigation and data entry
Data protection	The interface saves the data to an external spreadsheet that is picked up by, 'Extend™.' If all of the data was instead stored within Extend and any corruption occurred, then the model would need to be corrected and the malformed data would need to be updated/replaced. With an external data set, loss of data within the model ceases to be a problem, because the data is called in anew every time the simulation is activated. Hence such separation between the model and its data entry tables ensures that a degree of data protection is conferred to the system.

Table 6: Advantages of using an external interface

2.8.2 Structure of the user interface

The interface was written in VBA within Microsoft® Excel (on the 256 MB Pentium M 1.4 GHz computer) because Extend™ provides functions to connect the model to a spreadsheet containing the input data. When data values in the VBA interface are changed, they are automatically saved to a spreadsheet and the next time the Extend™ simulation is executed, it updates the new values from Excel. A spreadsheet also enables straightforward listing of results such as costs, times and final product yields that are exported back from Extend™. Parameters manipulated within the interfaces included:

- Feed volumes
- Fixed process times
- Types of process vessels
- Material balance and volumetric recovery data
- Numbers of manufacturing staff for each task in each unit operation
- Identities of buffers for each task
- Volumes of chromatography matrix
- Flowrates in centrifuge or column steps
- Resource costs for labour, matrices, buffers and disposable product bags
- Some data values are determined by equations – e.g. volumes of buffers or numbers of process vessels were dependent upon the process volume entering the unit operation e.g.:

$$Number\ of\ vessels = Round\ up\left[\frac{Feed\ volume}{Volume\ per\ tank}\right]$$

2.8.3 Overall model operation

Overall operation of the model can be summarised by the following:

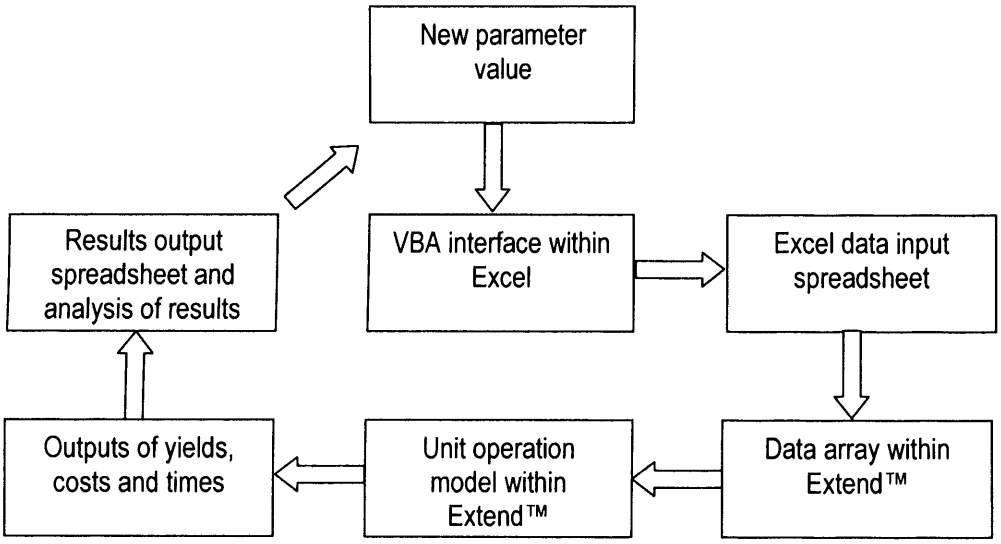


Figure 8: Depiction of how the different parts of the model integrate with each other

2.9 DigiFab™ experimental protocols

2.9.1 DigiFab™ affinity process

This section describes the experimental protocols used for the affinity column in the DigiFab™ process in order to populate the simulation described in Chapter 7 (laboratory studies conducted by Thillaivinayagalingam et al., 2007). The DigiFab™ production process is similar to CroFab™: initially, sheep are immunised with an analogue of digoxin, digoxin-dicarboxymethoxylamine (DDMA). After several months, when the antibody concentration has risen to a sufficient level, the sheep are bled and the serum is collected. Precipitation and centrifugation eliminate unwanted

proteins from the ovine serum, prior to IgG digestion by papain to leave a complex concentrated mixture consisting primarily of monovalent F_{AB}. No ion exchange chromatography is then undertaken (unlike CroFab™ manufacture) – instead, digoxin-specific F_{AB} is isolated by affinity chromatography directly, using a custom matrix consisting of DDMA coupled to 6-amino Sepharose (GE Healthcare) as per the protocol in Table 7 to separate digoxin-specific polyclonal F_{AB} from the mixed papain-digested antibody fragment population.

Stage	Buffer	Column volumes	Production flowrate (L/hr)	Scale-down flowrate (cm/hr)
Load	Filtered ovine serum	–	9.6	30
Initial wash	12mM disodium tetraborate pH 9.2	3	13	40
Acid wash	1% v/v propionic acid pH 3.7	14	13	40
Elution	5% v/v acetonitrile and 3.6% v/v propionic acid pH 2.6	19	4	13
Equilibration	12mM disodium tetraborate pH 9.2	4	13	40
Sanitisation	20% v/v ethanol and 0.1M sodium hydroxide	4	13	40

Table 7: Operating protocol for the DigiFab™ affinity purification step
The load range of total protein is between ~200 and 500 mg/mL. The loading flowrate achieves a residence time of over 11 minutes, sufficient to ensure complete pore penetration. Flowrates used in the scale-down model described later were chosen to maintain the same residence times as used at commercial scale

An A280 trace outlining the main purification steps and buffer changes is shown in Figure 9. A 12mM disodium tetraborate (pH 9.2) wash is used to equilibrate the column prior to loading the F_{AB} mixture. After feed application, a 1% v/v propionic acid wash is used to eliminate adsorbed impurities such as non-specific F_{AB} or in-process impurities including the digestion excipients, prior to the low pH elution of specific F_{AB} using a mixture of 5% v/v acetonitrile and 3.6% v/v propionic acid.

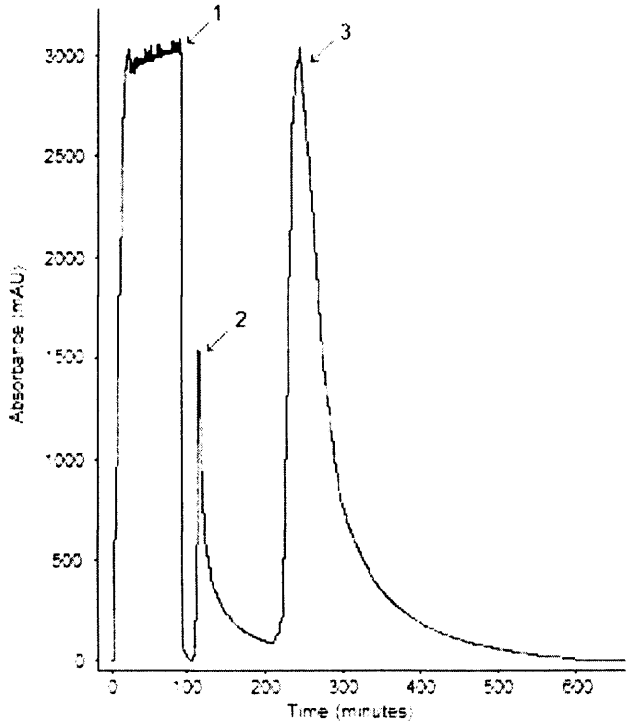


Figure 9: A 280nm chromatogram showing the main purification steps involved when applying the F_{AB} feed to the DigiFab column
(1) wash with 12mM disodium tetraborate (pH 9.2); (2) eliminate impurities with an acid wash: 1% v/v propionic acid (pH 3.7); (3) elute with 5% v/v acetonitrile and 3.6% v/v propionic acid (pH 2.5-2.7)

2.9.2. DigiFab™ scale-down mimics

2.9.2.1 Column studies

A complete experimental description can be found in Thillaivinayagalingam et al. (2007). A 500 mL frozen production sample of papain digested IgG was thawed overnight and stored at 2-8°C. The total protein concentration of the load material was 59mg/mL as determined by UV spectrophotometry at 280nm (using an absorption coefficient at 1.0 mg/mL of 1.4 - the value used for evaluating polyclonal antibody concentrations by the quality control laboratories at Protherics). To imitate production scale operations, the protein sample was filtered to remove any insoluble proteins using 0.2 µm syringe filters (Sartorius Limited, Surrey, U.K.). Scale down affinity chromatography was then undertaken by packing a C10/10 column (GE Healthcare) with a 5.5 cm high bed of the custom DigiFab affinity matrix (packed bed volume of 4.32 mL). The scale-down runs involved a 20-fold reduction in column diameter, while bed height and processing times were unchanged from the commercial-scale process. An ÄKTA Explorer (GE Healthcare) was used to run all separations, using the protocol in Table 7. Five consecutive affinity runs were undertaken at each of four different total protein loadings (174, 260, 347 and 500 mg of protein per millilitre of packed matrix). Freshly packed columns were used for each experiment with a given total protein concentration. Eluates from each column run were assayed for F_{AB} concentration by UV absorbance at 280 nm.

2.9.2.2 Determination of the maximum binding capacity

The maximum capacity of the adsorbent was calculated by batch experimentation. A stock of the DigiFab™ affinity matrix was equilibrated in PBS and 100 µL (packed volume) was pipetted into each of twelve eppendorf tubes. A stock of purified F_{AB} prepared from previous affinity runs (Thillaivinayagalingam et al., 2007) was diluted in PBS to create twelve solutions in the range of 1 to 22 mg/mL, from which 100 µL samples were withdrawn and added to the eppendorf tubes. These were placed on an end-over-end stirrer for 24 hours, after which the matrix was packed down by spinning the eppendorf tubes in a Beckman GS-6R centrifuge (Beckman Coulter U.K. Limited, Buckinghamshire, U.K.) at 1300 rpm for 2 min. 80 µL of the mobile phase was removed from each tube and analyzed for their total protein concentration. A Langmuir-Freundlich isotherm was selected as it provided the best fit to the adsorption data – polyclonal antibody feeds usually produce nonlinear Scatchard plots, indicating deviation from Langmuir behaviour and Langmuir–Freundlich provides a better fit when dealing with a polyclonal antibody (Thillaivinagagalingam et al., 2007):

$$C_s = \frac{q_{ms} C_m^{n_{LF}}}{\frac{1}{K_a} + C_m^{n_{LF}}} \quad (1)$$

C_s represents the concentration of bound solute at equilibrium with the mobile phase of concentration C_m. q_{ms} is the maximum capacity of the settled matrix volume, K_a is the association constant and n_{LF} is the Langmuir-Freundlich coefficient. The maximum capacity was calculated by iteration using SigmaPlot (Systat Software UK, Hounslow, UK) – Thillaivinayagalingam et al., 2007.

2.10 Synthetic ligand experimental protocols

2.10.1 Introduction

This section presents the methods used to evaluate the novel adsorbent for capturing polyclonal antibodies from crude ovine serum (Chapter 7). Process scale chromatography requires a robust protocol that achieves consistently high yields and purities over repeated feed cycles. To determine whether the prototype adsorbent was capable of achieving these goals, the following investigations were conducted:

- Calculation of static and dynamic capacities for the prototype and four commercially available matrices when challenged with purified IgG
- Determination of column re-use capacity when challenged with ovine serum
- Evaluation of cleaning protocols for the novel adsorbent
- Application of confocal microscopy to compare protein uptake patterns between the prototype and commercially available Protein G Sepharose (GE Healthcare, Buckinghamshire, U.K.)

Unless otherwise stated, all chemicals used during experimentation were obtained from Sigma-Aldrich (Poole, Dorset, U.K.).

2.10.2 Static capacity measurements

The static antibody binding capacities of the five matrices were determined in order to provide an indication of the amount of IgG needed to achieve dynamic breakthrough from packed columns. 100 µL of each adsorbent (Table 8) was pipetted into eppendorf tubes and spun down at 3000 rpm for two minutes, after which the supernatant was withdrawn and discarded. Each matrix was equilibrated by adding an excess (500 µL) of the relevant buffer (composition and concentration given in Table 9), agitated on a Minishaker (IKA Works, Wilmington, North Carolina, U.S.A.) at 2000 rpm for ten minutes and spun down by centrifugation at 3000 rpm for two minutes. This was carried out in triplicate to leave the matrix fully immersed in equilibration buffer. To ensure that the matrix was fully saturated with protein, each tube was then overloaded with 500 µL of 98% pure IgG, produced using MEP Hypercel (IgG concentration of 30 mg/mL in 10 mM sodium phosphate pH 7.6, as determined by absorbance at A280 and a molar extinction coefficient of 225,000 $M^{-1}cm^{-1}$).

Adsorbent	Manufacturer	Batch/lot number	Storage medium
Prototype	Millipore (Consett, Co Durham, U.K.)	-	20% ethanol
MEP Hypercel	Pall BioSepra (Cergy-Saint-Christophe, France)	12035-010 / 200920/C271	1M NaCl and 20% ethanol
ProSep G	Millipore (Consett, Co Durham, U.K.)	115113324 / 1000002	0.1M sodium acetate, 0.5M NaCl and 1% benzyl alcohol
MAbsorbent® A2P	Prometic Biosciences (Cambridge, U.K.)	3901-0100 / FA 0673	20% ethanol
ProSep-vA Ultra	Millipore (Consett, Co Durham, U.K.)	CAE003324 / 0320001	Sodium acetate and 1% benzyl alcohol

Table 8: Adsorbents evaluated

Adsorbent	Equilibration buffer	Washing buffer	Eluting buffer
Prototype	25mM sodium phosphate/HCl pH 7.6	25mM sodium phosphate/HCl pH 7.6	100mM glycine/HCl pH 2.8
MEP Hypercel	100mM sodium phosphate/HCl pH 7.8	150mM mixed buffer (sodium phosphate, sodium chloride, caprylic acid) pH 7.6	100mM citric acid monohydrate pH 2.3
ProSep G	0.01M PBS pH 7.4	0.01M PBS pH 7.4	100mM glycine/HCl pH 2.8
MAbsorbent® A2P	25mM sodium phosphate/HCl pH 7.6	25mM sodium phosphate/HCl pH 7.6	50mM sodium citrate pH 4
ProSep-vA Ultra	0.01M PBS pH 7.4	0.01M PBS pH 7.4	100mM glycine/HCl pH 2.8

Table 9: Buffer compositions
All chemicals were obtained from Sigma-Aldrich Limited (Poole, Dorset, U.K.). All columns were equilibrated for five column volumes; washing and elution steps were carried out to base line in the UV trace

Antibody purity was confirmed by non-reduced SDS PAGE gel analysis and scanning densitometry as described in section 2.8.7). Separate 100 µL aliquots of washed and equilibrated matrices were also incubated with ovine serum obtained from Protherics U.K. Limited (IgG concentration of 34 mg/mL determined by gel analysis, CCT stage 002, batch 4081, bag 03). Prior to use, the serum was filtered using 0.2 µm Minisart syringe filters (Sartorius Limited, Surrey, U.K., 16534/060031) to remove particulates that could potentially interfere with IgG binding. Both sets of eppendorf tubes were agitated on the Minishaker at 2000 rpm for twenty-four hours. This duration was selected to provide conditions for maximal protein uptake. The tubes were spun down at 3000 rpm for two minutes and the supernatants were withdrawn and assayed for IgG concentration by SDS PAGE gel analysis.

2.10.3 Dynamic capacity measurements

Dynamic capacities of the five adsorbents were determined at 10% breakthrough. All resins were packed into a C10/10 column (GE Healthcare) at 3.95 mL/min (300 cm/hr) with five column volumes of water and the beds were then operated with the buffers given in Table 9 using an ÄKTA Basic 100 (GE Healthcare). For each packed bed, the operational flowrate was selected to achieve a residence time of six minutes i.e. sufficient time to allow adequate diffusive penetration of the matrix particles by the load. Each bed was overloaded with 5 mL (150 mg) of 98% pure IgG (using the same MEP Hypercel-purified stock as above) in order to create breakthrough curves. Dynamic

capacities were then calculated as follows: the UV absorbance, A (arbitrary absorbance units: AU) of the flowthrough from the column was determined by the Beer-Lambert law:

$$A = E\,c\,l \tag{2}$$

E is the extinction coefficient (225,000 $M^{-1}cm^{-1}$), c is the flowthrough molar concentration of IgG at 10% breakthrough (3.0 mg/mL; 20μM) and the path length l is 0.2 cm in the UV cell of the ÄKTA Basic 100 (GE Healthcare). Based on this data it was determined that the 10% breakthrough point would occur at an absorbance of 900 mAU (280 nm). Capacities (DBC) were then determined from the following equation (Millipore Technical Brief, Literature Number TB1175EN00):

$$DBC = \frac{C_0 \times (V_L - V_0)}{V_{Bed}} \tag{3}$$

Where C_0 is the IgG feed concentration (30 mg/mL), V_L is the volume loaded up to 10% breakthrough at 900 mAU, V_0 is void volume of the beds (void fraction of 0.4 assumed for all matrices) and V_{Bed} is the packed bed (all volumes in millilitres). Bed volumes and flow rates used in the study are given in Table 10. DBC results were averaged over three repeat runs (± 4.5%).

Adsorbent	Bed height (cm)	Bed volume (mL)	Flowrate (mL/min)	Flowrate (cm/hr)
Prototype	4.2	3.3	0.55	42.0
MEP Hypercel	3.9	3.1	0.51	39.0
Prosep G	3.0	2.4	0.39	29.8
MAbsorbent® A2P	5.0	3.9	0.66	50.4
ProSep-vA Ultra	5.3	4.2	0.69	52.7

Table 10: Bed volumes and operational flowrates used in the dynamic capacity study
Flowrates were used for all washing, eluting and re-equilibrating steps with each matrix. All flowrates were set to achieve residence times of six minutes for every bed. The diameter of all packed beds was 1 cm (C10/10 column)

2.10.4 Prototype re-use study

The re-use capacity of the prototype was determined over ten cycles using the C10/10 column, again operating with a flow rate set to achieve a residence time of six minutes. Based on the results of the static and dynamic capacity measurements, 1 mL of 0.2 μm filtered ovine serum was applied to a 2.98 mL packed bed (3.8 cm high, packed at 3.95 ml/min) and the flowthrough was fractionated into 1 mL aliquots. The serum was washed through the column until the UV trace returned to

baseline (< 50 mAU), after which the IgG was eluted using 100 mM glycine/HCl pH 2.8 in a single eluted fraction. The flowthrough and elution portions were both analysed by SDS PAGE and the ten-cycle run was repeated on a 3.14 mL bed using a mixed sodium phosphate/caprylic acid wash (same as the washing buffer for MEP Hypercel – see Table 9 for composition) prior to elution in an attempt to remove adsorbed albumin and other minor impurities to improve eluted IgG purity (Newcombe et al., 2005). To determine the impact of repeated purification cycles, the dynamic capacity of the adsorbent was then determined again after ten repeat runs by the method in section 2.8.3.

2.10.5 Column cleaning

Relatively harsh cleaning agents are desirable for process-scale chromatography when purifying antibodies or proteins from crude materials in order to eliminate adsorbed protein and remove the potential risk of bioburden that could adversely affect subsequent column runs and product quality. Although chaotropic agents such as urea or guanidine hydrochloride are commonly used for cleaning Protein A columns and are highly efficient agents for removing unwanted impurities, their high costs and the disposal problems associated with their use can make them unattractive for commercial operation (Shukla et al., 2007). A major advantage of the prototype adsorbent is that it utilises a conjugated chemical ligand and may therefore show tolerance to alternative cleaning solutions. Alternatives were considered for this study: 0.5 M NaOH and 0.1M HCl were evaluated to determine their cleaning power and the impacts upon column capacity over multiple loading cycles.

2.10.6 Inverted confocal scanning laser microscopy

Inverted confocal laser microscopy was used to visualise the uptake of labelled IgG by individual beads of the prototype. Visualisation was achieved using the fluorescent Cy™3 dye (GE Healthcare – catalogue number PA23001, expiry July 2008, transported to UCL and stored at 2-8˚C) which was used to tag purified IgG prior to incubation with the matrix. As a control, the experiment was also repeated with Protein G Sepharose 4 Fast Flow (GE Healthcare – lot 310734, stored in 20% ethanol). The spherical beads of Protein G Sepharose were used instead of the ProSep G used in sections 2.1 and 2.2 because the latter consists of angular beads owing to its glass base matrix, complicating its direct comparison with the spherical particles of the novel adsorbent.

The method used by Thillaivinayagalingam et al. (2007) was employed to label the IgG. Two stocks of IgG were prepared by diluting a portion of the 30.0 mg/mL stock of MEP-Hypercel purified IgG described above down to 2.0 mg/mL – one in 0.1M sodium carbonate (buffer pH adjusted to 9.3) and the other in 0.01 M PBS pH 7.4. 1 mL of the sodium carbonate diluted IgG was then labelled fluorescently by addition to a vial of the Cy™3 dye and inverted gently by hand for forty minutes (owing to its light sensitivity, the dye was protected at all times by storing it inside its foil pouch). Subsequently, the contents of the vial were added to 4 mL of the PBS-diluted IgG to produce a

solution containing a total of 10 mg of IgG. To prepare the adsorbents, 100 µL of each resin was spun down at 3000 rpm for two minutes and washed three times using 1 mL of 0.01 M PBS. This resin volume was chosen to ensure that it would be completely saturated by the 10 mg IgG solution. After adding the PBS-washed matrix to the 10 mg of IgG, it was inverted gently and two 300 µL samples were withdrawn – one immediately after incubation and the other after 150 minutes (duration chosen to maximise uptake of IgG). Each 300 µL sample was spun down to recover the matrix and then washed three times with 1 mL of 0.01M PBS to remove any unbound IgG and fluorophore. Each sample of matrix was then resuspended in 20 µL of PBS and analysed using an inverted confocal laser microscope (Leica Microsystems U.K., Buckinghamshire, U.K.) at a twenty fold magnification using an excitation wavelength of 568 nm. The image of each bead was focused upon its central plane to ensure that the intensity of the monitored emissions accurately reflected IgG uptake along the entire cross section of the bead. Intensity values for each bead were averaged over three scans and three beads were analysed at each time point to verify that similar intensity profiles were obtained.

Two further control experiments were also performed. Firstly, the fluorescence signal from a sample of the novel adsorbent was monitored every three minutes over a fifteen minute period (the total duration taken to analyse a total of three beads at any time point) to determine whether there was any significant fluctuation over the course of the experiment as a result of photobleaching. Secondly, Sepharose CL 4B agarose beads without any ligand attached (GE Healthcare) were also analysed by the Cy™3 protocol to determine whether there was any non-specific binding of the fluorescently labelled IgG to the base matrix which would potentially interfere with the evaluation of the intensity profiles.

2.10.7 SDS PAGE

Antibody concentrations were determined by non reducing denaturing SDS PAGE, using pre-cast 4-20% Novex® Tris-Glycine gels (Invitrogen U.K., Paisley, U.K.). 10µL of each sample was mixed with an equal volume of 2×SDS sample buffer (Invitrogen) and heated in boiling water for two minutes. Each well in the gel was loaded with 10µL of the heated samples (approximate protein loadings of 5–10 µg per lane) and electrophoresed at 150 V for ninety minutes. Protein bands were visualised by staining with SafeStain (Invitrogen) for one hour, after which the gel was destained with water for one hour and then placed into fresh water overnight. The integrated optical densities (IOD) of each band were then determined by scanning densitometry (LabWorks™ version 4.5, Ultra-Violet Products Ltd, Cambridge, U.K.), and compared to calibrated IgG concentrations. IgG purities were calculated from the percentage band intensities as evaluated by densitometry analysis.

2.10.8 Matrix particle sizing

The size distribution of the prototype adsorbent particles were evaluated using a Mastersizer 2000 (detection range of 0.02 – 800 μm) equipped with a small volume sample dispersion unit (Malvern Instruments Ltd, Malvern, UK). The matrix was well mixed and added to the sample dispersion unit until an obscurance of 12% was reached. Sizes were evaluated with respect to the percentage of total particle volume. Measurements were repeated three times, with high levels of reproducibility observed between samples.

2.11 Conclusion

This chapter has described the simulation and experimental methods used to evaluate the effects of making changes to the commercial-scale CroFab™ process. Initially, the challenges involved in understanding a process and translating that knowledge into a functioning model were discussed, along with the selection of the most suitable modelling package for achieving these objectives. The ways in which the selected environment (Extend™) can then be used to create models was described, as well as how they were structured according to a hierarchical decomposition in order to capture every salient feature of the manufacturing process. Coupled to this was an explanation of how the user interface in Visual Basic for Applications was set-up in order to house the input database and interfaced with Extend™. Finally, the details of the experimental protocols used in the the DigiFab and synthetic affinity adsorbent experiments were given.

Having outlined the methodologies used in this thesis, the next chapter describes how the modelling techniques in Extend™ were utilised in constructing a model in order to illustrate the effects of changing Protherics' process.

3: FRAMEWORK FOR EVALUATING THE MANUFACTURING IMPACTS OF CHANGING LARGE SCALE PROCESSES

3.1 Introduction

As outlined in the introduction, the proven advantages of simulations in fields such as chemicals or polymers have established modelling as a commonplace technique for process design and evaluation in these sectors (Shanklin et al., 2001; Petrides et al., 2002). Although technical models of single stages (Pinto et al., 2001), such as fermentation (Richard and Margaritis, 2004), filtration (Reynolds et al., 2003), centrifugation (Clarkson et al., 1996; Maybury et al., 1998), homogenisation (Siddiqi et al., 1996; Wong et al., 1997) and ion exchange, size exclusion, affinity and expanded bed chromatography (Kaczmarski et al., 2001; Li et al., 1998; Vunnum et al., 1995; Bruce et al., 2002) exist, the development of multi stage simulations has been limited largely to 3–4 unit operations (e.g. Zhou et al., 1997). Whilst such models constitute a useful first step towards the production of robust and predictive modelling packages, it is the simultaneous consideration of all operations and replication of the interactions between them that is needed for the assessment of whole process feasibility. Advances made in this area thus far have centred primarily on establishing modelling methodologies through the study of generic process flowsheets (Zhou et al., 1999a/b). Bioprocess simulations remain considerably under–developed for the description of real production processes and without the presence of such examples in the literature, the pharmaceutical and biotechnology industries are unlikely to gain the confidence in modelling that is necessary for the potential advantages of the technique to be exploited fully. The aim of this chapter is to construct a simulation in order to evaluate the financial and technical impacts of implementing strategies designed to improve product yield and process duration, illustrating its utility by application to the CroFab™ process. The resulting work has been published (Chhatre et al.., 2006).

The studies presented in this chapter include:

- Comparing production scale data with model outputs to verify its accuracy and validity
- After this, a yield sensitivity analysis is used to identify the unit operations that provide the greatest potential for increasing product throughput
- A quantification of the interactions between process steps is also presented
- A process duration sensitivity analysis is used to identify the input factors that make the largest contribution towards total process duration
- Based on this data, the effects upon processing time as a result of changing properties such as feed or chromatography column volumes are then finally calculated

3.2 Modelling equations and data used in this chapter

3.2.1 Introduction

The model was set to run for four batches to simulate one complete pass per F_{AB} fragment to create all the material needed for a single final formulated product batch, with 500L of serum taken as the base case feed volume processed per batch (see Section 1.4.4 for further CroFab™ manufacturing details; the flowsheet (Figure 2 on Page 34) is reprinted below in Figure 10 for clarity).

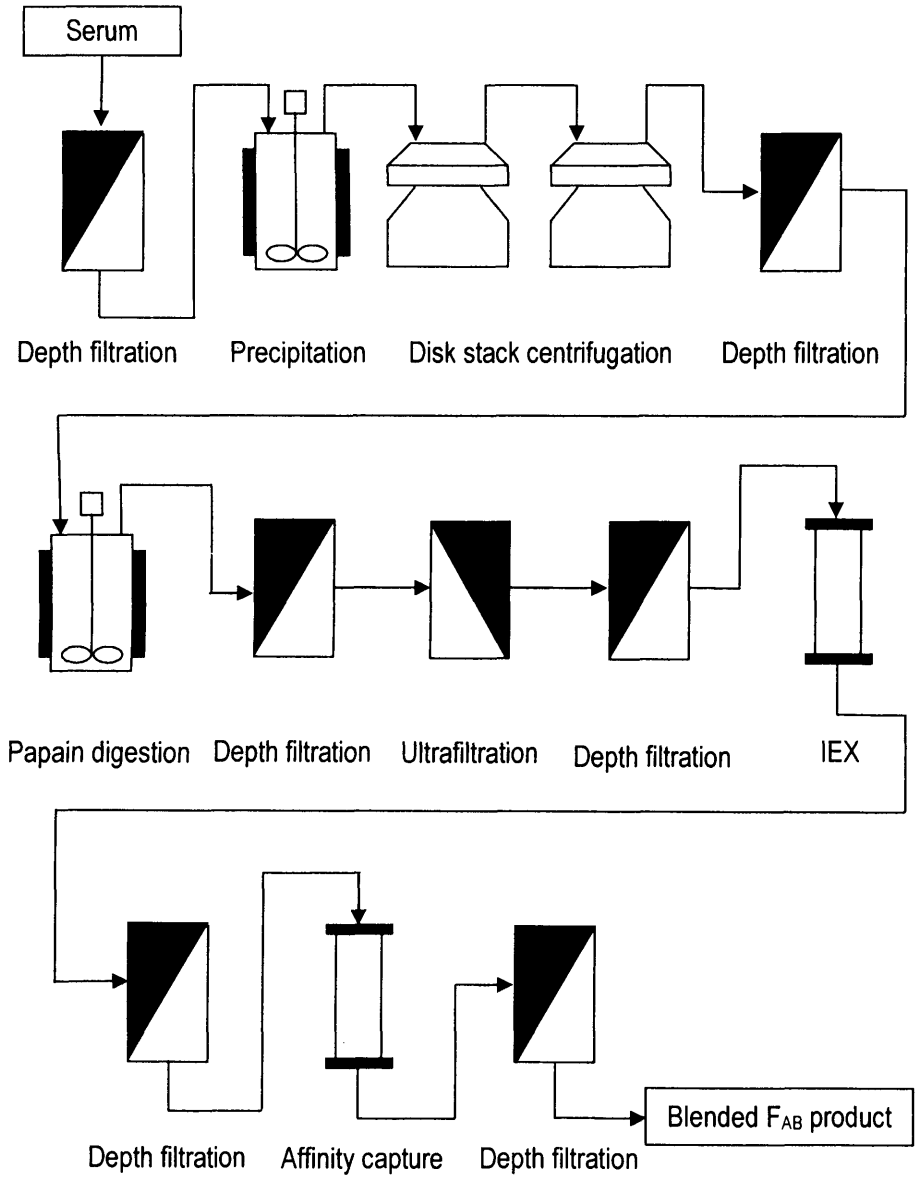


Figure 10: CroFab™ process flowsheet
(Reprint of Figure 2 on Page 34)

Assumptions made during model construction included:

- For this model, only the venom-specific F_{AB} component in each of the process streams was tracked i.e. the quantities of the impurities through the process were not taken into account
- Fugitive losses between stages were assumed negligible and were not incorporated
- For this model (and all subsequent models in this thesis), there was no value attached to the cost of using larger volumes of serum. Although using a higher amount of ovine serum would attract a greater expenditure (e.g. greater expenses at source in Australia or increasing the shipping costs), since the process envelope for these models concentrated upon those operations in the facility in Wales alone, the feedstock was assumed to cost nothing extra to the process as modelled
- Data used in model construction were provided by Protherics and included the following:

Initial serum feed volume for each F_{AB} stream	500 L
Number of passes simulated	4 (one complete blended F_{AB} batch)
Feed total IgG concentration for each F_{AB} stream	30 g/L
Feed venom specific IgG concentration per F_{AB} stream	8 g/L
Current affinity column linear flow rate	1.54 m/hr
Costs per litre for all buffers/chemical solutions were provided by Protherics*	
Buffer costs account for 5% of the total product cost per gram	

Table 11: Modelling data
Cost values were incorporated into the model via the Visual Basic user interface (see Section 3.2.9)

The following section lists the modelling equations used to calculate recoveries and durations from each process step operated by Protherics for CroFab™ manufacture. These equations were used as the basis of all models described in this thesis. Some of the relationships that calculate times and volumes use current manufacturing data and are scaled accordingly for the volume of operation; the current values are referred to as base case data below:

3.2.2 Precipitation

Steps modelled included: transfer feed to precipitating tank, heat vessels, add sodium sulphate and mix

For this model, as per operation at commercial-scale, the following data were employed:

- The volume of precipitating agent was set equal to the volume of the incoming feed serum
- The duration needed to transfer the feed into the precipitating tank was set to equal to the time needed for the preceding depth filtration
- Durations for heating the mixture, adding sodium sulphate and mixing were constant values

3.2.3 Centrifugation

Steps modelled included: CIP, connect and setup, flush, transfer feed, desludge, resuspend

The volume of the sedimented product ($V_{out,cent}$) was evaluated as a function of the fraction of solids in the input stream (f_{sed}), the volume of the input stream ($V_{in,cent}$), the volume of the resuspending agent used per cycle ($V_{precip,1\text{-}pass}$) and the number of cycles in centrifugation (N_{cent}). N_{cent} was equal to the number of precipitation vessels used in the previous step:

$$V_{out,cent} = f_{sed} V_{in,cent} + N_{cent} V_{precip,1-pass} \tag{4}$$

A number of the durations of different steps in centrifugal operation (CIP, connect & set-up, flush, desludging and resuspension) were constant. The time for transferring feed material into the bowl per cycle ($T_{transfer}$ [hr]) was calculated from the volume at the inlet per cycle and the volumetric flowrate ($Q_{v,cent}$):

$$T_{transfer} = \frac{V_{in,1-pass}}{Q_{v,cent}} \tag{5}$$

f_{sed}	Solids fraction in feed [–]
$V_{in,cent}$	Total feed volume [L]
N_{cent}	Number of centrifugal cycles [–]
$V_{precip,1\text{-}pass}$	Sodium sulphate volume per cycle for washing precipitate [L]
$V_{in,1\text{-}pass}$	Volume at the inlet per cycle [L]
$Q_{v,cent}$	Volumetric flow rate in bowl [L/hr]

3.2.4 Digestion

Steps modelled included: add papain, add activator, digest, terminate digestion

The volumes [L] of the 4% w/w papain/IgG (V_{papain}), its 11% v/v cysteine/EDTA activator and 11% v/v iodoacetamide terminator ($V_{activator}$ and $V_{terminator}$ respectively) used in the process were calculated as functions of the IgG mass and feed volumes to the step. The volume of papain was calculated from the mass of IgG fed into the digestion vessels ($M_{IgG,in,papain}$) [g] and the papain concentration (20g/L). $V_{in,papain}$ in the following refers to the inlet process volume [L] being transferred into the digestion vessels:

$$V_{papain} = 0.04 \times \frac{M_{IgG,in,papain}}{20} \quad (6)$$

$$V_{activator} = 0.11 \cdot V_{in,papain} \quad (7)$$

$$V_{terminator} = 0.11 \cdot \left(V_{in,papain} + V_{activator}\right) \quad (8)$$

Durations of each step during digestion were constant values (based on data from Protherics).

3.2.5 Ion exchange and affinity chromatography

Steps modelled included: CIP, equilibration, regeneration, load, wash and elution

Corresponding equations to the following were used for the venom-specific affinity step. The number of cycles required to process the feed to each column (load, wash and elute tasks) is calculated using the mass of antibody ($M_{IgG,in,IEX}$) [g], the volume of the bed matrix ($V_{matrix,a}$) [L] and the unit IgG loading used per cycle (L_{IEX}) [g/L]:

$$N_{IEX} = \left(\frac{M_{IgG,in,IEX}}{V_{matrix,a}}\right) \Big/ L_{IEX} \quad (9)$$

The volume of feed loaded onto the column per cycle ($V_{Load,IEX}$) [L] was then calculated from $V_{in,IEX}$ i.e. the total volume [L] applied to the column and N_{IEX}:

$$V_{load,IEX} = \frac{V_{in,IEX}}{N_{IEX}} \quad (10)$$

The relationship for the volume of eluted material was determined experimentally by Protherics (the value of 6 in Equation (11) is the hold-up volume) – personal communication, A. Newcombe, Protherics U.K. Limited):

$$V_{elute,IEX} = 3.5V_{load,IEX} - 6 \quad (11)$$

Volumes of buffers used for other steps such as preparing the column e.g. equilibration, regeneration or CIP were calculated as follows, using the CIP buffer as an example:

$$V_{CIP} = \left(\frac{B_{CIP}}{V_{matrix,B}}\right) V_{matrix,A} \quad (12)$$

B is the 'base case' buffer volume [L] determined experimentally by Protherics relative to a base case matrix volume ($V_{matrix,B}$). The actual volume of buffer required is then determined by scaling. Process times for a number of different steps in the process were calculated as a function of base case durations [hr], the height of the column (H) [m] and the linear flowrate (Q_L) [m/hr]. Thus for the duration of a CIP step [hr], the expression was:

$$T_{CIP} = \left(\frac{B_{CIP}}{V_{matrix,B}}\right)\left(\frac{H}{Q_L}\right) \tag{13}$$

3.2.6 Depth filtrations

Steps modelled included: setup filter, prime filter, filter the feed and flush

Common formulae were used for all depth filtration steps shown in Figure 2. The number of filter housings (F_{no}) was calculated from the total filtrate volume ($V_{out,filter}$) [L] and the maximum volume of the holding vessel for the filter ($V_{vessel\ max}$) [L]:

$$F_{no} = Round\ up\left\langle \frac{V_{out,filter}}{V_{vessel,max}} \right\rangle \tag{14}$$

The durations [hr] of the set-up, priming and flushing phases were calculated as functions of the number of filter housings needed. As standard practice, the setting up, priming and flushing steps take ¼ hour for every filter.

$$T_{setup} = T_{prime} = T_{flush} = 0.25 \times F_{no} \tag{15}$$

The time needed for filtration itself (T_{filter}) was calculated from the total inlet volume ($V_{in,filter}$ [L]) and the volumetric flowrate of material ($Q_{v,filter}$ [L/hr]):

$$T_{filter} = \frac{V_{in,filter}}{Q_{v,filter}} \tag{16}$$

3.2.7 Ultrafiltration

Steps modelled included: setup, sanitise, concentrate, diafilter, transfer, flush

The sanitisation buffer volume was set to a constant value, whilst the diafiltration buffer volume ($V_{diafilter}$) was a function of the concentrate volume ($V_{concentrate}$ – [L]) and the diafiltration ratio D (v/v):

$$V_{diafilter} = V_{concentrate} \times D \tag{17}$$

The volume [L] of flush buffer (V_{flush}) was scaled from the pooled inlet volume [L] prior to the initial depth filtration ($V_{in,pool}$) by dividing by 400 (based on experimental data from Protherics); the F_{AB} is present at this stage in 10mM phosphate and 10 mM sodium chloride: hence the value is multiplied by 10 + 10 = 20:

$$V_{flush} = \frac{V_{in,pool}}{400} \times 20 \tag{18}$$

The times [hr] for the setup, sanitise, transfer and flush steps were set as constant values, whilst for the concentration and diafiltration steps ($T_{concentrate}$ and $T_{diafilter}$) respectively:

$$T_{concentrate} = \frac{V_{out,UF} - V_{in,UF}}{Q_{v,UF}} \tag{19}$$

$$T_{diafilter} = \frac{V_{diafilter}}{Q_{v,UF}} \tag{20}$$

$V_{in,UF}$ Inlet process volume [L]

$V_{out,UF}$ Outlet process volume [L]

$Q_{v,UF}$ Volumetric flow rate in ultrafiltration [L/hr]

3.2.8 Definitions of efficiency and yield

The yield of F_{AB} from any given part of the process was defined using a step efficiency (E_S):

$$E_S = \frac{M_{F_{AB},out}}{M_{F_{AB},in}} \tag{21}$$

E values were defined for five parts of the process:

E_1 At the initial depth filtration step

E_2 At the filtration that follows the second centrifugation step

E_3 At the ultrafiltration step

E_4 At the ion exchange chromatography step

E_5 At the affinity chromatography step

Yield was then calculated as the product of the E factors:

$$\gamma_{F_{AB},final} = \prod_{i=1}^{5} E_i \qquad (22)$$

This is applicable to the four different F_{AB} products and the blended F_{AB} yield (across all four streams that are operated) was then taken to be the mean of the four $\gamma_{F_{AB},final}$ values [–].

3.2.9 Visual Basic data entry interface

This section gives an indication of how the Visual Basic interface enables the straightforward entry of data into the model and shows a sample screenshot. The shaded boxes in Figure 11 represent those data values that are automatically recalculated based upon other information about this or previous unit operations. (A), for example, shows the inlet feed volume, which is taken directly from the previous step, whilst (B) shows the number of separation cycles for which the centrifuge is run, which is equal to the number of precipitation vessels.

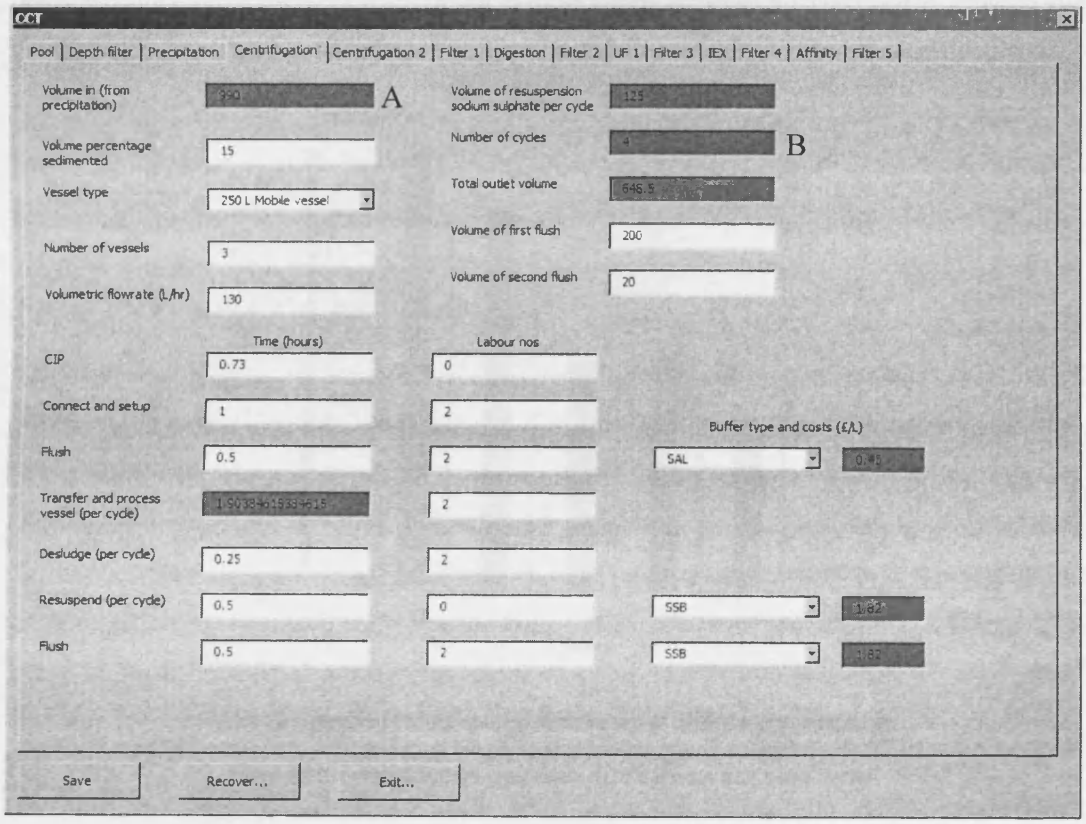


Figure 11: Sample screenshot of the Visual Basic interface
(Using the centrifugation step as an example)

This logic is used throughout the interface to facilitate updating of the values as a result of changes made elsewhere. The other entry boxes shown in white enable the entry of other values, such as identities and volumes of buffers or the durations of individual steps e.g. for CIP, desludging or

flushing etc; all the necessary data values are then saved to a spreadsheet, which is connected to Extend™ to enable the model to run.

3.3 Results and Discussion

3.3.1 Verification of model outputs

To verify whether the model accurately reflected operation at large scale, the total process yield as well as intermediate and final product concentrations were compared to production scale data. The predicted total process yield was 30%, whilst the actual production value is 25%. Although these values are not identical, the fact that this chapter was based upon an analysis with comparatively simple black box modelling equations, it may not be unreasonable to expect a 5% discrepancy between manufacturing and simulation outputs. Figure 12 plots the intermediate and final quantities of specific F_{AB} mass produced throughout the process, calculated by multiplying the model abd production specific Fab concentrations by the process volumes present at each stage. The predicted data values (black bars) are similar to those values achieved from production, implying that the simulation provides a good representation of the performance achieved in industrial operation.

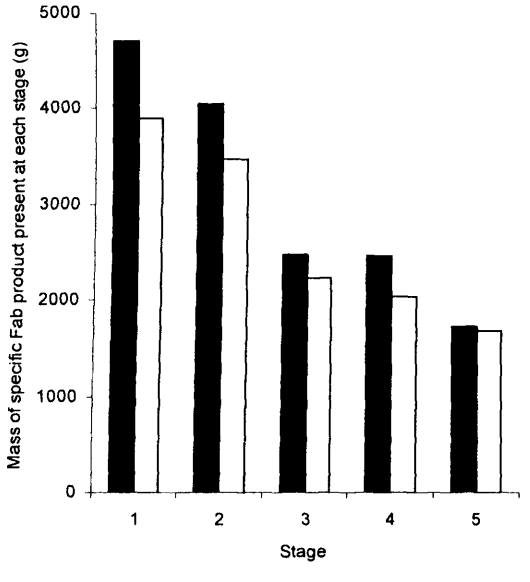


Figure 12: Venom-specific F_{AB} masses achieved at different process steps
Numbers refer to the amounts produced from the numbered groups of unit operations identified in Section 3.2.8 (black bars = model output data; white bars = production scale data)

Furthermore, the cost of goods value that was predicted by the model was shown to Protherics; although for reasons of commercial sensitivity the company was unwilling to disclose the exact production value, it *was* able to provide a qualitative indication that the model prediction was similar to the actual cost. Although it is recognised that this is not a quantitatively rigorous comparison, this assertion provided somewhat greater confidence in the simulation.

It should be noted that in this thesis, data provided by Protherics which were entered into the model were averages from production and were assumed to be fixed for simplicity i.e. a deterministic analysis was carried out. In reality, values might vary depending upon the errors inherent to typical analytical assays, such as for:

1) The feed and intermediate specific titres as judged by a small scale affinity column
2) The total IgG feed titre as determined by a Protein G column
3) The initial and in-process impurity levels (primarily albumin)
4) The final amount of product, for example judged by absorbance measurements

Protherics indicated that the analytical errors would lie in the range of +/- 10%, meaning that the exact values entered into the model could alter in that range and hence the outputs could potentially change, depending upon exactly which data were selected. Thus, although what follows below was developed on the basis of assuming a fixed set of values for the input parameters, if the modelling framework was to be applied to the CroFab process again, the impact of measurement errors upon the input data values and hence the quality of the modelling predictions would need to be taken into account. This could be achieved by running the models over the range of input data values obtained from analysis and hence determining a range of possible output values.

3.3.2 Yield sensitivity analysis

The relative contributions of the different steps in the process towards the blended F_{AB} yield (i.e. overall averaged final yield achieved across all four operated streams) were identified by examining the impact of changes made in the F_{AB} efficiencies E_i of the various unit operations. Values of E_i for the main groups of unit operations were increased relative to their original values and the blended yield recalculated each time, producing the linear lines in Figure 13. Normalised blended yields were defined as the gradient of each of these lines and specify the percentage improvement in blended F_{AB} yield achieved for every percentage improvement in step efficiency. Unit operations that have larger normalised values provide greater potential for increasing the final product mass and hence should form the focus of process development effort. Figure 14 represents these results and shows that the papain digestion and affinity stages are predicted to have a 30% greater impact on the normalised blended yield than do the precipitation, centrifugation or ion exchange steps.

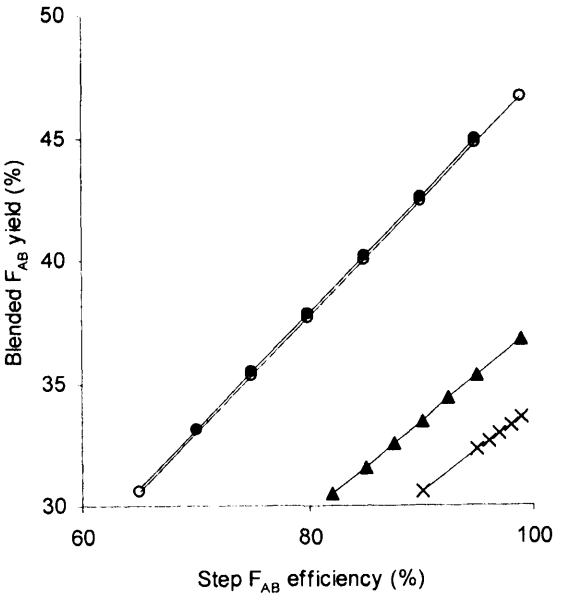


Figure 13: Impact of percentage step efficiencies upon percentage blended F_{AB} yield
Filled circles = affinity capture; empty circles = papain digestion; filled triangles = precipitation and two centrifugations; crosses = IEX. Precipitation and centrifugation were grouped together as this was how Protherics' manufacturing data were available

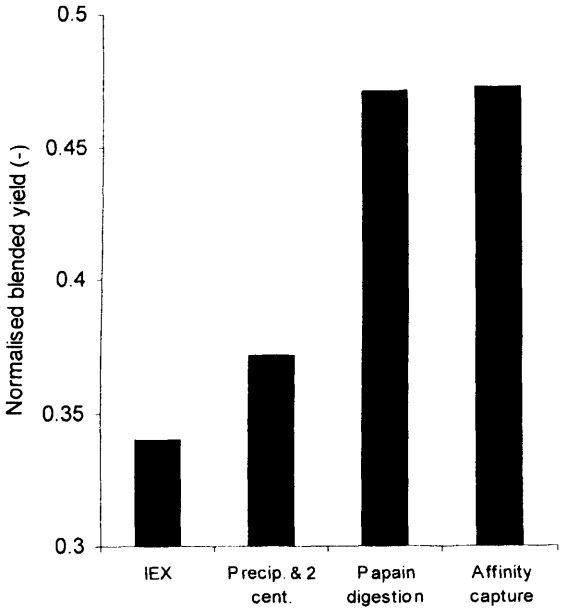


Figure 14: Sensitivity analysis of blended yield to changes in step efficiencies
(Expressed on a normalised basis - normalised blended yield = % increase in blended yield divided by % rise in step efficiency). The affinity and papain steps have the most significant impact upon blended yield)

3.3.3 Quantification of interactions between process steps

The effect of changing the efficiencies of both the papain and the affinity stages upon the blended F_{AB} yield was analysed to quantify the impact of simultaneous changes made to the two stages. The three lines on Figure 15 represent constant yield values. The extreme left hand line represents the current blended yield of 30% and was generated by altering papain stage efficiencies in the 45-95% range and at each alteration, manually iterating upon the affinity efficiency needed to retain 30% yield. 35% and 40% curves were created in the same fashion. Practical industrial limits of the affinity and papain stage efficiencies were set at 95%.

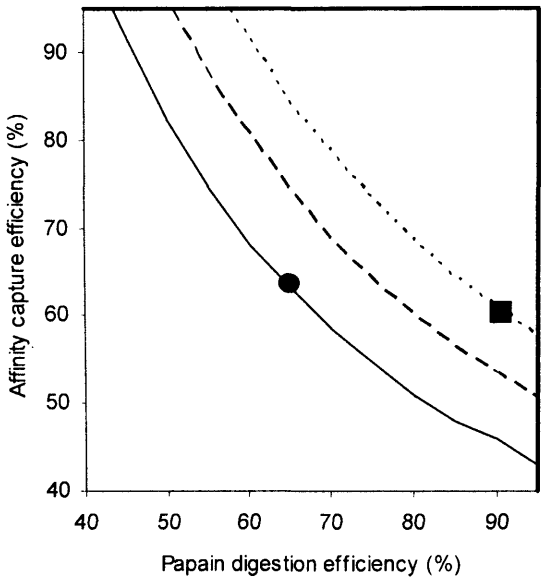


Figure 15: Blended F_{AB} yield achieved when improving the efficiencies of papain and affinity stages for all four products simultaneously
(30% —; 35% – – –; 40% -----). Values are bounded by 95% efficiency limits for both steps. • = current operating position on the 30% blended yield line; ■ = new location on the 40% blended yield line, as described in the text

Each yield line also represents constant blended product masses. Since total costs do not change on a given line, each one therefore also specifies constant costs per gram. Hence, the graph can additionally be used to assess the financial impact of changing the stage efficiencies. For this model, only buffer/solution costs were taken into account. Data from Protherics suggested that these represent approximately 5% of total costs and so for the calculation of cost of goods, total solution costs were adjusted accordingly to arrive at the final result. The calculated cost of goods values for the current manufacturing operation was verified to be in the range of actual commercial values (commercial data not shown). The points on Figure 15 show that if the papain recovery is increased from the current operating position by 25% whilst retaining approximately constant affinity efficiency, the overall yield rises by 10% whilst the cost per gram is cut by 25%.

3.3.4 Extent of optimisation required for yield improvements

The four F_{AB} specific matrices have different binding capacities for their products. The effect of varying the affinity capture efficiency value on the blended F_{AB} yield was investigated to estimate the level of improvement needed to produce a useful increase in final blended F_{AB} yield. Figure 16 shows that changing the efficiency of any one matrix alone has only a moderate effect upon yield. Significant changes to raise the yields to 85% or above for all four matrices simultaneously are required to achieve substantial increases overall. This implies upgrades in F_{AB} production cannot be based upon one venom-specific matrix alone, but that a more development-intensive route of successive improvements to several matrices is needed to achieve commercially useful results.

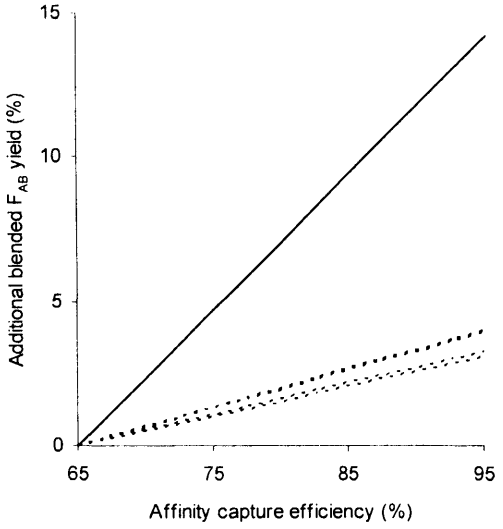


Figure 16: Blended F_{AB} yield achieved as a function of affinity capture efficiencies
In excess of current averaged value of 65% for F_{AB} 1 – 4: – = blended yield; --- = individual specific F_{AB} products. A 30% increase in the efficiency for each F_{AB} results in an increase of just less than 15% blended yield

3.3.5 Process duration sensitivity analysis

In addition to investigating the effects of process changes on blended yield, the model was also used to determine which input parameters have the greatest impact on total process duration. The key variables likely to affect duration were altered one at a time in the ranges shown below, whilst others were kept constant (Table 12 on Page 75).

The results in Figure 17 show that the feed volume, the affinity flow rate and the affinity matrix volume make the greatest impacts upon process duration, whilst the volume of ion exchange matrix and the centrifugal flow rate have the smallest effects. This reflects, in particular, the fact that the affinity operation is the rate limiting step in the process and that making alterations here results in the largest change in process duration compared to the other changes in Table 12. The following investigations therefore focused on the impacts of changing the feed volume, affinity matrix volume and affinity flow rate on process time.

Input variable	Percentage change relative to the base case
Feed volume (L)	–20/+100
Affinity flow rate (m/hr)	–30/+200
Volume of affinity matrix (L)	–50/+200
IEX flow rate (m/hr)	–30/+200
IgG titre (g/L)	–33/+33
Volume of IEX matrix (L)	–50/+200
Centrifuge flow rate (L/hr)	–25/+15

Table 12: Percentage ranges over which input variables were altered to determine process duration sensitivity ***Percentages were chosen based on discussions with Protherics. The IgG titre range was selected based on the variability observed in different production batches of ovine sera (Newcombe et al., 2006)***

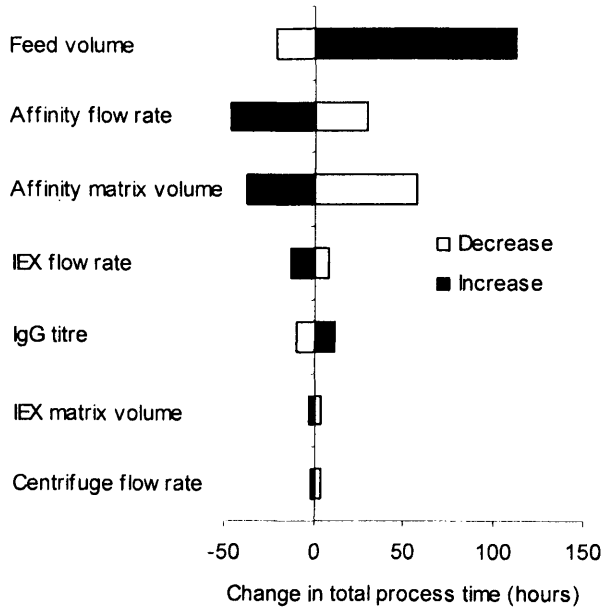


Figure 17: Tornado diagram showing sensitivities of process duration to input variables ***Values show the change in duration relative to the base case (current operation)***

3.3.6 Impact of increasing affinity matrix volume on unit operation time

Since the number of chromatography cycles (N) is calculated based upon the volume of affinity matrix (Equation (9)), increasing the quantity of resin results in a reduced value of N and reduced duration. The possibility of decreasing the chromatography cycle time by increasing the volume of the matrix was investigated. The results in Figure 18 show that, for example, a volume increase of nearly 70% (14 L) for each F_{AB} resin would be needed to provide a cumulative time saving of over 80 hours over the four process runs, although this would need to be balanced against the associated increase in matrix cost. In order to accommodate the extra resin volume, additional column shells would be needed. Although the availability of suitably sized columns and the validation of these have not been considered within this study, such constraints would need to be accounted for when assessing the feasibility of increasing the matrix volume by 70%.

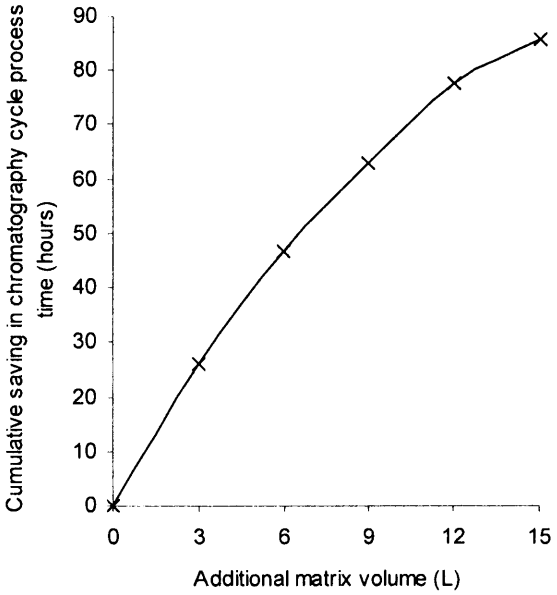


Figure 18: Cumulative saving in chromatography cycle process time
The cumulative saving is defined as the sum of the time savings for the load, wash and elution affinity cycles as the volumes of the four F_{AB} affinity matrices are increased

3.3.7 Comparison between papain and affinity improvements

Sections 3.3.3 and 3.3.4 showed that the blended yield can be raised from 30 to 40%, by increasing the levels of recovery in either the digestion or affinity capture steps. Data from Protherics indicated that the product loss currently observed in the papain stage is due to digested F_{AB} being inactivated or precipitated and thus lost during the depth or ultrafiltration steps. These effects might be mitigated if a suitable buffer such as phosphate buffered saline was used to wash the digested material, potentially resolubilising and stabilising the F_{AB}. Alternatively, recovery in the affinity step could be raised by using a larger volume of resin for each F_{AB} type. Assuming that step yield was directly proportional to matrix volume (increasing the matrix volume by, for example, 5% would increase the number of binding sites available for interaction with the product by 5% and hence raise product recovery by the same amount), then with reference back to Figure 16, using an extra 20% of the matrices for each F_{AB} type would achieve a recovery of 85% at the affinity stage and hence an additional 10% blended F_{AB} yield. Such results would need to be verified experimentally. Table 13 outlines the impacts of these changes.

	Change	Advantages	Disadvantages
Papain	Washing digested material with phosphate buffered saline	Achieve 40% blended yield with minimal additional manufacturing cost	May require PQ and PV, but probably at minimal expense of cost and time
Affinity	Increase in matrix volume by 20%	40% blended yield with cumulative time saving of 30 hours	Matrices' costs rise by 20% and extra validated chromatography columns would be required

Table 13: Comparison of changes to digestion and affinity stages
Given the heavier expenses of the affinity option, it is likely that the change to the papain step may be preferential

Washing the digested material with phosphate buffered saline would be a cheaper way of attaining the 10% improvement and has less impact upon cost of goods than increasing the volume of capture resin, especially when considering the significant expense of affinity matrices. Increasing the resin volume by 20% (4 L) would, however, save a total of 30 hours (see Figure 18).

