

# **Novel engineering tools to aid drug discovery processes**

A thesis submitted to the University of London  
for the degree of DOCTOR OF ENGINEERING

by

Rezwan Syed Islam BEng (Hons)

August 2007

The Advanced Centre for Biochemical Engineering  
Department of Biochemical Engineering  
University College London  
Torrington Place  
London WC1E 7JE

UMI Number: U592103

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U592103

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.  
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against  
unauthorized copying under Title 17, United States Code.



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

I, Rezwan Syed Islam, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

---

## Abstract

A major bottleneck in drug discovery is the production of soluble human recombinant protein for functional, biochemical and structural analyses. The level of recombinant protein expression is controlled by a complex relationship between both biological and engineering variables. Due to the inter-play between these variables and standard experimental methods, the identification of the key variables which lead to improved protein expression can sometimes be missed. This thesis presents a framework which underpins the generation of large quantities of soluble recombinant protein in *E. coli* in a rapid and cost-effective manner. To achieve this goal, Design of Experiments (DoE) was first employed in combination with microwell plate (MWP) fermentations to investigate the wide array of protein expression variables. These tools are well suited to high-throughput expression requirements as they afford large savings in time, cost and resource requirements. The information generated from these MWP experiments was then exploited to devise a strategy for reproducing the process within stirred-tank reactors (STRs).

The DoE methodology was first used to identify relevant protein expression variables including fermentation variables (media type and fermentation time), protein induction variables (inducer concentration and induction time) and environmental variables such as oxygen transfer rate, temperature and pH. Ten factors were screened overall at the microwell scale and three were investigated further through optimisation designs. The application of DoE led to a robust understanding of the process and resulted in protein yields five-fold greater than those obtained under standard shake-flask conditions. The most significant factors were post-induction period and shaking speed, the latter of which is strongly related to the mass transfer coefficient,  $k_La$ .

In order to translate this stable and optimised small-scale expression system to a production-scale stirred-tank reactor (STR), an understanding of the engineering parameters at both scales of operation was crucial. This need was complicated by significant differences between the MWPs and STRs such as geometry, mode of

---

aeration and agitation, and the effects of surface tension. In this work, the MWP fermentation results led to the hypothesis that operation at a constant  $k_La$  value would facilitate predictable scale translation. However, there currently exists very little published work on the characterisation of  $k_La$  within MWPs. Miniature oxygen probes were, therefore, used to characterise MWP  $k_La$  values directly via the static gassing-out method over a range of square-well MWP formats and shaking speeds.

This information was then used to translate the performance of a 3ml MWP *E. coli* fermentation, on the basis of matched  $k_La$ , to STR working volumes of 5l and 45l. The efficacy of scale-up was confirmed by performing *F* tests on pairs of profiles for cell growth and expression levels of recombinant firefly luciferase. This rapid, accurate and direct method of  $k_La$  characterisation within MWPs enabled a 15,000-fold direct scale-up of fermentation performance in terms of cell growth and protein expression from MWP to STR.

## **Acknowledgements**

It's a difficult task to thank all the people that have helped me through this and give them the recognition that they deserve.

I would first like to thank Parviz Shamlou for introducing me to this project and encouraging me from the start. A great debt of thanks must go to Dominic Tisi for all his help and enthusiasm. I could not have imagined having a better advisor and mentor for my EngD. I also thank Gary Lye for the tremendous effort he has put into the project and Susana Levy for her unwavering support.

I am grateful to my friends Waqar Hussain, Kais Lakhdar and Samir Ujam for all the tea breaks, lunches and holidays we shared; it goes without saying that they have been a remarkable source of motivation and support over the years.

I would also like to give special thanks to Sonia Sheikh for her endless generosity, inspiration and "raisin" encouragement.

A final thank-you must go to my parents, my brother Shelim and my sister Sophie for their continued support.

---

## Table of contents

<b>LIST OF TABLES .....</b>	<b>12</b>
<b>LIST OF FIGURES .....</b>	<b>17</b>
<b>GLOSSARY.....</b>	<b>24</b>
<b>1 INTRODUCTION.....</b>	<b>26</b>
1.1 DRUG DISCOVERY .....	26
1.1.1 <i>The current direction of drug discovery</i> .....	26
1.1.2 <i>Drug discovery process</i> .....	27
1.1.3 <i>Drug discovery challenges</i> .....	30
1.2 HIGH THROUGHPUT PROTEIN EXPRESSION .....	31
1.2.1 <i>Choice of expression system</i> .....	31
1.2.1.1 Bacterial expression systems.....	32
1.2.1.2 Yeast expression systems.....	33
1.2.1.3 Insect cell/baculovirus expression systems.....	33
1.2.1.4 Mammalian expression systems.....	34
1.2.1.5 Cell-free expression systems.....	34
1.2.2 <i>Mechanism of protein expression</i> .....	35
1.2.3 <i>Factors influencing soluble protein expression</i> .....	36
1.2.4 <i>Microscale processing</i> .....	40
1.2.5 <i>Reporter genes</i> .....	43
1.3 DESIGN OF EXPERIMENTS.....	46
1.3.1 <i>Description and benefits over traditional approaches</i> .....	46
1.3.2 <i>Pre-experimental steps</i> .....	49
1.3.2.1 Experimental objective.....	49
1.3.2.2 Specification of factors .....	49
1.3.2.3 Specification of responses.....	50
1.3.2.4 Model selection .....	50

---

---

1.3.2.5	Generation of an experimental design and creation of a worksheet .....	50
1.3.3	<i>Post-experimental steps</i> .....	51
1.3.3.1	Evaluation of raw data .....	51
1.3.3.2	Regression analysis and model refinement .....	53
1.3.3.3	Use of regression model .....	53
1.3.4	<i>Relevant studies incorporating DoE</i> .....	53
1.4	SCALE-UP OF PROTEIN EXPRESSION .....	54
1.5	THESIS AIMS AND OBJECTIVES .....	57
<b>2</b>	<b>MATERIALS AND METHODS</b> .....	<b>59</b>
2.1	INTRODUCTION .....	59
2.2	GENERAL METHODS .....	59
2.2.1	<i>Media preparation</i> .....	59
2.2.2	<i>Bacterial expression system</i> .....	60
2.2.3	<i>Inoculum preparation</i> .....	61
2.3	METHODS FOR OPTIMISATION OF SOLUBLE PROTEIN EXPRESSION AT THE MICROWELL SCALE .....	61
2.3.1	<i>Microwell plate (MWP) fermentations</i> .....	61
2.3.2	<i>Design of MWP familiarisation experiments</i> .....	64
2.3.3	<i>Design of MWP screening experiments</i> .....	66
2.3.4	<i>Design of MWP optimisation experiments</i> .....	66
2.3.5	<i>Reference shake flask fermentations</i> .....	67
2.4	METHODS FOR CHARACTERISATION OF BIOREACTOR MASS TRANSFER COEFFICIENTS ( $K_{LA}$ ) .....	68
2.4.1	<i>Measurement of MWP <math>k_{La}</math> values</i> .....	68
2.4.2	<i>Specific air-liquid surface area measurements in MWPs</i> .....	70
2.4.3	<i>Measurement of stirred-tank reactor (STR) <math>k_{La}</math> values</i> .....	71
2.5	SCALE-UP OF MICROWELL PLATE FERMENTATIONS .....	73
2.5.1	<i>MWP fermentations</i> .....	73
2.5.2	<i>STR fermentations at equivalent <math>k_{La}</math> conditions</i> .....	74
2.5.3	<i>Statistical analyses of fermentation kinetic profiles</i> .....	76

---



---

2.6	ANALYTICAL PROCEDURES .....	77
2.6.1	<i>Quantification of biomass concentration</i> .....	77
2.6.2	<i>Quantification of glycerol concentration</i> .....	79
2.6.3	<i>Quantification of relative soluble firefly luciferase activity</i> .....	80
2.6.4	<i>Analytical software</i> .....	83
<b>3</b>	<b>A FRAMEWORK FOR THE RAPID OPTIMISATION OF SOLUBLE PROTEIN EXPRESSION IN <i>E. COLI</i></b> .....	<b>84</b>
3.1	AIMS AND OBJECTIVES .....	84
3.2	FRAMEWORK DESIGN .....	84
3.2.1	<i>Introduction</i> .....	84
3.2.2	<i>Component specification</i> .....	84
3.2.3	<i>Framework structure</i> .....	86
3.3	FRAMEWORK IMPLEMENTATION.....	86
3.3.1	<i>Protein expression characterisation</i> .....	86
3.3.1.1	Background .....	86
3.3.1.2	Familiarisation .....	88
3.3.1.3	Screening.....	90
3.3.1.4	Optimisation and process mapping .....	91
3.3.2	<i>Scale-up strategy: laboratory scale</i> .....	91
3.3.3	<i>Scale-up strategy: pilot scale</i> .....	92
3.4	SUMMARY .....	92
<b>4</b>	<b>OPTIMISATION OF SOLUBLE PROTEIN EXPRESSION AT THE MICROWELL SCALE<sup>†</sup></b> .....	<b>93</b>
4.1	AIMS AND OBJECTIVES .....	93
4.2	EXPERIMENTAL FAMILIARISATION .....	93
4.2.1	<i>Specification of variables</i> .....	93
4.2.2	<i>Development of response measurement techniques</i> .....	94
4.2.3	<i>Method development through experimentation</i> .....	95
4.3	SCREENING EXPERIMENTS .....	97
4.4	OPTIMISATION EXPERIMENTS .....	101

---

---

4.5	REGRESSION ANALYSIS AND PROCESS MODELLING .....	106
4.6	SUMMARY .....	111
<b>5</b>	<b>SCALE-UP OF MWP FERMENTATION PERFORMANCE TO 7.5 L STR SCALE<sup>†</sup> .....</b>	<b>114</b>
5.1	AIMS AND OBJECTIVES .....	114
5.2	MEASUREMENT AND CORRELATION OF MWP $K_{LA}$ VALUES .....	115
5.3	MEASUREMENT AND CORRELATION OF 7.5 L STR $K_{LA}$ VALUES .....	118
5.4	FERMENTATIONS AT MATCHED $K_{LA}$ VALUES .....	120
5.4.1	<i>Selection of <math>k_{La}</math> values and experimental conditions</i> .....	120
5.4.2	<i>MWP fermentation kinetics</i> .....	122
5.4.3	<i>7.5 l STR fermentation kinetics</i> .....	125
5.4.4	<i>Scale comparisons</i> .....	128
5.5	SUMMARY .....	130
<b>6</b>	<b>SCALE-UP OF PROTEIN EXPRESSION TO 75 L STR SCALE<sup>†</sup> ...</b>	<b>132</b>
6.1	AIMS AND OBJECTIVES .....	132
6.2	MEASUREMENT AND CORRELATION OF 75 L STR $K_{LA}$ VALUES .....	132
6.2.1	<i>Measurement of <math>k_{La}</math> values</i> .....	132
6.2.2	<i>Comparison of <math>k_{La}</math> values with other scales</i> .....	135
6.3	FERMENTATIONS AT MATCHED $K_{LA}$ VALUES .....	136
6.3.1	<i>Selection of <math>k_{La}</math> values and experimental conditions</i> .....	136
6.3.2	<i>MWP fermentation kinetics</i> .....	137
6.3.3	<i>75 l STR fermentation kinetics</i> .....	137
6.3.4	<i>Scale comparisons</i> .....	140
6.4	SUMMARY .....	142
<b>7</b>	<b>MANAGEMENT BENEFITS OF THE PROPOSED FRAMEWORK<sup>†</sup></b>	<b>144</b>
7.1	AIMS AND OBJECTIVES .....	144
7.2	PROTEIN EXPRESSION OPTIMISATION .....	144

---

---

7.2.1	<i>Background and general assumptions</i> .....	144
7.2.2	<i>Scenario 1: current framework</i> .....	145
7.2.3	<i>Scenario 2: current framework minus the MWP component</i> .....	146
7.2.4	<i>Scenario 3: current framework minus the DoE component</i> .....	147
7.2.5	<i>Scenario comparison</i> .....	147
7.3	GENERATION OF PROTEIN SAMPLES.....	149
7.3.1	<i>Background and general assumptions</i> .....	149
7.3.2	<i>Scenario 1: current framework</i> .....	150
7.3.3	<i>Scenario 2: traditional SF fermentation approach</i> .....	152
7.3.4	<i>Scenario comparison</i> .....	154
7.4	SUMMARY .....	155
<b>8</b>	<b>DEVELOPMENT OF THE PROPOSED FRAMEWORK TO FACILITATE BIOPROCESS VALIDATION<sup>†</sup></b> .....	<b>156</b>
8.1	AIMS AND OBJECTIVES .....	156
8.2	BIOPROCESS DEVELOPMENT THROUGH SIX SIGMA (DMAIC).....	157
8.2.1	<i>Introduction</i> .....	157
8.2.2	<i>Definition phase</i> .....	157
8.2.3	<i>Measurement phase</i> .....	158
8.2.4	<i>Analysis phase</i> .....	160
8.2.5	<i>Improvement phase</i> .....	161
8.2.6	<i>Control phase</i> .....	162
8.3	SUMMARY .....	162
<b>9</b>	<b>CONCLUSIONS AND FUTURE WORK</b> .....	<b>164</b>
9.1	CONCLUSIONS .....	164
9.2	FUTURE WORK.....	167
	<b>APPENDIX A – BIOREACTOR MIXING DATA</b> .....	<b>169</b>

---

**APPENDIX B - FERMENTATION RESPONSE MODELLING .....171**

B.1	BASIS .....	171
B.2	MWP AND 7.5 L STR: [BIOMASS] COMPARISON .....	171
B.3	MWP AND 7.5 L STR: [GLYCEROL] COMPARISON .....	172
B.4	MWP AND 7.5 L STR: SOLUBLE PROTEIN YIELD COMPARISON .....	173
B.5	MWP AND 75 L STR: [BIOMASS] COMPARISON .....	174
B.6	MWP AND 75 L STR: [GLYCEROL] COMPARISON .....	175
B.7	MWP AND 75 L STR: SOLUBLE PROTEIN YIELD COMPARISON .....	176

**APPENDIX C – EXPERIMENTAL COSTS AND TIMEFRAMES .....177**

C.1	COST OF CONSUMABLES.....	177
C.2	OPTIMAL SCHEDULES FOR THE THEORETICAL MWP FERMENTATION SCENARIOS (SECTIONS 7.2.2 AND 7.2.4).....	178
C.3	OPTIMAL SCHEDULES FOR THE THEORETICAL SHAKE-FLASK (SF) FERMENTATION SCENARIO (SECTION 7.2.3) .....	180
C.4	DISPOSABLE MWP REQUIREMENTS .....	182
C.5	COST SCENARIOS.....	183
C.5.1	<i>Background</i> .....	183
C.5.2	<i>Scenario one: current framework (Section 7.2.2)</i> .....	184
C.5.3	<i>Scenario two: current framework minus the MWP component (Section 7.2.3)</i> .....	184
C.5.4	<i>Scenario 3: current framework minus the DoE component (Section 7.2.4)</i> .....	187

**APPENDIX D – LIST OF PUBLICATIONS.....189****REFERENCES ..... 190**

## List of tables

- Table 1.1** Summary of microscale research into mixing and oxygen transfer carried out in a range of small-scale vessels including the shake flask (SF), miniature stirred bioreactor (MSBR) and microwell plate (MWP). <sup>1</sup>The terms *static gassing-out* and *dynamic gassing-out* are used in accordance with the definitions provided by Stanbury and Whitaker (1993)..... 44
- Table 1.2** Pictorial representation of various experimental designs. The star at the centre of each design represents centre-point experiments..... 52
- Table 2.1** Description of MWP geometries used throughout this work. All dimensions are given in mm, correct to the nearest mm. Drawings not to scale..... 63
- Table 2.2** Specification of factors and settings investigated within familiarisation, screening and optimisation experiments. Pre-I and post-I are abbreviations for pre-induction and post-induction, respectively. Plate P1 had 48 rectangular wells with flat bases, plate P2 had 24 square wells with round bases and plate P3 had 24 square wells with pyramidal bases. <sup>a</sup>The familiarisation experiments had a lower shaking speed of 50rpm. <sup>b</sup>An upper liquid fill volume of 2 ml was used in conjunction with the 48-well plate to prevent splashing..... 63
- Table 2.3** D-optimal design matrix for the initial assessment of soluble protein yield at the familiarisation stage. Quantitative variables are coded at low (-1), medium (0) and high (+1) levels. .... 65
- Table 2.4** Specification of factors and settings investigated during the scale-up of microwell plate fermentations. Pre-I and post-I are abbreviations for pre-induction and post-induction, respectively. <sup>a,b</sup>The MWP fermentation conditions described here have been adopted from experiments N32 and N36, respectively, from the optimisation experiments (Section 4.4)..... 74
- Table 2.5** Description of functions used to model each fermentation response profile, where  $y$  is the fermentation response level at time  $t$ ,  $y_0$  is the initial response level and  $a$ ,  $b$  and  $c$  are arbitrary constants. .... 77
- Table 4.1** Specification of factors and settings investigated within familiarisation, screening and optimisation experiments. Pre-I and post-I are abbreviations for pre-induction and post-induction, respectively. Plate P1 had 48 rectangular wells with flat bases, plate P2 had 24 square wells with round bases and plate P3 had 24 square wells with pyramidal bases. <sup>a</sup>The familiarisation experiments had a lower shaking speed of 50rpm. <sup>b</sup>An upper liquid fill volume of 2 ml was used in conjunction with the 48-well plate to prevent splashing..... 95

- 
- Table 4.2** D-optimal screening design matrix for the initial assessment of soluble protein yield. Quantitative variables are coded at low (-1), medium (0) and high (+1) levels according to the factor levels specified in Table 4.1. .... 98
- Table 4.3** CCF design matrix for the optimisation of soluble protein yield.  $DCW_i$  and  $DCW_f$  refer to dry cell weight measurements immediately prior to induction and immediately following the induction period respectively. All variables are coded at low (-1), medium (0) and high (+1) levels according to the factor levels specified in Table 4.1. Predicted mean soluble protein yields are calculated from Equation 4.1..... 103
- Table 4.4** Estimated regression coefficients of the reduced optimisation model for soluble protein yield (Equation 4.1). The corresponding p-values and confidence intervals are also shown. Variables as described in Table 4.1. 107
- Table 4.5** ANOVA results for analysis of the reduced optimisation model (Equation 4.1) for soluble protein yield. .... 108
- Table 5.1** Summary of kinetic parameters for fermentations performed at the microwell and laboratory scales at matched  $k_{La}$  values. Parameters calculated from Figures 5.5 to 5.8 inclusive. NA indicates data not obtainable. .... 123
- Table 5.2** Comparison of microwell fermentation kinetic profiles (Figure 5.5) to those from the laboratory scale fermentation (Figure 5.7) at matched  $k_{La}$  conditions ( $247 \text{ h}^{-1}$ ).  $P$  values shown are from  $F$  tests on pairs of profiles for biomass concentration, soluble protein yield and substrate concentration as described in Section 5.4.4. The null hypothesis, which states that there is no overall difference between each pair of profiles, is accepted if  $p > 0.05$ ... 130
- Table 6.1** Summary of kinetic parameters for fermentations performed at the microwell and pilot scales at matched  $k_{La}$  values. Parameters calculated from Figures 6.4 and 6.5. NA indicates data not obtainable. .... 138
- Table 6.2** Comparison of microwell fermentation kinetic profiles (Figure 6.4) to those from the pilot plant scale fermentation (Figure 6.5) at matched  $k_{La}$  conditions ( $247 \text{ h}^{-1}$ ).  $P$  values shown are from  $F$  tests on pairs of profiles for biomass concentration, soluble protein yield and substrate concentration as described in Section 6.3.4. The null hypothesis, which states that there is no overall difference between each pair of profiles, is accepted if  $p > 0.05$ ... 141
- Table 7.1** Cost analysis summary for Scenario 1: current framework. These costs have been summarised from the more detailed analysis of Section C.5.2. 146
- Table 7.2** Cost analysis summary for Scenario 2: current framework minus the MWP component. These costs have been summarised from the more detailed analysis of Section C.5.3. .... 146
-

---

<b>Table 7.3</b> Cost analysis summary for Scenario 3: current framework minus the DoE component. These costs have been summarised from the more detailed analysis of Section C.5.4.....	147
<b>Table 7.4</b> Summary of costs for each expression optimisation scenario. Details of each scenario are described in Sections 7.2.2 through 7.2.4 inclusive..	148
<b>Table 7.5</b> Typical protein sample requirements of drug discovery. Each stage of drug discovery aims to reduce the number of investigated protein constructs by identifying those which demonstrate several key properties. These include functional activity and ease of protein expression, purification and characterisation. Protein requirements may differ widely between different drugs and different drug discovery companies; the requirements shown here are representative of the operating ranges of Astex Therapeutics Ltd.....	150
<b>Table 7.6</b> Sample calculation for scenario one: current framework. Calculations based on average protein sample requirements (Table 7.5) and a mean soluble protein yield of 3 mg.l <sup>-1</sup> .....	151
<b>Table 7.7</b> Sample calculation for scenario two: traditional SF fermentation approach. Calculations based on average protein sample requirements (Table 7.5) and a mean soluble protein yield of 0.5 mg.l <sup>-1</sup> .....	153
<b>Table 8.1</b> Statement of goals at the strategic, process and project levels.....	158
<b>Table 8.2</b> List of key responses which may be used to measure process performance.....	159
<b>Table 8.3</b> List of the sources of variation to which inoculum SF experiments may be susceptible with corresponding suggested control measures. ....	160
<b>Table 8.4</b> Possible sources of observed variation in the MWP fermentation performance (Chapter 4) with corresponding suggested control measures. ....	161
<b>Table 8.5</b> Possible sources of observed underperformance in the DoE analysis (Section 4.5) with corresponding suggested improvements. ....	162
<b>Table A.1</b> MWP $k_{LA}$ values (h <sup>-1</sup> ) measured as described in Section 2.4.1. ....	169
<b>Table A.2</b> MWP $a_f/a_i$ values (dimensionless) measured as described in Section 2.4.2. NA indicates data not obtainable. ....	169
<b>Table A.3</b> STR $k_{LA}$ values (h <sup>-1</sup> ) measured as described in Section 2.4.3. ....	169
<b>Table A.4</b> STR gassed-power, $P_g$ , values (W.m <sup>-3</sup> ) calculated using Equation 5.6 .....	170
<b>Table B.1</b> Summary of model parameters correct to 3 s.f.....	171

---

---

<b>Table B.2</b> Summary of modelling statistics correct to 3 s.f. ....	171
<b>Table B.3</b> Summary of model parameters correct to 3 s.f.....	172
<b>Table B.4</b> Summary of modelling statistics correct to 3 s.f. ....	172
<b>Table B.5</b> Summary of model parameters correct to 3 s.f.....	173
<b>Table B.6</b> Summary of modelling statistics correct to 3 s.f. ....	173
<b>Table B.7</b> Summary of model parameters correct to 3 s.f.....	174
<b>Table B.8</b> Summary of modelling statistics correct to 3 s.f. ....	174
<b>Table B.9</b> Summary of model parameters correct to 3 s.f.....	175
<b>Table B.10</b> Summary of modelling statistics correct to 3 s.f. ....	175
<b>Table B.11</b> Summary of model parameters correct to 3 s.f.....	176
<b>Table B.12</b> Summary of modelling statistics correct to 3 s.f. ....	176
<b>Table C.1</b> Cost of LB media .....	177
<b>Table C.2</b> Cost of TB media .....	177
<b>Table C.3</b> Cost of GM9Y media .....	177
<b>Table C.4</b> Cost of MWP's including cost of breathable membranes (£0.83 per membrane).....	177
<b>Table C.5</b> Cost of luciferase assay buffer (LAB).....	178
<b>Table C.6</b> Minimum number of MWP's required at the familiarisation stage when experiments are arranged as described in Figure C.1. ....	182
<b>Table C.7</b> Minimum number of MWP's required at the screening stage when experiments are arranged as described in Figure C.2. ....	183
<b>Table C.8</b> Minimum number of MWP's required at the optimisation stage when experiments are arranged as described in Figure C.3 .....	183
<b>Table C.9</b> Cost analysis for scenario one: current framework. <sup>1</sup> Unit costs calculated in Section C.1. <sup>2</sup> Requirements are in accordance with experimental methods (Chapter 2). <sup>3, 4</sup> Requirements are calculated in Sections C.4 and C.2, respectively.....	185
<b>Table C.10</b> Cost analysis for scenario two: current framework minus the MWP component. <sup>1</sup> Unit costs calculated in Section C.1. <sup>2</sup> Minimum total fermentation times calculated in Section C.3.....	186

---



**Table C.11** Cost analysis for scenario three: current framework minus the DoE component. <sup>1</sup>Unit costs calculated in Section C.1..... 188

## List of figures

- Figure 1.1** Generic process of drug discovery.....2
- Figure 1.2** Response contour plot of a system influenced by two factors, investigated using OFAT (a) and DoE (b). Individual experiments are also shown [ • ]. .....22
- Figure 1.3** The DoE procedure in which system or process is modelled, enabling the prediction of response values from defined factor levels.....23
- Figure 2.1** Cross-sectional diagram illustrating the setup of a microwell plate modified with the oxygen sensor system as described in Section 2.4.1.....44
- Figure 2.2** Dimensions of BioFlo 110 7.5 l STR used in this study. All dimensions are given in millimetres. *A* = impeller shaft; *B* = impeller disc; *C* = impeller blade; *D* = baffle. Impeller spacings are set according to the vessel manufacturer's guidelines. ....47
- Figure 2.3** Calibration curve for optical density and dry cell measurements of *E. coli* grown in LB ( ⊖ ), GM9Y ( ⊞ ) and TB ( ⊠ ) medium. The following lines of best-fit are obtained:  $y = 0.342x$ ,  $R^2 = 0.967$  (LB),  $y = 0.387x$ ,  $R^2 = 0.998$  (GM9Y),  $y = 0.497x$ ,  $R^2 = 0.964$  (TB). .....54
- Figure 2.4** Example of standard luminescence profiles from samples of the replicate MWP fermentations described in Section 2.3.1, taken at identical time points. Measurements made as described in Section 2.6.3. ....56
- Figure 2.5** Typical luminescence profiles for the replicate samples within each MWP fermentation. For samples with particularly high soluble protein yields, the reaction proceeded very fast and the peak was often missed (a). In these scenarios, the lysates were first diluted 20-fold in 100 mM Tris pH 7.8 prior to the addition of LAB which resulted in more standard profiles (b). .....58
- Figure 3.1** Overview of the proposed framework for large-scale optimised protein expression. The numbered arrows represent information streams. The specification of factors and responses is determined by the particular target protein and expression system (1) and DoE is used to design the subsequent microscale experiments (2). Several rounds of expression studies are required to determine the optimal process conditions in a time and cost efficient. The information gathered from each round is used to generate subsequent experimental designs (3) until the optimal conditions are found. The process knowledge which emerges from the DoE / microscale experiments should provide insight into larger (laboratory) scale performance and thus it should be incorporated into the appropriate scale-up strategy (4). Similarly, the knowledge gained from the laboratory scale

strategy should inform the subsequent pilot scale strategy (5), thereby enabling the production of gram quantities of target protein (6). ..... 62

**Figure 3.2** Implementation sequence of the stages of DoE applied in this work (arrows) towards the goal of obtaining a reliable process map. As each stage is completed, the number of potentially significant variables is reduced whilst the degree of process characterisation is increased. .... 64

**Figure 4.1** Column chart illustrating the variation of soluble protein yield from MWP familiarisation experiments. Experiments performed as described in Section 2.3.1..... 71

**Figure 4.2** Coefficient plot with confidence intervals (set at 95%) for soluble protein yield obtained from screening experiments. The size of each bar (or coefficient) represents the relative importance of that factor upon soluble protein expression. A positive coefficient for a quantitative factor indicates that an increase in that factor, relative to its centre-point level, would result in an increase in soluble protein yield. The opposite is true for negative coefficients. Qualitative factors, such as medium type, have no centre-point levels and so the coefficients displayed are relative to one another. Factors and experimental conditions as described in Table 4.1..... 74

**Figure 4.3** Coefficient plot for final optimisation model (Equation 4.1), complete with confidence intervals (set at 95 %) for soluble protein yield ( $Y$  transformed as  $Y^{0.25}$ ). The original model consisted of 27 terms. However, terms whose confidence interval included the value zero were deemed insignificant and so were removed from the model in a hierarchical manner. Factors and experimental conditions as described in Table 4.1..... 80

**Figure 4.4** Response surface plot showing influence of pre-I and post-I shaking speed on soluble protein yield (all other optimisation factors at their centre-point levels). Response surface generated using Equation 4.1. .... 81

**Figure 4.5** Parity plot showing measured and predicted soluble protein yields for optimisation experiments. Predicted values obtained from Equation 4.1. Solid line represents line of parity. .... 82

**Figure 4.6** Histogram of responses for (soluble protein yield)<sup>0.25</sup> from the optimisation experiments (Table 4.3). .... 84

**Figure 4.7** Contour plot showing variation of soluble protein yield ( $\text{RLU}\cdot\text{ml}^{-1}$ ) with respect to pre-induction dry cell weight ( $\text{DCW}_i$ ) and final dry cell weight ( $\text{DCW}_f$ ). Model used to generate contours was a least squares quadratic fit:  $\log_{10}(\text{soluble protein yield}) = 11.36 - 0.13x(\text{DCW}_i) + 0.50x(\text{DCW}_f) - 0.12x(\text{DCW}_i)^2 - 0.02x(\text{DCW}_f)^2 + 0.09x(\text{DCW}_i)x(\text{DCW}_f)$ . Data used to generate model are taken from Table 4.3 and are superimposed over the contour plot. Dashed line represents line of parity. .... 86

**Figure 4.8** Comparison of maximum soluble protein yields obtained from the familiarisation, screening and optimisation experiments. Dashed line

indicates the mean soluble protein yield measured from shake flask experiments under standard expression conditions (Section 2.3.5). ..... 88

**Figure 5.1** Example DOT profiles obtained across the full range of shaking / agitation speeds within (a) the MWP (plate P3) and (b) the 7.5 l STR. Experiments were performed as described in Section 2.4. A logarithmic transformation of time is displayed on the x-axes such that the full range of profiles for each bioreactor may be easily distinguished in one figure..... 93

**Figure 5.2** Comparison of predicted and measured MWP  $k_La$  and  $a_f/a_i$  values: (a) the correlation for  $k_La$  presented by Doig *et al.* (2005) (Equation 5.1) was used to predict  $k_La$  values for plates P1 [ ● ], P2 [ ▲ ] and P3 [ ■ ]; (b) Equation 5.2 was used to predict  $a_f/a_i$  values for plates P1 [ ○ ], P2 [ △ ] and P3 [ □ ] and the modified Doig correlation (Equation 5.3) was fitted to these data and used to predict  $k_La$  values for plates P1 [ ● ], P2 [ ▲ ] and P3 [ ■ ]. Values predicted from Equations 5.2 and 5.3 lie close to the line of parity [ ] and so are in good agreement with the corresponding experimental values.  $k_La$  values were measured as described in Section 2.4.1 and values for  $a_f/a_i$  were measured as described in Section 2.4.2 and. Error bars indicate one standard deviation around the mean..... 94

**Figure 5.3** Comparison of experimental 7.5 l STR  $k_La$  values with those predicted by van't Riet (1979) (Equations 5.4 and 5.5) for pure water [ ... ] and strong ionic solutions [ --- ] at an air flow rate of 1 vvm. Values for  $P_g$  are calculated from Equation 5.6. Error bars indicate one standard deviation around the mean. .... 96

**Figure 5.4** Variation of experimental  $k_La$  values with shaking speed for plate P3 [ ■ ] and with agitation speed for 7.5 l STR [ □ ]. The dashed lines represent boundaries on the common range of  $k_La$  values measured at both scales. The solid lines represent least-squares model fits of the  $k_La$  data, described by Equations 5.8 and 5.9. Error bars indicate one standard deviation around the mean. .... 98

**Figure 5.5** MWP (plate P3) fermentation kinetic profiles at high  $k_La$  ( $247 \text{ h}^{-1}$ ). Profiles show biomass concentration [ -⊖- ], soluble protein yield [ ● ], glycerol concentration [ -△- ] and DOT [ ... ]. Arrow indicates time of induction. Experiments performed as described in Section 2.5.1. .... 100

**Figure 5.6** MWP (plate P3) fermentation kinetic profiles at low  $k_La$  ( $55 \text{ h}^{-1}$ ). Profiles show biomass concentration [ -⊖- ], soluble protein yield [ ● ] and DOT [ ... ]. Arrow indicates time of induction. Experiments performed as described in Section 2.5.1. .... 101

**Figure 5.7** 7.5 l STR fermentation kinetic profiles at high  $k_La$  ( $247 \text{ h}^{-1}$ ): (a) offline data shows biomass concentration [ -⊖- ], soluble protein yield [ ● ] and glycerol concentration [ -△- ]; (b) online data shows DOT [ ... ], OUR [ — ], CER [ — ] and pH [ --- ]. Arrows indicate time of induction. Experiments performed as described in Section 2.5.2. .... 103

- Figure 5.8** 7.5 l STR fermentation kinetic profiles at low  $k_La$  ( $55 \text{ h}^{-1}$ ): (a) offline data shows biomass concentration [  $\ominus$  ] and soluble protein yield [  $\bullet$  ]; (b) online data shows DOT [  $\cdots$  ], OUR [  $—$  ], CER[  $—$  ] and pH [  $---$  ]. Arrows indicate time of induction. Experiments performed as described in Section 2.5.2..... 104
- Figure 5.9** Parity plots comparing fermentation kinetics between MWP (plate P3) and 7.5 l STR for biomass yield [  $\square$  ], soluble protein yield [  $\bullet$  ] and glycerol concentration ( $\times 2$ ) [  $\Delta$  ] at  $k_La$  values of (a)  $247 \text{ h}^{-1}$  and (b)  $55 \text{ h}^{-1}$ . Data taken from Figures 5.5 to 5.8 inclusive. .... 106
- Figure 6.1** Example DOT profiles obtained across the full range of agitation speeds within the 75 l STR. Experiments were performed as described in Section 2.4.3. A logarithmic transformation of time is displayed on the x-axes such that the full range of profiles for each bioreactor may be easily distinguished in one figure..... 110
- Figure 6.2** Comparison of experimental 75 l STR  $k_La$  values with those predicted by van't Riet (1979) (Equations 5.4 and 5.5) for pure water [  $\cdots$  ] and strong ionic solutions [  $---$  ] at an air flow rate of 1 vvm. Values for  $P_g$  are calculated from Equation 5.6. Error bars indicate one standard deviation around the mean. .... 111
- Figure 6.3** Variation of experimental  $k_La$  values with shaking speed for plate P3 [  $\blacksquare$  ] and with agitation speed for 7.5 l STR [  $\square$  ] and 75 l STR [  $\blacktriangle$  ]. The dashed lines represent boundaries on the common range of  $k_La$  values measured at both the MWP and 75 l STR scales. The solid lines represent least-squares model fits of the  $k_La$  data, described by Equations 5.8, 5.9 and 6.1. Error bars indicate one standard deviation around the mean. .... 112
- Figure 6.4** MWP (plate P3) fermentation kinetic profiles at high  $k_La$  ( $247 \text{ h}^{-1}$ ). Profiles show biomass concentration [  $\ominus$  ], soluble protein yield [  $\bullet$  ] and glycerol concentration [  $-\Delta-$  ]. Arrow indicates time of induction. Experiments performed as described in Section 2.5.1. .... 115
- Figure 6.5** 75 l STR fermentation kinetic profiles at high  $k_La$  ( $247 \text{ h}^{-1}$ ): (a) offline data shows biomass concentration [  $\ominus$  ], soluble protein yield [  $\bullet$  ] and glycerol concentration [  $-\Delta-$  ]; (b) online data shows DOT [  $\cdots$  ], OUR [  $—$  ], CER[  $—$  ] and pH [  $---$  ]. Arrows indicate time of induction. Experiments performed as described in Section 2.5.2. .... 116
- Figure 6.6** Parity plot comparing fermentation kinetics at 'high'  $k_La$  value of  $247 \text{ h}^{-1}$  between MWP (plate P3) and 75 l STR for biomass yield [  $\square$  ], soluble protein yield [  $\bullet$  ] and glycerol concentration ( $\times 2$ ) [  $\Delta$  ]. Data taken from Figures 6.4 and 6.5. .... 118
- Figure 7.1** Fermentation time window for the current framework over the full range of protein sample requirements (Table 7.5). .... 129

---

<b>Figure 7.2</b> Fermentation time window for the traditional SF fermentation framework over the full range of protein sample requirements (Table 7.5). .....	130
<b>Figure 7.3</b> Increased drug sales revenue window of the current framework in comparison to the traditional SF framework over the full range of protein sample requirements (Table 7.5), assuming average daily sales of \$1 million per drug, per day. The results displayed in this figure are calculated from the data in Figures 7.1 and 7.2. ....	131
<b>Figure 8.1</b> Typical flow of a DMAIC project (Creveling, 2006). In practice, there may be significant overlap between each phase and an iterative approach to Six Sigma (indicated by the dashed line) may be required (Hayler and Nichols, 2005). ....	134
<b>Figure 8.2</b> Process map illustrating the overall scope of the Six Sigma project, complete with the flow of process steps, factors and responses. The green arrows represent responses and the red and blue arrows represent controlled and uncontrolled factors, respectively; response measurements from the inoculum SF fermentation will serve as uncontrolled factors in the MWP fermentation step. ....	136
<b>Figure B.1</b> Logistic curves fitted separately to MWP [biomass] data ( ■ ) and 7.5 l STR [biomass] data ( □ ). ....	148
<b>Figure B.2</b> Single logistic curve fitted to combined MWP and 7.5 l STR [biomass] data. ....	148
<b>Figure B.3</b> One site competition curves fitted separately to MWP [glycerol] data ( ■ ) and 7.5 l STR [glycerol] data ( □ ). ....	149
<b>Figure B.4</b> Single one site competition curve fitted to combined MWP and 7.5 l STR [glycerol] data. ....	149
<b>Figure B.5</b> Extreme peak function curves fitted separately to MWP soluble protein yield data ( ■ ) and 7.5 l STR soluble protein yield data ( □ ). ....	150
<b>Figure B.6</b> Single extreme peak function curve fitted to combined MWP and 7.5 l STR soluble protein yield data. ....	150
<b>Figure B.7</b> Logistic curves fitted separately to MWP [biomass] data ( ■ ) and 75 l STR [biomass] data ( □ ). ....	151
<b>Figure B.8</b> Single logistic curve fitted to combined MWP and 75 l STR [biomass] data. ....	151
<b>Figure B.9</b> One site competition curves fitted separately to MWP [glycerol] data ( ■ ) and 75 l STR [glycerol] data ( □ ). ....	152

---

- 
- Figure B.10** Single one site competition curve fitted to combined MWP and 75 l STR [glycerol] data..... 152
- Figure B.11** Extreme peak function curves fitted separately to MWP soluble protein yield data ( ■ ) and 75 l STR soluble protein yield data ( □ )..... 153
- Figure B.12** Single extreme peak function curve fitted to combined MWP and 75 l STR soluble protein yield data..... 153
- Figure C.1** An optimal schedule for the MWP familiarisation experiments described in Table 2.3, showing minimum theoretical number of hours required to complete all fermentations. Experiments from Table 2.3 are arranged in fermentation batches according to those with identical pre-I and post-I shaking speeds. Numbers to the right of each bar represent number of fermentations per batch (including replicates). ..... 155
- Figure C.2** An optimal schedule for the MWP screening experiments described in Table 4.2, showing minimum theoretical number of hours required to complete all fermentations. Experiments from Table 4.2 are arranged in fermentation batches according to those with identical pre-I and post-I shaking speeds. Numbers to the right of each bar represent number of fermentations per batch (including replicates). ..... 156
- Figure C.3** An optimal schedule for the MWP optimisation experiments described in Table 4.3, showing minimum theoretical number of hours required to complete all fermentations. Experiments from Table 4.3 are arranged in fermentation batches according to those with identical pre-I and post-I shaking speeds. Numbers to the right of each bar represent number of fermentations per batch (including replicates). ..... 156
- Figure C.4** An optimal schedule for the MWP optimisation experiments, identical to that of Figure C.3, in which the *maximum* number of fermentations is performed (72 wells in parallel). Numbers to the right of each bar represent number of fermentations per batch (including replicates). ..... 157
- Figure C.5** An optimal schedule showing the minimum theoretical number of hours required to complete the familiarisation experiments described in Table 2.3, if performed in SFs. Experiments from Table 2.3 are arranged in fermentation batches according to those with identical pre-I and post-I shaking speeds. Numbers to the right of each bar represent number of SF fermentations per batch (including replicates). ..... 158
- Figure C.6** An optimal schedule showing the minimum theoretical number of hours required to complete the screening experiments described in Table 4.2, if performed in SFs. Experiments from Table 4.2 are arranged in fermentation batches according to those with identical pre-I and post-I shaking speeds. Numbers to the right of each bar represent number of SF fermentations per batch (including replicates). ..... 158
-

---

**Figure C.7** An optimal schedule showing the minimum theoretical number of hours required to complete the optimisation experiments described in Table 4.3, if performed in SFs. Experiments from Table 4.3 are arranged in fermentation batches according to those with identical pre-I and post-I shaking speeds. Numbers to the right of each bar represent number of SF fermentations per batch (including replicates)..... 159



## Glossary

$\alpha$	significance level
$\beta_0$	constant term
$\beta_i$	linear effect
$\beta_{ii}$	squared effect
$\beta_{ij}$	interaction effect
$\mu$	viscosity, $\text{kg.m}^{-1}.\text{s}^{-1}$
$\mu_{\text{max}}$	maximum specific growth rate, $\text{h}^{-1}$
$\rho$	density, $\text{kg.m}^{-3}$
$\sigma$	standard deviation
$\tau_p$	probe response time, s
$a_i$	initial specific surface area, $\text{m}^{-1}$
$a_f$	final specific surface area, $\text{m}^{-1}$
Bo	Bond number, $\rho d_v^2 g/W$ , dimensionless
CER	carbon-dioxide evolution rate, $\text{mmol.l}^{-1}.\text{h}^{-1}$
$C_p$	unsaturated oxygen fraction at time t, dimensionless
$d_t$	shaking amplitude, m
$d_v$	microwell vessel diameter, m
D	diffusion coefficient, $\text{m}^2.\text{s}^{-1}$
$D_i$	impeller diameter, m
DoE	Design of Experiments
DOT	dissolved oxygen tension, %
FFL	Firefly luciferase
Fr	Froude number $d_t(2\pi n)^2/(2g)$ , dimensionless
g	acceleration due to gravity, $\text{m.s}^{-2}$
HTES	High-throughput expression screening
HTPE	High-throughput protein expression
HTS	High-throughput screening
$k_{La}$	Mass transfer coefficient
MSBR	Miniature stirred bioreactor
MWP	Microwell plate
n	shaking frequency, $\text{s}^{-1}$

---

N	impeller speed, $s^{-1}$
OFAT	One-factor-at-a-time
OUR	oxygen uptake rate, $mmol.l^{-1}.h^{-1}$
$P_g$	power input to gassed liquid, W
$P_o$	power number, dimensionless
$P_{ug}$	power input to ungassed liquid, W
Q	air flow rate, $m^3.s^{-1}$
$Q^2$	fraction of the response variation predicted by the model (predictive $R^2$ )
$R^2$	fraction of the response variation explained by the model
Re	Reynolds number, $\rho n_d v 2/\mu$ , dimensionless
RLU	Relative light units
Sc	Schmidt number, $\mu/(\rho D)$ , dimensionless
SF	Shake-flask
STR	Stirred-tank reactor
t	time, s
$t_m$	mass transfer time, s
V	liquid volume, l
$v_s$	superficial gas velocity, $m.s^{-1}$
W	wetting tension, $N.m^{-1}$
$W_b$	impeller blade width, m
$x_i$	dimensionless coded value of the variable $X_i$
X	biomass concentration (dry cell weight), $g.l^{-1}$
$X_i$	DoE variable
Y	predicted response
$z_i$	actual variable value of $X_i$

## 1 Introduction

### 1.1 Drug discovery

#### *1.1.1 The current direction of drug discovery*

It is widely accepted that virtually all human diseases except trauma has some basis in our genes (Collins, 1999). In April 2003, the Human Genome Project (HGP) completed a rough draft of the human genome within which 20,000 to 25,000 genes were identified (Human Genome Sequencing Consortium, 2004). Extrapolations from the now complete genome sequence predict that between 2,000 and 3,000 genes favour interactions with drug-like chemical compounds (Russ and Lampel, 2005). Recent estimates state that there are only 268 therapeutic targets which are targeted by at least one marketed drug (Zheng *et al.*, 2006). The exposure of this shortfall has provided the main impetus for a new wave of drug discovery in the post-genomic era.

Since virtually all human diseases have a hereditary component, it is perhaps not so surprising that the vast majority of today's drug targets are gene-products, proteins (Dahl and Sylte, 2006; Tyers and Mann, 2003). For this reason, proteomics, the study of protein function, has recently become of burgeoning interest to the pharmaceutical industry.

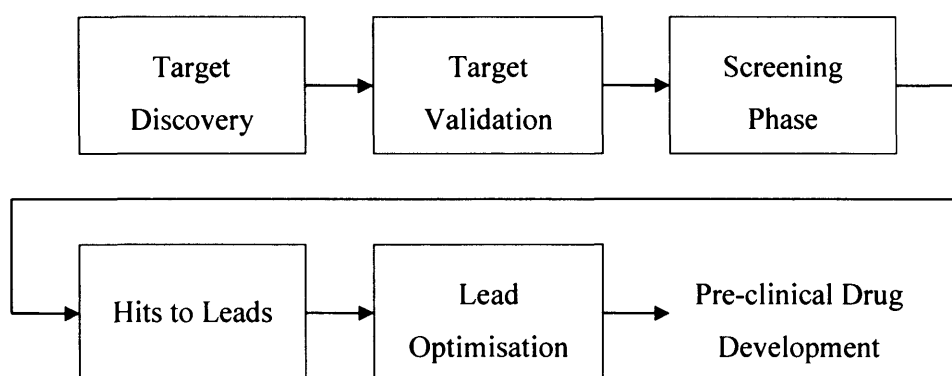
The sequencing of the human genome and that of many pathogens has provided a sequence based framework for investigating proteomes (Hanash, 2003), opening the gateway for proteomics to emerge. As a result, there is now intense interest in applying proteomics to increase our knowledge of disease processes, develop new biomarkers for early detection and diagnosis of disease, and accelerate drug development (Hanash, 2003).