3.3.8 Matching initial process volume to available time window

As the time sensitivity analysis in Figure 17 highlighted, the volume of starting serum and flow rate in the affinity column each have significant implications for total process duration. The effects of changing these parameters were studied so as to evaluate the trade off against the total process time needed to make all four F_{AB} products at a theoretical base case 30% yield. Data suggested that the rigid controlled pore glass matrix (Bioprocess Division, Millipore (U.K.) Limited, U.K.) used in the manufacture of the venom-specific affinity matrix would be able to withstand pressures from flow rates of over 10 m/hr (Millipore application note TB1010EN00, 2002) and hence process duration was evaluated at flow rates of the current 1.54 m/hr as well as 2– and 3–fold higher. Figure 19 shows that when operating at the lowest linear velocity in the affinity column, process time increases by 120 hr (25%) if feed serum volume is doubled. Similarly, a 20% increase in time (100 hr) is seen at the highest velocity for the same feed volume increase. The majority of the improvements in process duration can be achieved by making relatively small increases in flow rate. Hence if the feed volume is doubled and the linear flow rate is increased from the base case to 3 fold higher, of the total time saving, 75% is achieved between the current and doubled flow-rates, whilst the remainder occurs in the two-fold to three-fold range.

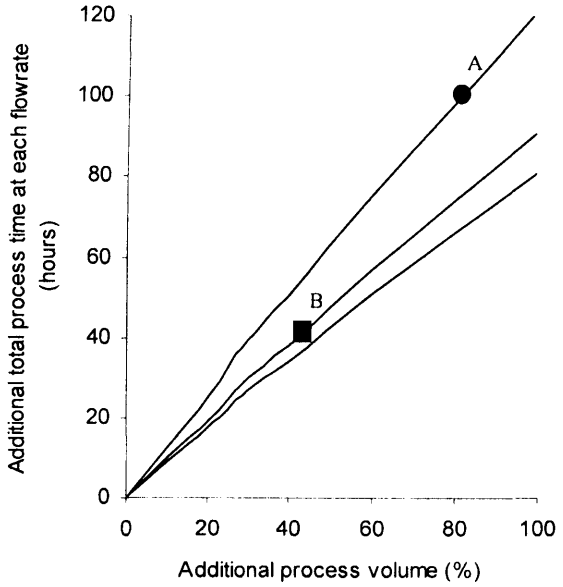


Figure 19: Impact of varying the initial serum volume above the base case 500L initial serum on the additional total process time needed to manufacture the four F_{AB} products at each linear flow rate.
Current affinity flow rate = 1.54 m/hr (upper line); 2-fold and 3-fold higher flow rates have also been plotted below respectively. Points A and B refer to calculations carried out in the main text

These results allow additionally for the initial feed volume of serum for processing to be matched to the amount of time available within the schedule of the plant. If, for example, a given batch takes longer than expected, perhaps due to resource constraints, then this may affect the time that is available to process subsequent batches. Figure 19 can be used in such a scenario to match processing volumes to the timeframe available. Point A shows operation occurring at a linear flow rate of 1.54 m/hr and with a feed volume of 80% over the base case conditions, which are located at the origin for the 1.54 m/hr line in Figure 19. This requires a total process time of approximately 566 hours, 100 more than the base case. If the flow rate is doubled and the additional process time available over the base case becomes only forty hours, a maximum feed volume of just over 40% over the base conditions can be processed (point B).

3.3.9 Impact of feed and affinity matrix volume on total process duration

The combined effects of the requirement to process additional feed material and the use of extra affinity matrix for the four F_{AB} products simultaneously are presented in Figure 20. The concave upwards plot shows that the process time reduces rapidly when the presence of small additional matrix volumes combines with the need to handle large additional process volumes. Thus, when doubling the feed volume, 35% of the reduction on the time axis is exhibited in the first 15% increase on the matrix volume axis. This, however, equates to only a 4% reduction in total process duration and suggests that increasing the volume of affinity matrix by relatively small amounts when processing large feed volumes does not reduce total process time significantly. Hence if this strategy is considered by Protherics, then it would need to make significant investment into additional column sizes in order to achieve appreciable reductions in process time.

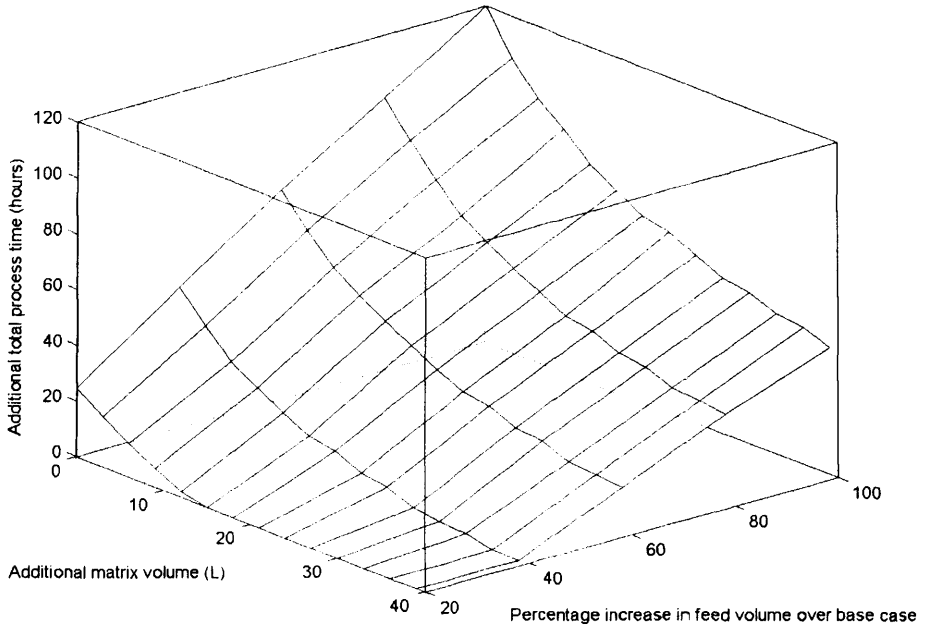


Figure 20: Impact of matrix and feed volume upon additional time required to process the material for the four F_{AB} species relative to the duration when processing 500L of feed using the current matrix volume

3.4 Conclusions

This chapter has presented an analysis of an FDA-approved antibody purification process and created a simulation which assesses a series of proposed strategies designed to achieve manufacturing improvements to determine their impact upon product yield and process durations. Unit operations have been ranked and the interactions between them captured. The research identified the papain digestion and affinity chromatography stages as being especially critical in determining blended F_{AB} yield and process time. The trade-offs presented between key attributes provide useful information in guiding enhancements to the manufacturing process. Nevertheless, the black box material balancing models employed are relatively simplistic and thus do not give a proper representation of the impacts of making alterations. Hence the next chapter introduces a new multi-layered modelling methodology based on mathematically rigorous mass balance expressions in order to enable more accurate calculations and thereby facilitate comparison of bioprocess options.

4: A MULTI-LAYERED SOFTWARE METHODOLOGY FOR THE EVALUATION OF BIO–MANUFACTURING PROCESS OPTIONS

4.1 Introduction

The current generation of commercially available bio–manufacturing models, such as those presented in the previous chapter, are predominantly black box in nature and utilise simple material balancing methods such as assigning yield fractions to quantify levels of product recovery from each unit operation (Petrides, 1994). By limiting models in this way, the impact of changing more fundamental parameters such as the flowrate in a solid–liquid separation step upon the composition of the outlet streams and hence recovery from that step cannot be investigated properly. In order for this type of analysis to be conducted, 'detailed' material balance models are required that use mathematically rigorous expressions to quantify recovery, thus providing greater insight into the likely manufacturing performance and so creating greater confidence in simulation outputs. An example is the use of the Rosin–Rammler–Sperling–Bennett grade efficiency expression for the quantification of particulate recovery in a centrifuge as a function of flow rate and particle size (Clarkson et al., 1996) or the Thomas model for calculation of product capture in a chromatography step as a function of feed concentration and matrix capacity (Montesinos et al., 2005). To date, very few simulations have used rigorous material balance models to quantify product recovery and those which have tended to be established within simulation environments such as MATLAB®. These environments do not account for dynamic resource constraints such as the levels of manufacturing staff or equipment available for processing (e.g. Groep et al., 2000 or Varga et al., 2001), although in some cases, additional software packages have been developed to handle dynamic resource constraints such as Stateflow® for MATLAB®. As mentioned in the first chapter, these constraints are key features of bio–manufacturing operations and should be captured in any simulation in order to give a more realistic representation of product throughput and batch times. A bioprocess simulation which was both based on rigorous material balance equations *and* established in a dynamic simulation environment would provide valuable process insight and allow the evaluation of bio–manufacturing strategies designed to optimise outputs such as product recoveries, costs and batch times. The development and application of this type of process simulation is the subject of this chapter. This work has been published (Chhatre et al., 2007c).

4.2 Challenges in using rigorous material balance models

As outlined above, rigorous material balancing models provide a more accurate representation of process operation than black box approaches and so give greater confidence in the predicted outputs. Despite their advantages, however, the use of rigorous material balance equations in simulations may necessitate extensive and time–consuming experimentation in order to collect all the data needed to populate these equations (Zhou and Titchener–Hooker, 2003). Additionally, if

there are many potential manufacturing options under consideration, there may be prolonged model execution times compared to black box simulations. Should any particular option turn out to be unfavourable, the experimental and computational time spent on assessing them will ultimately represent a waste of resources. This chapter proposes a software methodology that is designed to avoid this problem and thereby facilitate the rapid evaluation of the attractiveness of process options. The methodology comprises three layers, each containing its own dynamic simulation. The first layer contains a simplified version of the overall process model and requires relatively sparse input data. All the postulated manufacturing options are evaluated rapidly in this layer. Subsequently, the least favourable with respect to a series of pre–determined metrics, such as production levels, costs or times, are screened out and not evaluated any further. The more promising candidates are retained and then evaluated in a similar fashion in the lower two layers, each containing a more computationally–intensive and accurate model that is populated by larger quantities of input data and which therefore requires more data from time-consuming experimentation. By adopting this approach, the total amount of input data needed to evaluate all options can be kept to a minimum and time–consuming laboratory work can be focused on those manufacturing strategies which show the greatest potential. The final output emerging from the methodology is a set of the most favourable process options. This chapter describes the development of the methodology and illustrates it utility through its application to the CroFab™ process in order to identify the most favourable options for achieving manufacturing improvements. Strategies that have been examined include increasing the feed volume, replacing centrifugation with microfiltration and replacing precipitation/centrifugation with a Protein-G column.

4.3 Chapter structure

The remainder of this chapter is now structured as follows: a description of the modelling framework is given, followed by details of how the manufacturing models were set–up and how they were used to evaluate multiple process options. Application of the software is then demonstrated through the evaluation of a series of proposed changes to the CroFab™ process.

4.4 Methodology

4.4.1 Structure of the modelling methodology

Every layer in the methodology contains its own dynamic manufacturing model. Figure 21 (Page 83) shows a schematic of how the methodology operates and Table 14 summarises the differences between the layers. Since the first layer is less accurate than the second layer, which in turn is less accurate than the third, acceptance criteria specifying which manufacturing options should be screened in and out at each layer are set accordingly. Less stringent criteria are used in the first layer than the second, which in turn uses less stringent criteria than the third. This prevents

screening out options in the upper layers which are favourable, but which would be eliminated due to inaccuracies in the simulations. Details of acceptance criteria are provided later in section 4.5.5/6.

	First layer	Second layer	Third layer
Material balance models	Simplified versions of 'detailed' material balance equations are used e.g. a simplified grade efficiency expression for the centrifuge (Svarovsky, 2000) – see Table 17 on Page 92		Rigorous forms of equations e.g. a rigorous grade efficiency expression is used for the centrifuge (Table 17)
Volumetric recovery	Calculated by assigning a percentage recovery for every unit operation		Calculated from buffer volume addition/reduction by concentration
Model organisation	Individual sub–tasks in every unit operation are grouped together e.g. in chromatography, preparatory steps such as matrix packing are grouped into an 'initiation' subtask; operative steps (loading/washing/eluting) are grouped into an 'operation' subtask and post–operative steps such as re–generation are grouped into a 'termination' subtask	Every subtask for every unit operation is modelled explicitly, each with its own individually calculated duration. Costs are calculated based on resource usage as outlined below	
Resource constraints	For simplicity, only buffer consumption is tracked and an approximation for the fraction that buffer costs contribute to overall process expenditure is used to determine cost of goods (see below).	Resource pools exist for: labour, buffers, process vessels, equipment (centrifuges, filter housings and column shells), filter membranes and chromatographic media. Models accumulate costs of new resources purchased when previous batches are exhausted	
Time & cost calculations	For each of the three grouped sub–tasks, base case values for times and buffer costs are assigned relative to a base case feed volume. One cost and one time is assigned to each of the initiation, operation and termination sub-tasks for each unit operation. When the feed volume increases, times/buffer costs are scaled proportionately	Process costs are incurred dependent on the utilisation of resources. Resource utilisation is determined by equations that use manufacturing data such as process volumes, product concentrations, column capacities etc., to calculate quantities of resources needed. Step–specific equations are used to determine the durations of each unit operation	

Table 14: Organisation of models in the different layers of the methodology
See Table 17 for further details of material balance equations

Successive layers incorporate additional detail and refinement in terms of:

- the degree of complexity of the material balance and volumetric recovery equations used
- the number of resource constraints applied in the simulation
- the organisation of the model
- the specific calculations for determining batch times and costs

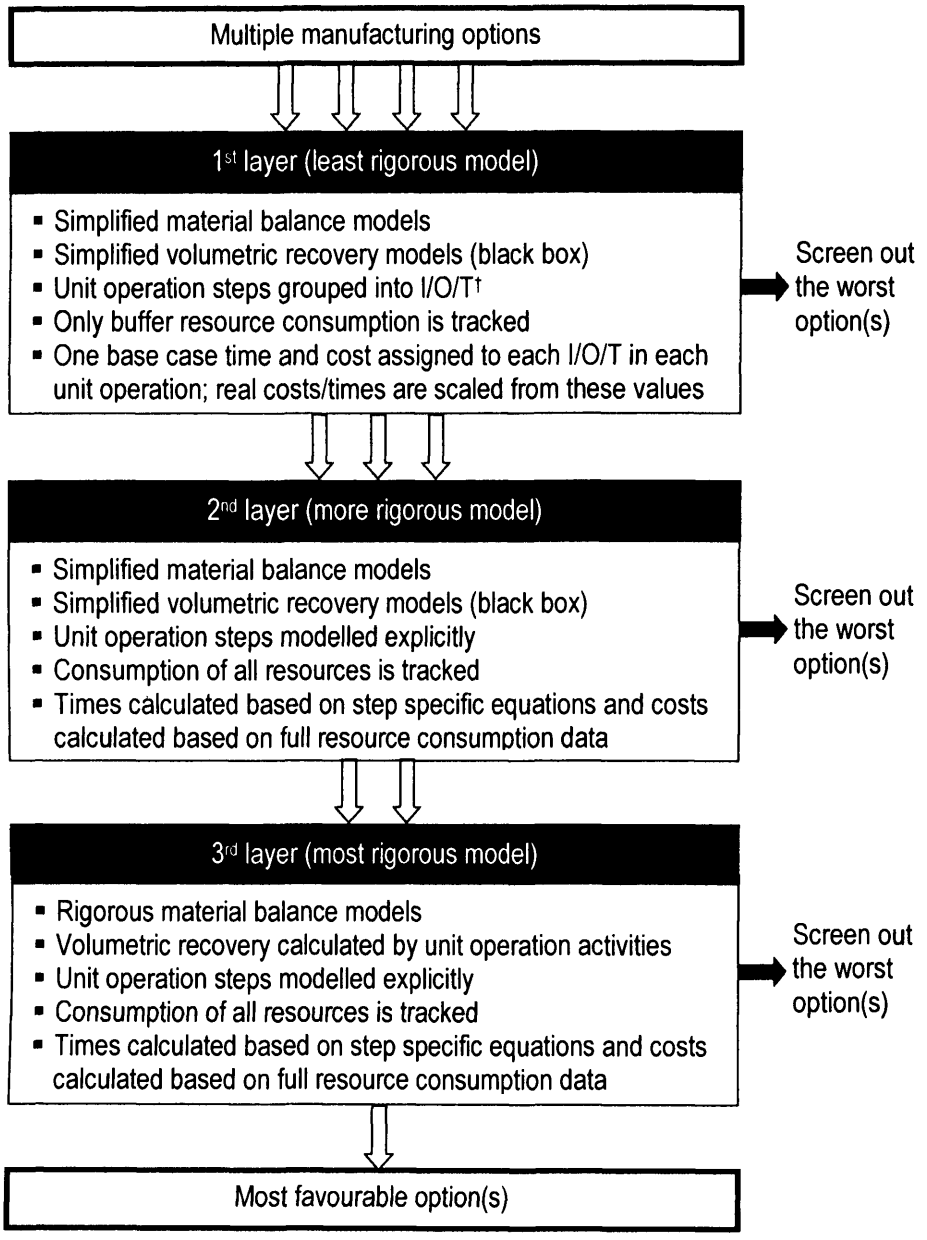


Figure 21: Schematic of how the methodology is used to evaluate process options
(See Table 14 for more details. †I/O/T = initiation, operation and termination; steps in each unit operation are grouped into these sub-tasks in the first layer)

4.4.2 Details of the different layers

4.4.2.1 First layer

The first layer contains a simplified version of the model used in the subsequent lower two layers and so allows an initial rapid evaluation of manufacturing options with minimal input data requirement. The first layer uses simplified versions of the 'detailed' material balance equations, such as a simplified expression for determining product recovery in a chromatography column [C_N = number of cycles; V_m = matrix volume; DBC = dynamic binding capacity (g/L)]:

Quantity bound = $C_N \times V_m \times DBC$ (23)

For simplicity, only one resource constraint is considered and an estimate for the percentage contribution of that resource towards total process expenditure is used to determine the cost of goods value (see the modelling assumptions section later for details on how this was applied in the industrial case study). This avoids the need to spend time gathering cost data for all other resources until the second layer is reached, by which point the least viable options have been screened out so as to focus effort on the more promising candidates. In the first layer, the model is organised such that steps required within each unit operation are grouped appropriately into initiation, operation and termination subtasks (see Table 14), with one buffer cost and one time value assigned to each of these. Inferior options at the first layer in terms of the metrics given in Table 15 are then screened out, leaving the remainder to be assessed by the second layer model.

Manufactured product mass (g)
Cost of goods (£/g)
Batch times (hours)

Table 15: Manufacturing metrics used to assess process options
For the industrial case study, batch time was taken to be the duration needed to manufacture the first blended, concentrated and filtered batch of product

4.4.2.2 Second layer

More rigorous models are deployed in the second layer, primarily to incorporate constraints for all resources required for the process. As a consequence, process costs are now calculated based upon resource utilisation levels, rather than scaling upon buffer consumption. In the problems simulated in this work, the second layer models took approximately four times longer to compute than the first layer models. The second layer uses the same material balance equations as the first layer. The simulation is also organised such that every sub–task is modelled explicitly (i.e. each task is modelled with its own time and cost, as opposed to the grouping of steps into initiation, operation and termination sub–tasks that occurred in the first layer). Again, the most inferior options emerging from the second layer with respect to Table 15 metrics are screened out, leaving the most promising options to be subjected to the third layer model.

4.4.2.3 Third layer

The third and final layer is much more computationally intensive and hence time–consuming than the upper two layers. The third layer takes an order of magnitude longer to execute than the second layer, owing to the use of the fully rigorous material balance equations, such as the Thomas model (Montesinos et al., 2005) for simulating chromatographic breakthrough curves. Since this layer is challenged with a reduced set of only the most promising manufacturing process options, the user makes the most efficient use of (a) the longer computational time needed to run a third layer model

and (b) in collecting only the relevant input data to the model, such as parameter estimation for the detailed material balance equations collected from laboratory experimentation.

4.4.3 Construction of manufacturing models in the methodology

For ease of data manipulation, input values for the manufacturing models in each of the layers were entered into Visual Basic for Applications user interfaces (Microsoft® Excel XP) and saved to spreadsheets. These were connected to the main manufacturing models themselves, which were constructed in Extend™. For the manufacturing models described in this chapter, the following entities were constructed by placing suitable combinations of blocks onto the Extend™ workspace:

- Resource pools for labour, buffers, process vessels, equipment, filter membranes and chromatographic media (buffers only for the first layer model)
- Methods to draw from and replenish the resource pools
- Methods to calculate durations of activities in every unit operation
- Material balance calculations for every unit operation

Each of the three models was modularised, such that there was one code block to represent the material balancing calculations, one block to represent the calculation of process durations for a given unit operation etc. Adopting this approach reduced duplication of model functions and enabled faster debugging and execution of the model. The simulation comprised icons for each unit operation, copies of which were cloned to a workspace and connected together in the appropriate order to simulate the desired flowsheet. As a product batches moved through the process, represented by an item, it called the relevant unit operation model, executed instructions encoded in the Extend™ blocks and accumulated time, cost and material balancing data. At the end of a simulation run, data acquired for all manufactured product batches were collected together and exported to a Microsoft® Excel spreadsheet automatically to enable further analysis.

4.4.4 Details of model set-up in Extend™

To illustrate the way in which a model is constructed, Figure 22 (Page 89) gives an example of what the Extend interface looks like and how it is used to develop simulations, using the third layer model of the CroFab™ model described later as an example. The screen is divided into two parts – the workspace on the left hand side onto which the unit operation blocks in the required sequence for the process, whilst on the right hand side, there are various palettes, each of which represents a combination of blocks that carry out specific functions such as resource allocation or material balancing. The first block placed onto the workspace is the Start block, which is responsible for 'creating' items that pass through the process and assigning feed properties such as volumes or concentrations. The Start block is then connected to a box marked 'unit operation'; double clicking on this brings up a dialog box that allows the end user to select a specific unit operation type.

Connection between the blocks is achieved by the thick white lines, which represent the streamlines down which items (i.e. batches) pass when moving from unit operation to unit operation. The remainder of the unit operation blocks are then added in the required order to simulate the process, with data transfer lines (the thin black lines) between unit operations carrying information such as process volumes. Finally, the End block removes items and sends data to the Output data block, which exports output information back to the spreadsheet.

4.4.5 Overall Ranking

Assessing the impacts of changing a process is complicated by the fact that some performance metrics may improve whilst others deteriorate – for example, increasing feed volumes will result in a higher product mass at the expense of extended batch times. Alternatively, removing a unit operation may potentially reduce manufacturing costs, but increase the purification burden downstream and hence necessitate revalidation of the remainder of the process. Such conflicts can make it difficult to assess whether to pursue a new process option or simply to retain the existing manufacturing process. In this chapter, the decision-making process was simplified by combining performance metrics into a single value using multi-attribute decision-making (MADM) in order to quantify the desirability of a strategy (Farid et al., 2005). A range of different MADM methods are available; for example, the lexicographic technique (Volgenant, 2002) involves ranking the different outputs in the order of their importance and then optimising each model output successively in this order for as long as a choice remains until a unique point is reached. The Electre method (Steuer and Na, 2003) involves the outranking principle, in which a solution to a model is said to outrank other options if it as good or better than others with respect to most of the key outputs and not significantly inferior for one or more of the remaining outputs. Another approach that is often seen in the literature is Pareto optimisation (e.g. Zhou and Titchener-Hooker, 2003), which involves the generation of a series of non dominated (optimal) solutions. A solution is said to be Pareto optimal if it is not possible to improve one output objective without causing a deterioration in at least one other. Although these are valuable techniques for allowing multiple-objective decision-making, due to the relative complexities involved with using these methods (George et al., 2007), this chapter instead uses a classical additive weighting approach, which is applied regularly as it provides a straightforward way to combine several output criteria into a single number that quantifies the feasibility of a strategy (Hwang and Yoon, 1981).

In the classical approach, results for each attribute (e.g. product mass, development costs or manufacturing process times) derived from simulations were normalised to a zero to one scale, with the zero bound representing the worst possible value and the one bound representing the best. For example, for product mass, the normalised value is given in Equation (24):

$$N_M = \frac{M_A - M_0}{M_1 - M_0} \tag{24}$$

N_M represents the normalised product mass, M_A is the actual product mass calculated by the simulation, M_0 is the lowest product mass out of all simulation runs (set to represent the zero bound) and M_1 is the highest product mass, thus set to represent the one bound. Similar formulae can be applied to all other metrics. Normalised values are each further multiplied by a zero to one weighting so as to place particular emphasis on those metrics deemed especially important, such that the more important metrics attract higher weightings. The sum of the weightings is equal to one. By applying this approach, it is possible to arrive at an assessment of a strategy consistent with commercial aims and objectives. Addition of the component weighted and normalised values, as calculated by Equation (24), creates a single metric, the Overall Rank (OR):

$$OR = \sum_{i=1}^{n} w_i N_i \tag{25}$$

Here, w_i and N_i respectively represent the weighting and the normalised values of the i^{th} performance metric. The Overall Rank has a value between zero and one, respectively representing the least and most attractive outcomes resulting from the development and implementation of process change strategies. The Overall Rank for the existing manufacturing process is also determined to provide a benchmark against which proposed options are compared.

It is recognised that the perceived importance of different output metrics as represented by the weighting value may change depending upon the circumstances facing a company from week-to-week or month-to-month. For example, if product throughput (mass of purified product generated per unit amount of time) is high and well above the level needed to maintain adequate market supply, then the impact of any process change upon throughput may not be that crucial. Conversely, if market demand increases appreciably, the ability to supply product by achieving a high enough throughput will become more critical and hence greater emphasis will be attached to that objective at that time. Since the process changes considered in this thesis are are intended for implementation into a facility for an extended duration (e.g. for several years of operation), the weighting value needs to reflect the average importance over this period instead and hence for this research, the weights were assigned with the intention that they would represent the average significance of each metric to a company over an extended period of time.

4.4.6 Criteria for screening options in and out

Acceptance criteria were specified for each layer in order to determine which options would be screened in and out, with the least attractive options in each layer being eliminated and the remainder passed to the next layer for further evaluation (or selected as amongst the best emerging

from the third layer). To reflect the differing levels of refinement and accuracy of the three layers, less stringent criteria were used for the first layer than for the second, which in turn used less stringent criteria than the third. This avoided screening out options in the first and second layers which *were* favourable, but which would otherwise be eliminated due to simulation inaccuracies. For layers one and two, the criteria for new process options to pass down to the next layer were:

$$OR_{New\ option} > OR_{Current\ operation} + M_1 \tag{26}$$

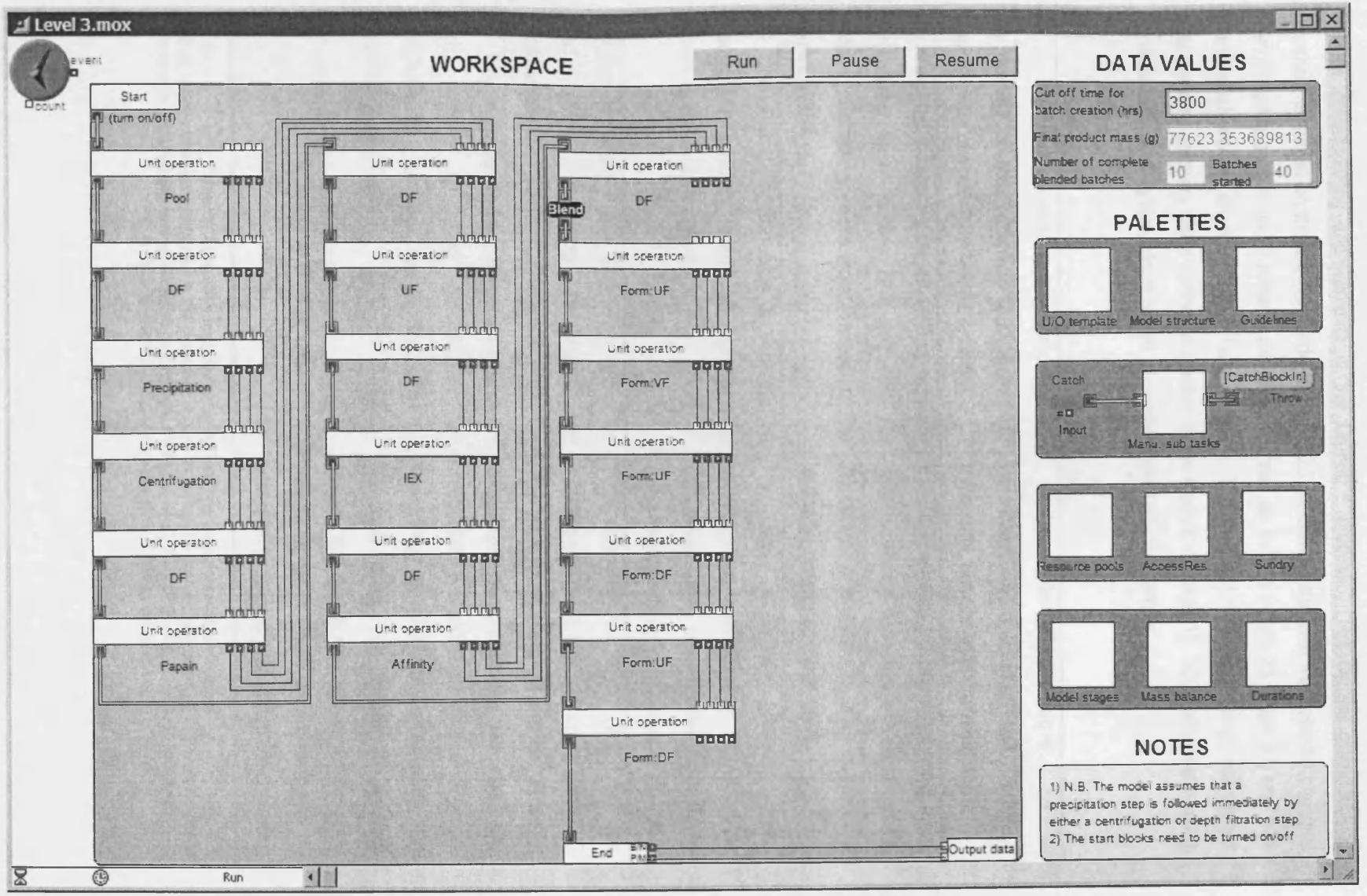
$$OR_{New\ option} > OR_{Current\ operation} + M_2 \tag{27}$$

In the third layer, the criterion by which options would be judged sufficiently superior to the current operation and so make them the best out of all under consideration was:

$$OR_{New\ option} > OR_{Current\ operation} + M_3 \tag{28}$$

M_1, M_2 and M_3 specify the margin in Overall Rank units by which an option should be greater than (and therefore superior to) the current operation in order to pass it into the next layer (or be selected as one of the best options from the third layer). Values are set such that $M_3 > M_2 > M_1$ (reflecting the greater stringency of lower layer models). For the industrial case study presented below, normalisation bounds and weightings used to calculate Overall Rank values, acceptance criteria and the values of M_{1-3} are specified in Section 4.5.9 and 4.6.2 later.

Figure 22: Sample screenshot from the third layer model



4.4.7 Industrial case study

The multi–layer modelling methodology was used to assess a range of options for their potential to achieve manufacturing improvements in the CroFab™ process. The current manufacturing flowsheet is shown on the left hand side of Figure 23 for comparison with the potential changes. Manufacturing options under consideration are indicated as A–E in Figure 23 (Page 91) and their potential advantages are presented in Table 16 below. The manufacturing methodology was used to evaluate and compare these options according to the metrics in Table 15, both individually and in combinations, leading to a total of fifteen alternative manufacturing options.

Alternative manufacturing option	Potential advantages
(A) Increase the feed volume to 1000L (maximum volume under consideration within the facility)	Increasing the volume will increase throughput, but also manufacturing times
(B) Vary the concentration of sodium sulphate precipitating agent (testing 24% and 48% w/v sodium sulphate concentrations instead of the current 36% w/v)	Experimental data indicated that using these sodium sulphate concentrations could make a significant difference to the quantity of antibody precipitated (Neal, 2005)
(C) Use a microfiltration step to separate the precipitated antibody from the contaminating albumin instead of the centrifuge	For the industrial process under consideration, microfiltration is a viable alternative to centrifugation, providing comparable antibody yields (Neal et al., 2004)
(D) Use a Protein-G step to capture IgG in the feed, instead of using precipitation and centrifugation	Data collected by Neal (2005) in the evaluation of ovine IgG capture by Protein G was used to populate the model. Potential advantages of consolidating two steps into one include a reduction in the cost of goods. The use of an additional concentration step will also lower the downstream process volume and hence may further reduce costs as well as batch times, although this may be mitigated by product loss during concentration
(E) Use a Protein-G step to capture IgG in the feed, instead of using precipitation and centrifugation, followed by concentration by ultrafiltration at 120 L/hr	

Table 16: Alternative manufacturing options and their potential advantages
Letters refer to strategies indicated in the flowsheet (Figure 23). The five individual options (A→E) were modelled, along with the following combinations: A→ B [@ 24% & 48%]; A → C; A → D; A → E; A → B [@ 24% & 48%] → C; B [@ 24% & 48%] → C, leading to a total of fifteen options

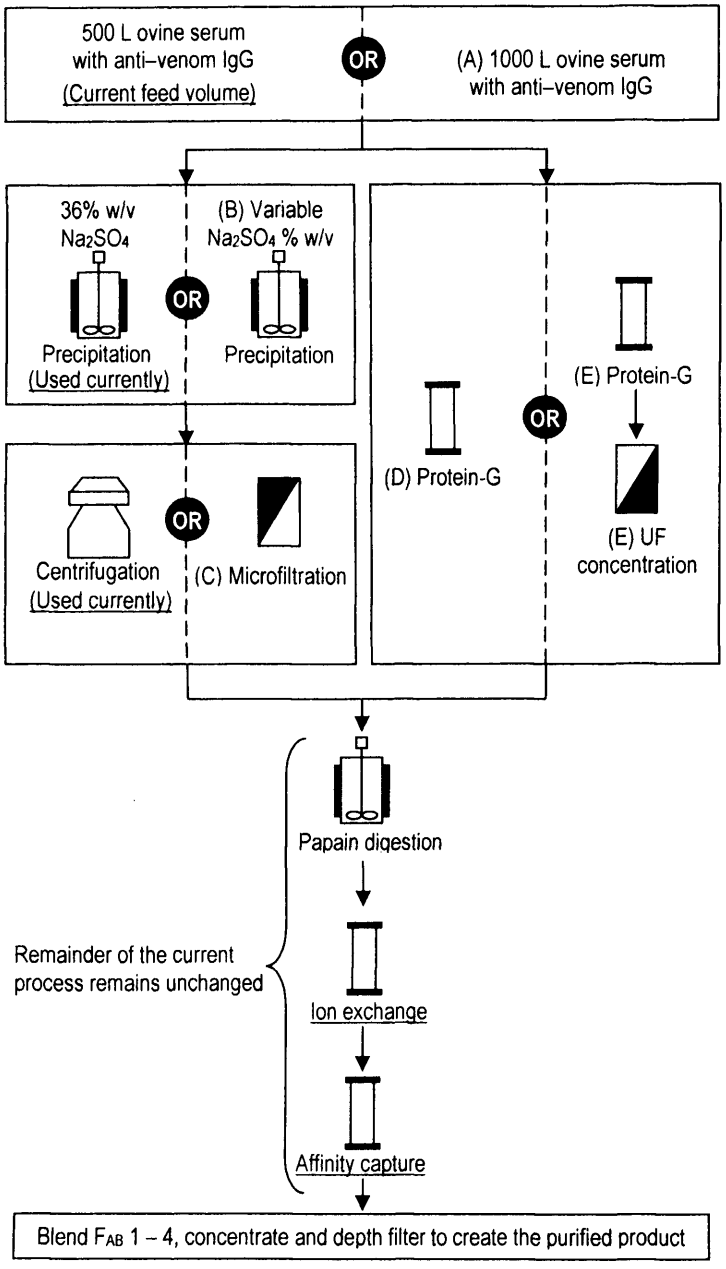


Figure 23: Process flowsheet

This shows the current process in Figure 2 for comparison with the alternatives under consideration (A to E) – see Table 16 on Page 90 for further details. The process is operated once for each of the four rattlesnake anti-venoms and then the outputs are blended, concentrated and filtered

The process changes being appraised here move the process in the opposite direction to that taken by other parts of the industry such as for the production of monoclonal antibodies, where the high expense of chromatography resins such as Protein A is instigating the study of alternatives such as precipitation or centrifugation. In CroFab™ manufacture, however, the existing precipitation and centrifugation steps are somewhat time-consuming and column steps are being considered actively by Protherics for their potential to reduce process times, whilst also theoretically decreasing manufacturing costs and reducing the number of places in the process where product can be lost by virtue of merging two steps into one (see the next chapter for the trade-off of this with the additional development burden needed to bring the change into the production facility). It is for this reason that chromatography steps were considered as the basis of the evaluation.

4.4.8 Modelling equations and data

4.4.8.1 Introduction

Material balancing equations and input values for the case study are provided below. Due to corporate restrictions, commercially sensitive details such as process costs that were used to construct the simulations have not been given in these tables. Model inputs were based upon laboratory studies and trend analysis of manufacturing data. The same feed composition data were used for all simulation runs in all three layers of the methodology.

Unit operation	First and second layers	Third layer
Precipitation	Data from Neal et al. (2003) related the sodium sulphate concentration to the amounts of IgG and albumin that were sedimented	
Centrifugation	Particle size distribution data (Neal *et al*, 2003) and physical and operational properties of the production scale centrifuge were used to determine the critical particle diameter (d_c) and recovery using a simplified Rosin–Rammler–Sperling–Bennett grade efficiency expression – see below	The same particle size distribution data was used as for the first two layers, in conjunction with the same equation for critical particle diameter and the mathematically rigorous version of the Rosin–Rammler–Sperling–Bennett expression – see below
Microfiltration (replacing centrifugation)	Particle size distribution and membrane pore size (0.2 µm) data were used to quantify recovery of the precipitated IgG particles	The same approach for material balancing as for the first and second layers was used, in conjunction with a cake deposition filtration model (Neal *et al*, 2004) to determine the duration of cake filtration of the precipitated IgG particles – see below
Papain digestion	A percentage defined the amount of IgG digested	Experimental data relating enzyme concentration, temperature and duration were used in conjunction with a GP† to create an equation relating input variables to quantity of IgG digested
All chromatography steps	Quantity bound = $C_N \times V_m \times DBC$	Quantity bound determined by simulated breakthrough curves produced by the Thomas model (Montesinos *et al*, 2005) – see below

Table 17: Principal unit operation material balance equations used at the different layers of the methodology
† GP – i.e. a Genetic Program (Goldberg, 1989) – see below for further details.

4.4.8.2 Precipitation data

The data relating precipitant concentration to IgG and albumin recovery was taken from Neal (2005):

Na_2SO_4 (% w/v)	% IgG recovered in the solid	Albumin (mg/ml)	Total protein (mg/ml)
12	0	0	0
16	0	0	0
20	0	0	0
24	40	2.5	20
28	75	3	30
32	75	4	35
36	75	5	35
40	70	6	35
44	65	7.5	35
48	60	12	35

Table 18: Data used to model recovery in precipitation as a function of sodium sulphate concentration

4.4.8.3 Centrifugation model

The simplified grade efficiency expression used for the centrifugation model in the first second layers was as follows (Svarovsky, 2000). The critical particle diameter was calculated by (see Table 20 below for definitions of the terms):

$$d_c = \sqrt{\frac{18 \cdot Q \cdot \mu}{\Delta\rho \sum g}} \tag{29}$$

The proportion of the solid material that was then recovered, G(d), was then determined by:

$$G(d) = \begin{cases} \left(\frac{d}{d_c}\right)^2 & d < d_c \\ 1 & d \geq d_c \end{cases} \tag{30}$$

Using the same expression for critical particle diameter i.e. Equation (29), the mathematically rigorous version of the grade efficiency expression required for the third layer was then used to determine recovery (Clarkson et al, 1996):

$$G(d) = 1 - e^{-\left(\frac{k \cdot d}{d_c}\right)^{\lambda}} \qquad (k = 1; \lambda = 2: \text{Values assumed from Varga et al (2001)} \tag{31}$$

The data used to populate these models are given in Table 20 and using the values, the critical particle diameter was calculated by Equation (29) to be 3.82 $\times 10^{-7}$ m i.e. 0.38 μm. The value of sigma in Table 20 (sedimentation area of the centrifuge) was evaluated as follows:

$$\sum = \frac{2\pi n\omega^2}{3g}\left(R_0^{\,3} - R_1^{\,3}\right)\cot(\theta) f_1 \tag{32}$$

Parameter	Description	Value
n	Number of disks (-)	95
N_R	Revolution speed (rps)	125
ω	Angular velocity = $2\pi N_R$	47,100
g	Gravitational acceleration (m/s^2)	9.81
R_o	Outer disk radius (m)	0.1
R_i	Inner disk radius (m)	0.045
θ	Half disk angle	55

Table 19: Centrifuge data

f_1 is a correction factor calculated by Equation (33) to account for the presence of the spacer ribs that interfere with the fluid flow (causing back mixing of sedimented material into the fluid) and reducing the sedimentation area. b_1 is the width of the caulk = 3.77 × 10^{-3} m and gives f_1 = 0.25, resulting in Σ = 8,400 m^2.

$$f_1 = 1 - \frac{3nb_1}{4\pi R_0}\frac{1-\left(\frac{R_i}{R_0}\right)^2}{1-\left(\frac{R_i}{R_0}\right)^3} \tag{33}$$

Variable		Data value
ρ_{solid} [kg/m^3]	(Solids phase density)	1300
ρ_{liquid} [kg/m^3]	(Liquid phase density)	1170
$\Delta\rho$ [kg/m^3]	(Density difference)	130
μ [NS/m^2]	(Feed viscosity)	2.4×10^{-3}
Σ [m^2]	(Sedimentation area)	8400
g [m/s^2]	(Gravitational acceleration)	9.81
Q [m^3/s]	(Volumetric flowrate)	3.61×10^{-5}

Table 20: Modelling data used for the centrifugation model
The first four values were taken from Neal (2005) and Neal et al., 2003; the sedimentation area was evaluated as described above, whilst Q was calculated from data provided by Protherics

Particle size distribution data relating particle size to volume percentage of the stream from Neal et al. (2003) and Neal (2005) was also used to populate the centrifugation model.

4.4.8.4 Papain digestion model

Cresswell et al. (2005) showed that papain digestion of ovine IgG was sensitive to the concentration of the enzyme, the temperature of the reaction and its duration. No equations could be found in the literature relating these variables to the extent of antibody digestion and in order to create such a relationship, experiments were conducted in which the impact of these parameters upon IgG cleavage was evaluated. F_{AB} concentrations in each sample were determined by using the Agilent 2100 Bioanalyzer system (Agilent Technologies Incorporated, Palo Alto, California, U.S.A.). In order to create an equation that calculates the extent of IgG digestion, these data were then encoded into a Genetic Program (GP). GPs are based on Genetic Algorithms (GA), which enable a population of trial solutions for a problem to be evolved to create individuals that more closely fit a desired solution (Holland, 1973). A GP allows the production of computer programs by application of GAs in order to investigate the solution space (Goldberg, 1989), computationally evolving a mathematical equation by random alteration, with the aim of giving it a structure that more accurately fits experimental data. Equations are exposed to multiple evolutionary rounds, with fitness judged by measures such as the root mean square difference (RMS). If the RMS fitness of the new equation is superior to the old one, the new equation is entered into a further round of evolution. Alternatively, if fitness is worse, the newly evolved equation is discarded and the old one re-used as the basis of the next evolutionary round.

For this work, an enzyme progress curve describing product concentration (and so yield) as a function of duration was entered into a GP:

$$\gamma_{Papain} = 1 + \Phi \cdot t_{enz} + \chi \cdot t_{enz}^2 + \Lambda \cdot t_{enz}^3 + \ldots \tag{34}$$

The starting values of φ, χ, and Λ were selected arbitrarily and for practical purposes, the equation was truncated to a four term partial sum. The GP was encoded in MATLAB® and evolved over 300 iterations, after which no further improvements in RMS fitness were observed. The final evolved equation was then encoded into the main manufacturing model.

4.4.8.5 Chromatography model

The Thomas model was used to simulate breakthrough curves for the chromatographic material balancing calculations in the third layer model (Montesinos et al., 2005). This model is based upon the following assumptions:

- There is no axial dispersion of the product i.e. plug flow operates in the column
- The column is initially completely free of the binding species of interest

- The process of product binding to the matrix can be represented by adsorption kinetics alone i.e. the rate of mass transfer into the model was assumed to be far greater than the rate of binding between solute molecules and the ligand

Initially, the following dimensionless terms are required (their definitions and data values are given below):

$$T = \frac{vt_c}{L} \qquad \text{(Dimensionless residence time)} \qquad (35)$$

$$N = \frac{(1-\varepsilon)\, q_m k_1 L}{\varepsilon v} \qquad \text{(Dimensionless number of transfer units)} \qquad (36)$$

$$Y = 1 + \frac{C_0}{K_d} \qquad \text{(Dimensionless separation factor)} \qquad (37)$$

$$G = \frac{\varepsilon \cdot K_d Y (T-1)}{(1-\varepsilon)\, q_m} \qquad \text{(Dimensionless mobile phase flowthrough velocity)} \qquad (38)$$

The equilibrium saturation capacity of the solid adsorbent (q_m) is determined from the equilibrium saturation capacity of the settled adsorbent (q_{ms}) by Equation (39):

$$q_m = \frac{q_{ms}}{(1-\varepsilon)\cdot(1-\varepsilon_i)} \qquad (39)$$

Breakthrough curves ($X = C/C_0$ as a function of time) can then be obtained from the following equation (valid for $T \geq 1$):

$$X = \frac{J(N/Y, NG)}{J(N/Y, NG) + [1 - J(N, NG/Y)]\cdot exp[(1-1/Y)(N-NG)]} \qquad (40)$$

$J(\alpha,\beta)$ is a function defined by Equation (41):

$$J(\alpha,\beta) = 1 - e^{-\beta}\int_0^{\alpha} e^{-\xi} I_0\left(2\sqrt{\beta\xi}\right) d\xi \qquad (41)$$

The I_0 function is defined by Equation (42):

$$I_0(x) = \sum_{k=0}^{\infty} \frac{\left(\frac{1}{4}x^2\right)^k}{(k!)^2} \quad (42)$$

The breakthrough curve was numerically integrated by the trapezium rule to determine the quantity of each component that passed into the flowthrough and hence from material balance the quantity that remained bound to the matrix. For all column steps, it was assumed that elution conditions were able to achieve complete recovery of all bound components.

Data values used for the different inputs in the Thomas model for the three columns are given below. Values which were not dependent upon the nature of each feed component (e.g. flowrate or bed voidage) were defined as follows:

		Data value		
Variable	**Definition**	**Ion exchange**	**Venom-specific affinity**	**Protein G**
v (m/s)	Interstitial velocity	3.94×10^{-4}	4.28×10^{-4}	4.39×10^{-4}
ε [-]	Bed voidage fraction	0.4	0.4	0.4
ε_i [-]	Bead porosity fraction	0.75	0.75	0.75
L [m]	Column length	0.2	0.34	0.2

Table 21: Data values used for the Thomas model

The bed voidage and bead porosity fractions were chosen as they represented typical values found in the literature for many commercially available matrices. Flowrate and bed height data were provided by Protherics. The concentration of each feed component was determined by the model. The equations used to calculate the duration of feed application to the column is given below:

Ion exchange column:

The capacity of the ion exchange column were assumed to be 120 mg/mL for all feed components i.e. F_{AB}, F_C, albumin, IgG and minor feed proteins (value taken from GE Healthcare literature number 71-5017-51 - description of HiTrap columns for the matrix used by Protherics (q_{ms} in the equation below). For a bead porosity of 0.75 and a void fraction of 0.4, the equilibrium capacity of the matrix skeleton is then determined by the following (Montesinos et al., 2005):

$$q_m = \frac{q_{ms}}{(1-\varepsilon)(1-\varepsilon_i)} \quad (43)$$

On this basis, the matrix capacity (q_m) was determined to be 800 mg/ml. This was divided the values by the molecular mass of the different feed components to give values in mol/L and converted to

mol/m^3 by multiplying by 1000 in order to be dimensionally consistent with the other data values used in the model. Molecular weights were assumed to be:

F_{AB} = 45,000 g/mol
F_C = 50,000 g/mol
Albumin = 55,000 g/mol
IgG = 150,000 g/mol

On this basis, the values for the capacity of the ion exchange matrices were calculated as follows (at commercial scale, only impurities bind to the ion exchange column and F_{AB} flows through with negligible interaction with the matrix; hence for the model, it was assumed that complete F_{AB} recovery could be achieved in the ion exchange column):

F_C = 16 mol/m^3
Albumin = 14.5 mol/m^3
Sundry proteins = 14.5 mol/m^3 (assumed the same as albumin since both are serum proteins)
Intact IgG = 5.33 mol/m^3

For the value of the equilibrium desorption constant, K_d [mol/m^3], the following logic was used to derive a value: at the post centrifugation stage, the IgG purity is approximately 90%. Given that the IgG concentration represents about 30 mg/ml, the total protein concentration is then approximately 33 mg/ml. F_C (being a third of the IgG molecule and hence with a concentration of 10 mg/ml) represents about 10/33 ~ 30% w/w of total protein and albumin is ~ 3/33 = 11% ~ 10% of the total protein. Across the filtrations and ultrafiltrations, it was assumed that none of the contaminating F_C or albumin was lost in the permeate (for the ultrafiltration step, the filter used has a cut off size that is significantly below the 50 KDa size of F_C and the 55KDa size of the albumin). Hence, the composition of the stream on a % w/w protein basis was assumed not to change. Based on data from Protherics, the ion exchange outlet typically contains F_C, albumin and sundry protein components that comprise a total of ~ 3% w/w of the total protein. This reflects the observation at commercial scale that the vast majority of the feed impurities are bound by the ion exchange matrix. Hence the concentration in the matrix phase is likely to be very high relative to the mobile phase. If the equilibrium is defined in terms of the following relationship:

Impurity + Matrix $\rightleftharpoons$ Impurity-Matrix complex

Hence the standard equilibrium (adsorption) constant can be defined as:

$$K_A = \frac{[Impurity\ (matrix\ phase)]}{[Impurity\ (soluble\ phase)]}\Bigg|_{eq} \quad (44)$$

From above, the impurities represent 30% + 10% = 40% w/w of total protein when applied to the column, whilst emerging from the column outlet, they represent only ~ 3%. Hence the ratio of bound concentration to mobile concentration in the above equilibrium expression = (40-3) = 37:3 equals approximately 12), meaning that the inverse i.e. the desorption (i.e. K_d) is very low (1/12 = 0.083 ~ 0.1). For the sake of simplicity, this value of K_d was assumed for all impurities fed to the ion exchange column. Since the primary impurity in the feed binding to the column is F_C and since this is also the part of the IgG molecule that is primarily responsible for the interaction with Protein A, it was assumed in the absence of other data for the forward rate constant k_1 [m^3/mol/s] that this value was likely to be similar between Protein A and ion exchange matrices. Values given by Hahn (2003) obtained at 100 cm/hr (closest flowrate to that used at commercial scale) for a series of Protein A adsorbents ranged between 0.035 and 0.24 m^3/mol/s, with an average of 0.129. As an approximation, this was rounded to 0.1 and entered into the model.

Venom-specific affinity column:

The capacities of thus column were determined in a similar way to that described above for both venom specific components (i.e. F_{AB} and IgG). Data from Protherics suggested that q_{ms} could be estimated to be 50 mg/ml. With a bead porosity of 0.75 and a voidage fraction of 0.4, this resulted in a q_m equal to 333 mg/ml. This was divided by the molecular weights of the two species to give a value in mol/L which was then converted to mol/m^3 using the following:

F_{AB} = 45,000 g/mol $\rightarrow$ q_m = 7.4 mol/m^3

IgG = 150,000 g/mol $\rightarrow$ q_m = 2.2 mol/m^3

There was no data available from Protherics regarding the binding of impurities to the affinity column, primarily because the matrix is considered to be highly specific to the anti-venom F_{AB} and without any significant propensity to interact with the other components. Investigations with the model also indicated that the impact of q_m of the impurities upon their recovery was minimal, because within the constraints of the other values, changing q_m did not influence predicted binding levels and hence predicted product impurity levels significantly. Thus q_m was assumed to be a significantly low value at which little to no impurity recovery occurred (giving close to 100% F_{AB} purity). Trial and iteration with the model indicated that a q_m capacity of 0.001 was suitable and hence this value was entered into the model.

The values for K_d were calculated in a similar way as for the ion exchange column: data from Protherics indicated that venom-specific F_{AB} recovery from the final affinity column is 0.56. Hence

the concentration in the mobile phase = 1 – 0.56 = 0.44, resulting in an equilibrium adsorption constant (K_A) of 0.56 / 0.44 = 1.27. Taking the inverse gives a K_d of 0.79 for the venom-specific F_{AB}.

For the impurities, K_d was determined as follows based on data from Protherics: the eluate from the ion exchange column contains approximately 3% w/w of impurities (97% total F_{AB}). Out of the total quantity of F_{AB} that is fed to the venom-specific affinity column, there is a 50:50 split between the specific and non-specific. Hence, the feed to the affinity column contains 97/2 = 48.5% as specific F_{AB}, leaving the remaining 51.5% as impurities. These impurities are eliminated to a large extent by the column and emerge as 10% of the total protein in the eluate. Hence the ratio of the bound to the mobile phase = 10: [51.5-10 = 41.5] = 10:41.5 = K_a. Taking the inverse gives K_d = 4.15. Again as a convenient approximation, this value was rounded to 4 and for simplicity, was assumed to be the same for all impurities fed to the venom-specific affinity column.

For this column, k_1 was not available in the literature and hence the value was determined by iterating on the breakthrough curve until the yield calculated by integration (see below) equalled that achieved at commercial scale (0.56). For the venom specific component, this value was equal to 0.045 and for simplicity, this value was assumed for all impurity components as well.

Protein G column:

As with the previous two columns, the capacity of the Protein G step was determined for both major feed components (IgG and albumin). The q_{ms} value was assumed to be 19 mg/ml (Neal. 2005), giving a q_m value of 130 mg/mL. Assuming a molecular weight for IgG of 150,000 Da and converting to mol/m^3 gives q_m = 0.87 mol/m^3. For the albumin, data from Protherics indicated approximately 3 g from the ovine serum could be expected to bind per litre of matrix (q_{ms}). Accounting for the bead and bed porosity, this results in q_m = 20 g/L. Assuming a molecular weight for albumin of 55,000, this gives a capacity of 0.36 mol/m^3. K_d and k_1 values were taken from Neal (2005):

Variable	IgG	Albumin
K_d [mol/m^3]	3.33×10^{-4}	9.09×10^{-2}
k_1 [m^3/mol/s]	0.5	0.055

Table 22: Values of K_d and k_1 used for the Protein G Thomas model

Column loading durations:

The durations of loading onto the column steps (t_c) were determined by:

$$t_c = \left(\frac{V_{load}}{V_{matrix}}\right)\left(\frac{H}{Q_L}\right) \qquad (45)$$

Yield calculation:

Figure 24 plots a theoretical breakthrough curve. 'X' ranges from zero to one and represents the fraction of the component concentration in the feed that flows through the column without interacting and which is therefore not recovered. The areas above and below the breakthrough curve respectively represent the mass of that component that is bound and so recovered by the matrix and the quantity that flows through the column without interacting with the matrix and which is thus lost. In order to determine recovery (yield), it was assumed that all molecules bound to the matrix were successfully recovered by elution from the column. Recovery could then be calculated directly by integration of the breakthrough curve in Figure 24 using the trapezium rule to calculate the proportion of antibody mass that is lost ($Area_{Lost}$). $Area_{Total}$ in Equation (46) represents the total amount of each component in the feed applied to each column:

$$Area_{Total} = Area_{Bound} + Area_{Lost} = t_c \times 1 = t \tag{46}$$

$$Area_{Bound} = Area_{Total} - Area_{Lost} \tag{47}$$

The proportion of antibody recovered i.e. the yield is then calculated from:

$$Yield = \frac{Area_{Bound}}{Area_{Total}} \tag{48}$$

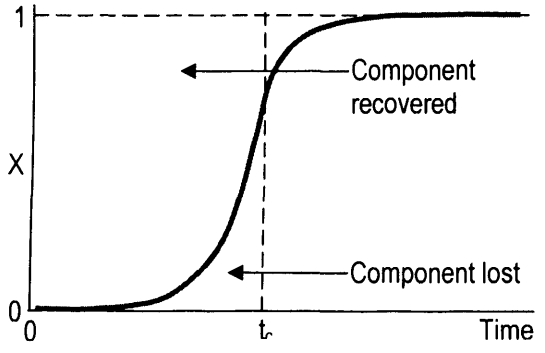


Figure 24: Theoretical breakthrough curve

4.4.8.6 Microfiltration model

A cake deposition model was employed to calculate the time needed for cake filtration of the precipitate (Neal et al., 2004) in the case that this option was used instead of the precipitation/centrifugation steps (when the filter was used elsewhere as in the current process, the same equations as given in Chapter 3 were used for time calculations). Darcy's law states that, for a volume of permeated fluid ($V_{permeate}$), the duration of filtration ($t_{permeate}$) can be determined by:

$$\frac{t_{permeate}}{V_{permeate}} = K_1\left(\frac{V_{permeate}}{2}\right) + \frac{1}{F_0} \quad (49)$$

Neal et al. (2004) fitted data at different transmembrane pressures relating $t/V_{permeate}$ to $V_{permeate}$, resulting in the following values for K_1 and F_0:

Transmembrane pressure (bar)	F_0	K_1
0.5	10	1.5×10^{-3}
1	25	1.0×10^{-3}
1.5	7.1	1.5×10^{-3}

Table 23: Fitted data to Darcy's law

Based on these data, Equation (49) was used to determine the duration of the microfiltration operation.

4.4.8.7 Other data inputs

Other key values that were used are given in the table below:

Variable	Assumed value
Current feed volume for each anti–venom IgG stream [L]	500
Number of F_{AB} types processed per blended batch [–]	4
Number of blended batches simulated	10
Initial total IgG feed titre [g/L]	25
Initial venom–specific IgG feed titre [g/L]	12
Percentage sodium sulphate used currently in precipitation [% w/v]	36
Percentage w/v sodium sulphate concentrations investigated	24 & 48
Microfiltration membrane pore size (replacing centrifugation) [µm]	3
Centrifugal flowrate [L/hr]	130
Volume of Protein-G matrix [L]	56
Protein-G IgG binding capacity [g/L]	19

Table 24: Key input assumptions

4.4.9 Modelling assumptions used for the CroFab™ process

1. The number of batches required to be manufactured was constant
2. There was no significant reduction in the capacity of a batch of chromatographic media over multiple loading cycles for the lifetime of that batch of media (assumption based on large scale data)
3. Complete recovery of all bound components was achieved in all chromatographic steps

4. Sufficient facility space exists to house any extra manufacturing equipment
5. Buffer costs comprise 5% of total CroFab™ cost of goods (Chhatre et al., 2006) and in the first layer model, this value was used to determine the annual cost of goods by multiplying the total buffer costs by 20. Buffer costs were used because they constitute a resource that every unit operation uses and hence scaling from their costs is likely to provide the most accurate final cost of goods value compared to all other resources
6. The normalisation of batch times required to computer the Overall Rank for the case study was based on the time needed to create the first blended, concentrated and filtered batch of purified product
7. Discussions with Protherics identified the most important output metrics (Table 15 on Page 84). Based on this, weightings for the Overall Ranking procedure were assigned, with manufactured product mass and cost of goods each assigned values of 0.4 and batch time assigned a weighting of 0.2

4.5 Results and Discussion

4.5.1 Simulation results

Graphs in this section (created using Microsoft® Excel XP) plot the change in Overall Rank (OR) relative to current operation, which involves processing a 500L feed with the existing flowsheet.

4.5.2 Case study acceptance criteria

The following values were industrially validated and set for the acceptance criteria:

Options assessed in the first layer which had OR values that were larger than for current operation were passed into the next layer, whilst the remainder were screened out i.e. the acceptance criterion was:

$$OR_{New\ option} > OR_{Current\ operation} \qquad \text{(Equation (26) with } M_1 = 0) \qquad (50)$$

In the second layer, more stringent acceptance criteria were set. For an option to be deemed sufficiently superior to current operation, and thus allow it to be passed into the next layer, it needed to provide improvements in the order of a 20% increase in product mass, a 15% reduction in cost of goods and a 5% reduction in batch time. Together, these equate to an Overall Rank that is at least 0.1 greater than current operation. Any option providing improvements in product mass, cost of goods and batch time which resulted in an Overall Rank that was 0.1 greater than current operation was passed into the next layer i.e. the acceptance criterion was:

$$OR_{New\ option} > OR_{Current\ operation} + 0.1 \quad \text{(Equation (27) with } M_2 = 0.1\text{)} \tag{51}$$

In the third layer, the most stringent criteria were set. For an option to be sufficiently superior to current operation and thus allow it to be chosen as one of the best options of all under consideration, it needed to provide improvements in the order of a 50% increase in product mass, a 40% reduction in cost of goods and a 5% reduction in batch time. Together, these equate to an Overall Rank that is at least 0.2 greater than current operation. Any option providing improvements in product mass, cost of goods and batch time which resulted in an Overall Rank that was 0.2 greater than current operation was selected as being amongst the best process options i.e. the acceptance criterion was:

$$OR_{New\ option} > OR_{Current\ operation} + 0.2 \quad \text{(Equation (28) with } M_3 = 0.2\text{)} \tag{52}$$

4.5.3 Results from the first layer

The current operation and the fifteen manufacturing alternatives were initially modelled in the first layer. Figure 25 plots the changes in Overall Rank for the alternatives relative to the current process. Eight options were superior to the current operation, with the three best results all involving doubling the feed volume to 1000 L. The most superior alternative is achieved when the feed volume is increased to 1000L and the Protein-G and concentration steps (option A→E) replace the existing precipitation and centrifugation combination. A breakdown of the weighted and normalised ranks for this option (black bars in Figure 26) shows this being due to an improvement in product mass and cost of goods, which significantly outweigh a slight increase in batch time. The second best level of improvement occurs when option A→D is employed, where a 1000 L feed is combined with the use of Protein-G IgG capture without the concentration step. When option A→D (white bars in Figure 26) is compared with option A→E, it can be seen that although the latter is inferior in terms of product mass owing to loss during the concentration step, this is more than outweighed by option A→E having a superior cost of goods value and also by the additional batch time taken by option A→D that results from not concentrating the process stream.