The HGP required high-throughput and cost-effective technology to process the immense amount of information contained within the human genome. The field of proteomics now presents a much greater challenge due to the far larger size,

complexity and dynamic nature of the human proteome. The need for high-throughput technology is, therefore, greatly amplified and lies at the core of today's drug discovery effort.

### 1.1.2 Drug discovery process

The field of drug discovery is highly diverse and complex and the exact process differs widely within the pharmaceutical industry. Figure 1.1 illustrates a generic drug discovery process. Although a sequential process is shown here, in reality a large number of iterative cycles may be involved. Thus, Figure 1.1 should be seen as the overall sequence of steps required, but not as a literal depiction of any given drug discovery program.



**Figure 1.1** Generic process of drug discovery

The process description that follows is a brief and generalised overview. The first stage involves the identification of a suitable disease target, towards which the Human Genome Project and, subsequently, functional genomics has contributed tremendously (Debouck and Metcalf, 2000; Mundy, 2001). The vast majority of therapeutic targets are proteins (Bleicher *et al.*, 2003; Dahl and Sylte, 2006; Lindsay, 2003) which originate either in humans or in infectious agents such as bacteria, fungi and viruses. The role and function of a therapeutic target is investigated and its chemical tractability or “druggability” evaluated.

Target Validation aims to demonstrate that a molecular target is critically involved in a disease process and that modulation of the target is likely to have a therapeutic effect. This stage involves *in vitro* and *in vivo* biochemical analysis and provides fundamental information which will direct and aid the screening phase. For example, target validation may generate information relating to the relevant form of the target protein needed for screening e.g active or inactive, membrane bound or cytoplasmic.

The Screening Phase (or Hit Generation) of the drug discovery process is broad in scope where various techniques can be applied to assess how effectively compounds interact with the target protein. Techniques such as bioassays, structure based methods (X-ray crystallography and Nuclear Magnetic Resonance) and biophysical techniques such as thermal denaturation and isothermal calorimetry are all used in the screening phase of modern drug discovery. The cost, time and throughput of these techniques vary as does the level of detail obtained regarding the binding of compounds to the target protein. These screening techniques require differing levels of target protein and so a flexible production capacity is required to accommodate these needs.

Traditionally, a functional biochemical assay would be used to screen compound libraries where a value for the inhibition of the target protein could be calculated. This type of *in vitro* assay is the basis for High-Throughput Screening (HTS) where the assay is formatted onto microwell plates (MWP) and compound libraries of >200,000 compounds can be screened in a timely fashion. Whilst the throughput of an HTS screen is very rapid, the sensitivity of the assay can be quite low with compounds requiring a low micro-molar potency to be detected. Therefore, smaller but more novel compounds could be missed in this type of screen. Although a value for the inhibition of the compound against the target protein is calculated in an HTS campaign, no information is generated regarding the actual binding mode of the compound to the protein. This information is highly valuable to the medicinal chemist when trying to optimise the potency and selectivity of the compound.

More recently, protein crystallography has emerged as a powerful screening tool for hit generation. Traditionally, protein crystallography has been used in the latter stages of the drug discovery process. However, advances in protein production, crystallisation and data collection have enabled this technique to be used in the screening phase. The generation of detailed structural information on how the compound interacts with the target protein is immensely powerful and is the key to designing more selective and potent drugs. The use of High Throughput Crystallography in structure based drug design has been enhanced by advances in *in silico* methods. Prior structural information can be used to enrich compound libraries with molecules that have been shown computationally to interact well with the target protein. Through iterative cycles of structure determination of protein/ligand complexes coupled to medicinal chemistry, detailed Structure Activity Relationships (SARs) can be generated.

The Hits to Leads phase builds up significant SAR data around the initial hits obtained from the screening phase. Chemical moieties with different functionalities are added onto the initial hit in an attempt to improve the binding of the compound to the target protein. These new compounds are then re-screened to establish which of the modifications have improved the potency of the compounds. The Lead Optimisation phase focuses on a few lead series that look the most attractive in terms of potency, selectivity and physicochemical properties. Medicinal chemists take these lead compounds and alter their structures using all the knowledge gained thus far, drawing heavily from SAR data, to improve the drug qualities of lead compounds such as stability, bioavailability and absorption, distribution, metabolism, and excretion (ADME) profiles. These drug candidates are then assessed in more extensive *in vivo* studies and animal models to generate pharmacokinetics (PK), pharmacodynamics (PD), ADME and toxicity data. If the profile of a lead compound is deemed to be promising then it progresses through to drug development.

### ***1.1.3 Drug discovery challenges***

A study of therapeutics in clinical trials between 1991 and 2000 revealed an 89% average failure rate across ten large pharmaceutical companies (Kola and Landis, 2004). The scale of this problem is compounded once the development time and cost for new drugs are considered. The average development time of all drugs that gained approval in 2002 was 12 years 10 months; the average development cost of a new drug in 2001 was approximately US \$802 million (DiMasi *et al.*, 2003) and some authors believe the true cost may be more than double this amount (Adams and Brantner, 2006). In addition to these costs is the human cost of the failure to discover new treatments, and the cost in animal lives throughout the research process. Understanding the root causes of clinical drug failure and minimising attrition is, therefore, one of the most important aims of today's pharmaceutical discovery programmes (Nicholson *et al.*, 2002).

The major causes of attrition in the clinic in 2000 were lack of efficacy, accounting for approximately 30 % of failures, and safety, accounting for a further 30 %, (Kola and Landis, 2004). The vast majority of attrition occurs in clinical development (Kola and Landis, 2004) at which stage the cost of failure is very high (Bleicher *et al.*, 2003; Nicholson *et al.*, 2002). Early measures are, therefore, required to increase the understanding of drug properties within the drug discovery process.

The key to overcoming this obstacle is to obtain a more rigorous understanding of how drug candidates bind and interact with their protein targets (Stewart *et al.*, 2002). To this end, it is critical that a more detailed knowledge of protein and drug structures is generated at the screening phase of drug discovery. This would enable medicinal chemists to optimise the drug properties of the lead molecules. For example, detailed knowledge of a target molecule's binding site would facilitate in the design of drug compounds with higher potency and selectivity (Blundell *et al.*, 2002). This in turn would decrease the occurrence of side reactions and might also improve efficacy and toxicity profiles, the leading factors in drug attrition.

X-ray crystallography is perhaps the most commonly used tool for three-dimensional molecular structure determination (Blundell and Patel, 2004; Congreve *et al.*, 2005; Kuhn *et al.*, 2002; Stewart *et al.*, 2002). The generation of sufficient quantities of soluble recombinant protein for the production of diffraction-quality crystals can present a bottleneck to the entire drug discovery process (Stewart *et al.*, 2002). Furthermore, there is a growing requirement to increase the throughput of X-ray crystallography (Blundell and Patel, 2004; Congreve *et al.*, 2005; Kuhn *et al.*, 2002; Stewart *et al.*, 2002). For instance, fragment-based lead-discovery is becoming increasingly popular for identifying better lead molecules so as to reduce attrition (Rees *et al.*, 2004). Molecular fragments by their very nature are of low chemical complexity and display limited functionality which results in a lower affinity in conventional biological assays (Gill *et al.*, 2005). X-ray crystallography is able to detect low-affinity binding and so has emerged as a complimentary/replacement approach to traditional high-throughput screening (Rees *et al.*, 2004). Furthermore, screening throughput itself may be driven higher as a result of the increased identification of potential therapeutic targets from functional genomics research, as outlined in Section 1.1.2. Overall, therefore, there is intense interest in enhancing throughput by automating all steps of protein crystallography, the first step of which is the rapid and cost-effective production of soluble protein samples (Congreve *et al.*, 2005).

## 1.2 High Throughput Protein Expression

### 1.2.1 Choice of expression system

The choice of expression system can greatly affect the yield and activity of a protein (Blundell and Patel, 2004; Stewart *et al.*, 2002). There is no universal expression system for heterologous proteins (Rai and Padh, 2001) and so the selection of an appropriate system involves an assessment of the benefits and disadvantages of each system in relation to the application requirements. This project aims to establish a framework that will underpin the generation of large quantities of soluble protein in a rapid and cost-effective manner. A model



---

expression system is required, therefore, the properties of which should facilitate the development of the framework. A model expression host should:

- i. grow rapidly;
- ii. require simple and inexpensive substrates;
- iii. express high yields of heterologous protein;
- iv. lend well to high-throughput production, MWP growth, automation and scale-up.

An immediate choice would be a bacterial expression host as it meets all these requirements. *Escherichia coli* in particular is an ideal host as it grows rapidly with a short doubling time of ~20 minutes (Singleton and Sainsbury, 2002), it is inexpensive to culture (Blundell and Patel, 2004; Hunt, 2005) and efficient expression of recombinant product to more than 50 % of total cell mass has been reported (Andersen and Krummen, 2002; Baneyx, 1999; Jana and Deb, 2005; Swartz, 2001). Moreover, a wealth of information exists regarding its genetics and physiology which greatly facilitates gene cloning and cultivation (Rai and Padh, 2001).

#### 1.2.1.1 Bacterial expression systems

Since *E. coli* is a simple organism, it is also highly amenable for growth within the MWP format using standard shaking incubators (Hunt, 2005). This utility allows many parameters to be analysed simultaneously in small volumes which, in turn, allows the rapid identification of optimal expression conditions for the generation of large quantities of recombinant protein.

These properties make *E. coli* the most widely used host in industry for recombinant protein expression, provided post-translational modifications of the product are not required (Andersen and Krummen, 2002; Blundell and Patel, 2004; Choi *et al.*, 2006; Graumann and Premstaller, 2006; Hunt, 2005; Rai and Padh, 2001; Stewart *et al.*, 2002; Walsh, 2002).

If diffraction-quality crystals are not readily obtainable using *E. coli*, then several alternative expression systems are available, each with its own unique combination of benefits and disadvantages in relation to cost, ease of use, and post-translational modification profiles.

#### 1.2.1.2 Yeast expression systems

Yeast systems present the favoured alternative host for expression of recombinant proteins for research, industrial and medical use (Rai and Padh, 2001) of which the most common strains are *Saccharomyces cerevisiae* and *Pichia pastoris* (Andersen and Krummen, 2002; Graumann and Premstaller, 2006; Stewart *et al.*, 2002). Yeasts have played a central role in many traditional biotechnological processes such as brewing and baking and so they, like *E. coli*, are very well characterised within modern biotechnology. Yeasts are simple eukaryotes and thus are able to perform a range of post-translational modifications required by many human proteins. They usually grow fast, but slower than *E. coli*, and are also able to secrete proteins to the culture medium (Graumann and Premstaller, 2006) which facilitates isolation and purification of the product. Although a variety of therapeutic proteins have been manufactured in *Saccharomyces cerevisiae*, expression levels of heterologous protein are often low, typically representing less than 5 per cent of total cellular protein (Walsh, 2002).

#### 1.2.1.3 Insect cell/baculovirus expression systems

Insect cell lines are also emerging as popular systems for overproducing recombinant proteins (Rai and Padh, 2001). Heterologous protein expression is usually mediated by the baculovirus expression system which is highly specific to the target cell line. The gene encoding the protein of interest is introduced into a non-essential region of the viral genome. The virus is then propagated within the insect cell culture and the recombinant protein is expressed in the process. The main advantage afforded by insect cells is that they are able to produce many

of the post-translational protein modifications that occur in human cells (Davies, 1994). The majority of proteins expressed within insect cells are, therefore, soluble (Rai and Padh, 2001). However, insect cells grow much slower than bacteria or yeasts with a doubling time of approximately 24 hours (O'Reilly *et al.*, 1992) and the cost and complexity of the growth medium is usually high. Furthermore, expression levels of heterologous protein have thus far been highly variable (Walsh, 2002).

#### 1.2.1.4 Mammalian expression systems

Proteins which require mammalian-specific post-translational modifications should ideally be expressed in mammalian cells (Rai and Padh, 2001). Chinese hamster ovary (CHO), human embryonic kidney (HEK 293) and non-secreting murine myeloma (NS0) cell expression systems have now established themselves as the predominant systems for mammalian expression (Andersen and Krummen, 2002). Mammalian cell culture requires complex and expensive culture media (Rai and Padh, 2001; Walsh, 2002) and low product yields are often obtained (Rai and Padh, 2001). Furthermore, cell growth is slower and more complicated than the expression systems reviewed thus far. As such, they are not the first choice for high-throughput protein production.

#### 1.2.1.5 Cell-free expression systems

Perhaps the most interesting emerging technology is that of cell-free expression systems. These include both more traditional systems using cell extracts and systems reconstituted from purified components (Kim and Swartz, 2001; Shimizu *et al.*, 2001). The main advantages of these systems over conventional *in vivo* systems are: (i) the target protein is the only protein synthesised which greatly simplifies the purification requirements, (ii) the reaction is rapid and can be carried out in small volumes and (iii) the absence of living cells makes the use of a wide range of different reaction conditions possible (Ozawa *et al.*, 2005). A further advantage of this system is that chemical modifications for structural

studies such as N<sup>15</sup> and C<sup>13</sup> labelling and Selino-Methionine labelling can be made in a more straight forward manner

Recombinant proteins are often expressed in more than one system during drug discovery (Janssen, 2004). A bacterial expression system may be used for initial studies to determine solubility or activity or to produce large amounts of protein for structural studies. In order to obtain higher activity or post-translational modifications, the same gene might then be expressed in eukaryotic cells (protein expression and purification techniques). Overall, *Escherichia coli*, remains the first choice of expression system for producing samples for protein crystallography (Blundell and Patel, 2004; Edwards *et al.*, 2005). It will, therefore, be used as the model expression host in this project.

### **1.2.2 Mechanism of protein expression**

In general, both constitutive and inducible systems have been described for *E. coli* where the latter dominate (Graumann and Premstaller, 2006). The expression of recombinant proteins induces a metabolic burden, which is defined as the amount of resources which are withdrawn from the host metabolism for maintenance and expression of foreign DNA (Bentley and Kompala, 1990). Maximum productivity in recombinant *E. coli* is often achieved when the growth and production phases are separated by delaying the induction time until the cell density reaches a suitable value (Choi *et al.*, 2006). This approach is especially important when the expression target is toxic to the cell (Balbas, 2001; Jana and Deb, 2005; Sorensen and Mortensen, 2005). The ideal expression vector tightly regulates gene expression to achieve rapid cell growth to sufficient densities before the induction phase.

*E. coli* strain BL21 (DE3) was chosen as the model expression system for this project. BL21 is a robust and commonly used strain for the expression of recombinant proteins (Sorensen and Mortensen, 2005). It is deficient of both the *lon* protease and the *ompT* outer membrane protease (Grodberg and Dunn, 1988). The lack of two key proteases reduces degradation of heterologous proteins. The

DE3 designation indicates that the strain contains the  $\lambda$ DE3 lysogen that carries the gene for T7 RNA polymerase under control of the *lacUV5* promoter. This promoter contains a mutation in the consensus region of the *lac* promoter which increases its strength (Reznikoff, 1980). Expression of T7 RNA polymerase is inducible by isopropyl- $\beta$ -D-thio-galactopyranoside (IPTG), a lactose analogue, which binds to the *lacI* repressor and reduces its affinity for the *lac* operator; the overall expression mechanism is summarised effectively by Sorensen and Mortensen (2005).

The pET (*plasmid for expression by T7 RNA polymerase*) vector system was developed by Studier *et al.* (1990) for cloning and expressing recombinant DNA under control of a T7 promoter. Due to the high selectivity of T7 RNA polymerase for the T7 promoter, transcription of target DNA by *E. coli* RNA polymerase in the absence of T7 RNA polymerase is low enough to enable the expression of very toxic genes (Studier *et al.*, 1990), which increases the applicability of the model expression system. Furthermore, pET vectors confer antibiotic resistance upon their host and so serve as a selectable marker. The pET expression system is by far the most widely used currently in recombinant protein expression (Sorensen and Mortensen, 2005) and high levels of success have been reported in obtaining diffraction-quality protein crystals originating from *E. coli* BL21 (DE3) cells transformed with the pET vector (Folli *et al.*, 2001; Goulding and Perry, 2003; Park *et al.*, 2005; You *et al.*, 2003).

### **1.2.3 Factors influencing soluble protein expression**

Much research has been carried out into optimising soluble heterologous protein expression through host strain and vector development (for detailed reviews, see Baneyx, 1999; Georgiou and Valax, 1996; Hannig and Makrides, 1998; Jana and Deb, 2005; Sorensen and Mortensen, 2005; Weickert *et al.*, 1996). A simpler and often equally effective approach is to manipulate the culture conditions (Georgiou and Valax, 1996). Relatively little research has been carried out in this area, however.

Batch cultivation is the simplest way to produce a recombinant protein (Jonasson *et al.*, 2002), and it is particularly appropriate for the MWP format. Several factors are reported to affect cell growth and soluble protein expression within this mode of fermentation. Fermentation variables such as media composition, pH, temperature and dissolved oxygen can affect transcription, translation, and proteolytic activities within a cell (Choi *et al.*, 2006; Bird *et al.*, 2004). Protein induction conditions such as pre-induction cultivation period, inducer concentration, post-induction cultivation period and the levels of fermentation variables during the induction phase also affect expression levels (Choi *et al.*, 2000; Jeong and Lee, 1999; Yim *et al.*, 2001). In order to investigate these effects on expression levels of heterologous protein, it is vital that all other minor variables are controlled between batches.

The composition of the growth medium has significant effects on both the rate of cell growth and level of product accumulation (Broedel *et al.*, 2001; Jana and Deb, 2005). The translation of different mRNAs, for example, is differentially affected by changes in the culture medium (Corisdeo and Wang, 2004). Fast growth due to high concentrations of rapidly metabolised sugars is often associated with low productivity of recombinant protein (Stanbury and Whitaker, 1993). Furthermore, Broedel *et al.* (2001) observed that the optimal medium composition for soluble protein expression differed between proteins.

It is well established (Riesenberg, 1991) that *E. coli* growth is inhibited when the following nutrients are present above certain concentrations (shown in brackets): glucose (50 g.l<sup>-1</sup>), phosphorous (10 g.l<sup>-1</sup>), magnesium (8.7 g.l<sup>-1</sup>), ammonia (3 g.l<sup>-1</sup>), iron (1.15 g.l<sup>-1</sup>) and zinc (0.038 g.l<sup>-1</sup>). It is, therefore, important to develop media in which these components are present at non-inhibitive concentrations.

While defined media are generally used to obtain high-cell densities (Sang, 1996), complex media are often employed in industrial settings because of lower costs and more robust cell growth (Diaz-Ricci *et al.*, 1990). Although fermentations in complex media are more prone to variation due to varying nutrient composition and quality (Jonasson *et al.*, 2002; Sang, 1996), the vitamins

and macromolecular precursors provided by the media are often necessary to obtain high levels of recombinant protein (Diaz-Ricci *et al.*, 1990; Jonasson *et al.*, 2002).

A major challenge in the production of heterologous protein at high cell density is the formation of acetate, a lipophilic agent that inhibits cell growth (Jana and Deb, 2005; Luli and Strohl, 1990). The wide range of detrimental effects on the growth of *E. coli* caused by acetate accumulation is reviewed in detail by Shiloach and Fass (2005). Acetate is produced when the culture is grown in the presence of glucose or under oxygen limiting conditions (Kleman and Strohl, 1994) and it is important to maintain the acetate concentration below a certain, inhibitory level (Jonasson *et al.*, 2002). One way of achieving this is by using glycerol in preference to glucose as the carbon substrate (Holms, 1996; Korz *et al.*, 1995); at least one study has demonstrated a two-fold increase in recombinant protein expression from *E. coli* cultures grown on glycerol compared to those grown on glucose (Lee *et al.*, 1997).

The pH of the culture medium can also affect product yield. One benefit of avoiding acetate formation is that it minimises the pH disturbance which is of particular concern with respect to MWP cultures where pH control is not routinely available. The optimal pH range for *E. coli* growth is approximately 6-8 (Ingraham and Marr, 1996) and several studies have also observed optimal recombinant protein production in *E. coli* within this pH range (Dien *et al.*, 2001; Ryan and Parulekar, 1990; Wang *et al.*, 2005). Although relatively little information exists on the behaviour of pH-uncontrolled systems, Calik *et al.* (2006) reported higher cell concentrations and product yields from batch fermentations with no pH control compared to those with pH control.

Temperature is another important variable. Cultivation at reduced temperature slows down cell growth and protein synthesis (Peng *et al.*, 2004). Although this generally leads to a lower final biomass concentration, the amount of soluble recombinant protein is usually increased. These factors contribute to a higher solubility of foreign proteins (Schein, 1989) and the technique has proved

---

effective in producing a number of proteins which are typically poorly expressed (Vasina and Baneyx, 1997). The reason why cultivation at a lower temperature of ~30 °C favours the native state is related to a number of factors, including a decrease in the propensity of hydrophobic interactions which cause proteins to aggregate (Georgiou and Valax, 1996; Kiefhaber *et al.*, 1991), a slower rate of protein synthesis (Georgiou and Valax, 1996), the partial elimination of heat shock proteases (Chesshyre and Hipkiss, 1989) and an increase in the expression of *E. coli* chaperones (Ferrer *et al.*, 2003; Mogk *et al.*, 2002).

Oxygen is an essential nutrient for the aerobic growth of *E. coli* (Andersen and von Meyenburg, 1980; McDaniel *et al.*, 1965), but at the same time it is the most difficult to supply because of its low solubility (Li *et al.*, 1992). For this reason the oxygen demand may become the limiting factor of *E. coli* growth, if it exceeds the oxygen transfer capacity of the fermentation vessel (Maier and Buchs, 2001). This is more likely to occur if rapidly metabolised sugars such as glucose, which lead to high oxygen demand, are available in high concentrations (Stanbury and Whitaker, 1993). The issue of oxygen demand is particularly significant to shaken MWP cultures, where oxygen supply, provided solely by surface aeration, may be poor.

There are no definitive rules regarding the effect of oxygen supply on expression levels of recombinant protein. This is because there are relatively few published studies in this area and furthermore, those that do exist observe differing trends. In a study by De Leon *et al.* (2003), the maximum production of active penicillin acylase by *E. coli* was obtained at a dissolved oxygen tension (DOT) of 1 %. In contrast, Bhattacharya and Dubey (1997) reported a drastic decrease in the level of target protein (*MspI* methylase) in recombinant *E. coli* under oxygen-deficient conditions. This variable trend was summarised by Li *et al.* (1992) who examined the effects of dissolved oxygen levels on the level of biomass, plasmid content and the levels of chloramphenicol acetyltransferase and beta-galactosidase within four recombinant strains of *E. coli*. The optimal dissolved oxygen concentration for the specific activity of recombinant proteins was found to be dependent on both the host strain and the particular recombinant product. In



---

light of these studies, oxygen should prove to be an interesting variable for investigation.

The optimal conditions for inducing recombinant protein expression are highly dependent upon the characteristics of the target protein and thus the extent of the imposed metabolic burden (Donovan *et al.*, 1996). In many cases the inducer concentration is optimised to balance the decreasing yields of recombinant cells following induction with increasing cellular levels of target protein (Bentley *et al.*, 1991). For soluble cytoplasmic proteins, this optimal level is usually approximately 1 mM IPTG (Bentley *et al.*, 1991; Donovan *et al.*, 1996). When product expression is low and/or does not significantly affect cell growth, overall heterologous protein yield is maximised by inducing expression throughout the entire growth phase (Donovan *et al.*, 1996). However, if expression of the recombinant protein does impose a significant metabolic burden, induction at or beyond middle stage of exponential growth may prove optimal (Bentley *et al.*, 1991; Donovan *et al.*, 1996; Lee, 1996; Peng *et al.*, 2004).

A common limitation of the studies mentioned so far is that they fail to adequately study the combined effect of variables on heterologous protein expression. Thus no account is made for interactions between the large number of variables. An objective of this project is to address this shortfall.

#### **1.2.4 Microscale processing**

A wide range of variables affect heterologous protein expression levels as described in Section 1.2.3 and numerous experiments are, therefore, required to adequately characterise the total complement of interactions and effects. Microscale processing techniques are rapidly emerging as a means of satisfying the demands placed by High Throughput Protein Expression (HTPE) studies. Through the miniaturisation and automation of fermentation, these techniques allow for vast savings in time, money (especially where expensive media or substrates are involved), space and manpower (Kumar *et al.*, 2004; Lye *et al.*,

2003; Maier and Buchs, 2001). This also enables a wider range of process variables to be examined (Lye *et al.*, 2003).

The pressing need for reproducible and scalable high-throughput cultivation technology has identified several areas of research interest (Weuster-Botz *et al.*, 2005). These interests draw a parallel with the direction of research that followed the emergence of conventional bioreactors. For example, a quantitative understanding of the engineering parameters associated with the operation of small-scale fermentation vessels such as the hydrodynamics and mass transfer characteristics, is of primary concern (Fernandes and Cabral, 2006). The ability to measure online process parameters such as dissolved oxygen, temperature and pH is also needed. Appropriate sensors have been developed towards these purposes for a range of small-scale bioreactors including shake-flasks, miniature stirred bioreactors (MSBRs), and MWPs.

The shake-flask represents the classical parallel reactor used in biotechnology (Weuster-Botz, 2005). Scientists have used cell cultivation in shake-flasks as a means of process development for the past fifty years (Betts and Baganz, 2006), with volumes ranging from ca. 10 ml to 500 ml (Buchs, 2001). These reactors are inexpensive, easy to use and largely impervious to mechanical complications (Betts and Baganz, 2006; Kumar *et al.*, 2004). Recently, instrumented shake-flasks have become available, which allow for the monitoring of pH and DOT levels online (Anderlei and Buchs, 2001; Wittmann *et al.*, 2003). Other parameters such as the oxygen transfer rate (OTR) and carbon dioxide evolution rate (CER) can now also be measured online. A major limitation, however, is that shake-flask fermentations are not typically automated. Sampling, for example, requires the removal of the shake-flasks from the shaking device, thus interrupting oxygen supply (Kumar *et al.*, 2004).

Miniature stirred bioreactors based on conventional STRs have been developed as an alternative to traditional shaken systems for early-stage process development (Betts and Baganz, 2006; Hsu and Wu, 2002; Harms *et al.*, 2006; Kostov *et al.*, 2001; Lamping *et al.*, 2003; Zanzotto *et al.*, 2004; Zhang *et al.*,

2006). Despite their small size, temperature, pH and dissolved oxygen can be monitored and controlled at desired levels within MSBRs. They are also relatively easy to scale-up due to similarities in geometry and in methods of aeration and agitation with conventional bioreactors. The use of these reactors is, however, severely limited by their high costs relative to other small-scale devices and also the difficulties of their integration with high-throughput screening technologies (Kumar *et al.*, 2004; Weuster-Botz, 2005).

MWPs are perhaps the most popular choice for the screening stages of high-throughput bioprocess development (Weuster-Botz *et al.*, 2005). They have been used to grow a wide range of cell lines including bacteria (see Kumar *et al.*, 2004 for a review), yeast (Hammonds *et al.*, 1998; Janssen, 2004), insect cells (Bahia *et al.*, 2005; Chambers *et al.*, 2004) and also mammalian cells (Davies *et al.*, 2005; Deshpande *et al.*, 2004; Micheletti *et al.*, 2006). They offer all the advantages of shake-flasks with the added benefit of being highly amenable to process automation. Their ability to perform many identical reactions in parallel with very small volumes (Betts and Baganz, 2006) endows them with the greatest high-throughput capability out of all the miniature bioreactor formats mentioned thus far.

MWPs typically contain 6, 12, 24, 48, 96 or 384 wells, with up to 1536 and 3456 wells now available for ultra high-throughput screening (UHTS) (Mere *et al.*, 1999). MWPs are available in a range of geometries, well volumes and materials of construction and are reviewed in detail by Lye *et al.* (2003) and Betts and Baganz (2006).

The processing of MWPs can be automated using robotics with modern pipetting and dispensing systems, centrifuges, etc (Kumar *et al.*, 2004). A wide range of commercially available automated platforms is reviewed by Lye *et al.* (2003). Instruments and methods have been developed for pH measurements (Elmahdi *et al.*, 2003; John *et al.*, 2003a; Weiss *et al.*, 2002) and dissolved oxygen measurements (see Table 1.1). Now, MWPs with integrated sensors are commercially available and are reported by Betts and Baganz (2006). A major

bottleneck in the use of MWP for cultivation is the risk of cross-contamination (Kumar *et al.*, 2004). Evaporation of medium is also a critical factor particularly for slow growing organisms. These limitations can be reduced with the use of oxygen-permeable sealing membranes, a variety of which are reviewed by Zimmermann *et al.* (2003).

Table 1.1 summarises the existing research into microscale mixing within the aforementioned reactor formats. The mass transfer coefficients ( $k_La$ ) of shake-flasks and MSBRs are well characterised over a range of geometries and working volumes and the two formats share similar  $k_La$  ranges. In contrast, the majority of research into MWPs has focussed on the 96 well format, which is probably due to the frequency of its use in industry (Kensy *et al.*, 2005). Subsequently, the research is also limited with respect to the range of working volumes investigated. For these plates, the majority of  $k_La$  values are in the lower range, below  $180 \text{ h}^{-1}$ . The notable exception here is the extreme  $k_La$  value reported Kensy *et al.* (2005) of  $1600 \text{ h}^{-1}$  for a 48 well plate, which is comparable to that of a conventional STR. Clearly, there is a need for further investigation in this area, but this research has already demonstrated the potential for MWPs to exhibit similar mixing characteristics to that of conventional STRs.

### **1.2.5 Reporter genes**

The term reporter gene is used to define a gene with a readily measurable phenotype that can be easily distinguished over a background of endogenous proteins (Alam and Cook, 1990). This characteristic is highly desirable when choosing a suitable product model for the investigation of high throughput protein expression. Moreover, reporter gene assays are largely quantitative and rapid, requiring no prior protein purification. This facilitates the accurate analysis and comparison of recombinant protein expression levels under a range of culture conditions.

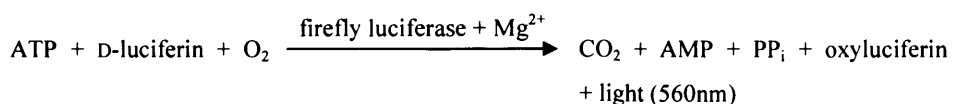
**Table 1.1** Summary of microscale research into mixing and oxygen transfer carried out in a range of small-scale vessels including the shake flask (SF), miniature stirred bioreactor (MSBR) and microwell plate (MWP). <sup>1</sup>The terms *static gassing-out* and *dynamic gassing-out* are used in accordance with the definitions provided by Stanbury and Whitaker (1993).

Reactor Format	Working Volume (ml)	$k_L a$ ( $\text{h}^{-1}$ )	Method <sup>1</sup>	Reference
SF (unbaffled)	2 – 160	< 565	Sulphite oxidation	Maier <i>et al.</i> (2004)
SF (unbaffled)	100	10 – 54	Static gassing-out	Van Suijdam <i>et al.</i> (1978)
SF (baffled)	200	144 - 482		
SF (baffled/ unbaffled)	50 – 200	35 - 400	Sulphite oxidation and oxygen balance ( <i>E. coli</i> / <i>S. tendae</i> )	Henzler and Schedel (1991)
MSBR	6	100-400	Static gassing-out	Lamping <i>et al.</i> (2003)
MSBR	1	< 500	Sulphite oxidation	Harms <i>et al.</i> (2006)
MSBR	2	10 – 44	n/a	Kostov <i>et al.</i> (2001)
MSBR	0.15	20 - 75	Static gassing-out	Zhang <i>et al.</i> (2006)
MSBR	18	< 480	n/a	Betts and Baganz. (2006)
MWP (96 well)	0.2	< 150	Bio-oxidation of catechol	Ortiz-Ochoa <i>et al.</i> (2005)
MWP (96 well)	0.2	< 130	Dynamic gassing-out ( <i>C. glutamicum</i> )	John <i>et al.</i> (2003b)
MWP (24 well)	0.07	< 180	Static gassing-out and	Doig <i>et al.</i> (2005)
MWP (96 well)	0.2	< 180	linear growth of strict	
MWP (384 well)	1.2	< 100	aerobe ( <i>B. subtilis</i> )	
MWP (96 well)	0.5	< 188	Oxygen balance ( <i>P. putida</i> )	Duetz <i>et al.</i> (2000)
MWP (48 well)	0.3	< 1600	Sulphite oxidation	Kensy <i>et al.</i> (2005)

The variety of available reporter genes is very broad. Some commonly used ones include chloramphenicol acetyltransferase (CAT),  $\beta$ -galactosidase, bacterial and firefly luciferases, alkaline phosphatase and green fluorescent protein (GFP) (Naylor, 1999; Welsh and Kay, 1997; Wood, 1995).

The *luc* gene from the firefly (*Photinus pyralis*) is one of the most widely used reporter genes (Roda *et al.*, 2004; Urbain, 2001; Welsh and Kay, 1997) and it is also well-suited for application within this project. The assay has a broad linear range of up to 7-8 orders of magnitude (Joyeux *et al.*, 1997) and, unlike  $\beta$ -galactosidase, no endogenous activity exists within *E. coli* for the luciferases (Wood, 1995), a property which contributes to the highly sensitive nature of the assay (Alam and Cook, 1990). The *luc* gene encodes a single polypeptide which imparts a low metabolic burden upon the expression host in contrast to the bacterial luciferase operon which codes for five polypeptides (Hakkila *et al.*, 2002). Furthermore, firefly luciferase shows little toxicity (Welsh and Kay, 1997) and requires no posttranslational modifications (Bronstein *et al.*, 1994).

Luciferase activity is proportional to the total light emission of the reaction scheme shown below (Bronstein *et al.*, 1994):



The requirement of D-luciferin, an exogenous substrate, makes the use of an endpoint assay for soluble protein yield possible. In contrast, if bacterial luciferase, was used as a reporter, it would be necessary to monitor light emission continuously because the corresponding reaction would proceed automatically from endogenous substrates upon induction of the reporter genes (Hakkila *et al.*, 2002).

The luciferase assay is also well suited to the high-throughput screening format unlike the CAT assay which is laborious, expensive and relatively insensitive

(Suto and Ignar, 1997). GFP is also relatively insensitive, because of a considerable background signal due to cellular autofluorescence (Hakkila *et al.*, 2002; Naylor, 1999). Moreover, the accumulation of GFP, narrow linear range and low turnover make it also unsuitable for any high throughput investigation (Naylor, 1999).

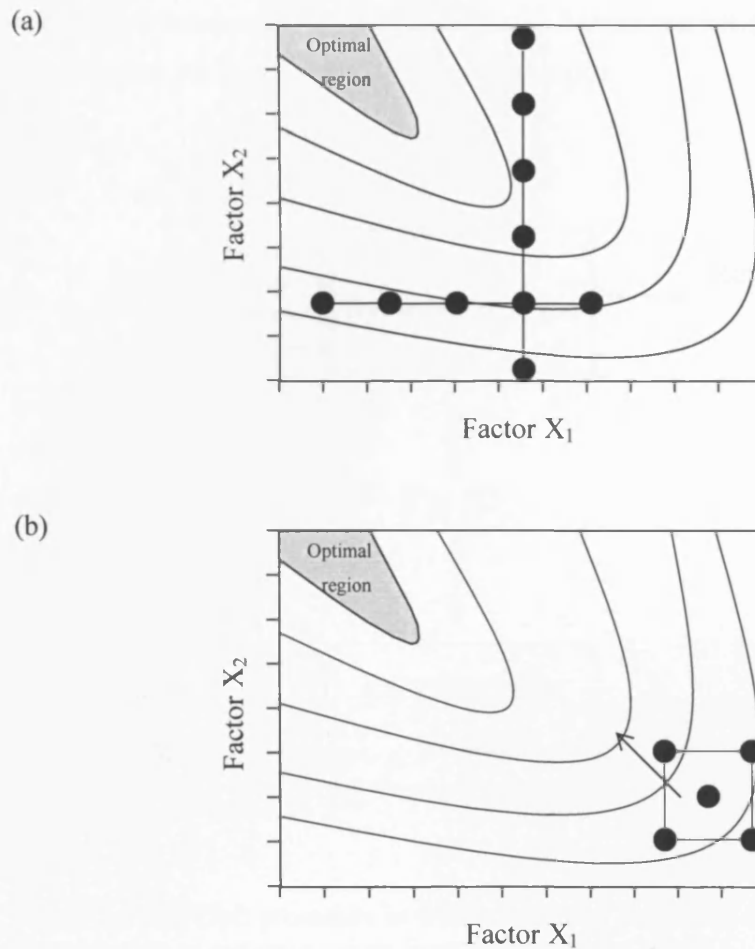
### **1.3 Design of Experiments**

#### ***1.3.1 Description and benefits over traditional approaches***

High-throughput protein expression (HTPE) studies rely on screening a large number of expression parameters. Most approaches, including high-throughput expression screening (HTES), are unstructured and result in only a limited understanding of the system. The task of accurately determining the optimal expression conditions, therefore, can be resource intensive. This requirement may prove prohibitive above a certain number of variables.

The one-factor-at-a-time (OFAT) method represents a more systematic approach to HTPE studies. This approach is illustrated in Figure 1.2 (a), for a system influenced by two parameters. This is representative of a simple biological system in which interactions between variables may exist. Here, the OFAT approach fixes factor  $X_2$  and varies factor  $X_1$  until a maximum response is observed. Then, factor  $X_1$  is fixed at its apparent optimum level and factor  $X_2$  is varied until a new optimum response is observed. In this scenario OFAT analysis fails to identify the optimal conditions as it is unable to account for the interaction of factors.

Statistical Design of Experiments (DoE) is able to overcome the limitations of traditional HTPE approaches. DoE is a structured and efficient approach for determining the mathematical relationship between multiple factors affecting a process and the output(s) of that process as illustrated in Figure 1.3.

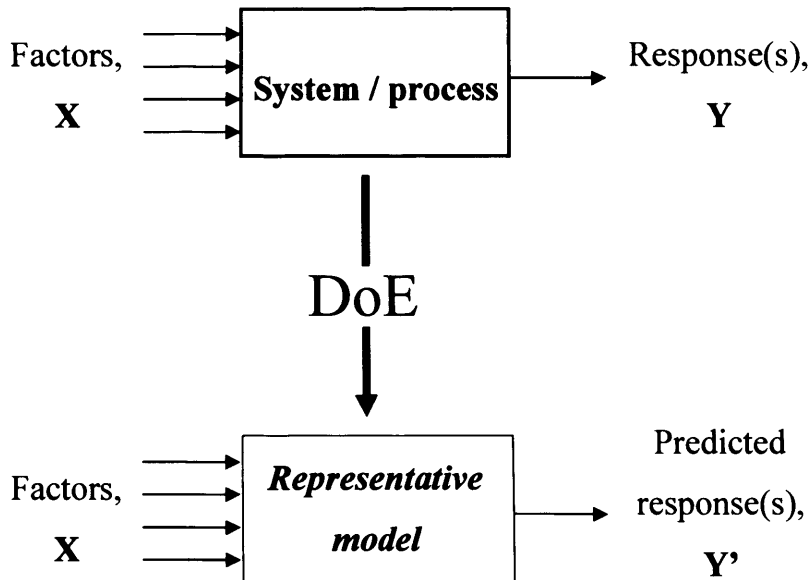


**Figure 1.2** Response contour plot of a system influenced by two factors, investigated using OFAT (a) and DoE (b). Individual experiments are also shown [•].

The DoE approach is illustrated in Figure 1.2 (b), the same system for which the OFAT methodology was previously demonstrated. Here, a small experimental design is applied in which the two factors are varied simultaneously. Due to the symmetry of the experimental design, certain predictions about the system's behaviour can be made. In this scenario, it is possible to predict the *direction* in



which the optimal region is likely to be found. Subsequent designs will be able to determine the precise *location* of the optimal region.



**Figure 1.3** The DoE procedure in which system or process is modelled, enabling the prediction of response values from defined factor levels.

In summary, DoE offers three clear advantages over other HTPE methods. Firstly, as mentioned previously, DoE is able to account for interactions between factors. Design of Experiments is also able to estimate experimental error by considering the overall difference between observed response values and those predicted by the mathematical model. In this way, real effects can be distinguished from experimental error. Finally, in quantifying the real effects, DoE is able to produce reliable maps of the investigated process or system from which the optimal region can be identified. This last point is of key importance to bioprocess development, where fermentation optimisation is a crucial precursor to successful scale-up; results obtained under sub-optimal conditions such as

oxygen limitation are likely to be misleading (Buchs, 2001; McDaniel *et al.*, 1965).

### **1.3.2 Pre-experimental steps**

The first step in implementing DoE is the problem formulation and its objective is to establish and clarify the intentions underlying an experimental investigation. The problem formulation considers six aspects: (1) the experimental objective, (2) factors, (3) responses, (4) model, (5) design, and (6) worksheet.

#### 1.3.2.1 Experimental objective

The exact aims of DoE depend on the stage of an investigation/experimental process to which it is applied. Design of Experiments need not be applied to every stage of the experimental process, but it is often prudent to break down an investigation into several stages. These stages, which are discussed further in Chapter 3, normally include familiarisation, screening, finding the optimal region, optimisation, robustness testing and mechanistic modelling (Eriksson *et al.*, 2000). By characterising the system gradually in this way, the cost of resources is minimised. By comparison, going straight into an optimisation design with little or no prior knowledge of a system would require huge amounts of resources and would defeat the overall purpose of DoE.

#### 1.3.2.2 Specification of factors

Here, the factors relevant to the current experimental objective are listed and categorised according to whether they are quantitative or qualitative, controlled or uncontrolled, etc. The range and levels of each factor are specified, and this is influenced by the experimental objective; in screening, factors have relatively large ranges and are varied over only two levels whereas during optimisation, factor ranges are narrow, involving three or more levels per factor. If factor transformations are required, they must be done at this stage, prior to generation of a worksheet. One heuristics approach is to transform a factor where the range

varies by more than 10-fold. Finally, factor levels are coded between values of -1 and +1 to facilitate later-stage data analysis.

#### 1.3.2.3 Specification of responses

Appropriate process performance indicators should have been identified at the familiarisation stage of the experimental process and are listed here.

#### 1.3.2.4 Model selection

Selection of a model at this stage is partly constrained by the experimental objective. If screening was selected, a linear or interaction polynomial model will suffice, whereas a quadratic polynomial model is the minimum requirement for optimisation. This is intuitive because linear and interaction polynomial models have no optimum.

#### 1.3.2.5 Generation of an experimental design and creation of a worksheet

The choice of experimental design is intimately linked to the specified model, and hence experimental objective, but also to the number of factors, their levels and whether they are quantitative or qualitative. Some of the most commonly used classical designs include fractional factorial, full factorial and composite designs (Table 1.2). Two-level fractional factorial and full factorial designs are used for screening and yield linear and interaction models. These designs can also be used for optimisation if each factor is varied over a minimum of three levels. The last row of Table 1.2 displays composite designs which are used in optimisation.

The abovementioned designs are all of regular geometry and so each is only applicable to a regular experimental region. Where the experimental region is irregular, i.e. when certain factor combinations are not permitted, computer-

generated D-optimal designs are employed. These designs allow for the removal of unfeasible experiments from the overall candidate set before a design is generated. An algorithm then computes the optimal design region from the remaining experiments. D-optimal designs offer several other advantages including the capability of investigating qualitative factors within the experimental region.

Once an appropriate design is selected, its corresponding algorithm is used to generate a worksheet in which the required experiments are specified.

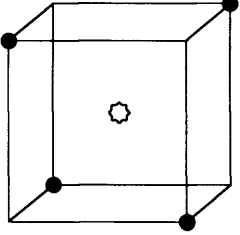
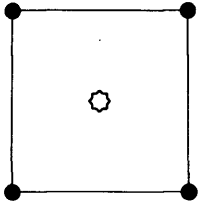
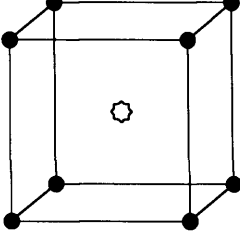
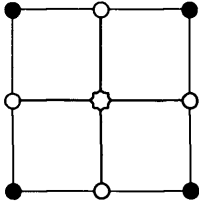
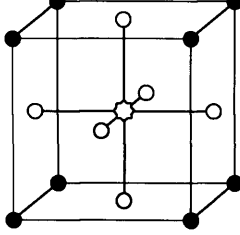
### ***1.3.3 Post-experimental steps***

The analysis of experimental data generated through DoE consists of three stages. The first stage, *evaluation of raw data*, focuses on the general trends and irregularities within the data. The second stage, *regression analysis and model refinement* involves calculating the actual terms within the regression model and evaluating its overall accuracy. The *use of the regression model* constitutes the final stage of analysis where the model is used to map the process and determine the optimal factor settings.

#### **1.3.3.1 Evaluation of raw data**

A large variety of tools are available at this stage of analysis. A *replicate plot*, for example, illustrates the size of replicate error relative to the overall response variation. A *histogram of response* illustrates the distribution of a response which should be continuous and normally distributed under ideal circumstances. Otherwise, the response may need to be transformed in which case a *Box-Cox plot* serves as a guide towards the correct response transformation. An *N-probability plot* is a useful tool for detecting deviating experiments, or outliers. If an outlier is detected, a decision can be made to repeat the experiment, omit it from the design or simply leave it as it is.

**Table 1.2** Pictorial representation of various experimental designs. The star at the centre of each design represents centre-point experiments.

	<i>2 factors</i>	<i>3 factors</i>	<i>&gt;3 factors</i>
<i>Fractional factorial</i>			<i>Balanced fraction of a hypercube</i>
<i>Full factorial</i>			<i>Hypercube</i>
<i>Composite</i>			<i>Hypercube + axial points</i>

### 1.3.3.2 Regression analysis and model refinement

Two important diagnostic tools at this stage include the  $R^2/Q^2$  performance indicators and analysis of variance, *ANOVA*. The  $R^2$  and  $Q^2$  values vary between 0 and 1 and represent the fraction of the response variation explained and predicted by the model, respectively. These performance indicators should be used in conjunction with *ANOVA*, which formally tests if a given model explains a significant fraction of the response variation and whether the unexplained fraction is due to noise or model error. During optimisation, these performance indicators can be used to monitor the performance of a model as terms are removed via the *coefficient plot* until the overall model performance no longer increases. A *parity plot* of observed vs. predicted response values can then be used to illustrate final model performance.

### 1.3.3.3 Use of regression model

The purpose of a regression model depends upon the experimental objective. For screening, the model is used as a guide to determine which factors may have strong effects on the process. For optimisation, the model is used to map the process via *response contour/surface plots* and predict the optimal factor settings.

## 1.3.4 *Relevant studies incorporating DoE*

Given the large number of culture parameters which are known to affect soluble protein expression, the task of optimising expression levels is large and complex. The possibility exists that a thorough understanding of this complex relationship will result in significant improvements in expression levels. Currently, however, this opportunity has not been fully exploited.

The recent application of methods which analyse the effects of simultaneously varying several parameters, such as the method of DoE, is certainly a step in the right direction. Unfortunately, the majority of relevant studies are limited by a small number of investigated parameters and/or incomplete application of the DoE toolset. For example, the parameters investigated by most studies can broadly be grouped by one of three categories: media formulation (Adinarayana and Ellaiah, 2002; Galindo *et al.*, 1990; Hounsa *et al.*, 1996; Nikerel *et al.*, 2005; Ren *et al.*, 2006; Sunitha *et al.*, 1999; Zhang, 2006), growth conditions (Dutta *et al.*, 2004; El Helow *et al.*, 2000; Roebuck *et al.*, 1995) or induction parameters (Cao, 2006; Swalley *et al.*, 2006; Urban *et al.*, 2003; Wang *et al.*, 2005; Xie *et al.*, 2003). By investigating only a small number of parameters, it is unclear whether these studies have fully optimised protein expression levels. Furthermore, these studies do not consider the effects of engineering parameters and so the results are of limited benefit to scale-up.

The mathematical model which results from the implementation of DoE serves as a very powerful predictive tool when applied correctly. However, further investigation into the abovementioned reports reveals that some authors conduct statistical analyses on their regression models, but then do not utilise this information to alter their models accordingly (Adinarayana and Ellaiah, 2002; Nikerel *et al.*, 2005; Ren *et al.*, 2006; Wang *et al.*, 2005). Other authors do not perform statistical analyses and so it is difficult to evaluate which of the investigated factors are important (Dutta *et al.*, 2004; El Helow *et al.*, 2000; Hounsa *et al.*, 1996; Sunitha *et al.*, 1999). In two other studies, the models are not stated (Galindo *et al.*, 1990; Urban *et al.*, 2003) and this severely limits the interpretation of results.

#### **1.4 Scale-up of protein expression**

The protein generated from small to medium scale expression studies may, in some instances, be sufficient for downstream purposes. In many cases the generation of protein for high throughput screening, secondary selectivity assays

---

or a structural biology campaign may require gram quantities of protein over a long period of time. It is therefore highly desirable to produce single preparations of the protein whenever possible, which minimises batch-to-batch variability and reduces the demand placed on resources. For such larger-scale production of protein, shake-flask cultures can be employed, but for volumes in excess of 10 l, the use of stirred-tank reactors (STRs) is more commonplace (Hunt, 2005).

Successful scale-up means that a process has been designed and built giving a predictable increase in production capacity (Reisman, 1993). More specifically, the aim here is to translate the data generated from HTPE studies to production scale rapidly and accurately (Freyer *et al.*, 2004). The optimal conditions determined from screening at small-scale are also likely to provide the environment for the production of high quantities of soluble protein (Hunt, 2005). To achieve accurate and rapid scale-up, it is first necessary to understand the total environment of the cell (Young, 1979) and consider how this changes with scale. The total environment may encompass the biological, chemical and physical variables relating to fermentation (Freyer *et al.*, 2004; Junker, 2004). A complete catalogue of these factors is also detailed extensively by Reisman (1993).

Biological factors are largely associated with the state of the inoculum, i.e. the number of generations associated with its development, phase of growth of inoculum cells, cell density and cell viability. These variables also extend to secreted products and their interactions, inhibitory factors and cell-cell interactions. Chemical factors typically include the type and concentration of pH control and protein induction agents, composition and quality of nutrients, water quality and foam formation. Physical factors include bioreactor configuration, aeration, agitation and chemical homogeneity, temperature control, pressure and method of medium sterilisation. This list is by no means exhaustive and serves as a general guide for all cell culture systems.

Some factors will have little effect on the system but may still cause an undesired deviation from predicted operation. These factors typically include ambient



---

temperature, humidity and variability in raw material composition. They should be identified and regulated as far as possible so as to establish a fine degree of process control.

Some variables are not easily translated. For example the host of physical variables which arise upon scale-up due to the physical geometrical difference between an MWP and an STR requires further investigation. Ordinarily, these variables are summarised by engineering parameters such as impeller tip speed, superficial gas velocity, mixing time or power per unit volume (Junker, 2004). Traditional scale-up techniques prescribe that one or more these parameters should remain constant upon scale-up. In the context of MWPs these parameters are either inapplicable or difficult to characterise. Moreover, the significance of certain phenomena such as surface tension effects may not be shared between scales. Thus many established scale-up correlations are unusable. Yet without proper process characterisation, it is likely that traditional scale-up problems such as those presented by Humphrey (1998) would be amplified.

Initial studies have suggested that scaling-up on the basis of a constant mass transfer coefficient ( $k_La$ ) appears to be the most appropriate approach when dealing with microorganisms growing under aerobic conditions (Ferreira-Torres *et al.*, 2005; Micheletti *et al.*, 2006). Adequate oxygen provision remains a significant challenge for aerobic growth in shaken MWP fermentations (Maier and Buchs, 2001) and results obtained under oxygen limiting conditions are likely to be misleading particularly for scale-up purposes (Buchs, 2001; McDaniel *et al.*, 1965). Considerable effort, therefore, has been put recently into the characterisation within MWPs of engineering parameters such as  $k_La$ .

Different  $k_La$  characterisation techniques such as sulphite oxidation (Hermann *et al.*, 2003; Kensy *et al.*, 2005), dynamic gassing-out (Duetz and Witholt, 2001) and enzymatic methods (Duetz and Witholt, 2004; Ortiz-Ochoa *et al.*, 2005) have been employed. These methods are, however, often time consuming, labour intensive and in the case of the sulphite oxidation technique, notoriously non-robust (Van't Riet and Tramper, 1991). Furthermore, unless experiments are

conducted in fermentation broth, results may be deceiving due to rheological differences (Stanbury and Whitaker, 1993).

Recently, however, miniature oxygen probes have been used to characterise oxygen transfer directly via the static gassing-out method (John *et al.*, 2003b; Doig *et al.*, 2005), which overcomes many of the aforementioned complexities and enables the same measurement principle to be applied at both microwell and larger scales. Doig *et al.* (2005) also presented a  $k_La$  correlation derived from round-well MWP. These advances represent valuable tools in the rapid and accurate scale-up of MWP fermentations.

## 1.5 Thesis aims and objectives

The overall aim of this project is to establish a generic framework that will underpin the generation of large quantities of soluble protein in *E. coli* in a rapid and cost-effective manner. Specific objectives will be:

- to determine the key components and overall structure of the framework and explore the practical issues of implementing the framework components within this project. The framework is outlined in Chapter 3 and then refined through practical experimentation in subsequent chapters;
- to demonstrate the sequential application of DoE in engineered microwell experiments for the optimisation of soluble protein expression in *E. coli*. This will be described in Chapter 4;
- to characterise the engineering environment in microwell cultures and define a reliable basis for the reproducible scale-up of optimised culture conditions. This will be described in Chapter 5;
- to show that the insights gained at the microwell scale can inform pilot-scale operation. This will be described in Chapter 6;

- to provide a thorough and quantitative analysis of the inherent time and cost savings provided by the current framework over traditional approaches. This will be discussed in Chapter 7.
- to explore how the current framework may be developed towards further industrial applications. This will be discussed in Chapter 8.

## 2 Materials and methods

### 2.1 Introduction

The aim of this chapter is to provide a clear description of all the materials and methods used in this research. Section 2.2 describes the preparation of fermentation media used in this work, the bacterial expression system and the preparation of inoculum for all subsequent fermentations. Section 2.3 describes the experimental design and fermentation procedures for microwell plate fermentations with the aim of identifying the optimal conditions for recombinant soluble protein expression. This section also illustrates the procedures for standard shake-flask fermentations from which the soluble protein yield serves as a reference. Section 2.4 describes the methods for characterising the mass transfer coefficient,  $k_La$ , as a basis for scale-up at all scales of fermentation. The scale-up of microwell plate fermentations on the basis of constant  $k_La$  is described in Section 2.5, together with statistical procedures for evaluating the similarity of performance between fermentations. Section 2.6 describes the quantitative assay procedures used in this work for biomass concentration, glycerol concentration and relative soluble firefly luciferase activity.

### 2.2 General methods

#### 2.2.1 Media preparation

All chemicals used were purchased from Sigma-Aldrich Chemical Company (Dorset, UK) unless stated otherwise and reverse osmosis (RO) water was used throughout. The three media types used were prepared as follows:

**Luria Bertani (LB):** 5 g.l<sup>-1</sup> yeast extract, 10 g.l<sup>-1</sup> tryptone and 10 g.l<sup>-1</sup> NaCl, pH 7.2;

**Terrific Broth (TB):** 24 g.l<sup>-1</sup> yeast extract, 12 g.l<sup>-1</sup> tryptone, 4.0 ml.l<sup>-1</sup> glycerol (100 %), 2.31 g.l<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub> and 12.54 g.l<sup>-1</sup> of K<sub>2</sub>HPO<sub>4</sub>, pH 7;

**Glucose M9Y (GM9Y):** 5 g.l<sup>-1</sup> yeast extract, 4.0 g.l<sup>-1</sup> glucose, 1 g.l<sup>-1</sup> NH<sub>4</sub>Cl, 0.5 g.l<sup>-1</sup> NaCl, 241 mg.l<sup>-1</sup> MgSO<sub>4</sub>, 11 mg.l<sup>-1</sup> CaCl<sub>2</sub>, 6.0 g.l<sup>-1</sup> Na<sub>2</sub>HPO<sub>4</sub> and 3.0 g.l<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, pH 6.9.

GM9Y media was obtained from Athena Environmental Sciences (Baltimore, MD, USA). The glycerol and glucose components were filter sterilised through a 0.2 µm filter (Sartorius Ltd, Epsom, U.K.). The phosphate components were heat sterilised separately and added to the respective media using aseptic technique. All other media components were combined as dry powders and were heat sterilised for 20 min at 121 °C. After sterilisation, all media formulations were supplemented with 30 mg.l<sup>-1</sup> kanamycin to prevent contamination and hence uncontrolled depletion of the media.

### 2.2.2 Bacterial expression system

The bacterial strain *E. coli* BL21 (DE3) (Invitrogen, Paisley, UK) was used as the expression host for all experiments. This strain is deficient in two key proteases (*lon* protease and the *ompT* outer membrane protease) which results in a reduced degradation of heterologous protein. This strain contains the λDE3 lysogen which carries the gene for T7 RNA polymerase under control of the *lacUV5* promoter. Expression of T7 RNA polymerase is inducible by isopropyl-β-D-thio-galactopyranoside (IPTG), a lactose analogue, which binds to the *lacI* repressor and reduces its affinity for the *lac* operator.

The gene encoding a thermally stable mutant of firefly luciferase (FFL) (Law *et al.*, 2006) was provided as a kind gift by Lumora Ltd (Cambridge, UK) and subsequently cloned into a pET30a expression vector (Novagen, Nottingham, UK). This vector, which confers kanamycin resistance upon its host, expresses recombinant DNA under the control of a T7 promoter. Due to the high selectivity of T7 RNA polymerase for the T7 promoter, transcription of target DNA by *E. coli* RNA polymerase in the absence of T7 RNA polymerase is very low.

### **2.2.3 Inoculum preparation**

*E. coli* BL21 (DE3) cells were transformed with pET30a-FFL using standard methods (Maniatis *et al.*, 1989) and grown overnight at 37 °C on Luria Bertani (LB) agar plates containing 30 mg.l<sup>-1</sup> kanamycin. A single colony of *E. coli* BL21 (DE3) transformed with pET30a-FFL was transferred to a 2 l shake-flask using a sterile loop. Shake-flasks contained 500 ml of either LB, TB or GM9Y depending on the experiments to be performed. Cultures were grown overnight at 37 °C and 200 rpm in a Multitron incubator shaker (Infors, Bottmingen, Switzerland). Cultures grown in the appropriate medium were used for subsequent inoculation of microwell plate (MWP) experiments described in Section 2.3.

For the MWP and stirred-tank reactor (STR) fermentations described in Section 2.5, glycerol-cell stocks were used for inoculation of these fermentations. Glycerol-cell stocks were prepared in the following manner. Aliquots of 0.85 ml of the abovementioned overnight culture were mixed thoroughly with 0.15 ml of glycerol previously sterilised by autoclaving at 121 °C for 20 min. These glycerol-cell stocks were stored in 2 ml eppendorf tubes (Eppendorf, Cambridge, UK) at -80 °C. When required, glycerol-cell stocks were thawed at room temperature and transferred to a 2 l shake-flask containing 500 ml TB. Cultures were grown overnight at 37 °C and 200 rpm in a ISF-1-V incubator shaker (Adolf Kühner AG, Birsfelden, Switzerland) and used to inoculate the fermentations described in Section 2.5.

## **2.3 Methods for optimisation of soluble protein expression at the microwell scale**

### **2.3.1 Microwell plate (MWP) fermentations**

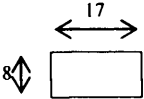
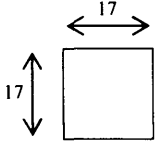
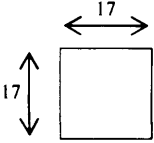
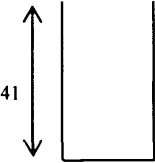
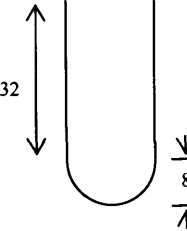
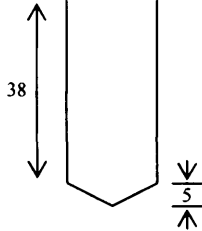
Three different deep-well polypropylene MWPs were used in this work: (P1) 48 rectangular wells, flat bases (Whatman PLC, Middlesex, U.K.), (P2) 24 square wells, round bases (Whatman PLC) and (P3) 24 square wells, pyramidal bases

---

(DOT Scientific Inc. Burton, MI, USA). These plates are described in further detail in Table 2.1.

For the initial familiarisation and screening experiments (Sections 2.3.2 and 2.3.3), an overnight *E. coli* culture was prepared as described in Section 2.2.3 and used to inoculate 40 ml of LB, TB or GM9Y sterile media such that an initial OD<sub>600</sub> of 0.1 was obtained (an initial OD<sub>600</sub> of 0.001 was used for the familiarisation experiments described in Section 2.3.2). Aliquots of the inoculated media were transferred into one of the three MWP's and the plates were sealed with gas permeable membranes (ABgene, Epsom, UK) to minimise the loss of liquid due to evaporation. Cultures were grown in a HiGro incubator-shaker (GeneMachines, Huntingdon, UK), with an orbital shaking diameter of 8 mm, under the experimental conditions described in Table 2.2. Optical density measurements at a wavelength of 600 nm (OD<sub>600</sub>) of broth samples were made both immediately prior to and again after protein induction and were subsequently converted into dry cell weight measurements as described in Section 2.6.1. FFL was expressed as an intracellular enzyme by the addition to each well of 20 µl isopropylthiogalactosidase (IPTG) solution at the appropriate concentration. After the induction period, cells were harvested by centrifugation at 4 °C for 10 min at 3000 rpm using an Eppendorf 5417R centrifuge (Cambridge, UK) and samples were stored at -80 °C. Identical methods were adopted for optimisation experiments (Section 2.3.4) but only TB media was used and the levels of the factors were altered as described in Table 2.2.

**Table 2.1** Description of MWP geometries used throughout this work. All dimensions are given in mm, correct to the nearest mm. Drawings not to scale.

	<b>Plate P1</b>	<b>Plate P2</b>	<b>Plate P3</b>
<b>Total volume</b>	48 wells x 5 ml.well <sup>-1</sup>	24 wells x 10 ml.well <sup>-1</sup>	24 wells x 10 ml.well <sup>-1</sup>
<b>Plan view of well</b>			
<b>Side view of well</b>			

**Table 2.2** Specification of factors and settings investigated within familiarisation, screening and optimisation experiments. Pre-I and post-I are abbreviations for pre-induction and post-induction, respectively. Plate P1 had 48 rectangular wells with flat bases, plate P2 had 24 square wells with round bases and plate P3 had 24 square wells with pyramidal bases. <sup>a</sup>The familiarisation experiments had a lower shaking speed of 50rpm. <sup>b</sup>An upper liquid fill volume of 2 ml was used in conjunction with the 48-well plate to prevent splashing.

Name	Abbr.	Units	Experimental settings	
			Familiarisation / Screening	Optimisation
Growth medium	x <sub>1</sub>	-	Glucose M9Y, LB Broth, Terrific Broth	Terrific Broth
Plate geometry	x <sub>2</sub>	-	P1, P2, P3	P3
Liquid fill volume	x <sub>3</sub>	ml	1, 2 <sup>a</sup> , 3	1, 2, 3
Pre-I temperature	x <sub>4</sub>	°C	17, 37	27
Pre-I shaking speed	x <sub>5</sub>	rpm	100 <sup>b</sup> , 500	100, 300, 500
Pre-I period	x <sub>6</sub>	h	2, 6	1, 4, 7
Inducer concentration	x <sub>7</sub>	μM	50, 1000	500
Post-I temperature	x <sub>8</sub>	°C	17, 37	23, 30, 37
Post-I shaking speed	x <sub>9</sub>	rpm	100 <sup>b</sup> , 500	100, 300, 500
Post-I period	x <sub>10</sub>	h	3, 15	6, 12, 18



### 2.3.2 Design of MWP familiarisation experiments

Ten factors were initially chosen for investigation towards experimental familiarisation. These factors are specified in Table 2.2 and the rationale for their selection will be discussed further in Section 4.2.1. A D-optimal design (Eriksson *et al.*, 2000) was chosen to generate an appropriate experimental plan. D-optimal designs are capable of handling both quantitative variables (time, concentration, etc.) and qualitative variables (media type, plate geometry, etc.). They also allow for the removal of experiments which have unfavourable factor combinations from the initial candidate set of experiments. For example, preliminary experiments had shown that excessive splashing would occur within the 48-well plates at a shaking speed of 500 rpm, above a liquid fill volume of 2 ml. Unfavourable factor combinations were thus removed from the candidate set prior to generation of the experimental plan. A  $\log_{10}$  transformation of the inducer concentration factor range was also made at this time. The general approach is to transform any factor range which spans one or more orders of magnitude, so as to preserve the orthogonal shape of the design region (Eriksson *et al.*, 2000).

Table 2.3 shows the D-optimal familiarisation design in which ten factors were investigated in 30 experimental runs. Two replicated centre-point experiments were included for estimation of pure error. The soluble protein yield results shown in Table 2.3 will be discussed in detail in Chapter 4. All factors were varied over two levels (low and high) and for statistical calculations the actual levels of each variable ( $X_i$ ) were coded as  $x_i$  according to the following relationship:

$$x_i = \frac{2z_i - z_i^{\max} - z_i^{\min}}{z_i^{\max} - z_i^{\min}} \quad (2.1)$$

where  $z_i$  is the actual variable value of  $X_i$  and the superscripts “max” and “min” denote the maximum and minimum values used.

**Table 2.3** D-optimal design matrix for the initial assessment of soluble protein yield at the familiarisation stage. Quantitative variables are coded at low (-1), medium (0) and high (+1) levels.

Observations	x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	x <sub>4</sub>	x <sub>5</sub>	x <sub>6</sub>	x <sub>7</sub>	x <sub>8</sub>	x <sub>9</sub>	x <sub>10</sub>	Mean soluble protein yield (RLU.ml <sup>-1</sup> )
N1	M9Y	P1	0	1	1	-1	0	-1	-1	-1	0
N2	M9Y	P1	0	1	-1	1	0	-1	-1	-1	1221
N3	M9Y	P1	0	-1	1	1	-1	1	1	1	404587
N4	LB	P1	0	-1	-1	1	-1	-1	-1	1	136
N5	TB	P1	0	1	-1	1	0	1	-1	1	6947
N6	LB	P1	-1	1	-1	-1	-1	1	1	-1	0
N7	TB	P1	-1	-1	1	-1	0	-1	1	1	0
N8	TB	P1	-1	-1	1	-1	-1	1	1	-1	129
N9	M9Y	P2	1	-1	-1	-1	-1	1	-1	1	2748
N10	M9Y	P2	1	-1	-1	1	0	1	1	-1	0
N11	LB	P2	1	1	-1	-1	0	-1	1	1	319
N12	TB	P2	1	1	1	-1	-1	-1	-1	-1	0
N13	TB	P2	1	-1	1	-1	0	-1	1	1	59
N14	M9Y	P2	-1	1	1	-1	-1	1	-1	-1	11260
N15	LB	P2	-1	-1	1	1	-1	-1	-1	1	128
N16	LB	P2	-1	-1	1	1	0	1	1	-1	0
N17	TB	P2	-1	1	-1	1	0	1	-1	1	9983
N18	TB	P2	-1	1	-1	1	-1	-1	1	-1	0
N19	M9Y	P3	1	1	1	1	-1	-1	1	1	60559
N20	LB	P3	1	-1	-1	-1	-1	-1	-1	-1	0
N21	LB	P3	1	1	1	1	0	1	1	-1	0
N22	TB	P3	1	-1	1	1	-1	1	-1	-1	0
N23	TB	P3	1	1	-1	-1	-1	1	1	1	374457
N24	M9Y	P3	-1	-1	-1	-1	0	1	-1	1	350
N25	M9Y	P3	-1	-1	-1	-1	0	-1	1	-1	0
N26	M9Y	P3	-1	1	1	1	-1	-1	1	1	6896
N27	LB	P3	-1	1	1	-1	0	1	-1	1	4922
N28	TB	P3	-1	-1	-1	1	0	-1	-1	-1	0
N29	M9Y	P1	0	0	0	0	0	0	0	0	78111
N30	M9Y	P1	0	0	0	0	0	0	0	0	4076

Each experimental run was performed in quadruplicate and the mean soluble protein yield was recorded. A linear model was then fitted to the data and assumed the following form:

$$Y = \beta_0 + \sum \beta_i x_i \quad (2.2)$$

where  $Y$  is the dependent variable (soluble protein yield),  $\beta_0$  is the independent term and  $\beta_i$  are the regression coefficients related to the main effects. The size of each coefficient is proportional to the influence of a particular factor on the response and can be used to determine which factors should be investigated further.

### ***2.3.3 Design of MWP screening experiments***

The ten factors identified at the familiarisation stage were investigated further through a set of screening experiments. The aim here was to elucidate which of the factors influenced soluble protein yield the most. Identical procedures used in the familiarisation stage were adopted here, but with altered factor ranges, as specified in Table 2.2. The D-optimal experimental design for the screening experiments is shown in Table 4.2.

### ***2.3.4 Design of MWP optimisation experiments***

A Central Composite Face (CCF) design was employed to determine the optimal levels of the key variables identified from initial screening experiments. This design requires each factor to be varied over three levels only (low, medium and high) and fewer runs are needed compared to an equivalent Box-Behnken or 3-level full factorial design (Eriksson *et al.*, 2000). All factors were coded as described in Section 2.3.2. Table 4.3 shows the experimental plan for optimisation in which six factors were investigated in 50 experimental runs. Six replicated centre-point experiments were also included for estimations of pure

error. Each experimental run was performed in quadruplicate and the mean response values were recorded.

The behaviour of the system was modelled using the following equation:

$$Y = \beta_0 + \sum \beta_i x_i + \sum \beta_{ij} x_i x_j + \sum \beta_{ii} x_i^2 \quad (2.3)$$

where  $\beta_{ij}$  and  $\beta_{ii}$  are the regression coefficients corresponding to the second-order interactions and squared main effect terms, respectively. Regression coefficients whose confidence intervals included zero were removed from the model in a step-wise manner, starting with the least significant terms. The resulting regression model for soluble protein yield was maximised and this enabled the prediction of the optimal factor settings. A verification experiment was then performed at these settings and factors which did not appear in the model were maintained at their mid-point values.

### **2.3.5 Reference shake flask fermentations**

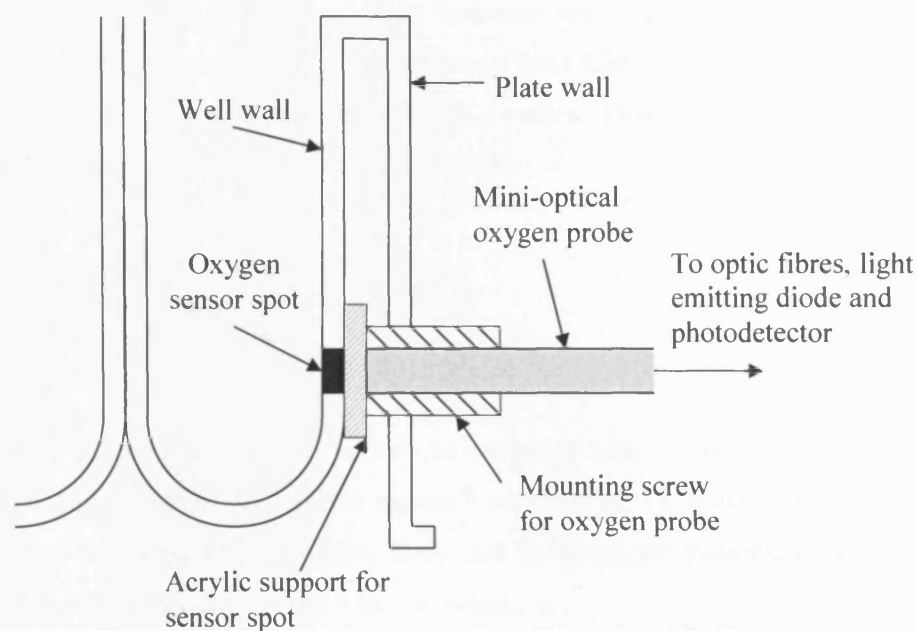
An aliquot of 50 ml overnight *E. coli* culture, prepared as described in Section 2.2.3, was used to inoculate 1 l of TB medium supplemented with 30 mg.l<sup>-1</sup> kanamycin in a Nalgene 2 l baffled shake flask (VWR International Ltd, Leicestershire, U.K.). This culture was grown at 27 °C and 185 rpm in a Multitron incubator shaker (Infors, Bottmingen, Switzerland). The culture was induced with 0.5 mM IPTG when the OD<sub>600</sub> reached 1.0 at approximately 2 h 40 min. After a 4 h induction period, cells were harvested by centrifuging 1 ml aliquots of culture at 4 °C for 10 min at 3000 rpm in an Eppendorf 5417R centrifuge (Cambridge, UK). Samples were subsequently stored at -80 °C.

## 2.4 Methods for characterisation of bioreactor mass transfer coefficients ( $k_La$ )

### 2.4.1 Measurement of MWP $k_La$ values

The MWPs described in Section 2.3.1 were modified such that three wells on each plate had a small oxygen sensor spot (Precision Sensing GmbH, Regensburg, Germany) mounted flush with the inside wall of each well. The principle of the sensor operation is based on the quenching of luminescence caused by collisions between molecular oxygen and luminescent dye molecules on the surface of each sensor spot. The sensor spots were located as close as possible to the base of each well. Consequently, a minimum liquid fill volume of 2 ml (rounded up to the nearest ml) was required to completely submerge the sensor spot in each well of plate P1 at all shaking speeds. Similarly, a minimum liquid fill volume of 3 ml was required for plates P2 and P3. Miniature optical oxygen probes (Precision Sensing GmbH) were then mounted in line with the sensors, through the outer wall of each plate. Each probe was connected via optic fibres to a light emitting diode (LED) to illuminate the sensor spots and also to a photodetector (Precision Sensing GmbH) to measure the light intensity. The overall setup is illustrated in Figure 2.1.

TB medium was used for all  $k_La$  characterisation experiments at the MWP scale. An aliquot 0.2 ml.l<sup>-1</sup> polypropylene glycol was first added to the medium after sterilisation, as would be used in subsequent STR  $k_La$  characterisation experiments to prevent foaming. Aliquots of 2ml of this TB medium were then added to the modified wells of plate P1 and 3ml to those of plates P2 and P3. Oxygen probes were calibrated at 0 % air saturation at room temperature immediately after exposing the medium in each well to a blanket of nitrogen for a period of 20 min. Similarly, the probes were calibrated at 100 % air saturation at room temperature immediately after a 20 min period of pumping air continuously into the headspace above the medium in each well.



**Figure 2.1** Cross-sectional diagram illustrating the setup of a microwell plate modified with the oxygen sensor system as described in Section 2.4.1.

$k_La$  measurements were made for each plate at five different shaking speeds (100, 200, 300, 400 and 500 rpm) on an Infors Orbitec orbital shaker (Bottmingen, Switzerland). This orbital shaker has an identical orbital shaking diameter (8 mm) to that of the HiGro incubator-shaker used in MWP fermentation experiments (Section 2.3.1). The dynamic gassing-out technique (Van't Riet, 1979) was used for all  $k_La$  measurements. The appropriate agitation speed was selected and a blanket of nitrogen was placed over the wells until the dissolved oxygen tension (DOT) reached zero. At this point the nitrogen supply was removed, the plate was sealed with a gas permeable membrane (ABgene, Epsom, UK) and the rate at which the DOT increased was recorded. Each measurement was performed in triplicate.

The response time for the oxygen probe was measured as follows. A 3 ml aliquot of water was first added to the modified well of plate P3 and the plate was shaken at 500 rpm so that the oxygen concentration reached 100 % saturation. A 2 ml aliquot of 2 M  $\text{Na}_2\text{SO}_3$  was then added to the well, and the time taken for the oxygen concentration to reach 37 % was recorded. This procedure was

repeated 4 times. The mean probe response time was relatively long (25 s) compared to those quoted in the literature (Van't Riet, 1979) so it was necessary to account for this when calculating  $k_La$  values. This was achieved using the following expression (Lamping *et al.*, 2003):

$$C_p = \frac{1}{t_m - \tau_p} \left[ t_m \exp\left(\frac{-t}{t_m}\right) - \tau_p \exp\left(\frac{-t}{\tau_p}\right) \right] \quad (2.4)$$

where  $C_p$  is the undissolved fraction of oxygen at time  $t$ ,  $\tau_p$  is the probe response time and  $t_m = 1/k_La$ . An iterative approach was then used to solve this equation for  $k_La$  at each time point between 20 % and 80 % oxygen saturation. From these results, a mean value for  $k_La$  was calculated.

#### 2.4.2 Specific air-liquid surface area measurements in MWP

Doig *et al* (2005) previously established a correlation for the prediction of  $k_La$  values in MWPs. In order to adapt the correlation for use with the MWP geometries used in this study, it was necessary to measure the specific air-liquid surface area in the wells over same the range of conditions used in the  $k_La$  experiments (Section 2.4.1). All wells in plates P2 and P3 were each filled with 3 ml RO water and shaken at one of six different speeds (0, 100, 200, 300, 400 and 500 rpm) on an Infors Orbitec orbital shaker at 23 °C for 4 h. The mean evaporation rate over this period was measured. The evaporation rate from plate P1 was measured in the same way, but using a liquid fill volume of 2 ml per well and an upper shaking speed of 400 rpm to avoid splashing. Once this evaporation rate data was measured, the specific air-liquid surface areas were calculated according to Equation 2.5:

$$\frac{a_f}{a_i} = \frac{\text{shaken evaporation rate}}{\text{unshaken evaporation rate}} \quad (2.5)$$

where  $a_f$  and  $a_i$  are the specific air-liquid surface areas ( $\text{m}^{-1}$ ) in the wells under shaken and unshaken conditions, respectively. The values for  $a_i$  were  $68 \text{ m}^{-1}$  for plate P1 and  $96 \text{ m}^{-1}$  for plates P2 and P3.

Previous trial experiments had shown the evaporation rates from MWP's covered with gas-permeable membranes to be very small and hence difficult to measure accurately. Consequently, and in contrast to all other MWP experiments conducted thus far, the experiments here were performed without the use of gas-permeable membranes. Although this would increase the rates of evaporation, the ratio of the shaken evaporation rate to the unshaken evaporation rate should remain constant for each MWP. Consequently, the values of  $a_f$  should also remain the same.

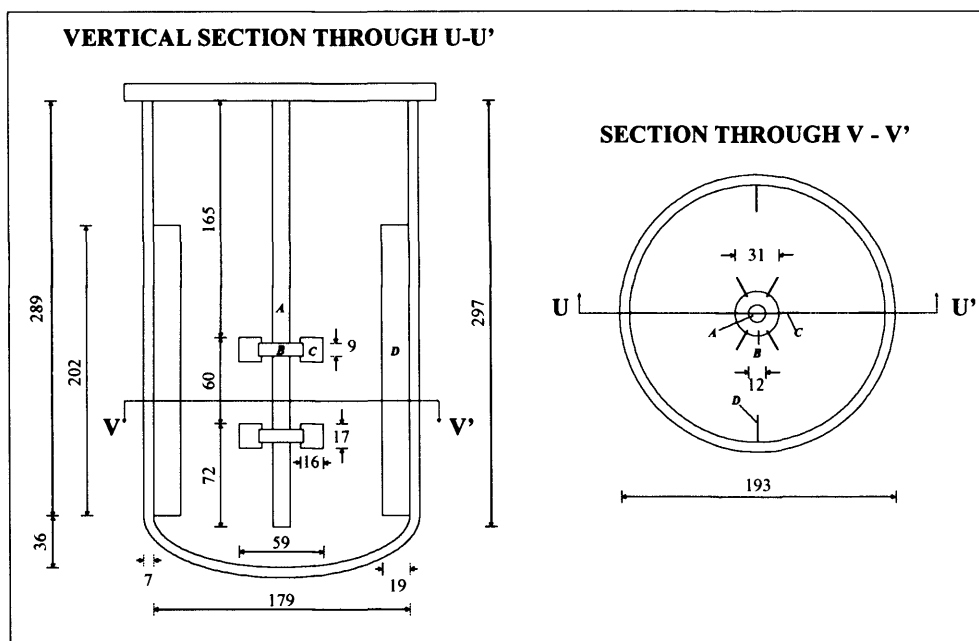
#### **2.4.3 Measurement of stirred-tank reactor (STR) $k_L a$ values**

Two different sizes of STR were used in this work. The first vessel (Figure 2.2) was a 7.5 l BioFlo 110 (New Brunswick Scientific Co., Inc., Edison, NJ, USA) which had two six-bladed Rushton turbine impellers connected to a top-driven impeller shaft. The second vessel was a 75 l LH (formerly LH Fermentation, Reading, U.K.) and had three six-bladed Rushton turbine impellers attached to a bottom-driven impeller shaft. The 75 l LH had the following approximate dimensions: 0.92 m internal vessel height, 0.33 m internal vessel diameter, 0.11 m impeller diameter.

In accordance with the MWP experiments, TB medium was used for all  $k_L a$  characterisation experiments at the STR scale. The 7.5 l STR was first filled with 4.5 l of TB medium, excluding the phosphate components, and the entire vessel was autoclaved at  $121 \text{ }^\circ\text{C}$  for 20 min. Similarly, the 75 l STR was filled with 40.5 l of TB medium excluding the phosphate components and the vessel was steam-sterilised in place at  $121 \text{ }^\circ\text{C}$  for 20 min. The phosphate components were sterilised separately in an autoclave and added aseptically to each vessel, raising the final working volumes to 5 l in the 7.5 l STR and 45 l in the 75 l STR. An



aliquot 0.2 ml.l<sup>-1</sup> polypropylene glycol was added to all media after sterilisation, to prevent foaming during experimentation.



**Figure 2.2** Dimensions of BioFlo 110 7.5 l STR used in this study. All dimensions are given in millimetres. *A* = impeller shaft; *B* = impeller disc; *C* = impeller blade; *D* = baffle. Impeller spacings are set according to the vessel manufacturer's guidelines.

The DOT probes (Mettler-Toledo Ltd., Leicester, U.K.) were calibrated at 100 % air saturation at 23 °C by sparging air into each vessel at a rate of 1 vvm for 20 min. The 0 % air saturation calibration was made by placing the probe in a stream of nitrogen gas outside the vessel.

In accordance with the MWP experiments, the dynamic gassing-out technique was again used to measure  $k_La$  values at the STR scale. Measurements were performed at four different impeller speeds (200, 400, 600 and 800 rpm) in the 7.5 l STR and five impeller speeds (200, 400, 600, 800 and 1000 rpm) in the 75 l STR. Nitrogen was first sparged through the medium in each vessel at a rate of 1 vvm at the predetermined impeller speed. Once the DOT level had fallen to zero, the gas supply was switched rapidly to air at a flow rate of 1 vvm and the rise in

DOT was recorded. All measurements were carried out in triplicate at room temperature. For each vessel, the response time for the oxygen probe was measured five times at each agitation speed using a standard procedure (Dunn and Einsele, 1975). The mean probe response time was relatively high, varying between 28 s and 46 s, and so it was again accounted for when calculating  $k_{La}$  values as described Section 2.4.1.

## 2.5 Scale-up of microwell plate fermentations

### 2.5.1 MWP fermentations

An overnight culture of *E. coli*, prepared as described in Section 2.2.3, was used to inoculate 40 ml of TB media such that an initial OD<sub>600</sub> of 0.1 was obtained. Aliquots of 3 ml inoculated media were then transferred into all wells of plate P3 and the plate sealed with a gas permeable membrane (ABgene, Epsom, UK). Cultures were grown in a HiGro incubator-shaker (Genomic Solutions, Ann Arbor, MI, USA) under the conditions described in Table 2.4 and FFL expression was induced by the addition to each well of 20 µl IPTG solution to a final concentration of 500 µM. The run order of all fermentations was randomised in order to prevent any uncontrolled variables from producing systematic effects on the results. For example, several plates were prepared simultaneously for each fermentation batch; the time delay in inoculating all the wells in sequence represented an uncontrolled variable, the effects of which were minimised through run-order randomisation.

Samples of 50 µl fermentation broth were taken at regular intervals for biomass growth measurements as described in Section 2.6.1. Samples of 1.2 ml fermentation broth were also taken at regular intervals for glycerol and FFL measurements. These samples were first clarified by centrifugation at 14,000 rpm for 10 min at 4 °C using an Eppendorf 5417R centrifuge (Cambridge, UK). The supernatant was assayed for glycerol, as described in Section 0, and the cell pellets were stored at -80 °C for subsequent FFL analysis, as described in Section 2.6.3. All sampling occurred in duplicate from separate sacrificial wells.

Dissolved oxygen profiles were obtained from the unsampled modified wells as described in Section 2.4.1. All MWP fermentations were performed in duplicate.

**Table 2.4** Specification of factors and settings investigated during the scale-up of microwell plate fermentations. Pre-I and post-I are abbreviations for pre-induction and post-induction, respectively. <sup>a,b</sup>The MWP fermentation conditions described here have been adopted from experiments N32 and N36, respectively, from the optimisation experiments (Section 4.4).

Factor	Units	$k_La = 247 \text{ h}^{-1}$			$k_La = 55 \text{ h}^{-1}$	
		Plate P3 <sup>a</sup>	7.5 l	75 l	Plate P3 <sup>b</sup>	7.5 l
Liquid fill volume	l	$3 \times 10^{-3}$	5	45	$3 \times 10^{-3}$	5
Pre-I temperature	°C	27	27	27	27	27
Pre-I shaking speed / agitation rate	rpm	500	721	645	300	242
Pre-I period	h	7	7	7	4	4
Inducer concentration	µM	500	500	500	500	500
Post-I temperature	°C	37	37	37	30	30
Post-I shaking speed / agitation rate	rpm	500	721	645	300	242
Post-I period	h	18	18	18	12	12

### 2.5.2 STR fermentations at equivalent $k_La$ conditions

These fermentations were performed within both the 7.5 l and the 75 l STRs. All probes were manufactured by Mettler-Toledo Ltd. (Leicester, U.K.). The pH probes on both vessels were calibrated outside the vessel using standard buffers at pH 4.01 and 7.00 supplied by Mettler-Toledo Ltd. Approximate calibrations for the DOT probes were made outside the vessel using gaseous air and nitrogen for the 100 % and 0 % air saturation calibrations, respectively. Once all calibrations were made, the probes were placed in their appropriate locations within each vessel. The vessels were then filled with TB medium and sterilised as described in Section 2.4.3. After the sterilisation procedure, the phosphate components were added aseptically to each vessel and an aliquot of kanamycin was added to each vessel through a 0.2 µm filter (Sartorius Ltd, Epsom, U.K.) to a final concentration of 30 mg.l<sup>-1</sup>.

At this stage the calibration of the pH and DOT probes were rechecked and recalibrated as appropriate. A 10 ml aliquot of media was sampled from each STR using aseptic techniques, and the pH was measured on a Mettler-Toledo MP220 pH meter. This measurement was used to adjust the pH readout from each STR accordingly. The 100 % air saturation calibration was checked after sparging air into each vessel at a rate of 1 vvm for 20 min at 27 °C using an agitation speed of 200 rpm.

An overnight *E. coli* culture, prepared as described in Section 2.2.3, was used to inoculate each vessel to an initial OD<sub>600</sub> of 0.1. At this point a sample of inoculated media was taken from the 75 l vessel and used for the parallel MWP fermentations as described in Section 2.5.1. The air supply was set to 1 vvm and cultures were grown under the conditions described in Table 2.4. The pH was left uncontrolled in accordance with the MWP fermentation operation. An aliquot of IPTG solution was added to each vessel to a final concentration of 500 µM at the appropriate time point through a 0.2 µm filter (Sartorius Ltd, Epsom, U.K.).

Feedback loops were set up on both vessels to monitor and control foam levels by way of automatic additions of polypropylene glycol MW = 2025 (VWR International Ltd, Leicestershire, U.K.). Temperature, DOT, pH and agitation speed were continuously recorded on both vessels using BioComand Plus software version 3.28 (New Brunswick Scientific Co.) on the 7.5 l STR and MTX-Propack software (formerly Acquisition Systems, Berkshire, UK) on the 75 l STR. On-line exit gas composition was measured using a Prima 600 mass spectrometer (formerly VG Gas Analysis Ltd., Cheshire, U.K.).

Samples of 50 µl fermentation broth were taken at regular intervals for OD<sub>600</sub> measurements as described in Section 2.6.1. Samples of 1.2 ml fermentation broth were also taken at regular intervals for glycerol and FFL measurements. These samples were first rapidly clarified by centrifugation at 14,000 rpm for 10 min at 4 °C using an Eppendorf 5417R centrifuge (Cambridge, UK). The supernatant was assayed for glycerol, as described in Section 0, and the cell

pellets were stored at -80 °C for subsequent FFL assaying, as described in Section 2.6.3.

### 2.5.3 Statistical analyses of fermentation kinetic profiles

Microcal Origin version 6.0 (Northampton, MA, USA) was used in the modelling of fermentation profiles. To evaluate the effect of scale on fermentation performance, response profiles of biomass growth, glycerol utilisation and soluble protein yield were compared between the MWP and 7.5 l scales and also between the MWP and 75 l scales. Assuming no difference (the null hypothesis), one model would adequately describe both sets of data for a given response, and any observed differences would be purely due to chance. However, if in fact fermentation scale did have an effect on performance (the alternative hypothesis), then each set of data for a given response would be discrete and separate models would be required to fit the two data sets. The basis for this approach is described further by Motulsky and Christopoulos (2003).

Both hypotheses were tested simultaneously using the models described in Table 2.5. The degrees of freedom (DF) from separate model fits were first summed and labelled  $DF_{sep}$ . The residual sum of squares (RSS) from each model were also summed and labelled  $RSS_{sep}$ . Then, both data sets were combined, the same model type was refitted and new values for DF and RSS, labelled  $DF_{comb}$  and  $RSS_{comb}$ , respectively, were determined.  $RSS_{sep}$  is expected to be smaller than  $RSS_{comb}$ , simply because more predictor variables are available when separate models are used. The key question is whether or not this difference lies within the expected range for normal experimental error, assuming the null hypothesis is true. To answer this, a statistical test was performed, the first part of which involved calculating the  $F$  ratio:

$$F = \frac{(RSS_{comb} - RSS_{sep}) / RSS_{sep}}{(DF_{comb} - DF_{sep}) / DF_{sep}} \quad (2.6)$$

The corresponding  $p$  value was then calculated. The  $p$  value is the probability of achieving a difference (in RSS values) as large as or larger than observed if the null hypothesis (no effect of scale) was true. If the  $p$  value were above a certain threshold, typically 0.05, it would be possible to conclude that the observed difference is due purely to chance and hence the effect of scale is insignificant.

**Table 2.5** Description of functions used to model each fermentation response profile, where  $y$  is the fermentation response level at time  $t$ ,  $y_0$  is the initial response level and  $a$ ,  $b$  and  $c$  are arbitrary constants.

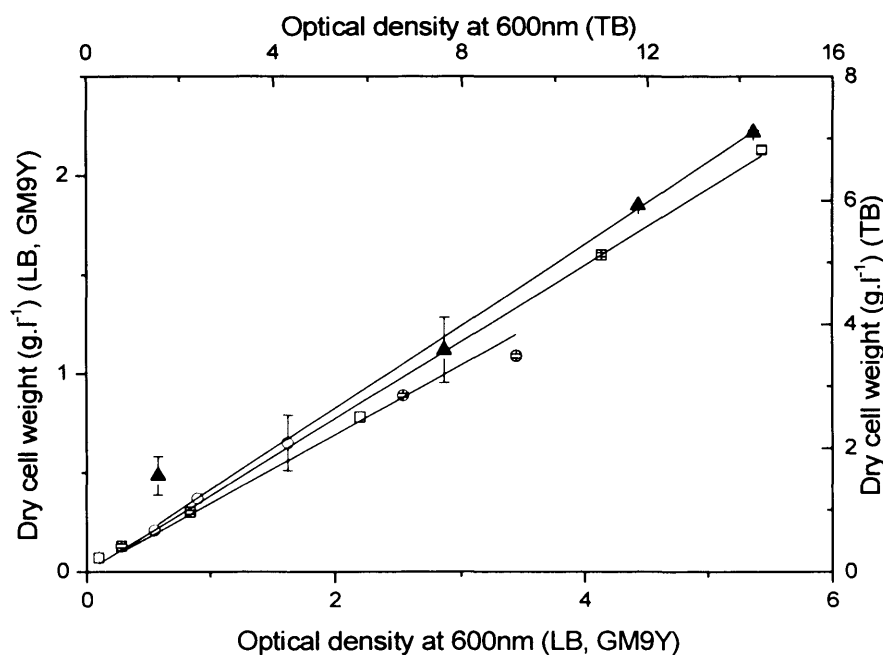
Response name	Model type	Model equation
[Biomass] (growth phase only)	Logistic	$\ln\left(\frac{y}{y_0}\right) = \frac{a}{1 + e^{-b(t-c)}}$
[Glycerol]	One site competition	$y = \frac{a}{1 + 10^{(t-\log b)}}$
[Soluble protein]	Extreme peak function	$y = y_0 + c \exp(-\exp(-d) - d + 1)$ where $d = \frac{(t-a)}{b}$

## 2.6 Analytical procedures

### 2.6.1 Quantification of biomass concentration

For the MWP fermentations, OD<sub>600</sub> measurements of broth samples were made using a Spectronic Helios Alpha spectrophotometer (Thermo Electron Corporation, Hemel Hempstead, U.K.). For the STR fermentations, OD<sub>600</sub> measurements were made using a Jenway 6400 spectrophotometer (Essex, U.K.). For each measurement, 50 µl of fermentation broth was first diluted in 1000 µl of corresponding media (1:21 dilution) before measuring OD<sub>600</sub>.