Amongst the other favourable options, using a 1000 L feed with a microfiltration step in place of the centrifugation (option A→C) is approximately as favourable as either using the Protein-G and concentration steps with a 500 L feed (option E) or simply employing a 1000 L feed with the current flowsheet (option A). The other superior options include using a 500 L feed with an altered flowsheet based on either replacing the centrifuge with the microfiltration or using a Protein-G column on its own (options C and D respectively). The final superior option using a 1000 L feed, 48% sodium sulphate and a microfiltration unit (A→B @ 48%→C) is only just better than current operation, with the advantages resulting from an improvement in product mass only just surpassing the disadvantage of additional batch time.

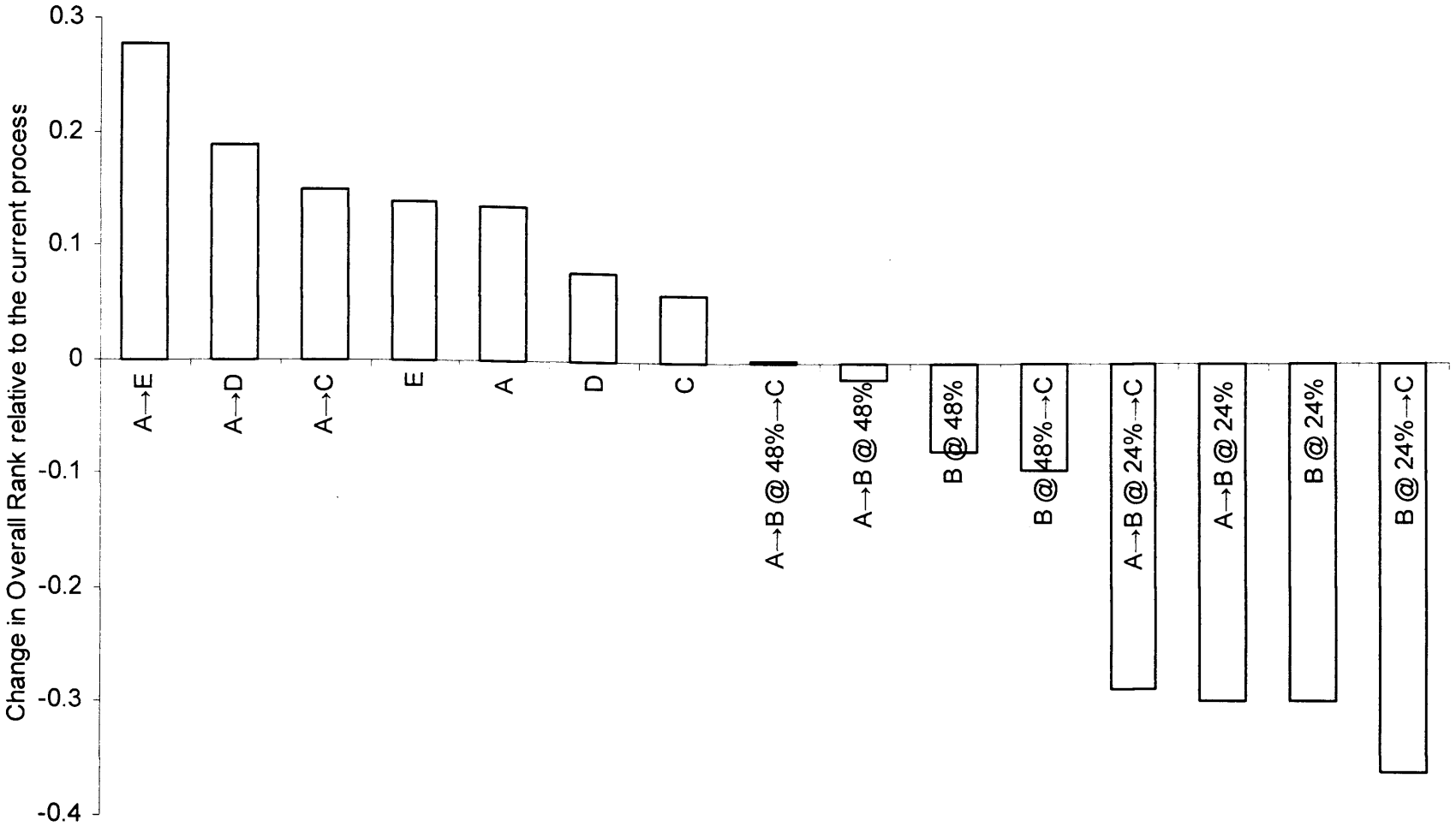


Figure 25: Results from the first layer

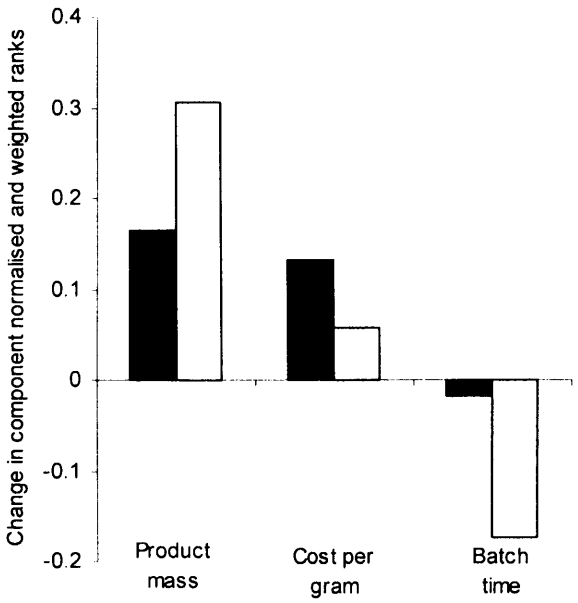


Figure 26: Breakdown of the individual component normalised and weighted ranks in the first layer for option A→E (black bars) and option A→D (white bars)

The remaining seven options, all involving a change in sodium sulphate concentration, were inferior to current operation. Use of 48% w/v sodium sulphate, either in isolation or in combination with any other process change, outweighed the use of any option involving 24% w/v sodium sulphate (Neal, 2005). Use of the latter precipitant concentration resulted in especially inferior alternatives, even when combined with an increase in feed volume and a change from the centrifuge to the microfiltration step. Option B @ 24% → C (i.e. using 24% w/v sodium sulphate and a microfiltration step) in particular provided the worst result, primarily due to a 50% reduction in product mass and a 70% increase in cost of goods compared to the current process. These inferior options were screened out in this layer, leaving the remaining eight to be evaluated by the model in the next layer.

4.5.4 Results from the second and third layers

Figure 27 plots the results from the second layer. Three options fall below the acceptance criterion and one other option is only just favourable. The more refined second layer model shows that option A→D (1000 L feed volume and a Protein-G step without the concentration step) has the best result and although the best option from the first layer (A→E – 1000 L feed with both the Protein-G and concentration steps) still meets the acceptance criterion, analysis using more accurate models has moved it to third place, with similar OR values to options A and A→C (1000 L feed without and with the microfiltration step respectively). The breakdown of options A→D and A→E for this layer is given in Figure 28 and shows that option A→D is more favourable in terms of product mass and cost of goods. Contrary to the conclusion drawn in the first layer, this now outweighs the disadvantage of the higher batch time which results from not concentrating the stream. This highlights a key point: owing to its simplifications, the first layer only returns an approximation of the feasibility of options and should only be used as a basis for screening and not final decision–making.

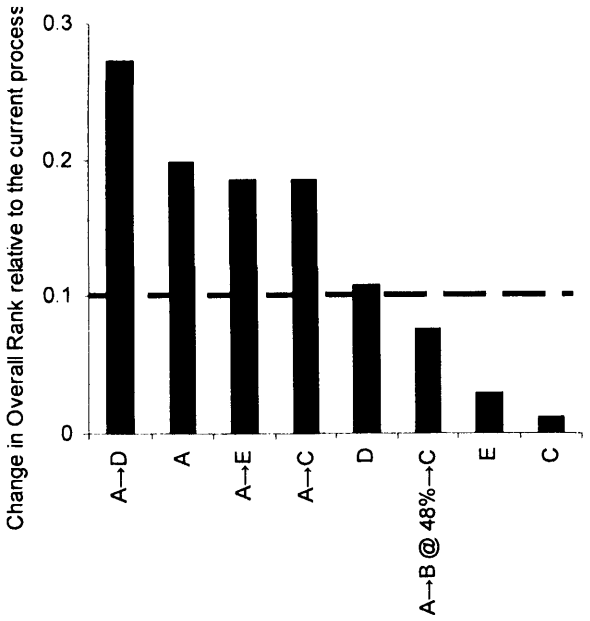


Figure 27: Results from the second layer
The dotted line represents the acceptance criterion for options in this layer

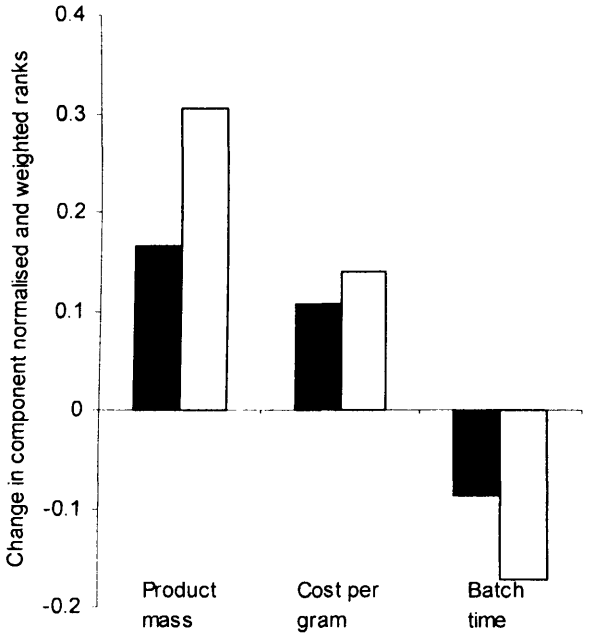


Figure 28: Breakdown of the individual component normalised and weighted ranks in the second layer for option A→E (black bars) and option A→D (white bars)

Options A→D, A, A→E, A→C and D were passed to the third layer for further analysis and the results are shown in Figure 29. Only options A→D and A→E exceed the acceptance criterion of being at least 0.2 units greater than the Overall Rank of the current process and hence out of all available options, these would be deemed the only feasible manufacturing strategies. Option A→D is significantly better than A→E and so would be the recommended process change out of all manufacturing alternatives under consideration.

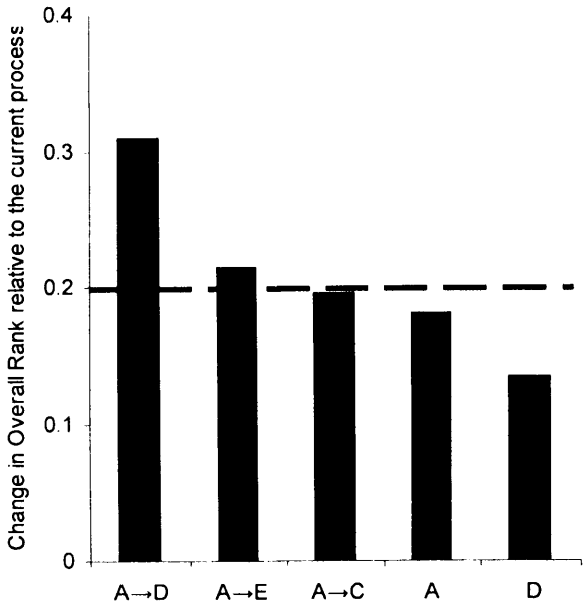


Figure 29: Results from the third layer
The dotted line represents the acceptance criterion for options in this layer

4.5.5 Verification of model consistency and compatibility between the three layers

In order to build confidence in the methodology and to verify how accurately the upper two layers reflected the manufacturing performance predicted by the third layer, every option was simulated and compared in all three layers. A summary of the results is provided in Table 25, which shows that all three layers predicted options A→D and A→E as being amongst the top three (based on their changes in Overall Rank relative to the current process). All three layers also made almost identical predictions for which options would be worse than the current process and which should therefore definitely be screened out.

	Layer		
	First	*Second*	*Third*
Best options	A→C A→D A→E	A A→D A→E	A→C A→D A→E
Worst options	A → B @ 24% A → B @ 24% → C B @ 24% B @ 24% → C B @ 48% B @ 48% → C A → B @ 48%	A → B @ 24% A → B @ 24% → C B @ 24% B @ 24% → C B @ 48% B @ 48% → C	A → B @ 24% A → B @ 24% → C B @ 24% B @ 24% → C B @ 48% B @ 48% → C

Table 25: Verification of which options were the best and worst in all three models
The best options were those in the top three out of all options tested, whilst the worst options were those which had Overall Rank values that were lower than the current process. Options are arranged in alphanumerical order for ease of comparison

4.5.6 Sensitivity analysis

To determine how robust the analysis was, the effect of varying the weighting values on the computed Overall Rank was determined. Since the weightings represent the relative importance attached to the different performance metrics, varying their values provides an indication of how alterations in priorities between the metrics affect the Overall Rank values and hence the standing of the different process options. By way of illustration, two studies were conducted to capture the impact of making hypothetical changes to the weightings. Weightings of two metrics for the most favourable options returned from the third layer analysis (A→D and A→E) were varied and the Overall Rank values were re-calculated. In each study, the weighting of the third metric remained unchanged.

a) Part of the reason why options A→D and A→E are superior to current operation stems from improvements in the cost of goods value, which help to outweigh the disadvantage of additional batch time caused by the need to process a larger feed volume. The situation is also helped by batch time having a lower weighting than cost of goods (0.2 and 0.4 respectively). This study therefore examines what would happen if the importance of batch time relative to the cost of goods was changed – Figure 30

b) Product mass and cost of goods are equally important in the calculation of Overall Rank (weightings of 0.4 each). This study examines what would happen if the importance of product mass relative to cost of goods was changed – Figure 31

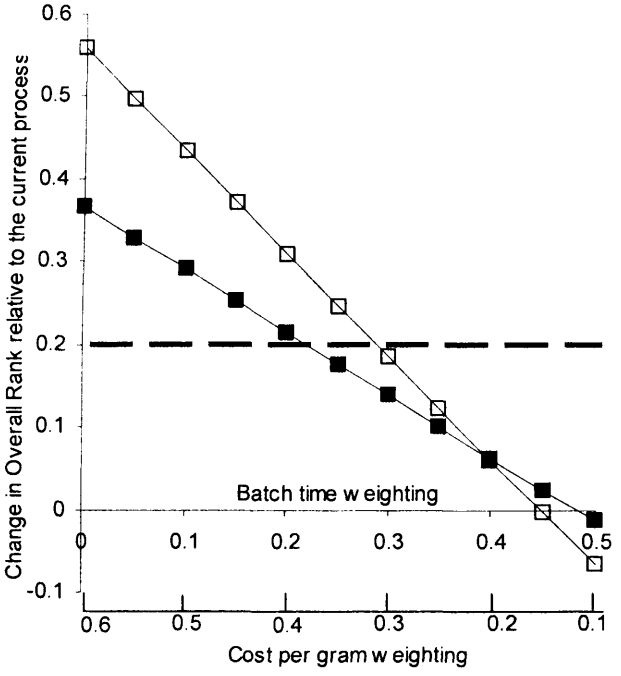


Figure 30: Impact of varying the batch time weighting on Overall Rank
(White squares = option A→D; black squares = option A→E). The cost per gram weighting was adjusted to suit as described in the text. The black dotted line represents the acceptance criterion in the third layer. The base case weightings for batch time and cost of goods are 0.2 and 0.4 respectively. The product mass weighting remained unchanged for this study

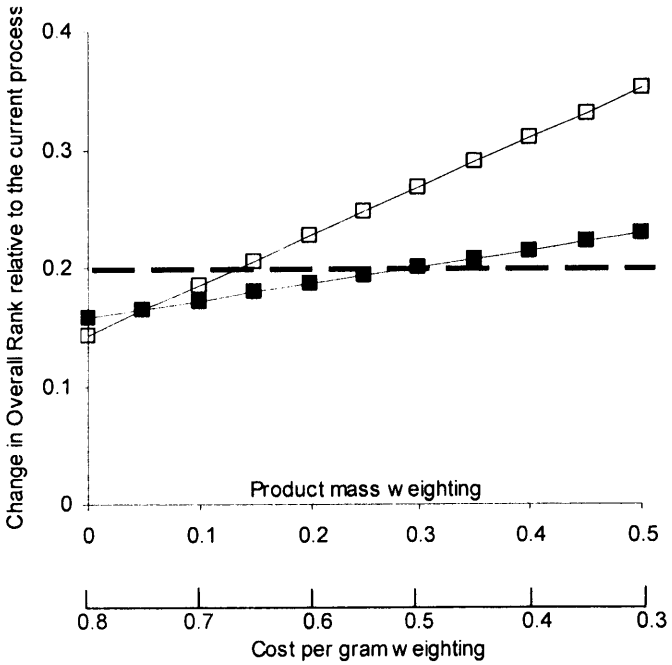


Figure 31: Impact of varying the product mass weighting on Overall Rank
(White square = option A→D; black square = option A→E). The cost per gram weighting was adjusted to suit as described in the text. The dotted line represents the acceptance criterion in the third layer. The base case weightings for both metrics were 0.4. The batch time weighting remained unchanged for this study

In study (a), increasing the batch time weighting from 0 to 0.5 and adjusting cost of goods to suit, leads to a reduction in the Overall Rank relative to the current process for both options. These fail to meet the third layer acceptance criterion once the batch time weighting exceeds approximately 0.2 for option A→E and approximately 0.3 for option A→D. This demonstrates that in order for these options to be favourable compared to the current process, the relative importance of batch time compared to the cost of goods needs to be maintained below these levels.

Study (b) shows that as the product mass weighting increases, option A→D becomes more favourable than the current process for weightings greater than approximately 0.15, whilst option A→E only becomes favourable at weightings greater than approximately 0.3. This is because option A→D results in a larger final product mass than option A→E, as product is lost in the latter case during the concentration step and consequently, as product mass becomes more important, option A→D becomes superior more rapidly than option A→E.

Additionally, it is also noticeable in both Figure 30 and Figure 31 show that over the majority of the weighting range examined, option A→D is superior to option A→E and mostly still exceeds the 0.2 acceptance criterion. This indicates that for a wide range of corporate priorities, option A→D would still be the more favourable process change to implement.

4.5.7 Conclusion

The above sections have described a deterministic analysis for identifying the most favourable manufacturing strategy to adopt for commercial operations. This approach does not account for the potential uncertainties that occur within a typical bioprocess, such as with respect to the level of market demand or the titre of product at the start of purification. Incorporation of uncertainty forms the subject of the next section.

4.6 Decision support modelling under uncertainty

4.6.1 Introduction

A follow-up study to the work described above was carried out by application of Monte Carlo modelling techniques and was applied to the third layer model of the CroFab™ process. Monte-Carlo analysis permits the incorporation of processing uncertainties into models to calculate the most likely values for outputs and works by assigning probability distributions to the uncertain input variables, running the models repeatedly and averaging the outputs to calculate the most likely values (Farid et al., 2005). This technique was used to examine two specific problems in relation to the CroFab™ process:

1. The current formulation consists of only four types of anti-venom Fabs and therefore only serves a limited percentage of the patient population. Adding extra rattlesnake Fabs to the CroFab™ formulation would increase the pharmacological relevance of the product and potentially open it up to a wider market and so more lucrative financial returns
2. Future vial requirements are likely to outstrip the current capacity to supply and hence strategies for satisfying market demands need to be identified and evaluated

For CroFab™, the key variable of uncertainty lies in the IgG titre and hence a range and a probability distribution (19-41 mg/mL normally distributed – Newcombe et al., 2006) were assigned to this input before entering it into the third layer model of Protherics' process. In relation to the two problems raised above, two specific scenarios were analysed using Monte Carlo simulation:

1. Firstly, if extra F_{AB} components, how does this affect key outputs such as COG, batch times and levels of resource utilisation? The study examined the effects of adding up to two extra Fabs
2. Secondly, what is the best operating strategy to adopt with respect to numbers of batches and volumes of feed per batch if demand for the product increases?

4.6.2 Study 1: Adding extra F_{AB} components

Turning to the first of the two studies, Figure 32 illustrates the effect of adding components upon the cost of goods for four, five and six component formulations. Adding one extra component results in a 17% reduction in the cost of goods between modal results, whilst adding another component (i.e. a total of six) results in a reduction by 26% relative to a four component formulation).

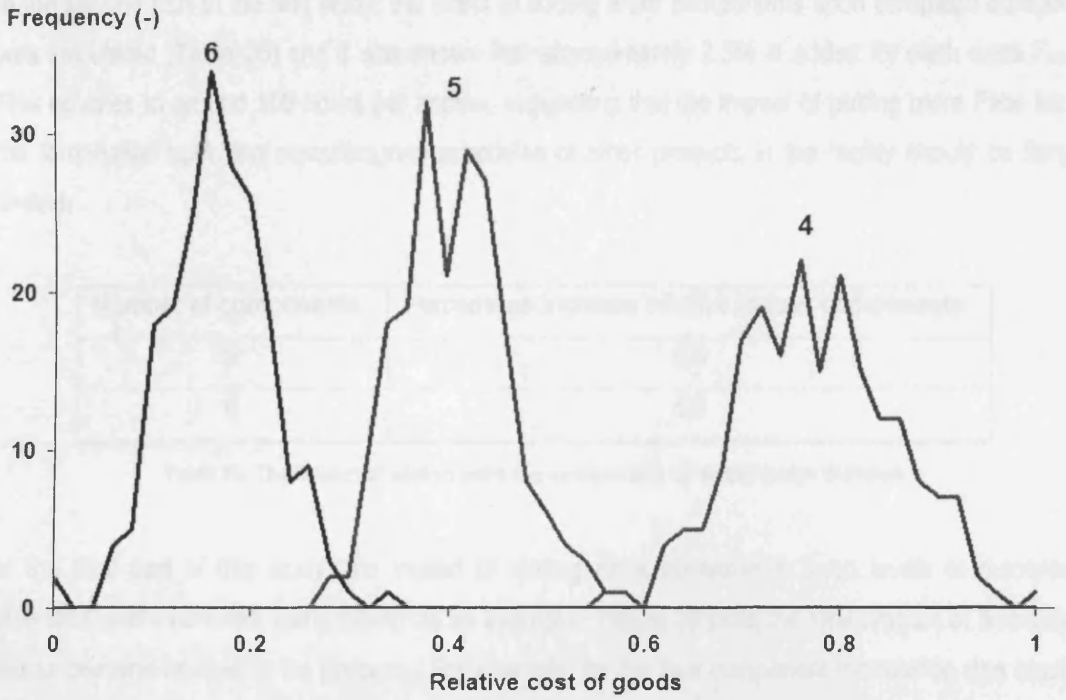


Figure 32: Probability distribution of cost of goods values resulting from adding up to two extra F_{AB} components
(Numbers indicate total number of F_{AB} species – four is the number of components used at present)

It is also noticeable that as the number of components increases, the data becomes more tightly distributed – something that is shown more clearly by Figure 32. This demonstrates that the standard deviation reduces markedly from four to five components and to a far smaller extent from five to six. If standard deviation is taken to be a measure of risk, it then follows that increasing the number of components increases the chance of obtaining the mean COG and so reduces the economic variability of the process.

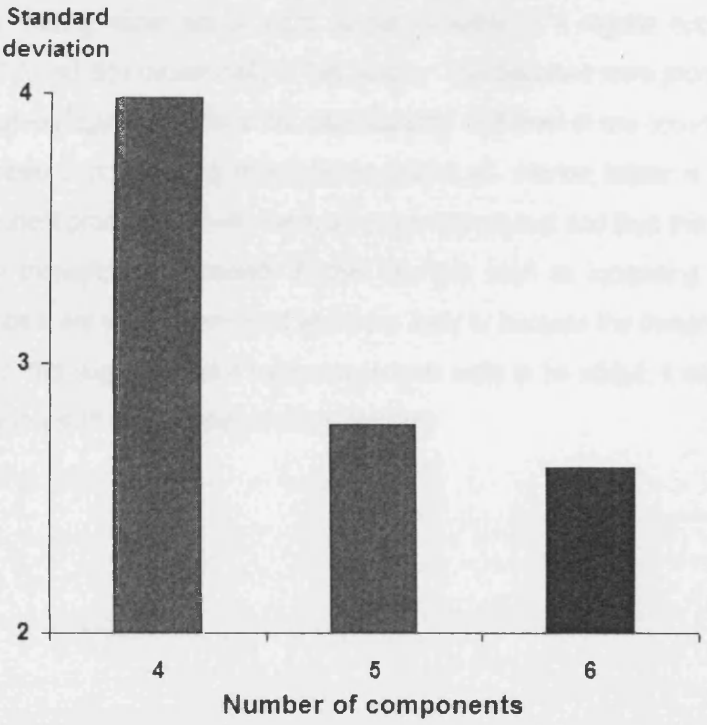


Figure 33: Standard deviation of the cost of goods distributions given in Figure 32

In the second part of the first study, the effect of adding more components upon campaign duration was calculated (Table 26) and it was shown that approximately 2.5% is added for each extra F_{AB}. This equates to around 100 hours per annum, suggesting that the impact of putting more Fabs into the formulation upon the manufacturing schedules of other products in the facility should be fairly limited.

Number of components	Percentage increase relative to four components
5	2.5
6	5.0

Table 26: The impact of adding extra F_{AB} components upon campaign duration

In the final part of this study, the impact of adding extra components upon levels of resource utilisation was examined, using labour as an example. Figure 34 plots the total amount of time that labour remains unused in the process. For example, for the four component formulation (the black bars), out of the total time needed for manufacturing all ten product batches, a total of 196 hours are spent with four manufacturing staff members remaining unused; similarly, when processing a six component formulation (grey shaded bars), a total of 166 hours are spent with ten staff remaining unused. For clarity, the absolute minimum numbers of manufacturing staff remaining available over the entire time-course of ten successive batches for the four, five and six component formulations are shown again in Figure 35. For the four component formulation, for the majority of the time, either six or eight people remain free in the labour resource pool, suggesting that these staff members could be redistributed to other tasks within the facility. At a minimum, there are always four people who are unused (Figure 35). Adding one extra component causes the labour to become used more frequently; again, having either six or eight people available is a regular occurrence, but the minimum pool remaining dips occasionally to two people. This becomes more pronounced when six components are used (Figure 35), where the manufacturing staff level of two occurs more frequently and on a few occasions, no members remain in the pool at all. Hence, labour is more of an issue with the six component product than with the four component product and thus this resource is more likely to constrain throughput – especially if other changes such as increasing feed volumes or annual batch numbers are also implemented which are likely to increase the demands on staff within the facility. Hence, this suggests that if more components were to be added, it might be prudent to consider training extra staff to cover this potential problem.

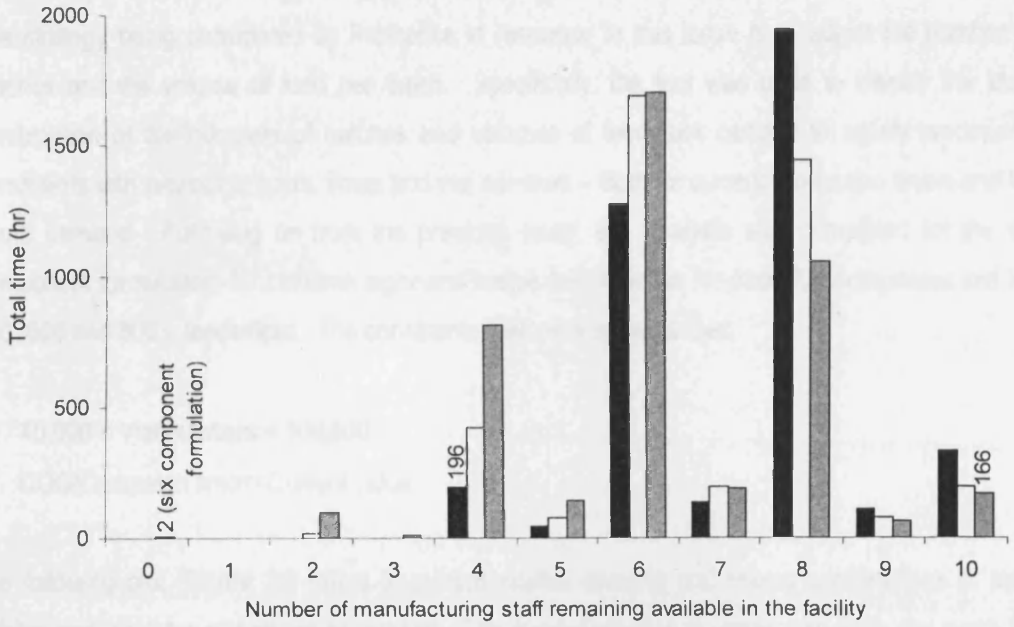


Figure 34: Facility labour usage when operating with four (black bar), five (white bar) and six (grey bar) component formulation processes
The bars plot the total amount of time for which each process has the specified number of manufacturing staff remaining available in the facility. Numbers above some of the bars are referred to in the text

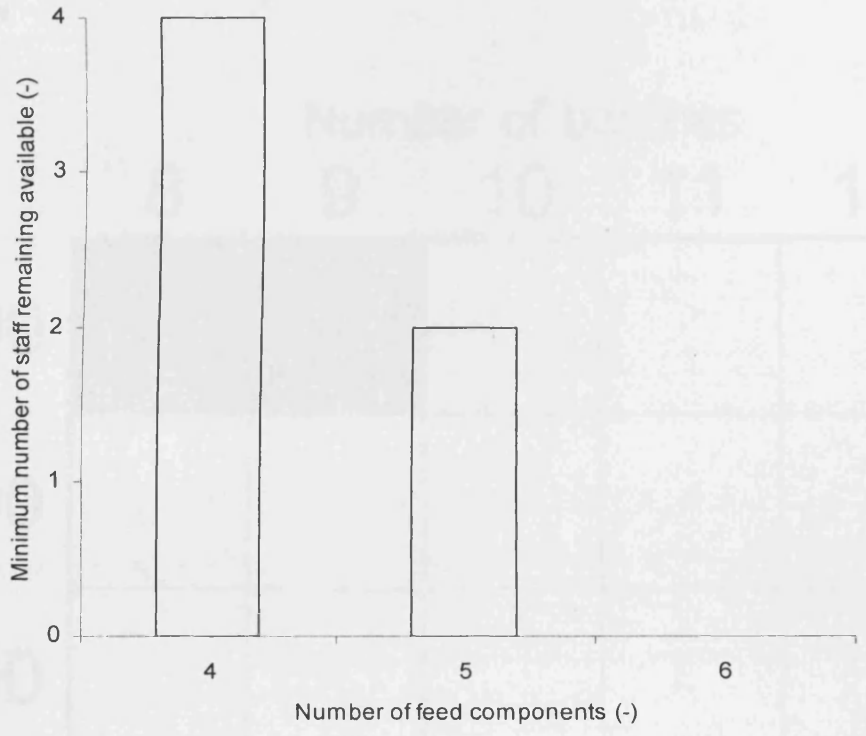


Figure 35: Minimum numbers of manufacturing staff remaining available over the entire time course of the simulation for four, five and six component formulations
Zero members remain available in the pool for a total of 2 hours, as indicated on the left hand side of Figure 34

Hence overall, the impact of adding more components is favourable, resulting in:

- A better COG
- A minimal impact on processing times
- The potential need to add resources e.g. labour

4.6.3 Study 2: Response to growing product demand

The strategy being considered by Protherics in response to this issue is to adjust the number of batches and the volume of feed per batch. Specifically, the tool was used to identify the ideal combination of the numbers of batches and volumes of feedstock needed to satisfy processing constraints with respect to costs, times and vial numbers – both for current production levels and for future demand. Following on from the previous study, this analysis was completed for the six component formulation, for between eight and twelve feed batches for each F_{AB} component and for 400, 600 and 800 L feedstocks. The constraints that were set were that:

- 40,000 < Vial numbers < 100,000
- COG/Campaign time<=Current value

The following plot (Figure 36) refers to current market demand and shows combinations of feed volume and numbers of batches processed. The current process operates with 400L per batch for 10 batches. The darkened areas represent those combinations of the two inputs that fail due to cost of goods. Each square reflects the mean value produced from repeatedly running the simulation at each combination of batch numbers and feed volumes until the rolling average of the outputs levelled off.

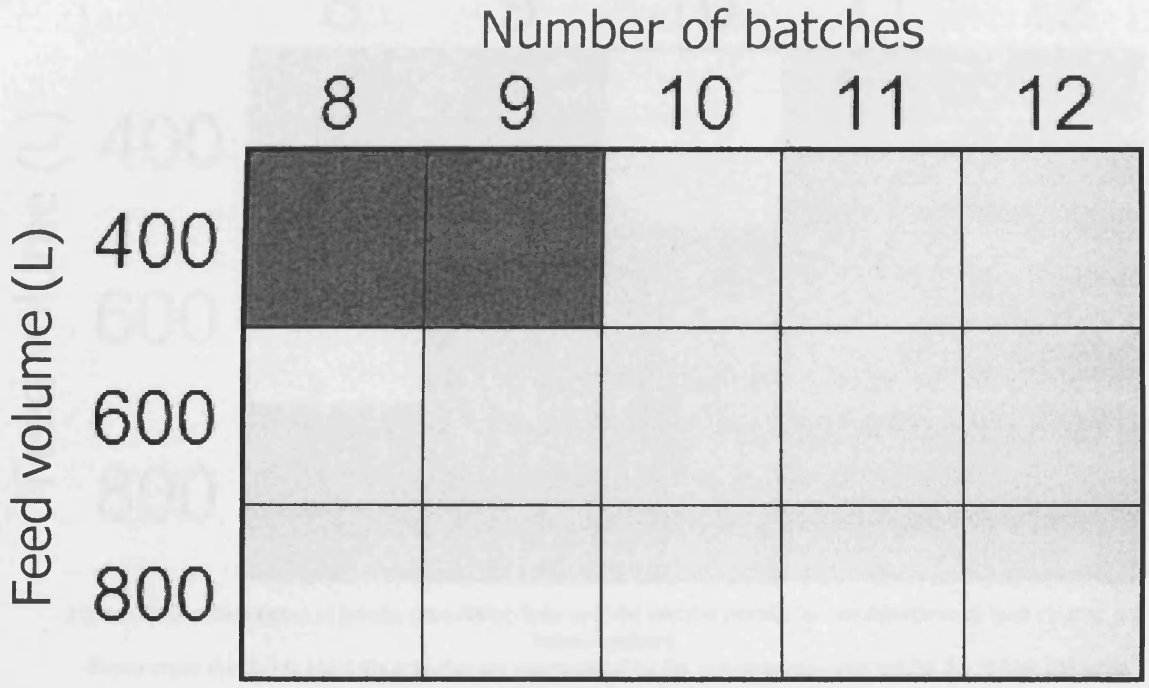


Figure 36: Cost of goods results for combinations of feed volume and batch numbers
Those areas that fail to meet the criterion are represented by the darkened squares, whilst the lighter coloured areas meet the criterion

If those areas that are unsatisfactory on the grounds of processing time are added in, Figure 37 results and if vial number constraints are then added in, the final plot in Figure 38 is achieved:

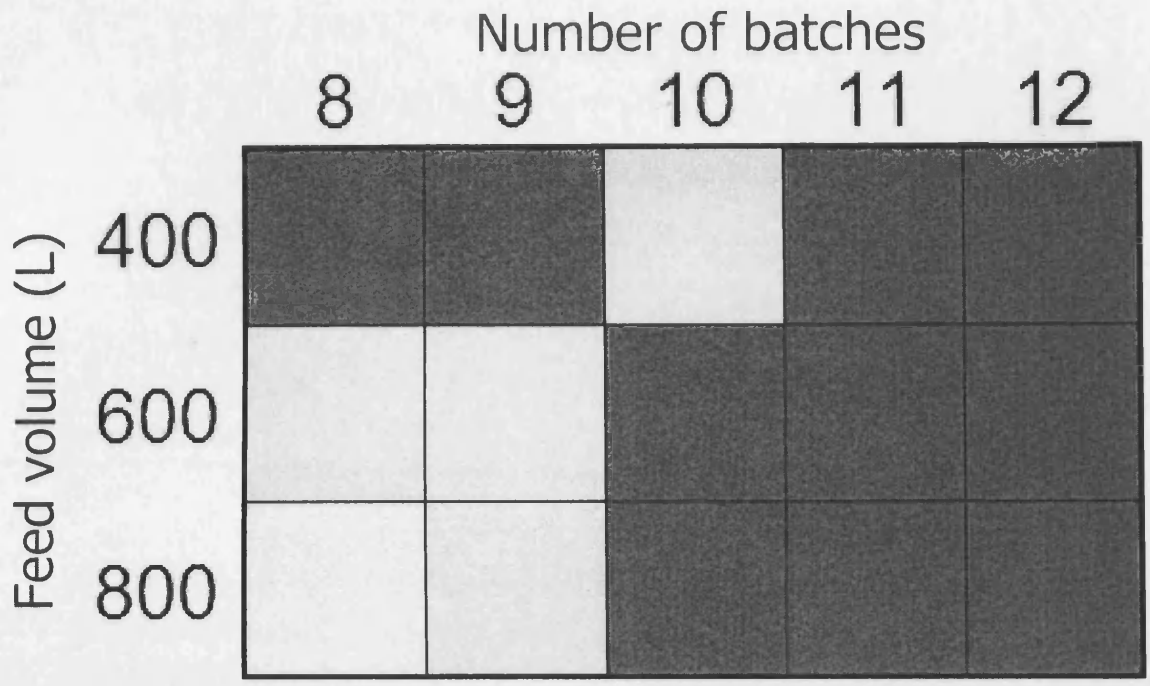


Figure 37: Combined cost of goods and processing time results for combinations of feed volume and batch numbers
Those areas that fail to meet the criterion are represented by the darkened squares, whilst the lighter coloured areas meet the criterion

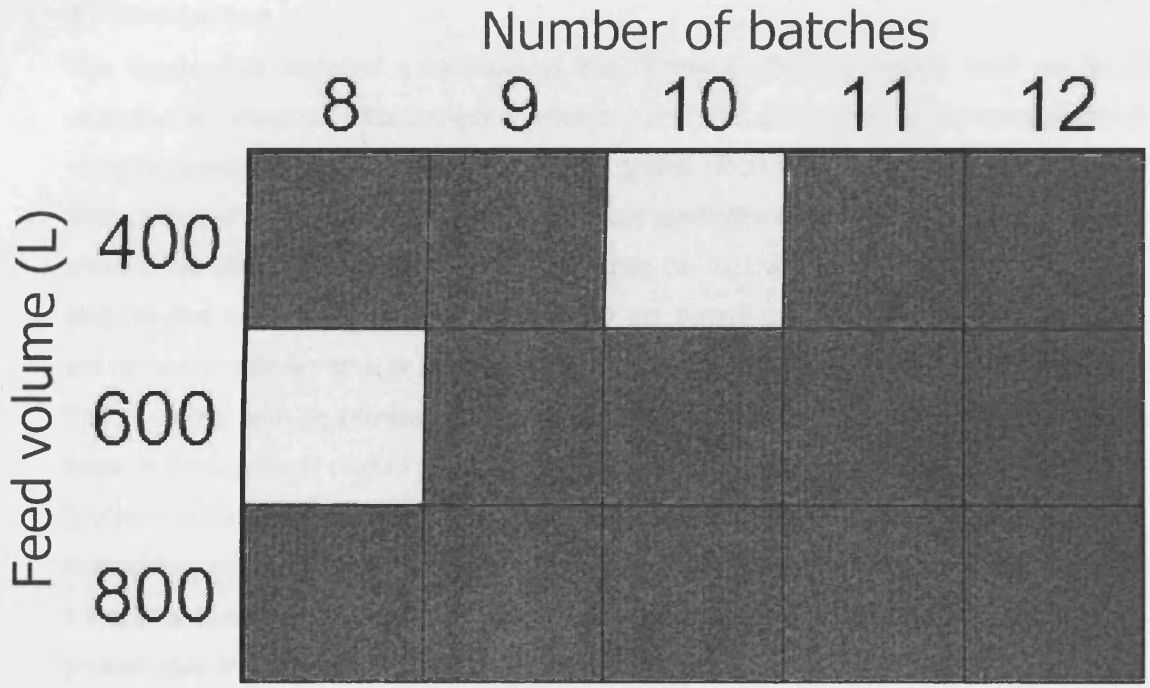


Figure 38: Combined cost of goods, processing time and vial number results for combinations of feed volume and batch numbers
Those areas that fail to meet the criterion are represented by the darkened squares, whilst the lighter coloured areas meet the criterion

Operating at 600 L per batch for 8 batches meets the acceptance criteria and in fact, is superior to the current position. Hence, for current product demand requirements, this is the best choice to make in terms of all three output metrics. Conversely, if the market demand increases by the 50% that it is anticipated to rise by, then the plot changes to Figure 39, showing that to meet the constraints, the necessary conditions are 800L per batch for 8 batches:

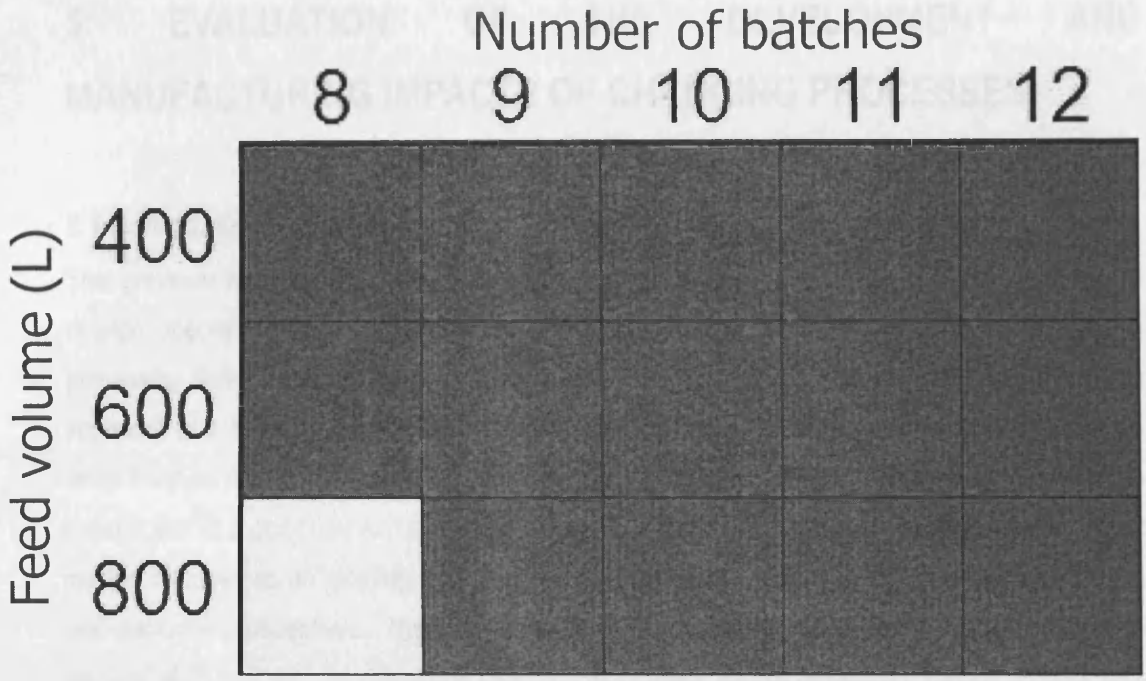


Figure 39: Combined cost of goods, processing time and vial number results for combinations of feed volume and batch numbers for a 50% higher market demand
Those areas that fail to meet the criteria are represented by the dark red squares

4.7 Conclusions

This chapter has described a methodology that minimises the computational effort needed to evaluate the commercial attractiveness of process options and also focuses experimental studies in order to provide model input data only for those options which show the greatest potential. The utility of the methodology has been illustrated through application to the CroFab™ process. Results showed that after assessment of all options, doubling the feed volume and employing a Protein-G step resulted in the most superior alternative to the current process and hence would be the recommended process change strategy. At the time of writing, Protherics is developing a new CroFab process with the intention of replacing the precipitation and centrifugation steps with column steps for direct antibody product capture from ovine serum. Hence the positive result for the Protein-G column option fits in well with what Protherics anticipates would be feasible. More generally, the methodology can be used to evaluate manufacturing options for any biotherapeutic and so identify the optimal flowsheet that meets a company's strategic priorities with respect to production levels, product costs and process times. As an extension to the study, Monte Carlo analysis was applied to the third layer model of the CroFab™ process, which showed that adding extra components to the product formulation would generally be favourable and which also identified the most suitable feed volumes and batch numbers for current and future levels of product demand.

Whereas the research presented in the previous two chapters assessed the impacts of changing processes in terms of manufacturing metrics alone, additionally evaluating the costs and times involved in developing and implementing changes would increase the utility of the techniques. This topic forms the subject of the next chapter.

5: EVALUATION OF THE DEVELOPMENT AND MANUFACTURING IMPACTS OF CHANGING PROCESSES

5.1 Introduction

The previous research chapters concentrated upon creating simulations that evaluate process change options in terms of manufacturing metrics such as product yields, cost of goods and processing times alone. Although the data generated by applying these models to a process is important to a company, greater value would be gained by additionally considering the costs and times involved in developing and implementing those changes. Hence, this chapter describes the construction of a computer simulation that evaluates the potential financial and technical impacts of making changes to an existing industrial bio-manufacturing process from both development and manufacturing perspectives. The methods presented in this chapter are illustrated by application to the CroFab™ process. Modifications designed to increase annual production, reduce process time and decrease production cost are assessed, including increasing the feed volume and titre, replacing initial downstream processing stages by an affinity column step operating in packed or expanded bed modes and removing the ion exchange step. These options are evaluated in terms of manufactured product mass, cost of goods and manufacturing batch times as well as the costs and timescales of development. The work in this chapter has been published (Chhatre et al., 2007a).

5.2 Assessment of developmental and manufacturing metrics

As with the previous chapter, the assessment of process change is complicated by the potential for certain outputs to improve whilst others deteriorate. To overcome this challenge, the multi-attribute decision-making technique employed in the previous chapter was used again to integrate several metrics into a single value in order to simplify their assessment.

5.3 Industrially relevant development strategies

5.3.1 Introduction

The following section outlines a series of proposed development strategies, each of which could, in isolation or in combination, provide a viable option for manufacturing improvement. These form the basis of the industrial case study, which establishes the ability of the methods proposed in this chapter to evaluate, quantify and rank process alternatives. Specific details of how these strategies were applied to CroFab™ are given in Section 5.6.

5.3.2 Increasing feed titres and batch volumes

A common strategy for improving annual production is to increase its initial feed titre. Methods of achieving this include metabolic engineering of an organism used in fermentation (Li and Townsend, 2006) or, for polyclonal processes, optimisation of the immunisation protocol and the selective

breeding of animals. Increasing the initial feed volume per batch will also generate larger quantities of final product, but if no extra equipment or vessels are available to hold the material, this change may generate bottlenecks in production and hence come at the expense of extended processing times per batch. Centrifuges and chromatography columns, for example, may have to operate for greater numbers of cycles in order to process all the material fed to those steps.

5.3.3 Replacing multiple steps with a single operation

Consolidating a series of steps into a single stage can reduce process costs and times and improve product yield. An option that attracts industrial interest is the replacement of potentially time consuming operations such as precipitation or centrifugation by an affinity step that selectively binds product molecules (Lowe et al., 2001), operating in either packed or expanded bed modes (Chase, 1998; González et al., 2003).

5.3.4 Removing a downstream operation

Removal of a step can increase yields and reduce costs and times as a consequence of the shorter process train, but may place a greater purification burden on subsequent stages. Hence elimination of a unit operation can necessitate alterations and re-validation of steps further downstream.

5.4 Construction of the development model

The development model used to evaluate the options in Section 5.3 was set-up in a spreadsheet. For strategies involving capital expenditure, correlations were used to relate equipment capacity to purchase costs (Harrison et al., 2003) and then updated to current day prices by assuming a 3% annual inflation. Bioprocess Lang factors (Novais et al., 2001) were used to calculate other costs e.g. validation (Table 27 – Page 120). Where a unit operation was removed, the cost of revalidating the remainder of the process was included in the model. For feed titre improvement strategies, the costs accounted for included those of purchasing assay systems to determine feed titres. For the purposes of the industrial case study, all developmental costs were treated as exceptional expenditures in year one. The durations of the different development strategies that were modelled were specific to the case study and are discussed later. The manufacturing models were constructed in Extend™ in the same way as described previously and used to track product mass, cost of goods and processing durations of the various manufacturing flowsheets that were evaluated.

5.5 Overall Ranking

As with the previous chapter, to enable comparisons between the available process options, a weighted sum additive weighting multi–attribute decision–making technique was used within each layer to combine manufacturing outputs into a single value. The technique initially involves normalising the simulation results for each output performance metric produced by the simulations in the methodology to a zero to one scale, where the zero and one bounds represent the least and most attractive values out of all manufacturing simulation runs respectively.

	Item	Lang factor (f_i)
1	Capital investment (e.g. for columns or filter housings = λ)	1
2	Qualification and validation	1.06
3	Installation	0.9
4	Instrumentation	0.6
5	Process control	0.37
6	Electrical supply	0.24
7	Detail engineering	0.77
8	Contingency factor	1.15
	Calculated cost	$\lambda \cdot f_8 \cdot \sum_{i=1}^{7} f_i$

Table 27: Lang factors used to calculate development costs (adapted from Novais et al., 2001)

5.6 Process change strategies for CroFab™

5.6.1 Introduction

The following sections provide more details about how the development strategies outlined in Section 5.3 were interpreted and thus modelled for the CroFab™ process.

5.6.2 Increase IgG feed titres and batch volumes to the process

Given sufficient time and funding, it is feasible that the venom-specific IgG feed titre could be increased by up to 80% by selective breeding of sheep and optimisation of the protocol used to immunise flocks, with respect to the quantity of antigen, the adjuvant and the frequency of immunisation. It was assumed that serum titres would increase linearly with development time (see Table 29 on Page 123). Although techniques such as ELISA can be used to determine feed IgG concentrations, their uses are limited owing to the long times required to analyse large numbers of samples (Newcombe et al., 2006). Alternative techniques such as biosensor assays (Biacore International Aktiebolag P.L.C., Uppsala, Sweden) permit rapid evaluation of specific antibody titres for many serum samples in order to screen and identify sheep which are high responders. The resulting improvements in final product mass are self evident, but this option requires significant investment in purchasing the assay units. In conjunction, a rise in feed volume up to the maximum of 1000 L that can be accommodated within the existing facility was also considered.

5.6.3 Replacement of precipitation and centrifugation by a single column capture step

Chromatographic capture of feed IgG operating either in packed or expanded modes has been proposed using a synthetic protein A resin (MAbsorbent® A2P – Prometic BioSciences Limited, Cambridge, U.K.), followed by ultrafiltration to concentrate and diafilter the eluate. Although commercial-scale use of expanded bed IgG capture is not widespread, the advantages it provides in

potentially purifying a crude feedstock without any prior clarification and in achieving concentration and clarification of product in a single step means that it is an attractive manufacturing option. It has been demonstrated that comparable IgG purities and recoveries can be achieved between the MAbsorbent® A2P and the sodium sulphate precipitation step used at present (Newcombe et al., 2005). A synthetic ligand was chosen to purify the IgG instead of the commonly used recombinant Protein A (Tejeda-Mansir et al., 1997), because the latter binds sheep antibodies only weakly (Huse et al., 2002) and because of the higher capacity and lower cost of the synthetic material, as well as its ability to tolerate the harsher cleaning conditions needed when subjected to crude feedstocks.

5.6.4 Removal of the ion exchange step

Eliminating the ion exchange step from the existing flowsheet (Figure 2) could potentially reduce manufacturing times and expenditure, but requires revalidation of the process downstream of the ion exchange step. Loading digested IgG directly onto the venom-specific affinity column may potentially result in non-specific adsorption to the affinity matrix and it was assumed that an extra wash step prior to elution from the affinity column would be needed to eliminate this and so ensure that the eluate met specification. This assumption was based upon the operating protocol for a similar process operated by Protherics for the purification of a different polyclonal F_{AB} product called DigiFab™ (Thillaivinayagalingam et al., 2007), in which digested IgG is loaded directly onto a F_{AB} specific affinity column without prior purification in an ion exchange step. When operating this affinity column, a wash step is used to eliminate non-specifically adsorbed material such as F_C.

5.7 Modelling details of the case study

5.7.1 Introduction

As outlined in earlier chapters, for convenience, input model data for calculating costs, times and material balances were entered into a Visual Basic for Applications user interface (within Microsoft® Excel XP) and connected to Extend™ (Table 28).

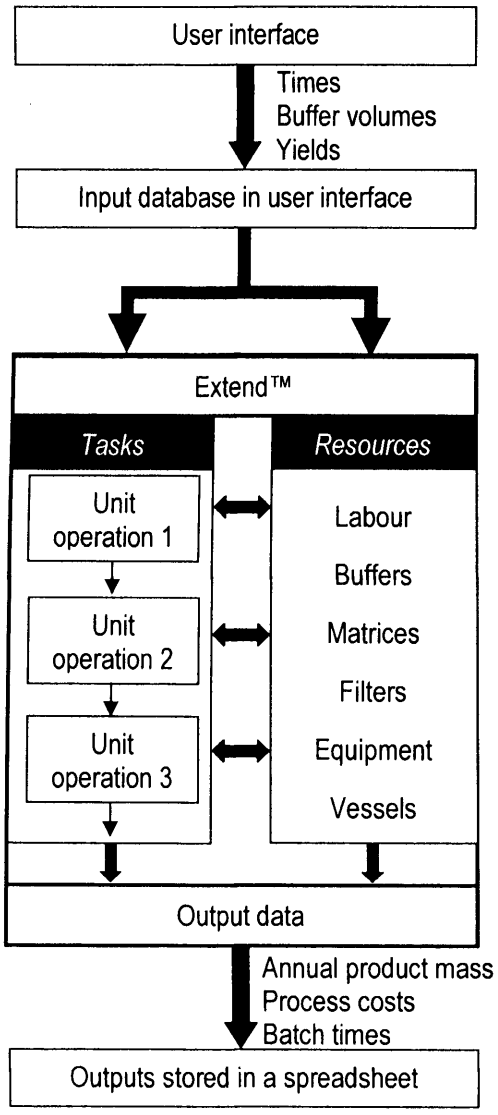
5.7.2 Manufacturing model

Figure 40 outlines the structure of the manufacturing model. Data used to calculate process costs, durations and mass balances were provided by Protherics (Table 28). Venom-specific F_{AB} mass balances were calculated in Extend™ using data provided by Protherics which specified yields achieved by groups of unit operations (see section 3.2.8). The manufacturing model accumulates costs for purchasing new matrices or membranes when previous stocks of those resources are exhausted. The model also automatically accounts for extra materials such as larger buffer volumes and the resultant higher costs incurred when processing larger volumes or higher titre feeds. For the titre improvement option, operational costs for the Biacore system were negligible (~0.1%) relative to purchasing costs and therefore were not included in the analysis.

Variable	Assumed value
Assumed current feed volume for each feed F_{AB} stream [L]	600
Number of F_{AB} types processed [-]	4
Assumed number of blended batches manufactured per annum [-]	12
Assumed current initial total IgG titre [g/L]	30
Assumed current initial venom-specific IgG titre [g/L]	8
Overall process duration for the first blended batch in current operation [hrs]	623
Volume of MAbsorbent® A2P matrix [packed and expanded columns – L]	56
Current overall yield per blended batch [%]	26%
Percentage IgG recovery using MAbsorbent® A2P [%]	95%

Table 28: Sample modelling data

Figure 40: Structure of the manufacturing model



5.7.3 Development model

The periods of time required to develop the different options and bring them into the manufacturing facility were estimated from discussions with Protherics (Table 29). Development costs were calculated as detailed previously. Where combinations of options were modelled, individual development costs were added together.

Process change strategy	Estimated duration (months)	Tasks involved
Titre improvement	Up to 19	One month to develop the Biacore assay; six months to write a validation protocol, undertake all the experiments and sign off the assay; two months to undertake trials on small sheep flocks and then scale up the immunisation protocol. Furthermore, based on an assumed 8% increase in titre per month, up to a maximum 80% rise, up to 10 further months would be needed
Packed and expanded bed IgG capture	6	Purchasing, validating and qualifying the column and the subsequent ultrafilter as well as installing the associated instrumentation, process control equipment and electrical supply and then revalidating the remainder of the process downstream of where the centrifuge was originally
Removing the ion exchange step	5	Development of the extra affinity wash step and revalidation of the affinity step and the filtration operations downstream

Table 29: Estimated durations of developing and implementing process changes

5.7.4 Modelling assumptions

The following assumptions were made based on discussions with Protherics:

Sufficient space is available in the facility to house any extra equipment required
The entire time-course of the development involved for any of the options was assumed to occur before their implementation in the manufacturing facility
No time slippage contingency would need to be allowed for in developing and implementing a given process change option
Where combinations of development options were modelled, sufficient resources (e.g. staff or funding) would be allocated to allow different projects to start simultaneously, meaning that the total duration for those combinations was set equal to the duration of the slowest strategy to be implemented

Table 30: Assumptions made for the development model

The flow rate used in the packed bed model was half of that used for expanded bed operation
Elution volumes in expanded mode were half of those used in the packed bed model (assumption based upon advice offered by Protherics)
Based upon the evaluation of the MAbsorbent® A2P resin (Newcombe et al., 2005)), it was assumed that recovery was constant across all operational cycles and that ligand leaching was negligible over those cycles
Currently, the MAbsorbent® A2P resin is not available on expanded bed adsorption beads, but for the purposes of the analysis, it was assumed that the binding and purification characteristics of the MAbsorbent® A2P matrix would be identical in packed and expanded modes (an assumption that would need to be verified experimentally)
Initial titres for each of the four rattlesnake anti-venom IgGs were the same for every batch of starting serum
The number of batches manufactured per annum remains unchanged (Table 28)

Table 31: Assumptions made for the manufacturing model

5.7.5 Application of the MADM technique

Table 32 provides the metrics used to assess the industrial development options and values assigned to the bounds for the normalisation as detailed in the previous chapter. Values quoted for the zero and one bounds are the highest and lowest values from the entire set of simulation results for all options and combinations of options examined. Development times were normalised relative to the strategy taking the longest time – namely increasing the IgG titre (maximum of 19 months). The current process with no need for development was assigned a normalised rank of 1 for each of the two development metrics. Overall Ranks for each process change option were then calculated by Equation (24) and Equation (25).

	Metric	Zero bound	One bound
Manufacturing	Annual product mass (g F_{AB})	~40,000	~ 147,000
	Cost per gram (£/g F_{AB})	Current level	70% lower
	Batch time (hours)	~ 790	~ 470
Development	Cost (£)	Maximum value from all simulation runs	0
	Time (months)	19 months	0 months

Table 32: Values used to normalise the output values of the five performance metrics onto a zero to one scale
The highest and lowest values from the entire set of simulation results were used to set the zero and one bound values for the five metrics. Manufacturing batch time was that measured for the first blended batch

Based on discussions with Protherics, the annual manufacturing cost per gram and F_{AB} product mass were taken as being the most important metrics and assigned weights of 1/4 each, whilst the other three metrics were given equal weighting (1/6), resulting in the following required relation:

$$\sum_{i=1}^{5} w_i = 1 \tag{53}$$

The Overall Rank benchmark for the current process was calculated to be 0.42 using the current manufacturing conditions given in Table 28 on Page 122.

5.8 Results and discussion

5.8.1 Simulation results

Simulation results were represented graphically using Microsoft® Excel XP and MATLAB® (version 6.5.2, The Math Works Incorporated, Massachusetts, U.S.A.). Where appropriate, graphs plot the change in Overall Rank relative to that of the current operation (i.e. the existing process flowsheet operated with the current feed volume and IgG titre – see Table 28). Outcomes of simulated options that are superior to the current operation appear above the x axis or the x-y plane (also known as the 'Titre-volume' plane at Overall Rank = 0), for 2 and 3D graphs respectively, whilst inferior options appear below.