A calibration curve (Figure 2.3) for biomass concentration versus  $OD_{600}$  was generated from separate shake-flask fermentations of *E. coli* in each media preparation (LB, TB and GM9Y media). Fermentation samples were collected in triplicate at a minimum of four different time points and their  $OD_{600}$  recorded. These samples were filtered through a 0.2  $\mu\text{m}$ , 25 mm diameter filter (Whatman PLC, Middlesex, U.K.) and the filter was dried to constant weight using a HG53 Halogen Moisture Analyser (Mettler-Toledo Ltd., Leicester, U.K.). The dry cell weight was calculated from the increase in weight from the original empty filter. A calibration curve was obtained and used to convert all  $OD_{600}$  measurements into dry cell weight per litre broth ( $\text{g.l}^{-1}$ ) measurements.

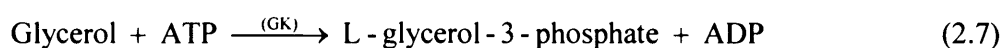


**Figure 2.3** Calibration curve for optical density and dry cell measurements of *E. coli* grown in LB ( $\circ$ ), GM9Y ( $\square$ ) and TB ( $\blacktriangle$ ) medium. The following lines of best-fit are obtained:  $y = 0.342x$ ,  $R^2 = 0.967$  (LB),  $y = 0.387x$ ,  $R^2 = 0.998$  (GM9Y),  $y = 0.497x$ ,  $R^2 = 0.964$  (TB).

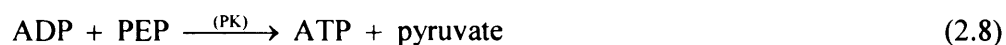
### 2.6.2 Quantification of glycerol concentration

Glycerol concentrations were measured using a commercially available kit (Megazyme International Ireland Ltd. Co. Wicklow, Ireland).

The general reaction principles of the assay are as follows. Glycerol is phosphorylated by adenosine-5'-triphosphate (ATP) to L-glycerol-3-phosphate in the reaction catalysed by glycerokinase (GK) (Equation 2.7):



The adenosine-5'-diphosphate (ADP) formed in the reaction is reconverted by phosphoenolpyruvate (PEP) with the aid of pyruvate kinase (PK) into ATP with the formation of pyruvate (Equation 2.8):



In the presence of the enzyme L-lactate dehydrogenase (L-LDH), pyruvate is reduced to L-lactate by reduced nicotinamide-adenine dinucleotide (NADH) with the production  $\text{NAD}^+$  (Equation 2.9):



The amount of  $\text{NAD}^+$  formed in the above reaction pathway is stoichiometric with the amount of glycerol. The NADH consumption is measured directly by the decrease in absorbance at 340 nm (Anon., 2005).

The change in absorbance ( $\Delta A_{\text{glycerol}}$ ) was then used to calculate the glycerol concentration according to the following formula:

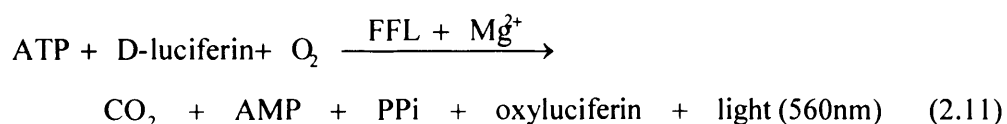
$$c = \frac{V \times MW}{\epsilon \times d \times v} \times \Delta A_{\text{glycerol}} \quad (2.10)$$



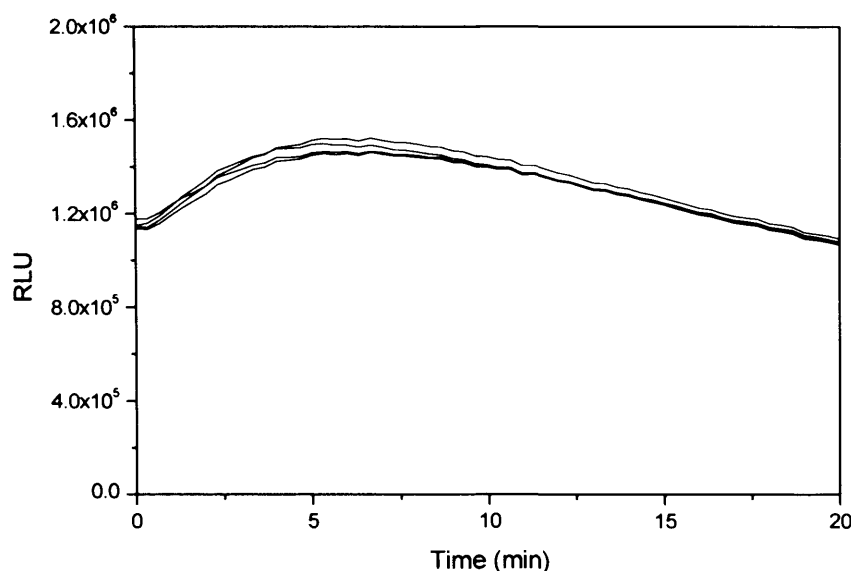
where  $c$  is the glycerol concentration ( $\text{g.l}^{-1}$ ),  $V$  = final volume (ml),  $MW$  = molecular weight of glycerol ( $\text{g.mol}^{-1}$ ),  $\epsilon$  = molar extinction coefficient =  $6300$  ( $\text{l.mol}^{-1}.\text{cm}^{-1}$ ),  $d$  = light path (cm),  $v$  = sample volume (ml).

### 2.6.3 *Quantification of relative soluble firefly luciferase activity*

The assay for FFL enables a quantitative value of expressed soluble protein to be derived and requires no prior purification of the enzyme. Cell pellets were thawed on ice and resuspended in  $500 \mu\text{l}$  BugBuster (Abgene, Epsom, UK). An aliquot of  $50 \mu\text{l}$  DNase I (Burgess Hill, UK) at a concentration of  $1 \text{ mg.ml}^{-1}$  was added and the samples were incubated for 20 min on ice. Lysates were clarified by centrifugation at  $14,000 \text{ rpm}$  for 10 min. A  $5 \mu\text{l}$  aliquot of lysate was mixed with  $200 \mu\text{l}$  of Luciferase Assay Buffer (LAB) containing  $10 \text{ mM MgSO}_4$ ,  $10 \text{ mM Tris pH } 7.8$ ,  $0.5 \text{ mM ATP}$ ,  $0.47 \text{ mM luciferin}$ ,  $0.27 \text{ mM co-enzyme A}$  and  $31.5 \text{ mM dithiothreitol}$ , in a Nalgene 96-well plate (VWR, Loughborough, UK). The activity of luciferase was measured using a SpectraMAX Gemini XS luminometer (Molecular Devices, Winnersh, UK) according to the reaction scheme shown below:



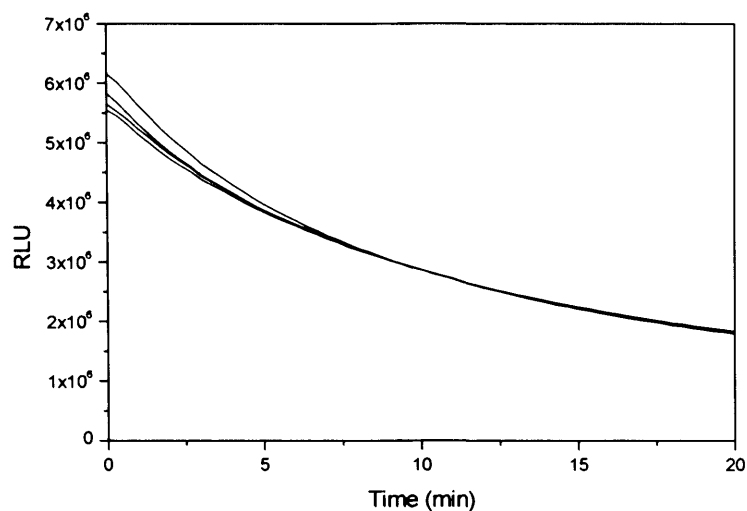
The luminescence profile of each sample was recorded at room temperature over a period of 20 min. An example of a typical luminescence profile, where measurements of Relative Light Units (RLU) are made, is shown in Figure 2.4.



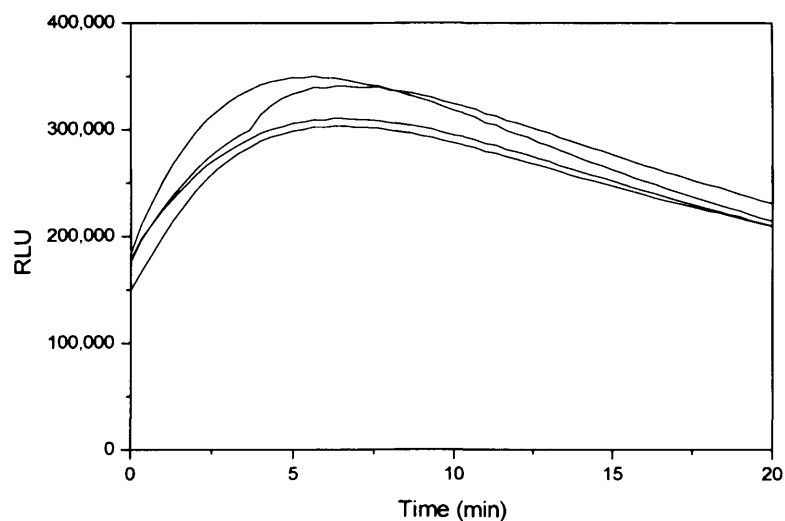
**Figure 2.4** Example of standard luminescence profiles from samples of the replicate MWP fermentations described in Section 2.3.1, taken at identical time points. Measurements made as described in Section 2.6.3.

For samples with particularly high soluble protein yields, the reaction shown in Equation 2.11 proceeded too rapidly and peak luminescence often occurred before measurements could be made as illustrated in Figure 2.5 (a). In order to obtain a usable luminescence profile from these samples, lysates were first diluted 20-fold in 100 mM Tris pH 7.8 prior to the addition of LAB. This resulted in a luminescence profile that enabled maximum peak heights to be obtained within the experimental time frame (Figure 2.5 (b)). The peak height of each luminescence profile is proportional to the luciferase activity (Bronstein *et al.*, 1994) and so for each luminescence profile, the mean of the four largest luminescence values was recorded. This value was then divided by the original microwell liquid fill volume to give the soluble protein yield in units of RLU.mL<sup>-1</sup>.

(a)



(b)



**Figure 2.5** Typical luminescence profiles for the replicate samples within each MWP fermentation. For samples with particularly high soluble protein yields, the reaction proceeded very fast and the peak was often missed (a). In these scenarios, the lysates were first diluted 20-fold in 100 mM Tris pH 7.8 prior to the addition of LAB which resulted in more standard profiles (b).

#### **2.6.4 Analytical software**

MODDE software version 7.0 (Umetrics, Windsor, UK) was used for experimental design of the MWP fermentations (Sections 2.3.2 to 2.3.4 inclusive) and the subsequent regression analysis and process modelling. Microcal Origin version 6.0 (Northampton, MA, USA) was used in the statistical data analysis of fermentation profiles (Section 2.5.3). It was also used for figure design. Microsoft Excel 2002 was used to solve Equation 2.4 for the calculation of all  $k_La$  values and also to calculate  $p$  values during the statistical analyses of fermentation profiles (Section 2.5.3).

## **3 A framework for the rapid optimisation of soluble protein expression in *E. coli***

### **3.1 Aims and objectives**

As described in Section 1.5 the overall aim of this project is to establish a generic framework which underpins the generation of large quantities of soluble recombinant protein in *E. coli* in a manner which is both rapid and cost-effective. To achieve this goal, various tools and technologies will need to be combined and applied in a novel manner. The aim of this chapter is to provide the conceptual framework to achieving this goal. Specific objectives are:

- to determine the key components of the framework and its overall structure (Section 3.2);
- to explore the practical issues of implementing the framework components within this project (Sections 3.3).

### **3.2 Framework design**

#### **3.2.1 Introduction**

The challenge of rapidly generating large quantities of protein may be split into two principal stages: (1) small scale protein expression characterisation followed by (2) scale-up of optimised protein expression. A wide range of development tools is available to aid in achieving these goals. In the following sections, the most appropriate tools are selected and a framework for their implementation is proposed.

#### **3.2.2 Component specification**

The first stage of the framework involves the characterisation of protein expression. A wide range of variables affects heterologous protein expression

levels, as described in Section 1.2.3, and numerous experiments are normally required to characterise such a complicated process. Design of Experiments is able to greatly reduce the number of these characterisation experiments (Section 1.3), whilst providing a high level of process understanding which also aids in subsequent scale-up. This high degree of characterisation coupled with the significant savings in time and costs places DoE at the heart of the project framework.

In order to generate the characterisation data in a rapid and cost-effective manner, experiments should be performed at small scale as discussed in Section 1.2.4. Microscale experimentation complements DoE in that material requirements are reduced and experimental throughput is increased through parallel operation and potential automation. A framework component which includes microscale experimentation would thus prove highly beneficial.

The second stage of the framework involves the scale-up of protein expression in order to access larger quantities of protein. The final scale of fermentation required is dependent on both the demands of the specific drug-discovery campaign and the expression characteristics of the target protein. The second stage of the framework will thus involve the reproduction of microscale protein expression levels over a range of scales. Separate strategies for laboratory scale and pilot scale expression will form the components of this stage of the framework.

In order to minimise the time and costs of scale-up, these strategies should incorporate the information generated from the DoE / microscale stage of experimentation. The optimal fermentation conditions determined from small-scale characterisation experiments, for example, are likely to mirror those at larger scales (Hunt, 2005). These optimal conditions should provide some indication of the key process scale-up parameters and the overall scale-up strategy should also take account of this.

### **3.2.3 Framework structure**

Having specified the key components, the next design step is to arrange these components within an optimal framework configuration such as that illustrated in Figure 3.1. The target protein and expression system proceed through to the first stage of the framework: protein expression characterisation. Here, suitable factors and responses are specified and experimental designs are generated through DoE. These fermentations are performed within microwell plates (MWP) and the data generated is used to guide subsequent experimental designs. This cycle continues until the optimum expression conditions have been identified.

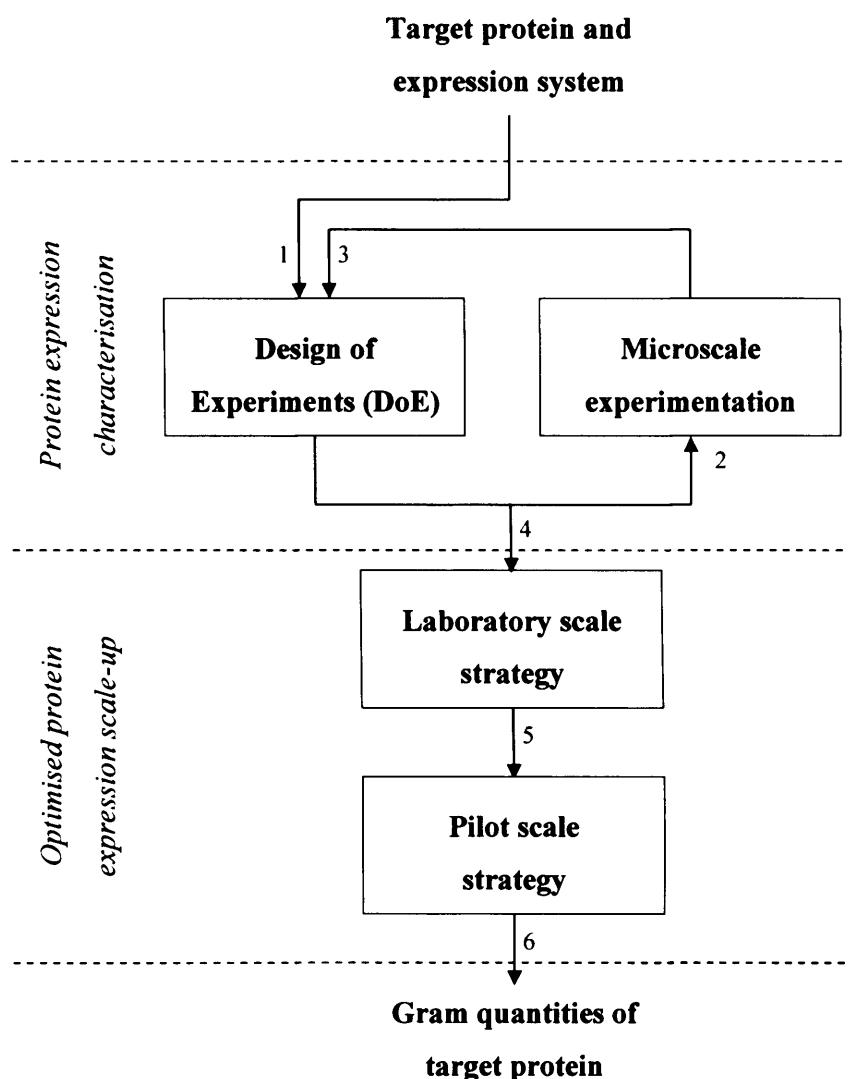
The target protein and expression system then proceed through to the second and final stage of the framework: scale-up of optimised protein expression levels. The process information gathered thus far, such as optimal expression conditions and potential key scale-up parameters is used to design an appropriate scale-up strategy and this strategy is implemented up to the laboratory scale. Then, the successful strategy is refined in light of any new information emerging from laboratory scale fermentations and applied to reproduce MWP protein expression levels at the pilot scale.

## **3.3 Framework implementation**

### **3.3.1 Protein expression characterisation**

#### *3.3.1.1 Background*

The experimental designs within this framework are greatly influenced by the properties of microscale experimentation. The implementation aspects of both components will, therefore, be considered simultaneously.



**Figure 3.1** Overview of the proposed framework for large-scale optimised protein expression. The numbered arrows represent information streams. The specification of factors and responses is determined by the particular target protein and expression system (1) and DoE is used to design the subsequent microscale experiments (2). Several rounds of expression studies are required to determine the optimal process conditions in a time and cost efficient manner. The information gathered from each round is used to generate subsequent experimental designs (3) until the optimal conditions are found. The process knowledge which emerges from the DoE / microscale experiments should provide insight into larger (laboratory) scale performance and thus it should be incorporated into the appropriate scale-up strategy (4). Similarly, the knowledge gained from the laboratory scale strategy should inform the subsequent pilot scale strategy (5), thereby enabling the production of gram quantities of target protein (6).

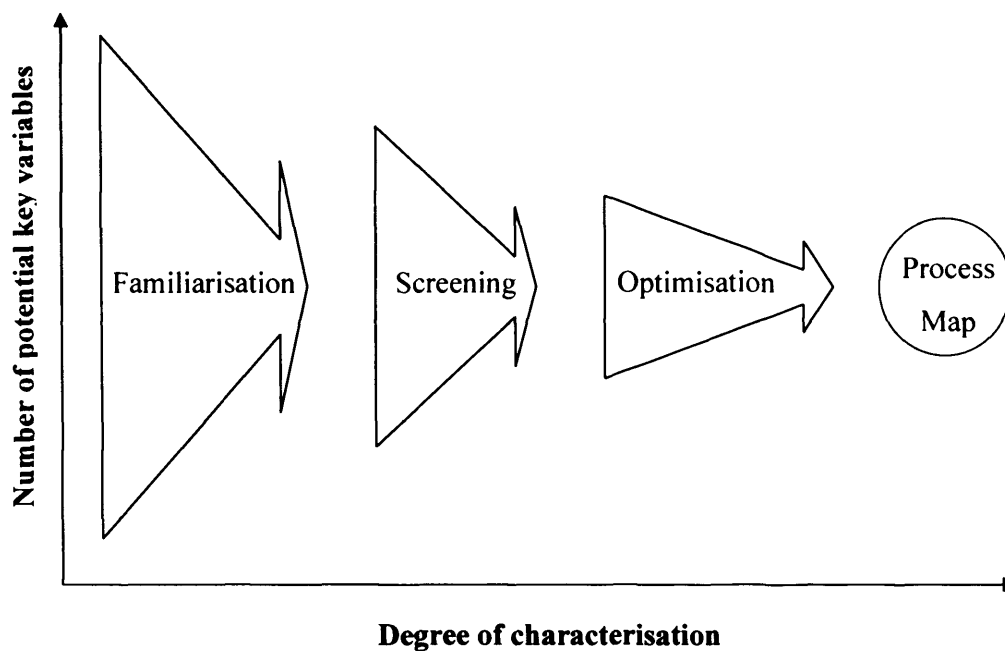


The stages involved in DoE were first reviewed in Section 1.3.2 and included familiarisation, screening, finding the optimal region, optimisation, robustness testing and mechanistic modelling. Not all stages are relevant to the objectives of this project, however. Mechanistic modelling, for example, is usually only attempted when there is a need to establish a theoretical model within a field and this lies beyond the scope of the current work. Similarly, robustness testing is mainly performed when reproducible product quality is of key importance. Since the expressed reporter protein has no application as a product, robustness testing is unnecessary. Finally, finding the optimal region does not constitute a primary DoE objective and it will thus be considered alongside the screening stage.

Overall, the three primary experimental objectives of this work are familiarisation, screening and optimisation. Figure 3.2 illustrates the implementation sequence and characteristics of these stages towards the goal of obtaining a process map. As each stage is completed, the number of potentially significant variables is reduced whilst the degree of process characterisation is increased. The specific implementation aspects of each stage are discussed in the following sections.

#### 3.3.1.2 *Familiarisation*

Familiarisation serves as perhaps the most important stage within the overall framework. As the name suggests, the purpose of this stage is to gain familiarity with the process under investigation. The accuracy of information generated here will affect the number of experiments required to locate and characterise the optimal region, the reproducibility of response data and the success of the subsequent scale-up strategy. Thorough planning of the familiarisation experiments is thus crucial.



**Figure 3.2** Implementation sequence of the stages of DoE applied in this work (arrows) towards the goal of obtaining a reliable process map. As each stage is completed, the number of potentially significant variables is reduced whilst the degree of process characterisation is increased.

The first step involves identifying all variables which may affect the process under investigation and this may be accomplished using insight gained from published literature, past experience or simple experimental designs. Some factors will have little effect on the system but may still cause an undesired spread around the ideal result. These factors typically include ambient temperature, humidity and variability in raw material composition. By developing techniques for controlling undesired sources of variation, it should be possible to establish a fine degree of process control and the accuracy of information gained from the process should thus be improved. In contrast, factors which are thought to have the potential to significantly affect the system, such as those described in Section 1.2.3, should be targeted for investigation through DoE.

A key requirement here is that factors are chosen to provide insight into larger scale performance and thus that the microwell experiments are performed in a defined engineering environment (Fernandes and Cabral, 2006). Initial studies have suggested that scaling-up on the basis of a constant oxygen mass transfer coefficient,  $k_La$ , appears to be the most appropriate approach when dealing with microorganisms growing under aerobic conditions (Ferreira-Torres *et al.*, 2005; Micheletti *et al.*, 2006). Factors which are thought to influence  $k_La$ , such as MWP shaking speed and fill volume should, therefore, also be included for investigation. Their optimisation should enhance culture conditions and provide insight into scale-up criteria for optimal fermentation performance.

The next step involves choosing responses that are relevant to the goals of the investigation and developing the appropriate response measurement techniques. Several responses will be considered throughout this project including cell growth and soluble protein yield. The corresponding assays will need to be compatible with both DoE and microscale processing techniques in that they should be rapid, sensitive, and quantitative, involving minimal processing to further reduce costs and facilitate potential automation.

Overall, the parameter assumptions and experimental methods developed thus far should be improved through experimentation before progressing through to the next stage of DoE, screening.

### 3.3.1.3 Screening

The key purpose of screening is to identify the most influential factors and provide an estimate for their appropriate ranges. Here, factors are commonly investigated over only two levels and so a relatively low number of experiments is required (Eriksson *et al.*, 2000). For soluble recombinant protein production, the appropriate screening design should be capable of handling not only a large number of factors, but also a mixture of qualitative and quantitative variables.

It is usually possible to gauge if a screening design has captured the optimal settings, but the results do not lend themselves well to extrapolation. Therefore, if the optimal region appears to lie outside of the screening window, it may be necessary to perform an additional number of extra scouting experiments prior to optimisation.

#### *3.3.1.4 Optimisation and process mapping*

The goal of optimisation is to reveal the nature of the mathematical relationship between factors and responses and thus produce an accurate map or model of the system under investigation. The corresponding designs vary factors over 3 to 5 levels and so involve more experiments per factor than screening designs. The model can then be used to accurately determine the optimal factor settings, assuming they lie within the optimisation window.

#### *3.3.2 Scale-up strategy: laboratory scale*

A strategy for translating optimal DoE results from MWP fermentations to a laboratory scale STR is then needed and is underpinned by the ability to predictively scale-up microwell results. This need is complicated by stark differences between microwell plates and stirred tank reactors such as geometry, the method of aeration and agitation and the effects of surface tension. Scaling-up on the basis of constant  $k_La$  has the potential of overcoming the aforementioned complications as it corresponds to an overall measure of agitation and aeration.

Regardless of the basis for scale-up, the key scale-up parameter(s) will first have to be characterised over a range of conditions at both the MWP scale and laboratory scale STR. It should then be possible to translate MWP fermentation performance to the larger scale. Several sets of fermentation conditions should be investigated to provide an estimate for the range of conditions over which successful scale-up can be achieved.

### **3.3.3 Scale-up strategy: pilot scale**

The scale-up strategy developed at the laboratory scale should also provide the foundation for an effective pilot scale strategy. For example, the techniques developed at the laboratory scale for measuring the key scale-up parameter(s) should also be applicable at this scale of operation. The range of scale-up conditions over which successful scale-up can be achieved may also be similar.

At this scale of operation, however, many minor sources of process variation may arise. These include method of inoculum preparation, method of sterilisation, quality of media components, etc. In combination these may cause a significant deviation from predicted performance. These variables should, therefore, be taken into account when designing the appropriate scale-up strategy. By this stage it should be possible to produce large amounts of target protein under optimal conditions.

## **3.4 Summary**

The aim of this chapter was to provide a conceptual framework for generating large quantities of soluble recombinant protein in *E. coli* in a manner which is both rapid and cost-effective. The key components and overall structure of the proposed framework were specified in Section 3.2 and in Section 3.3 the practical issues of framework implementation were explored. In the following chapters, the practical details of the framework are determined through experimentation. Chapter 4 will address protein expression characterisation and Chapters 5 and 6 will tackle the translation of microscale fermentation performance to the laboratory scale and pilot scale, respectively.

## 4 Optimisation of soluble protein expression at the microwell scale<sup>†</sup>

### 4.1 Aims and objectives

The first stage of the proposed framework (Section 3.3.1) involves the optimisation of soluble protein expression at the microwell scale. This is covered in this chapter which addresses the application of DoE and the performance of experiments at the microwell scale. The specific objectives of this chapter are:

- to develop appropriate experimental techniques and identify potential variables influencing protein expression through initial familiarisation experiments (Section 4.2);
- to screen the selected variables to identify those which produce the most significant effects on the expression system and estimate the factor ranges which are likely to encompass the optimal settings (Section 4.3);
- to confirm the identify of the important factors and model their effect on soluble protein expression (Section 4.4);
- to improve the predictive ability of the resultant model through statistical analysis, to pinpoint the optimal settings of the important factors and verify the model's practical capability (Section 4.5).

### 4.2 Experimental familiarisation

#### 4.2.1 *Specification of variables*

The first objective of the familiarisation experiments was to identify the variables influencing soluble protein expression and their appropriate ranges for the system under investigation. Ten factors were initially chosen for investigation and these are described in Table 4.1. This selection of factors is generic and could be

<sup>†</sup>The majority of the results presented in this chapter have been published as: Islam,R.S., Tisi,D., Levy,M.S., and Lye,G.J. (2007) Framework for the Rapid Optimization of Soluble Protein Expression in *Escherichia coli* Combining Microscale Experiments and Statistical Experimental Design. *Biotechnology Progress*. **23**, 785-793.

---

applied to the study of soluble protein production in microwells for any inducible *E. coli* expression system. Medium types were chosen from commonly used complex media and plate formats were chosen such that a range of geometries was present. The chosen screening temperatures, IPTG concentrations and post-induction periods were adopted from a study by Urban *et al.* (2003) who also investigated recombinant protein expression using the pET vector and an identical strain of *E. coli*. The IPTG concentrations used, 50-1000  $\mu\text{M}$ , are also in line with those used in other *E. coli* expression systems (Cao, 2006; Swalley *et al.*, 2006). Previous experiments near centre-point conditions had shown the mid-exponential growth phase to occur at  $\sim 4$  h and so the boundaries for the pre-induction growth period were positioned symmetrically around 4 h. In contrast to other DoE studies, shaking speed and liquid fill volume were included as factors since these strongly influence the oxygen transfer rate into the wells (Hermann *et al.*, 2003). The upper shaking speed was limited by the speed of the HiGro incubator-shaker and the upper liquid fill volume was chosen such that excessive splashing was avoided within the wells at the highest shaking speed.

#### **4.2.2 Development of response measurement techniques**

The second objective was to identify suitable responses and develop the appropriate measurement techniques. An obvious response was that of biomass concentration, for which optical density measurements would provide a good indication and which have the potential to be made rapidly in a high-throughput manner. These observations would be made during the optimisation experiments at specific time points throughout the course of each microwell plate fermentation. The time points immediately following the growth and induction periods were deemed most appropriate for these initial measurements.

Another key measurement was that of soluble protein yield. Although general assay procedures for FFL are described in the literature, it was necessary to develop an assay which met the specific requirements of this work. Firstly, all assay procedures were adapted to the MWP format. Following this, the relative quantities of assay reagents were adjusted to suit the range of expression levels

particular to these fermentation conditions. Finally, it was necessary to again develop assay procedures which were compatible with the demands of high-throughput automated processing (Lye *et al.*, 2003). The final version of the assay procedure used throughout this study is described in Section 2.6.3.

**Table 4.1** Specification of factors and settings investigated within familiarisation, screening and optimisation experiments. Pre-I and post-I are abbreviations for pre-induction and post-induction, respectively. Plate P1 had 48 rectangular wells with flat bases, plate P2 had 24 square wells with round bases and plate P3 had 24 square wells with pyramidal bases. <sup>a</sup>The familiarisation experiments had a lower shaking speed of 50rpm. <sup>b</sup>An upper liquid fill volume of 2 ml was used in conjunction with the 48-well plate to prevent splashing.

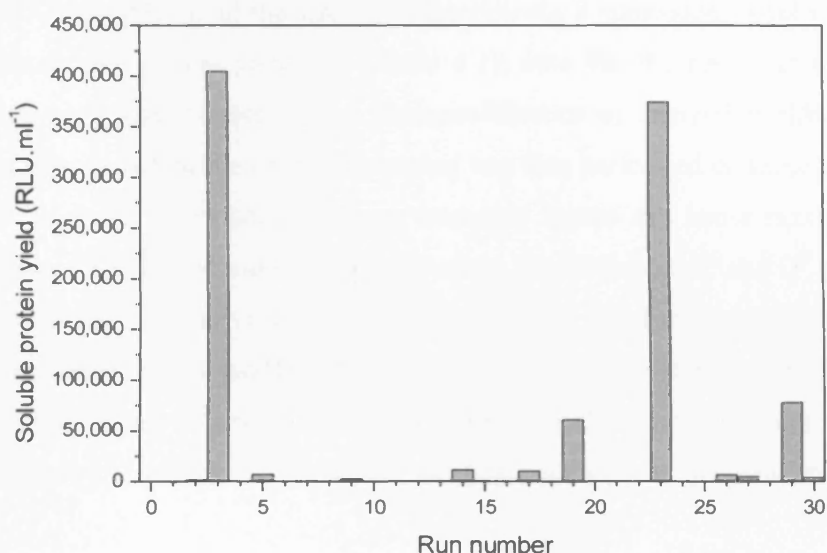
Factor name	Abbr.	Units	Experimental settings	
			Familiarisation / Screening	Optimisation
Growth medium	x <sub>1</sub>	-	LB, TB, GM9Y	TB
Plate geometry	x <sub>2</sub>	-	P1, P2, P3	P3
Liquid fill volume	x <sub>3</sub>	ml	1, 2 <sup>a</sup> , 3	1, 2, 3
Pre-I temperature	x <sub>4</sub>	°C	17, 37	27
Pre-I shaking speed	x <sub>5</sub>	rpm	100 <sup>b</sup> , 500	100, 300, 500
Pre-I period	x <sub>6</sub>	h	2, 6	1, 4, 7
Inducer concentration	x <sub>7</sub>	µM	50, 1000	500
Post-I temperature	x <sub>8</sub>	°C	17, 37	23, 30, 37
Post-I shaking speed	x <sub>9</sub>	rpm	100 <sup>b</sup> , 500	100, 300, 500
Post-I period	x <sub>10</sub>	h	3, 15	6, 12, 18

#### 4.2.3 Method development through experimentation

In order to evaluate the cell culture and analytical techniques developed methods thus far, a trial set of 30 DoE experiments was performed from a screening design according to the factor settings described in Table 4.1. The mean soluble protein yield results from these experiments are shown in Figure 4.1. Approximately 40 % of the experiments resulted in protein yields undetectable by the assay with the average soluble protein yield per fermentation being 32,230



RLU.ml<sup>-1</sup>. For comparison, shake flask fermentations were also performed under standard expression conditions as defined in Section 2.3.5. These fermentations were run in duplicate and a mean soluble protein yield of 552,000 RLU.ml<sup>-1</sup> ( $\sigma = 12.9\%$ ) was obtained.



**Figure 4.1** Column chart illustrating the variation of soluble protein yield from MWP familiarisation experiments. Experiments performed as described in Section 2.3.1.

Due to the low protein yield measurements and the high incidence of zero-value measurements, subsequent regression analysis was severely impeded. However, the goal of familiarisation is to gain some initial experimental experience with the system under investigation. Consequently, several changes were made to the experimental procedure, two of which were key. Firstly, the volume of inoculum was increased to raise the starting OD<sub>600</sub> from 0.001 to 0.1. Secondly, the lower shaking speed was raised to 100rpm, as shaking at 50rpm was observed to produce virtually no mixing in the wells.

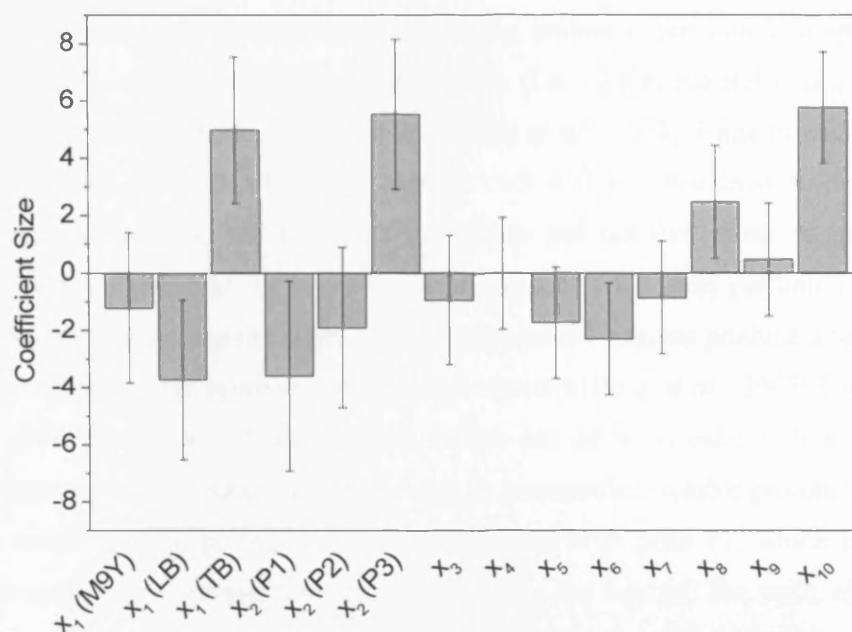
### 4.3 Screening experiments

Building upon the knowledge and experience gained from the familiarisation stage, the objective of the screening experiments was to rapidly explore the influence of a wide range of factors on soluble protein yield. Detailed information on biomass growth and protein expression kinetics was sacrificed at this stage in order to rapidly identify the main factors influencing soluble protein yield. On completion of the screening experiments, a regression model was fitted to the soluble protein yield data (Table 4.2); note that the particular screening design used here did not support the quantification of interaction effects, only main effects. A fourth root transformation was then performed on these values in order to render their distribution approximately normal and hence maximise the  $Q^2$  value for the fitted model. This subsequent model yielded  $R^2$  and  $Q^2$  values of 0.86 and 0.63 respectively which indicated that 86 % of the response variation was explained by the model and 63 % of the response variation was predicted by the model. These  $R^2$  and  $Q^2$  values suggested that the resulting model provided information of sufficient accuracy for creating a subsequent optimisation design.

The significance of each screening factor on soluble protein yield is illustrated in the coefficient plot shown in Figure 4.2. It can be clearly observed that, of the qualitative factors, LB medium and plate P1 contributed to the lowest soluble protein yields. In contrast, TB medium and plate P3 resulted in the highest soluble protein yields. There are several reasons why growth in TB could have resulted in the highest soluble protein yields. This medium is not only rich in nutrients such as yeast extract and tryptone, it also contains a defined carbon source, glycerol. Coupled to this, TB contains significant amounts of  $KH_2PO_4$  and  $K_2HPO_4$ , ( $2.31 \text{ g.l}^{-1}$  and  $12.54 \text{ g.l}^{-1}$  respectively) which provide some level of buffering capacity against any disturbance in pH within the medium during fermentation. In contrast, LB medium contains no defined carbon source and no phosphate components and has almost five times less yeast extract than TB.

**Table 4.2** D-optimal screening design matrix for the initial assessment of soluble protein yield. Quantitative variables are coded at low (-1), medium (0) and high (+1) levels according to the factor levels specified in Table 4.1.

Observations	x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	x <sub>4</sub>	x <sub>5</sub>	x <sub>6</sub>	x <sub>7</sub>	x <sub>8</sub>	x <sub>9</sub>	x <sub>10</sub>	Mean soluble protein yield (RLU.ml <sup>-1</sup> )
N1	M9Y	P3	-1	-1	1	1	1	-1	-1	-1	2,422
N2	M9Y	P3	-1	1	1	1	-1	-1	-1	1	228,491
N3	M9Y	P3	-1	-1	1	-1	1	1	1	-1	53,632
N4	LB	P3	-1	1	-1	1	-1	1	1	1	512,598
N5	TB	P3	-1	1	-1	-1	1	1	1	1	1,693,240
N6	M9Y	P3	1	-1	-1	-1	-1	-1	1	1	906,256
N7	LB	P3	1	-1	-1	-1	1	-1	-1	-1	440
N8	LB	P3	1	1	1	-1	-1	1	-1	-1	53,519
N9	TB	P3	1	1	-1	1	1	-1	1	-1	58,193
N10	TB	P3	1	-1	1	1	-1	1	1	1	329,012
N11	M9Y	P2	-1	1	-1	1	1	-1	1	-1	547
N12	LB	P2	-1	-1	-1	1	-1	-1	1	-1	1,366
N13	LB	P2	-1	-1	1	-1	1	-1	1	1	34,999
N14	TB	P2	-1	1	1	-1	-1	-1	-1	1	395,539
N15	TB	P2	-1	-1	-1	-1	-1	1	-1	-1	102,534
N16	M9Y	P2	1	-1	-1	1	-1	1	-1	-1	41,042
N17	M9Y	P2	1	1	1	1	-1	1	-1	1	171,085
N18	M9Y	P2	1	1	1	-1	1	-1	1	1	7,800
N19	LB	P2	1	1	1	1	1	1	1	-1	943
N20	TB	P2	1	-1	-1	-1	1	-1	-1	1	147,394
N21	M9Y	P1	-1	1	-1	-1	1	1	-1	-1	480
N22	M9Y	P1	-1	-1	-1	-1	-1	1	1	1	852,391
N23	LB	P1	-1	1	-1	1	-1	-1	-1	1	87,764
N24	LB	P1	-1	-1	1	1	1	1	-1	1	43,145
N25	TB	P1	-1	-1	-1	1	1	1	-1	1	321,696
N26	TB	P1	-1	1	1	-1	-1	-1	1	-1	10,359
N27	TB	P1	-1	-1	1	1	-1	-1	1	-1	11,020
N28	TB	P3	0	0	0	0	0	0	0	0	1,241,740
N29	TB	P3	0	0	0	0	0	0	0	0	1,220,350
N30	TB	P3	0	0	0	0	0	0	0	0	1,243,840



**Figure 4.2** Coefficient plot with confidence intervals (set at 95%) for soluble protein yield obtained from screening experiments. The size of each bar (or coefficient) represents the relative importance of that factor upon soluble protein expression. A positive coefficient for a quantitative factor indicates that an increase in that factor, relative to its centre-point level, would result in an increase in soluble protein yield. The opposite is true for negative coefficients. Qualitative factors, such as medium type, have no centre-point levels and so the coefficients displayed are relative to one another. Factors and experimental conditions as described in Table 4.1.

Intermediate levels of soluble protein expression were obtained with Glucose M9Y. Unlike LB medium, Glucose M9Y medium contains phosphate components and a defined carbon source in the form of glucose. This carbon source, however, provides limited benefits in comparison to glycerol which is present in TB medium. Glucose is a rapidly metabolised sugar which is associated with low recombinant protein yields (Stanbury and Whitaker, 1993). Unlike glycerol, it is also converted to acetate which is detrimental to the growth of *E. coli* (Jana and Deb, 2005; Luli and Strohl, 1990).

The plate geometry was also shown to cause diverse effects. This is especially significant since other publications examining protein expression in microwell formats have not fully investigated this factor (Cao, 2006; Nikerel *et al.*, 2005; Ren *et al.*, 2006; Swalley *et al.*, 2006; Urban *et al.*, 2003; Wang *et al.*, 2005; Zhang, 2006). Plate P3, which has 24 wells each with a square cross-section and a pyramidal bottom, had the most significant and positive effect on soluble protein yield. The large wells provide an increased surface area per unit volume for oxygen transfer and the edges along the pyramidal bottoms produce a baffling effect which further promotes mixing and aeration (Doig *et al.*, 2005; Duetz *et al.*, 2000; Lye *et al.*, 2003). Plate P2, which has 24 wells each with a square cross-section and a round bottom, resulted in intermediate soluble protein yields. The lowest soluble protein yields were observed with plate P1, which has 48 wells each with a rectangular cross-section and a flat bottom. The wells of plate P1 are small, each well with a cross-sectional area less than half those of the other plates. This would lead to a significant reduction in oxygen transfer capability due to the reduced air-liquid surface area within each well. The increased importance of surface tension effects as well as diameter decreases would exacerbate this problem.

The most significant quantitative factors (Figure 4.2) included the pre-induction (pre-I) and post-induction (post-I) periods and the post-I temperature. The post-I period had a positive coefficient, indicating that higher soluble protein yields are obtained at the higher setting (15 h). The highest soluble protein yields overall were obtained from the centre-point experiments where the post-I period was 9 h. This additional information suggested that the post-I period centre-point in the subsequent optimisation design should be positioned close to 9 h. The location of the window for post-I period would, therefore, be moved by +3 h for optimisation. The post-I temperature also had a positive coefficient. The upper limit was already at 37 °C, the optimum for *E. coli*, and so only the lower limit would be raised for optimisation experiments.

In many cases it is prudent to maximise the number of healthy cells prior to the addition of an inducer, which often retards cellular growth (Bentley *et al.*, 1991;

Donovan *et al.*, 1996; Lee, 1996; Peng *et al.*, 2004). Results from the screening experiments suggest differently, however. The negative coefficient for the pre-I period shows that, on average, higher soluble protein yields are obtained at the low setting of 2 h corresponding to lower cell densities prior to induction. For the optimisation experiments, the pre-I period range would be increased by 2 h overall symmetrically around the original centre-point.

While the coefficients for shaking speed, both pre- and post-induction, and also the liquid fill volume were not seen to have a significant impact on soluble protein yield (Figure 4.2), it was decided to retain them for subsequent optimisation experiments over their existing ranges. As previously mentioned in Section 4.2.1, these parameters are known to strongly influence  $k_{La}$  and hence the oxygen transfer rate into the wells (Hermann *et al.*, 2003). Since the optimisation experiments are expected to produce higher cell densities and protein yields, oxygen mass transfer limitations may appear and so it would be prudent to include these three factors in the next stage of the design procedure. If any of these parameters were to become significant under optimised conditions it would provide important insight for scale-up of the optimised process.

The remaining two quantitative factors of pre-I temperature and inducer concentration both appeared to be insignificant (Figure 4.2). Although some circumstances exist under which screening designs may return false-negative results for factor relevance the decision was made to dismiss these factors in favour of a reduction in the number of variables targeted for further investigation and optimisation. Since the centre-point experiments produced the highest soluble protein yields, both pre-I temperature and inducer concentration would be fixed near their centre-point levels for optimisation studies.

#### 4.4 Optimisation experiments

Having identified the key factors influencing soluble protein yield in Section 4.3, the aim of the optimisation experiments was to study a smaller number of factors,

over refined ranges, in more detail. The investigated factors are specified in Table 4.1. For the optimisation experiments the measured biomass dry cell weight pre- and post induction was also included as a response so as to provide extra insight into the relationship between cell growth and protein expression. The results for these experiments are summarised in Table 4.3. Here, it can be seen that there was a three-fold increase in the highest measured soluble protein yield ( $5.39 \times 10^6$  RLU.ml<sup>-1</sup>) compared to the maximum obtained in the initial screening experiments ( $1.69 \times 10^6$  RLU.ml<sup>-1</sup>).

On completion of the optimisation experiments, soluble protein yield values were transformed by taking the fourth root as described in Section 4.3. The initial optimisation model consisted of 27 terms (6 main effect terms, 6 square terms and 15 two-level interaction terms). Model terms that were not significant were then removed in a hierarchical manner; if a main effect term was removed then all higher order terms containing that factor, such as square terms and interaction terms were also removed. The remaining model coefficients for soluble protein yield are shown in Figure 4.3 and their associated statistics are given in Table 4.4. Confidence intervals, set at 95 %, show the intervals or limits within which there was a 95 % chance of finding the true values of the model coefficients. A *p* value indicates the probability that a given model coefficient has an insignificant effect on the response. It is evident, therefore, from Table 4.4 that all coefficients in the final optimisation model are significant at  $p < 0.05$ .