5.8.2 Impact of increasing the feed volume

Figure 41 shows the impact of increasing the feed volume to the current manufacturing process and to variants based on the use of MAbsorbent® A2P. Use of either packed or expanded bed column capture of IgG from a 600 L feed results in an inferior solution relative to the current manufacturing process. A breakdown of the normalised and weighted individual performance metrics in Figure 42 indicates that although the manufacturing ranks for the MAbsorbent® A2P column-based options exceed those for the current process, there is a heavy price to be paid in terms of additional development costs. The methodology used in this chapter calculates development costs as exceptional expenditures in year one and balances these against the manufacturing metrics shown in Table 32 for a single year's worth of production. Calculated in this way, the expenditure in development that has to be borne more than outweighs the advantages of increased annual product mass, reduced cost per gram and decreased batch times. Figure 41 also indicates that when operating with a 600 L feed, the Overall Rank of the packed bed option is slightly inferior to the expanded bed option, with Figure 42 showing that operation in expanded mode reduces the cost per gram and batch time to a greater extent than if using a packed bed.

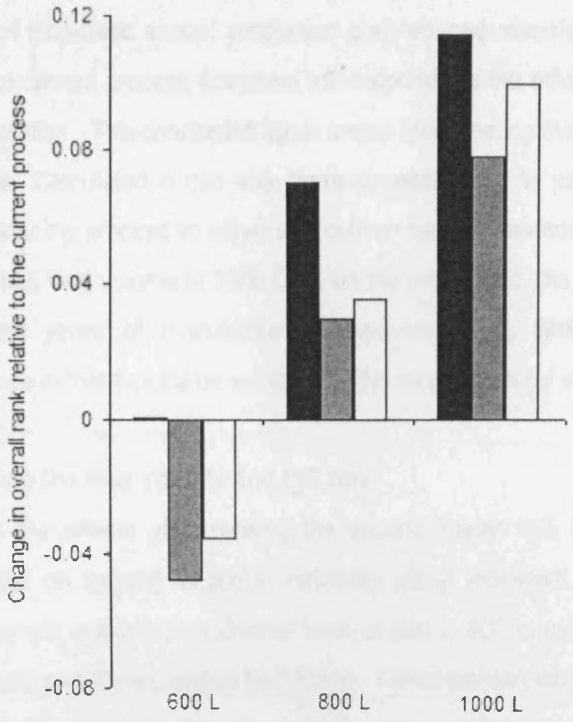


Figure 41: Impact of operating with 600 L, 800 L, 1000 L feed volumes on Overall Rank for the current (black bars), packed (grey bars) and expanded bed (white bars) processes

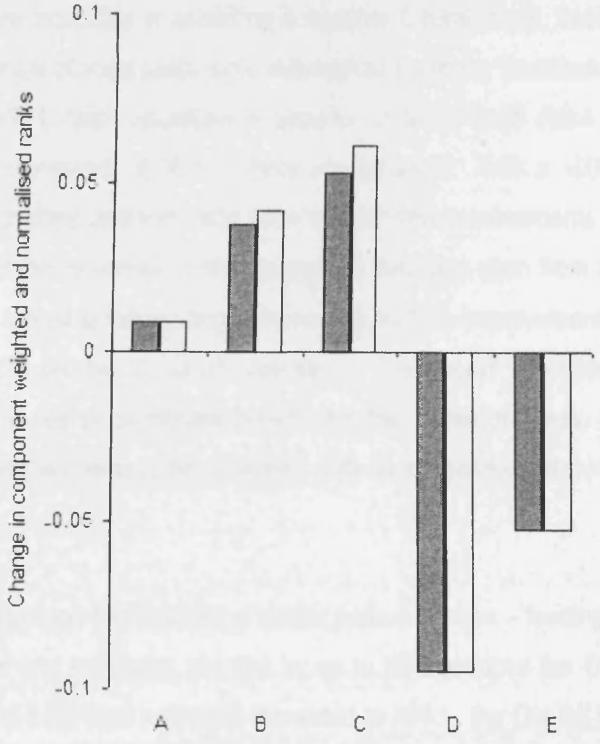


Figure 42: Breakdown of Figure 41 of the individual ranks for the five metrics for the packed bed (grey bars) and expanded bed (white bars) processes operating with a 600 L feed. Values plotted for each metric show the change relative to the weighted and normalised value for that metric in the current process flowsheet operated with a 600 L feed. A = manufacturing F_{AB} mass; B = manufacturing cost per gram; C = manufacturing batch time; D = development cost; E = development time

As the process volume is increased to 800 L and then 1000 L, the Overall Rank increases for all three options by virtue of increased annual production and reduced manufacturing cost per gram values. Nonetheless, the current process flowsheet still outperforms the other two options at 800 L and 1000 L scales of operation. This conclusion again arises from treating the development costs as exceptional expenditures. Calculated in this way, there appears to be no justification for changing from the current manufacturing process to either IgG column capture method on their own, even if combined with an increased feed volume to 1000 L. If, on the other hand, the development cost was depreciated over several years of manufacturing operations, it is likely that the improved manufacturing performance in that timeframe would merit the developmental expenditure.

5.8.3 Impact of increasing the feed volume and IgG titre

Figure 43 demonstrates the effects of increasing the venom-specific IgG titre by up to 80% (a realistic upper limit based on existing antibody variability within individual sheep and assuming sufficient time and money are available) on Overall Rank at 600 L, 800 L and 1000 L feed volumes both for the current process and the expanded bed option. For all combinations of feed volumes and titres that were tested and based on the assumptions given earlier, the expanded bed–based process outperforms the current flowsheet (the packed bed results lie in-between – data not shown for clarity). The levels of titre improvement required to generate a superior alternative can be determined from Figure 43. When operating with a 600 L of feed to the current process, increases in titre of up to 80% were incapable of achieving a superior Overall Rank, because improvements in cost per gram and annual product mass were outweighed by heavy developmental times and costs incurred. With an 800 L feed, operation is superior to the Overall Rank for current operation provided that titre improvements of 35% or more are achieved. With a 1000 L feed, operation is more favourable than current operation with even modest titre improvements. In part these results are due to an improvement in annual product mass, but they also stem from a reduction in cost per gram e.g. with an 800 L feed to the existing process and an 80% improvement in titre, cost per gram is reduced by over 50% relative to current operation. The overall implication is that in order for improvements in titre to realise significant benefits for the current process, they have to be large (greater than 35%) and they need to be combined with an increase in feed volume of at least one-third.

When operating the expanded bed process, a similar pattern is seen – feeding 600 L of serum to the expanded bed process and increasing the titre by up to 80% reduces the Overall Rank relative to current operation, whilst if the feed volume is increased to 800 L, the Overall Rank is superior to the current operation provided that the IgG titre is increased by a minimum of 20%. Expanded bed operation is favourable over the entire titre range if a 1000 L feed is utilised.

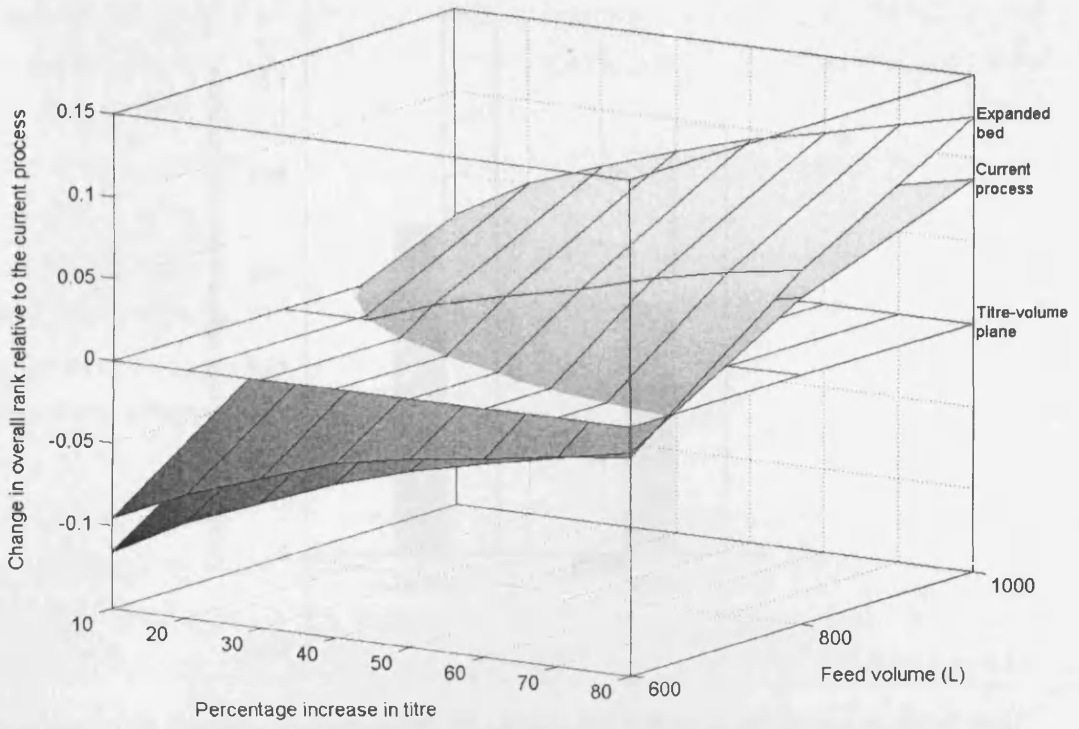


Figure 43: Impact of increasing the venom-specific IgG feed titre upon Overall Rank for the current process and the expanded bed variant operating with 600 L, 800 L and 1000 L feed volumes
Superior performance relative to current operation is achieved by combinations of feed volume and titre above the titre-volume plane i.e. the plane representing zero Overall Rank

5.8.4 Impact of removing the ion exchanger

Figure 44 examines the impact of eliminating the ion exchanger from the existing flowsheet when operating with 600 L, 800 L and 1000 L feed volumes. A small reduction in rank is seen at 600 L, because improvements achieved in the manufacturing metrics are just offset by development costs and durations. This suggests that removing the ion exchange step on its own from the current process would not be favourable. Increasing the feed volume to 800 L and 1000 L and eliminating the ion exchanger is beneficial, but still inferior to the option of simply employing the existing process with 800 L and 1000 L feed volumes. Conversely, Figure 45 demonstrates that at the feed volumes tested, the packed and expanded bed options without the ion exchange step outperform the current flowsheet and the packed and expanded bed options, when the ion exchanger is still in place. This is due to cumulative improvements in the manufacturing metrics which significantly outweigh the development costs. The implication is that using direct column capture whilst also eliminating the ion exchange step can achieve a superior process compared to current operation.

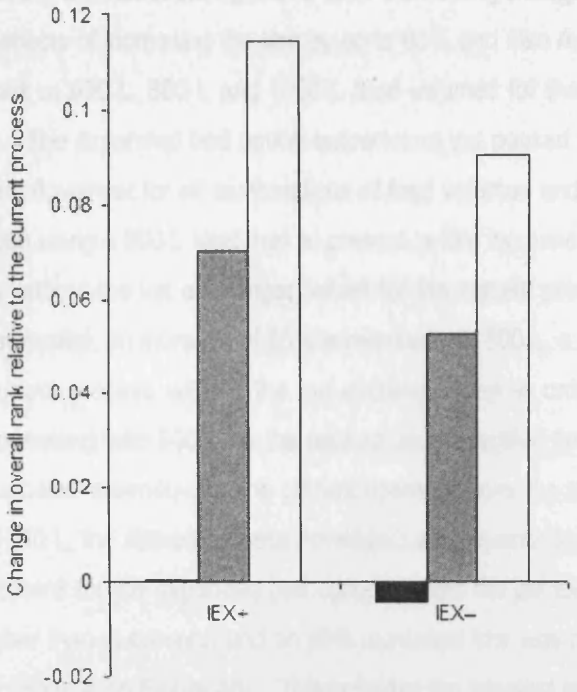


Figure 44: Impact of retaining and removing the ion exchange step from the current process on Overall Rank, operating with 600 L (black bars), 800 L (grey bars) and 1000 L (white bars) feed volumes
Presence or absence of the ion exchanger is denoted by IEX+ and IEX- respectively

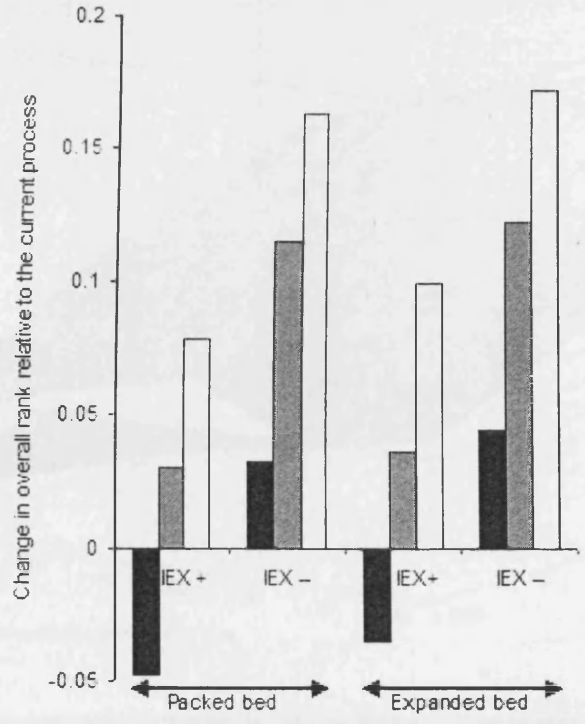


Figure 45: Impact of retaining and removing the ion exchange step from the packed and expanded bed processes on Overall Rank, operating with 600 L (black bars), 800 L (grey bars) and 1000 L (white bars) feed volumes.
Presence or absence of the ion exchanger step is denoted by IEX+ and IEX- respectively

5.8.5 Impact of removing the ion exchanger and also increasing the IgG titre

Figure 46 shows the effects of increasing the titre by up to 80% and also removing the ion exchange step upon Overall Rank at 600 L, 800 L and 1000 L feed volumes for the current process and the expanded bed option. The expanded bed option outperforms the packed bed option, which in turn outperforms the current flowsheet for all combinations of feed volumes and titres tested. In order to achieve better operation using a 600 L feed than at present, a titre increase of 20% is needed for the expanded bed option without the ion exchanger, whilst for the current process flowsheet operating without the ion exchange step, an increase of 65% is required. At 800 L, a titre improvement of 20% is required for the current process without the ion exchange step in order to achieve a superior Overall Rank, whilst operating with 800 L for the packed and expanded bed options without the ion exchanger results in a better alternative to the current operation over the complete 80% titre range. In all three cases at 1000 L, the altered process flowsheets are superior to the current operation. In particular, the Overall Rank for the expanded bed option without the ion exchanger operating with a 1000 L feed (66% higher than at present) and an 80% increased titre was the highest seen out of all the options examined (Point A on Figure 46) . This provides the greatest return in manufacturing for the investment in development and hence would be the most desirable replacement to the current operation.

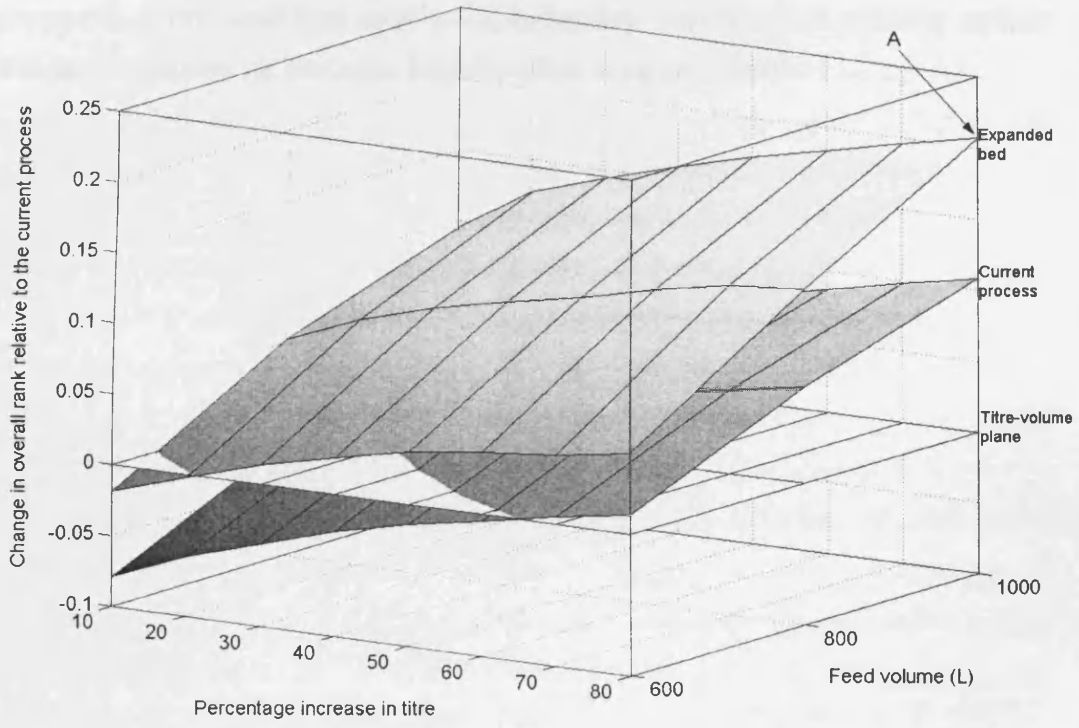


Figure 46: Impact of removing the ion exchange step and increasing the venom-specific IgG feed titre on Overall Rank for the current and expanded bed flowsheets, operating with 600 L, 800 L and 1000 L feed volumes
Superior performance relative to current operation is achieved by combinations of feed volume and titre above the titre-volume plane

5.9 Conclusions

This chapter has used a computer simulation to evaluate the impact of developing and implementing a series of alternative industrial production process scenarios, subject to predicted levels of changes in key input variables such as feed titre. Each option was assessed using a multi-attribute decision making technique, both in terms of annual manufactured product mass, costs of goods and manufacturing batch times as well as developmental costs and durations. Of all the process change options that were modelled, combining an expanded bed column with a 66% higher feed volume and an 80% higher titre together with the elimination of the ion exchange step provided the most desirable alternative to current operation. Amongst other changes evaluated, the simulation demonstrated that if the feed volume to the existing process was increased from 600 L to 800 L and the titre was increased by 80%, the cost per gram could be reduced by over a half. Such data can be used for the rapid assessment of process alternatives in line with commercial aims and objectives and thus ensure that the most attractive new process scenario is selected to replace current operation.

In these last three chapters, the process change scenarios examined through simulation were defined by Protherics; it is also possible, however, that the company may not know exactly which other factors or properties of the process provide the greatest potential for achieving manufacturing improvements. Such knowledge would be highly beneficial and the use of sensitivity analysis techniques for acquiring this information forms the subject of the next chapter.

6: GLOBAL SENSITIVITY ANALYSIS FOR DETERMINING PARAMETER IMPORTANCE IN BIOPROCESSES

6.1 Introduction

Although computer simulations such as those presented in previous chapters offer the potential for determining the impacts of changing manufacturing variables such as titres or flowrates, their effective use for process optimisation also requires knowledge of the relative influence that each input variable exerts upon outputs, allowing development studies to be concentrated upon the most important factors controlling manufacturing performance (Chhatre et al., 2007d). When simple material balance expressions such as black box models are employed, identifying the most critical variables can be achieved readily by visual inspection, but when more complicated, mathematically rigorous expressions (such as the non-linear equations given in Chapter 4) are used, inspection may prove difficult or impossible because of the complexity and size of the design space. In such circumstances, more sophisticated approaches are required.

One such method which has attracted recent interest is the use of sensitivity analysis techniques, which use calculus-based algorithms to determine how variations in model outputs are influenced by changes to the inputs (Saltelli et al., 2000). Local Sensitivity Analysis (LSA), for example, involves calculation of the partial derivative with respect to individual input variables at specific points on the output function (McRae et al., 1982). Although a partial derivative provides valuable data about what occurs at that point and in its immediate vicinity, this approach is inadequate for describing the behaviour elsewhere in the model. To determine sensitivities at other parts of the function using LSA would require the computationally intensive and time consuming evaluation of partial derivatives at every point of interest. Furthermore, LSA only analyses one input at a time (King et al., 2007), such that complete model characterisation would necessitate separate calculations for every variable. Global Sensitivity Analysis (GSA) is a more powerful method that assesses all input parameters simultaneously over their full ranges of uncertainty and ranks them on a zero to one scale to signify the average contribution that each one makes towards the variance of the model. Variable ranks can be used to quantify both the impact of individual variables on a model as well as their interactions. By identifying the most critical input factors in this way, GSA concentrates process development efforts upon the most important variables and aids identification of the most suitable conditions for operating a drug manufacturing process.

Although GSA has been used previously to assess the sensitivities of processes in sectors such as the chemical industry (Schaibly and Shuler, 1973), its application in the bioprocess arena has been limited to individual unit operations. Kontoravdi et al. (2005), for example, applied GSA to a model of mammalian cell culture producing monoclonal antibodies, while King et al. (2007) used the method

for the centrifugal clarification of fermentation broths. The wider application of GSA would bring significant benefits to the bioprocessing sector, as it can be used in conjunction with regulatory concepts such as Quality By Design (QBD) in order to achieve more efficient process development. QBD argues for a thorough characterisation of the design space by quantifying the influence of each variable upon the process. The resulting information can then be used to determine the most suitable operating position with respect to technical outcomes such as yield, purity, product quality and process robustness (i.e. operating away from edges of failure). GSA can help to achieve this goal by calculating the relative control exerted by different manufacturing variables and thus enabling unimportant variables to be screened out. By uncovering the most critical process parameters in this way, GSA establishes a basis for choosing the most suitable fractional factorial design for those key inputs alone. Hence in order to illustrate the utility of GSA in bio-manufacturing, this chapter describes the first application of the technique to mathematical models of bioprocesses, again using the CroFab™ process as a case study. Two sets of studies have been undertaken, both of which have been submitted to journals (Chhatre et al., 2007d; Chhatre et al., 2007e):

1) Applying the GSA methodology to the synthetic Protein A-based purification of ovine serum
2) Widening the first study to consider the variable sensitivities of the whole CroFab™ process

6.2 Description of case studies

6.2.1 Synthetic Protein-A purification

In this first study, GSA was used to determine variable importance for the synthetic Protein A purification of IgG. Although recombinant Protein A is regularly employed industrially for the purification of feeds such as monoclonal antibodies from clarified cell culture supernatant (Fahrner et al., 1999; Swinnen et al., 2007), high resin costs and the requirement for expensive cleaning solutions such as urea or guanidine hydrochloride have prompted the development of synthetic alternatives with potentially greater affinity for antibodies from feeds such as crude ovine serum (Hober et al., 2007). This study therefore models the synthetic adsorbent-based chromatographic capture of antibodies from sheep serum (Newcombe and Newcombe, 2007; Chhatre et al., 2007b; Chhatre et al., 2007f). As described in the previous chapter, the synthetic adsorbent step was simulated using the Thomas model (Montesinos et al., 2005), in order to predict the amounts of antibody that are captured upon serum application to the column. Assumed data values obtained from Protherics (Newcombe et al., 2005) were used to populate the Thomas model, after which it was subjected to GSA to quantify the influence exerted upon antibody yield by the IgG concentration, duration of feed application, matrix capacity and flowrate.

6.2.2 Whole process modelling

To extend the application of the technique, GSA was then used to quantify the impact of different input variables upon the yield and throughput of the final CroFab™ product, using one of the proposed processes described in Chapter 4. In the process, the hyperimmunised serum feed is initially subjected to a synthetic affinity ligand adsorbent step to recover the antibodies, after which they are digested by papain, concentrated and diafiltered by ultrafiltration, ion exchanged and finally affinity purified to recover the venom-specific product. This study modelled the purification of a single batch of ovine serum passed through this process.

6.3 Synthetic Protein-A purification

6.3.1 Mathematical basis to GSA

As outlined above, GSA calculates the relative importance of every input variable by assessing its contribution to the variance of the model function. This variance-based rank for each variable is called its sensitivity index (SI) and is expressed on a zero to one scale. The GSA method used in this part of the chapter to calculate the SI terms (Sobol', 2001) relies on defining the function $f(\boldsymbol{x})$ of n arguments within an n^{th} dimensional unit hypercube with each input value ranging from zero to one (its 'range of uncertainty'). The function can then be expressed as a series of summands of increasing dimensionality:

$$f(\bar{x}) = f_0 + \sum_{s=1}^{n} \sum_{i_1<\ldots<i_s}^{n} f_{i_1\ldots i_s}\left(x_{i_1\ldots i_s}\right); 1 \le i_1 < \ldots < i_s \le n \tag{54}$$

This equation implies that any function $f(\boldsymbol{x})$ can be represented as a sum of functions $f_{i1\ldots is}$ which take arguments from within $\boldsymbol{x}$. Upon expansion, Equation (54) becomes Equation (55):

$$f(\bar{x}) = f_0 + \sum_{i} f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j) + \sum_{i<j<k} f_{ijk}(x_i, x_j, x_k) + \ldots \tag{55}$$

Equation (55) requires that f_0 is a constant and that the integrals between unit limits of every summand of any of its own variables are equal to zero:

$$\int_0^1 f_{i_1\ldots i_s}\left(x_{i_1\ldots i_s}\right) dx_k = 0;\ k = i_1 \ldots i_s \tag{56}$$

From Equation (56), it follows that the members of Equation (54) are mutually orthogonal and can be expressed as integrals of $f(\boldsymbol{x})$:

$$\int_0^1 f(\bar{x}) d\bar{x} = f_0 \tag{57}$$

$$\int_0^1 f(\bar{x})dx_k\Big|_{k\neq i} = f_0 + f_i(x_i) \tag{58}$$

$$\int_0^1 f(\bar{x})dx_k\Big|_{k\neq i,j} = f_0 + f_i(x_i) + f_j(x_j) + f_{ij}(x_i,x_j) \tag{59}$$

In Equation (59), $f_{ij}(x_i,x_j)$ represents the interaction between the two input variables. This expansion can be continued for higher order terms. If $f(\bar{x})$ is assumed to be square integrable, then it follows that all of the $f_{i_1\ldots i_s}$ terms are square integrable as well. Squaring Equation (54) and integrating over the unit hypercube generates the following expression:

$$\int_0^1 f^2(\bar{x})dx - f_0^2 = \sum_{s=1}^{n}\sum_{i_1<\ldots<i_s}^{n}\int_0^1 f_{i_1\ldots i_s}^2\left(x_{i_1\ldots i_s}\right)dx_{i_1\ldots i_s} \tag{60}$$

The left and right hand sides of Equation (60) can be re-written as follows:

$$D_T = \int_0^1 f^2(\bar{x})dx - f_0^2 \tag{61}$$

$$D_{i_1\ldots i_s} = \int_0^1 f_{i_1\ldots i_s}^2\left(x_{i_1\ldots i_s}\right)dx_{i_1\ldots i_s} \tag{62}$$

D_T and $D_{i_1\ldots i_s}$ are the total and partial variances respectively, such that:

$$D_T = \sum_{s=1}^{n}\sum_{i_1\ldots i_s}^{n} D_{i_1\ldots i_s} \tag{63}$$

Taking the ratio between the partial variance and total variance in Equation (61) generates Sensitivity Indices ($S_{i_1\ldots i_s}$), each with values between zero and one. SI terms can be calculated to quantify both the importance of individual variables ('first order' indices) and their interactions ('interaction' or 'higher order' indices). The sum of all indices always equals 1.

$$S_{i_1\ldots i_s} = \frac{D_{i_1\ldots i_s}}{D_T};\ 0 \le S_{i_1\ldots i_s} \le 1 \tag{64}$$

6.3.2 Modelling details

The synthetic Protein A-based purification was modelled using the Thomas model (see previous chapter), simulating breakthrough curves for the amount of antibody that flows through without being captured by the column. Figure 47 plots a theoretical breakthrough curve, in which 'X' ranges from zero to one and represents the fraction of the antibody in the feed that flows through the column without interacting with the matrix and which is therefore not recovered. The areas above and below the breakthrough curve respectively represent the quantities of antibody that are bound and lost.

't_{Load}' is the duration for which feed is applied to the column. In order to determine antibody recovery (i.e. yield), it was assumed that all antibody molecules bound to the matrix were successfully recovered by elution from the column. Recovery was then calculated directly by integration of the breakthrough curve using the trapezium rule to determine the proportion of antibody mass that is lost ($Area_{Lost}$). $Area_{Total}$ in Equations (65) and (66) represents the total amount of antibody applied:

$$Area_{Total} = Area_{Bound} + Area_{Lost} = t_{Load} \times 1 = t_{Load} \tag{65}$$

$$Area_{Bound} = Area_{Total} - Area_{Lost} \tag{66}$$

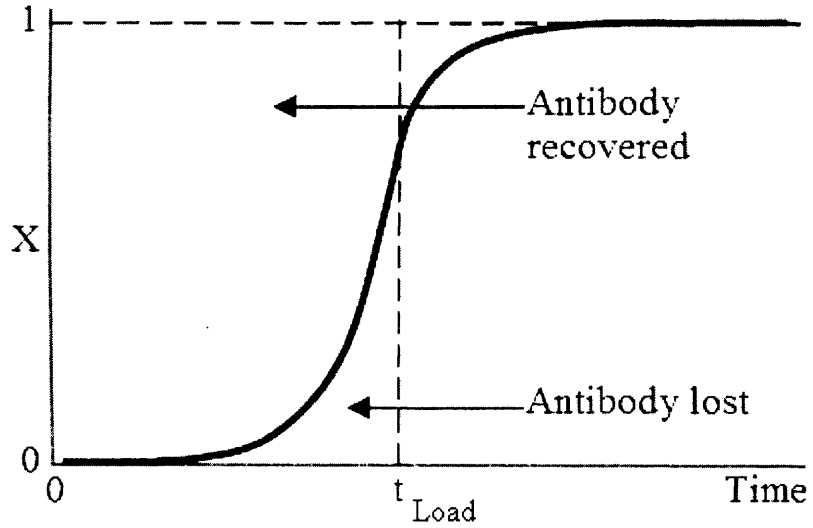


Figure 47: Theoretical breakthrough curve

The proportion of antibody recovered i.e. the yield is then determined by the following equation:

$$\gamma = \frac{Area_{Bound}}{Area_{Total}} \tag{67}$$

Input variables were split into two groups – one for those where sensitivity indices were required for reasons of commercial interest (see Table 33) and the other for the remaining inputs (Table 34):

Variable		Value
C_0	Feed concentration (mol/m^3)	0.15 - 0.2
t_{Load}	Duration of feed application (s)	900 - 1000
q_m	Matrix capacity (mol/m^3)	0.15 – 0.20
v	Linear flowrate (m/s)	3 - 4 × 10^{-4}

Table 33: Input uncertainty ranges of values for variables for which SI terms were calculated

For the remaining variables, the following values were assigned:

Variable		Value and comment	
K_d	Equilibrium desorption constant (-)	10^{-4}	These data were based upon literature values (Hahn et al., 2003)
K_1	Adsorption rate constant (m^3/mol/s)	0.129	
ε	Bed porosity (-)	0.4	Values assumed from Montesinos et al. (2005)
L	Column length (m)	0.2	Assumed column length (Protherics)

Table 34: Values for variables that were fixed during the GSA analysis

6.3.3 Input variables scaling

Although the inputs for variables in the GSA method range from zero to one, values for chromatography model do not take that range and were therefore scaled using Equation (68):

$$\lambda_{Scaled} = \lambda_{GSA} \times (\lambda_U - \lambda_L) + \lambda_L \tag{68}$$

λ_{Scaled} is the scaled input value, λ_{GSA} is the input value in the zero to one range and λ_L and λ_U are the real lower and upper bounds respectively representing the zero and one bounds in the GSA method. For example, if λ_{GSA} is 0.5, λ_L is 0.15 and λ_U is 0.2, then λ_{Scaled} is calculated by Equation (21):

$$\lambda_{Scaled} = 0.5 \times (0.2 - 0.15) + 0.15 = 0.175 \tag{69}$$

6.3.4 Implementation

The GSA methodology given in Section 6.3.1 had already been encoded into MATLAB® by Dr Josh King (Department of Biochemical Engineering, University College London) and hence this used to complete the analysis for this case study. In this analysis, the four variables for which sensitivity indices were calculated were the antibody feed concentration (C_0), matrix capacity (q_m), duration of feed application (t_{Load}) and mobile phase flowrate (v); all other variables were kept constant for the analysis. All integrals in the GSA method were evaluated using a four strip trapezium rule. The MATLAB® code has been placed into Appendix 2.

6.3.5 Manufacturing strategies evaluated in the study

6.3.5.1 Base case

The sensitivities of the antibody recovery calculation i.e. Equation (67) were initially established by applying the GSA method to the Thomas model using the data values given in Table 33 and Table 34 in order to create an initial base case set of results. After this, three process changes were made to investigate their impacts upon parameter importance

6.3.5.2 Vary the loading duration

To examine the effects of varying the loading duration upon the sensitivity indices, the following ranges were entered into the model instead of the base case 900-1000s: 700-800; 800-900; 1000-1100 and 1100-1200 seconds. There are both positive and negative outcomes from either increasing or decreasing the loading duration range. Higher durations result in moving further up the breakthrough curve, causing greater product loss in the flowthrough – although this needs to be balanced against more of the feed being processed per loading cycle. Conversely, lower loading times result in higher yields at the expense of extended processing times (resulting from the potential requirement to increase the number of operational cycles in order to process all of the feed). This suggests that a trade-off between the inputs is needed in order to identify the optimal operating conditions that satisfy requirements for outputs such as product yield and throughput. Based upon the Global Sensitivity Analysis data, this trade off was quantified through a window of operation.

6.3.5.3 Vary the flowrate

Again, varying this variable will have potentially conflicting effects in terms of processing times and recovery and hence establishing the impact of the different variables upon product yield is critical. Instead of using the base case range of 3 - 4 $\times 10^{-4}$ m/s, the effects of using the following ranges upon the sensitivity indices were determined: 4-5, 5-6 and 6-7 $\times$ 10^{-4} m/s (all ranges fall within the maximum flowrate that can be tolerated by the synthetic matrix upon which the study was based i.e. MAbsorbent® A2P – see Chapter 7 and Newcombe et al., 2005).

6.3.5.4 Increase the level of uncertainty in the inputs

Another application of the technique involves examining the impact of greater uncertainty in the input ranges. During process development, it is possible that values of variables may not be well defined and hence may be open to wider ranges than first thought. For example, if a feedstock has not been well-characterised, then the product concentration fed to the downstream process could vary over more extensively than originally anticipated. Alternatively, a development group may wish to examine the impact of using a wider range of flowrates upon product yield. To illustrate the effect of this uncertainty upon the sensitivity indices, the final study widened the input ranges for C_0 and v to 0.15 – 0.4 mol/m^3 and 2 – 6 $\times$ 10^{-4} m/s respectively, while all other variables remained unaltered (the upper end of the C_0 range was chosen as it is a value that can be achieved without causing solubility problems for the antibody [A.R. Newcombe, Protherics U.K. Limited, personal communication]).

6.3.6 Results and Discussion

6.3.6.1 Initial base case

Figure 48 plots the results for the initial base case, using the inputs given in Table 33. The antibody binding capacity was the single most important parameter affecting recovery (SI = 0.32), followed by

the flowrate (SI = 0.30) and then feed concentration (SI = 0.28). These factors should therefore form the focus of process optimisation studies. Conversely, the duration of feed application makes relatively little impact upon recovery (SI = 0.04) and thus if all other parameters were kept constant, yield would not change significantly over the 900-1000 second duration. It is also important to note that predominantly, the four input variables act independently, with little interaction between the terms (interactive SI of 0.06 only).

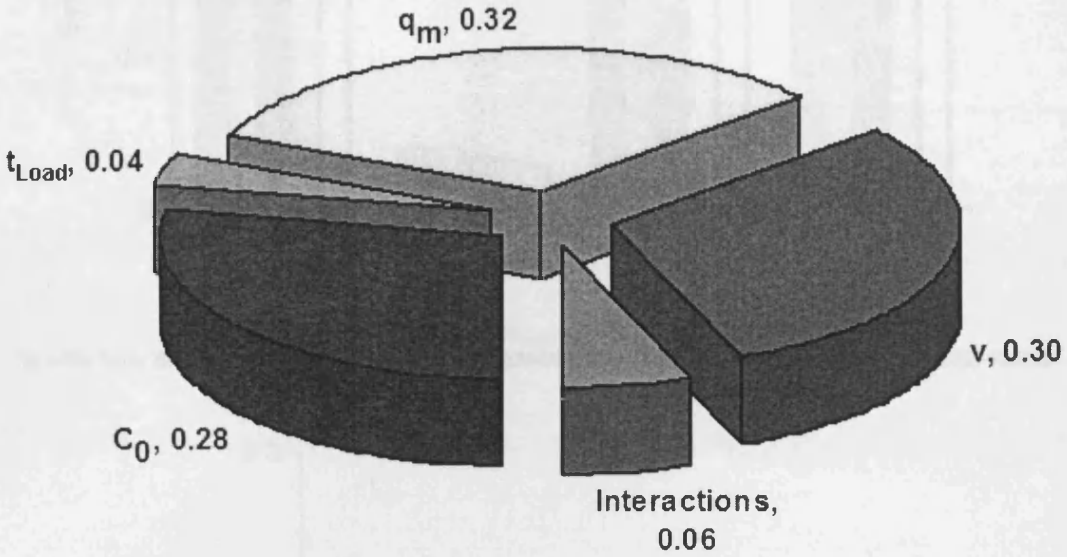


Figure 48: Base case sensitivity indices
The first order indices account for 94% of the model variance

6.3.6.2 Vary the loading duration

One of the simplest alternatives to the process involves changing how long the antibody feed is applied to the column. The duration of feed application ('t_{Load}') was varied in the ranges of 700-800; 800-900; 1000-1100 and 1100-1200 seconds as well as the base case (900-1000s). The impact upon the SI values is plotted in Figure 49, which shows that as the loading duration increases, its impact diminishes and the other three variables become progressively more important. Additionally, at the lowest loading duration range (700-800 seconds), the GSA method predicts that the first order indices sum to a value of approximately 0.65, with fairly significant interactions of just under 0.1 for each of the following pairs of variables: C_0/q_m; C_0/v and q_m/v (Figure 50). Hence if the duration of feed application was reduced to maximise antibody recovery, feed concentration, column capacity and flowrate would need to be optimised together, as their interactions make a significant contribution to the output yield values. In addition, this implies that the process is more complicated at lower loading durations – the interactions make it more difficult to control, resulting in implications for validation due to the process becoming harder to control. Thus any improvements in yield achieved by reducing the loading duration would need to traded off against potential process stability issues.

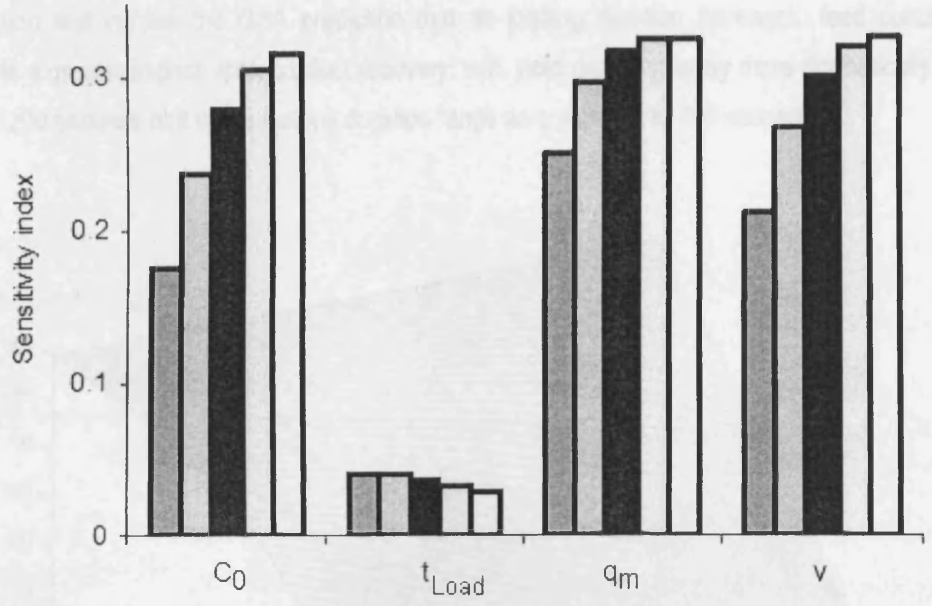


Figure 49: Impact of varying the loading duration (t_{Load}) on first order indices
In order from left to right for each variable:700-800, 800-900, 900-1000 (base case), 1000-1100 and 1100-1200 s

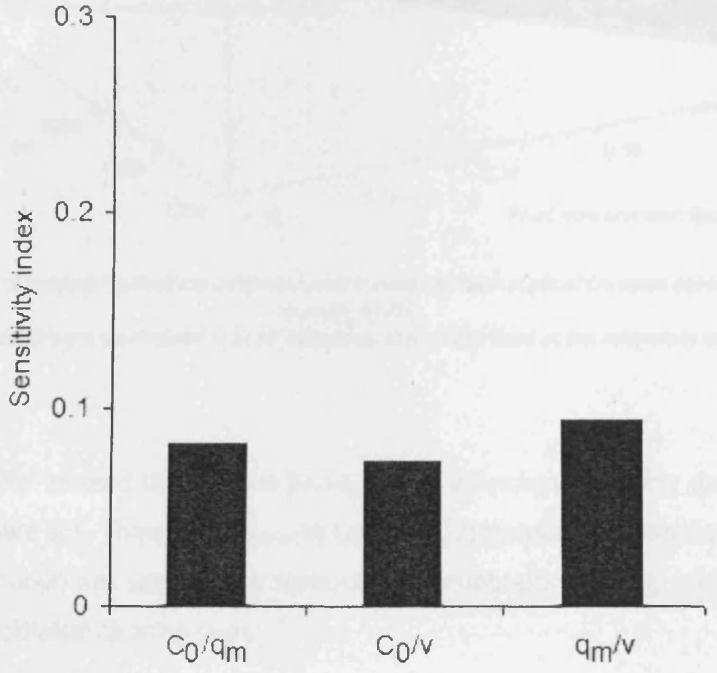


Figure 50: Interaction indices obtained at the lowest loading duration tested (700-800s)

6.3.6.3 GSA as a technique for developing windows of operation

Knowledge of variable sensitivities enables optimisation of the most important inputs in order to achieve desired manufacturing performance. For example, the knowledge that feed concentration becomes more important as loading duration increases (Figure 49) encourages further investigation into the model to determine the exact nature of the relationship between C_0, t_{Load} and antibody yield.

Figure 51 plots percentage antibody recovery as a function of feed concentration and loading duration and verifies the GSA prediction that as loading duration increases, feed concentration exerts a greater impact upon product recovery, with yield dropping away more dramatically towards the 1200 seconds end of the loading duration range as compared to 700 seconds.

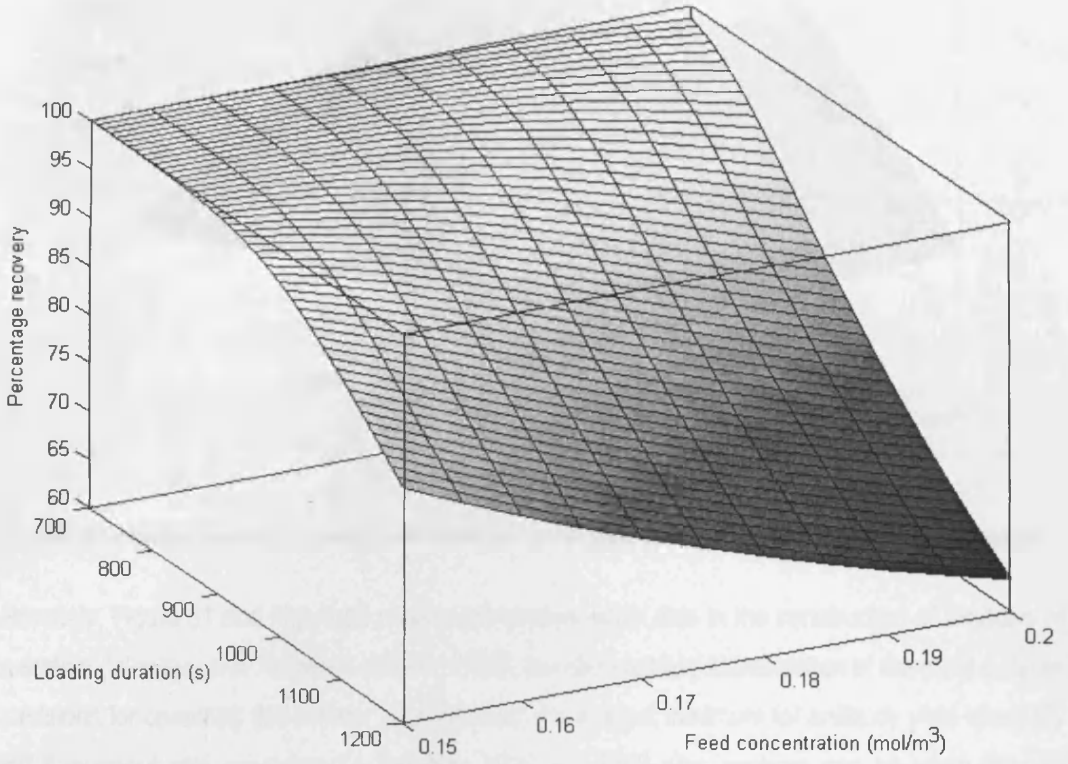


Figure 51: Impact of varying the feed concentration and duration of feed application upon percentage antibody recovery (yield)
All other parameters were maintained at fixed values (q_m and v were fixed at the midpoints of their ranges)

The model can also be used to determine throughput as a function of loading duration and feed concentration (Figure 52). Throughput (τ_{column} in Equation (22) measured in grams of IgG processed by the column per hour) was calculated by accounting for equilibration, loading, washing and eluting times as well as acid/alkali cleaning steps:

$$\tau_{column} = \frac{\gamma \cdot M_F}{t_{Equil} + t_{Load} + t_{Wash} + t_{Elute} + t_{HCl} + t_{NaOH}} \tag{70}$$

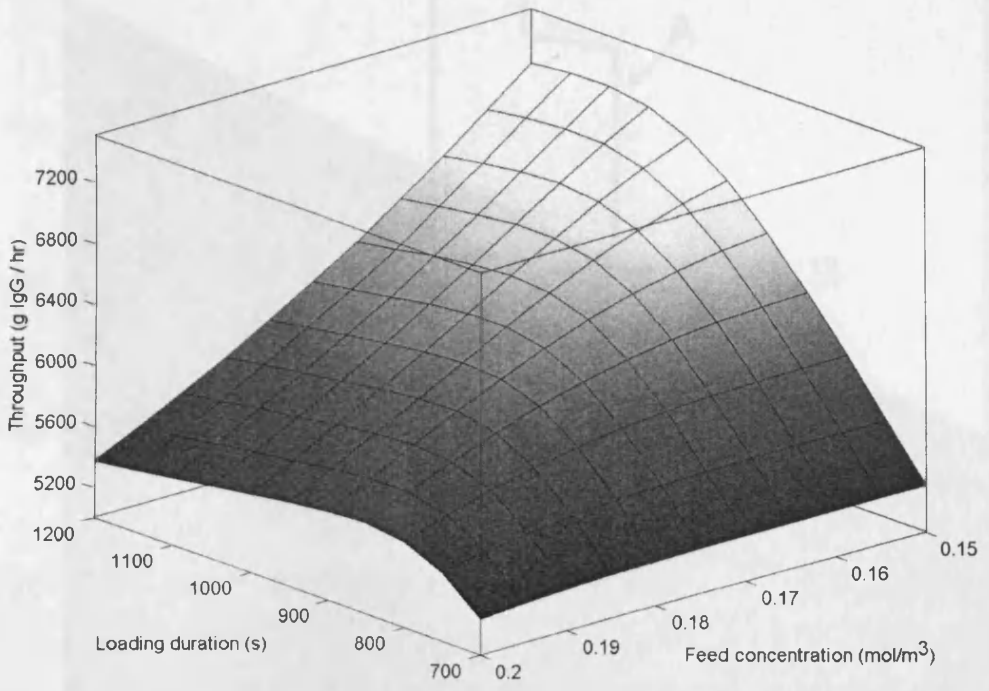


Figure 52: Impact of varying the feed concentration and duration of feed application upon antibody throughput

Ultimately, Figure 51 and Figure 52 provide information which aids in the construction of windows of operation (Woodley and Titchener-Hooker, 1996), thereby enabling identification of the most suitable conditions for operating this column. For example, if the target minimum for antibody yield was 85% and throughput was constrained to between 6000 and 6500 g/hr, sections can be taken through Figure 51 and Figure 52 and the feasible operating space which meets these criteria can then be plotted (Figure 53). The inverted L-shaped region indicated as A in this plot encloses an area that satisfies the throughput constraint, whilst the dark region (B) meets the yield target. Overlapping A and B produces the grey area in the centre of the graph (C), which represents the feasible window of operation for this column that satisfies both yield and throughput requirements. Hence using Global Sensitivity Analysis has allowed us to identify the key input variables, thus providing a rational basis for process design as visualised through windows such as Figure 53. A key problem often associated with modelling approaches is the lack of confidence in their outcomes – this can be overcome by techniques methods such as windows of operation, which enable complex modelling outputs to be visualised in an intuitive fashion, ultimately enabling identification of the most appropriate operating conditions for a given process specification (Titchener-Hooker et al., 2001).

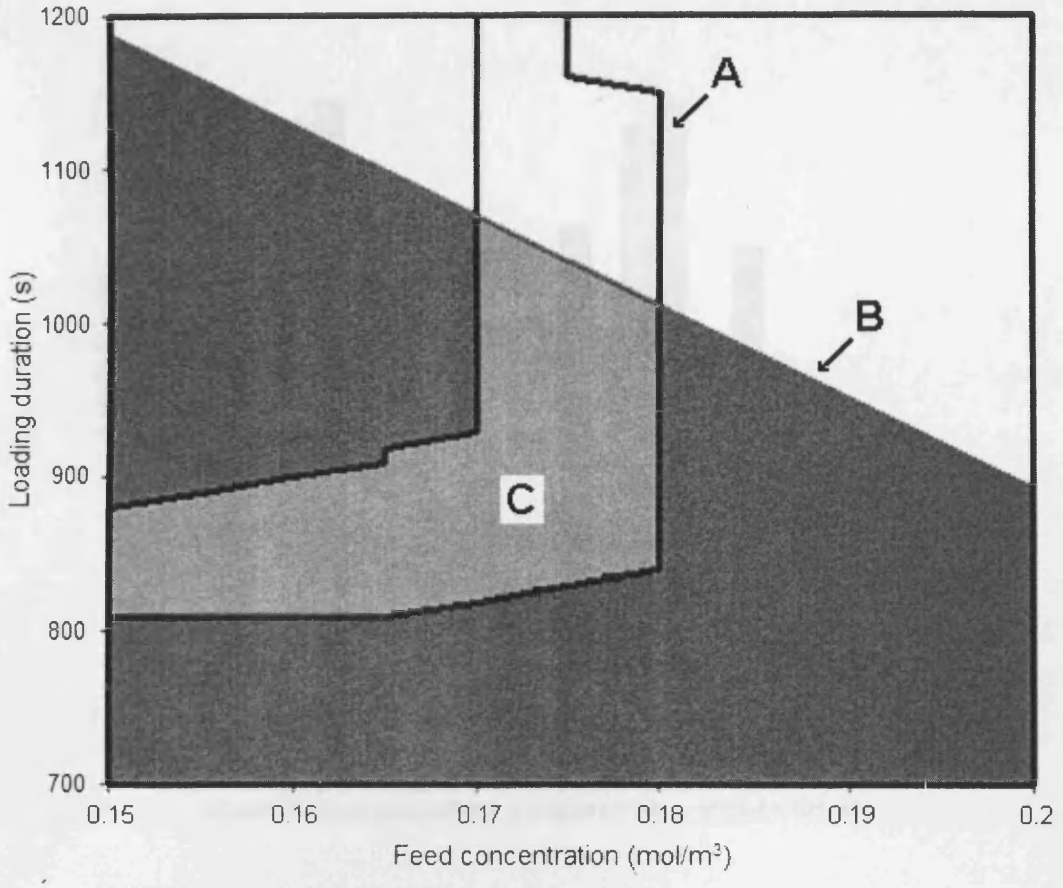


Figure 53: Window of operation showing the feasible operating region for feed concentration and loading duration. ***Areas A and B identify those regions which satisfy throughput and yield constraints respectively. Area C is the window created by overlapping regions A and B***

6.3.6.4 Vary the flowrate

The impact of employing flowrates in the ranges 3-4 (base case), 4-5, 5-6 and 6-7 × 10^{-4} m/s upon the sensitivity indices are shown in Figure 54. Increasing the flowrate diminishes its impact upon yield, whilst feed concentration and saturation capacity become more important to roughly the same extent as each other. At relatively low value ranges of v, the flowrate is able to exert significant control over the yield from the column, but once its value increases, any further influence it has is lost and the yield limiting variables shift further towards C_0 and q_m. Hence, the higher the flowrate, the less worthwhile it becomes to optimise this variable.

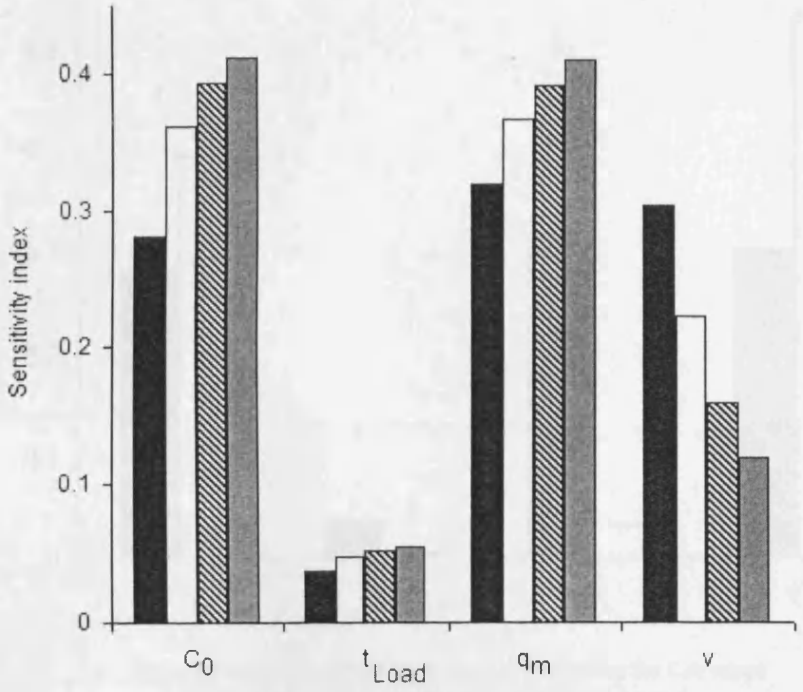


Figure 54: Impact of varying the flowrate upon the first order sensitivity indices
(Black = 3-4 (base case); white = 4-5; striped = 5-6; grey = 6-7 × 10^{-4} m/s)

6.3.6.5 Increase the level of uncertainty in the inputs

The final study examined the effect of increasing the range of uncertainty for the feed concentration and flowrate. The resulting SI values are presented in Figure 55 and show that these changes increase the influence of these variables significantly, at the expense of the other two inputs (especially q_m). This result was again illustrated by using the model to calculate yield as a function of feed concentration/flowrate (Figure 56) and again as a function of matrix capacity/loading duration (Figure 57) when employing greater uncertainty. Figure 56 shows that C_0 and v have a dramatic impact, confirming the GSA results for feed concentration and flowrate, whereas Figure 57 shows that recovery is now a far weaker function of q_m and t_{Load}. Ultimately, such data enables the efficient identification and selection of process optimisation strategies in order to achieve required goals for outputs such as yield or throughput.

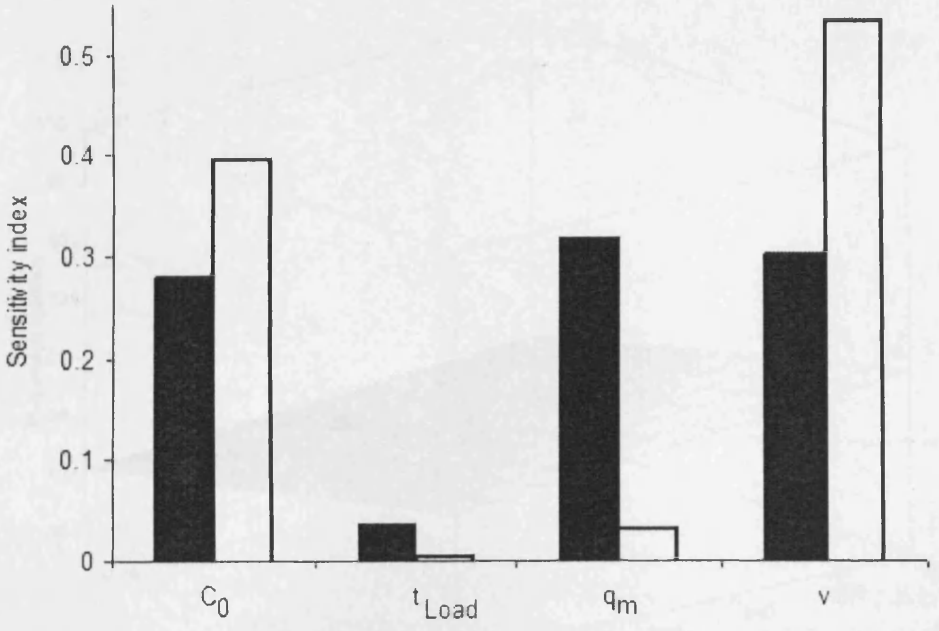


Figure 55: Impact upon first order indices of widening the C_0/v range
(Black = base case ranges; white = widened ranges)

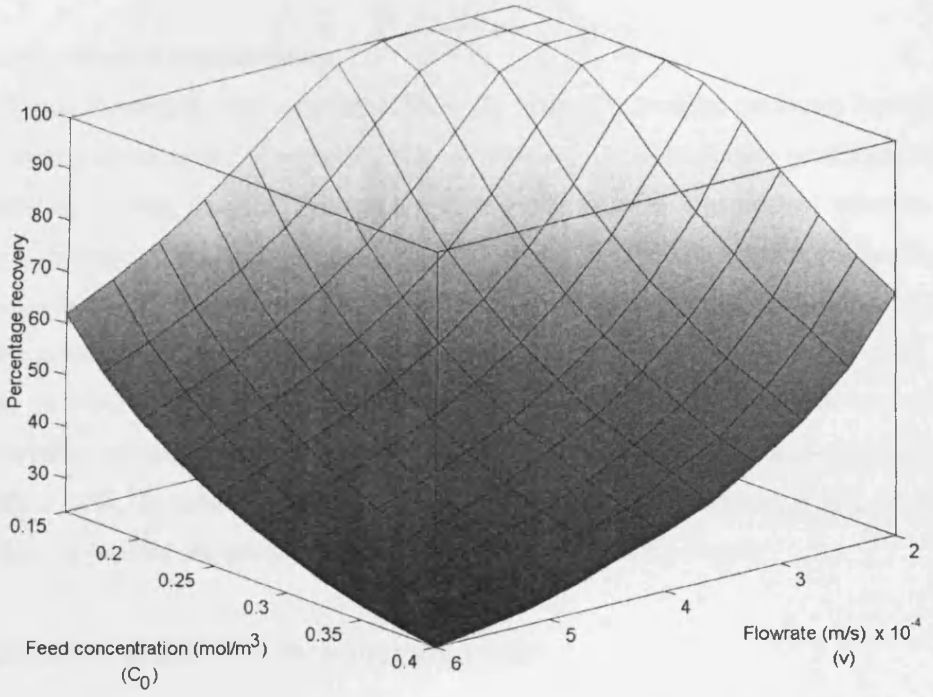


Figure 56: Impact upon recovery of varying C_0 and v when using a wider C_0/v range

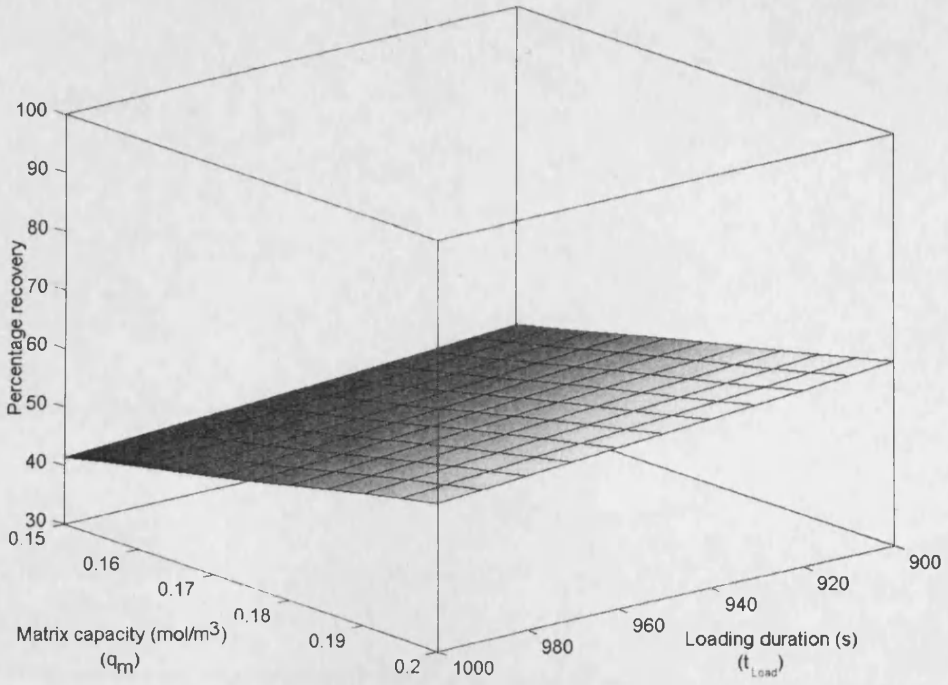


Figure 57: Impact upon recovery of varying q_m and t_{Load} when using a wider C_0/v range

6.3.7 Conclusions of the case study

This first GSA investigation has used Global Sensitivity Analysis to evaluate parameter importance in the chromatographic capture of antibodies from ovine serum. Under base case conditions, the feed concentration, binding capacity and flowrate were found to have the greatest influence upon recovery. Further studies investigated the effects of varying input values such as the loading time, which showed that at the lower end of the duration range, interactions between feed concentration, flowrate and matrix capacity made important contributions towards yield. Another study which widened the input ranges for feed concentration and flowrate showed that greater uncertainty in these variables increased their predicted influence upon recovery. Although such results illustrate the utility of GSA, its value would be enhanced by application of the technique to a whole drug manufacturing process and this is the subject of the next section of this chapter.

6.4 Application of GSA to a whole process model

6.4.1 Introduction

The second case study expands the GSA treatment to the whole manufacturing process for CroFab™, looking at which variables most significantly influenced both the yield and throughput of the product. The process used as the basis of the exercise was as follows: a synthetic adsorbent column was used initially to recover IgG, after which digestion by papain yielded a mixture containing

F_{AB} and F_C. After ultrafiltration, ion exchange and venom-specific affinity purification steps were employed to yield the final product. For simplicity, the intervening depth filtration steps that are found in the current CroFab™ process were not modelled. The unit operation equations used to simulate the process were as follows:

- All chromatography steps were simulated using the Thomas model, populated by data provided earlier. Throughput for each column was calculated using an equation of the form of Equation (70), accounting for loading, washing, eluting, regeneration and re-equilibration steps
- The papain digestion equation developed in the previous chapter was employed for this step
- A concentration factor (CF) was defined to calculate the concentration of material fed to the subsequent ion exchange step (V_F = feed volume; $V_{Retentate}$ = retentate volume). The duration of the ultrafiltration step included times for concentration, diafiltration, transferring processed material and filter flushing:

$$CF = \frac{V_F}{V_{Retentate}} \tag{71}$$

The final expression used to calculate the quantity of venom-specific F_{AB} emerging from the end of the process was defined as follows:

$$\gamma_{VS\,F_{AB},Final} = \frac{M_{VS\,F_{AB},\,Final}}{M_{VS\,F_{AB},\,Initial}} \tag{72}$$

The initial mass of specific F_{AB} fed to the start of the process $\left(M_{VS\,F_{AB},\,Initial}\right)$ was calculated from the initial IgG titre and feed volume (see Table 35 and Table 36 below). Throughput ($\tau_{Process}$) of the final venom specific F_{AB} (g/hr) was then determined by dividing $M_{VS\,F_{AB},\,Final}$ by the total processing time for all five unit operations:

$$\tau_{Process} = \frac{M_{VS\,F_{AB},\,Final}}{t_{Synthetic\ adsorbent} + t_{Papain} + t_{UF} + t_{IEX} + t_{CroFab\ affinity}} \tag{73}$$

Again, as with the first investigation, the inputs to these unit operation equations were divided into two groups – one for those variables for which SI terms were required and another for the remaining inputs (values obtained from Protherics).

Variable	Assumed ranges
Feed properties	
Initial total IgG feed concentration (mol/m^3)	0.15 – 0.20
Synthetic ligand column	
IgG matrix capacity (mol/m^3) Flowrate (m/s)	0. 5 – 0.7* $3 \times 10^{-4} - 5 \times 10^{-4}$
Ultrafiltration	
Concentration factor (-)	2–4
Ion exchange column	
Flowrate (m/s)	$3 \times 10^{-4} - 5 \times 10^{-4}$
CroFab™ venom-specific affinity column	
Venom-specific F_{AB} matrix capacity (mol/m^3) Flowrate (m/s)	6 – 8 $3 \times 10^{-4} - 5 \times 10^{-4}$

Table 35: Assumed base case data values for key variables which are mapped to the zero to one input ranges in the GSA analysis
Input ranges reflect uncertainty or variability in properties of the process or the feed modelled in the case study (* - a higher capacity matrix was used than in the previous study to determine the effects of this upon the sensitivity indices)

Variable	Assumed values
Feed properties	
Feed volume (m^3) Proportion of total IgG that is assumed to be venom specific (–)	0.50 0.5
Synthetic ligand column	
Column dimensions (diameter × height) (m × m) Voidage fraction (ε) (–)	0.60 × 0.20 0.4
Papain digestion	
Temperature of digestion (°C) Duration of digestion (hr) Concentration of enzyme (%)	40 7 1.75
Ultrafiltration	
Yield of F_{AB} (–)	0.65
Ion exchange column	
Column dimensions (diameter × height) (m × m)	0.30 × 0.20
F_{AB}-specific affinity column	
Column dimensions (diameter × height) (m × m) Voidage fraction (ε) (–)	0.27 × 0.34 0.4

Table 36: Assumed data values for variables which were fixed during the GSA analysis

6.4.2 Modelling assumptions

- The purification process was simulated for a single batch of feed
- Fugitive losses between steps in the process were assumed to be negligible
- Before any of the purification cycles, all columns were assumed to be free of any adsorbate

- There was assumed to be no loss in matrix capacity between chromatography cycles
- As with Chapter 5, papain digestion conditions were chosen to achieve near-complete IgG digestion
- Sufficient resources were assumed to be available in the facility to allow all processing steps to occur sequentially without any delay or holding time between them – hence times used in the throughput calculation were only affected by the durations of running each unit operation

6.4.3 GSA methodology

6.4.3.1 Fourier Amplitude Sensitivity Test

Although the Sobol' method implemented in the previous investigation was adequate for systems of up to four variables, it was found that the time taken by the MATLAB® code to determine the SI terms when more than five inputs were involved was excessive (in the order of several hours). In order to model these equations more efficiently, an alternative GSA method called Fourier Amplitude Sensitivity Test (FAST) was used (Chan et al., 1997; Saltelli et al., 2000). As described later, FAST enables more rapid calculation of sensitivity indices than the Sobol' method.

Fourier Amplitude Sensitivity Test (FAST) involves initially converting the *n*-dimensional integral of the model function $f(\boldsymbol{x})$ into a single dimension $f(s)$, through use of a transformation function (G_i) and a set of integer frequencies (ω_i). The resulting function is then Fourier analysed to determine the partial variance due to individual inputs and the total model variance. One transformation function and one frequency are assigned to each input variable *i* for which an SI needs to be determined. The general form of the *G* function is given in Equation (74):

$$x_i = G_i(\sin(\omega_i s)); i = 1...n \tag{74}$$

s is a scalar parameter ranging from $-\pi$ to π and x_i ranges from 0 to 1, with $f(s)$ as the resulting transformed function (see Equation (79) for details):

$$f(s) = f(\bar{x}) = f(G_1(sin(\omega_1 s)), G_2(sin(\omega_2 s)), G_3(sin(\omega_3 s)), \ldots, G_n(sin(\omega_n s))) \tag{75}$$

As with the previous sections, although the inputs to x_i in FAST range from zero to one, these need to be scaled to the real input values. The total model variance of $f(s)$, D_{FAST}, can then be calculated using Equations (76) – (78):

$$D_{FAST} = 2\sum_{j=1}^{P\omega_{max}} A_j^{\,2} + B_j^{\,2} \tag{76}$$

$$A_j = \frac{1}{2\pi}\int_{\pi}^{\pi} f(s)\cos(js)ds \quad (77)$$

$$B_j = \frac{1}{2\pi}\int_{\pi}^{\pi} f(s)\sin(js)ds \quad (78)$$

where A_j and B_j are the j^{th} sine and cosine Fourier coefficients respectively, P is the maximum harmonic considered and ω_{max} is the highest value in the set of angular frequencies used. P was set to 4, because any additional contributions made towards D_{FAST} by terms computed at higher harmonics are insignificant (Saltelli et al., 2000).

SI calculation requires definition of the transformation function, *G*, as well as the frequency set: several different *G* functions have been proposed in the past (e.g. Cukier et al., 1973; Koda et al., 1979). This chapter uses Equation (79) proposed by Saltelli et al. (2000), because out of all transformations, it provides the most uniform coverage of G–space. The transform relies on the periodicity of the inverse sine function when implemented in programming languages.

$$x_i = \frac{1}{2} + \frac{1}{\pi}\sin^{-1}(\sin(\omega_i s)) \; ; -\pi < s < \pi \quad (79)$$

In theory, values of the different frequencies ω_i should be incommensurate i.e. the values should be chosen such that none of them can be formed by a linear combination of any of the other frequency values using integer coefficients, because otherwise ω_i will interfere and lead to inaccuracies. In practice, incommensurate sets are impossible and instead, frequencies are chosen judiciously such that they are free of interference to a certain number of harmonics (the value stipulated by P), above which further contributions to the variance become negligible. Frequency sets used in this work were taken from literature (Schaibly and Shuler, 1973).

The contribution that the i^{th} variable makes to the variance of *f(s)* can then be determined by evaluating A_j and B_j for the fundamental frequency ω_i and its higher harmonics, up to the value of P – namely $2\omega_i$, $3\omega_i$ and $4\omega_i$, by Equation (80):

$$D_{\omega_i} = 2\sum_{j=1}^{P} A_{j\omega_i}^2 + B_{j\omega_i}^2 \quad (80)$$

The SI value for the i^{th} input variable is then calculated by Equation (81):

$$SI_i = \frac{D_{\omega_i}}{D_{FAST}} \; ; 0 \leq SI_i \leq 1 \quad (81)$$

The values of the first order SI terms calculated by the Sobol' and FAST methods in Equations (64) and (81) respectively are equivalent (Saltelli and Bolado, 1998). FAST is far more computationally efficient than Sobol' when calculating the first order indices, because the partial variances can be determined directly from the same calculations used to evaluate the total model variance (D_{FAST}). Unlike Sobol', however, FAST does not permit straightforward calculation of the interaction indices (Saltelli and Bolado, 1998), which is necessary for studying the significant interactions that often occur between unit operations in bioprocess flowsheets (Groep et al., 2000). Hence for the research presented in this part of the chapter, both FAST and Sobol' methods are employed – the first order indices are calculated by FAST and their sum is determined to see whether it exceeds a pre-specified threshold. If this is the case, no further calculations are undertaken, but otherwise the Sobol' method is used to evaluate the interaction indices. This way, the time consuming Sobol' calculation of interactions is only undertaken if necessary. Nevertheless, any calculation of interaction terms by the previously implemented Sobol' method would still be too time consuming and hence in an attempt to accelerate the calculation, Monte Carlo integration was used instead.

6.4.3.2 Monte Carlo integration

Monte Carlo methods rely on evaluating the area of the bounded region by picking points at random from the input space and calculating the number that lie above and below the function $f(x)$. Dividing the number of points below the function by the total number of points selected and then multiplying the result by the volume of the input space provides an approximation for the value of the integral. For example, for the function $f(x) = x^2$, the following diagram may result:

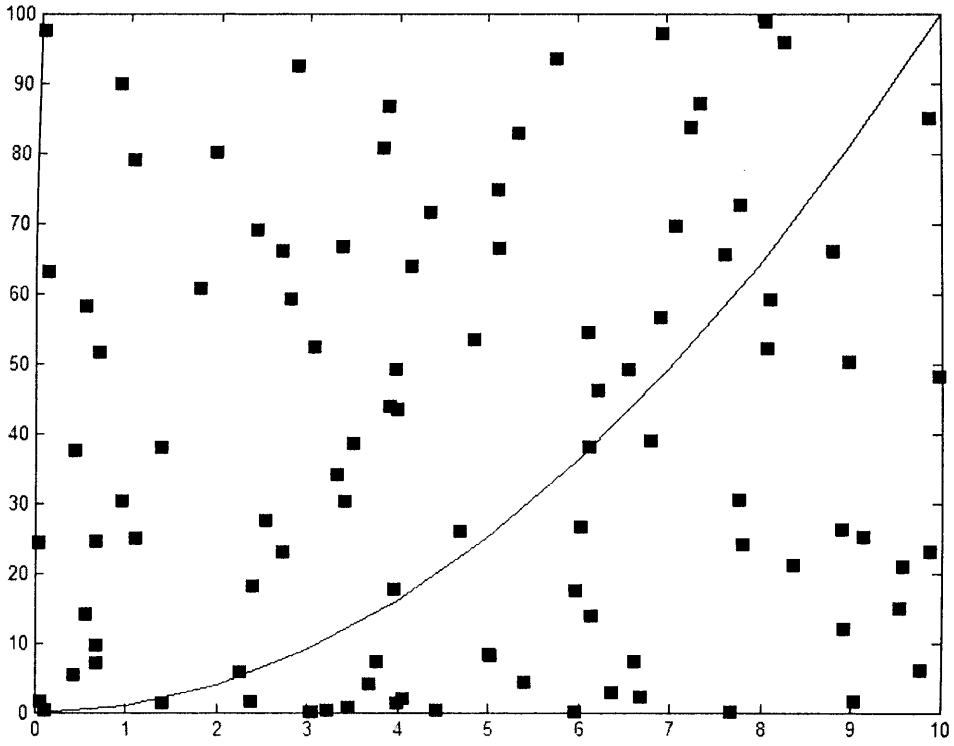


Figure 58: A sample result from a Monte Carlo integration of the function $y = x^2$
The solid line is the analytical function whilst the 100 points are random samples of the input space (graph produced using MATLAB®)

In total, there are 100 points, of which 39 fall below the line. Hence the integral can be calculated as (39 ÷ 100) × (100 – 0) × (10 – 0) = 390. The actual value of the integral can be determined analytically:

$$\int_0^{10} x^2 dx = \left[\frac{x^3}{3}\right]_0^{10}$$
$$= \frac{1000}{3}$$
$$= 333\frac{1}{3}$$

In general, the accuracy increases as the number of points that is selected rises. The theory is equally applicable to multidimensional cases, again by picking points at random and determining whether they lie above or below the hyper-dimensional surface. Sobol' (1993) describes a Monte Carlo method by which the values of f_0, D_T and D_i can be evaluated by calculation of $f(x)$ alone rather than the time consuming process of calculating every summand in Equation (54). The independent variables are partitioned into two groups denoted by, 'y,' and, 'z,' resulting in the following:

$$f(x) = f_0 + f_1(y) + f_2(z) + f_{12}(y,z) \tag{82}$$

The terms on the right hand side of Equation (82) are defined as follows:

$$f_1(y) = \int_0^1 \ldots \int_0^1 f(x)dz - f_0 \tag{83}$$

$$f_2(z) = \int_0^1 \ldots \int_0^1 f(x)dy - f_0 \tag{84}$$

$$f_{12}(y,z) = f(x) - \left(f_0 + f_1(y) + f_2(z)\right) \tag{85}$$

The variances of these functions can then be expressed as the following integrals:

$$D = \int_n f^2(x)dx - f_0^2 \qquad (\textit{Total model variance}) \tag{86}$$

$$D_1 = \int_0^1 \ldots \int_0^1 f_1^2(y)dy \qquad (\textit{Total variance due to subset 1}) \tag{87}$$

$$D_2 = \int_0^1 \ldots \int_0^1 f_2^2(z)dz \qquad (\textit{Total variance due to subset 2}) \tag{88}$$

The integrals in Equations (86) to (88) are then calculated by Monte Carlo integration, selecting the values of y and z uniformly from [0,1]:

$$f_0 = \frac{1}{N}\sum_{j=1}^{N} f(y,z) \tag{89}$$

$$D + f_0^2 = \frac{1}{N}\sum_{j=1}^{N} f^2(y,z) \tag{90}$$

$$D_1 + f_0^2 = \frac{1}{N}\sum_{j=1}^{N} f(y,z)\cdot f(y,v) \tag{91}$$

The *v* subset of inputs represents the same group of variables as z, but with different random realisations within [0,1]. For example, to calculate the interaction variance between the first and second inputs in a seven variable system, the 'y' input set is used to represent variables 1 and 2, whilst the z and v sets are used to represent variables 3-7; Equation (90) is then used to determine the total model variance, while Equation (91) calculates the total variance due to the first and second inputs. The ratio of D_1/D is defined by the following equation:

$$\frac{D_1}{D} = S_{T12} = S_1 + S_2 + S_{12} \tag{92}$$

S_{T12} represents the contribution made towards the total model variance by the first and second variables, in both a first order *and* an interactive manner. Thus to calculate the interaction SI term (S_{12}), initially S_{T12} is determined by the Monte Carlo Sobol' method, after which the lower order terms calculated by FAST (S_1 and S_2) are subtracted to compute S_{12}. This procedure is continued for all other second order terms (e.g. S_{13}, S_{14}, S_{25}, S_{67}...) and higher orders (e.g. S_{236}, S_{1457}, S_{23567}...).

Use of Equations (89) to (91) also requires a sufficient number of random samples of the input space to be taken to ensure that the values of f_0, D and D_1 converge. Although the maximum number of iterations set in the code was N = 1,000,000 (a high value set for its strong likelihood of achieving convergence), it was anticipated that convergence might potentially occur earlier and hence the following three parameters were also defined:

- Lower limit = 50000
- Range = 1000
- Variation = 5%

The absolute minimum number of points calculated was 50,000, after which the following test was applied: if the difference between the averaged variance calculated over the previous 1,000 iterations was less than 5%, then adequate convergence had occurred. Values of the Lower limit, Range and Variation were chosen after experimentation with the program showed that these consistently achieved convergence. To provide a further check on the convergence, the variance for

each sensitivity index was plotted as a function of *N*. For the industrial case study, these graphs showed that stability was achieved consistently when calculating the interaction terms.

This software was developed in MATLAB® and the code has been placed in the appendix. As with the previous study, input values were rescaled to match the zero to one input scale of the GSA methods, as per Equation (68) on Page 137.

6.4.4 Base case and process change strategies

6.4.4.1 Base case

Initially, the values for parameters given in Table 35 and Table 36 were used to populate the model and create a base case set of sensitivity indices, after which the impacts of making a series of process changes of industrial relevance were evaluated.

6.4.4.2 Increase the synthetic affinity ligand adsorbent capacity

Several different matrices are available commercially which provide higher capacities than the 0.5-0.7 mol/m^3 range used in the base case (Chhatre et al., 2007f). The impacts of using matrices with capacities of 0.7-0.9 and 0.9-1.1 mol/m^3 were investigated whilst keeping all other parameters unchanged.

6.4.4.3 Pre-concentrate the feed

The base case range of antibody feed concentration was 0.15–0.20 mol/m^3 (Table 35). Pre-concentrating the feed at source in Australia prior to its shipment to the U.K. could potentially reduce the costs involved in transporting ovine serum to the manufacturing facility. For the purposes of this study, it was assumed that the feed concentration could be increased up to three times it base case value and the impact of doubling and tripling the IgG concentration about the mid point of its base case range to 0.325–0.375, 0.5–0.55 mol/m^3 whilst leaving the feed volume unchanged was examined. This study also examined the combined effect of increasing both feed concentration and synthetic affinity ligand adsorbent capacity upon the sensitivity indices.

The model was also used to assess the impact of greater variability in the IgG feed titre between batches. Although this normally lies in the 0.15 – 0.20 mol/m^3 range, it is possible for it to fluctuate more significantly to between 0.13 and 0.27 mol/m^3 as a result of natural variability between high and low responding sheep (Newcombe et al., 2006). Hence this study examined the effect of widening the range of IgG feed titre upon the sensitivity indices whilst leaving all other variables unchanged.

6.4.4.4 Vary the synthetic affinity ligand adsorbent flowrate

Varying the flowrate will have conflicting impacts on yield and throughput – a reduction will result in more IgG binding to the column at the expense of extended process times, whilst increasing it will reduce yield but also process the feed more rapidly. In this situation, a trade-off will be required and establishing sensitivities will help to decide which variable(s) exert the greatest influence.

6.4.4.5 Vary the capacity and volume of the CroFab™ affinity column

The CroFab™ column uses an expensive custom-made affinity resin and a reduction in its volume could make significant cost savings to the process, although this needs to be offset against the potential capacity loss observed at commercial scale. Hence, the effects of simultaneously varying the capacity and volume of the column upon the sensitivity indices were determined.

6.4.5 Results and Discussion

6.4.5.1 Base case

The base case results in Figure 59 shows that product yield is most strongly controlled by the flowrate used in the synthetic affinity ligand step (SI = 0.57), followed by the capacity of this adsorbent and the IgG feed concentration at SI = 0.24 and 0.17 respectively. None of the other inputs were significant (SI values $<7.6\times10^{-4}$).

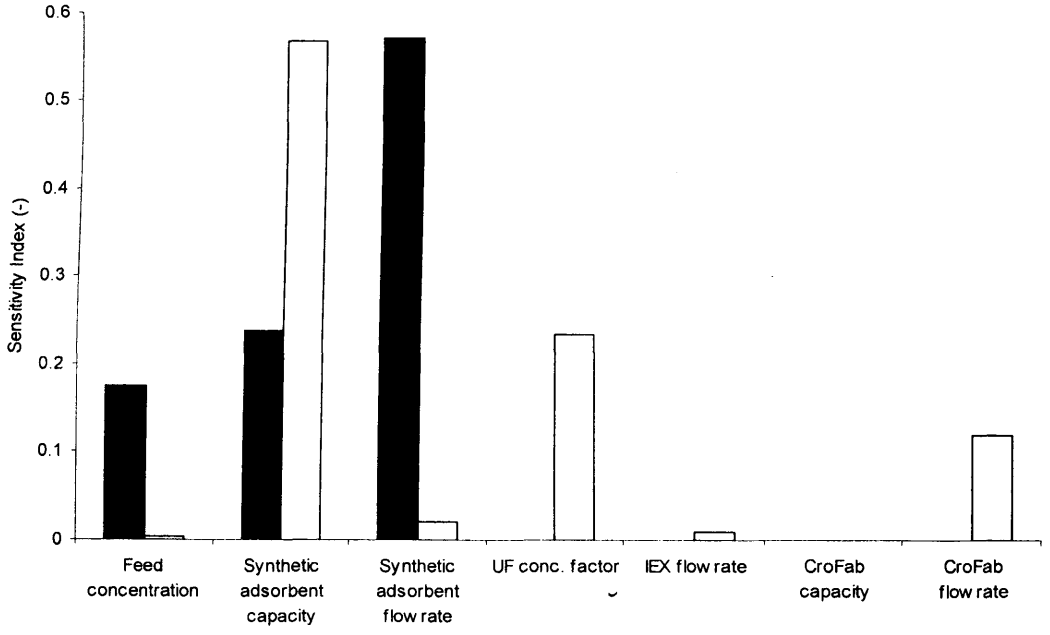


Figure 59: Base case sensitivity indices
(■ = yield; □ = throughput)

Conversely, throughput is primarily controlled by the capacity of the synthetic ligand adsorbent step (SI = 0.57), followed by the ultrafiltration rig concentration factor (SI = 0.23) and then the CroFab™ affinity column flowrate (SI = 0.12). Overall, the synthetic adsorbent column is the single most important unit operation controlling process performance. It should also be noted that under base

case conditions, neither yield nor throughput exhibited any significant interactions between the input variables, suggesting that the process could be optimised by considering each variable independently.

To verify the accuracy of the information provided by the sensitivity indices, the yield output was chosen as an example and the model was used to identify how F_{AB} recovery varied across the full range of the inputs given in Table 35. The results in Figure 60 plot the change in yield relative to that obtained at the lowest input value for each variable. This demonstrates that the SI values correctly predicted the relative importance of the variables, with the synthetic column flowrate shown to have the most significant impact on yield, whilst the synthetic affinity capacity and feed concentration exert smaller but still notable effects. None of the other inputs resulted in any measurable change in yield, reflecting their negligible SI values.

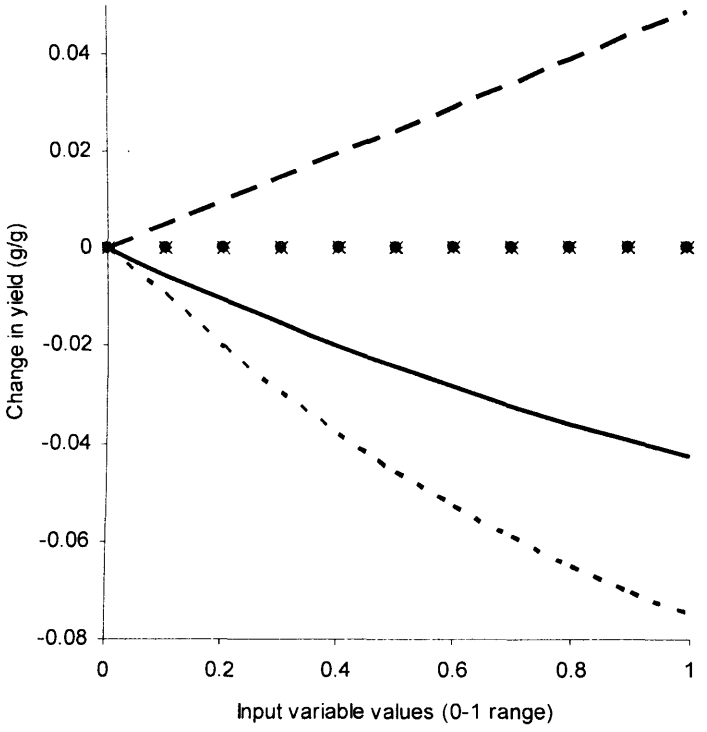


Figure 60: Verification of the base case yield sensitivity indices
This shows the change in yield over the ranges of the inputs relative to the yield obtained at the lowest input value for each variable: solid line = feed concentration; long dashed = synthetic adsorbent capacity; short dashed = synthetic adsorbent flowrate. Other inputs are represented by the symbols on the zero change in yield)

6.4.5.2 Increase the synthetic affinity ligand adsorbent capacity

When the use of improved adsorbent capacities was simulated, the synthetic adsorbent flowrate remained the single most important input influencing yield, but its importance reduced significantly (Figure 61), along with those for the feed concentration and synthetic matrix capacity. None of the SI terms for the other variables changed, but the interactions rose as capacity increased, with significant interaction indices observed between feed concentration and synthetic adsorbent flowrate

as well as synthetic adsorbent capacity and flowrate. This was especially true at the 0.9 – 1.1 mol/m^3 binding capacity (Table 37).

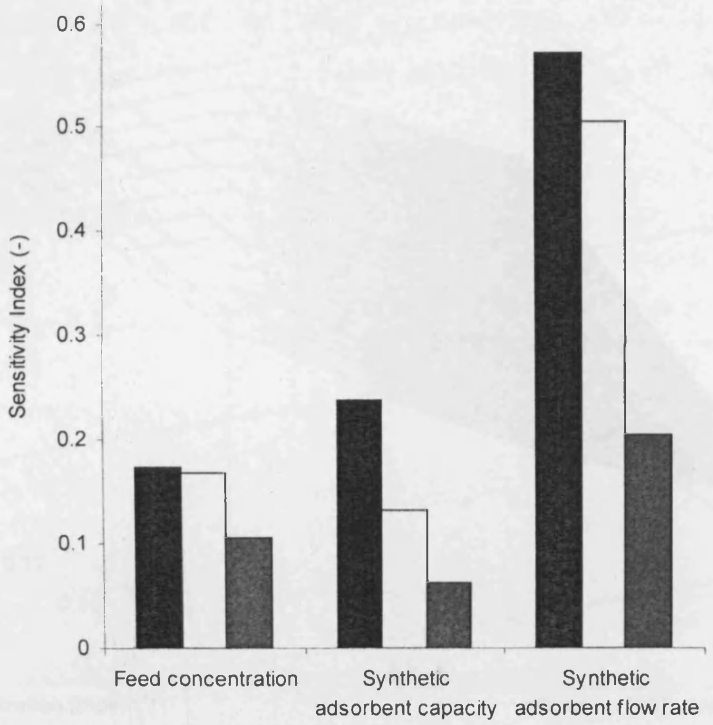


Figure 61: Impact of increasing the synthetic adsorbent capacity upon the yield sensitivity indices ***(Black = 0.5 - 0.7; white = 0.7 - 0.9; grey = 0.9 - 1.1 mol/m^3)***

	Synthetic adsorbent capacity (mol/m^3)	
Variables	*0.7 – 0.9*	*0.9 – 1.1*
Feed concentration/synthetic adsorbent flowrate	0.03	0.13
Synthetic adsorbent capacity/flowrate	0.03	0.09

Table 37: Interaction sensitivity indices for yield resulting from increasing the synthetic adsorbent capacity

There were also a number of notable interactions between synthetic adsorbent flowrate and the other four inputs (UF concentration factor, IEX flowrate, CroFab™ capacity and CroFab™ column flowrate) with SI values of around 0.04 to 0.05 each at 0.9 – 1.1 mol/m^3. This suggests that using a higher capacity synthetic adsorbent complicates process optimisation efforts by making yield more strongly dependent upon combinations of inputs. The nature of these interactions was explored further by using the feed concentration and synthetic adsorbent flowrate variables as an example. Figure 62A simulates the yield results obtained from the 0.7-0.9 mol/m^3 matrix capacity and shows that increasing the flowrate at a constant IgG concentration of 0.15 mol/m^3 results in a moderate reduction in yield – a drop which becomes more pronounced as feed concentration rises simultaneously to 0.2 mol/m^3. Hence the synthetic adsorbent flowrate is still a controlling variable in its own right (reflected by its first order sensitivity index of 0.50 – see Figure 61), but its interaction

with feed concentration also becomes important towards the upper end of the IgG titre range (interaction SI = 0.03).

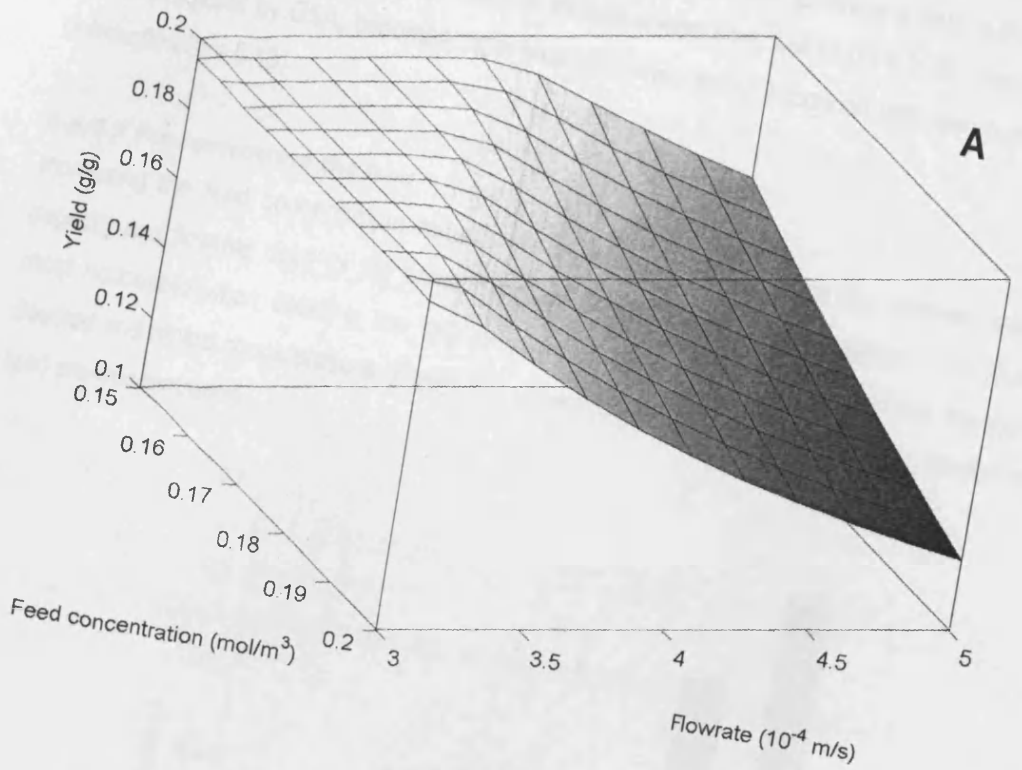


Figure 62: Impact of varying the feed concentration and synthetic adsorbent flowrate upon yield *(A = 0.7 - 0.9 mol/m3; B = 0.9 - 1.1 mol/m3 synthetic adsorbent capacity)*

Conversely, when employing the 0.9-1.1 mol/m^3 matrix (Figure 62B), yield remains virtually unchanged across the flowrate range when using a 0.15 mol/m^3 feed concentration and it requires the simultaneous use of both a high flowrate *and* a high IgG titre to cause a drop in F_{AB} recovery. Flowrate now has a relatively small effect on its own compared to before (SI = 0.20 – see Figure 61) and as predicted by GSA, becomes more important when acting in concert with feed concentration (interaction SI = 0.13).

6.4.5.3 Pre-concentrate the feed

Increasing the feed concentration reduces its impact upon yield, whilst the synthetic adsorbent capacity and flowrate used for the primary capture of IgG become more important. The change is most noticeable when doubling the IgG titre, with a more modest effect observed between the doubled and tripled concentrations (Figure 63). The throughput sensitivities remained unaffected by feed pre-concentration.

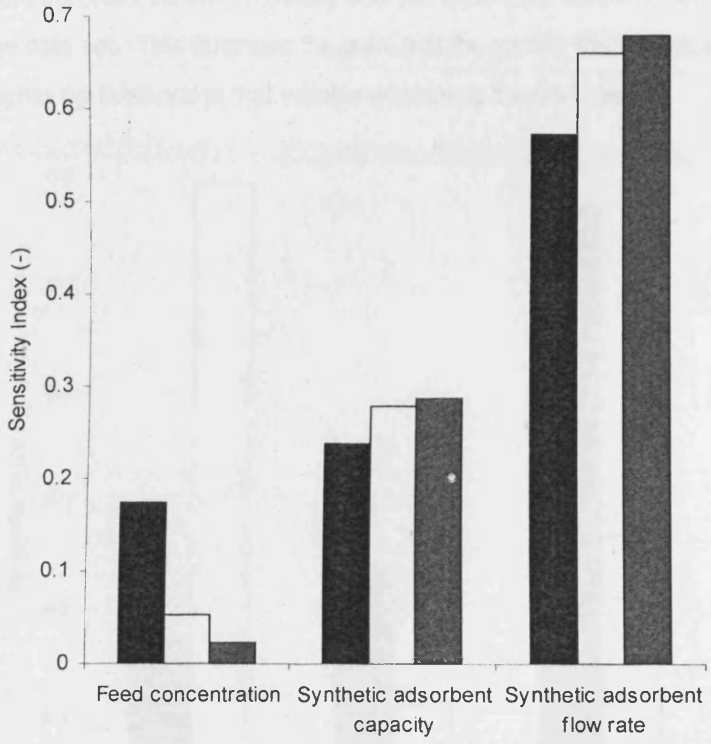


Figure 63: Impact of pre-concentrating the feed on the yield sensitivity indices
Black = 0.15-0.2; white = 0.325-0.375 (2 × base case); grey = 0.5-0.55 mol/m^3 (3 × base case)

Combining the use of higher feed concentrations with higher capacity synthetic adsorbent matrices resulted in no interactions between any of the variables for either yield or throughput. For yield, increasing titre and capacity together most notably caused the synthetic adsorbent flowrate to become far more important (SI = 0.83 at 0.9-1.1 mol/m^3 matrix capacity and a tripled concentration, compared to 0.57 in the base case). This contrasts with Section 6.4.5.2, where only increasing the capacity resulted in a reduction in the importance of flowrate and a growth in its interaction with feed

concentration. Hence the use of higher capacity synthetic adsorbent matrices in conjunction with high titre feedstocks enable the process to be optimised by considering each input variable independently, with the major focus on the synthetic adsorbent flowrate. For throughput, using higher capacities and feed concentrations together caused a reduction in the importance of the synthetic adsorbent capacity compared to base case (where SI was 0.57), although this still remained a more important variable than if capacity was increased without feed pre-concentration (e.g. at the highest capacity and titre, the synthetic adsorbent capacity SI was 0.39, as compared to 0.001 at the highest capacity and base case titre). The ultrafiltration concentration factor and CroFab™ column flowrate became more important factors when using higher feed concentrations and capacities.

Finally, the effects of incorporating greater uncertainty into the IgG titre were assessed and the results for the yield SI values are shown in Figure 64. Higher levels of uncertainty in the feed titre make this a far more important variable, resulting from the wider input concentration range causing a wider output range data set. This illustrates the point that the greater the biological variability for a given input, the higher the likelihood of that variable dominating the GSA results.

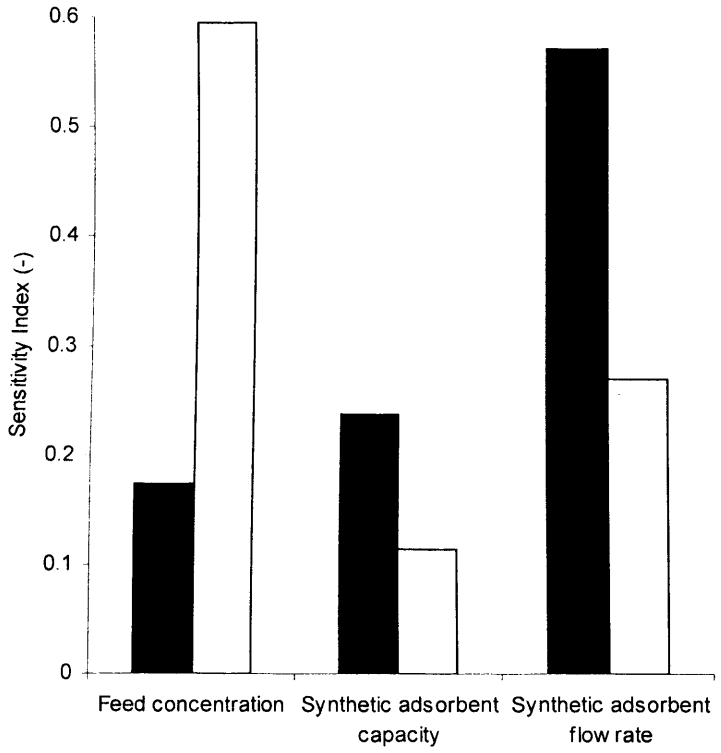


Figure 64: Impact of widening the feed concentration range on the yield SI terms
(■ = base case; □ = widened feed titre)

6.4.5.4 Vary the synthetic affinity column flowrate

Figure 65 plots the impact of varying the synthetic adsorbent column flowrate upon the yield sensitivity indices. Flowrate remains a reasonably significant factor over the full range of values

tested, although its importance peaks at the base case range (3-5 × 10^{-4} m/s). It is the dominant variable up to 5-7 × 10^{-4} m/s, after which synthetic adsorbent capacity becomes more important. The process displays few interactions except at the lowest flowrate range (1-3 × 10^{-4} m/s), where the sum of the first order indices is only equal to 0.36 and feed concentration, synthetic adsorbent capacity and flowrate interact strongly. The flowrate also displays modest second order interactions of around 0.05 with the remaining input variables at the 1-3 × 10^{-4} m/s flowrate range (Table 38). Hence although lowering the flowrate is likely to increase the product yield, this again results in the need to optimise several variables simultaneously in order to achieve maximal product recovery. Conversely, there are no interactions between variables for throughput at any flowrate range and furthermore, Figure 66 shows that the synthetic adsorbent flowrate is only significant at the lowest flowrate range – above this, its importance diminishes rapidly, whilst the synthetic adsorbent capacity and the UF concentration factor become the key variables. Hence, this analysis suggests that if optimisation was to be undertaken on the basis of throughput alone, the impact of flowrate can be ignored above 3 × 10^{-4} m/s, whereas if yield is the more important output, the flowrate needs to be selected carefully whatever its operational range. Thus, for example, if throughput is the key output and operation within 1 – 3 × 10^{-4} m/s is planned, then it is especially important to check that the correct flowrate within that range has been chosen, whereas if operation within 5 – 7 × 10^{-4} m/s is planned, optimising the flowrate within this range will make little difference.

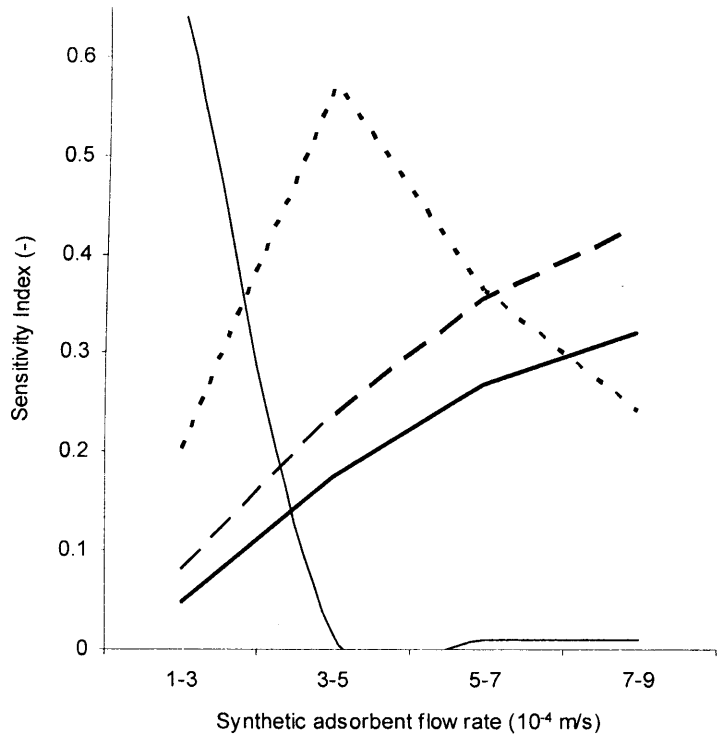


Figure 65: Impact of varying the synthetic adsorbent flowrate upon the yield sensitivity indices
(Thick solid line = feed concentration; long dashed = synthetic adsorbent capacity; short dashed = synthetic adsorbent flowrate; thin solid line = sum of all interactions)

Variables	Interaction sensitivity index
Feed concentration/ synthetic adsorbent flowrate	0.09
Synthetic adsorbent capacity/flowrate	0.13
Synthetic adsorbent flowrate and the other four inputs	~0.05 for all

Table 38: Yield interactions seen when reducing the synthetic adsorbent flowrate range to 1-3 × 10^{-4} m/s

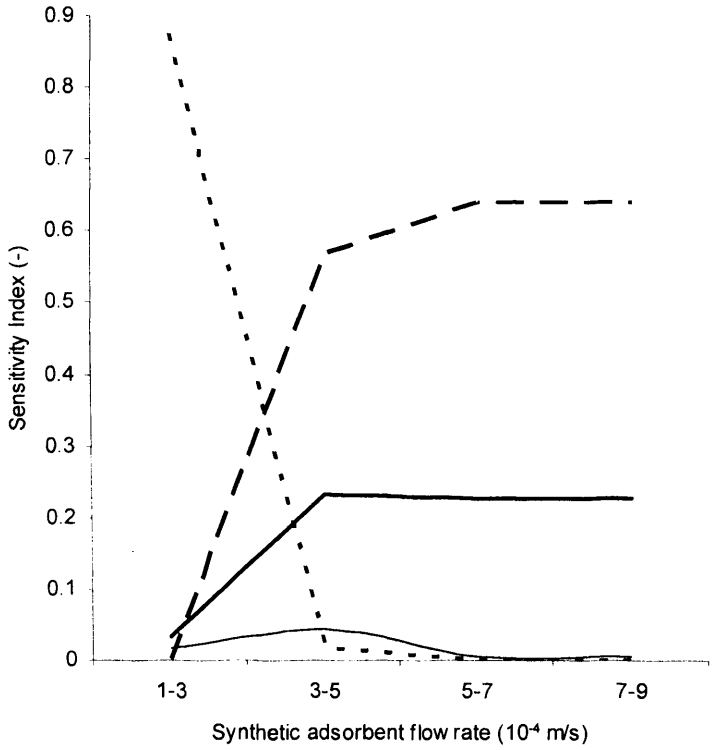


Figure 66: Impact of varying the synthetic affinity flowrate upon the throughput sensitivity indices
(Thick solid line = UF concentration factor; long dashed = synthetic adsorbent capacity; short dashed = synthetic adsorbent flowrate; thin solid line = sum of all interactions)

6.4.5.5 Vary the capacity and volume of the CroFab™ affinity column

Although previous studies indicated the importance of the synthetic adsorbent column to both yield and throughput, it was hypothesised that if sufficiently small column capacities and bed volumes were used for the CroFab™ affinity step, this unit operation might become a more important factor in the process. Hence GSA was used to determine the SI values at various combinations of CroFab™ column volumes (-90 to +30% around the base case) and capacity ranges (from the base case value down to around 30% of that value) whilst leaving all other variables unchanged. The results were used to identify the column volume at each capacity where the CroFab™ affinity column became more important than the synthetic adsorbent step. The results for both yield and throughput in Figure 67 demonstrate that using small volume CroFab™ columns that have lost a significant amount of capacity leads to that step being a greater determinant of process performance, whereas if the capacity or bed volume is increased, this increases the potential for product capture in the CroFab™ affinity step and means that yield and throughput switch back to being predominantly

controlled by the synthetic affinity column. Consequently, this plot ultimately facilitates identification of the most suitable basis for process optimisation.

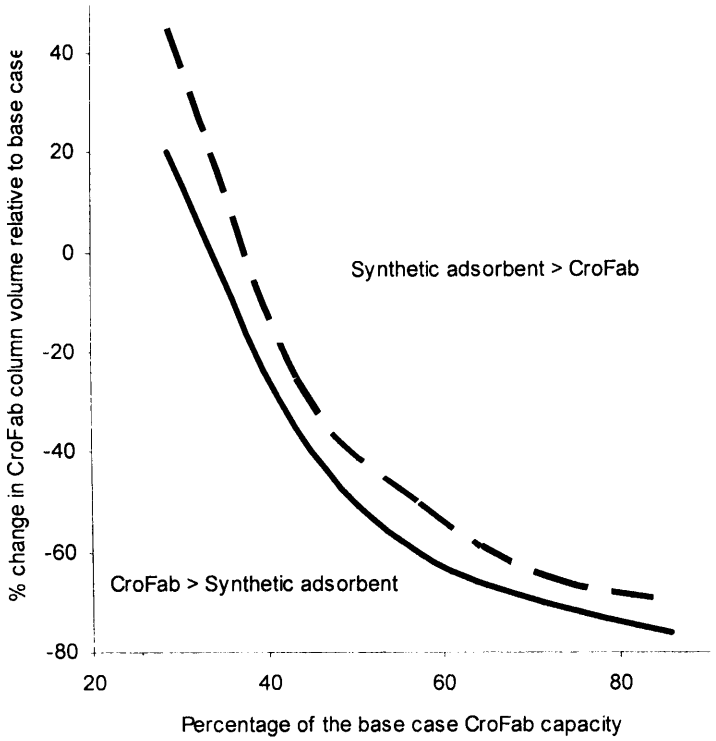


Figure 67: Plot indicating where the synthetic adsorbent column is more important ***(Indicated by the '>' symbol) than the CroFab™ affinity step and vice versa (solid line = yield; dashed line = throughput). Axes plot percentage change in CroFab™ column capacity and volume relative to base case conditions***

6.4.6 Whole process case study summary

This section has described the application of Global Sensitivity Analysis in identifying the most important attributes influencing the performance of industrial bio-manufacturing processes and the interactions within and between unit operations. Results produced from modelling the CroFab process showed that the synthetic ligand adsorbent step was a key factor in determining both final product yield and throughput. Data from the study also showed that process interactions only occurred in the yield output and were predominantly restricted to the feed concentration, synthetic adsorbent capacity and flowrate.

6.5 Chapter conclusions

This chapter has presented the first application of Global Sensitivity Analysis (GSA) to the determination of parameter importance in a synthetic Protein A column and also to an entire production process. The feed concentration, synthetic adsorbent capacity and flowrate were identified as being critical parameters for optimisation in CroFab™ manufacture – a fact which is in line with the expectation expressed by Protherics that this step is likely to be the most critical one for

controlling product recovery and throughput. Other steps are already operated in an optimal fashion e.g. the conditions for achieving complete Fab recovery in papain digestion are known and are used at commercial scale already, whilst the IEX step is deliberately underloaded to achieve 95% Fab recovery and complete elimination of impurities. The operation of the final venom-specific affinity step also involves deliberately underloading it to avoid capacity loss and thus achieve consistently high recoveries across multiple cycles. Conversely, Protherics has spent a great deal of time optimising the operation of the synthetic adsorbent step in an attempt to achieve high binding capacities and thus IgG recoveries from the serum. Hence the GSA prediction that this is likely to be a key step fits in with what is anticipated by the company.

From a more generic perspective, GSA can also be applied to any other production process and the resulting sensitivity data can allow a research and development group to focus optimisation efforts on the most significant controlling factors. Hence the use of GSA as described in this chapter promotes the efficient design and development of manufacturing operations.

Thus far, the thesis has focused upon whole process modelling and the evaluation of strategies to achieve improvements in product yield, cost and time. The next chapter instead concentrates ostensibly upon one unit operation in particular – chromatography – and describes methods developed for assessing the performance of column steps used to purify antibody products.

7: DEVELOPMENT OF METHODS FOR THE INDUSTRIAL-SCALE CHROMATOGRAPHIC PURIFICATION OF ANTIBODIES

7.1 Introduction

Although this thesis was primarily aimed at developing simulations for investigating whether specific alterations to a sequence of process steps was feasible, the full remit of the research involved looking at a range of different methods (both *in silico* as well as experimental) in order to evaluate the impacts of process change. Since a large percentage of the development work carried out by Protherics focuses upon optimising expensive chromatographic purifications, this chapter concentrates upon this operation and presents a combination of simulation and experimental methods for evaluating the performance of column steps for antibodies and their fragments. Two studies are presented below:

1) Development of decision-support software for the chromatographic purification of antibody fragments. The method is tested by its application to an affinity column operated by Protherics for the manufacture of polyclonal F_{AB} used to treat digoxin toxicity (product name DigiFab™)
2) Evaluation of a novel agarose-based synthetic ligand adsorbent for antibody recovery from ovine serum, using the CroFab™ feedstock to test the matrix

7.2 Software for the industrial-scale chromatographic purification of antibodies

7.2.1 Introduction

As outlined in previous chapters, polyclonal antibodies and their fragments purified from antisera have been used for many years to treat the harmful effects of many acute clinical indications and as a consequence, have resulted in significant financial gains for companies that manufacture these therapeutics. One of the major challenges in purifying such feedstocks lies in separating the product-specific IgGs from the remainder of the contaminating antibody repertoire and chromatographic purifications using highly specific, custom-made affinity matrices are commonly employed for this purpose (Huse et al., 2002). The high cost of such supports necessitates efficient design in order to maximise the economic value of the process and thus techniques which identify the most suitable operating conditions for achieving desired levels of manufacturing performance offer significant value. Many factors influence the design of such steps, including the likely protein loading range, matrix capacity loss over repeated loading cycles and flowrate. The values of these variables must be chosen carefully so as to achieve yields and throughputs that will satisfy market demand and ensure a financially viable process. The potential complexity of this task has driven the development of methods which identify the optimal operating location. Joseph et al. (2006), for

example, developed a computer-based design tool for the initial sizing of columns and selection of flowrates for achieving threshold levels of throughput and cost productivity. Many of these computer simulations have focused predominantly upon predicting the technical performance of the purification of highly clarified solutions consisting of either one or two components alone. Such models fail to represent adequately the challenge involved in separating product species from concentrated and highly complex protein feedstocks (Bak et al., 2007).

This chapter instead describes a simulation which seeks to optimise the operating conditions of pre-existing industrial-scale columns whilst trading off potentially conflicting requirements for yield and throughput. The method accounts for the impact of mass of protein loaded on to the column upon dynamic binding capacities over repeated loading cycles in order to predict breakthrough curves at a range of loading flowrates. The method is illustrated by application to the affinity purification step used in the manufacture of DigiFab™, another FDA-approved polyclonal antibody fragment purified from ovine serum by Protherics for the treatment of digoxin toxicity. The model was populated with data obtained from scale-down experimental studies of the commercial-scale affinity purification step, which correlated measured changes in matrix capacity with the total protein load and number of resin re-uses. To enable a trade-off between yield and throughput, output values were integrated together using multi-attribute decision making techniques in order to identify the most suitable flowrate and feed concentration required for achieving target levels of DigiFab™ yield and throughput. This work has been published (Chhatre et al., 2007g).

7.2.2 Section structure

The remainder of this section of the chapter is structured as follows: the DigiFab™ affinity purification process is described, along with the ultra scale-down experimentation undertaken in relation to this step. The mathematical description of this step is then provided, outlining how the simulated outputs were used to evaluate yield and throughput and how these were integrated together by multi-attribute-decision-making to simplify their evaluation. Finally, a discussion of the results concludes the work.

7.2.3 Methodology

For the DigiFab affinity step, breakthrough curves were modelled by the Thomas model presented in earlier chapters. The data used to populate the model was obtained from the methods described in Section 2.9. As with previous work, it was assumed that complete elution from the column could be achieved and that product yield (γ) could therefore be calculated by direct integration of the breakthrough curve. Throughput (τ_{column}) was determined by dividing the product mass recovered from the column by the total duration for which the column was operated (θ_{total}), accounting for the loading time (calculated according to the feed volume and operating flowrate), as well as the elution, equilibration and sanitisation steps:

$$\theta_{total} = t_{equilibrate} + t_{load} + t_{wash} + t_{elute} + t_{sanitise} \quad (93)$$

$$\tau_{column} = \frac{\gamma \cdot M_{feed}}{\theta_{total}} \quad (94)$$

The main design problem addressed by the model was to examine the impact of varying the loading flowrate and level of protein loading in order to identify the optimal combinations of those input variables which meet industrial goals for product yield and throughput. The model uses the ultra scale-down experimental data to relate product yield to the number of operational cycles and the mass of protein loaded on to the column per cycle. For each of these yield values, the model iterates upon the matrix capacity to locate values which satisfy the yield requirement. No other factors are altered in the iteration procedure. The calculated capacity values are then fixed for a given combination of protein load and number of matrix re-uses when conducting any further investigations.

In order to simplify the process of evaluating the effects of changing flowrates and protein loadings, model outputs were combined together into a single value by multi-attribute decision-making in a similar fashion to earlier chapters. Initially, the yield and throughput values were each normalised to a zero to one scale, which respectively represent the least and most attractive values:

$$N_{\gamma} = \frac{\gamma_A - \gamma_0}{\gamma_1 - \gamma_0} \quad (95)$$

$$N_{\tau} = \frac{\tau_A - \tau_0}{\tau_1 - \tau_0} \quad (96)$$

N_γ and N_τ are the normalised yield and throughput, γ_A and τ_A are the actual yield and throughput values calculated by the model, γ_0 and τ_0 are the lowest yield and throughput values out of all model runs and thus set to represent the zero bound and γ_1 and τ_1 are the highest values and set to represent the upper unity bound. Each normalised value was then further weighted by multiplying it by a value between zero and one (w) to place particular emphasis on the more important outputs Addition of the weighted (w) and normalised (N) values generated the Overall Rank (OR):

$$OR = w_{\gamma} N_{\gamma} + w_{\tau} N_{\tau} \quad (97)$$

As with earlier work, OR varies between zero and one, respectively representing the least and most attractive combined yields and throughputs.

7.2.4 Results and Discussion

7.2.4.1 Model inputs and outputs

Assumed input values to the DigiFab™ affinity model are given in Table 39.

Parameter	Assumed value
ε (–)	0.6
ε_i (–)	0.75
K_d (mol/m^3)	0.066
q_{ms} (mol/m^3)	1.3
v_{base} (m/s)	8.5×10^{-5}
L (m)	0.055
k_1 (m^3/mol/s)	3.12×10^{-4}

Table 39: Assumed inputs to the DigiFab™ model
ε and ε_i were chosen as typical values found in the literature. K_d was calculated from the batch binding data (inverse of the equilibrium adsorption constant). v_{base} = base case loading flowrate; other assumptions are explained below. Simulation results were generated by varying the flowrate (range = 0.9 v_{base} to 2 v_{base}).

The capacity of the settled affinity bed volume (q_{ms}) was determined by the method in section 7.2.3.2 to be 45 ± 15 mg/mL, corresponding to a maximum capacity of 60 mg/mL (1.3 mol/m3). The equilibrium desorption constant (K_d) required for the Thomas model was determined by calculating the inverse of K_a in Equation (1). The adsorption rate constant (k_1) given in Table 39 was calculated by keeping all other parameters fixed and manually iterating upon k_1 until the calculated yield reached 82% (y_{max}). This is the highest recovery possible when challenging a fresh batch of matrix with the lowest protein load tested by Protherics (100 mg protein/mL matrix) at the same loading flowrate of 30 cm/hr used in this study (Thillaivinayagalingam et al., 2007). The value of k_1 was then fixed for all further investigations. The yield at any given protein load was calculated as the average over five consecutive runs to determine overall recovery and throughput with a given batch of matrix. Five cycles were modelled as this is close to the seven for which a batch of matrix is used at large scale before it is replaced. The simulation results were visualised in Microsoft® Excel and MATLAB® to identify combinations of feed volume and flowrate which met preset criteria for product yield and throughput.

7.2.4.2 Results of scale down studies

Figure 68 plots the change in specific F_{AB} yield observed when the columns were challenged with different protein loads over five consecutive loading cycles. In general, yield dropped over the five cycles, such that the rate of yield loss increased with protein concentration. These data values were used as outlined in section 7.3.3 to determine the capacities of the columns which would achieve the corresponding yields; capacity values were then used by the model to determine the impact of flowrate upon yield and throughput as outlined below.

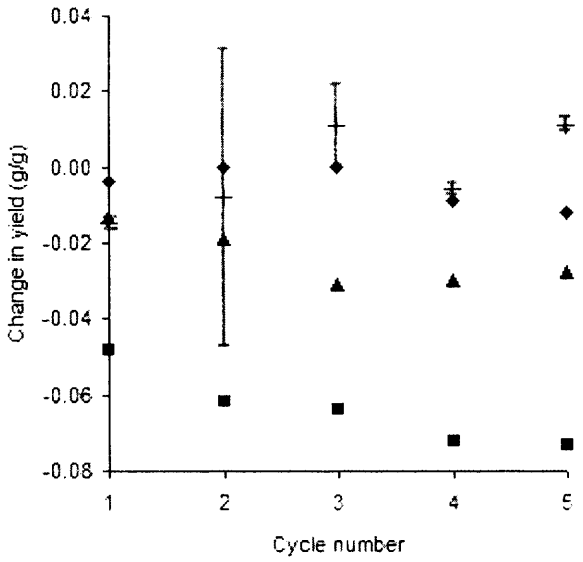


Figure 68: Experimental results obtained when challenging the affinity matrix with feed over five loading cycles ***The graph plots the change in yield relative to the base case 82% (y_{max}). Each line refers to one protein load: + = 174 mg/mL; ♦ = 260 mg/mL; ▲ = 347 mg/mL; ■ = 500 mg/mL***

7.2.4.3 Impact of flowrate upon yield

The flowrate used in the simulation was varied to between -90 to +100% around the base case value (v_{base}) given in Table 39 (equivalent to 30 cm/hr) to determine the impact upon average yield over five loading cycles (Figure 69). As expected, the yield declines at higher flowrates and protein loadings, as product flows through the bed without being captured by the resin. The yields achieved at 174 and 260 mg/mL loadings are very similar for all flowrates tested, while yields at 347 and 500 mg/mL are much lower. This result is consistent with earlier data (Thillaivinayagalingam et al., 2007) which showed that loading more than 300 mg/mL oversaturates the matrix and so leads to deterioration in column performance. The difference in yields becomes wider at higher flowrates e.g. the gap between the yields at 174 and 500 mg/mL increases from 3% to 30% across the full range of flowrates tested. It is also noticeable that the yield initially drops away more quickly at 500 mg/mL compared to the other three protein loads.

These data can be used to identify how operation needs to be changed dependent upon the protein load in order to achieve preset goals for yield. For example, in order to achieve y_{max} using a 500 mg/mL protein load, the flowrate has to be reduced by over 70% compared to v_{base}. Conversely, operating at the two lowest protein loads (174 and 260 mg/mL), the y_{max} is realised at a flowrate close to v_{base}, meaning there is far greater flexibility in selecting the flowrate when operating at the lower end of the feed loading range.

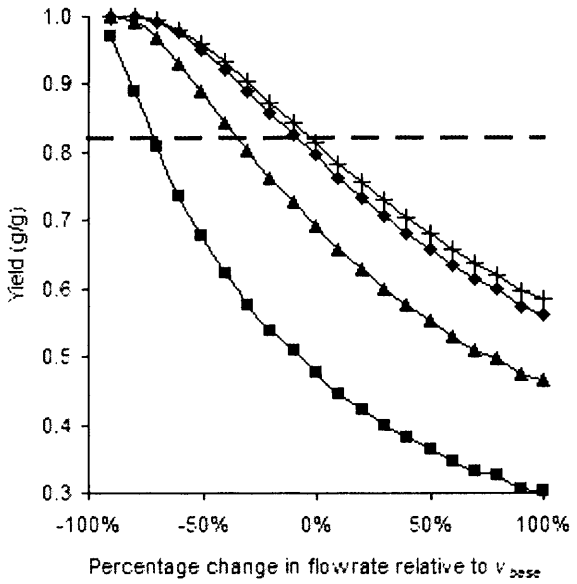


Figure 69: Impact of varying the simulated flowrate from $0.9v_{base}$ to $2v_{base}$ on average predicted yield over five runs ***The dotted line marks the maximum possible yield (y_{max}) obtained with a fresh batch of matrix. + = 174 mg/mL; ♦ = 260 mg/mL; ▲ = 347 mg/mL; ■ = 500 mg/mL***

7.2.4.4 Impact of flowrate upon throughput

As outlined in the previous chapter, reducing the flowrate will increase product yield, but potentially at the expense of extended processing times and so a lower throughput. Hence, the impact of simultaneously varying both the flowrate and the protein load upon product throughput over the five loading cycles was investigated (Figure 70). For all protein loading values, throughput initially increases sharply as a function of flowrate towards an optimum level the optimal flowrates lie at between 50 and 70% below v_{base}. Throughput then falls away more gently as flowrate rises to double that of the base case. In the -30 to +100% flowrate range, throughputs are at their highest when loading at 347 mg/mL. Conversely, the highest throughput observed from all simulation runs occurs when applying a 500 mg/mL feed at a flowrate 70% below v_{base}.