Only three of the original six factors were found to influence soluble protein yield significantly: the pre-I and post-I shaking speed and the post-I period. The square terms for both shaking speeds were significant indicating that the optimum levels for these factors lay within the ranges studied. The fact that both these shaking speeds have become significant factors in the optimisation experiments is a consequence of the higher biomass concentrations achieved (>8 g.l<sup>-1</sup>) and the further enhancement obtained in soluble protein yield. Both of these will increase the oxygen demand of the culture making oxygen transfer, and hence shaking speed, a key parameter.

**Table 4.3** CCF design matrix for the optimisation of soluble protein yield.  $DCW_i$  and  $DCW_f$  refer to dry cell weight measurements immediately prior to induction and immediately following the induction period respectively. All variables are coded at low (-1), medium (0) and high (+1) levels according to the factor levels specified in Table 4.1. Predicted mean soluble protein yields are calculated from Equation 4.1.

Observations	$x_3$	$x_5$	$x_6$	$x_8$	$x_9$	$x_{10}$	Mean $DCW_i$ ( $g.l^{-1}$ )	Mean $DCW_f$ ( $g.l^{-1}$ )	Mean soluble protein yield ( $RLU.ml^{-1}$ )	
									Measured	Predicted
N1	-1	-1	-1	-1	-1	-1	0.09	0.60	199,625	124,038
N2	-1	1	-1	-1	1	-1	0.09	1.10	144,839	245,718
N3	1	-1	-1	-1	1	-1	0.07	0.55	106,421	489,412
N4	1	1	-1	-1	-1	-1	0.09	0.32	75,637	45,207
N5	-1	-1	1	-1	1	-1	0.74	3.34	710,852	489,412
N6	-1	1	1	-1	-1	-1	4.90	4.87	15,679	45,207
N7	1	-1	1	-1	-1	-1	0.47	0.81	210,553	124,038
N8	1	1	1	-1	1	-1	3.02	3.60	77,194	245,718
N9	-1	-1	-1	-1	1	1	0.08	5.16	1,955,840	980,053
N10	-1	1	-1	-1	-1	1	0.09	1.05	705,421	147,452
N11	1	-1	-1	-1	-1	1	0.07	0.77	527,654	319,833
N12	1	1	-1	-1	1	1	0.12	3.30	1,270,210	553,720
N13	-1	-1	1	-1	-1	1	0.74	1.17	149,192	319,833
N14	-1	1	1	-1	1	1	3.01	3.38	84,983	553,720
N15	1	-1	1	-1	1	1	0.52	5.19	2,804,730	980,053
N16	1	1	1	-1	-1	1	2.81	2.44	106,785	147,452
N17	-1	-1	-1	1	1	-1	0.05	0.87	150,716	489,412
N18	-1	1	-1	1	-1	-1	0.11	0.77	132,055	45,207
N19	1	-1	-1	1	-1	-1	0.09	0.45	135,652	124,038
N20	1	1	-1	1	1	-1	0.11	1.61	429,209	245,718
N21	-1	-1	1	1	-1	-1	0.75	1.13	87,900	124,038
N22	-1	1	1	1	1	-1	2.85	3.67	289,720	245,718
N23	1	-1	1	1	1	-1	0.61	5.33	1,481,220	489,412
N24	1	1	1	1	-1	-1	2.89	2.42	27,639	45,207
N25	-1	-1	-1	1	-1	1	0.11	1.32	128,921	319,833
N26	-1	1	-1	1	1	1	0.13	5.05	368,363	553,720
N27	1	-1	-1	1	1	1	0.08	5.52	564,253	980,053
N28	1	1	-1	1	-1	1	0.12	0.90	114,381	147,452
N29	-1	-1	1	1	1	1	0.44	2.13	593,970	980,053
N30	-1	1	1	1	-1	1	2.96	2.25	95,179	147,452

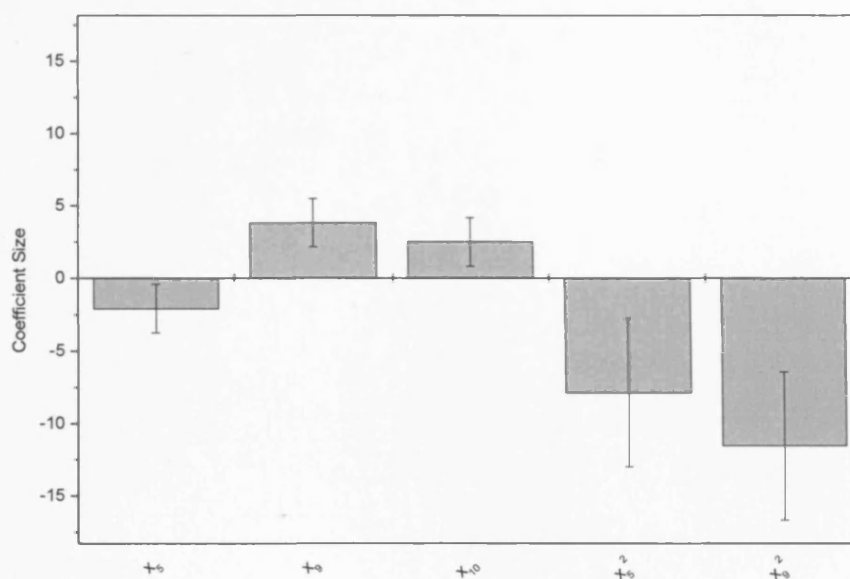


N31	1	-1	1	1	-1	1	0.54	1.01	164,592	319,833
N32	1	1	1	1	1	1	2.86	3.17	565,823	553,720
N33	0	-1	0	0	0	0	0.24	5.26	2,233,070	1,804,860
N34	0	1	0	0	0	0	0.66	3.46	1,858,960	1,111,250
N35	-1	0	0	0	0	0	0.78	5.44	1,726,340	3,248,830
N36	1	0	0	0	0	0	0.68	5.31	2,341,450	3,248,830
N37	0	0	-1	0	0	0	0.10	2.75	1,115,710	3,248,830
N38	0	0	1	0	0	0	2.18	5.19	4,264,530	3,248,830
N39	0	0	0	0	0	-1	0.91	4.82	3,257,280	2,546,740
N40	0	0	0	0	0	1	1.07	8.18	5,388,800	4,086,950
N41	0	0	0	-1	0	0	0.65	5.24	2,945,420	3,248,830
N42	0	0	0	1	0	0	0.91	7.37	2,751,670	3,248,830
N43	0	0	0	0	-1	0	0.59	1.37	498,272	537,454
N44	0	0	0	0	1	0	0.91	5.45	3,031,590	1,459,700
N45	0	0	0	0	0	0	0.44	5.61	3,423,950	3,248,830
N46	0	0	0	0	0	0	0.59	5.64	3,249,220	3,248,830
N47	0	0	0	0	0	0	0.53	5.60	3,630,600	3,248,830
N48	0	0	0	0	0	0	0.73	5.72	3,317,380	3,248,830
N49	0	0	0	0	0	0	0.98	6.24	4,068,960	3,248,830
N50	0	0	0	0	0	0	0.89	5.80	4,126,540	3,248,830

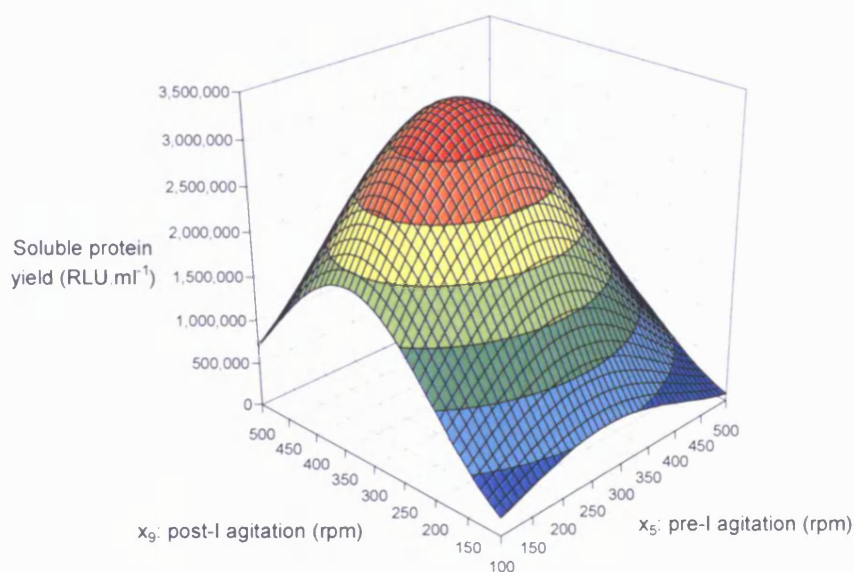
This also provides an early indication of the significance of agitation and aeration conditions upon scale-up if the optimised results are to be reproduced at larger scales. Although the post-I period was significant, the lack of a significant square term implied that its optimum setting lay outside of the range studied; the positive coefficient suggested it lay beyond 18h.

Figure 4.4 shows the response surface plot of pre- and post induction shaking speed with respect to soluble protein yield. Since shaking speed was shown to have such a significant effect on soluble protein yield it is logical to deduce that mixing and aeration are important factors. The existence of an optimum shaking speed was at first surprising. However, visual observations indicated that at the higher speeds splashing occurred within each well leading to the formation of a liquid film on the underside of the gas permeable membrane. This would have increased resistance to oxygen transfer (Zimmermann *et al.*, 2003) resulting in oxygen limitation of the culture and decreased soluble protein yield. Also, while

high oxygen transfer rates may benefit cell growth, high growth rates do not normally favour expression of heterologous protein which is the main response variable in this study (Broedel *et al.*, 2002; Stanbury and Whitaker, 1993).



**Figure 4.3** Coefficient plot for final optimisation model (Equation 4.1), complete with confidence intervals (set at 95 %) for soluble protein yield ( $Y$  transformed as  $Y^{0.25}$ ). The original model consisted of 27 terms. However, terms whose confidence interval included the value zero were deemed insignificant and so were removed from the model in a hierarchical manner. Factors and experimental conditions as described in Table 4.1.



**Figure 4.4** Response surface plot showing influence of pre-I and post-I shaking speed on soluble protein yield (all other optimisation factors at their centre-point levels). Response surface generated using Equation 4.1.

Overall the three key variables identified at this stage strongly affect both cell growth rate and final biomass concentration. Analysis of the raw factor combinations (Table 4.3) suggests that the highest soluble protein yields are obtained under conditions which promote both a slow growth rate during protein synthesis and a high final biomass yield.

#### 4.5 Regression analysis and process modelling

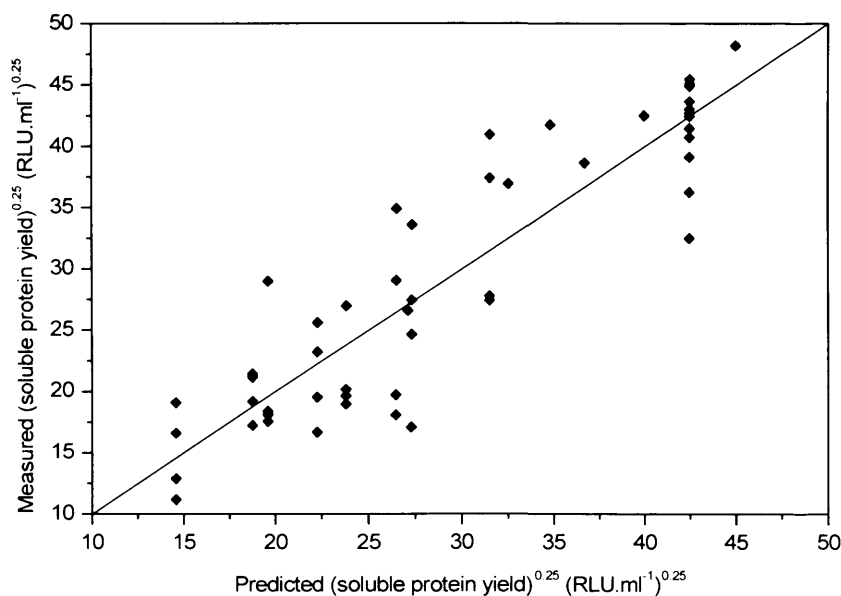
The final stage in the application of DoE involves analysis and verification of the model. Model analysis helps to provide insight into any experimental issues that may not be accounted for in the model. Model validation involves experimental verification of model predictions under conditions not specifically tested in the experimental design. Regression analysis of the reduced optimisation model for soluble protein yield (Table 4.4) indicated that 82 % of the response variation was explained by the model and 76 % of the response variation was predicted by the model. Figure 4.5 shows the correlation between measured and predicted

soluble protein yield values. The predictive power of the regression model is generally good but the 76 % value suggests that aspects of the system being studied were behaving in an unpredictable manner.

**Table 4.4** Estimated regression coefficients of the reduced optimisation model for soluble protein yield (Equation 4.1). The corresponding *p*-values and confidence intervals are also shown. Variables as described in Table 4.1.

Variable	Coefficient	Standard error	<i>p</i> value	95% confidence interval ( $\pm$ )
Constant	42.455	1.250	< 0.001	2.520
Pre-I shaking speed	-2.093	0.830	0.015	1.672
Post-I shaking speed	3.841	0.830	< 0.001	1.672
Post-I period	2.507	0.830	0.004	1.672
(Pre-I shaking speed) <sup>2</sup>	-7.895	2.538	0.003	5.115
(Post-I shaking speed) <sup>2</sup>	-11.538	2.538	< 0.001	5.115

$$R^2 = 0.817; Q^2 = 0.760$$



**Figure 4.5** Parity plot showing measured and predicted soluble protein yields for optimisation experiments. Predicted values obtained from Equation 4.1. Solid line represents line of parity.

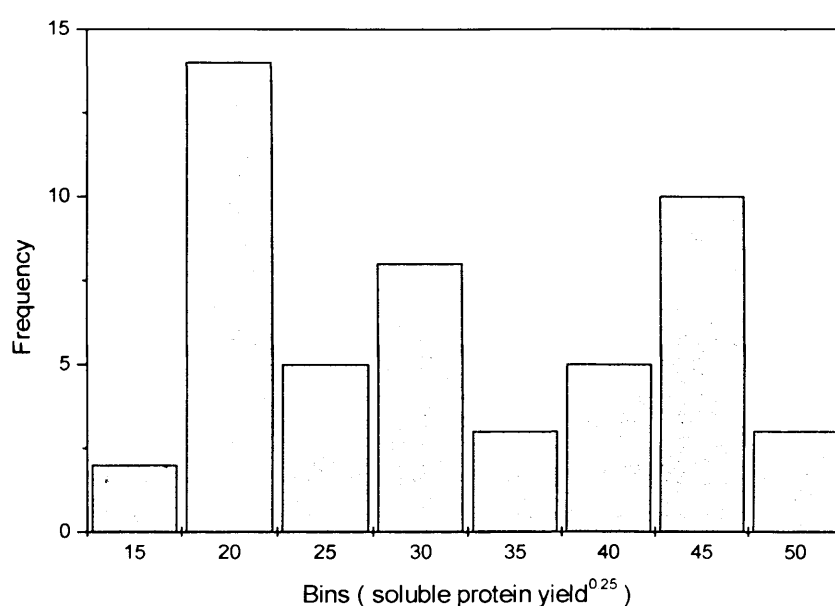
Analysis of Variance (ANOVA) was performed on the residual sum of squares ( $SS_{\text{resid}}$ ), the results for which are shown in Table 4.5.  $SS_{\text{resid}}$  is comprised of the sum of two components: the experimental error, or the pure error sum of squares ( $SS_{\text{pe}}$ ) and the model error, the lack of fit sum of squares ( $SS_{\text{lof}}$ ). ANOVA revealed that  $SS_{\text{lof}}$  was the major contributor to  $SS_{\text{resid}}$  ( $SS_{\text{lof}} \approx 1023$ ), with the pure error associated with the replicate experiments being relatively very low ( $SS_{\text{pe}} \approx 6$ ).

**Table 4.5** ANOVA results for analysis of the reduced optimisation model (Equation 4.1) for soluble protein yield.

Source	Degrees of freedom (DF)	Sum of squares (SS)	Mean square (MS)	<i>F</i> statistic	<i>p</i> value
Total	50	48390.100	967.802		
Constant	1	42752.100	42752.100		
Total Corrected	49	5637.980	115.061		
Regression	5	4608.540	921.709	39.396	< 0.001
Residual	44	1029.440	23.396		
Lack of Fit	39	1023.050	26.232	20.547	0.002
Pure Error	5	6.384	1.277		

It's possible this lack of fit was due to the removal of too many terms from the model; it may have been prudent to leave significant interaction and square terms in the model even if their corresponding main effect terms were shown to be insignificant. Uncontrolled factors and/or mechanisms may also have been present. A number of experiments were seen to generate extremely low soluble protein yields despite having relatively high biomass concentrations. On inspection of these cultures, cells were seen to have aggregated at the end of the fermentation and the broth had a viscosity consistent with that of lysed cells. If cells within these cultures did indeed lyse disproportionately to other

experiments, it would have represented a new mechanism, unaccounted for by the existing experimental design. Secondly, in experiments where the culture was shaken gently, cells were seen to adhere to the surface of the wells complicating OD measurements and sampling of the wells. Overall, however, the distribution of responses (Figure 4.6) shows that the current experimental window is ideally placed around a highly variable region of the system. Furthermore, the general trends seem to have been captured well since both the R2 and Q2 values are high and acceptable for such a complex system



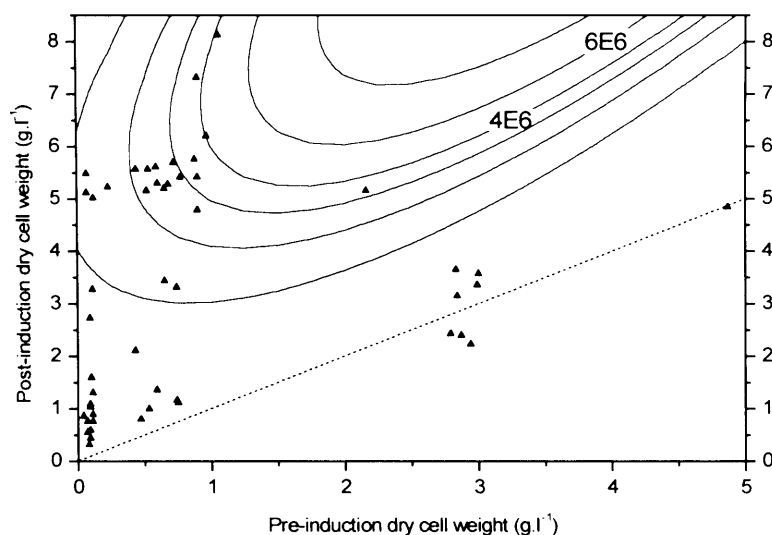
**Figure 4.6** Histogram of responses for (soluble protein yield)<sup>0.25</sup> from the optimisation experiments (Table 4.3).

To test the accuracy of the regression model, two verification experiments were performed at the predicted optimal factor settings. The final regression model was comprised of the coefficients shown in Table 4.4 and assumed the form:

$$Y^{0.25} = 42.46 - 2.09x_5 + 3.84x_9 + 2.51x_{10} - 7.90x_5^2 - 11.54x_9^2 \quad (4.1)$$

From this equation the predicted optimal factor settings were pre-I shaking speed  $\approx 270$  rpm, post-I period  $\approx 18$ h and post-I shaking speed  $\approx 330$  rpm. All other factors were maintained at their mid-point levels (Table 4.1). The mean soluble protein yield measured experimentally at these predicted optimal settings was  $3.26 \times 10^6$  RLU.ml<sup>-1</sup> ( $\sigma = 14.5$  %) and the corresponding pre-I and post-I mean biomass concentrations were 1.01 g.l<sup>-1</sup> and 6.88 g.l<sup>-1</sup> respectively. The mean soluble protein yield obtained was equivalent to just over 76 % of the predicted optimum of  $4.25 \times 10^6$  RLU.ml<sup>-1</sup>. This result suggested the model is highly predictive however, some lack of fit of the model does occur. For comparison, shake flask fermentations were also performed under standard expression conditions (Section 2.3.5). These fermentations were run in duplicate and a mean soluble protein yield of 552,000 RLU.ml<sup>-1</sup> ( $\sigma = 12.9$  %) was obtained. Overall, the soluble protein yield measured under the predicted optimum conditions is almost 6-fold greater than that obtained under reference shake flask conditions.

Finally, Figure 4.7 shows the variation of soluble protein yield modelled as a quadratic function of pre- and post-induction biomass yield, over which lies a scatter plot of experimental observations. The model is able to largely explain the variation in soluble protein yield, with an  $R^2$  value of 0.87 and thus several conclusions can be drawn. When the pre-induction biomass is high,  $\sim 3$  g.l<sup>-1</sup>, the resultant post-induction biomass concentration remains relatively constant. The addition of IPTG at high biomass concentrations seems to prevent further growth of cells. Furthermore, the soluble protein yields measured at these levels are all relatively small ( $<1 \times 10^6$  RLU.ml<sup>-1</sup>). In contrast, the highest soluble protein yields generally correspond to both a relatively low pre-induction biomass concentration, below  $\sim 1$  g.l<sup>-1</sup>, and high post-induction biomass concentration, above  $\sim 5$  g.l<sup>-1</sup>.



**Figure 4.7** Contour plot showing variation of soluble protein yield ( $\text{RLU}\cdot\text{ml}^{-1}$ ) with respect to pre-induction dry cell weight ( $\text{DCW}_i$ ) and final dry cell weight ( $\text{DCW}_f$ ). Model used to generate contours was a least squares quadratic fit:  $\log_{10}(\text{soluble protein yield}) = 11.36 - 0.13x(\text{DCW}_i) + 0.50x(\text{DCW}_f) - 0.12x(\text{DCW}_i)^2 - 0.02x(\text{DCW}_f)^2 + 0.09x(\text{DCW}_i)x(\text{DCW}_f)$ . Data used to generate model are taken from Table 4.3 and are superimposed over the contour plot. Dashed line represents line of parity.

## 4.6 Summary

The aim of this chapter was to optimise soluble protein expression at the microwell scale. Towards this aim a generic framework was developed, comprising four key stages: familiarisation, screening, optimisation and regression analysis towards generating a process map.

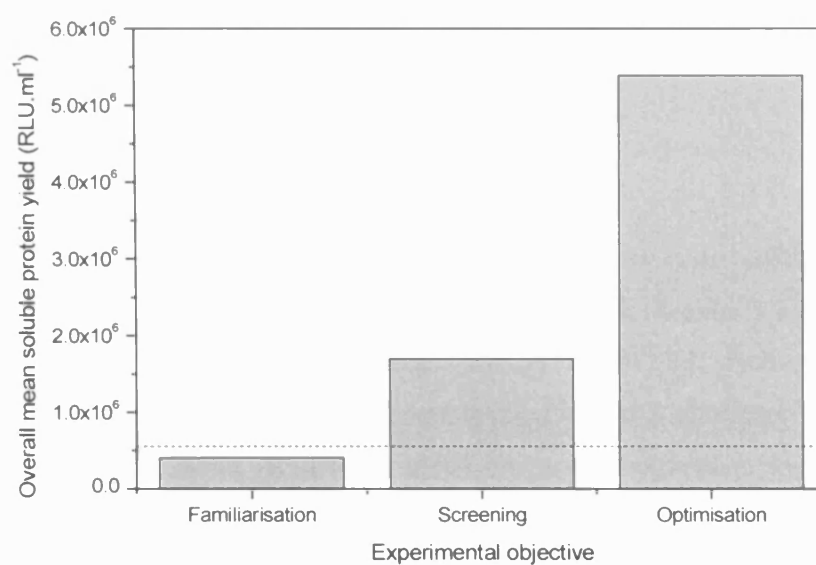
Ten factors were initially identified at the familiarisation stage (Section 4.2) using a combination of experience and literature studies. Relevant system responses were identified and the appropriate measurement techniques were developed. A simple experimental design was also applied to further develop the experimental methods. The ten selected factors were subsequently screened



(Section 4.3) and six were chosen for further investigation through optimisation. The optimisation stage (Section 4.4) identified three significant factors and their effects on soluble protein expression levels were modelled. This model was subsequently improved through statistical analysis and the model's predictive ability verified through experimentation (Section 4.5).

In comparison to other HTPE approaches, the application of DoE led to a significant decrease in the number of experiments performed. Ten variables were studied at both the familiarisation and screening stages (2 qualitative variables at 3 levels and 8 quantitative variables at 2 levels) and six were studied at optimisation (6 quantitative variables at 3 levels). The total number of factor combinations for these three experimental designs is  $2(3^2 \cdot 2^8) + 3^6 = 5,337$ . This is a large, potentially expensive, and time consuming number of experiments to perform. In contrast, the DoE approach required just 110 experiments (a 98 % reduction) and led to a high degree of process understanding, as evidenced by the results for verification of the process model. At each stage of the DoE process, an increased understanding of the expression system was achieved. This is clearly illustrated in Figure 4.8 which demonstrates the significant improvements in maximum soluble protein yields obtained from the familiarisation, screening and optimisation experiments. Overall, the maximum soluble protein yield was increased more than 9-fold compared to reference conditions.

The use of parallel microwell scale fermentations also helped to reduce the experimental timeframes. These timeframes could be further reduced if all experiments required by DoE were automated. Importantly, shaking speed and liquid fill volume, two factors that have a strong influence on oxygen mass transfer into microwells (Hermann *et al.*, 2003) were included as factors in the experimental design. The fact that shaking speed was one of the most significant factors in the optimisation model illustrates its importance when performing microwell fermentations. It also suggests that careful control of agitation and aeration conditions upon scale-up of the optimised culture conditions will be important in maintaining the optimal balance between cell growth and protein expression.



**Figure 4.8** Comparison of maximum soluble protein yields obtained from the familiarisation, screening and optimisation experiments. Dashed line indicates the mean soluble protein yield measured from shake flask experiments under standard expression conditions (Section 2.3.5).

## 5 Scale-up of MWP fermentation performance to 7.5 l STR scale<sup>†</sup>

### 5.1 Aims and objectives

Following the optimisation of soluble protein expression at the microwell scale in Chapter 4, the next stage of the proposed framework (Section 3.3.2) involves implementing an appropriate scale-up strategy which will enable the rapid reproduction of MWP fermentation performance within a laboratory scale STR. Given the significance of shaking speed on protein expression in Chapter 4 (Figure 4.3) and its impact on oxygen transfer (Hermann *et al.*, 2003),  $k_La$  was considered to be an important scale-up parameter. Initial studies have also suggested that scaling-up on the basis of constant  $k_La$  appears to be the most appropriate approach when dealing with microorganisms growing under aerobic conditions (Ferreira-Torres *et al.*, 2005; Micheletti *et al.*, 2006).

The aim of this chapter is, therefore, to establish  $k_La$  as a suitable basis for scale-up of MWP results and to determine the range of  $k_La$  values over which this scale-up strategy is valid. Specific objectives are:

- to measure the variation of  $k_La$  over a range of operating conditions within both the MWP and 7.5 l STR scales (Sections 5.2 and 5.3);
- to validate these measurements through comparisons to existing correlations (Sections 5.2 and 5.3);
- to perform MWP and 7.5 l STR scale fermentations at matched  $k_La$  values (Section 5.4);
- to use established statistical techniques to demonstrate the equivalence, or otherwise, of fermentation kinetic profiles at the two scales (Section 5.4).

<sup>†</sup>The majority of the results presented in this chapter have been submitted for publication as: Islam,R.S., Tisi,D., Levy,M.S., and Lye,G.J. (2007) Scale-up of *E. coli* growth and recombinant protein expression conditions from microwell to laboratory and pilot scale based on matched  $k_La$ . *Biotechnology and Bioengineering*. In press. doi:10.1002/bit.21697

## 5.2 Measurement and correlation of MWP $k_La$ values

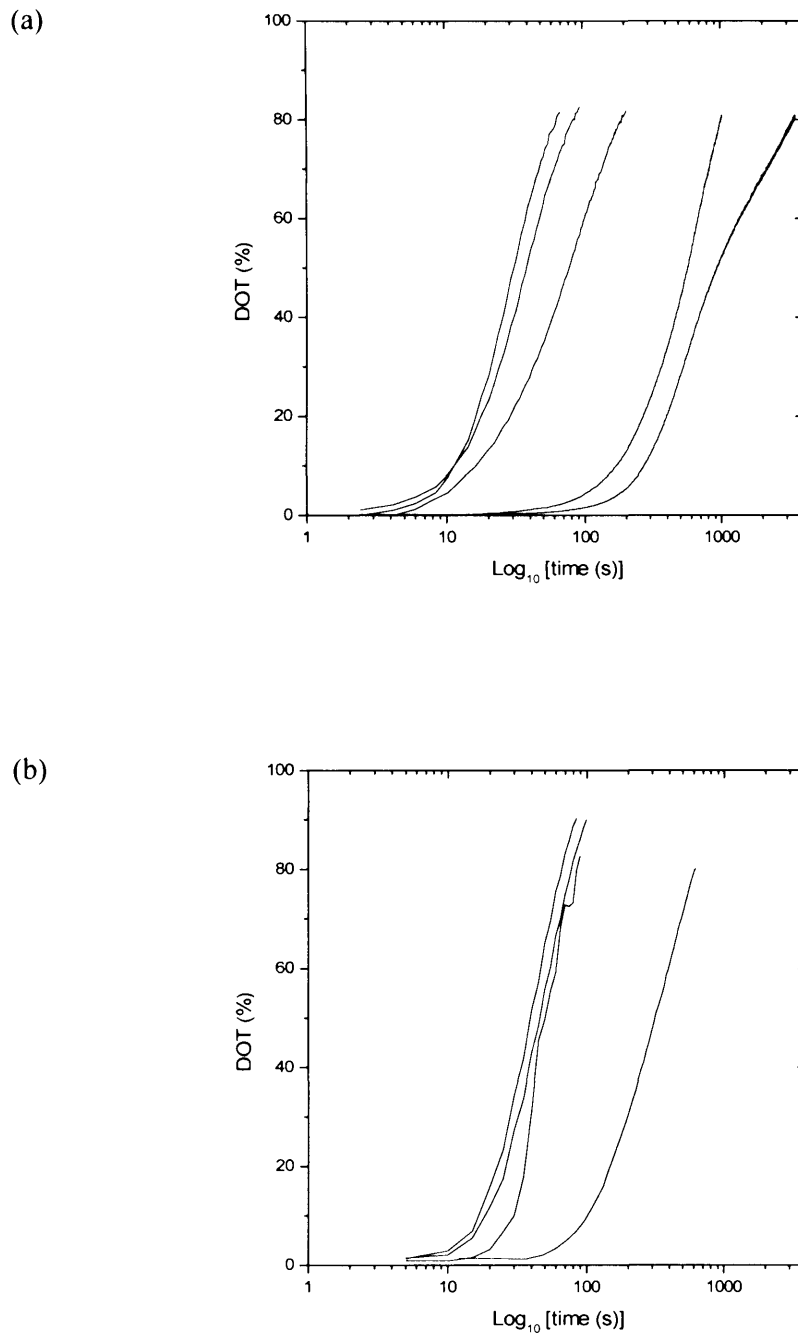
The initial stage of the scale-up strategy involved characterising and correlating MWP  $k_La$  values over a range of operating conditions. First, the static-gassing out technique was used to generate dissolved oxygen tension (DOT) profiles for each of the MWP geometries studied in Chapter 4. These experiments are described in Section 2.4.1 and a sample of the resultant DOT profiles is shown in Figure 5.1 (a). The range of  $k_La$  values measured from such profiles is illustrated in Figure 5.2 (a) and  $k_La$  is observed to increase with increasing shaking speed in all cases (the raw data from which Figure 5.2 was constructed is given in Appendix A). This figure also compares the measured  $k_La$  values with those predicted from the correlation of Doig *et al.* (2005) for a 24-well plate with a circular cross-section and flat base (the closest geometry to those used here):

$$k_La = 31.35.D.a_i.Re^{0.68}.Sc^{0.36}.Fr^{0.86}.Bo^{0.03} \quad (5.1)$$

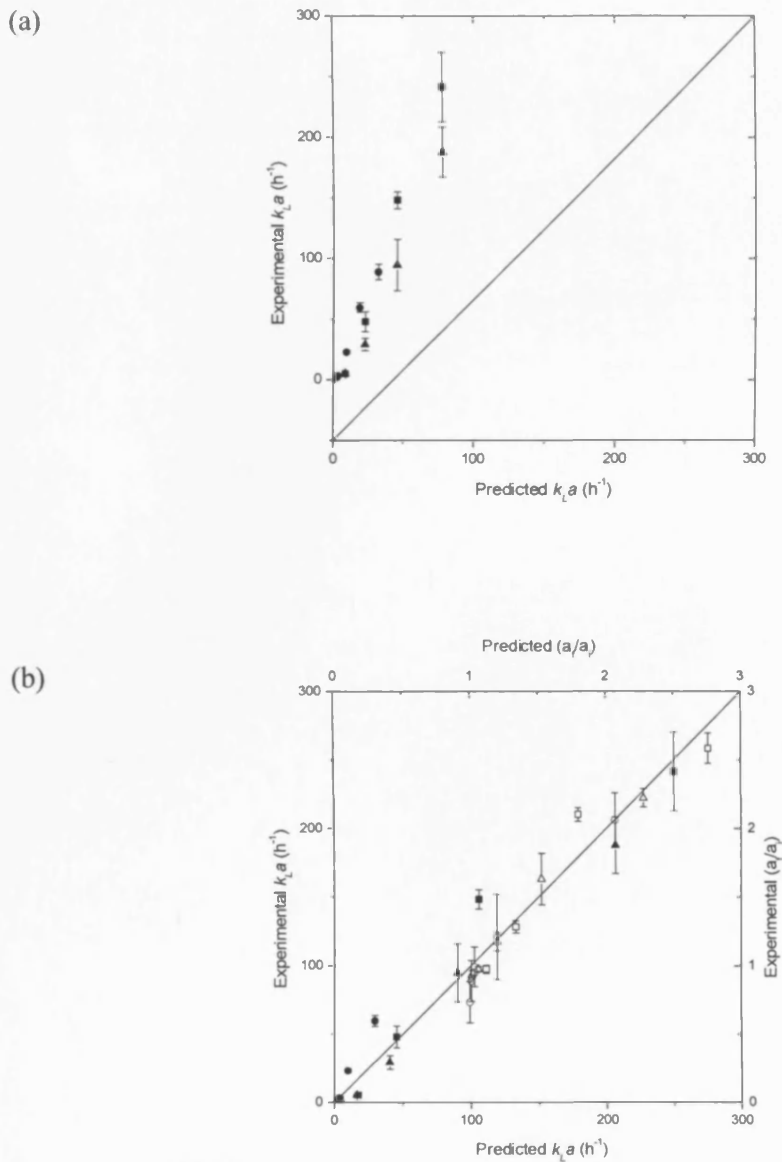
The experimental  $k_La$  values shown in Figure 5.2 (a) are significantly larger than those predicted. This is most probably due to the improved oxygen mass-transfer characteristics of square versus round well plate formats (Duetz and Witholt, 2004; Hermann *et al.*, 2003).

In order to broaden the application of the Doig *et al.* (2005) correlation to include square well plates, the original correlation was modified. The Bond and Schmidt numbers were first removed from the correlation since these groups do not vary with shaking speed (the only variable investigated here in the  $k_La$  experiments). The sub-correlation for prediction of  $a_f / a_i$  was then altered to the following form:

$$\ln\left(\frac{a_f}{a_i}\right) = a.Fr^b \quad (5.2)$$



**Figure 5.1** Example DOT profiles obtained across the full range of shaking / agitation speeds within (a) the MWP (plate P3) and (b) the 7.5 l STR. Experiments were performed as described in Section 2.4. A logarithmic transformation of time is displayed on the x-axes such that the full range of profiles for each bioreactor may be easily distinguished in one figure.



**Figure 5.2** Comparison of predicted and measured MWP  $k_L a$  and  $a_f/a_i$  values: (a) the correlation for  $k_L a$  presented by Doig *et al.* (2005) (Equation 5.1) was used to predict  $k_L a$  values for plates P1 [●], P2 [▲] and P3 [■]; (b) Equation 5.2 was used to predict  $a_f/a_i$  values for plates P1 [○], P2 [△] and P3 [□] and the modified Doig correlation (Equation 5.3) was fitted to these data and used to predict  $k_L a$  values for plates P1 [●], P2 [▲] and P3 [■]. Values predicted from Equations 5.2 and 5.3 lie close to the line of parity [—] and so are in good agreement with the corresponding experimental values.  $k_L a$  values were measured as described in Section 2.4.1 and values for  $a_f/a_i$  were measured as described in Section 2.4.2 and. Error bars indicate one standard deviation around the mean.

where  $Fr$  is the Froude number and  $a$ , and  $b$  are constants. This alteration ensured that both sides of the equation would tend towards zero as the shaking speed approached zero. Equation 5.2 was then fitted to each set of specific air-liquid surface area measurements (Appendix A) within the three MWP geometries. The  $a_f / a_i$  values predicted are in excellent agreement with those obtained experimentally as illustrated by the open symbols in Figure 5.2 (b).

These specific air-liquid surface area measurements were also used to model the variation of  $k_L$  with respect to shaking speed in the same manner as Doig *et al.* (2005). This model was combined with Equation 5.2 and the following modified correlation for square well plates as used here was obtained:

$$k_L a = 3.94 \times 10^{-4} \cdot (D/d_v) \cdot a_i \cdot Re^{1.91} \cdot \exp(aFr^b) \quad (5.3)$$

The corresponding values for  $a$  and  $b$ , respectively, were 1.66 and 2.47 for plate P1, 0.70 and 1.51 for plate P2, and 0.88 and 1.24 for plate P3. Figure 5.2 (b) also illustrates the good agreement between  $k_L a$  values (solid symbols) predicted from Equation 5.3 with those obtained experimentally. Plate P3 is seen to yield the highest  $k_L a$  values under the conditions studied. As it also showed the highest soluble protein yields (Figure 4.2) during protein expression characterisation, this MWP format was chosen for subsequent scale-up studies.

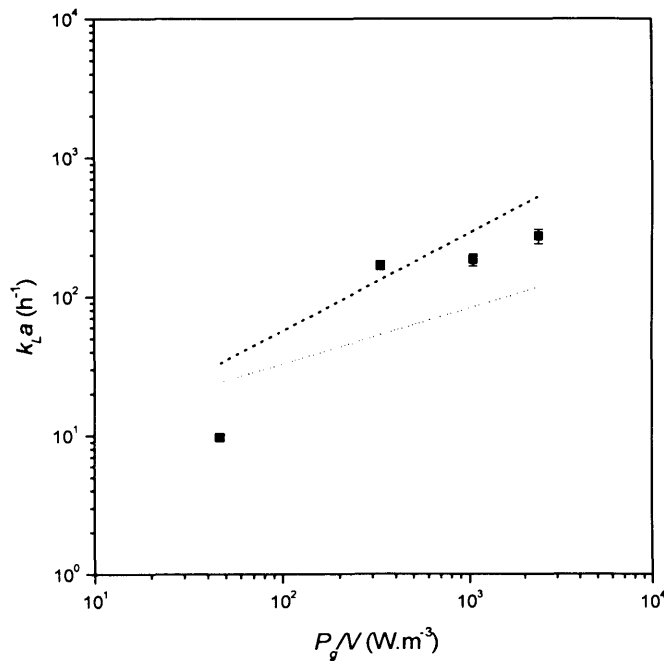
### 5.3 Measurement and correlation of 7.5 l STR $k_L a$ values

The next stage of the scale-up strategy involved characterising and correlating laboratory scale STR  $k_L a$  values over a range of operating conditions. The static-gassing out technique was again used to generate DOT profiles, in this case within a conventional 7.5 l STR as described in Section 2.4.3, and a sample of the resultant DOT profiles is shown in Figure 5.1 (b). The range of  $k_L a$  values measured from such profiles (Figure 5.3) span a similar range to those obtained in the MWP experiments and  $k_L a$  is again observed to increase with increasing agitation speed (the raw data from which Figure 5.3 was constructed is given in

Appendix A). Figure 5.3 also compares the experimental  $k_La$  values with those determined from the correlations of van't Riet (1979). For pure water (coalescing medium) and strong ionic solutions (non-coalescing medium), respectively:

$$k_La = 0.026 \left( \frac{P_g}{V} \right)^{0.4} v_s^{0.5} \quad (5.4)$$

$$k_La = 0.002 \left( \frac{P_g}{V} \right)^{0.7} v_s^{0.2} \quad (5.5)$$



**Figure 5.3** Comparison of experimental 7.5 l STR  $k_La$  values with those predicted by van't Riet (1979) (Equations 5.4 and 5.5) for pure water [ ... ] and strong ionic solutions [ --- ] at an air flow rate of 1 vvm. Values for  $P_g$  are calculated from Equation 5.6. Error bars indicate one standard deviation around the mean.



In both cases  $P_g$  in the 7.5 l STR was estimated according to the correlation proposed by Hughmark (1980):

$$P_g = 0.1P_{ug} \left( \frac{Q}{NV} \right)^{-1/4} \left( \frac{N^2 D_i^4}{gW_b V^{2/3}} \right)^{-1/6} \quad (5.6)$$

$$\text{where } P_{ug} = P_0 \times \rho \times N^3 \times D_i^5 \quad (5.7)$$

The experimental  $k_L a$  values shown in Figure 5.3 lie mainly within the boundaries enclosed by the correlations for pure water and strong ionic solutions. This is expected as  $k_L a$  measurements made here were performed in a complex nutrient medium (as described in Section 2.2.1), which is of an ionic strength intermediate to the phases used by van't Riet.

Visual observations of the 7.5 l STR showed that all  $k_L a$  values above  $\sim 200 \text{ h}^{-1}$  were obtained under conditions where there was good gas-liquid distribution within the vessel. Under the agitation conditions used at lower  $k_L a$  values, however, there appeared to be poor gas dispersion within the vessel.

## 5.4 Fermentations at matched $k_L a$ values

### 5.4.1 Selection of $k_L a$ values and experimental conditions

The next stage of the scale-up process involved performing MWP and 7.5 l STR fermentations at two sets of  $k_L a$  values, 'high' and 'low' so as to provide an estimate for the *range* of  $k_L a$  values over which successful scale-up could be achieved.

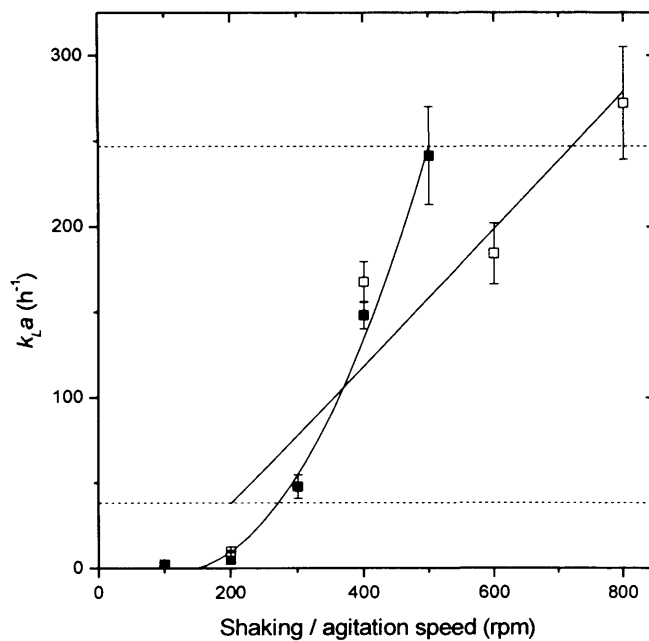
Figure 5.4 illustrates the variation of  $k_L a$  values achieved in relation to shaking / agitation speed in both the MWP and the 7.5 l STR. From this figure, the observed range of common  $k_L a$  values is between  $38 \text{ h}^{-1}$  and  $247 \text{ h}^{-1}$ , corresponding to MWP shaking speeds of between 270 and 500 rpm. Figure 5.4

also shows the least squares model fits of the  $k_La$  data with respect to shaking / agitation speed. For the MWP and the 7.5 l STR, respectively, these are:

$$k_La = 0.00170 n^2 - 0.400 n + 21.5 \quad (5.8)$$

$$k_La = 0.402 N - 42.6 \quad (5.9)$$

These models yielded  $R^2$  values of 0.99 and 0.90 respectively, indicating good overall explanations of the response variation. The difference in  $k_La$  profiles observed here (curved versus linear) is most probably due to the increased significance of surface tension effects at the microwell scale (Hermann *et al.*, 2003).



**Figure 5.4** Variation of experimental  $k_La$  values with shaking speed for plate P3 [ ■ ] and with agitation speed for 7.5 l STR [ □ ]. The dashed lines represent boundaries on the common range of  $k_La$  values measured at both scales. The solid lines represent least-squares model fits of the  $k_La$  data, described by Equations 5.8 and 5.9. Error bars indicate one standard deviation around the mean.