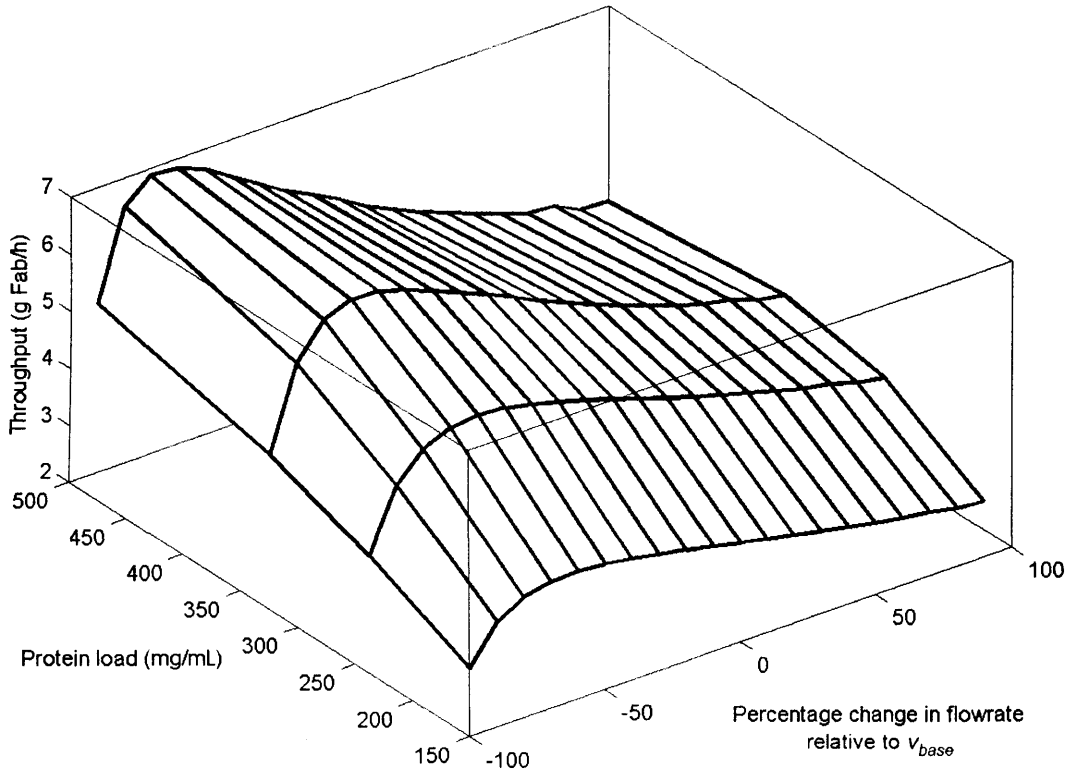


Figure 70: Impact of varying the flowrate by $0.9v_{base}$ to $2v_{base}$ on predicted throughput over five runs

7.2.4.5 Trade-off between yield and throughput

By comparing Figure 69 and Figure 70, it is clear that the operating conditions which maximise yield do not coincide with those that maximise throughput - the highest yield is achieved at the lowest flowrate and protein load, whereas throughput peaks at a flowrate that is 70% below the base case and at the highest protein load. Hence there is a need to trade off flowrate and feed load in order to identify suitable combinations of values for these variables that achieve industrial goals for a minimum yield and throughput. Based on discussions with Protherics, the minimum yield was set to 80% and minimum throughput was set to 3.38 g F_{AB}/h (achieved at the current production flowrate with the lowest protein loading in this study of 174 mg/mL). Combinations of feed load and flowrate fulfilling these criteria are represented by the white squares in Figure 71, whilst the black squares correspond to conditions which fail to meet the criteria. In all cases, operation needs to take place at or below the base case flowrate in order to meet both threshold values for yield and throughput. In general, as the flowrate is decreased from v_{base}, the protein load needs to increase in compensation. To identify the most suitable operating location from within the white square region, the outputs of yield and throughput were combined together using the multi-attribute decision making technique detailed in section 7.2.3.3. Discussions with Protherics identified yield and throughput to be as equally important as each other and as a result, equal weightings (w) of 0.5 were assigned to each and then used to calculate the Overall Rank from the yield and throughput data (Equation (97)). The results in Figure 72 indicate that the most favourable conditions with respect to the Overall Rank occur at a load concentration of 347 mg/mL and at a flowrate 70% lower than the base case. This

suggests that operating at significantly below the base case flowrate and in the middle of the protein loading range is the most suitable location. Although reducing the flowrate by 70% increases processing times, the advantage of higher product recovery more than compensates for this, resulting in the most favourable Overall Rank.

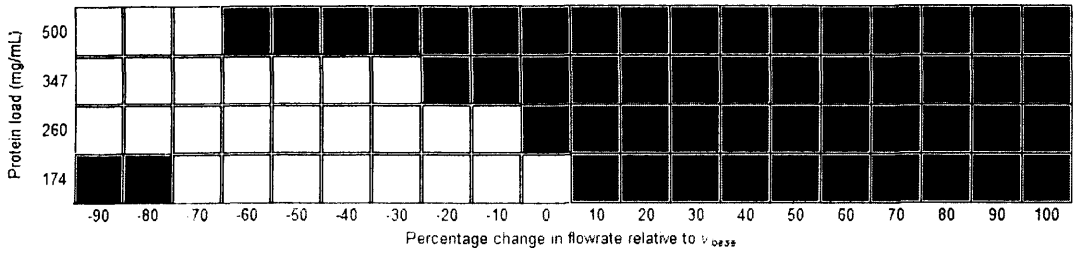


Figure 71: Plot indicating which combinations of operational flowrate and protein load achieve a yield and throughput over five runs that matches or exceeds a yield of 80% and a throughput of 3.38 g F_{AB}/h respectively
□ = *satisfactory*; ■ = *unsatisfactory*

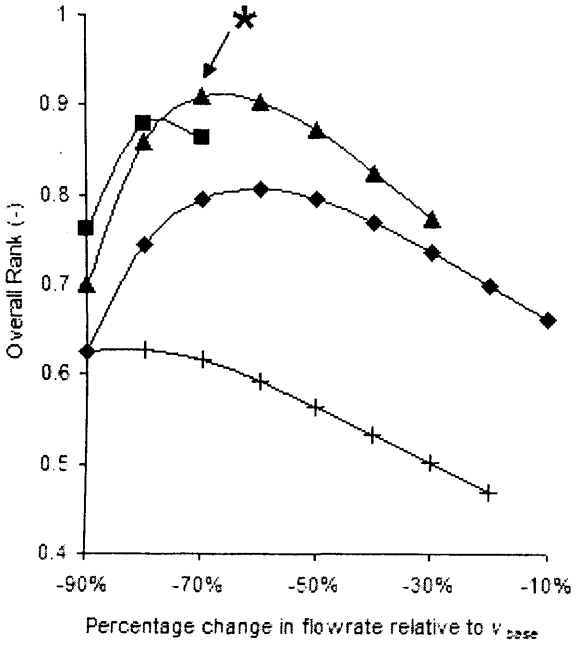


Figure 72: Overall Rank for those operational conditions that meet acceptance criteria as identified in Figure 71
+ = 174 mg/mL; ♦ = 260 mg/mL; ▲ = 347 mg/mL; ■ = 500 mg/mL. The asterisk marks the most favourable conditions according to the Overall Rank compared to any other location (at a flowrate that is 70% below the base case and with a 347 mg/mL protein load). This provides a yield of 0.97 and a throughput of 6.28 g F_{AB}/h

7.2.5 Case study summary

This section of the chapter has described the development of software which aids decision making during chromatographic separations of antibodies. The software simulates breakthrough behaviour in order to predict product yield and throughput, based upon scale-down column data which relate yield to the protein load and number of matrix re-uses. Outputs are combined together into a single value using multi-attribute decision-making techniques to identify the most suitable protein load and flowrate at which to operate. The application of the methodology to the DigiFab™ affinity step identified the most favourable operating conditions for achieving target yield and throughput values.

7.3 Evaluation of a novel agarose-based synthetic ligand adsorbent for the recovery of antibodies from ovine serum

7.3.1 Introduction

The increasing interest in the use of antibodies as clinical treatments (Flatman et al., 2007) has resulted in them becoming one of the largest categories of biotherapeutics currently in development (Roque et al., 2004; Shukla et al., 2007). The initial recovery step for many antibody manufacturing processes involves either fractionation or chromatographic purification using ion exchange or affinity matrices (Farid et al., 2006). Recombinant Protein A, derived from the cell wall of Staphylococcus aureus (Luo et al., 2002; Hober et al., 2007), for example, is an industrial standard that is commonly used to recover mAbs from clarified cell culture liquor (Fahrner et al., 1999; Swinnen et al., 2007). Unfortunately, several problems exist with this resin including the high purchase cost of Protein A and Protein G, poor stability when subjected to standard CIP solutions such as sodium hydroxide and the low capacities exhibited when challenged with crude, highly viscous feedstocks (Newcombe et al., 2005). Such issues have driven the manufacture of a series of next-generation synthetic ligand adsorbents (Hober et al., 2007) which have demonstrated considerable potential in capturing antibodies (Chhatre et al., 2007b). These are typically cheaper than recombinant Protein A or Protein G and permit robust alkaline cleaning protocols which do not adversely affect binding capacity, thus making them highly attractive for the purification of antibodies from a crude feed, such as recovering pAbs from hyperimmunised ovine serum (Bak and Thomas, 2007; Newcombe and Newcombe, 2007). This chapter evaluates a prototype affinity adsorbent utilising an undisclosed synthetic ligand (Millipore, Consett, Co Durham, U.K.) and compares its performance with four commercially-available resins (Table 40), as well as the precipitation and centrifugation steps employed at present for CroFab™ manufacture, to determine the ability of the prototype to recover and purify polyclonal IgG from ovine serum.

Adsorbent	Manufacturer	Base matrix	Ligand	Particle size (μm)
Prototype	Millipore (Consett, Co Durham, U.K.)	Agarose	Not disclosed due to commercial limitations	20-100†
MEP Hypercel	Pall BioSepra (Cergy-Saint-Christophe, France)	Cellulose	4-Mercapto-Ethyl-Pyridine	80-100
ProSep G	Millipore (Consett, Co. Durham, U.K.)	Controlled Pore Glass (CPG)	Recombinant Protein G	75-125‡
MAbsorbent® A2P	Prometic Biosciences (Cambridge, U.K.)	Agarose	Di-substituted phenolic derivative of tri-chlorotriazine	75-125
ProSep-vA Ultra	Millipore (Consett, Co Durham, U.K.)	CPG	Native protein A	75-125

Table 40: Adsorbents used in the study and details of their composition and size
(† = determined by method in 2.7; ‡ = assumed to be the same as ProSep A since both have a CPG base matrix)

A robust method is required for the operation of a commercial scale column in order to achieve high recoveries and impurity clearance over repeated feed applications. To establish the ability of the prototype to satisfy these objectives, the following scale-down studies were conducted:

- Calculation of static and dynamic matrix capacities when challenged with purified IgG
- Determination of novel adsorbent column re-use capacity when challenged with ovine serum
- Evaluation of cleaning protocols for the prototype
- Application of confocal microscopy to compare protein uptake patterns between the prototype and commercially available Protein G Sepharose (GE Healthcare)

The methods used to carry out these studies can be found in Section 2.9.

7.3.2 Results and Discussion

7.3.2.1 Static and dynamic capacity measurements

Figure 73 shows the static IgG capacities of the five matrices and demonstrates that the prototype has the highest binding capacity when challenged with both purified IgG and ovine serum. Except for ProSep G, the capacities are all higher for ovine serum compared to the pure IgG, with the prototype having a capacity of 54 mg/mL – a value that is comparable to ProSep-vA Ultra and higher than MAbsorbent® A2P, MEP Hypercel and ProSep G. Based on these results and to determine accurate dynamic IgG binding capacities, 150 mgs of purified IgG (5 mL of 30 mg/mL) was applied to overload a packed bed of each matrix and breakthrough curves were generated. The volume of feed loaded up to 10% breakthrough at 900 mAU (see section 2.8.3) was measured and inserted into Equation (3) (see page 58) in order to calculate the dynamic binding capacities of the matrices. The results in Table 41 show that the prototype adsorbent has the highest capacity out of the five evaluated, with a dynamic binding capacity of 29.2 mg IgG/mL matrix under the conditions described – approximately 30% higher than the best commercial adsorbent evaluated.

Adsorbent	Dynamic binding capacity (mg/mL)
Prototype	29.2 (±1.5)
MEP Hypercel	21.3 (±0.3)
ProSep G	18.7 (±0.4)
MAbsorbent® A2P	16.3 (±1.1)
ProSep-vA Ultra	13.7 (±0.1)

Table 41: Calculated dynamic capacities of each adsorbent
The standard deviation (1 d.p.) is shown in brackets

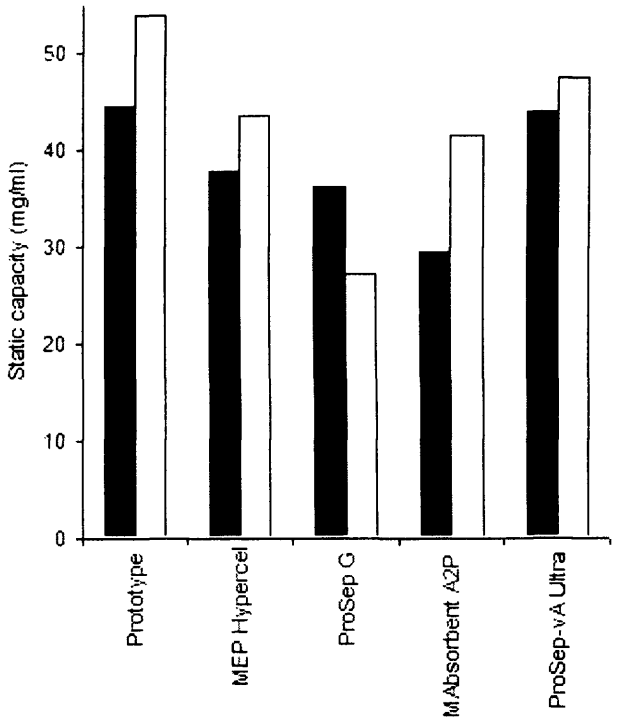


Figure 73: Static capacities of the five adsorbents
(■ = purified IgG; □ = serum)

7.3.2.2 Re-use study of the novel adsorbent

A packed column (C10/10 column, GE Healthcare) of the prototype was operated as described in section 2.8.4 for a total of ten cycles and the main flowthrough and elution peaks were fractionated and subjected to non reduced SDS PAGE analysis. The sample chromatogram in Figure 74 was taken from the first run and shows that both the main flowthrough and elution peaks have small shoulders next to them (F1 and E1 respectively in Figure 74). The flowthrough was collected in three 1 mL fractions, with the first consisting of the shoulder, whilst the eluate was collected in two separate fractions – the first containing the shoulder (E1) and the second containing the primary peak (E2). A gel of the flowthrough peaks from the third cycle is shown in Figure 75, along with an elution shoulder. The flowthrough shoulder contains a small amount of both IgG and albumin, whereas the main flowthrough peaks and the shoulder of the elution peak mainly contain only albumin.

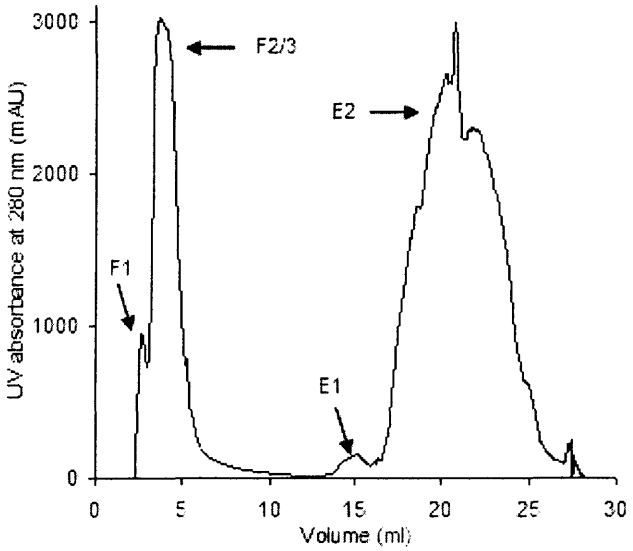


Figure 74: A 280 nm chromatogram produced when challenging a packed bed of the prototype adsorbent with 0.2 micron filtered ovine serum

Labels refer to lanes in the gel image in Figure 75. F1 refers to the fraction taken consisting of the shoulder on the flowthrough, whilst F2 and F3 lie within the main flowthrough peak. E1 refers to the shoulder on the elution peak, whilst E2 is the primary elution peak

The ten cycle re-use experiment was also repeated with a 3.14 mL column using a post serum-load caprylic acid wash to determine whether this would improve the purity of the eluted IgG. Figure 76 shows a gel of the primary elution peaks (i.e. peak E2 in Figure 74) from both re-use studies. Yields of 85% or above were obtained consistently, both with and without the caprylic acid wash – comparable to the precipitation and centrifugation steps employed currently by Protherics for manufacture of CroFab™. Notably, after the fourth loading cycle when operating without the caprylic acid wash, the albumin-rich elution shoulder merged with the main IgG elution peak, reducing its purity and overall, purities ranged from 86-91% (cycles 1-4) and 79-84% (cycles 5-10). Addition of caprylic acid substantially removed residual albumin remaining on the column after serum application, resulting in high IgG purities of 92-98%, comparing well with the 91% value achieved using the precipitation and centrifugation steps employed at present (Newcombe et al., 2005). The dynamic binding capacity of the caprylic acid-washed bed was then evaluated again after ten consecutive runs and an average of 27.2 mg/mL was obtained – close to the 29.2 mg/mL obtained in section 3.1, suggesting minimal loss of binding capacity. It should be noted that when operating under the conditions described, the volume of the eluate was around 10-11 mLs from each column run (10x the load volume; it is proposed that this is due to gradual elution of polyclonal isoforms from the column – this is based on the fact that in general, monoclonal antibodies would be expected to elute into relatively small volumes [personal communication, A.R. Newcombe, Protherics U.K. Limited]) suggesting that concentration would be needed before subsequent processing.

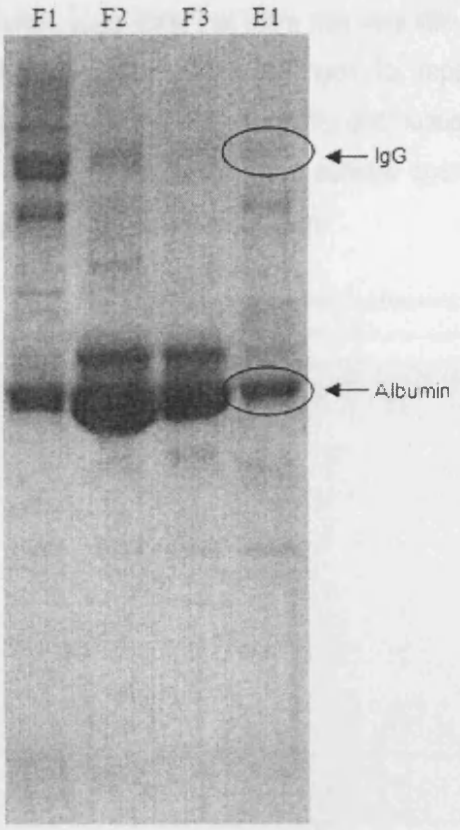


Figure 75: SDS PAGE gel showing the fractions identified in Figure 74 (Taken from the third cycle of the ten cycle run. F1 = shoulder of the main flowthrough peak; F2/F3 = second and third fractions of the flowthrough peak; E1 = shoulder of the main elution peak)

7.3.2.3 Column cleaning

Given that sodium hydroxide is commonly used to clean non-protein ligands, initial experiments to evaluate the cleaning and sanitisation of the prototype adsorbent used five column volumes of 0.5M NaOH. A 3.14 mL packed bed of the prototype adsorbent was subjected to 1 mL of ovine serum, washed and eluted as described in section 2.8.4 and then cleaned and sanitised with 0.5M NaOH, prior to re-equilibration and application of another 1 mL of ovine feed. The chromatogram in Figure 77 was then obtained, with no initial shoulder and only a small elution peak. The flowthrough was collected in 1 mL fractions and was shown by SDS PAGE analysis to contain large amounts of IgG and serum proteins whilst the eluate lane only contained a faint band of IgG (Figure 78). The gel also revealed a significant amount of 'smearing' along the flowthrough lanes. Given that the only difference between this bed and those described in section 3.2 was the application of a 0.5M NaOH sanitisation step, the smearing may therefore represent leached ligand arising from NaOH wash that in turn results in a reduced column capacity. In this respect, the prototype behaves similarly to conventional Protein A resins in terms of its low tolerance of high pH conditions. Conversely, when five column volumes of 0.1M HCl were used to clean a 3.38 mL column of the prototype and the subsequent 1 mL ovine feed was applied, a similar chromatogram to Figure 74 was obtained. Fractions of the flowthrough and the elution peaks were also shown by gel analysis to contain similar proportions of IgG and serum proteins to those obtained previously in section 7.3.2.2 (Figure 75 and Figure 76). Closer examination of the UV trace during the HCl wash showed small peaks appearing

within one to two column volumes, suggesting that there was very little build-up of adsorbed protein to the bed. The HCl washing experiment was continued for repeated loading cycles, which demonstrated that consistently high yields (just under 90%) and purities (95%) were achieved over nine affinity cycles. Hence 0.1M HCl appears to be a suitable agent for cleaning the prototype adsorbent when purifying polyclonal antibodies from serum.

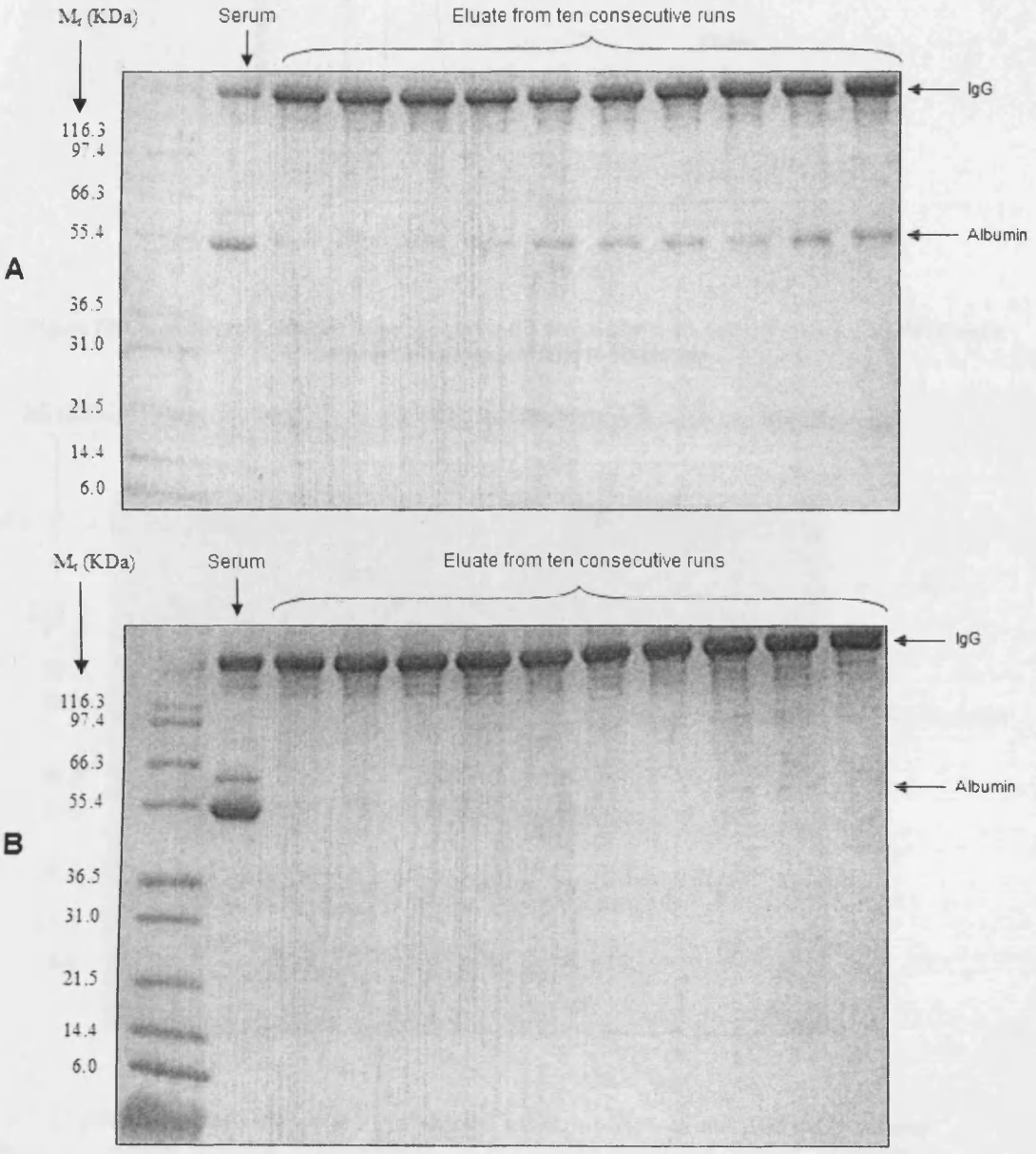


Figure 76: SDS PAGE gel of the eluates

Primary peaks i.e. E2 in Figure 74, collected from ten consecutive runs when a packed bed of the prototype was challenged with 1 mL of 0.2 µm filtered ovine serum: (A) without a caprylic acid wash; (B) with caprylic acid. Purity drops appreciably during runs 5-10 when operating without a caprylic acid wash

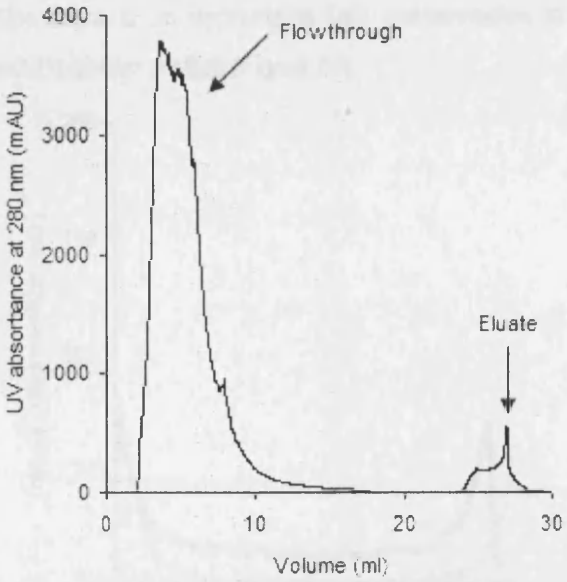


Figure 77: Chromatogram obtained when applying ovine serum after a five column volume 0.5M NaOH wash
Minimal IgG appears to bind to the column

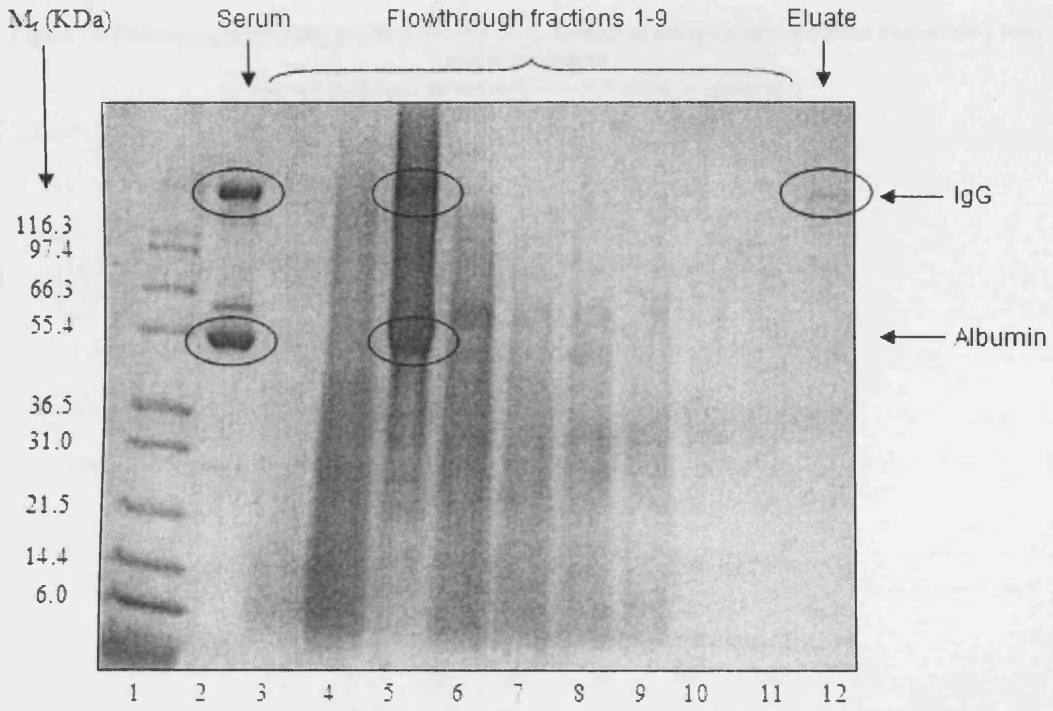


Figure 78: SDS PAGE gel of the flowthrough and elution fractions after 0.5M NaOH washing
Lanes: 1 – marker; 2 – ovine serum; 3 to 11 – the nine flowthrough fractions; 12 - eluate

7.3.2.4 Inverted confocal microscopy

Confocal microscopy was used to evaluate binding and subsequent diffusion of fluorescent IgG over time. Figure 79 shows the fluorescence intensity profiles along the cross sections of the beads of both the prototype and Protein G Sepharose immediately after matrix incubation, whilst Figure 80 shows the equivalent profiles after 150 minutes. Immediately after incubation, the prototype shows similar intensities at the edge of the beads compared to the Protein G Sepharose. Conversely, the Protein G adsorbent shows no protein binding at all across most of the central 50% of the bead,

whereas for the prototype there is an appreciable IgG concentration in the centre. After 150 minutes, both matrices exhibit similar profiles (Figure 80).

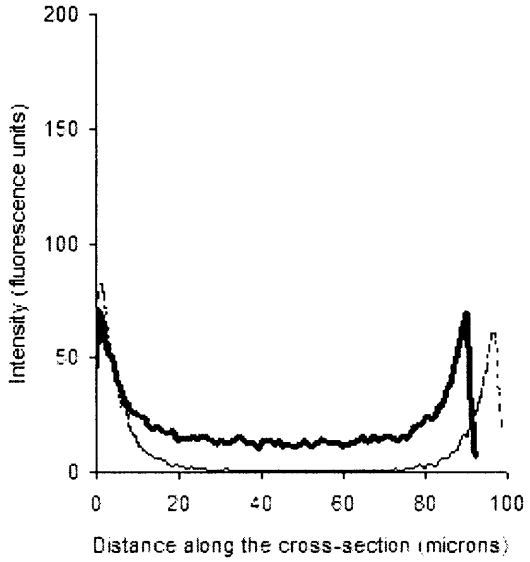


Figure 79: Fluorescence intensity profile along the cross section of the adsorbent particles immediately after matrix incubation
(—— = prototype adsorbent; - - - = Protein G Sepharose)

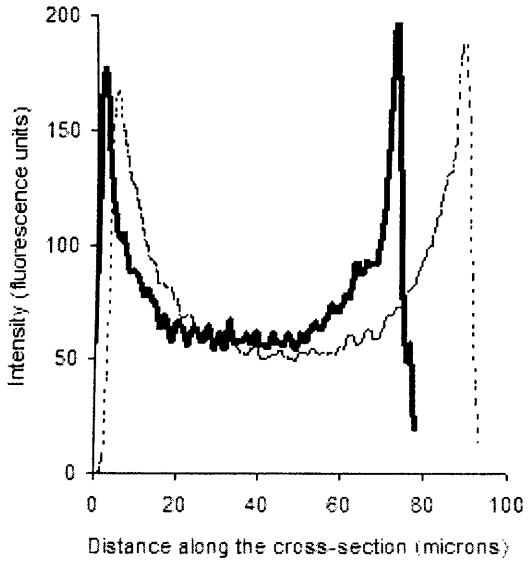


Figure 80: Fluorescence intensity profile along the cross section of the adsorbent particles after incubation for 150 minutes
(—— = prototype adsorbent; - - - = Protein G Sepharose)

Despite deliberately over-saturating with protein, however, the IgG appears to concentrate primarily along the outer 20% of the beads in the case of both resins, leaving a central core that exhibits only a third of the intensity seen on the outside. The experiment was repeated with a 24 hour incubation of the prototype adsorbent. The results in Figure 81 show that after this period of time, the intensity profile is similar to that obtained after 150 minutes, suggesting that equilibrium has been reached and no further IgG can bind. These data therefore imply that the IgG uptake characteristics of the

novel adsorbent are similar to those of Protein G Sepharose and that the prototype may be a viable option for the commercial-scale capture and purification of polyclonal antibodies from crude industrial feedstocks such as ovine serum. Due to commercial limitations at the time of writing, details of the bound synthetic chemical ligand (structure, charge, bead density and hydrophobicity) have not been disclosed and hence no further characterisation was undertaken.

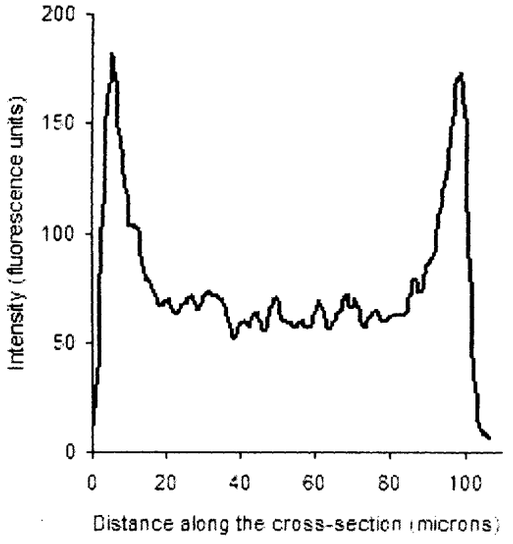


Figure 81: Fluorescence intensity profile along the cross section of a bead of the prototype adsorbent after incubation for 24 hours

In control experiments, the intensity profile was shown to remain constant in the fifteen minutes during which analysis of the prototype adsorbent occurred, implying that there was no photobleaching over the course of the experiments. The following graph was taken from a prototype adsorbent bead after 150 minutes incubation and the intensity profile was scanned at five 3-minute intervals. In all cases, the fluorescence traces were found to be very similar:

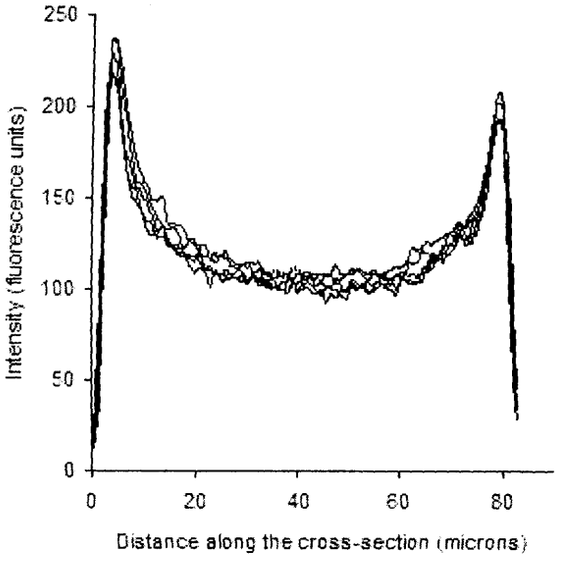


Figure 82: Constant intensity profiles over a fifteen minute period during the control experiment

It was also shown that there was no binding of the fluorescently labelled IgG to the Sepharose CL 4B agarose beads, implying that the fluorescence patterns seen with the prototype and Protein G Sepharose were due to the binding of IgG to the ligand and not due to non-specific binding of fluorophore. Intensity profiles were also comparable for each of the three beads analysed at each time point, suggesting that the data that formed the basis of the evaluation were a representative sample.

7.3.3 Case study summary

This section of the chapter has evaluated a prototype adsorbent for its ability to recover and purify polyclonal antibodies from crude hyperimmunised ovine serum. The adsorbent was shown to have a higher dynamic capacity for polyclonal IgG than a range of other commercially available matrices and with a post-load caprylic acid wash, was shown to achieve consistently high yields and purities over ten purification cycles with minimal drop in binding capacity. It was also demonstrated that than 0.1M hydrochloric acid could be applied to clean the column efficiently. Inverted confocal microscopy was used to visualise the antibody binding characteristics for individual beads of the prototype matrix, producing similar results to those achieved with a commercially available Protein G. Although the adsorbent evaluated is not currently available commercially and the details of the bound ligand have not been disclosed, these data indicate that the prototype adsorbent is a feasible option for the recovery of polyclonal antibodies from ovine serum and may therefore have wider applicability to other complex feedstocks of industrial relevance. The high binding capacities and eluted IgG purities observed suggest that this 'next generation' synthetic alternative to Protein A or G chromatography adsorbents may be suitable for the commercial scale purification of other products such as therapeutic monoclonal antibodies from mammalian cell culture or from feedstocks produced using transgenic technologies.

7.4 Chapter conclusions

This chapter has focused upon the development of simulation and experimental methods for evaluating the performance of chromatographic matrices in recovering polyclonal antibodies and their fragments from complex protein feeds. Thus, for example, capacity, re-use and cleaning data obtained from the assessment of the prototype adsorbent facilitate good column design and selection of the optimal operating conditions in order to achieve required goals for outputs such as yield, throughput and purity of the product.

8: COMMERCIALISATION, MANAGEMENT AND VALIDATION ISSUES RELATING TO THE RESEARCH

8.1 Introduction

As part of the Engineering Doctorate scheme, it is a requirement for the thesis to include sections describing commercialisation and validation issues arising from the thesis. For example, the research described in the previous chapters has the potential to provide substantial industrial benefits by enabling the rapid identification and appraisal of manufacturing strategies which satisfy both business and market driven goals. Hence this work could form the basis of a start-up company providing these services to other organisations such as pharmaceutical, biotechnological, or contract manufacturers that are actively seeking to exploit new approaches for drug development. This chapter focuses upon how to commercialise the technology described previously.

8.2 Commercialisation of the technology

8.2.1 Introduction

This section of the chapter addresses the following topics associated with setting up a new company based upon the software technologies proposed previously:

- Identify the industrial need for the models developed in this thesis
- Determine the value of these technologies and how they would meet these needs
- Specify the services offered by the start-up
- Identify the companies which constitute the best targets for marketing
- Determine the potential advantages of the start-up compared to competitors
- Summarise the costs involved in setting up the company and how funding would be acquired
- Identify potential risks to the new business

8.2.2 The need for software technology

The development of pharmaceutical and biotechnological processes has become an increasingly time-consuming and expensive process which can still fail to locate the conditions needed to optimise financial and technical performance. Although the operation of sub-optimal systems was acceptable in the past, owing to the high premiums that new drugs were able to attract, costs have now risen and revenues have fallen to the point where drug manufacture is becoming progressively less economically viable. This is compounded by other factors which increase the difficulty that companies have in planning for commercial-scale operation at early stages of development:

- Uncertainties in titres achieveable upon scale-up or dosage needed to for acceptable efficacy
- The requirement for large process volumes during pilot studies which may be costly or unavailable early-on
- The lack of sufficient time available for process optimisation (at pilot and commercial scale) during development, meaning that companies are forced to sacrifice development in favour of rapid market access in order to exploit patent cover. This can often give rise to processes that lack robustness and cost-effectiveness
- Growing global competition, increasing numbers of drug substitutes, reimbursement pressures, regulatory moves to apply cGMP standards to materials in early-stage development and growing clinical burdens mean that operating inefficient processes is becoming increasingly uneconomic

These factors have all increased the need for designing processes quickly and correctly, especially when considering the growing complexity of drugs that are now being developed, which tends to increase the time and money needed to bring them to full-scale production.

8.2.3 Value proposition in commercialising the software

New methods now being investigated for the design of pilot- and commercial-scale production systems include high throughput screening, ultra-scale down techniques and modelling software. The latter has attracted renewed interest as a means for overcoming the current challenges and typical questions that can be answered by software fall into three groups:

- Process-related e.g. what is the most suitable flowrate to optimise yield and throughput?
- Facility-related e.g. what is the best approach to resource deployment to eliminate bottlenecks?
- Business-related e.g. which operating strategy minimises the cost of goods?

A software company based on the research in this thesis can be set-up to provide simulation methods to major pharmaceutical and biotechnological companies in order to answer these types of questions. Ultimately this would help to get drugs to market more quickly and cheaply, giving benefits to both companies and patients alike.

8.2.4 Intellectual property rights

The intellectual property rights for the start-up company take the form of the copyright that subsists within the software. For each individual client, the models produced will be bespoke solutions and provided on licence. In all cases, the copyright will be retained by the start-up company.

8.2.5 Services provided by the company and its novelty

The services being offered by the new software company are wide ranging and include:

- Identifying the most critical input variables in a unit operation or whole process by Global Sensitivity Analysis, in order to determine the most suitable process development strategies
- Investigating and evaluating process development strategies by using models developed for individual unit operations (e.g. chromatography) or whole process models constructed in Extend™. The results can then be used to make decisions, such as determining the best unit operations for a given process duty and the operating conditions required to achieve required goals. The software can also be used to address related problems e.g. highlighting constraints upon throughput, pinpointing possible solutions and evaluating their relative merits
- Using multi-attribute decision-making techniques to integrate development and manufacturing model outputs together in order to trade off their values and so determine the best overall option
- Models can also be offered to simulate a drugs portfolio of at various stages of development and manufacture, enabling a more holistic representation of the activities of a typical company

The novelty of the business lies in troubleshooting pre-existing processes that need improvement by using model-driven implementation of Quality By Design. Growing emphasis is being placed by regulatory authorities upon using a systematic approach for exploring the design space to work out the most robust operating region. As indicated earlier in the thesis, performing a full factorial design by experimentation can be time-consuming and hence simulation approaches can be a useful way to narrow down the size of the design space to those key factors of critical importance. A fractional factorial experimental design can then be set-up focusing upon those factors or those process options that have been highlighted by the model to be promising approaches.

8.2.6 The target market

The target market for these software methodologies includes medium to large size pharmaceutical and biotechnology companies such as Pfizer, Eli Lilly or Genentech and also contract manufacturers (such as Avecia or Lonza) that develop or modify processes for pilot- or full-scale production.

- Where a new process is being designed, USD studies can be undertaken during early development to gather the data needed to populate simulations. Experimental studies at this scale are best suited to early development as there is normally insufficient process material available to conduct pilot- or full-scale trials and reluctance to procure larger scale equipment owing to a lack of efficacy data. The populated models can then be used for optimisation e.g. making choices for the type, number and sequence of unit operations in a flowsheet. As these studies are being undertaken far earlier than in 'normal' development, this leaves a larger window of time to optimise the process, increasing the chances of doing so satisfactorily
- Alternatively, where changes to an existing process are being countenanced, simulations can investigate development options, followed by a limited number of full-scale verification runs

Clients benefit from this approach in several ways:

- A company could potentially reduce the time to market for a new candidate or reduce the time taken to re-start pre-existing processes that are undergoing optimisation. By accelerating development, the money saved could be diverted to other company activities such as new drug research and development or implementing facility improvements
- Models can help with process robustness (e.g. windows of operation can determine how wide the operating boundaries need to be to satisfy production goals). Ultimately, this can lead to better process design, higher manufacturing consistency and thus fewer regulatory recalls
- Higher quality processes increase regulatory compliance and therefore enhances the reputation of a client with government bodies such as the NHS. This reputation also encourages other companies to go into partnership with the client for the development/manufacture of other products i.e. the benefits can extend into future projects

8.2.7 Competition to the new software company

Commercially-available software for modelling bio-manufacturing operations includes SuperPro Designer (Intelligen) or BioProcess Simulator (AspenTech) and there are several examples in the literature describing the application of these packages to the production of commercially relevant bioproducts e.g. Harrison et al. (2003) describe the use of SPD in simulating the production of monoclonal antibodies and insulin. Nevertheless, these programs are hampered by restrictions:

- Their static nature is unable to account for dynamic resource allocation/deallocation in a process
- The currently available packages are generic, thus making it harder to capture the unique characteristics of each process
- Unit operation models are often highly simplified, meaning that they provide little process insight

The methods developed by the new company instead utilise dynamic simulation software which incorporate mathematically rigorous design equations in order to generate more comprehensive process understanding. In addition, the models will be custom-built for each client, allowing any unique process characteristics to be incorporated.

Further competition comes in the form of consultancies such as Biopharm Services, which already provides custom-built Extend modelling solutions to organisations such as Cancer Research U.K. or GlaxoSmithKline. Possible competitive advantages compared to Biopharm Services include:

- Biopharm Services does not provide the actual final model to the client, whereas the new company will do so, with full training and after-support in how to use and maintain the simulation
- The company is founded at University College London, with access to other process development technologies such as Ultra Scale Down that can assist with acquiring input data

8.2.8 Company start-up

Initially, staff members with a background in Biochemical Engineering would be recruited. These individual would have a good working knowledge of how difficult unit operations work, how they are modelled as well as how processes are developed and optimised. Other requirements would include an understanding of cost evaluation and facility analysis. These people could come from the Department of Biochemical Engineering at University College London, with an additional source of employees coming from institutions such as the London Business School, who develop the marketing skills needed in order to advertise the company.

University contacts would be used to gain the initial contract in the first year of company operation, which would need to be for a major pharmaceutical or biotechnological client. This would provide an illustrative 'showcase' which demonstrates the utility of the methods and thus give the start-up a high profile. This company could be Protherics, since the collaborative EngD has demonstrated the potential of the modelling techniques to the company. Furthermore, a partnership that Protherics has entered into with AztraZeneca for another novel product currently going through development could also be a useful source of future contracts. Immediately after this, additional business would be solicited by direct application to companies including Pfizer, GlaxoSmithkline, Eli Lilly, Amgen and also contractors. Again, the high profile of these organisations would be useful in giving wider exposure to the software technology. Advertisements placed in the trade press, on websites (both self and others e.g. Pharmiweb, Bioindustry Association and BioprocessUK), through published papers and at conferences would also be used to market the company.

8.2.9 Expenditure and revenues

8.2.9.1 Expenditure

- In the initial six month period of the company, seed finance of around £20,000 would be needed to acquire equipment (computers etc.), rent out premises, conduct initial market research and advertise the company. This would be sought from sources including self financing (including friends and family of the founder), DTI grants and angels

During this period, the copyright of the technology which is currently owned by University College London would need to be licensed out or transferred, in return for payment such as a nominal one-off fee in addition to a royalty stream in the event of the company being successful. In the subsequent six to twelve months, another £50,000 would be needed for company expansion, involving further marketing and building up the client base.

8.2.9.2 Revenues

All projects will incur revenue by charging a percentage of future product sales – typically, the industry norm is approximately 1% (Neal, 2005). No upfront charge would be made (at least during the early years of company operation) in order to attract companies. Based on the work carried out by this thesis, it is known that it takes around three months to complete each modelling exercise and hence potentially, over a ten year period, a total of forty contracts could be undertaken. From its website, Biopharm Services has worked with around fifty companies from 1992 to 2007 and hence forty contracts over a ten year period for the new company seems reasonable. If it is assumed that the start-up works on one to two drugs which become blockbusters, while the remainder of the forty contracts attract lower revenue streams, then the following approximate calculations can be made (the exact values depend upon future sales values, which by their nature are uncertain and depend upon a range of factors, including the value and demand of other competing products in the marketplace). Hence, the values given are approximate:

Blockbuster?	Numbers of candidates	Sales (£M)	Value at 1% of sales (£M)
✓	2	500	10
✗	38	125	47.5
		Total (£M):	57.5

Table 42: Sales values for the new company
The £125M sales value was chosen as this is below the 'radar' for most major pharmaceutical companies which normally seek blockbusters (£0.5 billion is the normal threshold for a new product to be defined as a blockbuster). Hence £125M is used to represent a typical 'modest' sales revenue figure

Hence the new company could stand to make around £57.5 million over a ten year period. Even if sales revenue was far lower e.g. if there were no blockbusters and the sales revenue for each candidate was only (say) £25M instead of £125M, the start-up would still stand to make 0.01 × 40 × 25 = £10M, which is still a reasonable amount for a small software consultancy.

8.2.10 Potential risks to the business

- Not attracting enough clients (e.g. due to an inability to acquire a sufficiently high profile)
- Other companies e.g. Biopharm Services, or future competition such as other spin-out companies set-up by future doctoral students
- In the experience of the author, the industry is sceptical generally about the use of software for process design and this needs to be overcome before a viable business could be created

Initial marketing efforts will be critical in mitigating these factors to ensure success, as outlined further in the next section.

8.2.11 Case study

In order to make the set-up described above more specific and to demonstrate the utility of the start-up organisation, this section describes a hypothetical scenario in which the services of the company are called upon to help a client. The scenario is *not* real and is only used to illustrate how commercialising the techniques developed in this thesis *could* be of benefit to industry.

I) Scenario:

It is assumed that a client company has a process used currently for the full-scale production of a soluble therapeutic protein (e.g. a monoclonal antibody), but which is at risk of losing market share due to the following:

- Product yield is inconsistent, meaning that market supply can be difficult to maintain sometimes
- Batches need to be re-processed on occasion due to an inability of the process to meet purity specifications. Problems include leached ligand that may not be cleared completely
- The multi-product facility used to manufacture the product is employed continuously for other products and so there is little room to increase the number of feed batches processed per year for the monoclonal antibody in question.
- The low yield and the need for re-processing means that the cost of goods is high and hence the price to the consumer is large. There is therefore a risk that agencies such as the National Institute for Health and Clinical Excellence in the U.K. may re-examine whether the monoclonal is a viable product to buy, especially since product alternatives may become available in the next few years with similar efficacy profiles but at a lower cost.
- The company has very little time or money to spend on improving these issues itself due to other projects – but it nevertheless wishes to improve the prospects for the monoclonal antibody in the long term

II) Company strategy for dealing with problems:

The software consultancy has been approached to try and find ways that can increase production levels and decrease costs. The approach taken by the company involves:

a) Examining the process and identifying possible problems e.g. where costs are high or where product is lost
b) Using process data to determine likely data values and equations. Laboratory experiments may also have to be performed to populate equations fully. Precise equations to choose will be determined by the type of unit operation being considered - for example in a column step, molecular properties such as size, charge, hydrophobicity, kinetic and mass transport properties will affect the type of equation that is suitable (e.g. pore diffusion, film mass transfer or solid diffusion etc.). Genetic algorithms may have to be encoded to fit literature equations to the data
c) Applying GSA to those equations generated in part (b) to identify the key variables for all flowsheet development options

d) Using an Extend model to explore the response curve for those key variables (as determined by GSA) and for all flowsheet options, thereby determining likely operating regions

e) Using this information to confirm the exact operating points/regions where robust operation can be achieved by fractional factorial design (taking into account only those variables within an experimental set-up that have been determined by GSA to be relevant)

III) Application of the strategy to the scenario:

For this scenario, it is assumed that the monoclonal antibody production follows a standard process – e.g. a cell culture step, followed by an initial clarification by centrifugation, Protein A chromatographic purification, intermediate clarifying steps (e.g. filtration), purifying ion exchange chromatography and final polishing such as by size exclusion. The key problems with the process (part (a)) have already been identified from the scenario. Following the process of fitting equations to data from this process (as per part (b) above), it is assumed that GSA (i.e. part (c)) has revealed the following variables to be critical to the process:

- Fermentation titre
- Clarifying centrifuge flowrate
- Capacity of the direct capture Protein A column
- Flowrate and volume of the purifying ion exchange column

This information forms the basis of potential solutions to the problems faced by the client company. Simulations can then be run in Extend to determine exactly which problems exist in the process (part (d) in the strategy above). Hypothetically, this might show that:

- Cell culture titres vary depending upon the nutrient feed strategy (amount, rate and frequency), thereby affecting product yield
- Poor efficiency of clarification prior to the column capture step hampers the latter step's effective dynamic binding capacity due to the high solids concentraton being present
- The effective capacity of the purifying ion exchange step is too low owing to the use of only a small column size and a high flowrate

Based on the knowledge of exactly which factors are most important and how they affect the process, scale-down studies of the process can then be completed to optimise these variables:

- The nutrient feeding strategy
- The clarifying centrifuge flowrate
- The volume and flowrate used in ion exchange

A Design of Experiments approach (part (e)) can be taken to investigate these factors and so identify suitable operating values that work in flat (robust) regions of the response curve. The overall result of this approach should be an improved process that operates with greater consistency and

robustness with respect to product mass and manufacturing costs, without significantly changing the time needed for manufacture (i.e. without affecting the schedule of other manufactured products) and providing a better resistance against the influx of product alternatives into the market.

8.2.12 Conclusion

This section has described how the software research proposed in this thesis could be developed into a business and given the potential benefits of the technology, it stands a good chance of succeeding and making a profit. The next section goes into further details about how the software methods provided by the company would help with process development.

8.3 Software approaches to managing process development

In the past, most companies were unwilling to acquire or commit pilot or full-scale facilities to a project until convincing signs of drug efficacy had been established. This often left very little time prior to market launch in which to conduct process optimisation and so this was often sacrificed in order to minimise the time to market. Although this strategy *was* economically viable in the past, this is no longer the case and optimisation is becoming increasingly important to making drug manufacture profitable. This change has been caused by a variety of factors:

- Reductions in drug premiums owing to reimbursement pressures from bodies such as the National Institute of Health and Clinical Excellence when deciding whether to use a new drug
- The complexity of products being developed has increased (partly because the industry is focusing more upon complex biopharmaceuticals rather than simpler new chemical entities), resulting in longer and more expensive development
- Clinical trials now take far longer to complete than before, resulting from larger numbers of patients being tested and greater numbers of trials in each phase. This also necessitates a more extensive data management exercise, increasing the costs involved
- Nowadays, patents now provide fewer barriers to new competition entering the marketplace

Hence the industry generally needs to develop processes in a better way by starting their optimisation long before the acquisition of capital. This leaves a larger window of time in which to identify the most suitable manufacturing flowsheet and operating conditions. Development needs to cover issues of cost effectiveness, robustness, recoveries and purities i.e. business and process factors need to be integrated together in decision-making. The software methods proposed in this thesis enable a fast, simple and inexpensive means of achieving these goals and unlike a laboratory experiment, a simulation can easily be re-run at minimal expense of time or money. Nevertheless, the current status of the bioprocess software field still presents certain problems that need to be resolved before modelling becomes more widely accepted in the industry:

- How can an end user be certain that the results from a simulation accurately reflect 'real-life' performance? The only true way to verify model outputs at present would be to validate at scale, which may be impossible if no such process exists
- It can be very difficult to capture every possible aspect of a facility which could ever influence its performance and in addition, every process is unique - hence the use of a generic modelling framework may only approximate likely manufacturing performance
- One way of trying to solve these problems would be to develop more generic models which simulate performance of each unit operation at the molecular level, rather than the more general mathematical (and especially calculus-based) approaches used to date

8.4 Software validation issues arising from the thesis

8.4.1 Introduction

Technically, since the modelling tools proposed in this thesis are for process development only and hence are 'offline', there is no explicit need to validate them in the same way as for 'online' process control software. Nevertheless, if they were to be analysed in the same way, this would give regulators confidence that a rigorous basis was used for process development. FDA software guidelines formed the basis of the research presented in the previous chapters and the steps taken to validate software can be expressed as the 'V' model, which encompasses the planning, design, commissioning, testing and on-going operation/maintenance of the software (Figure 83):

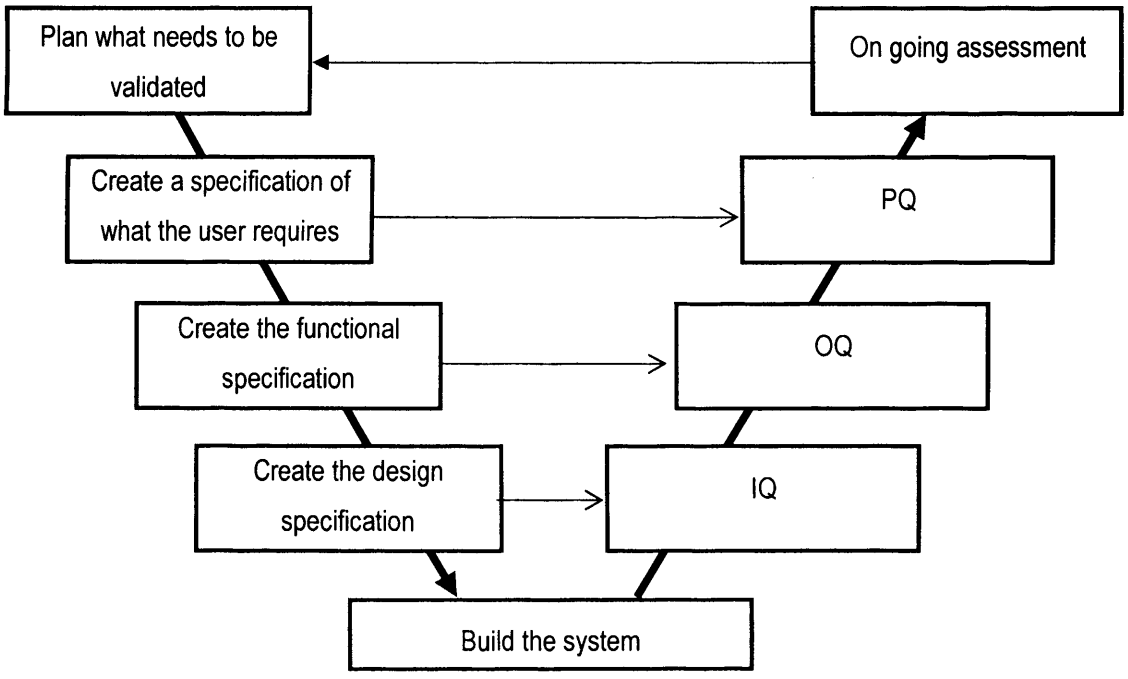


Figure 83: 'V' model of validation
The thin horizontal lines indicate how planning/design tasks (on the left) are implemented (on the right)

8.4.2 Software planning phase

The first step in the 'V' model identifies the key things to be delivered in the project, the personnel responsible for doing so and the timescales involved. The outcome from the planning stage is the development of a User Requirements Statement (URS), setting out what the program needs to do.

- In the case of each model set-up for Protherics, there was always a thorough discussion with their development and management staff in order to identify their key requirements, often formulated in the guise of questions that need to be answered – e.g. what are the advantages and disadvantages of using a synthetic adsorbent step at the start of the CroFab process?

8.4.3 Specification and design phases

A functional specification is used to determine how the URS is met, by defining the programming code needed to build the system.

- The component parts of each model were always identified on paper and the outline (pseudo-English) code was always written down before coding. The efficiency of the code was also a key factor at this point and a modular approach was typically employed in order to reduce the amount of re-coding, thus also making it easier to identify and correct bugs.

8.4.4 Qualification phases

The qualification phases are used to review the quality of the program, eliminating any redundant code and documenting any changes after model construction. As part of this, PQ involves documented testing against the URS to show that the program works correctly in routine use.

- As per the defect prevention guidance given by the FDA, processes were put into place when generating the research to ensure quality by preventing the entry of defects into the code in the first place, rather than testing quality into the system after it has been built. Although testing *is* important, the complexities of a typical program make it difficult to check every feature exhaustively, meaning that defects may remain hidden. For this reason, software construction requires the use of suitable methods which detects errors when the code is first generated and in the Extend™ or MATLAB® models, the software engines have routines in place to trap illegal errors (e.g. division by zero, making a call to a file that does not exist etc.). The Visual Basic for Applications engine also includes this protection and as an additional safeguard, further code was added e.g. when entering the number of vessels needed to hold the volume present at a particular stage, if too small a number was inserted, then the interface would flag this up as an error and stop further processing. Nevertheless, it was realised at the start of the doctorate that such methods may still allow certain unexpected errors through and hence whenever a model was built, each component part was also tested using 'dummy' and 'real' data before being used in the main simulation. Each completed model was also run through step by step (e.g. using debug points in the MATLAB® code, or by using Pause or Count blocks to check on the progress of items between different sections of Extend codeblocks) in order to ensure that the

behaviour was as expected. It was only after such a process that any model was then used to generate the research results.

8.4.5 Ongoing assessment

This is used to ensure that any changes required are identified and implemented in a controlled manner before retesting, both when problems are encountered and also as a periodic review to check that the system still functions as expected.

- After model construction, the author periodically reviewed its operation to see if any further changes were needed. Due to the complexity of software, any change – however small – could potentially have a wider impact on the model and hence after implementing an alteration, its effect upon the validation status of the software always needs to be re-established in order to show that nothing adverse or unexpected occurs and that the intended effects of the change are observed. Where a change was required in Protherics' models, it was always planned and its implementation accompanied by a thorough re-test of the relevant component part of the model; if the change was also likely to affect other parts of the model, the whole simulation was then re-tested to check for any possible deleterious effects.

8.4.6 Software as a tool for scale-up

The software developed in this thesis could help with the scale-up from laboratory to pilot or commercial-scale operation and thus models could be built to predict the impact of increasing scale upon processing characteristics. For example the chromatography 'wall effect', which describes the support that the side of a column provides to a bed at small scale, is lost when moving up to a larger column – models could be developed to predict the resulting impacts upon performance.

8.5 Synthetic adsorbent column validation

8.5.1 Introduction

If the synthetic adsorbent evaluated in Chapter 7 was to be used as the basis of commercial operation, the following validation tasks would need to be carried out (in addition to others such as IO, OQ and PQ of the full-scale equipment as mentioned above):

8.5.2 Critical parameter identification

The impact of varying inputs such as flowrate, temperature, pH, conductivity etc would need to be evaluated through either single parameter investigation or factorial analysis, in order to identify the critical process parameters that affect performance. On this basis, acceptable levels for these variables would need to be set in order to achieve robust operation (Figure 84).

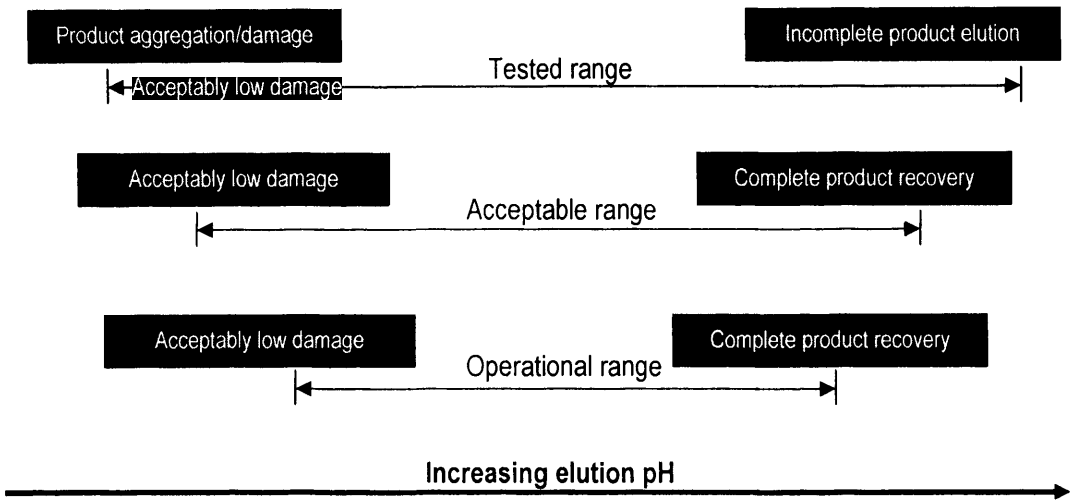


Figure 84: Different types of limits set for critical process parameters (using elution pH as an example)
The operational range inside the acceptable range minimises damage, maximises recovery and also allows for potential variability during processing)

8.5.3 Consistency

This is one of the most important aspects of validation and normally requires three to five sequential full scale lots in order to characterise the recovery of product from a chromatography step and the whole process, the level of impurity/contaminant clearance and product stability.

8.5.4 Resin re-use

This requires a combination of laboratory and full-scale data and test to check whether consistent results are achieved in the following properties across multiple runs with a given batch of matrix:

- The chromatographic profiles need to be comparable (e.g. for an elution profile, the number of peaks, the relative peak heights and the elution peak volume should all be similar)
- Levels of ligand leakage into the product
- Product recovery and impurity reduction
- Product integrity and activity
- Viral reduction (see below)
- Residual impurity levels after processing (checked with blank buffer runs)

8.5.5 Spiking studies

These involve operating a scale-down mimic of the commercial column. Effective scaling requires the following rules to be observed in order to achieve the same product yield, purity, retention time and elution profile:

- Same residence time (i.e. linear flowrate)
- Same buffer composition and numbers of column volumes for each step
- Same feed materials as used at full scale

The study then involves spiking a sample of the feed material with an impurity in order to determine whether the log clearance of that component is acceptable in the scale down mimic.

8.5.6 Viral clearance

The raw material to the CroFab process (ovine serum) is a potential source of endogenous viruses and hence the synthetic adsorbent step could also act as a viral reduction step in the process, in addition to the two currently validated methods used at full-scale (i.e. a low pH hold after venom-specific affinity purification and a viral filtration during the pre-formulation stages). Some of the aspects which need to be considered in a viral reduction study are given below (based on a lecture delivered by Julian Bonnerjea (Lonza) to the Planning and Implementation of the Development and Manufacture of Biopharmaceuticals MBI module, UCL, April 2004):

Aspect	Description
Selection of model viruses	• Viruses that resemble those which may contaminate the serum need to be selected. To provide the most comprehensive testing, a number of different viruses (of different sizes, morphologies, resistance to inactivation, DNA/RNA contents etc.) need to be selected • These need to be viruses which could be present at high concentrations in the feed stream and for which a reliable assay is available
Validation of viral assays	• Assays that can be used include infectivity assays and those based on transmission electron microscopy as well as the polymerase chain reaction • Several conditions that occur during normal processing could potentially interfere with these assays e.g. high conductivities, the presence of cytotoxic species etc. Pre-treatment can be used to eliminate these factors before conducting the assay, although it is necessary to show that pre-treatment does not reduce the viral content prior to assaying • The same needs to be demonstrated for storage conditions e.g. does retention at low temperatures affect viral titre?
Statistical evaluation of results	• Viral assays experience the same degree of variability as any other biological assay and hence in order to demonstrate reproducibility, 95% confidence limits need to be set at $\pm$ ½ $\log_{10}$ difference • When low sample volumes or low viral concentration samples are taken, there is always a probability that an infection could be missed – hence more steps may need to be taken e.g. calculating the probability of a viral particle being missed and therefore taking more samples to compensate, developing a superior assay, designing the process carefully to reduce the risk of possible contamination etc. The results from the viral clearance study of the synthetic adsorbent column would need to be integrated together with the data from the equivalent studies on other process steps in order to achieve the industrial standard of reducing viral infection to only one in a million final doses

Table 43: Viral clearance validation studies

8.6 Chapter conclusions

This chapter has concentrated upon the commercialisation and validation issues that arise from the research described in this thesis:

- Setting up a company based upon the proposed simulation methods was shown to be an economically viable proposition, although a number of issues surrounding the use of bioprocess models still need to be resolved before the financial potential can be realised fully
- Validation studies in relation to both the software and the chromatography experiments were outlined, showing how product quality, robustness and regulatory compliance could be achieved

9: CONCLUSIONS AND FUTURE WORK

9.1 Summary of the doctoral research

The research in the previous chapters has focused upon the development of techniques designed to evaluate the impacts of changing commercial-scale bio-manufacturing processes. The first chapter started by outlining some of the challenges facing the pharmaceutical/biopharmaceutical sector and then described a range of process changes which are typically considered for their potential to overcome these problems and so improve financial and technical performance. The utility of modelling techniques in achieving this was summarised, along with the gaps that currently exist within the field and how the methods proposed in this thesis set out to bridge them. A summary of the experimental work into the synthetic adsorbent step was also given.

The second chapter introduced the modelling and laboratory techniques needed to carry out the studies described in the previous chapter. After a review of available software, Extend™ was selected as it provided the most suitable environment for constructing the dynamic, discrete event models needed to represent a bio-manufacturing process accurately. The hierarchical structure used to organise and structure all modelling efforts in Extend™ was also put forward. Finally, the experimental methods such as static/dynamic capacity measurements and confocal microscopy were outlined.

In the third chapter, a simulation methodology for assessing the impacts of making moderate changes within the current production framework for CroFab™ was introduced, focusing primarily upon a yield/time sensitivity analysis and the effects of increasing the volumes of feed material, columns flowrates and volumes of affinity matrices. In particular, results showed that the final product yield was most strongly controlled by the papain and affinity steps, but that due to the cost involved, changing the way in which the papain step is operated was likely to be the more cost effective option. This work was extended in the fourth chapter, in which a novel three-layered software technique was proposed in order to enable the most efficient use of simulations and the underlying experimentation needed to collect the data required to populate them. In each layer, inferior changes are screened out, while the more promising options are sent to the subsequent, more refined layer, for further evaluation using more rigorous models that require additional data inputs from potentially time-consuming laboratory work. Screening ensures that laboratory experimentation concentrate solely upon those strategies which show the greatest potential. Alternatives to the CroFab™ process that were evaluated included increasing the feed volume, replacing centrifugation with microfiltration and replacing precipitation/centrifugation with a Protein-G column. Results showed that after assessing every option, doubling the volume of feedstock and using a Protein-G column was the best way to alter the existing process and so would be the

recommended process change strategy. As a check, every option was also run through every layer and this showed that all three layers were very similar in terms of which options were predicted to be particularly good or unsuitable. Building upon this work and since all manufacturing models to that point had been deterministic, this chapter also utilised Monte Carlo techniques to incorporate the effects of uncertainty into the simulations. These were applied to the main variable that tends to change – namely the IgG titre. Scenarios investigated included the effect of adding up to two extra anti-venom components to the formulation; this was shown to be beneficial in terms of cost of goods with minimal impact upon processing times, although it was found that the change might place additional burdens upon resources such as labour availability in the facility.

Whereas the the previous two chapters concentrated upon assessing the manufacturing impacts of process change, the fifth chapter sought to analyse both the development and manufacturing consequences of implementing alterations in order to find the best overall strategy which traded off development effort (time and money) against improvements in manufacturing (product mass, cost of goods and processing time). The method was illustrated through some radical alterations to CroFab™, including changing the precipitation/centrifugation steps to a synthetic adsorbent column step operating in either packed or expanded modes, increasing the feed titre/volume and removing downstream purification steps. Outputs from the study showed that the best performance in terms of manufacturing gains that outweighed development burdens was achieved by implementing a range of strategies simultaneously involving the use of an expanded bed column with the synthetic adsorbent, the highest feed volume and titre possible whilst operating without the ion exchange step.

Although the simulations described in previous chapters determine the financial and technical impacts of changing input variables, their utility would be enhanced by knowledge of which inputs were the more important ones to focus upon during development studies. Hence the sixth chapter examined the use of GSA, which provides a rigorous sensitivity analysis technique for determining the importance of both individual variables and their interactions. This allows a company to prioritise development activities upon the essential parameters alone. In this chapter, GSA was used to evaluate the parameters that have the greatest impact upon the performance of a manufacturing flowsheet, using both a synthetic adsorbent column and also the CroFab™ process as an example to illustrate the technique. Inputs that were assessed for their influence on product yield and throughput included flowrates, matrix capacities, loading durations and feed concentrations. Variables such as the feed concentration and the capacity/flowrate used in the synthetic adsorbent column were found to be especially critical in determining the level of yield and throughput. Ultimately, this type of knowledge facilitates Quality By Design by enabling a more efficient exploration of the operating space through fractional factorial design focusing upon only those variables deemed relevant by GSA.

The seventh chapter was divided into two parts; firstly, a simulation of the DigiFab™ affinity process was created, identifying the best conditions that achieved target levels for yield and throughput and it was shown that for this column, operating in the middle of the protein loading range (200-500 mg/mL) and at a flowrate that is 70% below the current level resulted in the best performance. Secondly, an experimental evaluation of a novel synthetic ligand adsorbent was used to show that it could achieve comparable or superior results to other commercially available matrices. The resin is capable of a high dynamic binding capacity (29.2 mg/mL), a consistently high yield (85%) over repeated cycles and comparable purification characteristics to another resin (Protein G) as shown by confocal microscopy. The resin may also have validity in applications to other feeds such as monoclonal antibodies. After this, the eighth chapter described the commercialisation and validation issues arising from the work.

9.2 Practical conclusions from the industrial case study

The key issues that have to be addressed by Protherics include the need to increase production levels, reduce manufacturing costs, minimise disruption to the existing facility from implementing CroFab process changes to the facility and speed up processing. In light of these and after application of the techniques proposed in this thesis to the operations carried out by Protherics, the following conclusions have been drawn:

- Increasing the feed volume and titre are generally beneficial, with the improvements in product mass outweighing factors such as additional processing times or development costs
- At the front end of the process, there are several options available for IgG capture. For example, replacing the existing precipitation and centrifugation steps during CroFab™ manufacture with column capture stages using synthetic ligands or Protein G matrices can help to improve processing times and costs. This conclusion is in line with the general development strategies of Protherics, which seeks to use column steps at the start of its polyclonal antibody processes. Optimising the capacities and flowrates of these steps is especially important to ensure adequate yield
- Changing the number of feed batches or venom components can affect cost of goods, times and resource usage and hence the effects of these changes need to be considered if implemented
- Eliminating downstream steps such as ion exchange columns also results in better processes that are faster and cheaper to operate

9.3 Future work

Research ideas that follow on from the work in this thesis include the following:

- More work is required in modelling individual unit operations based on fundamental engineering understanding e.g. kinetics, rather than the more generic mathematical approach that is used at

present. Ultimately, this will enable more robust modelling and so provide greater understanding of how to optimise a bioprocess

- The modelling approaches need to be extended to consider fermentation-based processes in order to make the techniques developed in this thesis more generically applicable
- More work is needed in simulating the campaign manufacture of multiple products, especially in light of the added burden that any newly developed candidates would place on a manufacturing facility already making approved products. The impacts of adding novel therapeutics to a pre-existing manufacturing portfolio would be evaluated in terms of factors such as resource availability or risks of a campaign overrunning
- Simulation work to date has concentrated upon scenario analysis i.e. testing strategies presented to the model. A better idea would be to use the model to work out the optimal operating strategy itself with respect to target levels for outputs such as resource utilisation, yield and throughput, purity and capital/production costs
- Whole process ultra scale-down mimics that could be developed include:
  a) Linking the synthetic affinity column to the remainder of the process, either with or without the ion exchange step
  b) Evaluating an alternative such as using an ion exchange step for IgG capture (with a caprylic acid wash to eliminate adsorbed albumin), followed by digestion and an affinity step
  c) Using immobilised papain instead of soluble enzyme
- The prototype adsorbent should be tested with feeds other than ovine sera to identify its potential wider applicability

REFERENCES

Abbas, A.K., Lichtman, A.H., Basic immunology: functions and disorders of the immune system, 2nd edition (2004), Saunders (imprint of Elsevier), Philadelphia, U.S.A, 3-4

Asenjo, J.A., Montagna, J.M., Vecchietti, A.R., Iribarren, O.A., Pinto, J.M., Strategies for the simultaneous optimisation of the structure and the process variables of a protein production plant, Computers and Chemical Engineering, 24 (2000), 2277-2290

Atkinson, B., Mavituna F., Principles of costing and economic evaluation for bioprocesses, Biochemical Engineering and Biotechnology handbook, 2nd edition (1991), Stockton Press, Basingstoke, U.K., 997–999; 1059-1109

Aubrey, N. Muzarda, J., Peter, J. C., Rochat, H., Goyffon, M., Devaux, C., Billiald, P., Engineering of a recombinant Fab from a neutralizing IgG directed against scorpion neurotoxin AahI, and functional evaluation versus other antibody fragments, Toxicon, 43 (2004), 233–241

Bailey, J.E., Ollis, D.F., Biochemical Engineering Fundamentals, Bioprocess Economics, 2nd edition (1986), McGraw Hill International, London, U.K., 798-849

Bak, H., Thomas, O.R.T., Evaluation of commercial chromatographic adsorbents for the direct capture of polyclonal rabbit antibodies from clarified antiserum, Journal of Chromatography B, 848, 1 (2007), 116-130

Bak, H., Thomas, O.R.T., Abildskov, J., Lumped parameter model for prediction of initial breakthrough profiles for the chromatographic capture of antibodies from a complex feedstock. Journal of Chromatography B, 848, 1 (2007), 131–141

Baker, S.J., Wheelwright, S.M., Financially Based Modelling of Recovery Process Alternatives, BioProcess International, May 2004, 42-54

Basu, P.K., Quaadgras, J., Mack, R.A., Noren, A.R., Achieve the right balance in pharmaceutical pilot plants, Chemical Engineering Progress, 94 (1998), 67 – 74

Bauer, H.H., Fischer, M., Product life cycle patterns for pharmaceuticals and their impact on R&D profitability of late mover products, International Business Review, 9 (2000) 703–725

Boychyn, M., Doyle, W., Bulmer, M., More, J., Hoare, M. (2000), Laboratory Scaledown of Protein Purification Processes Involving Fractional Precipitation and Centrifugal Recovery. Biotechnology and Bioengineering, 69, 1, 1–10

Boychyn, M., Yim, S. S. S., Ayazi Shamlou, P., Bulmer, M., More, J., Hoare, M., Characterization of flow intensity in continuous centrifuges for the development of laboratory mimics, Chemical Engineering Science, 56 (2001), 4759–4770

Boyer L.V., Seifert S.A., Cain J.S., Recurrence phenomena after immunoglobulin therapy for snake envenomations: Part 2. Guidelines for clinical management with crotaline F_{AB} antivenom, Annals of Emergency Medicine, 37, 2 (2001), 196-201

Brekke, O.H., Sandlie, I., Therapeutic Antibodies For Human Diseases At The Dawn Of The Twenty-First Century, Nature Reviews: Drug Discovery, 2 (2003), 52-62

Bruce, L.J., Chase, H.A., The combined use of in-bed monitoring and an adsorption model to anticipate breakthrough during expanded bed adsorption, Chemical Engineering Science, 57 (2002) 3085 – 3093

Chan, K., Saltelli, A., Tarantola, S., Sensitivity analysis of model output: Variance based methods make the difference, Proceedings of the 1997 Winter Simulation Conference

Chan, S.H., Kiang, S., Brown, M.A., One-Dimensional Centrifugation Model, AIChE Journal, 49, 4 (2003), 925-938

Chase, H.A., The use of affinity adsorbents in expanded bed adsorption, Journal of Molecular Recognition, 11 (1998), 217-221

Chhatre, S., Jones, C., Francis, R., O'Donovan, K., Titchener–Hooker, N. J., Newcombe, A. R., Keshavarz–Moore, E., The integrated simulation and assessment of the impacts of process change in biotherapeutic antibody production, Biotechnology Progress, 22 (2006), 1612-1620

Chhatre, S., Francis, R., O'Donovan, K., Titchener-Hooker, N.J., Newcombe, A.R., Keshavarz-Moore, E., A decision-support model for evaluating changes in biopharmaceutical manufacturing processes, Bioprocess and Biosystems Engineering, 30 (1) (2007a), 1-11

Chhatre, S., Titchener-Hooker, N.J., Newcombe, A.R., Keshavarz-Moore, E., Purification of antibodies using the synthetic affinity ligand absorbent MAbsorbent A2P, Nature Protocols, 2, 7 (2007b), 1763-1769

Chhatre, S., Francis, R., O'Donovan, K., Titchener–Hooker, N.J., Newcombe, A.R., Keshavarz–Moore, E., A prototype software methodology for the rapid evaluation of bio–manufacturing process options, Biotechnology and Applied Biochemistry, 48 (2007c), 65–78

Chhatre et al, Global Sensitivity Analysis for the determination of parameter importance in bio-manufacturing processes, Biotechnology and Bioengineering 2007d, in preparation

Chhatre et al, Global Sensitivity Analysis for the determination of, parameter importance in the chromatographic purification of polyclonal antibodies, Journal of Chemical Technology and Biotechnology 2007e, submitted

Chhatre, S., Francis, R., Titchener–Hooker, N.J., Newcombe, A.R., Keshavarz–Moore, E., Evaluation of a novel agarose-based synthetic ligand adsorbent for the recovery of antibodies from ovine serum, Journal of Chromatography B 2007f, submitted

Chhatre, S., Thillaivinayagalingam, P., Francis, R., O'Donovan, K., Titchener-Hooker, N.J., Newcombe, A.R., Keshavarz-Moore, E., Decision-support software for the industrial-scale chromatographic purification of antibodies, Biotechnology Progress, 23 (2007g), 888-894

Chowdhury, B.R., Chakraborty, Chaudhuri, U.R., Modelling and simulation of diffusional mass transfer of glucose during fermentative production of pediocin AcH from Pediococcus acidilactici H, Biochemical Engineering Journal, 16 (2003), 237-243

Clarkson, A.I., Bulmer, M., Titchener-Hooker, N.J., Pilot-scale verification of a computer-based simulation for the centrifugal recovery of biological particles, Bioprocess and Biosystems Engineering, 14, 2 (1996), 81-89

Clewlow, P., Drugs in transit, Chemistry and Industry, 12 (2004), 22-24

Coico, R., Sunshine, G., Benjamini, E., Immunology: a short course, 5th edition (2003), John Wiley and Sons Inc., Hoboken, New Jersey, U.S.A.

Coleman, L., Mahler, S., Purification of Fab fragments from a monoclonal antibody papain digest by Gradiflow electrophoresis, Protein Expression and Purification, 32 (2003), 246-251

Cresswell, C., Newcombe, A.R., Davies, S., Macpherson, I., Nelson, P., O'Donovan, K., Francis, R. (2005), Optimal conditions for the papain digestion of polyclonal ovine IgG for the production of biotherapeutic Fab fragments. Biotechnology and Applied Biochemistry, 42, 163–168

Davies, H., Introductory Immunobiology, Recognition of pathogens, 1997, Chapman and Hall, 44-45

Deeds, D.L., Decarolis, D., Coombs, J., Dynamic capabilities and new product development in high technology ventures: An empirical analysis of new biotechnology firms, Journal of Business Venturing, 15, 3, (2000), 211-229

Deschaine, L.M., Francone, F.D. (2005). Extending the Boundaries of Design Optimisation by Integrating Fast Optimisation Techniques with Machine Code Based, Linear Genetic Programming, in: Graňa, M., Duro, R., d'Anjou, A., Wang, P.P., (Eds.), Information Processing with Evolutionary Algorithms. Springer–Verlag London Limited, London, pp. 13–14

DiMasi, J.A., Hansen, R.W., Grabowski, H.G., The price of innovation: new estimates of drug development costs, Journal of Health Economics, 22 (2003), 151–185

Doblhoff-Dier, O., Bliem, R., Quality control and assurance from the development to the production of biopharmaceuticals. Trends in Biotechnology, 17 (1999), 266–270

Drews, J., Ryser, S., Pharmaceutical innovation between scientific opportunities and economic constraints, Drug Discovery Today, 2, 9 (1997), 365-372

Ernst, S., Garro, O.A., Winkler, S., Venkataraman, G., Langer, R., Cooney, C.L., Sasisekharan, R., Process simulation for recombinant protein production: Cost estimation and sensitivity analysis for heparinase I expressed in Escherichia coli, Biotechnology and Bioengineering, 53, 6 (1997), 575-582

Fahrner, R.L., Blank, G.S., Zapata, G.A., Expanded bed protein A affinity chromatography of a recombinant humanized monoclonal antibody: process development, operation, and comparison with a packed bed method, Journal of Biotechnology, 75 (1999), 273–280

Farid, S., Novais, J.L., Karri, S., Washbrook, J., Titchener-Hooker, N.J., A Tool for Modelling Strategic Decisions in Cell Culture Manufacturing, Biotechnology Progress, 16 (2000a), 829-836

Farid, S., Washbrook, J., Birch, J., Titchener-Hooker, N., A Hierarchical Framework for Modelling Biopharmaceutical Manufacture to Assess Process and Business Needs, European Symposium on Computer Aided Process Engineering, 10 (2000b), 673-678

Farid, S. A decisional support tool for simulating the process and business perspectives of biopharmaceutical manufacture, PhD thesis submitted to the University of London, 2001

Farid, S., Washbrook, J., Titchener-Hooker, N., Decision-support tool for risk analysis in biopharmaceutical manufacture, CAB, 8 (2001), 167-171

Farid, S.S., Washbrook, J., Titchener–Hooker, N.J., Combining Multiple Quantitative and Qualitative Goals When Assessing Biomanufacturing Strategies under Uncertainty, Biotechnology Progress, 21 (2005), 1183–1191

Farid, S.S., Process economics of industrial monoclonal antibody manufacture, Journal of Chromatography B, 848 (2007), 8–18

Fee, C.J., Economics of wash strategies for expanded bed adsorption of proteins from milk with buoyancy-induced mixing, Chemical Engineering and Processing, 40 (2001), 329-334

Flatman, S., Alam, I., Gerard, J., Mussa, N., Process analytics for purification of monoclonal antibodies, Journal of Chromatography B, 848 (2007) 79-87

Fogel, D. B. (2006). Evolutionary Computation. New Jersey: John Wiley and Sons Incorporated

Gebicke, K.W., Johl, H.-J., Sternad, W., Trösch, W., Chmiel, H., Application of modelling and simulation for optimisation of a continuous fermentation process, Computers and Chemical Engineering, 17, supplement 1 (1993), 177-182

George, E., Titchener-Hooker, N.J., Farid, S.S., A multi-criteria decision-making framework for the selection of strategies for acquiring biopharmaceutical manufacturing capacity, Computers and Chemical Engineering 31 (2007) 889–901

Giannasi, F., Lovett, P., Godwin, A.N., Enhancing confidence in discrete event simulations, Computers in Industry, 44 (2001), 141-157

González, Y., Ibarra, N., Gómez, H., González, M., Dorta, L., Padilla, S., Valdés, R., Expanded bed adsorption processing of mammalian cell culture fluid: comparison with packed bed affinity chromatography, Journal of Chromatography B, 784 (2003) 183–187

Groep, M.E., Gregory, M.E., Kershenbaum, L.S., Bogle, I.D.L., Performance Modelling and Simulation of Biochemical Process Sequences with Interacting Unit Operations, Biotechnology and Bioengineering, 67, 3 (2000), 300-311

Harrison R G, Todd P W, Rudge S R, Petrides D P Bioprocess Design. In Bioseparations science and engineering, 1st edition (2003), Oxford University Press, New York, U.S.A., 319-372

Hausner, B., Outsourcing: evaluating alternatives, Pharmaceutical Science & Technology Today, 1, 4 (1998), 148-152

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

Holford, N.H.G., Kimko, H.C., Monteleone, J.P.R., Peck, C.C., Simulation of clinical trials, Annual Review of Pharmacology and Toxicology, 40 (2000), 209-234

Holland, J.H. (1973), Genetic algorithms and the optimal allocation of trials. Society for Industrial and Applied Mathematics: Journal of Computation, 2, 88–105

Huse, K., Böhme, H-J., Scholz, G.H., Review article: Purification of antibodies by affinity chromatography, Journal of Biochemical and Biophysical Methods, 51 (2002) 217–231

Hwang, Ch.-L., Yoon, K., Lecture Notes in Economics and Mathematical Systems volume 186: *Multiple Attribute Decision Making – Methods and Applications* A State-of-the-Art Survey (1981), Springer-Verlag, Berlin, Germany, 99-102

Atkinson, B., Mavituna F., Principles of costing and economic evaluation for bioprocesses, Biochemical Engineering and Biotechnology handbook, 2nd edition (1991), Stockton Press, Basingstoke, U.K., 997–999; 1059-1109

Jang, J.D., Barford, J.P., An unstructured kinetic model of macromolecular metabolism in batch and fed-batch cultures of hybridoma cells producing monoclonal antibodies, Biochemical Engineering Journal, 4 (2000), 153-168

Joseph, J.R., Sinclair, A., Titchener-Hooker, N.J., Zhou, Y., A framework for assessing the solutions in chromatographic process design and operation for large-scale manufacture, Journal of Chemical Technology and Biotechnology, 81 (2006), 1009–1020

Juckett, G., Hancox, J.G., Venomous Snakebites in the United States: Management Review and Update, American Family Physician, 65 (2002), 1367-1374

Kaczmarski, K., Antos, D., Sajonz, H., Sajonz, P., Guichon, G., Comparative modelling of breakthrough curves of bovine serum albumin in anion exchange chromatography. Journal of Chromatography A, 2001, 925, 1–17

Karri, S., Davies, E., Titchener-Hooker, N., Washbrook, J., Biopharmaceutical Process Development: Part III, A Framework to Assist Decision Making, Biopharm Europe, September 2001a, 76-82

Karri, S., Foo, F., Titchener-Hooker, N.J., Dunnill, P., Biopharmaceutical process development: Part I: Information from the first product generation, Biopharm Europe, June 2001b, 58-64

King, J.M.P., Griffiths, P., Zhou, Y., Titchener-Hooker, N.J., Visualising bioprocesses using, '3D-Windows of Operation.', Journal of Chemical Technology and Biotechnology, 79 (2004), 518-525

King, J.M.P., Titchener-Hooker, N.J., Zhou, Y., Ranking bioprocess variables using global sensitivity analysis: a case study in centrifugation, Bioprocess and Biosystems Engineering, 30 (2007), 123 – 134

Lamping, S., Zhang, H., Allen, B., Ayazi Shamlou, P., Design of a prototype miniature bioreactor for high throughput automated bioprocessing, Chemical Engineering Science, 58, 3–6, February – March 2003, 747-758

Latham, P., Sinclair, A., Computer modelling for biomanufacturing, Presentation given to the Modular Training for the Bioprocess Industries module, University College London, May 2004

Levin, A.A., Papageorgiou, L.G., A hierarchical solution approach for multi-site capacity planning under uncertainty in the pharmaceutical industry, Computers and Chemical Engineering, 28 (2004), 707 – 725

Li, Z., Gu, Y., Gu, T., Mathematical modelling and scale-up of size exclusion chromatography, Biochemical Engineering Journal, 2, 2 (1998), 145-155

Li, R., Townsend, C.A., Rational strain improvement for enhanced clavulanic acid production by genetic engineering of the glycolytic pathway in Streptomyces clavuligerus, Metabolic Engineering, 8, (2006), 240-252

Lim, A. C., Zhou, Y., Washbrook, J., Titchener-Hooker, N.J., Farid, S., A decisional-support tool to model the impact of regulatory compliance activities in the biomanufacturing industry, Computers and Chemical Engineering, 28 (2004), 727-735

Lim, A. C., Zhou, Y., Washbrook, J., Sinclair, A., Fish, B., Francis, R., Titchener-Hooker, N. J., Farid, S. S., Application of a Decision-Support Tool to Assess Pooling Strategies in Perfusion Culture Processes under Uncertainty, Biotechnology Progress (2005), 1231 – 1242

Lim, A.C., Washbrook, J., Titchener–Hooker, N.J., Farid, S.S. (2006), A Computer–Aided Approach to Compare the Production Economics of Fed–Batch and Perfusion Culture Under Uncertainty. Biotechnology and Bioengineering, 93, 4, 687–697

Lowe, C.R., Lowe, A.R., Gupta, G., New developments in affinity chromatography with potential application in the production of biopharmaceuticals, Journal of Biochemical and Biophysical Methods, 49, 1-3 (2001), 561-574

Luo, Q., Mao, X., Kong, L., Huang, X., Zou, H., High-performance affinity chromatography for characterisation of human immunoglobulin G digestion with papain, Journal of Chromatography B, 776 (2002), 139 – 147

Mackean, D., Jones, B., Immunity and Immunisation, Introduction to human and social biology, 2nd ed., John Murray Limited, 252-253, 1985

Maybury, J.P., Mannweiler, K., Titchener-Hooker, N.J., Hoare, M., Dunnill, P., The performance of a scaled down industrial disk stack centrifuge with a reduced feed material requirement, Bioprocess Engineering, 18 (1998), 191-199

McGill, T., Klobas, J.E., The role of spreadsheet knowledge in user-developed application success, Decision Support Systems (2004), 39, 3 (2005), 355-369

McRae, G.J., Tilden, J.W., Seinfeld, J.H., Global Sensitivity Analysis – a computational implementation of the Fourier Amplitude Sensitivity Test, Computers and Chemical Engineering, 6, 1 (1982), 15-25

Montagna, J.M., Iribarren, O.A., Galiano, F.C., The design of multiproduct batch plants with process performance models, Chemical and Engineering Research Design, 72 (1994), 783-791

Montagna, J.M., Vecchietti, A.R., Iribarren, O.A., Pinto, J.M., Asenjo, J.A., Optimal Design of Protein Production Plants with Time and Size factor Process Models, Biotechnology Progress, 16 (2000), 228-237

Montesinos, R.M., Tejeda–Mansir, A., Guzman, R., Ortega, J., Schiesser, W.E. (2005), Analysis and simulation of frontal affinity chromatography of proteins. Separation and Purification Technology, 42, 75–84

Mustafa, M.A., Washbrook, J., Lim, A.C., Zhou, Y., Titchener–Hooker, N.J., Morton, P., Berezenko, S., Farid, S.S. (2004), A Software Tool to Assist Business–Process Decision–Making in the Biopharmaceutical Industry. Biotechnology Progress, 20, 1096 – 1102 [Correction: Biotechnology Progress, 21 (2005), 320]

Mustafa, M. A., Washbrook, J., Titchener-Hooker, N.J., Farid, S.S., Retrofit decisions within the biopharmaceutical industry: An EBA Case Study, Food and Bioproducts Processing, 84, C1 (2006), 84–89

Neal, G., Christie, J., Keshavarz-Moore, E., Ayazi Shamlou, P., Ultra Scale-Down Approach for the prediction of Full-Scale recovery of Ovine Polyclonal Immunoglobulins Used in the Manufacture of Snake Venom-Specific Fab Fragment, Biotechnology and Bioengineering, 81 (2003), 149-157

Neal, G., Francis, R., Ayazi Shamlou, P., Keshavarz-Moore, E., Separation of immunoglobulin G precipitate from contaminating proteins using microfiltration, Biotechnology and Applied Biochemistry, 39 (2004), 241-248

Neal, G., Integrated Scale Down Of The Primary Downstream Purification Process And Alternative Process Options For The Production Of An Ovine Polyclonal Antibody Based Anti-Venom, EngD thesis submitted to the University of London, 2005

Newcombe, A.R., Cresswell, C., Davies, S., Watson, K., Harris, G., O'Donovan, K., Francis, R., Optimised affinity purification of polyclonal antibodies from hyper immunised ovine serum using a synthetic Protein A adsorbent, MAbsorbent® A2P, Journal of Chromatography B, 814 (2005), 209–215

Newcombe, A.R., Cresswell, C., Davies, S., Pearce, F., O'Donovan, K., Francis, F., Evaluation of a biosensor assay to quantify polyclonal IgG in ovine serum used for the production of biotherapeutic antibody fragments, Process Biochemistry (2006), 41, 842–847

Newcombe, C., Newcombe, A.R., Antibody production: Polyclonal-derived biotherapeutics, Journal of Chromatography B, 848 (2007) 2–7

Novais, J.L., Titchener-Hooker, N.J., Hoare, M., Economic Comparison Between Conventional and Disposables-Based Technology for the Production of Biopharmaceuticals, Biotechnology and Bioengineering, 75 (2001), 143-153

Nortcliffe, A.L., Thompson, M., Shaw, K.J., Love, J., Fleming, P.J., A framework for modelling in S88 constructs for scheduling purposes, ISA Transactions 40 (2001), 295-305

Pandey, M., Investment decisions in pharmaceutical R&D projects, Drug Discovery Today, 8, 21 (2003), 968-971

Petrides, D., Cooney, C.L., Evans, L.B., Field, R.P., Snoswell, M., Bioprocess simulation: an integrated approach to process development, Computers and Chemical Engineering, 13, 4/5 (1989), 553-561

Petrides, D., Biopro Designer: An advanced computing environment for modelling and design of integrated biochemical processes, Computers and Chemical Engineering, 18 (*supplement* - 1994), S621 – S625

Petrides, D., Koulouris, A., Siletti, C., Throughput Analysis and Debottlenecking of Biomanufacturing Facilities: A Job for Process Simulators, BioPharm, August 2002a, 2-7

Petrides, D.P., Koulouris, A., Lagonikos, P.T., The Role of Process Simulation in Pharmaceutical Process Development and Product Commercialisation, Pharmaceutical Engineering, 22, 1 (2002b), 1-8

Pinto, J.M., Montagna, J.M., Vecchietti, A.R., Iribarren, O.A., Asenjo, J.A., Process Performance Models in the Optimization of Multiproduct Protein Production Plants, Biotechnology and Bioengineering, 74, 6 (2001), 451-465

Pisano, G., Wheelwright, S. C., High-tech R&D, Harvard Business Review, 73 (1995), 93-105

Pisano, G., Learning-before-doing in the development of new process technology, Research Policy, 25 (1996), 1097-1119

Porter, R.R., The hydrolysis of rabbit γ-globulin and antibodies with crystalline papain, Biochemistry Journal, 73 (1959), 119-126

Rajapakse, A., Titchener-Hooker, N.J., Farid S.S., Modelling of the biopharmaceutical drug development pathway and portfolio management, Computers and Chemical Engineering, 29 (2005) 1357–1368

Rajapakse, A., Titchener-Hooker, N.J., Farid S.S., Integrated approach to improving the value potential of biopharmaceutical R&D portfolios while mitigating risk, Journal of Chemical Technology and Biotechnology, 81 (2006), 1705–1714

Reynolds, T., Boychyn, M., Sanderson, T., Bulmer, M., More, J., Hoare, M., Scale-Down of Continuous Filtration for Rapid Bioprocess Design: Recovery and Dewatering of Protein Precipitate Suspensions, Biotechnology and Bioengineering, 83, 4 (2003), 454–464

Richard, A., Margaritis, A., Empirical Modelling of Batch Fermentation Kinetics for Poly(glutamic acid) Production and Other Microbial Biopolymers, Biotechnology and Bioengineering, 87, 4 (2004), 501-515

Rooney, K.F., Snoeck, E., Watson, P.H., Modelling and simulation in clinical drug development, Drug Discovery Today, 6, 15 (2001), 802 – 806

Roque, A. C. A., Lowe, C. R., Taipa, M. A., Antibodies and Genetically Engineered Related Molecules: Production and Purification, Biotechnology Progress, 20 (2004), 639-654

Rouf, S.A., Douglas, P.L., Moo-Young, M., Scharer, J.M., Computer simulation for large scale bioprocess design, Biochemical Engineering Journal, 8 (2001), 229-234

Salomone, H.E., Iribarren, O.A., Posynomial modelling of batch plants, Computers and Chemical Engineering, 16 (1992), 173-184

Saltelli, A., Bolado, R., An alternative way to compute the Fourier amplitude sensitivity test (FAST), Computational Statistics and Data Analysis, 26 (1998), 445 – 460

Saltelli A, Chan K and Scott EM, Sensitivity Analysis, Wiley Series in Probability and Statistics, Wiley, U.K. (2000)

Samsatli, N.J., Shah, N., Optimal integrated design of biochemical processes, Computers and Chemical Engineering, 20, supplemental (1996), S315-320

Schaibly, J.H., Shuler, K.E., Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients, II Applications, The Journal of Chemical Physics, 59, 8 (1973)

Schmid, E.F., Smith, D.A., Discovery, innovation and the cyclical nature of the pharmaceutical business, Drug Discovery Today, 7, 10 (2002), 563-568

Shah, N., Pharmaceutical supply chains: key issues and strategies for optimisation. Computers and Chemical Engineering, 28 (2004), 929–941

Shanklin, T., Roper, K., Yegneswaran, P.K., Marten, M.R. (2001), Selection of Bioprocess Simulation Software for Industrial Applications. Biotechnology and Bioengineering, 72, 4, 483–489

Shillingford, C.A., Vose, C.W., Effective decision-making: progressing compounds through clinical development, Drug Discovery Today, 6, 18 (2001), 941-946

Shukla, A.A., Hubbard, B., Tressel, T., Guhan, S., Low, D., Downstream processing of monoclonal antibodies–Application of platform approaches, Journal of Chromatography B 848 (2007), 28-39

Siddiqi, S. F., Titchener-Hooker, N.J., Ayazi Shamlou, P., Simulation of particle size distribution changes occurring during high-pressure disruption of bakers' yeast, Biotechnology and Bioengineering, 50, 2 (1996), 145-150

Sinnott, R.K., Chemical Engineering, 6: An introduction to chemical engineering design, Eds. Coulson, J.M. and Richardson, J.F., 1993, Pergamon Press, Oxford, U.K.

Skerra, A., Imitating the humoral immune response, Current Opinion in Chemical Biology, 7 (2003), 683 – 693

Skrepnek, G.H., Accounting- Versus Economic-Based Rates of Return: Implications for Profitability Measures in the Pharmaceutical Industry, Clinical Therapeutics, 26, 1 (2004), 155-174

Sobol', I.M., Sensitivity Estimates for Nonlinear Mathematical Models, MMCE, 1, 4 (1993), 407-414

Sobol', I.M., Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Mathematics and Computers in Simulation, 55 (2001), 271–280

Spalding, B.J., Downstream processing: key to slashing production costs 100 fold, Bio/technology, 9 (1991), 229-233

Steffens, M.A., Fraga, E.S., Bogle, I.D.L., Multicriteria process synthesis for generating sustainable and economic bioprocesses, Computers and Chemical Engineering, 23 (1999), 1455-1467

Steuer, R.E., Na, P., Multiple criteria decision making combined with finance: A categorized bibliographic study, European Journal of Operational Research, 150, 3 (2003), 496-515

Svarovsky, L., Solid–Liquid separation, Separation by centrifugal sedimentation, 4th ed., Oxford: Butterworth–Heinemann (2000), 264–266

Swinnen, K., Krul, A., Van Goidsenhoven, I., Van Tichelt, N., Roosen, A., Van Houdt, K., Performance comparison of protein A affinity resins for the purification of monoclonal antibodies, Journal of Chromatography B, 848, 1 (2007), 97-107

Tejeda-Mansir, A., Espinoza, R., Montesinos, R.M., Guzman, R., Modelling regeneration effects on protein A affinity chromatography, Bioprocess Engineering, 17 (1997), 39-44

Terzi, S., Cavalieri, S., Simulation in the supply chain context: a survey, Computers in Industry, 53 (2004), 3–16

Thillaivinayagalingam, P., O'Donovan, K., Newcombe, A. R., Keshavarz-Moore, E., Characterisation of an industrial affinity process used in the manufacturing of digoxin-specific polyclonal Fab fragments, Journal of Biotechnology, 848, 1 (2007), 88-96

Titchener-Hooker N. J., Zhou Y., Hoare M., Dunnill P., Biopharmaceutical Process Development: Part II, Methods of Reducing Development Time, Biopharm Europe (2001), September: 68–74

Torphy, T.J., Monoclonal antibodies: boundless potential, daunting challenges, Current Opinion in Biotechnology, 13 (2002), 589–591

Tranter, D., Preparing for change, Pharmaceutical Science & Technology Today, 2, 3 (1999), 89-90

Ungarala, S., Co., T.B., Time-varying system identification using modulating functions and spline models with application to bioprocesses, Computers and Chemical Engineering, 24 (2000), 2739-2753

Varga, E.G., Titchener–Hooker, N.J., Dunnill, P. (1998), Use of scale–down methods to rapidly apply natural yeast homogenisation models to a recombinant strain. Bioprocess Engineering, 19, 373–380

Varga, E.G., Titchener–Hooker, N.J., Dunnill, P. (2001), Prediction of the Pilot–Scale Recovery of a Recombinant Yeast Enzyme Using Integrated Models. Biotechnology and Bioengineering, 74, 96–107

Venkatasubramanian, V., Prognostic and diagnostic monitoring of complex systems for product lifecycle management: Challenges and opportunities, Computers and Chemical Engineering 29 (2005), 1253–1263

Volgenant, A., Solving some lexicographic multi-objective combinatorial problems, European Journal of Operational Research, 139 (2002), 578–584

Vunnum, S.; Gallant, S.R.; Kim, Y.J.; Cramer, S.M. Immobilised metal affinity chromatography: modelling of nonlinear multicomponent equilibrium. Chemical Engineering Science 1995, 50, 11, 1785–1803

Watson, D.R., Dzirbik, D., Rudolph, J., Joglekar, G.S., The integration of process simulation in operational analysis, Computers and Chemical Engineering, 24 (2000), 607-612

Watt, G.M., Lund, J., Levens, M., Kumar Kolli, V.S., Jefferis, R., Boons, G-J, Site-Specific Glycosylation of an Aglycosylated Human IgG1-Fc Antibody Protein Generates Neoglycoproteins with Enhanced Function, Chemistry & Biology, 10 (2003), 807–814

Williams, H.P., Model building in mathematical programming, 4th edition, Chichester, Wiley-Interscience (1999)

Wong, H.H., O'Neill, B.K., Middelberg, A.P.J., A mathematical model for *Escherichia coli* debris size reduction during high pressure homogenisation based on grinding theory, Chemical Engineering Science, 52, 17 (1997), 2883-2890

Woodley, J., Titchener-Hooker, N.J., The use of windows of operation as a bioprocess design tool, Bioprocess Engineering, 14 (1996), 263-268

Workman, R.W., Simulation of the drug development process: A case study from the pharmaceutical industry, Proceedings of the 2000 Winter Simulation Conference

Yavuz, H., Akgöl, S., Say, R., Denizli, A., Affinity separation of immunoglobulin G subclasses on dye attached poly(hydroxypropyl methacrylate) beads, International Journal of Biological Macromolecules, 39 (2006), 303-309

Zhou, Y.H., Holwill, I.L.J., Titchener-Hooker, N.J., A study of the use of computer simulations for the design of integrated downstream processes, Bioprocess Engineering, 16 (1997), 367-374

Zhou, Y.H., Titchener-Hooker, N.J., Simulation and optimisation of integrated bioprocesses: a case study, Journal of Chemical Technology and Biotechnology, 74 (1999a), 289-292

Zhou, Y.H., Titchener-Hooker, N.J., Visualising Integrated Bioprocess Designs Through, "Windows of Operation," Biotechnology and Bioengineering, 65, 5 (1999b), 550-557

Zhou, Y.H., Titchener–Hooker, N.J. (2003), The application of a Pareto optimisation method in the design of an integrated bioprocess. Bioprocess and Biosystems Engineering, 25, 349–355

APPENDIX 1: EVALUATION OF SOFTWARE ENVIRONMENTS

The following gives an outline of the qualitative assessment of each simulation package assessed, which formed the basis of the scores given in Table 3 (Page 44).

Analytica (Lumina Decision Systems - Los Gatos, California, U.S.A.)

- Simple model construction – graphical with dragging and dropping of icons to create the objects on the screen
- Pre-programmed modules would aid swift replication of the CroFab™ process
- Run time debug is allowed – hence parameters could be changed, 'on the fly'
- More attuned to the business aspects of modelling, with risk analysis and decision making being at the core of the program
- Very low computational burden – this software should therefore run fairly quickly
- Includes a reasonably powerful package for analysing output data
- Does not provide any tools to allow dissemination of the final simulation to the end user who may not have this piece of modelling software
- Provides methods of evaluating the effects of risk and uncertainty within the process setup

Anylogic (XJ Technologies - St. Petersberg, Russian Federation)

- Graphical structure, with pre-programmed modules and run time debug available
- Predictive piece of software – useful in answering, 'what if,' scenarios
- Supports a large number of experimental types e.g. Monte Carlo and sensitivity analysis
- Includes a wide range of statistical functions that can be used to operate on output values, as well as graphing and reporting utilities
- The complete model can also be packaged such that users who do not possess Anylogic software can still look at the simulation
- The program is very large, requiring a large amount of memory (512 MB recommended) and can only run on Microsoft® Windows® 2000 and XP. If an inferior computer specification was used, the modelling could prove slow and inefficient
- Not specifically targeted at a manufacturing process and so many need a fair amount of work to modify the existing pre-programmed modules into a useful form

Arena (Rockwell Software - Warrendale, Pennsylvania, U.S.A.)

- A reasonably powerful product, possessing a high amount of functionality
- A product that is targeted far more at manufacturing processes than Anylogic
- Does not require too much memory (64 MB) and can run on a number of different operating systems from Microsoft® Windows® 95 upwards i.e. the software should run fairly quickly

- Dragging and dropping of icons for object creation
- Pre-programmed and user defined modules available; run time debug allowed
- Completed models can be packaged such that any users who do not have Arena can look at the simulation
- Lacks any significant output data statistical analysis tools; sensitivity analysis and Monte Carlo setups are not available either

Crystal Ball (Decisioneering Incorporated - London, U.K.)

- Targeted at the process industries (pharmaceutical companies specifically listed as a key target market for the software), but primarily involved in business planning analysis, evaluation of costs and benefits, risk and project management, rather than manufacturing simulation
- Create turnkey applications with this Microsoft® Excel add in – Excel models enhanced with Crystal Ball can be viewed by anyone with Excel
- Very small memory requirements and can easily run on a variety of Windows platforms – important if the model is moved from UCL to Protherics
- Possesses a tool that allows automatic optimisation of a given model and transfer back into an Excel spreadsheet (this can lead to excessive dependency by the user if too much trust is placed in the model's optimisation routines)
- Use of automatic optimisation routines also assumes that the modelling software provides the best fit that the user would wish to see to a given problem – if it does not and if this is not noticed by a end user who is not well versed to the technology, this inaccuracy may be carried through to another part of the model
- Provides a method for forecasting of future outcomes based on historical data
- Can run multiple simulations simultaneously to test different assumptions
- Dragging and dropping of icons is not available = rather difficult to model
- No pre-programmed modules available: run time debug disallowed

Decision Pro (Vanguard Software Corporation - Cary, North Carolina, U.S.A.)

- More attuned to business management decisions than manufacturing objectives
- Strategic analysis in relation to, for example, competitive threat, financial outlook and marketing are all covered
- Decision tree analysis and Monte Carlo simulation are present
- Some operational modelling can be done in relation to process simulation, optimisation and scheduling
- Only moderate amounts of processing power are needed
- Graphical construction, with models being represented in an easy to visualise hierarchical tree format
- Both pre-programmed modules and simulation run time debug are present

- Graphical presentation of simulation statistics and results of sensitivity analysis
- Models can easily be distributed to end users

emPower (Technomatix Technologies Incorporated - Surrey, U.K.)

- Designed to model a number of different scenarios, from manufacturing and material handling through to scheduling
- Very high computer requirements (a recommended 512 MB of memory). If the software is upgraded, then this requirement is likely to increase, potentially beyond the specification of the computer
- Graphical construction (as opposed to using a programming language to create the model), pre-programmed modules and run time debug are present
- Provides a statistical data tool for analysis of output values
- Technology allows for dissemination of model to end users
- Allows for optimisation of material flows, plant resource utilization and planning
- Object-oriented technology and object libraries that are present can be readily customized, whilst user defined objects are also permitted
- Object orientation allows for creation of hierarchical trees to aid modelling
- Provides tools for, 'what if,' scenarios as well as charts to evaluate graphically differing manufacturing strategies or test proposed changes to given process parameters
- Primarily focused on non process industry related modelling and its intrinsic objects reflect this – such as its use for car manufacture. Although objects for Protherics' model could be created from scratch and added in, there are other types of modelling software available providing ready to use objects, targeted at the pharmaceutical industry

Enterprise Dynamics (Incontrol Enterprise Dynamics - Cheshire, U.K.)

- Contains a dedicated simulation suite for manufacturing and material handling
- Does not require significant quantities of computational power
- Graphical construction/pre-programmed modules/run time debug are present
- Capacity for graphing and reporting the results of simulations, but this does not seem to be as extensive as for some of the other packages
- Object oriented programming is present to allow for class inheritance
- Claims to be able to integrate with other software systems – could be useful if development in a particular part of the project can be carried out more readily in another development environment – this could allow seamless integration of any existing models into this project

Extend™ (Imagine That Incorporated - San Jose, California, U.S.A.)

- Good for modelling large scale simulations in manufacturing in a similar fashion to that envisaged for this project

- Moderate demands for computational power
- Drag and drop/pre-programmed modules/ run time debug are present
- Supports execution of a range of scenarios
- Technology allows for dissemination of model to end users
- Allows for modification or addition of objects as well as their reuse
- Programming language/graphical simulator hybrid
- Graphically represents the flow of materials and associated property values and hence leads to an interface that is relatively straightforward to understand
- Allows for data stored in an Excel spreadsheet to be picked up and used

Flexsim (Flexsim Software Products Incorporated - Orem, Utah, U.S.A.)

- Suited to manufacturing and material handling, as well as a wide range of other functions, including transport and storage of products
- Software is able to integrate modelling of manufacturing, logistics and administration (i.e. can provide a business and manufacturing viewpoint)
- Fairly heavy computational load
- Seems to be quite a complicated program to learn
- Graphical and hierarchical construction provide a class structure that allows for property and variable values inheritance from higher order entities
- Objects within the environment can be customised
- Code reuse can be readily achieved
- Model can be converted into its own executable
- Incorporates C++ into a graphical modelling environment. Hence C++ can be used directly in defining model logic and compiled into the application – providing the power to change things when such cannot be done graphically
- Scenario analyser can provide answers to, 'what if,' questions and information gathered can be automatically put out into the reports, tables and graphs
- A fairly extensive data analysis package allows use of pre-defined and user-defined indicators (e.g. utilization, throughput, lead-time and costs) allowing for analysis of performance of each scenario
- Data import or export from/to Microsoft® Word or Excel is possible if a data source is available – especially useful to Protherics' case if that data can simply be picked up from an Excel spreadsheet

GPSS World for Windows (Minuteman Software)

- Seems more attuned to modelling telecommunications systems, although it claims a general purpose capacity for modelling manufacturing

- Low requirements for computer memory or hard disk space (and hence can run multiple concurrent simulations fairly quickly if needed)
- Graphical user interface is provided, along with pre-programmed modules
- Some data analysis available (but less extensive than for other programs)
- Provides no tools to support model distribution to end user
- Own proprietary programming language (which would need to be learnt from scratch)
- Object oriented interface with many 'windows' onto the process available

Pasion Simulation System (Stansilaw Raczynski - Mexico D.F.Mexico)

- Universities listed specifically as part of targeted audience i.e. may be suited to students not especially familiar with such simulation software
- Fairly low computational needs
- Graphical construction is present, but run time debug is not allowed
- Fairly extensive data analysis is provided
- Reuse of coded elements seems to be limited and there are no tools for packaging the model into a form that is accessible by the end user
- Object oriented, but based on Pascal (author has no experience with this language)
- Presence of inheritance mechanism so as to allow addition of new properties to existing processes without affecting previously declared elements

PIMSS (MJC2 Limited - Crowthorne, Berkshire, U.K.)

- Main focus is on scheduling and planning of complicated manufacturing operations, producing detailed statistics and graphs on resource usage and labour requirements
- Extremely memory intensive – likely to be rather slow to run
- *Pre-programmed modules/run time debug are all available*
- Reasonably friendly graphical user interface
- A limited amount of graphical analysis of output information
- No tools for model packaging to end user
- Process costing is not a prominent part of the program

ProVision (Proforma Corporation - Southfield, Michigan, U.S.A.)

- Seems less focused on manufacturing and more on telecommunications and administration based projects
- Medium to heavy load on the computer
- Simple, graphical construction, but there are no pre-programmed modules, nor is runtime debug allowed
- Extensive charting functionality for usage of money, time and resources
- Full user packaging features for the model

- A highly structured and object orientated system
- Allows extensive, "what if," style testing of scenarios and storing/reuse of process objects
- Includes the capacity to look at the probabilities of using manufacturing strategies and assess them using the Monte Carlo simulation system etc
- Has a system for identifying process improvements, along with supporting costs, benefits, risks and required resources
- Can allow you to see whether a change in the process is feasible given the pool of resources available (and hence if not, what resources are needed)
- Graphical representation can allow for visualisation on screen of any bottlenecks that may develop during the course of a given simulation run
- Output simulation data can be easily migrated to other programs such as Excel

Resource Manager (User Solutions Incorporated)

- Primarily concerned with resource planning and utilisation as well as scheduling of a process, rather than manufacturing or costing aspects
- Has a rather heavy computational load
- Pre-programmed modules and runtime debug are allowed, but construction is not especially graphical (i.e. no icons or drag and drop controls are available)
- Limited graphical or statistical interpretation of output data
- Model packaging is available
- Designed to integrate as a Microsoft® Office add-in
- Provides for relatively straightforward, "what if," scenario analyses

SansGUI Modelling and Simulation Environment (ProtoDesign Incorporated - Bolingbrook, Illinois, U.S.A.)

- A simulation program built for a wide variety of purposes (i.e. not specifically targeted at a pharmaceutical manufacturing base) – so may need a fair amount of effort to use for the intended model
- Low demands placed on the computer
- Graphical construction, pre-programmed modules and runtime debug allowed
- Can connect to C++ code and has a wide range functions for producing graphs
- Provides an interactive and object-oriented simulation environment, without necessitating any code being written
- Little capacity for code reuse, but does allow packaging of model into a form that can be disseminated to the end user
- Model building blocks are specified in the code, within a hierarchical structure, which are then compiled into a supported programming language in order to produce the desired behaviour of the blocks

- Appears by reading available literature to be rather complicated – would need a long time to learn the relevant jargon and methodology

ShowFlow (Webb Systems Limited - Warrington, U.K.)

- A simulation software that focuses on bottleneck analysis, time reduction, resource consumption, staffing level analysis etc. in addition to material handling
- Low computational burden
- Reasonably swift in building a model
- Presence of simple graphical user interface/pre-programmed modules and runtime debug
- Objects that are present can be modified simply
- Reports and graphs are defined by the user and can be exported to Excel
- Feeds back statistics during the course of process run, not just at the end
- Packaging system for the model is available
- Suited to generating scenarios to test different manufacturing strategies
- Can simulate perishable or time limited items such that the real resource pool in the plant can be more accurately represented
- Has an attached costing module to perform any financial calculations

VisSim (Visual Solutions - Westford, Massachusetts, U.S.A.)