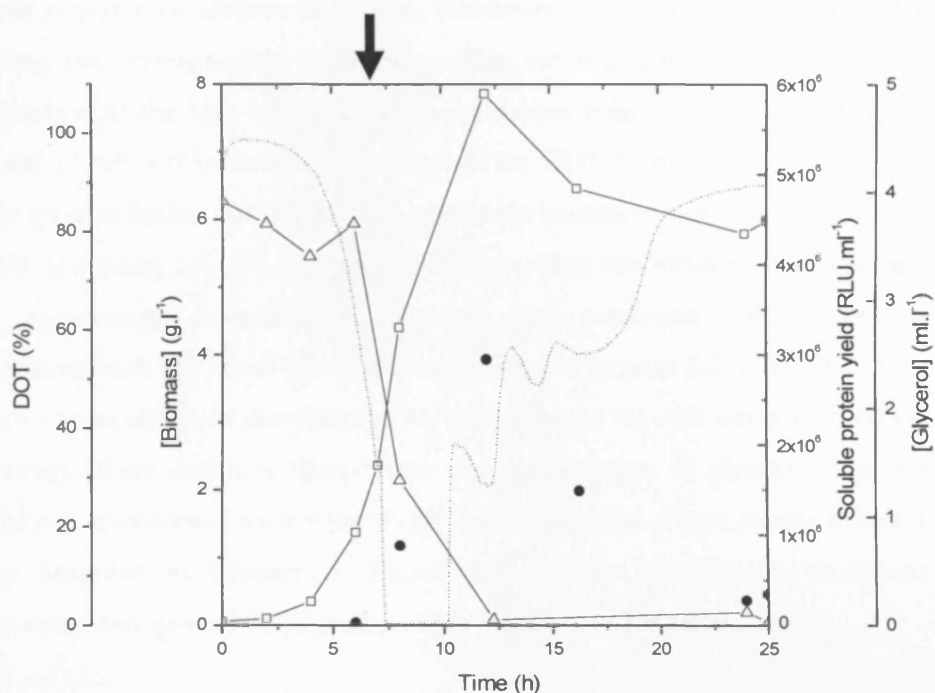
Equations 5.8 and 5.9 were then combined and rearranged to obtain a relationship between MWP and 7.5 l STR shaking / agitation speeds at matched  $k_La$ :

$$N = 0.00424 n^2 - 0.996 n + 160 \quad (5.10)$$

This equation enabled the prediction of suitable agitation speeds within the 7.5 l STR on fermentation scale-up. Appropriate fermentation conditions were then chosen from the optimisation experiments of Section 4.4 using the following criteria. A single shaking speed would be required throughout each fermentation in order to minimise the variation in  $k_La$ . This shaking speed should correspond to a scalable  $k_La$  value and should, therefore, be in the range of 270 to 500 rpm. The next requirement was for a minimum fill volume of 3ml per well - the same volume in which MWP  $k_La$  values were determined (Section 2.4.1). The final requirement was for a long overall fermentation time so that large comparative data sets could be generated. These criteria narrowed down the choice of experiments to N36 for the 'low'  $k_La$  control and N32 for the 'high'  $k_La$  control, the conditions of which are described in Table 2.4.

#### 5.4.2 MWP fermentation kinetics

Figure 5.5 illustrates a typical MWP fermentation profile at a 'high'  $k_La$  value of  $247 \text{ h}^{-1}$  and Table 5.1 shows a summary of kinetic parameters for the same fermentation. Exponential growth started at  $\sim 2 \text{ h}$  exhibiting a maximum specific growth rate,  $\mu_{\max}$ , of  $0.63 \text{ h}^{-1}$ . At 7 h, the cells were induced with  $500 \mu\text{M}$  IPTG and  $\sim 30 \text{ min}$  later, the dissolved oxygen tension fell to a minimum of 0 %. The DOT began to rise again at  $\sim 10 \text{ h}$  which coincided with the end of the growth phase. The biomass concentration reached a maximum of  $\sim 7.9 \text{ g.l}^{-1}$  around 12 h. This coincided with the depletion of glycerol, the main carbon substrate. At the same point, the soluble protein yield also appeared to reach a maximum of  $3.0 \times 10^6 \text{ RLU.ml}^{-1}$ . After this time the level of expressed FFL fell rapidly suggesting utilisation of proteins by *E. coli* as a more complex source of carbon and energy once the glycerol was depleted (Garcia-Arrazola *et al.*, 2005).

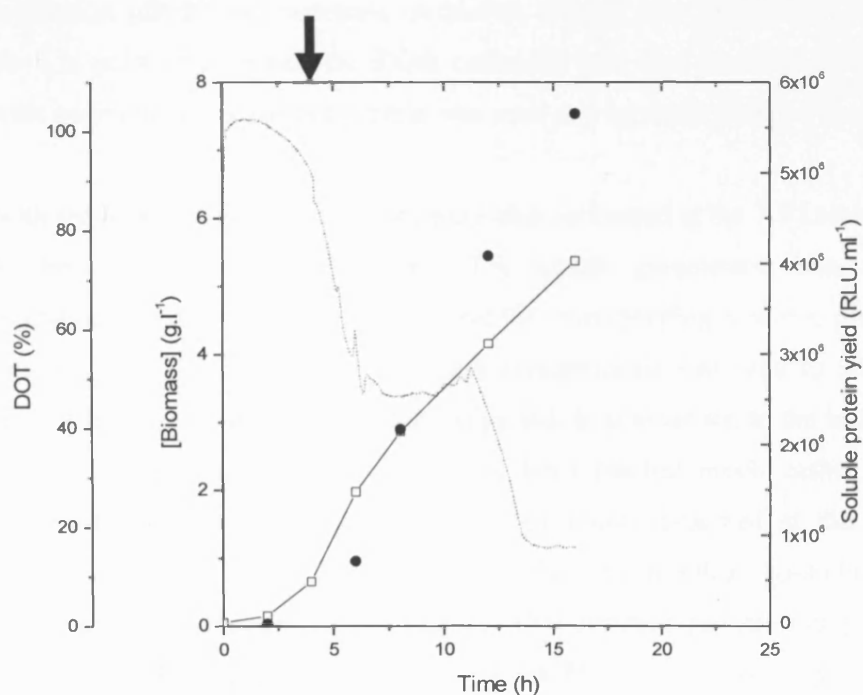


**Figure 5.5** MWP (plate P3) fermentation kinetic profiles at high  $k_L a$  ( $247 \text{ h}^{-1}$ ). Profiles show biomass concentration [  $\square$  ], soluble protein yield [  $\bullet$  ], glycerol concentration [  $\triangle$  ] and DOT [  $\cdots$  ]. Arrow indicates time of induction. Experiments performed as described in Section 2.5.1.

**Table 5.1** Summary of kinetic parameters for fermentations performed at the microwell and laboratory scales at matched  $k_L a$  values. Parameters calculated from Figures 5.5 to 5.8 inclusive. NA indicates data not obtainable.

Variable	Units	$k_L a = 247 \text{ h}^{-1}$		$k_L a = 55 \text{ h}^{-1}$	
		MWP vs. 7.5 l STR	MWP vs. 7.5 l STR	MWP vs. 7.5 l STR	MWP vs. 7.5 l STR
$CER_{\max}$	$\text{mmol.l}^{-1}.\text{h}^{-1}$	NA	41	NA	5.4
$OUR_{\max}$	$\text{mmol.l}^{-1}.\text{h}^{-1}$	NA	51	NA	5.0
$\mu_{\max}$	$\text{h}^{-1}$	0.63	0.59	0.63	0.49
$X_{\max}$	$\text{g.l}^{-1}$	7.9	7.7	5.4	2.3
$[\text{Soluble protein}]_{\max}$	$\text{RLU.ml}^{-1}$	$2.9 \times 10^6$	$3.2 \times 10^6$	$5.7 \times 10^6$	$2.2 \times 10^6$
Specific $[\text{soluble protein}]_{\max}$	$\text{RLU.mg}^{-1}$	$3.8 \times 10^5$	$4.2 \times 10^5$	$1.1 \times 10^6$	$9.8 \times 10^5$

MWP experiments were also performed at a 'low'  $k_{La}$  value of  $55 \text{ h}^{-1}$ . The kinetic response parameters from these fermentations are also described in Table 5.1 and the corresponding response profiles are illustrated in Figure 5.6. In comparison to the high  $k_{La}$  experiment, cells were found to be still growing by the end of the post-induction period though the DOT profile suggested this was under oxygen limited conditions. The maximum biomass concentration measured at this time point,  $5.4 \text{ g.l}^{-1}$ , is significantly lower than that obtained from the high  $k_{La}$  experiment. Interestingly, however, the maximum soluble protein concentration at the lower  $k_{La}$  was almost twice as large at  $5.7 \times 10^6 \text{ RLU.ml}^{-1}$ . There was no observed decrease in FFL expression as the cells never reached the stationary phase and it is likely there was no depletion of glycerol (data not available) as observed within the 'high'  $k_{La}$  experiment. These results reflect a trend observed in Chapter 4 (Figure 4.7), where fermentation conditions promoting slow growth often produced the highest soluble protein yields.

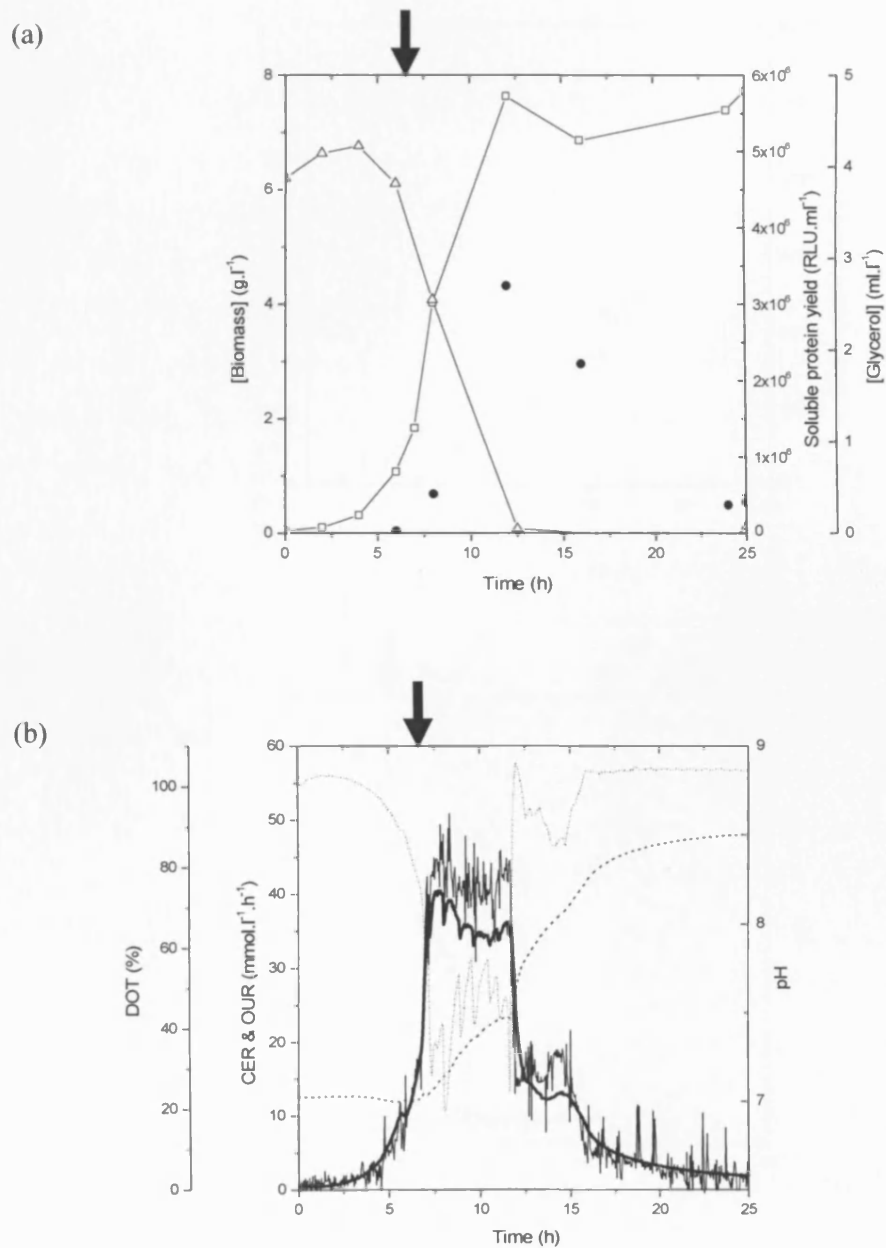


**Figure 5.6** MWP (plate P3) fermentation kinetic profiles at low  $k_{La}$  ( $55 \text{ h}^{-1}$ ). Profiles show biomass concentration [  $\square$  ], soluble protein yield [  $\bullet$  ] and DOT [  $\cdots$  ]. Arrow indicates time of induction. Experiments performed as described in Section 2.5.1.

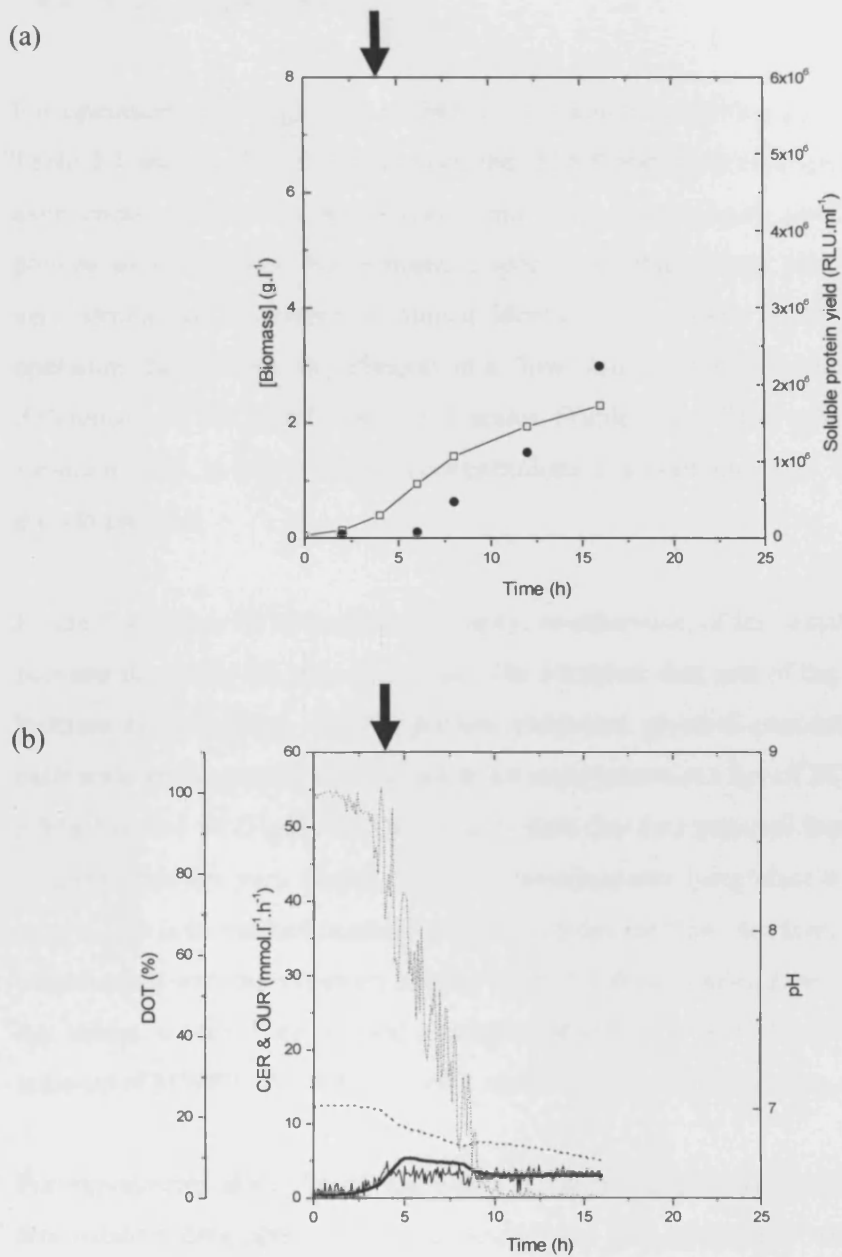
### 5.4.3 7.5 l STR fermentation kinetics

Figure 5.7 shows the fermentation profiles obtained within the 7.5 l STR, at a matched  $k_La$  of  $247 \text{ h}^{-1}$ . Exponential growth started at  $\sim 2 \text{ h}$  with a corresponding  $\mu_{\max}$  of  $0.59 \text{ h}^{-1}$  (Table 5.1). At 7 h, the cells were induced as in the MWP experiment. Approximately 1 h post-induction, the DOT fell to a minimum of  $\sim 20 \%$ , and the CER and OUR reached a maximum of  $41 \text{ mmol.l}^{-1}.\text{h}^{-1}$  and  $51 \text{ mmol.l}^{-1}.\text{h}^{-1}$  respectively. At  $\sim 12 \text{ h}$  there was a sharp change in these profiles, coinciding with the depletion of glycerol. The biomass and soluble protein concentrations both reached a maximum at this time of  $7.7 \text{ g.l}^{-1}$  and  $3.2 \times 10^6 \text{ RLU.ml}^{-1}$  respectively. The measured rise in pH after this point is consistent with protein deamination as a result of protein being used as a secondary carbon and energy source (Garcia-Arrazola *et al.*, 2005) and could again account for the decrease in the expressed levels of FFL. Further insight may have been gained through analysis of the rates of biomass accumulation ( $dX/dt$ ), product accumulation ( $dP/dt$ ) and substrate utilisation ( $dS/dt$ ). For example, if whilst  $dS/dt=0$ , a point of inflection for  $dX/dt$  coincided with that for  $dP/dt$ , it would provide additional evidence that protein was used as a secondary carbon source.

As with the MWP studies experiments were also performed at the 7.5 l scale at a 'low' matched  $k_La$  value of  $55 \text{ h}^{-1}$ . The kinetic parameters from these fermentations are described in Table 5.1 and the corresponding response profiles are illustrated in Figure 5.8. The biomass concentration was seen to be still increasing by the end of the post-induction period, in comparison to the high  $k_La$  experiment where the stationary phase had been reached much earlier. The maximum biomass and soluble protein concentrations measured at this time point,  $2.3 \text{ g.l}^{-1}$  and  $2.2 \times 10^6 \text{ RLU.ml}^{-1}$  respectively, are lower than those obtained from the high  $k_La$  experiment (Table 5.1). This retarded growth and protein expression in the stirred-tank bioreactor is attributed to the poor gas-liquid dispersion observed at the low impeller speed. The maximum CER and OUR values were also  $\sim 10$ -fold lower than those observed within the 'high'  $k_La$  experiment.



**Figure 5.7** 7.5 l STR fermentation kinetic profiles at high  $k_{La}$  ( $247 \text{ h}^{-1}$ ): (a) offline data shows biomass concentration [  $\square$  ], soluble protein yield [  $\bullet$  ] and glycerol concentration [  $\triangle$  ]; (b) online data shows DOT [  $\cdots$  ], OUR [  $-$  ], CER [  $- -$  ] and pH [  $- \cdot -$  ]. Arrows indicate time of induction. Experiments performed as described in Section 2.5.2.



**Figure 5.8** 7.5 l STR fermentation kinetic profiles at low  $k_{La}$  ( $55 \text{ h}^{-1}$ ): (a) offline data shows biomass concentration [  $\square$  ] and soluble protein yield [  $\bullet$  ]; (b) online data shows DOT [  $\cdots$  ], OUR [  $\text{—}$  ], CER [  $\text{---}$  ] and pH [  $\text{-.-}$  ]. Arrows indicate time of induction. Experiments performed as described in Section 2.5.2.

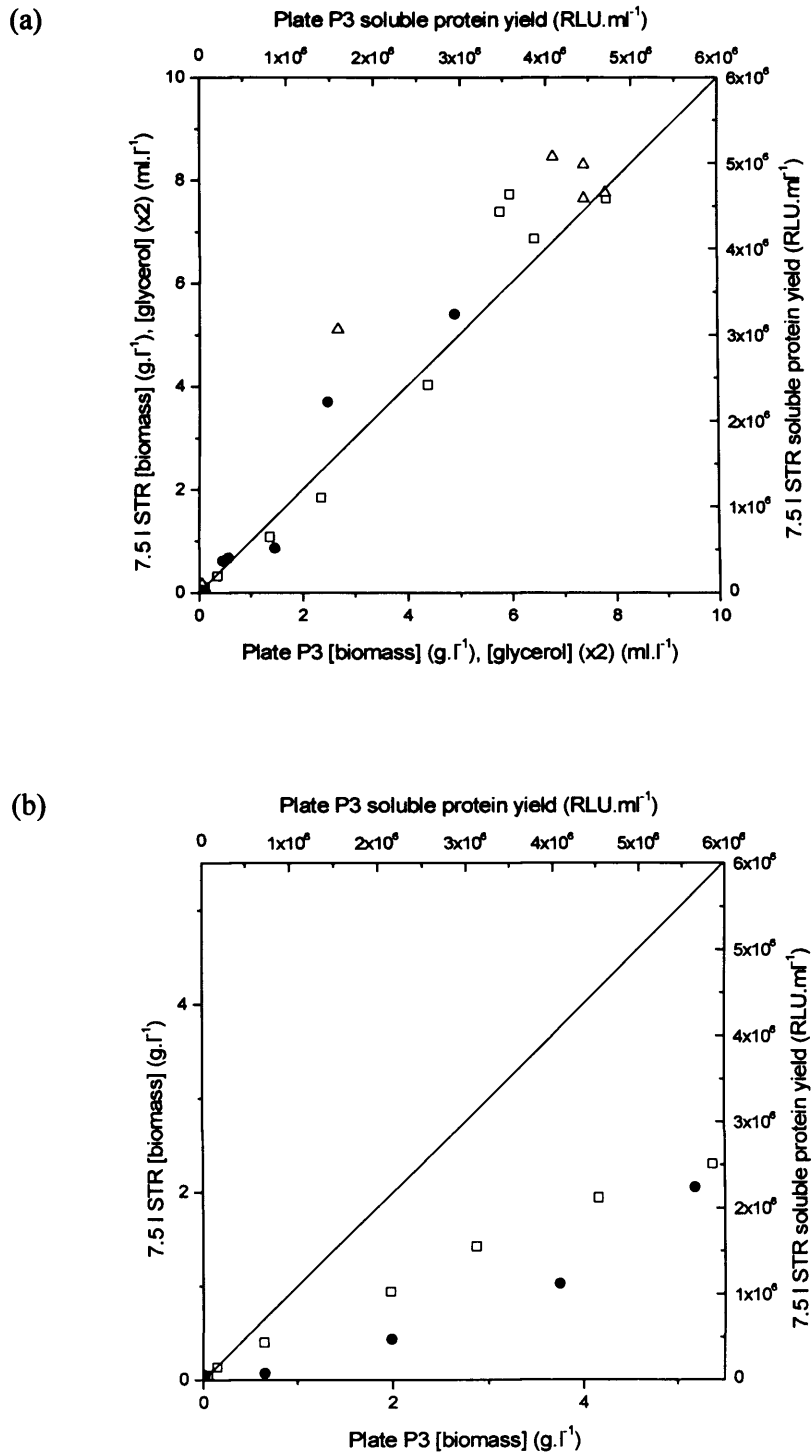


#### 5.4.4 Scale comparisons

For operation at a 'high'  $k_{La}$  of  $247\text{ h}^{-1}$ , the kinetic parameters summarised in Table 5.1 are largely similar between the MWP and 7.5 l STR scales. In these experiments similar final biomass concentrations and maximum yields of soluble protein were obtained. The maximum specific soluble protein yields were also very similar and occurred at almost identical time points at both scales of operation. In contrast, experiments at a 'low'  $k_{La}$  of  $55\text{ h}^{-1}$  showed significant differences at the MWP and 7.5 l scales (Table 5.1). There was significant variation both in the biomass concentrations achieved and also the biomass growth profiles.

Figure 5.9 further illustrates the similarity, or otherwise, of fermentation profiles between the scales for each  $k_{La}$  value. The complete data sets of the variation of biomass concentration, soluble protein yield and glycerol concentration from each scale are presented as parity plots for experiments at a  $k_{La}$  of  $247\text{ h}^{-1}$  (Figure 5.9 (a)) and  $55\text{ h}^{-1}$  (Figure 5.9 (b)). It is evident that data gathered from 'high'  $k_{La}$  fermentations are very similar, with all measurements lying close to the line of parity. This is in marked contrast to the data from the 'low'  $k_{La}$  fermentations. In combination with the summary data in Table 5.1 these results show that at 'high'  $k_{La}$  values, where there is good gas-liquid distribution at both scales, effective scale-up of MWP fermentation results can be achieved at matched  $k_{La}$  values.

For experiments at the 'high'  $k_{La}$  value, a rigorous statistical comparison of the fermentation data obtained at both scales was also performed with regard to kinetic profiles for cell growth, protein expression and substrate utilisation. Appropriate models were fitted to each profile data set (Appendix B) and analysed as described in Section 2.5.3. This analysis examines the similarity of data sets over the entire course of cell growth and provides a better indication of how well the MWP fermentation results are reproduced at larger scales.



**Figure 5.9** Parity plots comparing fermentation kinetics between MWP (plate P3) and 7.5 l STR for biomass yield [  $\square$  ], soluble protein yield [  $\bullet$  ] and glycerol concentration ( $\times 2$ ) [  $\Delta$  ] at  $k_L a$  values of (a)  $247\text{ h}^{-1}$  and (b)  $55\text{ h}^{-1}$ . Data taken from Figures 5.5 to 5.8 inclusive.

Table 5.2 lists the  $p$  values resulting from statistical analyses on the effect of scale on profiles of biomass concentration, soluble protein yield and substrate concentration at the MWP and 7.5 l STR scales. All profiles apart from those of glycerol were statistically similar at a significance level of 0.05 indicating an excellent agreement between all growth and protein expression. Although the glycerol utilisation kinetics were not found to be significantly similar at this level, the overall fermentation profiles shown in Figures 5.5 and 5.7 do show the complete utilisation of glycerol occurring at the same time point and coinciding with the maxima for cell growth.

**Table 5.2** Comparison of microwell fermentation kinetic profiles (Figure 5.5) to those from the laboratory scale fermentation (Figure 5.7) at matched  $k_La$  conditions ( $247 \text{ h}^{-1}$ ).  $P$  values shown are from  $F$  tests on pairs of profiles for biomass concentration, soluble protein yield and substrate concentration as described in Section 5.4.4. The null hypothesis, which states that there is no overall difference between each pair of profiles, is accepted if  $p > 0.05$ .

Kinetic profile	$P$ values
	Plate P3 vs. 7.5 l STR
Biomass growth	0.56
Glycerol utilisation	$1.63 \times 10^{-4}$
Soluble protein expression	0.08

## 5.5 Summary

The aim of this chapter was to design an effective scale-up strategy for rapidly translating MWP fermentation performance to larger scales. In Chapter 4, shaking speed was identified as one of the most significant factors affecting recombinant protein expression in *E. coli* MWP formats. Shaking speed is known to affect oxygen transfer (Hermann *et al.*, 2003), thus  $k_La$  was assumed to be an important scale-up parameter.

Consequently, in this chapter, MWP  $k_{La}$  values were determined over a range of shaking speeds via the static gassing-out technique (Figure 5.2) and the results were used to establish a predictive correlation applicable to the square-well MWP formats used here (Equation 5.3). A similar range of  $k_{La}$  values was obtained within the 7.5 l STR, using the same  $k_{La}$  measurement technique, and these values also compared well to those predicted by existing correlations (Figure 5.4).

Fermentations were then performed at ‘high’ and ‘low’ matched  $k_{La}$  values. For scale-up at a ‘high’ matched  $k_{La}$  value of  $247 \text{ h}^{-1}$ , MWP fermentation kinetic profiles and yields were accurately reproduced at the larger scale of operation, equivalent to a 1,700 fold scale-up. These results are significant for several reasons. Firstly, they confirm the initial scale-up hypothesis which proposed that  $k_{La}$  was an important scale-up parameter. This hypothesis emerged in Chapter 4 following the application of Design of Experiments (DoE) and so these results also provide further support for the use of DoE as a bioprocess development tool. Finally, these results also confirmed a number of recent studies which have suggested  $k_{La}$  to be an important scale-up parameter for aerobic fermentations (Ferreira-Torres *et al.*, 2005; Micheletti *et al.*, 2006).

Scale-up of MWP fermentations at a ‘low’ matched  $k_{La}$  value, in contrast, did not provide reproducible performance at the 7.5 l scale. This difference was attributed to the poor gas-liquid distributions observed within the larger vessel at the agitation speed used to provide this ‘low’  $k_{La}$  value. Overall, a scale-up strategy based on relatively high matched  $k_{La}$  values should be applicable to larger scales of fermentation. Further scale-up of MWP results to the 75 l scale using  $k_{La}$  as a basis for scale-up is addressed in the next chapter.

## 6 Scale-up of protein expression to 75 l STR scale<sup>†</sup>

### 6.1 Aims and objectives

The final stage of the proposed framework (Section 3.3.3) is to develop a scale-up strategy for translating MWP fermentation performance to a pilot scale STR. In Chapter 5, key fermentation kinetics such as those for cell growth and soluble protein yield were scaled up successfully on the basis of relatively high matched  $k_{La}$  values, i.e.  $k_{La}$  values corresponding to good gas-liquid distributions within the vessels.

The aim of this chapter, therefore, is to further develop this scale-up strategy for application at the pilot scale of operation where larger quantities of protein can be generated. The specific objectives of this chapter are:

- to measure the variation of  $k_{La}$  within a pilot scale (75 l) STR over a range of operating conditions and compare to values measured at the 7.5 l STR and MWP scales (Section 6.2);
- to validate the 75 l  $k_{La}$  measurements through comparisons to existing correlations (Section 6.2);
- to determine if equivalent fermentation performance can be obtained between the MWP and 75 l STR scales on the basis of 'high' matched  $k_{La}$  values (Section 6.3).

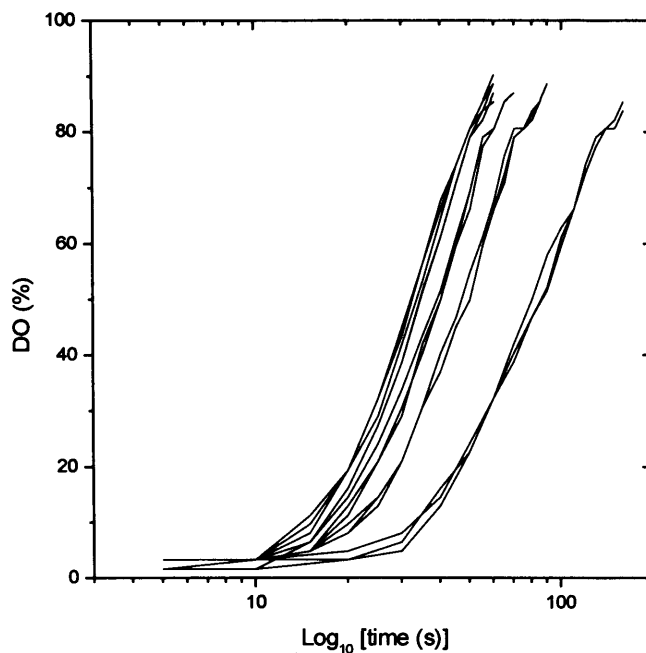
### 6.2 Measurement and correlation of 75 l STR $k_{La}$ values

#### 6.2.1 Measurement of $k_{La}$ values

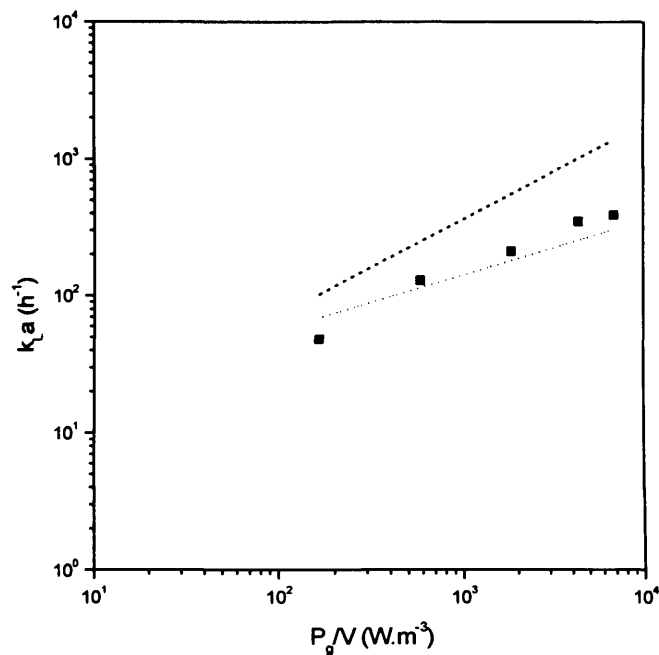
As described in Section 5.3 it was necessary to first measure and then correlate  $k_{La}$  values within the stirred-tank reactor over a range of operating conditions. The static-gassing out technique applied in Chapter 5 was again used to generate

<sup>†</sup>The majority of the results presented in this chapter have been submitted for publication as: Islam, R.S., Tisi, D., Levy, M.S., and Lye, G.J. (2007) Scale-up of *E. coli* growth and recombinant protein expression conditions from microwell to laboratory and pilot scale based on matched  $k_{La}$ . *Biotechnology and Bioengineering*. In press. doi:10.1002/bit.21697

dissolved oxygen tension (DOT) profiles, in this case within a conventional 75 l STR as described in Section 2.4.3, and a sample of the resultant DOT profiles is shown in Figure 6.1. The range of  $k_{La}$  values measured from such profiles (Figure 6.2) spans a similar range to those obtained in the MWP and 7.5 l STR experiments (Chapter 5) and  $k_{La}$  is again observed to increase with increasing agitation speed (the raw data from which Figure 6.2 was constructed is given in Appendix A). In Figure 6.2 the  $k_{La}$  values determined at the 75 l scale are also compared to those obtained from the correlations of van't Riet (1979) for pure water (Equation 5.4) and strong ionic solutions (Equation 5.5). In both cases  $P_g$  in the 75 l STR was estimated according to the correlation proposed by Hughmark (1980) (Equation 5.6).



**Figure 6.1** Example DOT profiles obtained across the full range of agitation speeds within the 75 l STR. Experiments were performed as described in Section 2.4.3. A logarithmic transformation of time is displayed on the x-axes such that the full range of profiles for each bioreactor may be easily distinguished in one figure.

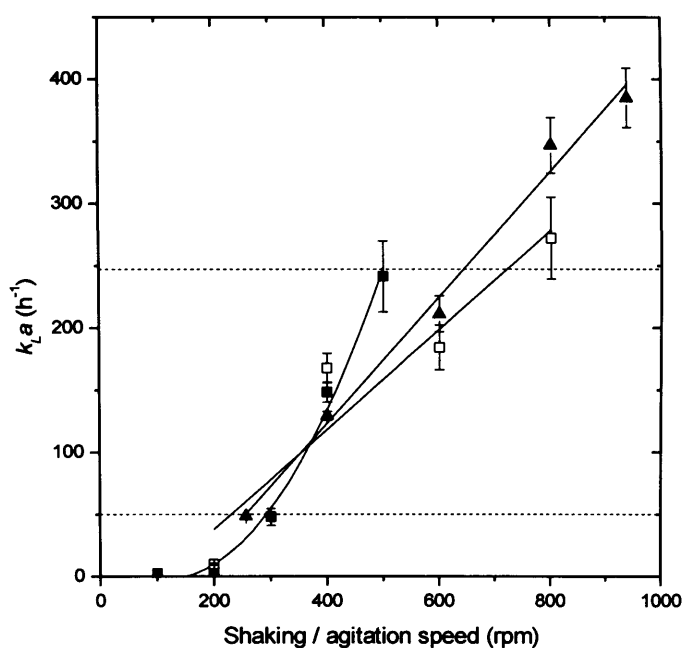


**Figure 6.2** Comparison of experimental 75 l STR  $k_{La}$  values with those predicted by van't Riet (1979) (Equations 5.4 and 5.5) for pure water [  $\cdots$  ] and strong ionic solutions [  $---$  ] at an air flow rate of 1 vvm. Values for  $P_g$  are calculated from Equation 5.6. Error bars indicate one standard deviation around the mean.

The experimental  $k_{La}$  values shown in Figure 6.2 lie mainly within the boundaries enclosed by the correlations for pure water and strong ionic solutions, similar to the trend observed within the 7.5 l STR (Chapter 5). They are also of a similar magnitude to  $k_{La}$  values determined at the 7.5 l scale since measurements were performed at similar  $P_g/V$  values. Visual observations of the 7.5 l STR showed that all  $k_{La}$  values below  $\sim 200 \text{ h}^{-1}$  corresponded to poor gas dispersion throughout the vessel. Visual observations of gas-liquid mixing within the 75 l STR, however, were not possible due to the stainless steel construction of the vessel.

### 6.2.2 Comparison of $k_{La}$ values with other scales

Figure 6.3 compares the variation of  $k_{La}$  values achieved in the 75 l STR, in relation to agitation speed, with those achieved in the MWP and 7.5 l STR (Chapter 5). The maximum  $k_{La}$  value measured in the 75 l STR ( $396 \text{ h}^{-1}$ ) is significantly higher than those measured in the MWP and 7.5 l STR ( $247 \text{ h}^{-1}$  and  $279 \text{ h}^{-1}$ , respectively). From this figure, the observed range of common MWP / 75 l STR  $k_{La}$  values is between  $50 \text{ h}^{-1}$  and  $247 \text{ h}^{-1}$ , corresponding to MWP shaking speeds of between 290 and 500 rpm.



**Figure 6.3** Variation of experimental  $k_{La}$  values with shaking speed for plate P3 [ ■ ] and with agitation speed for 7.5 l STR [ □ ] and 75 l STR [ ▲ ]. The dashed lines represent boundaries on the common range of  $k_{La}$  values measured at both the MWP and 75 l STR scales. The solid lines represent least-squares model fits of the  $k_{La}$  data, described by Equations 5.8, 5.9 and 6.1. Error bars indicate one standard deviation around the mean.



Figure 6.3 also shows the least squares model fits of the  $k_{La}$  data with respect to shaking / agitation speed. For the 75 l STR this is:

$$k_{La} = 0.505 N - 78.7 \quad (6.1)$$

The  $R^2$  value for this model was 0.99, indicating a good overall explanation of the response variation.

Equation 6.1 was then combined with the corresponding MWP equation (Equation 5.8) and rearranged to obtain a relationship between MWP and 75 l STR shaking / agitation speeds at matched  $k_{La}$ :

$$N = 0.00337 n^2 - 0.792 n + 198 \quad (6.2)$$

This equation enabled the prediction of suitable agitation speeds within the 75 l STR on fermentation scale-up.

## 6.3 Fermentations at matched $k_{La}$ values

### 6.3.1 Selection of $k_{La}$ values and experimental conditions

The next stage of the scale-up process involved performing fermentations at both the MWP and 75 l STR scales. In Chapter 5, key fermentation kinetics such as those for cell growth, soluble protein yield and substrate utilisation were scaled up successfully on the basis of relatively high matched  $k_{La}$  values, i.e.  $k_{La}$  values corresponding to good gas-liquid distributions within the vessels. To determine if this strategy also applied at the pilot scale of operation, the 'high'  $k_{La}$  ( $247 \text{ h}^{-1}$ ) MWP fermentation conditions from Chapter 5 would again be scaled up on the basis of matched  $k_{La}$ . This fermentation was identical to experiment N32 from the optimisation experiments of Section 4.4 and the specific fermentation settings for this experiment are described in Table 2.4.

### 6.3.2 MWP fermentation kinetics

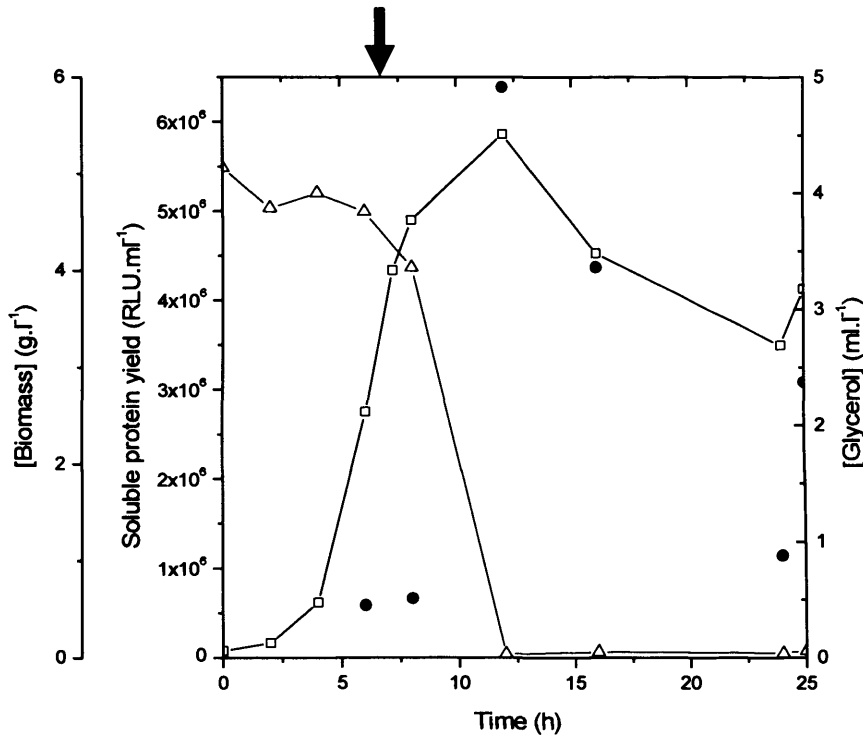
Initial experiments showed that fermentations within the MWP produced lower final biomass concentrations than those in the corresponding 75 l STR fermentations (data not shown). This was attributed to differences in the method of medium sterilisation between the two vessels, as described in Section 2.4, and the different temperature-time profiles achieved. Consequently, when performing MWP and 75 l fermentations at a matched  $k_{La}$  of  $247 \text{ h}^{-1}$ , the 75 l vessel was first inoculated and then a sample of the inoculated medium was rapidly removed and used to fill the wells in the parallel MWP fermentations as described in Section 2.5.

Figure 6.4 illustrates a typical MWP fermentation profile at the ‘high’  $k_{La}$  value of  $247 \text{ h}^{-1}$  and Table 6.1 shows a summary of kinetic parameters measured for the same fermentation. Exponential growth started at  $\sim 2 \text{ h}$  exhibiting a maximum specific growth rate,  $\mu_{\max}$ , of  $0.66 \text{ h}^{-1}$ . At 7 h, the cells were induced with  $500 \mu\text{M}$  IPTG and this coincided with the end of the exponential growth phase. At around 12 h, the biomass and soluble protein concentrations reached a maximum of  $\sim 5.4 \text{ g.l}^{-1}$  and  $\sim 6.4 \times 10^6 \text{ RLU.ml}^{-1}$  respectively, coinciding with the depletion of glycerol. After this time the level of expressed FFL fell rapidly suggesting utilisation of proteins by *E. coli* as a more complex source of carbon and energy once the glycerol was depleted as was seen previously at the MWP and 7.5 l scales (Section 5.4).

### 6.3.3 75 l STR fermentation kinetics

Figure 6.5 shows the fermentation profiles obtained within the 7.5 l STR, at a matched  $k_{La}$  of  $247 \text{ h}^{-1}$ . Exponential growth started at  $\sim 2 \text{ h}$  with a corresponding  $\mu_{\max}$  of  $\sim 0.62 \text{ h}^{-1}$  (Table 6.1) and the cells were induced at 7 h in the same way as for the MWP experiment and. Approximately 45 min post-induction, the DOT fell to a minimum of  $\sim 27 \%$ , and the CER and OUR reached a maximum of  $48 \text{ mmol.l}^{-1}.\text{h}^{-1}$  and  $50 \text{ mmol.l}^{-1}.\text{h}^{-1}$  respectively. At 12 h the biomass and soluble

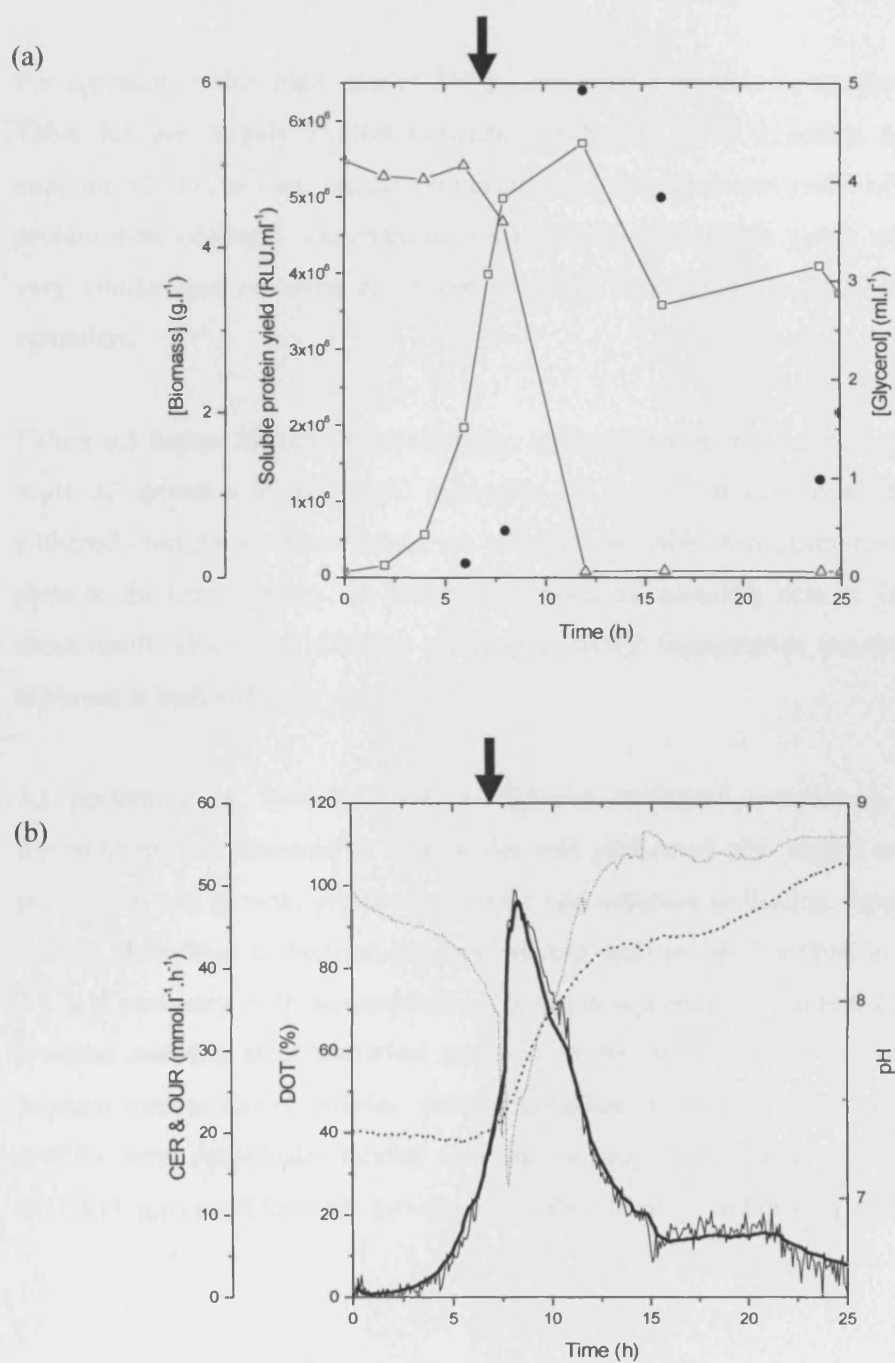
protein concentrations reached a maximum of  $\sim 5.3 \text{ g.l}^{-1}$  and  $\sim 6.4 \times 10^6 \text{ RLU.ml}^{-1}$ , respectively, which again coincided with the depletion of glycerol.



**Figure 6.4** MWP (plate P3) fermentation kinetic profiles at high  $k_{La}$  ( $247 \text{ h}^{-1}$ ). Profiles show biomass concentration [  $\square$  ], soluble protein yield [  $\bullet$  ] and glycerol concentration [  $\triangle$  ]. Arrow indicates time of induction. Experiments performed as described in Section 2.5.1.

**Table 6.1** Summary of kinetic parameters for fermentations performed at the microwell and pilot scales at matched  $k_{La}$  values. Parameters calculated from Figures 6.4 and 6.5. NA indicates data not obtainable.

Variable	Units	$k_{La} = 247 \text{ h}^{-1}$	
		MWP vs. 75 l STR	
$CER_{max}$	$\text{mmol.l}^{-1}.\text{h}^{-1}$	NA	48
$OUR_{max}$	$\text{mmol.l}^{-1}.\text{h}^{-1}$	NA	50
$\mu_{max}$	$\text{h}^{-1}$	0.66	0.62
$X_{max}$	$\text{g.l}^{-1}$	5.4	5.3
$[\text{Soluble protein}]_{max}$	$\text{RLU.ml}^{-1}$	$6.4 \times 10^6$	$6.4 \times 10^6$
Specific $[\text{soluble protein}]_{max}$	$\text{RLU.mg}^{-1}$	$1.2 \times 10^6$	$1.5 \times 10^6$



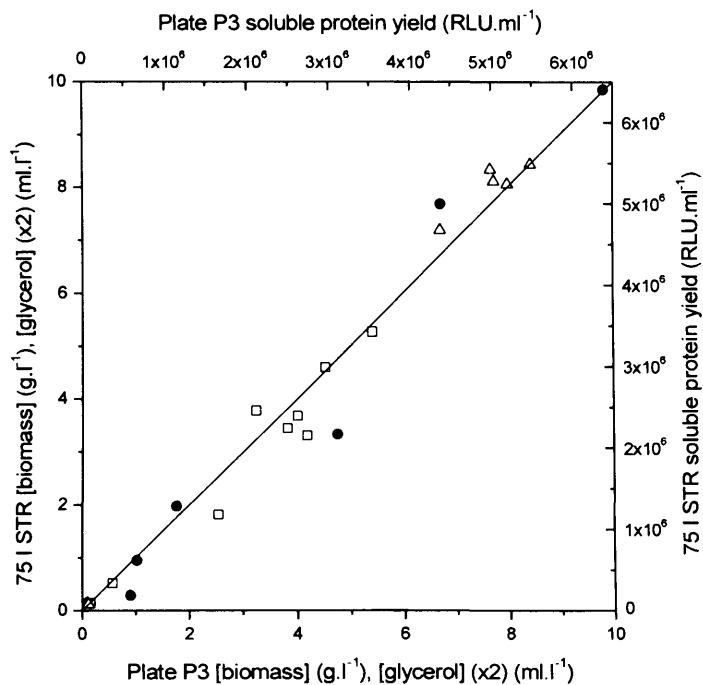
**Figure 6.5** 75 l STR fermentation kinetic profiles at high  $k_{La}$  ( $247 \text{ h}^{-1}$ ): (a) offline data shows biomass concentration [  $\square$  ], soluble protein yield [  $\bullet$  ] and glycerol concentration [  $\triangle$  ]; (b) online data shows DOT [  $\cdots$  ], OUR [  $-$  ], CER [  $\text{—}$  ] and pH [  $---$  ]. Arrows indicate time of induction. Experiments performed as described in Section 2.5.2.

#### 6.3.4 Scale comparisons

For operation at the 'high'  $k_{La}$  of  $247 \text{ h}^{-1}$ , the kinetic parameters summarised in Table 6.1 are largely similar between the MWP to 75 l scales. In these experiments similar final biomass concentrations and maximum yields of soluble protein were obtained. The maximum specific soluble protein yields were also very similar and occurred at almost identical time points at both scales of operation.

Figure 6.6 further illustrates the similarity of fermentation profiles between each scale of operation at the 'high'  $k_{La}$  value of  $247 \text{ h}^{-1}$ . It is evident that data gathered from these fermentations are very similar, with all measurements lying close to the line of parity. In combination with the summary data in Table 6.1 these results show that effective scale-up of MWP fermentation results can be achieved at matched  $k_{La}$  values.

As performed in Section 5.4.4, a rigorous statistical comparison of the fermentation data obtained at both scales was performed with regard to kinetic profiles for cell growth, protein expression and substrate utilisation. Appropriate models were fitted to each profile data set and analysed as described in Section 2.5.3. A summary of these model fits is given in Appendix B. Table 6.2 lists the  $p$  values resulting from statistical analyses on the effect of scale on profiles of biomass concentration, soluble protein yield and substrate concentration. All profiles were statistically similar at a significance level of 0.05 indicating an excellent agreement between growth, substrate utilisation and protein expression.



**Figure 6.6** Parity plot comparing fermentation kinetics at 'high'  $k_La$  value of  $247 \text{ h}^{-1}$  between MWP (plate P3) and 75 l STR for biomass yield [  $\square$  ], soluble protein yield [  $\bullet$  ] and glycerol concentration (x2) [  $\Delta$  ]. Data taken from Figures 6.4 and 6.5.

**Table 6.2** Comparison of microwell fermentation kinetic profiles (Figure 6.4) to those from the pilot plant scale fermentation (Figure 6.5) at matched  $k_La$  conditions ( $247 \text{ h}^{-1}$ ).  $P$  values shown are from  $F$  tests on pairs of profiles for biomass concentration, soluble protein yield and substrate concentration as described in Section 6.3.4. The null hypothesis, which states that there is no overall difference between each pair of profiles, is accepted if  $p > 0.05$ .

Kinetic profile	$P$ values
	Plate P3 vs. 75 l STR
Biomass growth	0.26
Glycerol utilisation	0.06
Soluble protein expression	0.80

## 6.4 Summary

In Chapter 5, a scale-up strategy was developed which enabled the reproduction of MWP fermentation performance within a laboratory scale STR on the basis of relatively high matched  $k_La$  values. The aim of this chapter was to further develop this strategy for application at the pilot scale of operation. Consequently,  $k_La$  values within a conventional 75 l STR were first rapidly quantified over a range of agitation speeds via the static gassing out technique and these values compared well to those predicted by existing correlations (Figure 6.2). This measurement technique was identical to those used at the MWP and laboratory scales and so the results were directly comparable. All three scales shared a wide range of common  $k_La$  values ( $50 \text{ h}^{-1}$  to  $247 \text{ h}^{-1}$ ), but overall significantly higher  $k_La$  values were achievable within the 75 l STR (Figure 6.3).

In Section 3.3.3 several factors were identified which could lead to variations in fermentation performance between the MWP and pilot scales. To minimise variation all media components were sourced from the same manufacturer and the method of inoculum preparation was preserved on scale-up. Initial experiments, however, still showed a difference in fermentation performance between the MWP and 75 l scales. This was attributed to differences in the method of medium sterilisation between the two vessels, as described in Section 2.4, and hence the different temperature-time profiles achieved. The scale-up strategy was thus modified by using media sterilised in the 75 l STR for subsequent fermentations at both scales.

MWP fermentations at a 'high'  $k_La$  value of  $247 \text{ h}^{-1}$  (Figure 6.4) were then scaled-up to the 75 l STR (Figure 6.5) on the basis of matched  $k_La$  and kinetic profiles and yields were accurately reproduced (Figure 6.6). No studies have previously demonstrated the reproduction of MWP fermentation profiles to this scale, equivalent to a 15,000 fold scale-up, and this result provides final indication of the efficacy of the scale-up strategy developed here. In so doing, the work covered in this chapter completes the proposed framework for underpinning the generation of large quantities of soluble protein in a rapid and

cost-effective manner. The validation of this approach and its application in bioprocess development are covered further in the next two chapters.



## **7 Management benefits of the proposed framework<sup>†</sup>**

### **7.1 Aims and objectives**

This thesis has addressed the major challenge of rapidly generating sufficient quantities of soluble recombinant protein for analysis during drug discovery. The design of the proposed framework (Chapter 3) was optimised towards this goal and, therefore, offers time and cost savings over alternative approaches.

The aim of this chapter is to provide a thorough and quantitative analysis of these management benefits over both stages of the proposed framework: the rapid optimisation of protein expression (Section 7.2) and the rapid scale-up of fermentation and generation of protein samples (Section 7.3).

### **7.2 Protein expression optimisation**

#### **7.2.1 Background and general assumptions**

Relatively few published studies exist regarding the optimisation of protein expression within drug discovery and this suggests that it may be an uncommon industrial activity. This is most probably due to the shortcomings of the traditional characterisation approaches. High-throughput expression screening (HTES), for example, is resource intensive and the one-factor-at-a-time (OFAT) approach is unable to model complex systems reliably (Section 1.3.1). Thus there are no published benchmarks against which the time and cost savings of the present work can be evaluated and it is, therefore, necessary to develop hypothetical alternatives.

Within the current framework, the DoE and MWP experimentation components were key to the rapid optimisation of protein expression. Reasonable hypothetical alternatives may, therefore, include approaches which lack one or the other of these components. In the following sections, three cost scenarios are

---

<sup>†</sup>The inclusion of this chapter relates to the University of London EngD requirement for consideration of the bioprocess management implications of the research work performed.

---

presented: (1) the current framework, (2) the current framework minus the MWP experimentation component, and (3) the current framework minus the DoE component which is akin to the HTES approach. The costs associated with each scenario have been separated into four categories: inoculum preparation, fermentation, assay and labour. A complete breakdown of costs is provided in Section C.5. The general assumptions for all three scenarios are:

- Material costs (Section C.1) are limited to:
  - the cost of media (including antibiotic and inducer);
  - the cost of disposable MWP;
  - the cost of luciferase assay buffer components;
- Labour costs are proportional to minimum fermentation time (Section C.2) which assumes:
  - 24 h experimentation and instant turnaround;
  - inoculum preparation and assays are performed in parallel with fermentations;
  - the availability of one technician payable at £10.h<sup>-1</sup>;
- Rate of fermentation limited to 72 microwells or 24 shake-flasks (SFs) in parallel assuming:
  - a maximum sampling time of three minutes;
  - a maximum sampling rate of 24 microwells or 8 SFs per minute per technician;
- All other costs including utilities, labware and equipment maintenance are negligible;
- All fermentations are performed in quadruplicate;

### **7.2.2 Scenario 1: current framework**

In this scenario, both the DoE and MWP framework components are applied and a cost analysis for this scenario is provided in Table 7.1.

**Table 7.1** Cost analysis summary for Scenario 1: current framework. These costs have been summarised from the more detailed analysis of Section C.5.2.

Cost category	Cost (£)
Inoculum preparation	94
Fermentation	261
Assay	252
Labour	2,750
<b>Total</b>	<b>3,357</b>

### 7.2.3 Scenario 2: current framework minus the MWP component

In this scenario, the DoE component is implemented but the MWP component is replaced by shake-flask (SF) experimentation and the costs associated with this change are estimated. Here it is assumed that SFs are processed at a slower rate than microwells (24 SFs versus 72 microwells) for the following reasons. Each set of four replicate MWP fermentations were situated on the same MWP, covered with a single gas-permeable membrane. By comparison, four separate SFs were required to process the same set of experiments, each of which was covered with cotton wool and aluminium foil. In combination, these differences would result in longer processing times. A cost analysis for this scenario is provided in Table 7.2.

**Table 7.2** Cost analysis summary for Scenario 2: current framework minus the MWP component. These costs have been summarised from the more detailed analysis of Section C.5.3.

Cost category	Cost (£)
Inoculum preparation	94
Fermentation	637
Assay	252
Labour	4,260
<b>Total</b>	<b>5,243</b>

#### 7.2.4 Scenario 3: current framework minus the DoE component

In this scenario, the MWP experimentation component is applied but the DoE component is removed and the costs associated with this change are estimated. Here, there are no statistical tools available to enable the selection of a representative subset of experiments, and so all experimental combinations need to be performed in order to *locate with confidence* the optimum experimental region. A cost analysis for this scenario is provided in Table 7.3.

**Table 7.3** Cost analysis summary for Scenario 3: current framework minus the DoE component. These costs have been summarised from the more detailed analysis of Section C.5.4.

Cost category	Cost (£)
Inoculum preparation	1,987
Fermentation	62,222
Assay	135,266
Labour	984,150
<b>Total</b>	<b>1,183,625</b>

#### 7.2.5 Scenario comparison

Table 7.4 displays a summary of the costs associated with each scenario presented in the previous sections. Scenario one which represents the current framework is clearly the cheapest approach, in terms of both materials and labour, with an overall cost of £3,357. It is interesting to note here that the total cost of disposable MWPs (£256) contributes to over 98 % of the fermentation cost (Table C.9).

**Table 7.4** Summary of costs for each expression optimisation scenario. Details of each scenario are described in Sections 7.2.2 through 7.2.4 inclusive.

Cost Scenario	Associated costs		Total (£)
	Materials (£)	Labour (£)	
1 Current framework ( <i>CF</i> )	607	2,750	3,357
2 <i>CF</i> - MWP	983	4,260	5,243
3 <i>CF</i> - DoE	199,474	984,150	1,183,624

The overall cost associated with scenario two, in which the MWPs have been replaced with traditional SF fermentations, is approximately one and a half times larger at £5,243. This increased cost is mainly attributed to the increased labour requirements due to the lower processing rate of SFs vs microwells. Moreover, although the media requirements of scenario two are much greater than for scenario one, the media savings in scenario one are largely offset by the relatively high costs of MWPs (Table C.4). Furthermore, the costs of purchasing and cleaning the reusable SFs were ignored in scenario two in order to simplify calculations which, otherwise, would have increased the overall costs of scenario two.

Scenario three, a partial representation of the conventional HTES approach, has the largest overall cost by far, which at £1,183,624 is over 350 times higher than that of scenario one. In reality, this approach would not be attempted as the experimental workload is simply too high.

The purpose of including this scenario was to emphasise the importance of DoE. In scenario one, where DoE was used, only 110 experiments were performed (Section 4.6). In scenario three by contrast, approximately 60,000 experiments would be required (Section C.5.4) to obtain a *similar level of process characterisation*. This assumes that the optimal region lies within the experimental window; if not, the experiments would have to be repeated under new factor settings. It would also be difficult to analyse the results without using basic statistical tools, in which case, they could have been employed from the start in the form of DoE.

---

Overall the costs involved in protein characterisation are minimised when the current framework is applied. Interestingly, the major costs involved in this stage of the framework are attributed to the labour requirements and hence the overall experimental timeframes. By comparing the labour costs of scenarios two and three, it is possible to conclude that the application of DoE results in far greater time savings than that of MWP experimentation. In the following section, the long-term effect of these time savings on the profitability of a new drug is investigated.

### **7.3 Generation of protein samples**

#### **7.3.1 Background and general assumptions**

Unlike protein expression optimisation, the generation of sufficient quantities of protein for drug discovery is a common and indeed essential activity within industry. It presents a major bottleneck within drug-discovery (Stewart *et al.*, 2002) which, consequently, impacts negatively on the time to market for a new drug. Aside from the obvious human costs, the delayed launch of an important new drug can cost a company on average between \$1 million (Kerns, 2001; Martin, 2002) and \$15 million (Noffke, 2007) per drug, per day.

In the following exercise, the speed of protein generation is estimated for two different scenarios and the potential difference in sales revenue is calculated. The first scenario represents the current framework which consists of DoE/MWP protein expression optimisation followed by optimised fermentations within a laboratory scale (7.5 l) stirred-tank reactor (STR). The second scenario consists of parallel shake-flask fermentations at a single set of conditions, which presents perhaps the most common method of protein generation within drug-discovery (Kumar *et al.*, 2004). This exercise includes no estimation of materials and labour costs, as they are assumed to be insignificant in comparison to the daily sales revenue of a new drug. The general assumptions for both scenarios are:

- 24 h experimentation with instant turnaround;
- Rate of experimentation limited only by equipment availability (same as that of Astex Therapeutics Ltd.):
  - one HiGro incubator shaker;
  - three SF incubator shakers, each capable of housing eight 2 l SFs (working volume = 0.5 l);
  - four STRs (working volume = 5 l);
  - Protein sample requirements as in Table 7.5;

**Table 7.5** Typical protein sample requirements of drug discovery. Each stage of drug discovery aims to reduce the number of investigated protein constructs by identifying those which demonstrate several key properties. These include functional activity and ease of protein expression, purification and characterisation. Protein requirements may differ widely between different drugs and different drug discovery companies; the requirements shown here are representative of the operating ranges of Astex Therapeutics Ltd.

Drug discovery stage	Number of protein constructs		Mass of protein required per construct (mg)	
	Min.	Max.	Min.	Max.
1 Pre-screening (protein expression)	10	20	5	10
2 Pre-screening (purification + characterisation)	4	6	10	50
3 Screening, hits-to-leads and lead optimisation	1	1	1000	2000

### 7.3.2 Scenario 1: current framework

In this scenario, the protein constructs examined at each stage of drug discovery first undergo DoE/MWP protein expression optimisation studies as in Chapter 4. Protein samples of each construct are then generated at the STR scale under the optimised conditions as in Chapter 5. Key assumptions are:

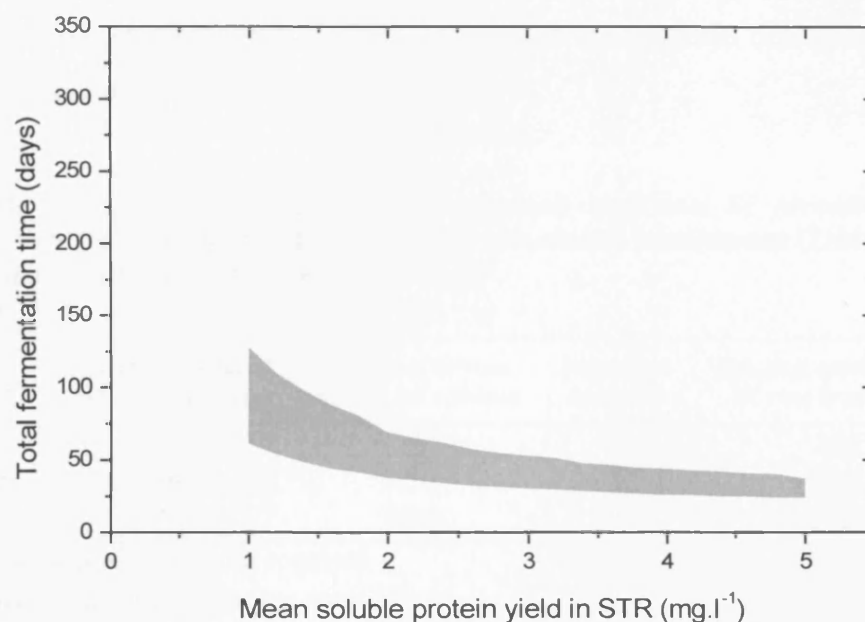
- DoE/MWP protein expression optimisation studies for all protein constructs are carried out in parallel;
- Minimum fermentation time for protein expression optimisation is same as that for the current framework, i.e. 275 h  $\approx$  11 days (Table C.9);
- Optimised protein expression levels are reproduced within the STR;
- Range of optimised soluble protein expression levels is same as that typically obtained at Astex Therapeutics Ltd.: 1-5 mg.l<sup>-1</sup>;
- STR fermentation time is same as the optimised fermentation time of the current framework, i.e. = 22 h (Section 4.5).

A sample calculation of the total fermentation time required by this scenario at a single set of conditions is provided in Table 7.6. This calculation was then repeated for the complete range of protein sample requirements (Table 7.5) and the resulting time window is illustrated in Figure 7.1 as a function of the mean soluble protein yield obtained in the STR (1-5 mg.l<sup>-1</sup>). The general trend observed here is that the total fermentation time required decreases with increasing optimised soluble protein expression levels.

**Table 7.6** Sample calculation for scenario one: current framework. Calculations based on average protein sample requirements (Table 7.5) and a mean soluble protein yield of 3 mg.l<sup>-1</sup>.

Stage	Protein requirement per construct (mg)	Number of STR runs required per construct	Number of constructs	Sub-total number of STR runs required
1	7.5	1	15	15
2	30	2	5	10
3	1500	100	1	100
Total number of STR runs required				125
Minimum number of STR batches required				32
Minimum MWP fermentation time (days)				11
Minimum STR fermentation time (days)				29
<b>Total fermentation time (days)</b>				<b>41</b>





**Figure 7.1** Fermentation time window for the current framework (DoE / MWP experimentation) over the full range of protein sample requirements (Table 7.5).

### 7.3.3 Scenario 2: traditional SF fermentation approach

In this scenario, there are no protein expression optimisation studies. Protein samples of each construct are generated immediately through SF fermentations at a single set of experimental conditions. Key assumptions are:

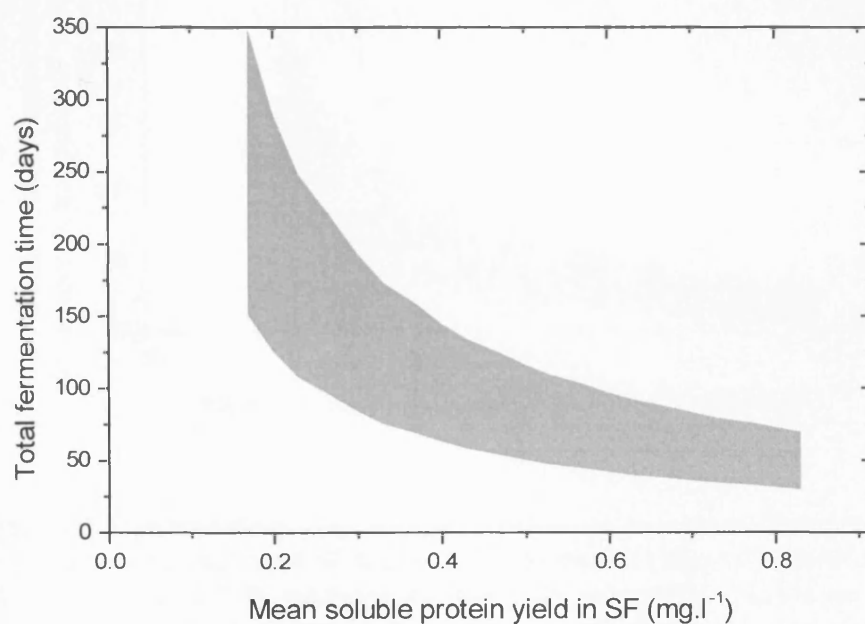
- SF fermentation time is the same as that used in the reference SF experiments (6 h 40 min) as described in Section 2.3.5;
- Ratio of optimised MWP protein yield to SF protein yield is the same as that obtained experimentally (Section 4.5), approximately 6:1, i.e. SF soluble protein yield range assumed to be: 0.17-0.83 mg.l<sup>-1</sup>.

A sample calculation of the total fermentation time required by this scenario at a single set of conditions is provided in Table 7.7. This calculation was then repeated for the complete range of protein sample requirements (Table 7.5) and

the resulting time window is illustrated in Figure 7.2 as a function of the mean soluble protein yield obtained in the SF ( $0.17\text{-}0.83\text{ mg.l}^{-1}$ ). Again, the general trend observed here is that the total fermentation time required decreases with increasing soluble protein expression levels.

**Table 7.7** Sample calculation for scenario two: traditional SF fermentation approach. Calculations based on average protein sample requirements (Table 7.5) and a mean soluble protein yield of  $0.5\text{ mg.l}^{-1}$ .

Stage	Protein requirement per construct (mg)	Number of SF runs required per construct	Number of constructs	Sub-total number of SF runs required
1	7.5	30	15	450
2	30	120	5	600
3	1500	6000	1	6000
Total number of SF runs required				7050
Total number of SF batches required				294
<b>Total fermentation time (days)</b>				<b>82</b>

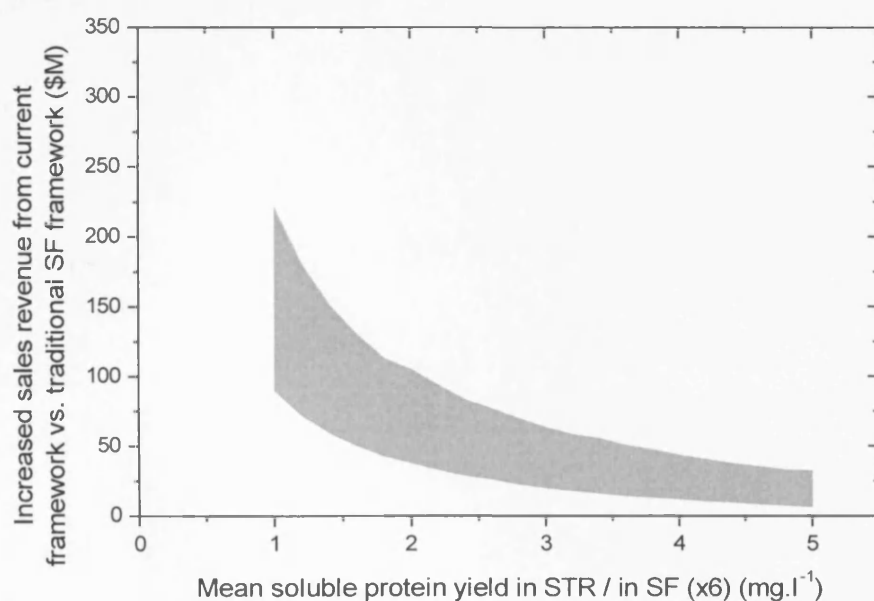


**Figure 7.2** Fermentation time window for the traditional SF fermentation framework over the full range of protein sample requirements (Table 7.5).

### 7.3.4 Scenario comparison

Direct comparison of Figures 7.1 and 7.2 illustrates that the total fermentation time of the current framework is generally less than that of the traditional SF framework, over the entire range of conditions tested (Table 7.5). The total fermentation time of the current framework is also less sensitive to decreasing protein yields. This implies that the current framework would provide the largest time savings when applied to proteins which express poorly.

Figure 7.3 illustrates the overall cost savings provided by the current framework relative to the traditional SF framework, in terms of increased drug sales revenue. Overall, the time savings provided by the current framework create the potential for significantly increased sales revenues.



**Figure 7.3** Increased drug sales revenue window of the current framework in comparison to the traditional SF framework over the full range of protein sample requirements (Table 7.5), assuming average daily sales of \$1 million per drug, per day. The results displayed in this figure are calculated by subtracting the boundary results illustrated in Figure 7.1 from those displayed in Figure 7.2.

## **7.4 Summary**

The aim of this chapter was to provide a thorough and quantitative analysis of the time and cost savings provided by the current framework over traditional alternatives. In Section 7.2 the main costs of protein expression optimisation were attributed to labour costs. The DoE component of the current framework was then identified as a key factor in significantly reducing overall experimental timeframes and hence labour costs. In Section 7.3 these time savings were also seen to apply to the larger scale generation of protein samples for drug discovery research. Within this context, the current framework creates the potential for significantly increased drug sales revenues by accelerating drug discovery. These benefits would be especially large when the target protein expresses poorly. In the following chapter, further potential benefits of the current framework, including the application towards manufacturing scale fermentation and bioprocess validation, are explored.

## **8 Development of the proposed framework to facilitate bioprocess validation<sup>†</sup>**

### **8.1 Aims and objectives**

Although the framework presented in this thesis has been optimised for application within drug discovery, it also provides a foundation for generating much larger quantities of protein. The key challenge here would be to maintain a fine degree of process control at the larger scales of fermentation and so the current framework would have to be developed towards achieving greater process understanding.

Moreover, the products of large scale fermentations are often used in pharmaceutical drugs where product quality is of paramount importance. The regulatory authorities would thus require documented evidence (bioprocess validation) of a well-defined manufacturing process which performs consistently within predetermined specifications (U.S. Food and Drug Administration, 1987). A thorough characterisation of the fermentation process during the early development stages would, therefore, greatly facilitate in the subsequent and mandatory validation process.

The aim of this chapter is to explore how the current framework may be developed towards this goal. The development approach is structured using a process improvement tool called Six Sigma and applied to a hypothetical scenario in which the MWP fermentation process described in Chapter 4 is to be scaled-up for manufacture of a therapeutic protein.

---

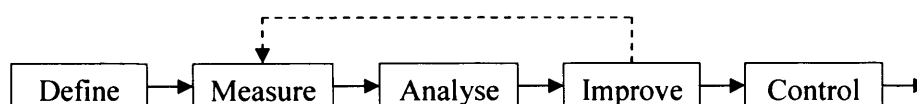
<sup>†</sup>The inclusion of this chapter relates to the University of London EngD requirement for consideration of the bioprocess validation implications of the research work performed.

## 8.2 Bioprocess development through Six Sigma (DMAIC)

### 8.2.1 Introduction

Six Sigma is a disciplined, data-driven methodology for reducing the number of defects in any process to 3.4 defects per million opportunities (DPMO) (Brue, 2002). A Six Sigma defect may be defined as any process output which lies outside of specifications.

Six Sigma has two key methodologies: DFSS and DMAIC. DFSS is an acronym for “Design For Six Sigma” and it is used to create new process designs. DMAIC is an acronym for “Define, Measure, Analyze, Improve and Control” and this methodology should be used when an existing product or process is not performing adequately. The flow of a typical DMAIC project is illustrated in Figure 8.1. The following section explores how the existing fermentation process may be developed, using the DMAIC methodology as a guide.



**Figure 8.1** Typical flow of a DMAIC project (Creveling, 2006). In practice, there may be significant overlap between each phase and an iterative approach to Six Sigma (indicated by the dashed line) may be required (Hayler and Nichols, 2005).

### 8.2.2 Definition phase

The purpose of this phase is to introduce the overall problem and define the project goals. As already mentioned in Section 8.1, a hypothetical scenario is presented in which the MWP fermentation process described in Chapter 4 is to be scaled-up for manufacture of a therapeutic protein. An appropriate strategic goal would thus be to achieve greater control of the MWP fermentation process

in order to facilitate the subsequent scale-up and validation of the process. Corresponding process and project goals are proposed in Table 8.1.

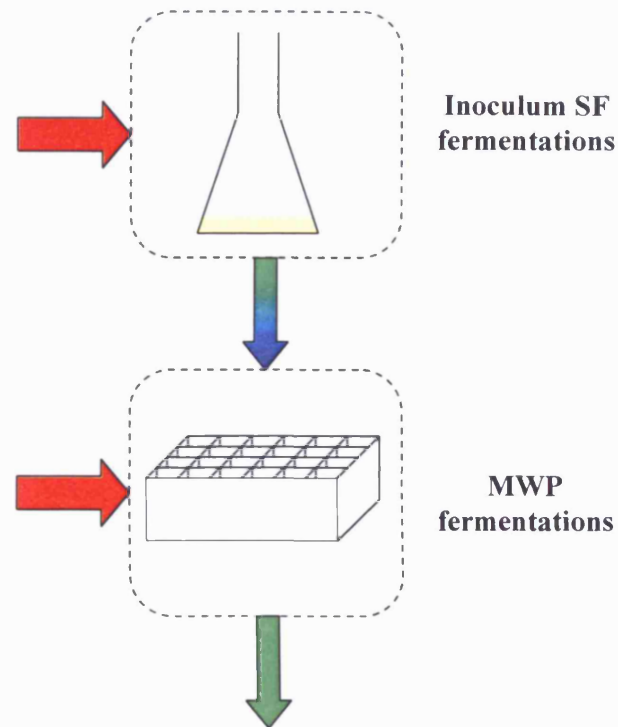
The overall scope of the project is limited to the development of the MWP fermentations and the upstream inoculum shake-flask (SF) fermentations (Figure 8.2). All analytical methods and all processes upstream of the inoculum preparation are assumed to be already developed / validated.

**Table 8.1** Statement of goals at the strategic, process and project levels.

Strategic goal	Process goals	Project goals
To achieve greater control of the MWP fermentation process in order to facilitate the subsequent scale-up and validation of the process.	<ul style="list-style-type: none"> <li>• Increased characterisation</li> </ul>	<ul style="list-style-type: none"> <li>• Identification of more responses</li> </ul>
	<ul style="list-style-type: none"> <li>• Increased reproducibility and predictivity</li> </ul>	<ul style="list-style-type: none"> <li>• All response measurements lie within 2% of predicted levels and/or optimisation <math>Q^2</math> for all fermentation response measurements <math>&gt; 0.9</math></li> </ul>
	<ul style="list-style-type: none"> <li>• Increased robustness</li> </ul>	<ul style="list-style-type: none"> <li>• Robustness <math>Q^2</math> for all fermentation response measurements <math>&lt; 0.1</math></li> </ul>

### 8.2.3 Measurement phase

The goal of this phase is to measure the current performance of the process which consists here of the inoculum SF and MWP fermentations. Key process responses should be identified (Table 8.2) and subsequently measured over an appropriate set of characterisation experiments. In order to obtain the relevant  $Q^2$  values (Table 8.1), these experiments should be organised through DoE. The standard methods and conditions for inoculum preparation (Section 2.2.3) and those optimised for MWP fermentations (Chapter 4) would serve as references here. It should then be possible to define a process defect and calculate the DPMO of the current process.



**Figure 8.2** Process map illustrating the overall scope of the Six Sigma project, complete with the flow of process steps, factors and responses. The green arrows represent responses and the red and blue arrows represent controlled and uncontrolled factors, respectively; response measurements from the inoculum SF fermentation will serve as uncontrolled factors in the MWP fermentation step.

**Table 8.2** List of key responses which may be used to measure process performance.

Biological responses	Engineering responses
<ul style="list-style-type: none"> <li>• biomass growth</li> <li>• cell viability</li> <li>• soluble protein yield</li> <li>• metabolite consumption</li> <li>• pH, DOT</li> </ul>	<ul style="list-style-type: none"> <li>• <math>k_{La}</math></li> <li>• power per unit volume</li> <li>• mixing / circulation time</li> </ul>



### 8.2.4 Analysis phase

At this stage of the Six Sigma process, the gaps between current performance and project goal performance are quantified, and potential root causes of underperformance are identified. Various tools, such as the *Ishikawa* diagram (Brue, 2002; Creveling, 2006; Hayler and Nichols, 2005) are available at this stage to aid in root cause identification. Sources of variation to which future inoculum SF experiments may be susceptible are described in Table 8.3. Tables 8.4 and 8.5 list root causes for the variation / underperformance already observed from the MWP experiments and DOE analysis (Chapter 4), respectively.

**Table 8.3** List of the sources of variation to which inoculum SF experiments may be susceptible with corresponding suggested control measures.

Root causes of variation	Suggested control measures
<ul style="list-style-type: none"> <li>• Variable medium quality;</li> <li>• Un-optimised fermentation performance;</li> <li>• Human error;</li> <li>• Variable incubator-shaker performance.</li> </ul>	<ul style="list-style-type: none"> <li>• Use a defined medium;</li> <li>• Perform DoE and include all key fermentation responses:               <ul style="list-style-type: none"> <li>○ Screen all critical factors – include all relevant factors from DoE screen (Table 2.2);</li> <li>○ Determine optimal factor ranges;</li> <li>○ Determine acceptable response ranges;</li> </ul> </li> <li>• Implement robotic automation where possible;</li> <li>• Monitor incubator performance.</li> </ul>

### 8.2.5 *Improvement phase*

Here, all the root causes of process underperformance identified during the Analysis phase should be controlled through implementation of the suggested improvement measures. Thus, new rounds of DoE-driven experimentation are required to optimise each stage of the process in sequence.

Once all improvements have been implemented, process performance should be re-evaluated and further improvements made if necessary. An iterative approach to the Measure-Analyse-Improve phases may, therefore, be required until all project goals (Table 8.1) are met as illustrated in Figure 8.1.

**Table 8.4** Possible sources of observed variation in the MWP fermentation performance (Chapter 4) with corresponding suggested control measures.

Root causes of variation	Suggested control measures
<ul style="list-style-type: none"> <li>• Variable inoculum quality;</li> </ul>	<ul style="list-style-type: none"> <li>• Use improved inoculum;</li> </ul>
<ul style="list-style-type: none"> <li>• Variable medium quality;</li> </ul>	<ul style="list-style-type: none"> <li>• Use a defined medium;</li> </ul>
<ul style="list-style-type: none"> <li>• Human error;</li> </ul>	<ul style="list-style-type: none"> <li>• Implement robotic automation where possible;</li> </ul>
<ul style="list-style-type: none"> <li>• Variable incubator-shaker performance.</li> </ul>	<ul style="list-style-type: none"> <li>• Monitor incubator performance.</li> </ul>

**Table 8.5** Possible sources of observed underperformance in the DoE analysis (Section 4.5) with corresponding suggested improvements.

Root causes of underperformance	Suggested improvements
<ul style="list-style-type: none"> <li>• Hidden effect of unidentified factor(s).</li> </ul>	<ul style="list-style-type: none"> <li>• Include responses from previous process step as uncontrolled factors;</li> <li>• Conduct further scouting experiments;</li> </ul>
<ul style="list-style-type: none"> <li>• Factor ranges too large;</li> </ul>	<ul style="list-style-type: none"> <li>• Establish narrower factor ranges;</li> </ul>
<ul style="list-style-type: none"> <li>• Model lack of fit;</li> </ul>	<ul style="list-style-type: none"> <li>• Remove qualitative factors from screen;</li> <li>• Optimise fermentation responses using a cubic model if necessary;</li> </ul>
<ul style="list-style-type: none"> <li>• Robustness of process uncertain</li> </ul>	<ul style="list-style-type: none"> <li>• Perform robustness testing around key factors and responses.</li> </ul>

### 8.2.6 Control phase

The goal of this phase is to control future process performance. All operating procedures associated with the new process should be documented and a monitoring and control plan should be implemented for sustaining improvements (short and long-term). Since small-scale processes are usually part of a short-term development programme, application of this phase may be more appropriate to the large-scale manufacturing process.

## 8.3 Summary

The aim of this chapter was to explore how the framework presented in this thesis could be developed to facilitate bioprocess validation. The need for greater process understanding and control was immediately identified and the principles of Six Sigma were applied to the current framework towards this goal. Using this methodology various framework improvements were proposed in the form of

additional key responses (Table 8.2) and control measures against sources of process variation (Tables 8.3 and 8.4) and DoE analysis underperformance (Table 8.5). Overall, the suggested improvements should greatly facilitate the subsequent hypothetical scale-up and validation of the process. In the following chapter, the main findings of this thesis are summarised together with suggestions for future work.

---

## 9 Conclusions and future work

### 9.1 Conclusions

At present, a major bottleneck within drug discovery involves the provision of sufficient quantities of recombinant protein for high-throughout screening (Stewart *et al.*, 2002). This thesis has established a practical and generic framework to address this challenge which underpins the generation of large yields of recombinant protein in *E. coli* in a rapid and cost-effective manner.

The underlying design of the framework was first established in Chapter 3 (Figure 3.1). This consisted of two principle stages of experimentation: (1) small scale protein expression characterisation followed by (2) scale-up of optimised protein expression. A wide range of variables was known to affect heterologous protein expression (Section 1.2.3). Consequently, during the initial characterisation stage, Design of Experiments (DoE) and experiments in microwell plate (MWP) formats were used in order to minimise the experimental costs and timeframes. A key requirement of stage one was to provide insight into larger scales of performance (Fernandes and Cabral, 2006). Thus, in contrast to previous studies (Section 1.3.4), experiments were performed within a defined engineering environment through the manipulation of factors such as MWP shaking speed and fill volume. The information generated from stage one of the framework then aided in the identification of key scale-up parameters and the subsequent design of a scale-up strategy. In order to meet the varying protein demands of different drug-discovery campaigns, separate strategies for laboratory scale and pilot scale protein expression were established.

In Chapter 4 the overall methodology for rapid protein expression characterisation was refined through practical experimentation and DoE was shown to lead to the rapid optimisation of culture conditions. A mixture of ten biological and engineering variables was initially chosen for investigation (Table 4.1). After only three rounds of DoE, the protein induction period and shaking speed were identified as key factors (Figure 4.3) and a predictive model of

---

protein expression was developed (Equation 4.1). In comparison to other high-throughput protein expression (HTPE) approaches, the application of DoE led to a significant decrease (up to 98 %) in the number of characterisation experiments required. At each stage of the DoE process a greater understanding of the expression system was achieved and, overall, the soluble protein yield was increased approximately 9-fold over reference shake-flask conditions (Figure 4.8). The use of parallel MWP fermentations further helped to reduce the experimental timeframes. As mentioned previously, the shaking speed, which is known to strongly affect the oxygen mass transfer into microwells (Hermann *et al.*, 2003), was identified as a key factor of protein expression. This suggested that careful control of agitation and aeration conditions would be important in maintaining optimal protein expression levels upon scale-up.

In Chapter 5 the oxygen mass transfer coefficient,  $k_{La}$ , was shown to be a suitable basis for scale-up and microwell results were accurately reproduced at the laboratory scale. Here, MWP fermentations were scaled-up to a 7.5 l stirred-tank reactor (STR) over a range of matched  $k_{La}$  values. MWP fermentation kinetic profiles and yields were accurately reproduced at a 'high' matched  $k_{La}$  value of  $247 \text{ h}^{-1}$ , equivalent to a 1,700 fold scale translation (Figure 5.9 (a)). This result confirmed the initial DoE-driven hypothesis that  $k_{La}$  was an important scale-up parameter and in so doing, demonstrated the importance of DoE within the current framework. This result also confirmed a number of recent studies which have suggested  $k_{La}$  to be an important scale-up parameter for aerobic fermentations (Ferreira-Torres *et al.*, 2005; Micheletti *et al.*, 2006). In contrast, scale-up of MWP fermentations at a 'low' matched  $k_{La}$  value of  $55 \text{ h}^{-1}$  did not provide reproducible performance at the 7.5 l scale. This difference was attributed to the poor gas-liquid distributions observed within the larger vessel at the 'low'  $k_{La}$  conditions.

This information was exploited in Chapter 6 to develop a scale-up strategy for translating MWP fermentation performance to pilot scale where  $k_{La}$  was again shown to be a suitable basis for scale-up. Here, MWP fermentations were scaled-up to a 75 l stirred-tank reactor (STR) on the basis of a 'high' matched  $k_{La}$  value

---

of  $247 \text{ h}^{-1}$ . Initial experiments showed a difference in fermentation performance between the MWP and 75 l scales. This was attributed to differences in the method of medium sterilisation between the two vessels (Section 2.4) and hence the different temperature-time profiles achieved. The scale-up strategy was thus further refined by first sterilising all fermentation media within the STR. This subsequently led to the accurate reproduction of MWP fermentation kinetic profiles at the 75 l scale (Figure 6.6). No previous studies have translated MWP fermentation performance to this scale, equivalent to a 15,000 fold scale-up. In so doing, this result clearly demonstrates the validity and efficacy of the overall framework for underpinning the generation of large yields of recombinant protein in a rapid and cost-effective manner.

Chapter 7 presented a thorough and quantitative analysis of the inherent time and cost savings provided by the current framework in which the DoE component was shown to be of principal benefit. Overall, in comparison to the traditional approach for protein sample generation the current framework was predicted to provide significantly increased drug sales revenues by accelerating drug discovery. These benefits would be especially large when the target protein expresses poorly.

Finally, Chapter 8 explored how this framework could be developed towards the generation of manufacturing scale quantities of protein and validation of the subsequent process. The need for greater process understanding and control was immediately identified and various minor framework improvements were proposed towards this goal in the form of additional key responses (Table 8.2), control measures against sources of process variation (Tables 8.3 and 8.4) and measures to improve DoE analysis performance (Table 8.5). Overall, the suggested improvements would greatly enhance process control, thereby widening the applications of the framework to include the production of larger quantities of protein and subsequent bioprocess validation.

## 9.2 Future Work

Throughout the course of this research, several areas for further work were identified. Based upon the experimental results obtained in this work, immediate follow-up studies should address the following points:

- The scale-up of protein expression within this framework was achieved on the basis of a single, high matched  $k_La$  value. Further work is thus necessary to establish more accurately the *range* of  $k_La$  values over which MWP fermentation kinetics can be reliably scaled-up;
- The cause(s) of the observed lack-of-fit of the optimisation model (Section 4.5) should be investigated. Analysis of why some MWP experiments yielded unusually low protein yields and why other experiments resulted in high degrees of cell lysis (Section 4.5) should facilitate this investigation. Ultimately, the process may have been affected by unidentified factors or the identified factor ranges may have been unsuitably high;
- Within the current work,  $k_La$  values were measured in cell-free media under static conditions. Measurements of  $k_La$  values within live media over the course of a fermentation would provide greater insight into the oxygen mass transfer characteristics of the process which should in turn facilitate scale-up;
- The underlying cause of why initial experiments within the MWP produced lower final biomass concentrations than those in the corresponding 75 l STR fermentations should be further studied.

Based on the wider considerations of this framework the following points could also be addressed:



- 
- Further characterisation of the engineering environment within the MWP and STRs through, for example, mixing time studies, towards the visualisation of gas-liquid dispersion;
  - Having reproduced near-optimal MWP protein expression levels at laboratory and pilot scales, further optimisation of protein expression could be performed at these larger scales. This may, for example, be accomplished through additional applications of DoE combined with the implementation of fed-batch culture to overcome the decrease in soluble protein yield once the glycerol supply is exhausted (Figures 5.5, 5.7, 6.4 and 6.5);
  - The key principles of the current framework are also likely to apply to other expression hosts, recombinant proteins and expression mechanisms such as extracellular expression. In this case, further work would be required to identify additional key factors, responses and experimental techniques involved in small-scale protein expression and the subsequent scale-up strategy would have to be refined as appropriate;
  - The level of process understanding and control provided by the current framework could also be easily developed. Chapter 8 describes in detail simple improvements to fermentation and DoE methods towards these goals. Ultimately, the improved framework could be applied to the rapid generation of manufacturing scale quantities of protein in addition to the subsequent validation of such a process.

## Appendix A – Bioreactor mixing data

**Table A.1** MWP  $k_{La}$  values ( $\text{h}^{-1}$ ) measured as described in Section 2.4.1.

Plate geometry	Shaking speed (rpm)				
	100	200	300	400	500
P1	$0.7 \pm 0.2$	$3.2 \pm 0.1$	$22.7 \pm 1.5$	$59.5 \pm 3.9$	$88.9 \pm 6.4$
P2	$1.6 \pm 0.4$	$5.1 \pm 0.6$	$29.0 \pm 5.1$	$95.5 \pm 21.2$	$187.6 \pm 20.6$
P3	$2.1 \pm 0.2$	$5.0 \pm 0.7$	$47.7 \pm 8.0$	$148.0 \pm 7.0$	$241.3 \pm 28.7$

**Table A.2** MWP  $a_f/a_i$  values (dimensionless) measured as described in Section 2.4.2. NA indicates data not obtainable.