- A reasonably powerful package that targets, amongst other aspects, process simulation
- Moderate computational requirements
- Graphical construction/pre-programmed modules/runtime debug available
- Some features for producing graphs of the process, but this seems to be fairly narrow in reach
- Monte Carlo simulation support allowing effects of risk and probabilities of particular events occurring to be incorporated
- Packaging function available
- Does not need programming code in order to create simulation i.e. is fully graphical
- Can perform extensive, "what if," testing scenarios
- Can be compiled into the programming language, 'C,' in order to be run on a compatible system

Visual Simulation Environment (Orca Computer Incorporated - Blacksburg, Virginia)

- A general purpose simulation environment that lacks the necessary building blocks which could be adapted for the project – this would make the simulation more time consuming and awkward
- Reasonably high requirements of memory
- Graphical construction and run time debug are available
- Output analyser tool provides support for evaluating data generated from the system
- No packaging function is available for model distribution
- Object-oriented and picture-based visual modelling and simulation tool

APPENDIX 2: MATLAB® CODE

GSA code

The structure of the GSA program used when modelling the whole process is as follows:

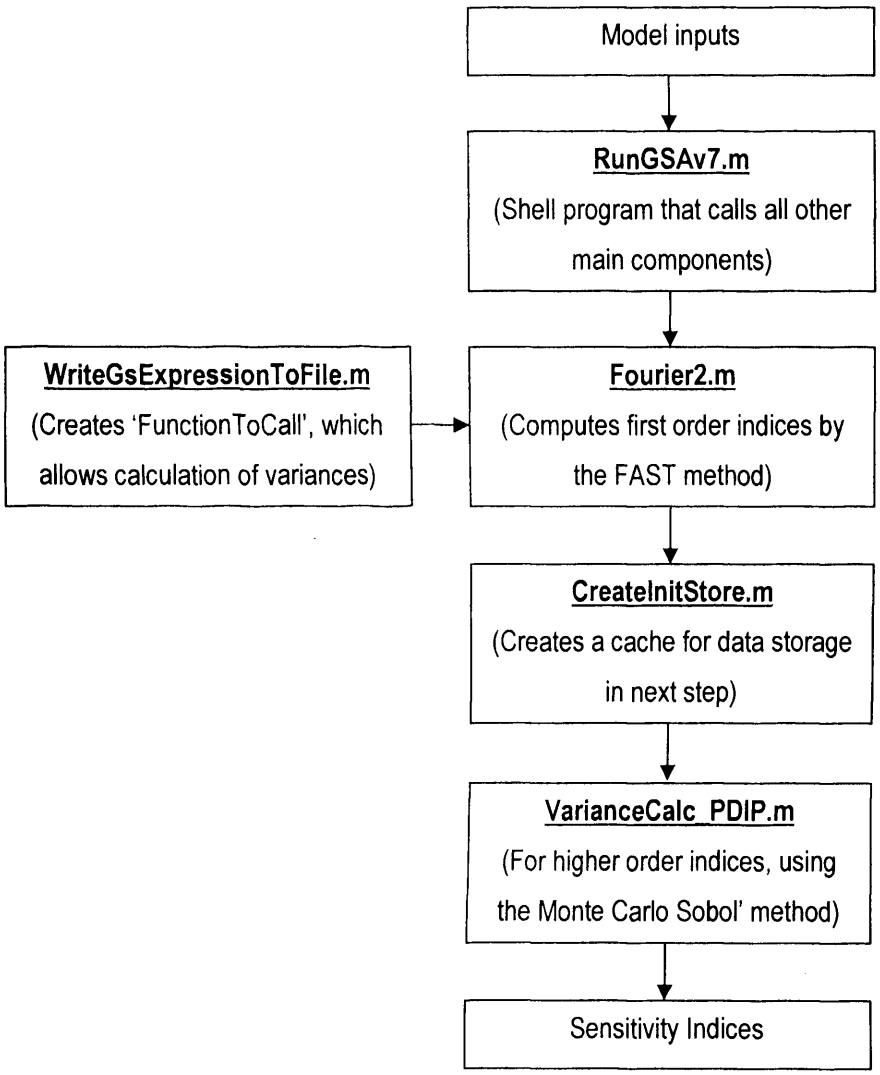


Figure 85: Schematic of GSA program operation

The MATLAB® code created for these steps (please see next pages) are aided by other files (Roundup, TrapeziumRule [for trapezium rule integration] and CreateRandom_0_to_1_Number [generates random inputs in the zero to one range]).

```
function RunGSAv7(NumberOfInputVariables,NumberOfOutputVariables,Limit,ModelFileName)
%THIS VERSION USES A DEFINED INPUT MESH (AND CALCULATES AT ALL POINTS) FOR DETERMINING HIGHER ORDER SENSITIVITY INDICES
%!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
%THIS FILE ONLY WORKS FOR A MODEL WITH A MAXIMUM OF 19 VARIABLES - THIS IS THE MAXIMUM FROM THE CUKIER PAPER
%!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
clear global
global SensitivityIndexMatrix%Used in the fWriter5.m file to calculate interaction indices with the MC Sobol
global OverallVariance% Used in fWriter5.m/VarianceCalc.m to multiply the lower order SIs (1st order) to calc. their variances
tic

%Delete previous files which stored the sensitivity index values (____SX_Y.m)
delete('GSA____S*.m');
delete('SIM.mat');
rehash;

%Limit (as a decimal) sets the lower limit below which first order variances do not
%adequately approximate the overall variance.  N.B. this, 'Limit,' is NOT the
%same as, 'LimitsLower,' used to calculate f_0
global StringOrIndiv;
StringOrIndiv=0;%Set this to 0 in order to return all outputs as a string or 1 for expressions/results emerging individually
global CurrentOPVar;
CurrentOPVar=0;%Set to 0 in order to initialise variable

format long
OverallVariance=0;
strFunction=ModelFileName;%Assign to use later
SecondHoldTerm='';%This string is set to be empty in preparation for its use in Monte Carlo integration, if required

%First order sensitivity indices as calculated by FAST
SensitivityIndexMatrix=Fourier2(NumberOfInputVariables,ModelFileName,NumberOfOutputVariables);
NeedToCalculateMCSobol=0;%Will use this variable to determine whether the MC Sobol method needs to run or not

SumOfFirstOrders=[];
for intCounter=1:1:NumberOfOutputVariables
    SumOfFirstOrders=[SumOfFirstOrders;sum(SensitivityIndexMatrix(:,intCounter))];
end
for intCounter=1:1:NumberOfOutputVariables
    if sum(SensitivityIndexMatrix(:,intCounter))<Limit
        NeedToCalculateMCSobol=1;
        break
```

```
    end
end

pack;

%Need to save the values of the sensitivity indices to file to be picked up
%later if necessary by Monte Carlo Sobol - file name SX_Y where X is the
%number of the input variable and Y is the number of the output variable
for intRowCounter=1:1:NumberOfInputVariables
    for intColCounter=1:1:NumberOfOutputVariables
        FunctionName=strcat('GSA____S',num2str(intRowCounter), '_',num2str(intColCounter));
        StringToFile=strcat('function Result=',FunctionName, '()');
        StringToFile=strcat(StringToFile,'\r\n','Result = ',num2str(SensitivityIndexMatrix(intRowCounter,intColCounter)),';');
        SavedIndexFile=fopen(strcat(FunctionName,'.m'),'w');
        fprintf(SavedIndexFile, StringToFile);
        fclose('all');
        rehash;
    end
end

rehash;

if NeedToCalculateMCSobol==1
    %Execute the Monte Carlo Sobol code for the higher order terms
    global gnsteps;
    gnsteps = 4;%Number of steps in the trapezium rule - this number is set to 25 for purposes of accuracy

    global DefinedMesh;%Used to refer to the defined inputs into the function file
    DefinedMesh=[0:(1/gnsteps):1];%Number of strips in the calculation

    CreateInitStore(NumberOfInputVariables);%Generate a store to hold the outputs in a cache when recalled later

    global f_0

    limitsLower = zeros(NumberOfInputVariables,1);
    limitsUpper = ones(NumberOfInputVariables,1);
    DT_DefinedInput=[];
    DT=[];
    f_0Str='';
    strFunction = strrep(ModelFileName,'inner','');%Eliminate the 'inner' literal from the file name, to operate on the file suc↙
```

```
h as 'WholeProcess,' rather than, 'innerWholeProcess.'
    pack;

    NumberOfIterations=(gnsteps+1)^NumberOfInputVariables;%e.g. for four strips and three inputs = 5^3 = 125

    %Create the f_0_DT_FunctionToCall file
    String=strcat('function Result = f_0_DT_FunctionToCall()','\r\n','global gnsteps','\r\n','\r\n','Temp=0;','TempDT=0;','\r\n'↙
);
    for intNumberOf_f_0Variables=1:1:(NumberOfInputVariables)
        String=strcat(String,'for InputMatrix',num2str(intNumberOf_f_0Variables),'=0:(1/gnsteps):1','\r\n');
    end
    String=strcat(String,'TempValue=',strFunction,'(');

    for intf_0Counter=1:1:NumberOfInputVariables
        String=strcat(String,'InputMatrix',num2str(intf_0Counter),',');
    end

    SizeOfString=size(String);%Get rid of final comma
    String=String(1:SizeOfString(2)-1);

    String=strcat(String,');');
    String=strcat(String,'\r\n','Temp=Temp+TempValue;');
    String=strcat(String,'\r\n','TempDT=TempDT+TempValue^2;');
    for intNumberOf_f_0Variables=1:1:(NumberOfInputVariables)
        String=strcat(String,'\r\n','end');
    end
    String=strcat(String,'\r\n','Result=[Temp,TempDT];');
    filename=strcat('f_0_DT_FunctionToCall.m');
    file=fopen(filename, 'w');
    fprintf(file, String);
    fclose('all');
    rehash;
    %Use the above file to calculate f_0 and DT
    for intCounter=1:1:NumberOfOutputVariables
        CurrentOPVar=intCounter;%Required by the main outer function file
        f_0_DT_value=f_0_DT_FunctionToCall;%Get the values of f_0 and DT
        f_0=[f_0;f_0_DT_value(1)/NumberOfIterations];%f_0 value
        DT_DefinedInput=[DT_DefinedInput;((f_0_DT_value(2)/NumberOfIterations)-f_0(intCounter))];%DT value
    end
    f_0
```

```
    DT_DefinedInput
    pause(10)

    global DFAST
    OverallVariance=DFAST;
    %Calculate the ratio of the pseudo to the FAST - derived DT
    global DTRatio
    DTRatio = OverallVariance'./DT_DefinedInput;

    %Run the method to calculate higher order variances
    PreTerms='';%Initialise matrix to hold the IDs of the relevant nos which get concatenated with 'f' to produce 'f12' etc.
    Terms=[];
    NumberOfTermsAtEachDepth=[];% Variable set to det how many combos of f12 (1st depth), f123 (2nd depth), f1234 (3rd depth)...
                                % there are such that the MC Sobol can be
                                % run for each depth in full and can then
                                % stop if need be
    %NoOfIndivNosInTerms=[];%Sets the number of individual numbers in each nchoosek call - used later in fWriter2 and then Varia↙
nceCalc
    for intOutsideCounter=2:1:NumberOfInputVariables
        Combination = nchoosek(1:1:NumberOfInputVariables,intOutsideCounter);%Creates all combos of f12, f13, f123, f1345 etc.
        SizeOfMatrix=size(Combination);
        SizeOfMatrix=SizeOfMatrix(1);
        NumberOfTermsAtEachDepth=[NumberOfTermsAtEachDepth;SizeOfMatrix];
        for intInsideCounter=1:1:SizeOfMatrix
            PreTerms=strvcat(PreTerms,num2str(Combination(intInsideCounter,:)));%vert. concatenates string versions of f12, f13.↙
..
            %NoOfIndivNosInTerms=[NoOfIndivNosInTerms;NumberOfInputVariables];%Sets the number of individual numbers in each nch↙
oosek call - used later in fWriter2 and then VarianceCalc
        end
    end
    SizeOfMatrix=size(PreTerms);
    SizeOfMatrix=SizeOfMatrix(1);

    HoldTerm='';
    SizeOfElementsInTempTerms=[];
    %The string version of the various f functions relating to the higher
    %order residuals (f12, f13, f123...) are not held together e.g. the
    %string version of the first is, '1  2,' not, '12.'
    %The following code concatenates the terms to convert, '1  2,' to, '12'
    %etc.
```

```
    for intOutsideCounter=1:1:SizeOfMatrix
        TempTerms = str2num(PreTerms(intOutsideCounter,:));%Creates a numerical version of the intOutsideCounter row of PreTerms
        NewSize=size(TempTerms);
        NewSize=NewSize(2);%Determine the number of rows in the TempTerms vector for the following for...next loop
        TempSizeOfElementsInTempTerms=size(TempTerms);
        TempSizeOfElementsInTempTerms=TempSizeOfElementsInTempTerms(2);
        SizeOfElementsInTempTerms=[SizeOfElementsInTempTerms;TempSizeOfElementsInTempTerms];
        for intInsideCounter=1:1:NewSize
            HoldTerm=strcat(HoldTerm,num2str(TempTerms(1,intInsideCounter)));
        end
        SecondHoldTerm=strvcat(SecondHoldTerm,HoldTerm);
        HoldTerm='';
    end
    SizeOfTerms=size(SecondHoldTerm);
    SizeOfTerms=SizeOfTerms(1);
    save('SecondHoldTermList','SecondHoldTerm');
    rehash;
    %Adjust the size of the SensitivityIndexMatrix to accommodate extra zeros, which will hold the final SecondHoldTerms numbers
    %'NumberOfOutputVariables'
    %NumberOfOutputVariables
    SensitivityIndexMatrix=[SensitivityIndexMatrix;zeros(SizeOfTerms,NumberOfOutputVariables)];

    rehash;
    OriginalNumberOfTermsAtEachDepth=NumberOfTermsAtEachDepth;%Reassign this so that when the code reloops for the next output v↙
ariable, it return to the original list of numbers of terms at each depth
    SavingVarianceTemp=[];
    for intOuterCounter=1:1:NumberOfOutputVariables
        rehash;
        NumberOfTermsAtEachDepth=OriginalNumberOfTermsAtEachDepth;
        rehash;
        if sum(SensitivityIndexMatrix(:,intOuterCounter))>=(Limit)
            strcat('First order variances (variable number:',num2str(intOuterCounter), ') are greater than', ' '' '' ', num2str(↙
Limit),' of the overall variance')
            'Do not recalculate';
        else
            CurrentOPVar=intOuterCounter;%Referenced to collect the correct f_0 value in higher order term calculation later
            StringOrIndiv=0;
            for intCounter=1:1:SizeOfTerms
                if NumberOfTermsAtEachDepth(1)==0
                    NumberOfTermsAtEachDepth(1)=[];
```

```
                end
                rehash;%Ensure that active memory content is refreshed with files from disk
                pack;
                %Create the relevant expression and integrate it using the variance function
                Expr=deblank(SecondHoldTerm(intCounter,:))
                rehash;
                %DO **NOT** REMOVE THE FOLLOWING PAUSE STATEMENT!!!!
                pause(10)
                Variance=((VarianceCalc_PDIP(Expr,NumberOfInputVariables,strFunction))/OverallVariance(intOuterCounter));
                SensitivityIndexMatrix((intCounter+NumberOfInputVariables),intOuterCounter)=Variance;
                save('SIM','SensitivityIndexMatrix');
                %Save the value of the latest Sensitivity Index Matrix to file
                FunctionName=strcat('GSA____S',Expr, '_',num2str(intOuterCounter));
                StringToFile=strcat('function Result=',FunctionName, '()');
                StringToFile=strcat(StringToFile,'\r\n','Result = ',num2str(SensitivityIndexMatrix((intCounter+NumberOfInputVari↙
ables),intOuterCounter)),';');
                SavedIndexFile=fopen(strcat(FunctionName,'.m'),'w');
                fprintf(SavedIndexFile, StringToFile);
                fclose('all');
                rehash;
                NumberOfTermsAtEachDepth(1)=NumberOfTermsAtEachDepth(1)-1;
                rehash;
                if (sum(SensitivityIndexMatrix(:,intOuterCounter))>=Limit & NumberOfTermsAtEachDepth(1)==0)
                    break
                else
                    'Do nothing';
                end
                pack;%Re-organise memory for efficiency
            end
        end
    end
    rehash;
else
    'Do nothing';%Does not calculate MC Sobol if all SI values are greater than the limit
end

%Adjust the SI data to scale back (else will give a sum(SI) of > 1
SizeOfSIMatrix=size(SensitivityIndexMatrix);
SizeOfSIMatrix=SizeOfSIMatrix(1);
for intCounter=1:1:NumberOfOutputVariables
```

```
    HOI_THEORY=1-SumOfFirstOrders(intCounter);%What the Higher Order Indices should add up to
    HOI_REALITY=sum(SensitivityIndexMatrix(:,intCounter))-SumOfFirstOrders(intCounter); %What higher order SI really add up to
    SensitivityIndexMatrix((NumberOfInputVariables+1):SizeOfSIMatrix,intCounter)=(HOI_THEORY/HOI_REALITY)*SensitivityIndexMatrix↙
((NumberOfInputVariables+1):SizeOfSIMatrix,intCounter);
end
'The following is a list of the sensitivities in the following order:'

FirstOrderLabel='';
for intCounter=1:1:NumberOfInputVariables
    FirstOrderLabel = strvcat(FirstOrderLabel,num2str(intCounter));
end

Label=FirstOrderLabel;

if isempty(SecondHoldTerm)==1
    strcat('Do nothing');
else
    Label=strvcat(FirstOrderLabel,SecondHoldTerm);
end
Label
SensitivityIndexMatrix
save(strcat('SIM_',strrep(strrep(datestr(now),':','_'),' ','_')),'SensitivityIndexMatrix');
toc
```

```
function OutputValues = Fourier2(NumberOfInputVariables,FileName,NumberOfOutputVariables)

%Sensitivity index computed by dividing variance for one specific variable
%by the variance of the entire function; the variances are determined by
%numerical integration of the Fourier coefficients' integrals

%Run the Fourier2.m file for each variable across the number of j iterations
%specified in 'NumberOfJIterationsMaxForOverall'

%NumberOfForwardIterations = number of steps forward from the lower bound to the upper bound
MaxHarmonicVarOfInt=4;%Assume that we only require the first four harmonics
load('OmegaTable','OmegaTable');%Loads minimal sets of Omega, taken from Cukier et at al 1973 paper II

Omega = OmegaTable((NumberOfInputVariables),:);%Delimits Omega to correct row
OmegaMax = max(OmegaTable((NumberOfInputVariables),:));%Gets the maximum harmonic value in the omega set
NumberOfForwardIterations = ((2*MaxHarmonicVarOfInt*OmegaMax)+1);%Calculated by the Nyqvist criterion
NumberOfJIterationsMaxForOverall = OmegaMax*MaxHarmonicVarOfInt;
Value=zeros(NumberOfJIterationsMaxForOverall,NumberOfOutputVariables);


intCounter=0;%For...Next loop counter variables
intRowCounter=0;%For...Next loop counter variables
intColCounter=0;%For...Next loop counter variables
IndivVar=zeros(NumberOfInputVariables,NumberOfOutputVariables);%Sets up a zeros matrix to hold the variances of individual varia↙
bles at the fundamental
%Gs vector for holding the transformations:
%This vector must contain sufficient rows to hold all of the GS values
%corresponding to one variable.  The number of columns must be equal to the
%number of variables.  Set a zero based array aside for holding this data
Gs=zeros(NumberOfForwardIterations+1,NumberOfInputVariables);
svector=[]; %Vector for holding the s values
%Vectors to hold the results of the Aj and Bj calculations for later
%integration by the trapezium rule
FinalCOSVector=zeros(NumberOfForwardIterations,NumberOfOutputVariables);
FinalSINVector=FinalCOSVector;%Same dimensions - hence copy the zeros matrix as for the cosine vector

TrapRuleNum=NumberOfForwardIterations;%Set the number of trapezium rule strips for numerical integration
LowerBoundForIntegration = -pi;%This sets the lower bound for the integration - as per Saltelli et al (2000)
UpperBoundForIntegration = pi;%This sets the upper bound for the integration - as per Saltelli et al (2000)
interval = (UpperBoundForIntegration-LowerBoundForIntegration)/NumberOfForwardIterations; %Sets up intervening values for s vect↙
```

```
or
svector=[(-pi):interval:pi]';%Choose s between -pi to pi and as determined by Saltelli equations

%Calculate the Gs terms
for intColCounter=1:1:NumberOfInputVariables
    for intRowCounter=1:1:(NumberOfForwardIterations+1)
        Gs(intRowCounter,intColCounter)=(0.5+(1/pi)*asin(sin((Omega(intColCounter)*svector(intRowCounter)))));
    end
end

WriteGsExpressionToFile(FileName,NumberOfInputVariables);%Creates a file called FunctionToCall which in turn calls the model 'Fi↙
leName'
rehash;

%Perform the Aj and Bj calculations

TempVector=zeros(NumberOfForwardIterations+1,NumberOfOutputVariables);
for intCounter=1:1:(NumberOfForwardIterations+1)
    %Calculates the function value at the specific value of Gs (as
    %determined by the relevant point in s space and the relevant variable)
    Temp=(FunctionToCall(Gs,intCounter));
    for intOPVarCounter=1:1:NumberOfOutputVariables
        TempVector(intCounter,intOPVarCounter)=Temp(intOPVarCounter);
    end
    %intCounter
end

for LoopVariable = 1:1:NumberOfJIterationsMaxForOverall
    for intCounter=1:1:(NumberOfForwardIterations+1)
        %The following lines calculate the f(s) values, for multiplication
        %later by either cos(js) or sin(js)
        for intInternalCounter=1:1:NumberOfOutputVariables
            TempCOS=TempVector(intCounter,intInternalCounter)*cos(LoopVariable*svector(intCounter));
            TempSIN=TempVector(intCounter,intInternalCounter)*sin(LoopVariable*svector(intCounter));
            FinalCOSVector(intCounter,intInternalCounter)=TempCOS;
            FinalSINVector(intCounter,intInternalCounter)=TempSIN;
        end
    end

    for intCounter=1:1:NumberOfOutputVariables
```

```
        AreaA = (1/(2*pi))*TrapeziumRule(FinalCOSVector(:,intCounter),svector(1),svector(NumberOfForwardIterations+1),TrapRuleNu↙
m);
        AreaB = (1/(2*pi))*TrapeziumRule(FinalSINVector(:,intCounter),svector(1),svector(NumberOfForwardIterations+1),TrapRuleNu↙
m);
        Value(LoopVariable,intCounter)=(2*(AreaA^2+AreaB^2));
    end
    %strcat(num2str(LoopVariable), '/',num2str(NumberOfJIterationsMaxForOverall), ' Fourier iterations completed')%Reports the o↙
utputs
end

format long
%YOU NEED TO BE ABSOLUTELY SURE THAT THESE VALUES HAVE CONVERGED AND THAT
%YOU ARE NOWHERE NEAR A POINT OF SUPERPOSITION WHERE FREQUENCIES MAY
%INTERFERE, RESULTING IN A SPIKE IN THE OUTPUT VALUE OF THE Aj/Bj INTEGRALS
%CHECK THE OUTPUT OF 'VALUE' AFTER EVERY RUN
SizeOfValue=size(Value);
SizeOfValue=SizeOfValue(1);
OverallVariance=zeros(1,NumberOfOutputVariables);
%Calculates the overall variance of the model function f(x)
for intExternalCounter=1:1:NumberOfOutputVariables
    for intCounter=1:1:SizeOfValue
        OverallVariance(1,intExternalCounter)=OverallVariance(1,intExternalCounter)+Value(intCounter,intExternalCounter);
    end
end

%Determine the individual variances of specific variables as the sum over
%the first 'MaxHarmonicVarOfInterest' number of harmonics
for intIPCounter=1:1:NumberOfInputVariables
    for intOPCounter=1:1:NumberOfOutputVariables
        for intRowCounter=1:1:MaxHarmonicVarOfInt
            IndivVar(intIPCounter,intOPCounter) = IndivVar(intIPCounter,intOPCounter) + Value((intRowCounter*Omega(intIPCounter)↙
),intOPCounter);
        end
    end
end

%Fourier results
OutputValues=zeros(NumberOfInputVariables,NumberOfOutputVariables);
for intOPCounter=1:1:NumberOfOutputVariables
        OutputValues(:,intOPCounter)=IndivVar(:,intOPCounter)/OverallVariance(1,intOPCounter);
```

```
end
%Make the overall variance global so that it can be accessed in the
% main RunGSAvX program
global DFAST
DFAST=OverallVariance;
```

```
function WriteGsExpressionToFile(FileName, NumberOfVariables)
%Works in the Fourier method
%Creates the file for 'FunctionToCall', which calls the model 'FileName'
header= sprintf('function result=%s(%s)', 'FunctionToCall', 'Gs,intCounter');

LoopCounter=0;
Argument='';

for LoopCounter=1:1:NumberOfVariables
    if LoopCounter==1
        Argument='Gs(intCounter,1)';
    else
        Argument=strcat(Argument, ',Gs(intCounter,', num2str(LoopCounter),')');
    end
end

func = strcat(header,'\n\n', 'result=', FileName, '(', Argument , ');');
FileName='FunctionToCall.m';
file=fopen(FileName, 'w');
fprintf(file, func);
fclose('all');
rehash;
```

```
function CreateInitStore(dims)
%Create the valueStore variables to the dimensionality of 'dims' - recalled
%later in the calculation of the higher order indices
global valueStoreYield;
global valueStorePurity;
global gnsteps

dimensionsArray=[];

for i=1:dims
    dimensionsArray=[dimensionsArray (gnsteps+1)];
end

valueStoreYield = zeros(dimensionsArray);
valueStorePurity = zeros(dimensionsArray);
```

```
function Variance = VarianceCalc_DefinedInput(Terms,NumberOfInputVariables,strFunction)
%This file creates the file 'D_FunctionToCall' (for higher order indices)
%It is the same as VarianceCalc_Partial_DefinedInput, but the name was
%shortened to allow it to run on UNIX
global f_0;
global CurrentOPVar;
global OverallVariance

%Set-up the y, z and v vectors
% global y
% global z
% global v

%Calculate the number of terms present
SizeOfTerms=size(Terms);
SizeOfTerms=SizeOfTerms(2);%Get the number of numbers in the Terms argument e.g. for 1246, SizeOfTerms = 4

y=zeros(SizeOfTerms,1);
z=zeros(NumberOfInputVariables-SizeOfTerms,1);
v=zeros(NumberOfInputVariables-SizeOfTerms,1);

%Set up the following variables for use later
LowerOrderValues=0;
LatestValue=0;
LatestVariance=[];

%The following code calculate the lower order variance values (to be taken
%away from the calculated higher order one at the end of this file
HoldValues1=[];
for intCounter=1:1:SizeOfTerms
    HoldValues1=[HoldValues1 , str2num(Terms(intCounter))];%Values of the input terms which are used to construct the lower orde↙
r subtraction
end
HoldValues2=[];
for intFirstCounter=1:1:(SizeOfTerms-1)
    HoldValues2=nchoosek(HoldValues1,intFirstCounter);
    SizeOfHoldValues2=size(HoldValues2);
    NumberOfRows=SizeOfHoldValues2(1);
    NumberOfCols=SizeOfHoldValues2(2);
    for intSecondCounter=1:1:NumberOfRows
```

```
       Expression='';
        for intThirdCounter=1:1:NumberOfCols
            Expression=strcat(Expression,num2str(HoldValues2(intSecondCounter,intThirdCounter)));
        end
        rehash;
        LowerOrderValues=LowerOrderValues+eval(strcat('GSA____S',Expression,'_',num2str(CurrentOPVar)));
    end
end
LowerOrderValues=LowerOrderValues*OverallVariance(CurrentOPVar);

VarConstant=[];
for intCounter=1:1:SizeOfTerms;
VarConstant=[VarConstant;str2num(Terms(intCounter))];%Create a matrix containing the numbers within 'Terms'
end

%The expression to be called is f(y,z)*f(y,v)
%Hence need to set up/initialise arguments
VarCounter=1;
ConstantCounter=1;
FirstArg='';
SecondArg='';

%Create f(y,z)
for intFirstCounter=1:1:NumberOfInputVariables
    if VarCounter<=SizeOfTerms
        if intFirstCounter==VarConstant(VarCounter)
            FirstArg=strcat(FirstArg, 'y', num2str(VarCounter), ',');
            VarCounter=VarCounter+1;
        else
            FirstArg=strcat(FirstArg,'z' , num2str(ConstantCounter), ',');
            ConstantCounter=ConstantCounter+1;
        end
    else
        FirstArg=strcat(FirstArg,'z' , num2str(ConstantCounter), ',');
        ConstantCounter=ConstantCounter+1;
    end
end

VarCounter=1;
ConstantCounter=1;
```

```
%Create f(y,v)
for intSecondCounter=1:1:NumberOfInputVariables
    if VarCounter<=SizeOfTerms
        if intSecondCounter==VarConstant(VarCounter)
            SecondArg=strcat(SecondArg, 'y', num2str(VarCounter), ',');
            VarCounter=VarCounter+1;
        else
            SecondArg=strcat(SecondArg,'v' , num2str(ConstantCounter), ',');
            ConstantCounter=ConstantCounter+1;
        end
    else
        SecondArg=strcat(SecondArg,'v' , num2str(ConstantCounter), ',');
        ConstantCounter=ConstantCounter+1;
    end
end

%Need to strip out the final comma from the two strings; since the two arguments are the same size, only do this once
SizeOfArg=size(FirstArg);
FirstArg=FirstArg(1:SizeOfArg(2)-1);%Does this work????
SecondArg=SecondArg(1:SizeOfArg(2)-1);

% Parameters used to carry out the integration
Range=1000;
LowerLimit=50000;
FractionalVariation=0.05;
NumberOfIterations=1000000;%Set this value to 1M to provide a good chance of convergence

%Generate the D_FunctionToCall file to evaluate the variance
FunctionName='D_FunctionToCall';
StringToFile=strcat('function Result=',FunctionName, '(NumberOfIterations, Range, LowerLimit, FractionalVariation, Terms)');
StringToFile=strcat(StringToFile,'\r\n','LatestVariance=[];');%Use this to hold the accumulating variance values
StringToFile=strcat(StringToFile,'\r\n','global gnsteps;');
StringToFile=strcat(StringToFile,'\r\n','global CurrentOPVar');
StringToFile=strcat(StringToFile,'\r\n','Temp=0;');

StringToFile=strcat(StringToFile,'\r\n','for Counter=1:1:NumberOfIterations');
%StringToFile=strcat(StringToFile,'\r\n','Counter');
StringToFile=strcat(StringToFile,'\r\n');
%Choose random number in the range [0;1/gnsteps;1]
```

```
for int_y_Counter=1:1:SizeOfTerms
    StringToFile=strcat(StringToFile,'\r\n','y', num2str(int_y_Counter) ,'=CreateRandom_0_to_1_Number;');
end

for int_z_Counter=1:1:(NumberOfInputVariables-SizeOfTerms)
    StringToFile=strcat(StringToFile,'\r\n','z', num2str(int_z_Counter) ,'=CreateRandom_0_to_1_Number;');
end

for int_v_Counter=1:1:(NumberOfInputVariables-SizeOfTerms)
    StringToFile=strcat(StringToFile,'\r\n','v', num2str(int_v_Counter) ,'=CreateRandom_0_to_1_Number;');
end

StringToFile=strcat(StringToFile,'\r\n','Temp=Temp+',strFunction,'(',FirstArg,')','*',strFunction,'(',SecondArg,');');
StringToFile=strcat(StringToFile,'\r\n','LatestVariance=[LatestVariance;Temp/Counter];');

StringToFile=strcat(StringToFile,'\r\n','if Counter<(LowerLimit+Range)');
StringToFile=strcat(StringToFile,'\r\n',' ''Do nothing'';  ');
StringToFile=strcat(StringToFile,'\r\n','elseif     abs(((max(LatestVariance(Counter-Range:Counter))-min(LatestVariance(Counter-↙
Range:Counter)))/min(LatestVariance(Counter-Range:Counter))))<FractionalVariation;');
StringToFile=strcat(StringToFile,'\r\n','break');
%StringToFile=strcat(StringToFile,'\r\n','else');
%StringToFile=strcat(StringToFile,'\r\n',' ''Do nothing;'';  ');
StringToFile=strcat(StringToFile,'\r\n','end');

StringToFile=strcat(StringToFile,'\r\n','end');
StringToFile=strcat(StringToFile,'\r\n','save(strcat(''LatestVariance_'',Terms,''_'',num2str(CurrentOPVar)),''LatestVariance'');↙
');%Save the accumulated variance values
StringToFile=strcat(StringToFile,'\r\n','Result = ','Temp/Counter;');
SavedIndexFile=fopen(strcat(FunctionName,'.m'),'w');
fprintf(SavedIndexFile, StringToFile);
fclose('all');
rehash;

global DTRatio %Adjust the value of D to suit (else overestimates the value of higher order indices.  DTRatio calculated in RunG↙
SAvX)
Variance=((D_FunctionToCall(NumberOfIterations, Range, LowerLimit, FractionalVariation, Terms))-(f_0(CurrentOPVar))^2)-LowerOrde↙
rValues;
Variance=DTRatio(CurrentOPVar)*Variance;
if Variance<0
    Variance=0;
```

```
end
```

```
function Result = RoundUp(InputNo)
if(InputNo<0)
    'This function only works with positive numbers'
else
    if InputNo==0
        Result=0
    elseif InputNo<1
        Result=1;
    elseif int16(InputNo) < InputNo
        Result = double(int16(InputNo))+1;
    elseif int16(InputNo) == InputNo;
        Result = InputNo;
    end
end
```

```
function Result = TrapeziumRule(XHolder,a,b,n)

%Area = [(b-a)/2n]*[f(a) + f(b) + 2*(f(all the rest))]
Tmp=0;
intCounter=0;

for intCounter = 2:1:(n)
    Tmp = Tmp+XHolder(intCounter);
end
Result = ((b-a)/(2*n))*(XHolder(1) + XHolder(n+1) + 2*Tmp);
```

```
function Result=CreateRandom_0_to_1_Number()
%Creates a number between 0 and 1
global gnsteps
Result=round(gnsteps * rand)/gnsteps;
```

Genetic Program

The structure of the MATLAB® code used for the Genetic Program employed to create the digestion equation is given below. Each term in the equation (i.e. either a constant number or one of the input variables [temperature, time and enzyme concentration]) of the initial/evolved equations was held in a separate disk file. These were then altered at random by the GP to achieve greater fitness.

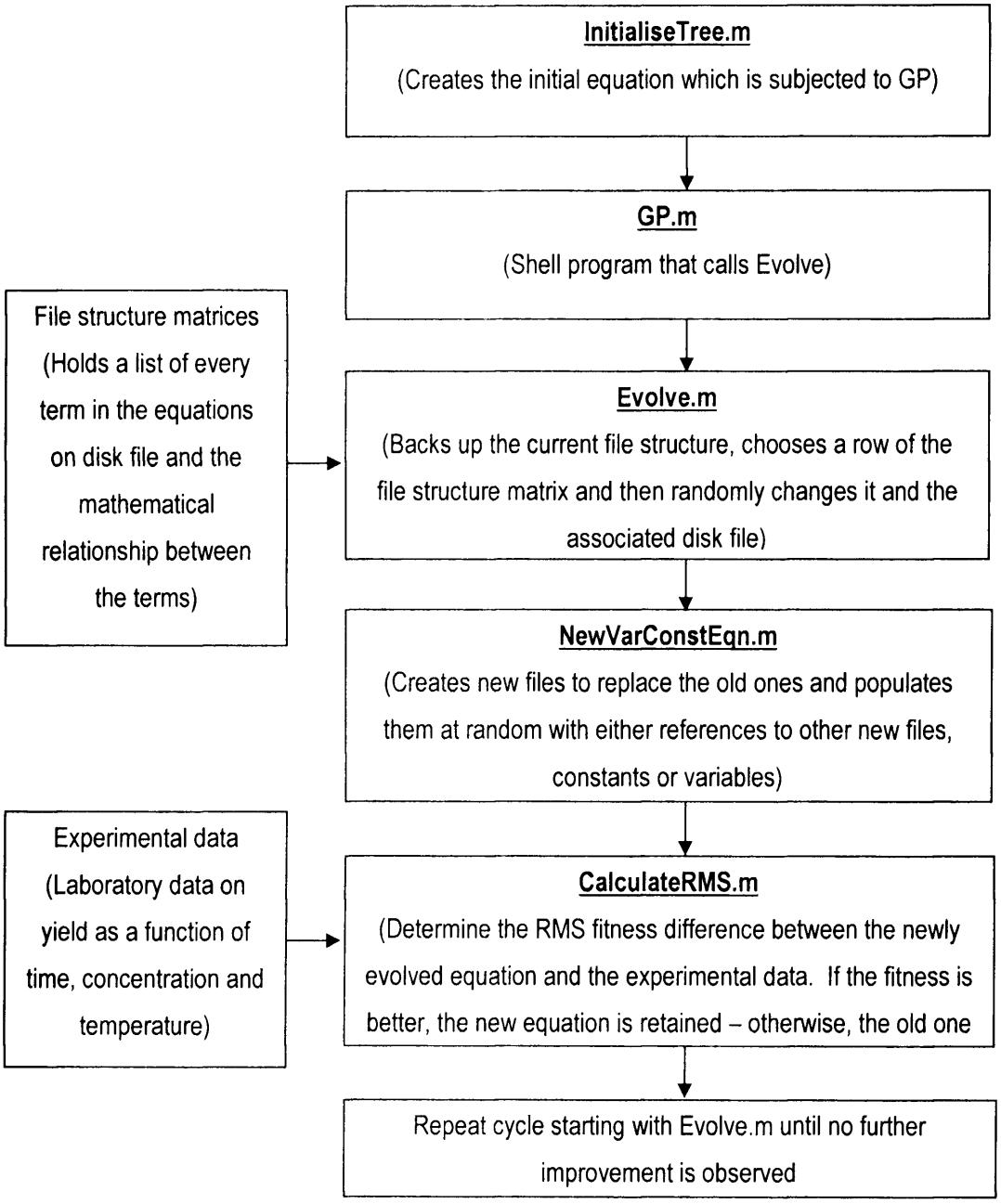


Figure 86: Schematic of GP program operation

Other files used by the program include:

MathsOperator.m - determines which mathematical operator to choose when evolving

NewFile.m - creates new files when evolving

fWriter.m - writes file to disk

```
function InitialiseTree()
clear all
load ('InitialFileRefTable2')
load ('Data')
%Delete the existing Ref???????.m files and CurrentDiskFile from line 183 (Evolve.m) from disk
DirList=dir('Ref*.m');
SizeOfDirList=size(DirList);
SizeOfDirList=SizeOfDirList(1);
ConstantFiles=[];
if SizeOfDirList>0
    delete('Ref*.m')
end

if exist('CurrentDiskFile.mat')~=0
    delete ('CurrentDiskFile.mat')
end
%This section of the file reestablishes the original equation
%The equation used comes from Ainsworth (Steady State Enzyme kinetics) and
%expresses the progress curve of the reaction (trace of P [and hence yield]
%against t) as a polynomial function which has been approximated to a
%third order partial sum:

%ProductConc = alpha + beta*time + gamma*time^2 + delta*time^3   [+ other
%terms, unnecessary as they do not make a significant contribution]

body = '(Ref0000002(ConcIP,TempIP,TimeIP))*(Ref0000003(ConcIP,TempIP,TimeIP));'; %Temp * rest
fWriter('Ref0000001', body)

body = 'TempIP;';
fWriter('Ref0000002', body)

body = '(Ref0000004(ConcIP,TempIP,TimeIP))+(Ref0000005(ConcIP,TempIP,TimeIP));';  %alpha plus rest
fWriter('Ref0000003', body)

body = '-0.02;'; %Value of alpha - originally -1 - then -1 - then 0
fWriter('Ref0000004', body)
ConstantFiles=[ConstantFiles;'Ref0000004'];

body = 'Ref0000006(ConcIP,TempIP,TimeIP) + Ref0000007(ConcIP,TempIP,TimeIP);'; %beta*time plus rest of string
fWriter('Ref0000005', body)
```

```
body = 'Ref0000008(ConcIP,TempIP,TimeIP) * Ref0000009(ConcIP,TempIP,TimeIP);'; %beta*time
fWriter('Ref0000006', body)

body = 'Ref0000010(ConcIP,TempIP,TimeIP) + Ref0000011(ConcIP,TempIP,TimeIP);';%gamma*time^2 plus rest of string
fWriter('Ref0000007', body)

body = '0.5;'; %beta - originally 0.5 - then 0.05
fWriter('Ref0000008', body)
ConstantFiles=[ConstantFiles;'Ref0000008'];

body = 'TimeIP;'; %time
fWriter('Ref0000009', body)

body = 'Ref0000012(ConcIP,TempIP,TimeIP) * Ref0000013(ConcIP,TempIP,TimeIP);'; %gamma*time^2
fWriter('Ref0000010', body)

body = 'Ref0000014(ConcIP,TempIP,TimeIP) * Ref0000015(ConcIP,TempIP,TimeIP);'; %delta*time^3
fWriter('Ref0000011', body)

body = '0.5;'; %gamma - originally 0.05
fWriter('Ref0000012', body)
ConstantFiles=[ConstantFiles;'Ref0000012'];

body = 'Ref0000016(ConcIP,TempIP,TimeIP) * Ref0000017(ConcIP,TempIP,TimeIP);'; %gamma*time^2
fWriter('Ref0000013', body)

body = '0.0001;'; %delta
fWriter('Ref0000014', body)
ConstantFiles=[ConstantFiles;'Ref0000014'];

body = 'TimeIP^3;'; %Time
fWriter('Ref0000015', body)

body = 'TimeIP;'; %time
fWriter('Ref0000016', body)

body = 'TimeIP;'; %time
fWriter('Ref0000017', body)
```

```
%The following is the original equation - it has been substituted for by
%the equation above in order to achieve a better fit with temperature
% body = '(Ref0000002(ConcIP,TempIP,TimeIP))+(Ref0000003(ConcIP,TempIP,TimeIP));';  %alpha plus rest
% fWriter('Ref0000001', body)
%
% body = '(Ref0000002(ConcIP,TempIP,TimeIP))+(Ref0000003(ConcIP,TempIP,TimeIP));';  %alpha plus rest
% fWriter('Ref0000001', body)
%
% body = '0;'; %Value of alpha - originally -1 - then -1
% fWriter('Ref0000002', body)
% ConstantFiles=[ConstantFiles;'Ref0000002'];
%
% body = 'Ref0000004(ConcIP,TempIP,TimeIP) + Ref0000005(ConcIP,TempIP,TimeIP);'; %beta*time plus rest of string
% fWriter('Ref0000003', body)
%
% body = 'Ref0000006(ConcIP,TempIP,TimeIP) * Ref0000007(ConcIP,TempIP,TimeIP);'; %beta*time
% fWriter('Ref0000004', body)
%
% body = 'Ref0000008(ConcIP,TempIP,TimeIP) + Ref0000009(ConcIP,TempIP,TimeIP);';%gamma*time^2 plus rest of string
% fWriter('Ref0000005', body)
%
% body = '0.5;'; %beta - originally 0.5 - then 0.05
% fWriter('Ref0000006', body)
% ConstantFiles=[ConstantFiles;'Ref0000006'];
%
% body = 'TimeIP;'; %time
% fWriter('Ref0000007', body)
%
% body = 'Ref0000010(ConcIP,TempIP,TimeIP) * Ref0000011(ConcIP,TempIP,TimeIP);'; %gamma*time^2
% fWriter('Ref0000008', body)
%
% body = 'Ref0000012(ConcIP,TempIP,TimeIP) * Ref0000013(ConcIP,TempIP,TimeIP);'; %delta*time^3
% fWriter('Ref0000009', body)
%
% body = '0.5;'; %gamma - originally 0.05
% fWriter('Ref0000010', body)
% ConstantFiles=[ConstantFiles;'Ref0000010'];
%
% body = 'Ref0000014(ConcIP,TempIP,TimeIP) * Ref0000015(ConcIP,TempIP,TimeIP);'; %gamma*time^2
% fWriter('Ref0000011', body)
```

```
%
% body = '0.0001;'; %delta
% fWriter('Ref0000012', body)
% ConstantFiles=[ConstantFiles;'Ref0000012'];
%
% body = 'TimeIP^3;'; %Time
% fWriter('Ref0000013', body)
%
% body = 'TimeIP;'; %time
% fWriter('Ref0000014', body)
%
% body = 'TimeIP;'; %time
% fWriter('Ref0000015', body)

RMSError = CalculateRMS('InitialFileRefTable2')

FileNumber=[17];

cd('C:/My Documents/Sunil/Papain_evolution/GP2/Backup')
delete('Ref*.m')
cd('C:/My Documents/Sunil/Papain_evolution/GP2')

save('FileRefTable','Row','LHS','Operator','RHS','FileNumber')
save('RMSError','RMSError')
save('ConstantFiles','ConstantFiles')
```

```
function GP(NoOfRounds)
for intCounter = 1:NoOfRounds
    clear all
    Evolve
    %ChangeConstants
end
```

```
function Evolve()
global NewlyCreatedFiles
NewlyCreatedFiles=[];
%Keep existing file structure
copyfile('Ref*.m','C:/My Documents/Sunil/Papain_evolution/GP2/Backup')
load ('FileRefTable')
load ('RMSError')
rand('state',sum(100*clock));
%Select one of the rows
NumberOfFirstEntries=[];
NumberOfRows = size(Row);
NumberOfRows = NumberOfRows(1);
ChoiceOfRow = round(NumberOfRows * rand(1));
if ChoiceOfRow==0
    ChoiceOfRow=1;
end
Operator(ChoiceOfRow,:)=MathsOperator;

%If a file exists in the LHS or RHS matrices of that chosen row, then need to access them and
%mark them for deletion
FilesToBeDeleted=[];

if LHS(ChoiceOfRow,:)~='          '
    FilesToBeDeleted=[FilesToBeDeleted;LHS(ChoiceOfRow,:)];
    NumberOfFirstEntries = 1;
end
if RHS(ChoiceOfRow,:)~='          '
    FilesToBeDeleted=[FilesToBeDeleted;RHS(ChoiceOfRow,:)];
    NumberOfFirstEntries=2;
end
if LHS(ChoiceOfRow,:)=='          '
    FileNumber=FileNumber+1;
    LHS(ChoiceOfRow,:)=NewFile(FileNumber);
    fWriter(Row(ChoiceOfRow,:),strcat(LHS(ChoiceOfRow,:),'(ConcIP,TempIP,TimeIP);'));
end
%Go through the row matrix and look for the entry which corresponds to the
%first (and also second, if it exists) in the FilesToBeDeleted matrix.
if isempty(FilesToBeDeleted)==0
    for intCounter = 1:NumberOfRows
        if strcmp(Row(intCounter,:),FilesToBeDeleted(1,:))==1
```

```
            %Add any referenced files to FilesToBeDeleted matrix and eliminate
            %the first entry in this matrix
            if LHS(intCounter,:)~='          '
                FilesToBeDeleted=[FilesToBeDeleted;LHS(intCounter,:)];
                LHS(intCounter,:)='          ';
            end
            if RHS(intCounter,:)~='          '
                FilesToBeDeleted=[FilesToBeDeleted;RHS(intCounter,:)];
                RHS(intCounter,:)='          ';
            end
            Operator(intCounter,:)=' ';
            fWriter(FilesToBeDeleted(1,:),'')
            FirstReferencedFileLocation=intCounter; %Use later to refill the matrix
        end
        if NumberOfFirstEntries==2
            if strcmp(Row(intCounter,:),FilesToBeDeleted(2,:))==1
                %Add any referenced files to FilesToBeDeleted matrix and eliminate
                %the first entry in this matrix
                if LHS(intCounter,:)~='          '
                    FilesToBeDeleted=[FilesToBeDeleted;LHS(intCounter,:)];
                    LHS(intCounter,:)='          ';
                end
                if RHS(intCounter,:)~='          '
                    FilesToBeDeleted=[FilesToBeDeleted;RHS(intCounter,:)];
                    RHS(intCounter,:)='          ';
                end
                Operator(intCounter,:)=' ';
                fWriter(FilesToBeDeleted(2,:),'')
                SecondReferencedFileLocation=intCounter; %Use later to refill the matrix
            end
        end
    end
end
StartPtForRepop=FilesToBeDeleted;
if NumberOfFirstEntries==1
    %delete(strcat(FilesToBeDeleted(1,:)))
    FilesToBeDeleted(1,:)=[];
end
if NumberOfFirstEntries==2
    %delete(strcat(FilesToBeDeleted(1,:)))
```

```
    FilesToBeDeleted(1,:)=[];
    %delete(strcat(FilesToBeDeleted(1,:)))
    FilesToBeDeleted(1,:)=[];
end

%Now need to find and remove all of the other references to the other files
%Also need to empty the disk files of all of these references (as well as
%those above)
NumberOfRows=size(Row);
NumberOfRows=NumberOfRows(1);
while isempty(FilesToBeDeleted)==0
    %Delete any files in the FilesToBeDeleted matrix from the three
    %matrices and delete the operator as well
    for intCounter=1:NumberOfRows
        if intCounter <= NumberOfRows
            if isempty(FilesToBeDeleted)==0
                if strcmp(Row(intCounter,:),FilesToBeDeleted(1,:))==1
                    Row(intCounter,:)=[];
                    if LHS(intCounter,:)~='          '
                        FilesToBeDeleted=[FilesToBeDeleted;LHS(intCounter,:)];
                    end
                    if RHS(intCounter,:)~='          '
                        FilesToBeDeleted=[FilesToBeDeleted;RHS(intCounter,:)];
                    end
                    fWriter(FilesToBeDeleted(1,:),'')
                    delete(strcat(FilesToBeDeleted(1,:),'.m'))
                    FilesToBeDeleted(1,:)=[];
                    LHS(intCounter,:)=[];
                    RHS(intCounter,:)=[];
                    Operator(intCounter,:)=[];
                    NumberOfRows=size(Row);
                    NumberOfRows=NumberOfRows(1);
                end
            end
        end
    end
    intCounter = 0;
end

%Now need to repopulate, starting from the the StartPtForRepop matrix.
```

```
%Refill the LHS, RHS and operator matrices for the one or two entries in
%here first with new files (assuming that you actually want to do this - need to decide at random).
if NumberOfFirstEntries==1
    FileNumber=FileNumber+1;
    LHS(FirstReferencedFileLocation,:) = NewFile(FileNumber);
    DoYouHaveAFileOnRHS = rand(1);
    if DoYouHaveAFileOnRHS >0.5 %Determines whether you are going to add a file reference to the RHS/operator matrix if necessar↙
y
        FileNumber=FileNumber+1;
        RHS(FirstReferencedFileLocation,:) = NewFile(FileNumber);
        Operator(FirstReferencedFileLocation) = MathsOperator;
        fWriter(StartPtForRepop(1,:),strcat(LHS(FirstReferencedFileLocation,:),'(ConcIP,TempIP,TimeIP)',Operator(FirstReferenced↙
FileLocation,:),RHS(FirstReferencedFileLocation,:),'(ConcIP,TempIP,TimeIP);'))
    else
        fWriter(StartPtForRepop(1,:),strcat(LHS(FirstReferencedFileLocation,:),'(ConcIP,TempIP,TimeIP);'))
    end
end

if NumberOfFirstEntries==2;

    FileNumber=FileNumber+1;
    LHS(FirstReferencedFileLocation,:) = NewFile(FileNumber);
    DoYouHaveAFileOnRHS = rand(1);
    if DoYouHaveAFileOnRHS >0.5 %Determines whether you are going to add a file reference to the RHS/operator matrix if necessar↙
y
        FileNumber=FileNumber+1;
        RHS(FirstReferencedFileLocation,:) = NewFile(FileNumber);
        Operator(FirstReferencedFileLocation) = MathsOperator;
        fWriter(StartPtForRepop(1,:),strcat(LHS(FirstReferencedFileLocation,:),'(ConcIP,TempIP,TimeIP)',Operator(FirstReferenced↙
FileLocation,:),RHS(FirstReferencedFileLocation,:),'(ConcIP,TempIP,TimeIP);'))
    else
        fWriter(StartPtForRepop(1,:),strcat(LHS(FirstReferencedFileLocation,:),'(ConcIP,TempIP,TimeIP);'))
    end



    FileNumber=FileNumber+1;
    LHS(SecondReferencedFileLocation,:) = NewFile(FileNumber);
    DoYouHaveAFileOnRHS = rand(1);
    if DoYouHaveAFileOnRHS >0.5
```

```
        FileNumber=FileNumber+1;
        RHS(SecondReferencedFileLocation,:) = NewFile(FileNumber);
        Operator(SecondReferencedFileLocation,:) = MathsOperator;
        fWriter(StartPtForRepop(2,:),strcat(LHS(SecondReferencedFileLocation,:),'(ConcIP,TempIP,TimeIP)',Operator(SecondReferenc↙
edFileLocation,:),RHS(SecondReferencedFileLocation,:),'(ConcIP,TempIP,TimeIP);'))
    else
        fWriter(StartPtForRepop(2,:),strcat(LHS(SecondReferencedFileLocation,:),'(ConcIP,TempIP,TimeIP);'))
    end
end

%Now have new filenames - add file references, constants, variable or a combination and add their names
%to Row list.  For any new file references, create those files and adjust
%the LHS, RHS and Operator matrices as needed

while isempty(NewlyCreatedFiles)==0
    Row=[Row;NewlyCreatedFiles(1,:)]; %Add the uppermost entry from the NewlyCreatedFiles matrix to the Row matrix
    IsThisFileGoingToReferToAnything = rand(1);
    if IsThisFileGoingToReferToAnything > 0.3 %No reference to anything - changed from 0.5 to make it more likely
                                              %that the branches will terminate
        %Add an empty reference
        LHS=[LHS;'          '];
        RHS=[RHS;'          '];
        Operator=[Operator;' '];
        %Then just populate this file with a variable, constant or a combination of the two
        fWriter(NewlyCreatedFiles(1,:),newVarConstEqn)
%     elseif IsThisFileGoingToReferToAnything > 0.33 %A single reference (and hence LHS in nature)
%         FileNumber=FileNumber+1;
%         AnotherFile=NewFile(FileNumber);
%         LHS=[LHS;AnotherFile];
%         %Populate the NewlyCreatedFiles(1) file with a reference to the AnotherFile
%         fWriter(NewlyCreatedFiles(1,:),strcat(AnotherFile,'(ConcIP,TempIP,TimeIP)'));
%         RHS=[RHS;'          '];
%         Operator=[Operator;' '];
    elseif IsThisFileGoingToReferToAnything > 0 %Reference to two files
        FileNumber=FileNumber+1;
        AnotherFile=NewFile(FileNumber);
        LHS=[LHS;AnotherFile];
        FileNumber=FileNumber+1;
        AnotherFile2=NewFile(FileNumber);
        RHS=[RHS;AnotherFile2];
```

```
        NewOperator=MathsOperator;
        Operator=[Operator;NewOperator];
        fWriter(NewlyCreatedFiles(1,:),strcat(AnotherFile,'(ConcIP,TempIP,TimeIP)',NewOperator,AnotherFile2,'(ConcIP,TempIP,Time↙
IP);'));
        %Populate the NewlyCreatedFiles(1) file with a reference to two files - hence the two fWriter statements in above codebl↙
ock
    end
    NewlyCreatedFiles(1,:)=[];    %Remove this file from consideration and loop to the next
end

save('CurrentDiskFile','Row','LHS','Operator','RHS','FileNumber')
CurrentDiskFile=('CurrentDiskFile');
LatestRMSValue=CalculateRMS(CurrentDiskFile)
if LatestRMSValue < RMSError
    if isreal(LatestRMSValue)==1
        %save('FileRefTable','Row','LHS','Operator','RHS','FileNumber')
        copyfile('CurrentDiskFile.mat','FileRefTable.mat')
        RMSError = LatestRMSValue;
        save('RMSError','RMSError')
        strcat('Latest best RMS value is: ', num2str(LatestRMSValue))
    end
else
    %Copy all the files back
    delete('Ref*.m')
    cd('C:/My Documents/Sunil/Papain_evolution/GP2/Backup')
    copyfile('Ref*.m','C:/My Documents/Sunil/Papain_evolution/GP2')
    delete('Ref*.m')
    cd('C:/My Documents/Sunil/Papain_evolution/GP2')
    clear
end
```

```
function result = newVarConstEqn()
global NewlyCreatedFiles
load('ConstantFiles')
rand('state',sum(100*clock));
%(A) Decide on the degree of complexity needed (i.e. how many terms are you going to have in the equation?)
%(B) Need to allow constants as well as part of this
%(C) Also need to decide on the terms themselves - from TimeIP, TempIP, ConcIP - all of which need to be equally likely
%(D) Need to allow for combinations of the time, temp and conc terms in equations
%(E) Assume that the upper limit for number of terms is 2 in any single round of evolution

Term=0;
Constant=0;
MaxNumberOfTerms = 2;
ActualNumberOfTerms = round(MaxNumberOfTerms * rand(1));
if ActualNumberOfTerms ==0
    ActualNumberOfTerms =1;
end
if ActualNumberOfTerms >MaxNumberOfTerms
    ActualNumberOfTerms =MaxNumberOfTerms;
end

%Construct equation string
NewEquationString=[];
for intCounter = 1:ActualNumberOfTerms
    %Are you going to have a constant number or a variable?
    ConstOrVar = rand(1);
    if ConstOrVar < 0.5
        Constant=1;
    else
        Term=1;
    end

    if Term==1
        %Choose from one of the three variable types present - BIAS THIS
        %TOWARDS TEMPERATURE AND ENZYME CONC - AND AWAY FROM TIME ALTOGETHER
        TermType = rand(1);
        if TermType > 0.875 %ORIGINALLY 0.8
            NewTerm = 'TimeIP';
        elseif TermType > 0.75 %ORIGINALLY 0.4
            NewTerm = 'TempIP';
```

```
        else
            NewTerm = 'ConcIP';
        end
    elseif Constant==1
        NewTerm = 10*rand(1);
        ConstantFiles = [ConstantFiles;NewlyCreatedFiles(1,:)];
        save('ConstantFiles','ConstantFiles')
    end

    if isempty(NewEquationString)==1
        NewEquationString = num2str(NewTerm);
    else
        NewEquationString = strcat('(',NewEquationString, ')', MathsOperator, num2str(NewTerm));
    end
    Term=0;
    Constant=0;
end
result = strcat(NewEquationString,';');
```

```
function result=CalculateRMS(DiskFile)

load (strcat(DiskFile))
load('Data')
fWriter('FirstRef',strcat(Row(1,:),'(ConcIP,TempIP,TimeIP);'));
%fWriter('FirstRef',strcat('Ref0000001','(ConcIP,TempIP,TimeIP);'));
MaximumFabConc=20;
NumberOfData=size(ConcData);
NumberOfData=NumberOfData(1);
CalcYield=[];

for intCounter = 1:NumberOfData
    NextValue = FirstRef(ConcData(intCounter),TempData(intCounter),TimeData(intCounter));
    CalcYield=[CalcYield;NextValue];
end

Difference=[];
%Find the difference between the values at each step
%Square them
%Take means of squares
%Take the root
for intCounter = 1:NumberOfData
    Difference=[Difference;YieldMeasured(intCounter)-CalcYield(intCounter)];
end
%Difference = Difference./MaximumFabConc;
Difference=Difference.^2;

MeanValue=mean(Difference);

RMS = MeanValue.^0.5;
result=RMS;
```

```
function result = MathsOperator()
rand('state',sum(100*clock));
Holder=rand(1);
NumberOfMathsOps=5;

if Holder<1/NumberOfMathsOps
    result='+';
elseif Holder <2/NumberOfMathsOps
    result='-';
elseif Holder <3/NumberOfMathsOps
    result='*';
elseif Holder<4/NumberOfMathsOps
    result='/';
elseif Holder<5/NumberOfMathsOps
    result='^';
end
```

```
function result=NewFile(LatestFileNumber)
global NewlyCreatedFiles

LatestFileName = strcat('Ref',MillionNoFormat(LatestFileNumber));
NewlyCreatedFiles = [NewlyCreatedFiles;LatestFileName];
NewValue = rand(1);
fWriter(LatestFileName, NewValue) %Originally nothing, but this seems to cause some problems
result=LatestFileName;
```

```
function fWriter(name, body)

header= sprintf(strcat('function result=%s(','ConcIP,TempIP,TimeIP','%s)'), name);
func = strcat(header, ')', '\r\n', '\r\n','result=',    num2str(body)      );
filename=strcat(name, '.m');
file=fopen(filename, 'w');
fprintf(file, func);
fclose('all');
```

The Integrated Simulation and Assessment of the Impacts of Process Change in Biotherapeutic Antibody Production

Sunil Chhatre,[†] Carl Jones,[‡] Richard Francis,[‡] Kieran O'Donovan,[‡] Nigel Titchener-Hooker,[†] Anthony Newcombe,[‡] and Eli Keshavarz-Moore*,[†]

The Advanced Centre for Biochemical Engineering, University College London, Torrington Place, London WC1E 7JE, United Kingdom, and Process Development Group, Protherics U.K. Limited, Blaenwaun, Ffostrasol, Llandysul, Wales, SA44 5JT, United Kingdom

Growing commercial pressures in the pharmaceutical industry are establishing a need for robust computer simulations of whole bioprocesses to allow rapid prediction of the effects of changes made to manufacturing operations. This paper presents an integrated process simulation that models the cGMP manufacture of the FDA-approved biotherapeutic CroFab, an IgG fragment used to treat rattlesnake envenomation (Protherics U.K. Limited, Blaenwaun, Ffostrasol, Llandysul, Wales, U.K.). Initially, the product is isolated from ovine serum by precipitation and centrifugation, before enzymatic digestion of the IgG to produce F_{AB} and F_C fragments. These are purified by ion exchange and affinity chromatography to remove the F_C and non-specific F_{AB} fragments from the final venom-specific F_{AB} product. The model was constructed in a discrete event simulation environment and used to determine the potential impact of a series of changes to the process, such as increasing the step efficiencies or volumes of chromatographic matrices, upon product yields and process times. The study indicated that the overall F_{AB} yield was particularly sensitive to changes in the digestive and affinity chromatographic step efficiencies, which have a predicted 30% greater impact on process F_{AB} yield than do the precipitation or centrifugation stages. The study showed that increasing the volume of affinity matrix has a negligible impact upon total process time. Although results such as these would require experimental verification within the physical constraints of the process and the facility, the model predictions are still useful in allowing rapid "what-if" scenario analysis of the likely impacts of process changes within such an integrated production process.

1. Introduction

Within the pharmaceutical and biotechnological industries, there is currently significant interest in optimizing the process development pathway (*1*) as a result of the high costs and long times to market involved (*2*). Companies have to bear significant risks of failure from competitive, regulatory, economic and clinical factors (*3*–*7*). Against this backdrop, a number of techniques now exist that will serve to accelerate and enhance R&D, including genomics and high throughput screening (*8*) and the application of scale-down methodologies (*9*–*12*). Computer simulations have demonstrated a similar potential, enabling rapid and inexpensive exploration of different process options (*13*) and facilitating comparison of alternatives (*14*). Manufacturing models, in particular, can assist in the synthesis of process flowsheets and the study of the knock-on effects of changes in one operation on downstream operations (*15*). Consequently, segments of the process that need improvement and the ways in which this could be achieved can be highlighted.

The proven advantages of simulations in fields such as chemicals or polymers have established modeling as a commonplace technique for process design and evaluation in these sectors (*16*, *17*). The application of process modeling within the bioprocess sector is in its infancy (*18*). Technical models of single stages (*19*), such as fermentation (*20*), filtration (*21*), centrifugation (*22*, *23*), homogenization (*24*, *25*), and ion exchange, size exclusion, affinity and expanded bed chromatography (*26*–*29*) do exist, but development of multistage simulations has been limited largely to 3–4 unit operations, for example (*30*). While such models constitute a useful first step toward the production of robust and predictive modeling packages, it is the simultaneous consideration of all operations and replication of the interactions between them that is needed for the assessment of whole process feasibility. Advances made in this area thus far have centered primarily on establishing modeling methodologies through the study of generic process flowsheets (*31*, *32*). Bioprocess simulations remain considerably underdeveloped for the description of real production processes, and without the presence of such examples in the literature, the industry is unlikely to gain the confidence in modeling that is needed for the potential advantages of the technique to be exploited fully. The aim of this study is to apply modeling to an industrial manufacturing process. The study considers a process operated by Protherics U.K. Limited (Blaenwaun, Ffostrasol, Llandysul, Wales, U.K.) for the production of polyclonal F_{AB} purified from ovine serum used for the treatment of rattlesnake (*Crotalidae*) envenomation (*33*). The FDA-approved product, designated as CroFab, is generated by enzymatic digestion of whole antibody molecules just above the hinge region on their heavy chains to produce two separate

* To whom correspondence should be addressed. Tel: (+44) 20 7679 2961. Fax: (+44) 20 7916 3943. E-mail: e.keshavarz-moore@ucl.ac.uk.
† University College London.
‡ Protherics U.K. Limited.

10.1021/bp0602000 CCC: $33.50
Published on Web 09/23/2006