Plate geometry	Shaking speed (rpm)				
	100	200	300	400	500
P1	$0.89 \pm 0.15$	$0.99 \pm 0.15$	$1.21 \pm 0.31$	$2.06 \pm 0.20$	NA
P2	$0.93 \pm 0.03$	$0.97 \pm 0.02$	$1.17 \pm 0.07$	$1.63 \pm 0.19$	$2.22 \pm 0.07$
P3	$0.95 \pm 0.02$	$0.97 \pm 0.03$	$1.28 \pm 0.04$	$2.10 \pm 0.05$	$2.58 \pm 0.11$

**Table A.3** STR  $k_{La}$  values ( $\text{h}^{-1}$ ) measured as described in Section 2.4.3.

STR size	Agitation speed (rpm)					
	200	256	400	600	800	937
7.5 l	$9.7 \pm 0.6$		$167.5 \pm 12.0$	$184.1 \pm 17.9$	$272.1 \pm 32.9$	
75 l		$48.4 \pm 1.8$	$129.1 \pm 3.6$	$211.1 \pm 14.5$	$346.7 \pm 22.3$	$384.9 \pm 23.9$

**Table A.4** STR gassed-power,  $P_g$ , values ( $\text{W}\cdot\text{m}^{-3}$ ) calculated using Equation 5.6

STR size	Agitation speed (rpm)					
	200	256	400	600	800	937
7.5 l	0.23		1.66	5.26	11.95	
75 l		7.6	26.8	84.4	199.0	311.6

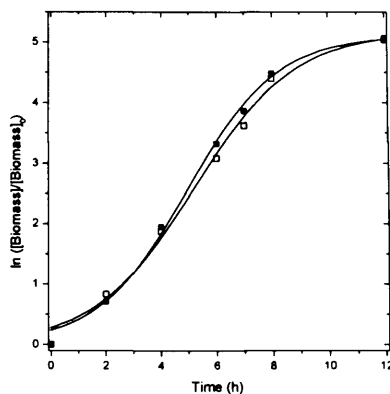
## Appendix B - Fermentation response modelling

### B.1 Basis

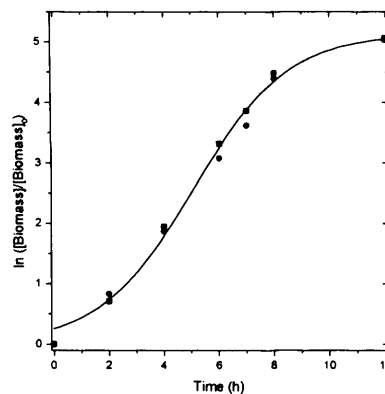
All statistical modelling and analytical methods are described in Section 2.5.3.

### B.2 MWP and 7.5 l STR: [Biomass] comparison

Equation of model used:  $\ln\left(\frac{y}{y_0}\right) = \frac{a}{1 + e^{-b(t-c)}}$  (logistic curve)



**Figure B.1** Logistic curves fitted separately to MWP [biomass] data (■) and 7.5 l STR [biomass] data (□).



**Figure B.2** Single logistic curve fitted to combined MWP and 7.5 l STR [biomass] data.

**Table B.1** Summary of model parameters correct to 3 s.f.

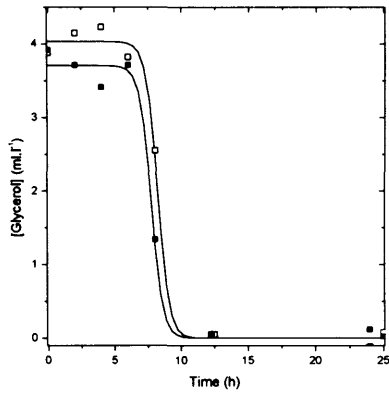
	Parameter	Value	Std. Error
MWP	<i>a</i>	5.11	0.158
	<i>b</i>	4.96	0.180
	<i>c</i>	0.615	0.0510
7.5 l	<i>a</i>	5.17	0.224
	<i>b</i>	5.21	0.257
	<i>c</i>	0.551	0.0589
Comb.	<i>a</i>	5.14	0.130
	<i>b</i>	5.08	0.149
	<i>c</i>	0.581	0.0378

**Table B.2** Summary of modelling statistics correct to 3 s.f.

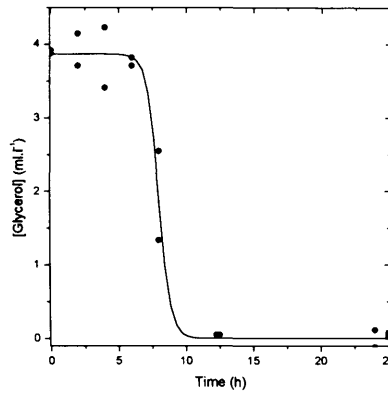
RSS <sub>sep</sub>	0.231
RSS <sub>comb</sub>	0.294
DF <sub>sep</sub>	8
DF <sub>comb</sub>	11
<i>F</i>	0.733
<i>p</i>	<b>0.561</b>

### B.3 MWP and 7.5 l STR: [Glycerol] comparison

Equation of model used:  $y = \frac{a}{1 + 10^{(t - \log b)}}$  (one site competition curve)



**Figure B.3** One site competition curves fitted separately to MWP [glycerol] data ( ■ ) and 7.5 l STR [glycerol] data ( □ ).



**Figure B.4** Single one site competition curve fitted to combined MWP and 7.5 l STR [glycerol] data.

**Table B.3** Summary of model parameters correct to 3 s.f.

	Parameter	Value	Std. Error
MWP	<i>a</i>	3.70	0.0814
	<i>log b</i>	7.76	0.0833
7.5 l	<i>a</i>	4.03	0.0748
	<i>log b</i>	8.24	0.0726
Comb.	<i>a</i>	3.86	0.106
	<i>log b</i>	8.00	0.0977

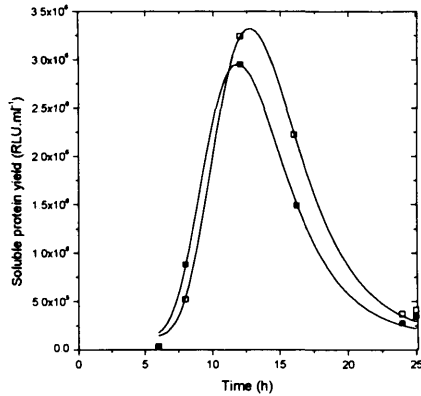
**Table B.4** Summary of modelling statistics correct to 3 s.f.

RSS <sub>sep</sub>	0.288
RSS <sub>comb</sub>	1.233
DF <sub>sep</sub>	12
DF <sub>comb</sub>	14
<i>F</i>	<b>19.7</b>
<i>p</i>	<b>1.63E-04</b>

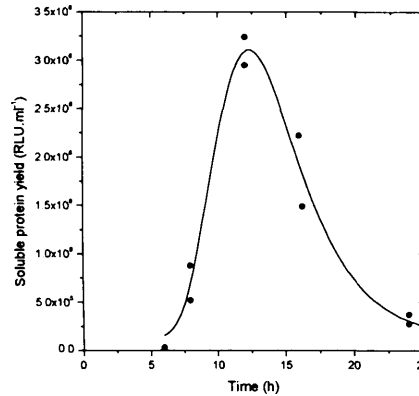
### B.4 MWP and 7.5 l STR: Soluble protein yield comparison

Equation of model used:  $y = y_0 + c \exp(-\exp(-d) - d + 1)$

where  $d = \frac{(t - a)}{b}$  (extreme peak function curve)



**Figure B.5** Extreme peak function curves fitted separately to MWP soluble protein yield data (■) and 7.5 l STR soluble protein yield data (□).



**Figure B.6** Single extreme peak function curve fitted to combined MWP and 7.5 l STR soluble protein yield data.

**Table B.5** Summary of model parameters correct to 3 s.f.

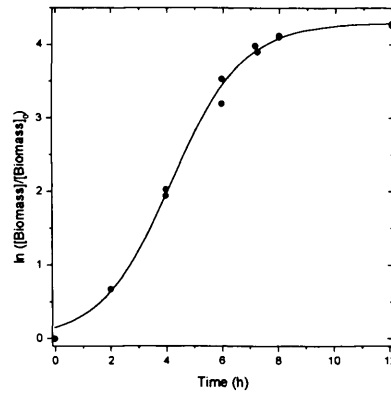
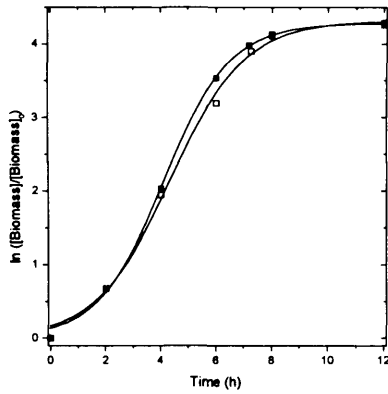
	Parameter	Value	Std. Error
MWP	$y_0$	1.29E+05	1.02E+05
	$a$	11.8	0.222
	$b$	2.96	0.230
	$c$	2.82E+06	1.64E+05
7.5 l	$y_0$	1.32E+05	9.14E+04
	$a$	12.7	0.170
	$b$	3.09	0.209
	$c$	3.19E+06	1.32E+05
Comb.	$y_0$	1.27E+05	1.16E+05
	$a$	12.3	0.236
	$b$	3.07	0.254
	$c$	2.98E+06	1.76E+05

**Table B.6** Summary of modelling statistics correct to 3 s.f.

$RSS_{sep}$	6.59E+10
$RSS_{comb}$	3.77E+11
$DF_{sep}$	4
$DF_{comb}$	8
$F$	4.72
$p$	<b>0.0811</b>

### B.5 MWP and 75 l STR: [Biomass] comparison

Equation of model used:  $\ln\left(\frac{y}{y_0}\right) = \frac{a}{1 + e^{-b(t-c)}}$  (logistic curve)



**Figure B.7** Logistic curves fitted separately to MWP [biomass] data ( ■ ) and 75 l STR [biomass] data ( □ ).

**Figure B.8** Single logistic curve fitted to combined MWP and 75 l STR [biomass] data.

**Table B.7** Summary of model parameters correct to 3 s.f.

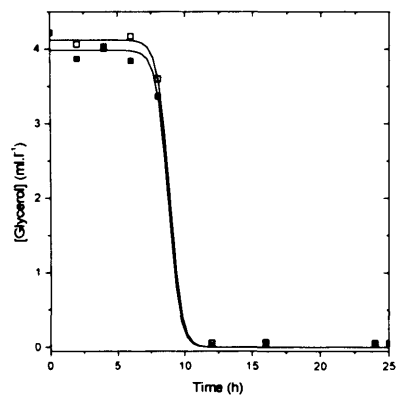
	Parameter	Value	Std. Error
MWP	<i>a</i>	4.28	0.0577
	<i>b</i>	4.12	0.0803
	<i>c</i>	0.836	0.0460
75 l	<i>a</i>	4.31	0.110
	<i>b</i>	4.35	0.153
	<i>c</i>	0.740	0.0654
Comb.	<i>a</i>	4.29	0.0653
	<i>b</i>	4.23	0.0910
	<i>c</i>	0.784	0.0447

**Table B.8** Summary of modelling statistics correct to 3 s.f.

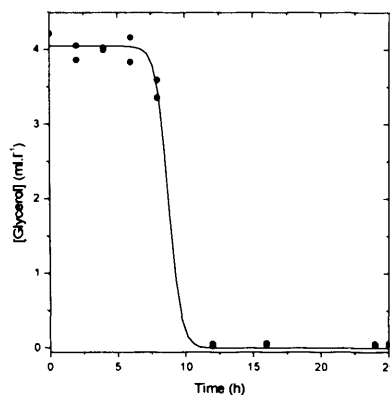
RSS <sub>sep</sub>	0.082
RSS <sub>comb</sub>	0.131
DF <sub>sep</sub>	8
DF <sub>comb</sub>	11
<i>F</i>	1.60
<i>p</i>	<b>0.264</b>

## B.6 MWP and 75 l STR: [Glycerol] comparison

Equation of model used:  $y = \frac{a}{1 + 10^{(t - \log b)}}$  (one site competition curve)



**Figure B.9** One site competition curves fitted separately to MWP [glycerol] data (■) and 75 l STR [glycerol] data (□).



**Figure B.10** Single one site competition curve fitted to combined MWP and 75 l STR [glycerol] data.

**Table B.9** Summary of model parameters correct to 3 s.f.

	Parameter	Value	Std. Error
MWP	$a$	3.98	0.0588
	$\log b$	8.73	0.105
75 l	$a$	4.12	0.038
	$\log b$	8.84	0.0792
Comb.	$a$	4.05	0.0400
	$\log b$	8.78	0.0767

**Table B.10** Summary of modelling statistics correct to 3 s.f.

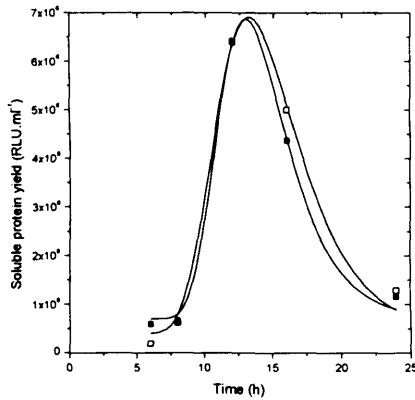
$RSS_{\text{sep}}$	0.136
$RSS_{\text{comb}}$	0.203
$DF_{\text{sep}}$	14
$DF_{\text{comb}}$	16
$F$	3.438
$p$	<b>0.061</b>



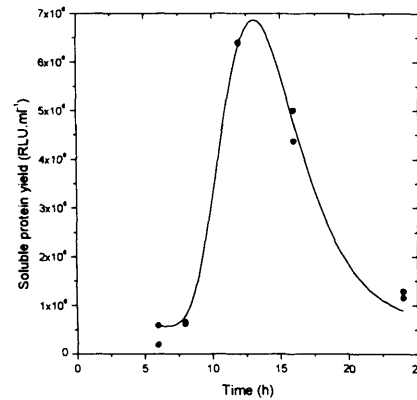
## B.7 MWP and 75 l STR: Soluble protein yield comparison

Equation of model used:  $y = y_0 + c \exp(-\exp(-d) - d + 1)$

where  $d = \frac{(t-a)}{b}$  (extreme peak function curve)



**Figure B.11** Extreme peak function curves fitted separately to MWP soluble protein yield data (■) and 75 l STR soluble protein yield data (□).



**Figure B.12** Single extreme peak function curve fitted to combined MWP and 75 l STR soluble protein yield data.

**Table B.11** Summary of model parameters correct to 3 s.f.

	Parameter	Value	Std. Error
MWP	$y_0$	6.95E+05	2.77E+05
	$a$	13.0	0.434
	$b$	2.51	0.775
	$c$	6.17E+06	6.65E+05
7.5 l	$y_0$	3.95E+05	4.70E+05
	$a$	13.2	0.354
	$b$	3.03	0.672
	$c$	6.51E+06	5.85E+05
Comb.	$y_0$	5.60E+05	2.07E+05
	$a$	13.1	0.195
	$b$	2.77	0.374
	$c$	6.31E+06	2.98E+05

**Table B.12** Summary of modelling statistics correct to 3 s.f.

$RSS_{sep}$	3.38E+11
$RSS_{comb}$	6.14E+11
$DF_{sep}$	2
$DF_{comb}$	6
$F$	0.410
$p$	<b>0.797</b>

## Appendix C – Experimental costs and timeframes

### C.1 Cost of consumables

All media costs shown below include the cost of an average addition of 500  $\mu\text{M}$  (= 0.12  $\text{g.l}^{-1}$ ) IPTG for all fermentations. All components were purchased from Sigma-Aldrich Chemical Company (Dorset, UK) unless otherwise stated in Chapter 2.

**Table C.1** Cost of LB media

<i>Component</i>	<i>Concentration</i>	<i>Unit cost (£)</i>	<i>Unit quantity</i>	<i>Cost (£.l<sup>-1</sup>)</i>
Yeast extract	5 $\text{g.l}^{-1}$	126.50	1000 g	0.63
Tryptone	10 $\text{g.l}^{-1}$	24.10	1000 g	0.24
NaCl	10 $\text{g.l}^{-1}$	16.40	1000 g	0.16
Kanamycin	30 $\text{mg.l}^{-1}$	59.60	5000 mg	0.36
IPTG	0.12 $\text{g.l}^{-1}$	216.60	10 g	2.58
				<b>3.98</b>

**Table C.2** Cost of TB media

<i>Component</i>	<i>Concentration</i>	<i>Unit cost (£)</i>	<i>Unit quantity</i>	<i>Cost (£.l<sup>-1</sup>)</i>
Yeast extract	24 $\text{g.l}^{-1}$	126.50	1000 g	3.04
Tryptone	12 $\text{g.l}^{-1}$	24.10	1000 g	0.29
Glycerol	4 $\text{ml.l}^{-1}$	29.80	1000 ml	0.12
$\text{KH}_2\text{PO}_4$	2.31 $\text{g.l}^{-1}$	25.00	1000 g	0.06
$\text{K}_2\text{HPO}_4$	12.54 $\text{g.l}^{-1}$	76.40	2500 g	0.38
Kanamycin	30 $\text{mg.l}^{-1}$	59.60	5000 mg	0.36
IPTG	0.12 $\text{g.l}^{-1}$	216.60	10 g	2.58
<b>Total</b>				<b>6.82</b>

**Table C.3** Cost of GM9Y media

<i>Component</i>	<i>Concentration</i>	<i>Unit cost (£)</i>	<i>Unit quantity</i>	<i>Cost (£.l<sup>-1</sup>)</i>
GM9Y	15.5 $\text{g.l}^{-1}$	25.69	500 g	0.80
Kanamycin	30 $\text{mg.l}^{-1}$	59.60	5000 mg	0.36
IPTG	0.12 $\text{g.l}^{-1}$	216.60	10 g	2.58
<b>Total</b>				<b>3.73</b>

**Table C.4** Cost of MWP's including cost of breathable membranes (£0.83 per membrane)

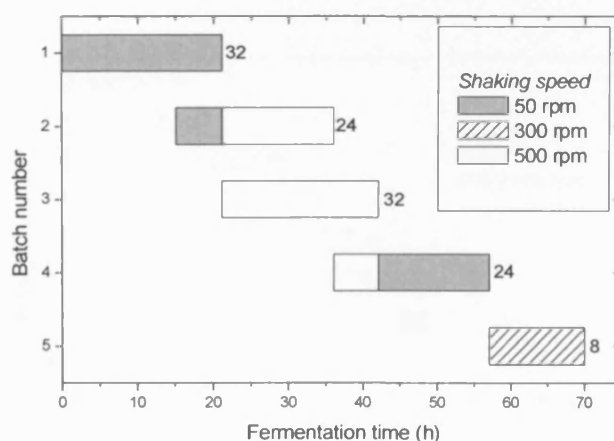
<i>MWP geometry</i>	<i>Unit cost (£)</i>	<i>Unit quantity</i>	<i>Cost per MWP</i>
P1	220.00	25	<b>9.63</b>
P2	220.00	25	<b>9.63</b>
P3	74.86	25	<b>3.82</b>

**Table C.5** Cost of luciferase assay buffer (LAB)

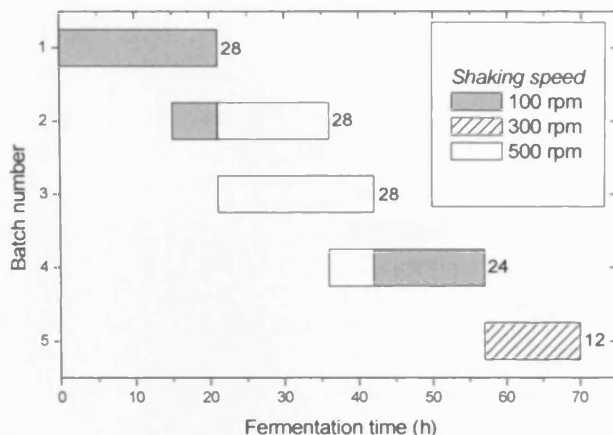
Component	Concentration	Unit cost (£)	Unit quantity	Cost (£.l <sup>-1</sup> )
MgSO <sub>4</sub>	1.20 g.l <sup>-1</sup>	15.80	100 g	0.19
Tris	1.21 g.l <sup>-1</sup>	15.60	100 g	0.19
ATP	0.28 g.l <sup>-1</sup>	38.60	1 g	10.64
Luciferin	0.13 g.l <sup>-1</sup>	643.90	0.05 g	1696.68
Co-A	0.22 g.l <sup>-1</sup>	424.70	0.1 g	928.33
DTT	4.86 g.l <sup>-1</sup>	11.70	0.25 g	227.40
Total				2863.42
<b>Per 200µl assay (Section 2.6.3)</b>				<b>0.57</b>

## C.2 Optimal schedules for the theoretical MWP fermentation scenarios (Sections 7.2.2 and 7.2.4)

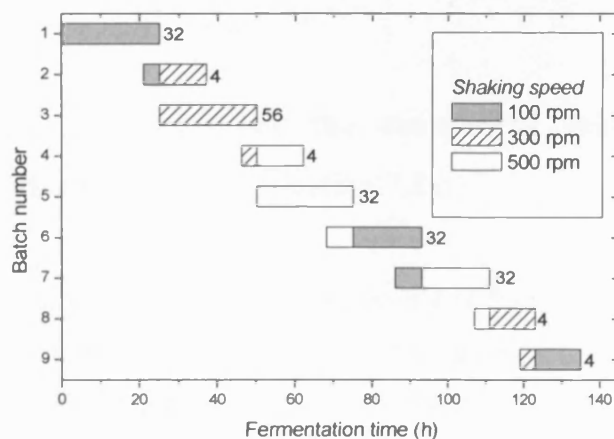
The following schedules assume 24 h experimentation, instant turnaround and a maximum processing throughput of 72 wells in parallel. These schedules also account for the fact that the HiGro incubator shaker can only be operated at one shaking speed at a time.



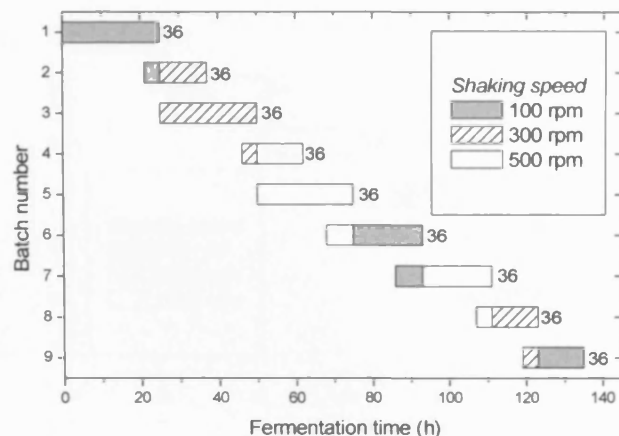
**Figure C.1** An optimal schedule for the MWP familiarisation experiments described in Table 2.3, showing minimum theoretical number of hours required to complete all fermentations. Experiments from Table 2.3 are arranged in fermentation batches according to those with identical pre-I and post-I shaking speeds. Numbers to the right of each bar represent number of fermentations per batch (including replicates).



**Figure C.2** An optimal schedule for the MWP screening experiments described in Table 4.2, showing minimum theoretical number of hours required to complete all fermentations. Experiments from Table 4.2 are arranged in fermentation batches according to those with identical pre-I and post-I shaking speeds. Numbers to the right of each bar represent number of fermentations per batch (including replicates).



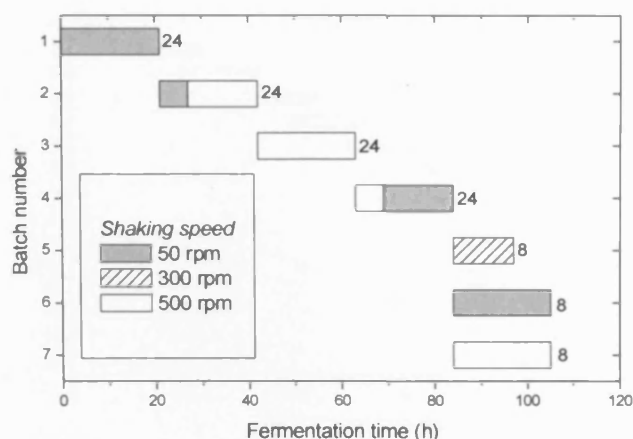
**Figure C.3** An optimal schedule for the MWP optimisation experiments described in Table 4.3, showing minimum theoretical number of hours required to complete all fermentations. Experiments from Table 4.3 are arranged in fermentation batches according to those with identical pre-I and post-I shaking speeds. Numbers to the right of each bar represent number of fermentations per batch (including replicates).



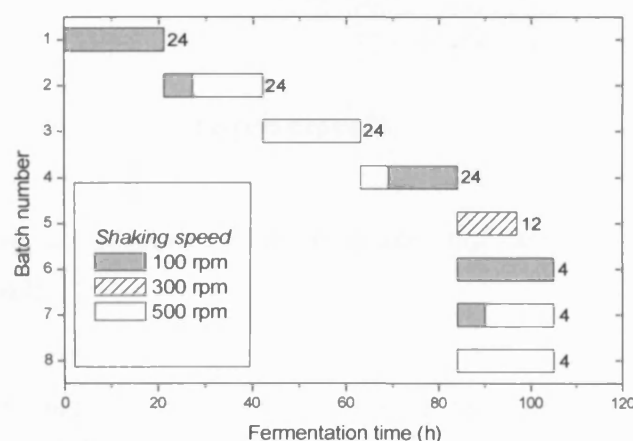
**Figure C.4** An optimal schedule for the MWP optimisation experiments, identical to that of Figure C.3, in which the *maximum* number of fermentations is performed (72 wells in parallel). Numbers to the right of each bar represent number of fermentations per batch (including replicates).

### C.3 Optimal schedules for the theoretical shake-flask (SF) fermentation scenario (Section 7.2.3)

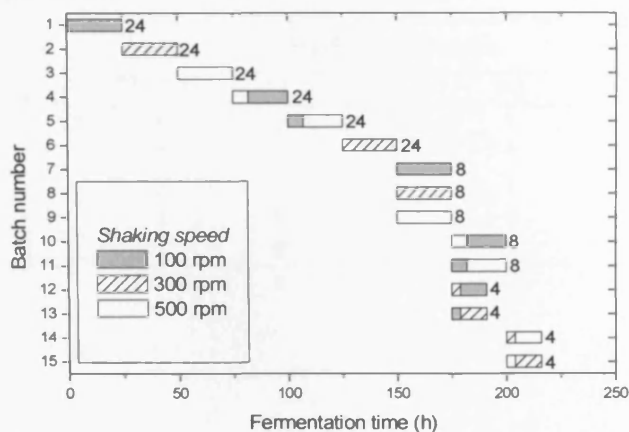
The following schedules assume the availability of three SF incubator shakers, each of which can be operated at a separate shaking speed. It is also assumed that a maximum of 24 SF experiments can be processed in parallel.



**Figure C.5** An optimal schedule showing the minimum theoretical number of hours required to complete the familiarisation experiments described in Table 2.3, if performed in SFs. Experiments from Table 2.3 are arranged in fermentation batches according to those with identical pre-I and post-I shaking speeds. Numbers to the right of each bar represent number of SF fermentations per batch (including replicates).



**Figure C.6** An optimal schedule showing the minimum theoretical number of hours required to complete the screening experiments described in Table 4.2, if performed in SFs. Experiments from Table 4.2 are arranged in fermentation batches according to those with identical pre-I and post-I shaking speeds. Numbers to the right of each bar represent number of SF fermentations per batch (including replicates).



**Figure C.7** An optimal schedule showing the minimum theoretical number of hours required to complete the optimisation experiments described in Table 4.3, if performed in SFs. Experiments from Table 4.3 are arranged in fermentation batches according to those with identical pre-I and post-I shaking speeds. Numbers to the right of each bar represent number of SF fermentations per batch (including replicates).

#### C.4 Disposable MWP requirements

The following requirements for MWPs assume that each MWP is disposed of after each batch.

**Table C.6** Minimum number of MWPs required at the familiarisation stage when experiments are arranged as described in Figure C.1.

Batch Number	MWP requirements		
	P1	P2	P3
1	1	1	1
2	1	1	1
3	1	1	1
4	1	1	1
5	1	0	0
<b>Total</b>	<b>5</b>	<b>4</b>	<b>4</b>

**Table C.7** Minimum number of MWPs required at the screening stage when experiments are arranged as described in Figure C.2.

Batch Number	MWP requirements		
	P1	P2	P3
1	1	1	1
2	1	1	1
3	1	1	1
4	1	1	1
5	0	0	1
<b>Total</b>	<b>4</b>	<b>4</b>	<b>5</b>

**Table C.8** Minimum number of MWPs required at the optimisation stage when experiments are arranged as described in Figure C.3

Batch Number	MWP requirements
	P3
1	2
2	1
3	3
4	1
5	2
6	2
7	2
8	1
9	1
<b>Total</b>	<b>15</b>

## C.5 Cost scenarios

### C.5.1 Background

The cost analyses in the following sections refer to the protein expression optimisation scenarios defined in Section 7.2. Section 7.2 also provides the general assumptions for all three cost analyses. In each analysis, 1 unit = 1 l media / 1 MWP / 1 assay / 1 h pay.



### ***C.5.2 Scenario one: current framework (Section 7.2.2)***

In this scenario, both the DoE and MWP framework components are applied and the experimental requirements are considered at each stage of the DoE process: familiarisation (*F*), screening (*S*) and optimisation (*O*). A detailed cost analysis for this scenario is provided in Table C.9. Specific assumptions are as follows:

- Total number of fermentations performed = 440 (Chapter 4);
- Mean microwell fermentation volume = 2 ml.

### ***C.5.3 Scenario two: current framework minus the MWP component (Section 7.2.3)***

In this scenario, the MWP component is replaced by shake-flask (SF) experimentation and the costs associated with this change are estimated. Again, the experimental requirements at each stage of the DoE process are considered. A detailed cost analysis for this scenario is provided in Table C.10. Specific assumptions are as follows:

- The same number of fermentations (440) is performed as in scenario one (Section C.5.2);
- The DoE factor related to MWP geometry is exchanged for SF geometry;
- Mean SF fermentation volume = 250 ml.
- Same number of inoculum SFs and assays as for scenario one (Table C.9).

**Table C.9** Cost analysis for scenario one: current framework. <sup>1</sup>Unit costs calculated in Section C.1. <sup>2</sup>Requirements are in accordance with experimental methods (Chapter 2). <sup>3, 4</sup>Requirements are calculated in Sections C.4 and C.2, respectively.

Variable	Requirements by stage				Unit cost (£) <sup>1</sup>	Cost of materials (£)	Cost of labour (£)
	F	S	O	Total			
<b>Inoculum preparation</b>							
<i>Number of 500 ml (working volume) SFs<sup>2</sup></i>							
LB	4	4	0	8	3.98	15.90	
TB	4	5	9	18	6.82	61.41	
M9Y	5	4	0	9	3.73	<u>16.81</u>	
						<b>94.12</b>	
<b>Fermentation</b>							
<i>Minimum number of MWP<sup>3</sup></i>							
P1	5	4	0	9	9.63	86.67	
P2	4	4	0	8	9.63	77.04	
P3	4	5	15	24	3.82	<u>91.79</u>	
						<b>255.50</b>	
<i>Number of 2 ml MWP fermentations<sup>2</sup></i>							
LB	32	32	0	64	3.98	0.51	
TB	40	48	200	288	6.82	3.93	
M9Y	48	40	0	88	3.73	<u>0.66</u>	
						<b>5.10</b>	
<b>Assay</b>							
<i>Number of assays<sup>2</sup></i>							
	120	120	200	440	0.57	<b>251.98</b>	
<b>Labour</b>							
<i>Minimum fermentation time (h)<sup>4</sup></i>							
	70	70	135	275	10		2,750
<b>Total costs</b>						<b>607</b>	<b>2,750</b>

**Table C.10** Cost analysis for scenario two: current framework minus the MWP component. <sup>1</sup>Unit costs calculated in Section C.1. <sup>2</sup>Minimum total fermentation times calculated in Section C.3.

Variable	Requirements by stage				Unit cost (£) <sup>1</sup>	Cost of materials (£)	Cost of labour (£)
	F	S	O	Total			
<b>Inoculum preparation</b>							
<i>Number of 500 ml (working volume) SFs</i>							
LB	4	4	0	8	3.98	15.90	
TB	4	5	9	18	6.82	61.41	
M9Y	5	4	0	9	3.73	<u>16.81</u>	
						<b>94.12</b>	
<b>Fermentation</b>							
<i>Number of 250 ml SF fermentations</i>							
LB	32	32	0	64	3.98	63.61	
TB	40	48	200	288	6.82	491.31	
M9Y	48	40	0	88	3.73	<u>82.16</u>	
						<b>637.09</b>	
<b>Assay</b>							
<i>Number of assays</i>							
	120	120	200	440	0.57	<b>251.98</b>	
<b>Labour</b>							
<i>Minimum total fermentation time (h)<sup>2</sup></i>							
	105	105	216	426	10.00		4,260
<b>Total costs</b>						<b>983</b>	<b>4,260</b>

#### ***C.5.4 Scenario 3: current framework minus the DoE component (Section 7.2.4)***

In this scenario, the DoE component is removed and the costs associated with this change are estimated. Here, the DoE stages of familiarisation, screening and optimisation do not exist and so it is necessary to investigate each factor at three levels from the start. A detailed cost analysis for this scenario is provided in Table C.11. Specific assumptions are as follows:

- Total number of factor combinations  
= 2 qualitative factors (3 levels each) x 8 quantitative factors (3 levels each)  
=  $3^2 \times 3^8$   
= 59,049
- Total number of experiments (including replicates)  
= 4 x 59,049  
= 236,196
- Mean microwell fermentation volume = 2 ml;
- 1 SF inoculates 9 MWP = 3x48 + 6x24 exps = 288 exps;
- Minimum number of MWPs required = total number experiments / individual MWP capacity;
- The schedule of experimentation is the same as that of the optimisation experiments of scenario one (Section C.5.2), where the maximum rate of experimentation = 324 fermentations in 135 h (Figure C.4).

**Table C.11** Cost analysis for scenario three: current framework minus the DoE component. <sup>1</sup>Unit costs calculated in Section C.1.

<b>Variable</b>	<b>Total requirements</b>	<b>Unit cost (£)<sup>1</sup></b>	<b>Cost of materials (£)</b>	<b>Cost of labour (£)</b>
<b>Inoculum preparation</b>				
<i>Number of 500 ml (working volume) SFs</i>				
LB	273	3.98	543	
TB	273	6.82	933	
M9Y	273	3.73	<u>510</u>	
			<b>1,987</b>	
<b>Fermentation</b>				
<i>Minimum number of MWP</i> s				
P1	1,640	9.63	15796	
P2	3,281	9.63	31591	
P3	3,281	3.82	<u>12546</u>	
			<b>59,933</b>	
<i>Number of 2 ml MWP fermentations</i>				
LB	78,732	3.98	626	
TB	78,732	6.82	1074	
M9Y	78,732	3.73	<u>588</u>	
			<b>2,289</b>	
<b>Assay</b>				
<i>Number of assays</i>				
	236,196	0.57	<b>135,266</b>	
<b>Labour</b>				
<i>Minimum total fermentation time (h)</i>				
	98,415	10		984,150
<b>Total costs</b>			<b>199,474</b>	<b>984,150</b>

## Appendix D – List of publications

- **Chapter 4:** Islam,R.S., Tisi,D., Levy,M.S., and Lye,G.J. (2007) Framework for the Rapid Optimization of Soluble Protein Expression in *Escherichia coli* Combining Microscale Experiments and Statistical Experimental Design. *Biotechnology Progress*. **23**, 785-793.
- **Chapters 5 + 6:** Islam,R.S., Tisi,D., Levy,M.S., and Lye,G.J. (2007) Scale-up of *E. coli* growth and recombinant protein expression conditions from microwell to laboratory and pilot scale based on matched  $k_{La}$ . *Biotechnology and Bioengineering*. In press. doi:10.1002/bit.21697

---

## References

- Adams,C.P. and Brantner,V. (2006) Estimating the cost of new drug development: is it really 802 million dollars? *Health Affairs Web Exclusive* **25**, 420-428.
- Adinarayana,K. and Ellaiah,P. (2002) Response surface optimization of the critical medium components for the production of alkaline protease by a newly isolated *Bacillus sp.* *Journal of Pharmacy and Pharmaceutical Sciences* **5**, 272-278.
- Alam,J. and Cook,J.L. (1990) Reporter genes: Application to the study of mammalian gene transcription. *Analytical Biochemistry* **188**, 245-254.
- Anderlei,T. and Buchs,J. (2001) Device for sterile online measurement of the oxygen transfer rate in shaking flasks. *Biochemical Engineering Journal* **7**, 157-162.
- Andersen,D.C. and Krummen,L. (2002) Recombinant protein expression for therapeutic applications. *Current Opinion in Biotechnology* **13**, 117-123.
- Andersen,K.B. and von Meyenburg,K. (1980) Are growth-rates of *Escherichia coli* in batch cultures limited by respiration? *Journal of Bacteriology* **144**, 114-123.
- Anon. (2005) *Megazyme glycerol assay procedure*. Technical Brochure.
- Bahia,D., Cheung,R., Buchs,M., Geisse,S., and Hunt,I. (2005) Optimisation of insect cell growth in deep-well blocks: development of a high-throughput insect cell expression screen. *Protein Expression and Purification* **39**, 61-70.
- Balbas,P. (2001) Understanding the art of producing protein and nonprotein molecules in *Escherichia coli*. *Molecular Biotechnology* **19**, 251-267.
- Baneyx,F. (1999) Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology* **10**, 411-421.
- Bentley,W.E., Davis,R.H., and Kompala,D.S. (1991a) Dynamics of induced CAT expression in *E. coli*. *Biotechnology and Bioengineering* **38**, 749-760.
- Bentley,W.E., Davis,R.H., and Kompala,D.S. (1991b) Dynamics of Induced Cat Expression in *Escherichia-Coli*. *Biotechnology and Bioengineering* **38**, 749-760.
- Bentley,W. and Kompala,D. (1990) Optimal induction of protein synthesis in recombinant bacterial cultures. *Annals of the New York Academy of Sciences* **589**, 121-138.
- Betts,J.I. and Baganz,F. (2006) Miniature bioreactors: current practices and future opportunities. *Microbial Cell Factories* **5**.
- Bhattacharya,S.K. and Dubey,A.K. (1997) Effects of dissolved oxygen and oxygen mass transfer on overexpression of target gene in recombinant *E. coli*. *Enzyme and Microbial Technology* **20**, 355-360.

- Bird,P.I., Pak,S.C., Worrall,D.M., and Bottomley,S.P. (2004) Production of recombinant serpins in *Escherichia coli*. *Methods* **32**, 169-176.
- Bleicher,K.H., Bohm,H.J., Muller,K., and Alanine,A.I. (2003) Hit and lead generation: Beyond high-throughput screening. *Nature Reviews Drug Discovery* **2**, 369-378.
- Blundell,T.L., Jhoti,H., and Abell,C. (2002) High-throughput crystallography for lead discovery in drug design. *Nature Reviews Drug Discovery* **1**, 45-54.
- Blundell,T.L. and Patel,S. (2004) High-throughput X-ray crystallography for drug discovery. *Current Opinion in Pharmacology* **4**, 490-496.
- Broedel,S., Papciak,S., and Jones,W. The selection of optimum media formulations for improved expression of recombinant proteins in *E. coli*. 2001. Athena Enzyme Systems™ Group.
- Broedel,S., Papciak,S., and Jones,W. (2002) The selection of optimum media formulations for improved expression of recombinant proteins in *E. coli*. *Athena Enzyme Systems Technical Bulletin* **2**, 1-6.
- Bronstein,I., Fortin,J., Stanley,P.E., Stewart,G.S., and Kricka,L.J. (1994) Chemiluminescent and bioluminescent reporter gene assays. *Analytical Biochemistry* **219**, 169-181.
- Brue,G. (2002) Chapter 1: What is Six Sigma? In *Six Sigma for managers* McGraw-Hill, New York; London.
- Buchs,J. (2001) Introduction to advantages and problems of shaken cultures. *Biochemical Engineering Journal* **7**, 91-98.
- Calik,P., Yilgor,P., and Demir,A.S. (2006) Influence of controlled-pH and uncontrolled-pH operations on recombinant benzaldehyde lyase production by *Escherichia coli*. *Enzyme and Microbial Technology* **38**, 617-627.
- Cao,Y. (2006) Optimization of expression of *dhaT* gene encoding 1, 3-propanediol oxidoreductase from *Klebsiella pneumoniae* in *Escherichia coli* using the methods of uniform design and regression analysis. *Journal of Chemical Technology and Biotechnology* **81**, 109-112.
- Chambers,S.P., Austen,D.A., Fulghum,J.R., and Kim,W.M. (2004) High-throughput screening for soluble recombinant expressed kinases in *Escherichia coli* and insect cells. *Protein Expression and Purification* **36**, 40-47.
- Chesshyre,J.A. and Hipkiss,A.R. (1989) Low temperatures stabilize interferon-alpha-2 against proteolysis in *Methylophilus methylotrophus* and *Escherichia coli*. *Applied Microbiology and Biotechnology* **31**, 158-162.
- Choi,J.H., Jeong,K.J., Kim,S.C., and Lee,S.Y. (2000) Efficient secretory production of alkaline phosphatase by high cell density culture of recombinant *Escherichia coli*



---

using the *Bacillus sp* endoxylanase signal sequence. *Applied Microbiology and Biotechnology* **53**, 640-645.

Choi,J.H., Keum,K.C., and Lee,S.Y. (2006) Production of recombinant proteins by high cell density culture of *Escherichia coli*. *Chemical Engineering Science* **61**, 876-885.

Collins,F.S. (1999) Medical and Societal Consequences of the Human Genome Project. *N Engl J Med* **341**, 28-37.

Congreve,M., Murray,C.W., and Blundell,T.L. (2005) Keynote review: Structural biology and drug discovery. *Drug Discovery Today* **10**, 895-907.

Corisdeo,S. and Wang,B.Y. (2004) Functional expression and display of an antibody Fab fragment in *Escherichia coli*: study of vector designs and culture conditions. *Protein Expression and Purification* **34**, 270-279.

Creveling,C. (2006) Chapter 1: Introduction to Six Sigma for technical processes. In *Six Sigma for technical processes: an overview for R&D executives, technical leaders, and engineering managers*. Prentice Hall, London.

Dahl,S.G. and Sylte,I. (2006) From genomics to drug targets. *Journal of Psychopharmacology* **20**, 95-99.

Davies,A., Greene,A., Lullau,E., and Abbott,W.M. (2005) Optimisation and evaluation of a high-throughput mammalian protein expression system. *Protein Expression and Purification* **42**, 111-121.

Davies,A.H. (1994) Current methods for manipulating baculoviruses. *Bio-Technology* **12**, 47-50.

De Leon,A., Hernandez,V., Galindo,E., and Ramirez,O.T. (2003) Effects of dissolved oxygen tension on the production of recombinant penicillin acylase in *Escherichia coli*. *Enzyme and Microbial Technology* **33**, 689-697.

Debouck,C. and Metcalf,B. (2000) The impact of genomics on drug discovery. *Annual Review Of Pharmacology And Toxicology* **40**, 193-207.

Deshpande,R.R., Wittmann,C., and Heinzle,E. (2004) Microplates with integrated oxygen sensing for medium optimization in animal cell culture. *Cytotechnology* **46**, 1-8.

Diaz-Ricci,J.C., Hitzmann,B., Rinas,U., and Bailey,J.E. (1990) Comparative studies of glucose catabolism by *Escherichia coli* grown in a complex medium under aerobic and anaerobic conditions. *Biotechnol.Prog.* **6**, 326-332.

Dien,B.S., Nichols,N.N., and Bothast,R.J. (2001) Recombinant *Escherichia coli* engineered for production of L-lactic acid from hexose and pentose sugars. *Journal of Industrial Microbiology and Biotechnology* **V27**, 259-264.

DiMasi,J.A., Hansen,R.W., and Grabowski,H.G. (2003) The price of innovation: new estimates of drug development costs. *Journal of Health Economics* **22**, 151-185.

- Doig,S.D., Pickering,S.C.R., Lye,G.J., and Baganz,F. (2005) Modelling surface aeration rates in shaken microtitre plates using dimensionless groups. *Chemical Engineering Science* **60**, 2741-2750.
- Donovan,R.S., Robinson,C.W., and Glick,B.R. (1996a) Optimizing inducer and culture conditions for expression of foreign proteins under the control of the *lac* promoter. *Journal of Industrial Microbiology* **16**, 145-154.
- Donovan,R.S., Robinson,C.W., and Glick,B.R. (1996b) Review: Optimizing inducer and culture conditions for expression of foreign proteins under the control of the *lac* promoter. *Journal of Industrial Microbiology* **16**, 145-154.
- Duetz,W.A., Ruedi,L., Hermann,R., O'Connor,K., Buchs,J., and Witholt,B. (2000) Methods for intense aeration, growth, storage, and replication of bacterial strains in microtiter plates. *Applied and Environmental Microbiology* **66**, 2641-2646.
- Duetz,W.A. and Witholt,B. (2001) Effectiveness of orbital shaking for the aeration of suspended bacterial cultures in square-deepwell microtiter plates. *Biochemical Engineering Journal* **7**, 113-115.
- Duetz,W.A. and Witholt,B. (2004) Oxygen transfer by orbital shaking of square vessels and deepwell microtiter plates of various dimensions. *Biochemical Engineering Journal* **17**, 181-185.
- Dunn,I.J. and Einsele,A. (1975) Oxygen transfer coefficients by the dynamic method. *Journal Of Applied Chemistry And Biotechnology* **25**, 707-720.
- Dutta,J.R., Dutta,P.K., and Banerjee,R. (2004) Optimization of culture parameters for extracellular protease production from a newly isolated *Pseudomonas* sp using response surface and artificial neural network models. *Process Biochemistry* **39**, 2193-2198.
- Edwards,A., Berman,J., and Sundstrom,M. (2005) Structural Genomics and Drug Discovery. In *Annual Reports in Medicinal Chemistry* (Edited by Annette,M.D.) Pp. 349-369. Academic Press.
- El Helow,E.R., Abdel-Fattah,Y.R., Ghanem,K.M., and Mohamad,E.A. (2000) Application of the response surface methodology for optimizing the activity of an *aprE*-driven gene expression system in *Bacillus subtilis*. *Applied Microbiology and Biotechnology* **54**, 515-520.
- Elmahdi,I., Baganz,F., Dixon,K., Harrop,T., Sugden,D., and Lye,G.J. (2003) pH control in microwell fermentations of *S. erythraea* CA340: influence on biomass growth kinetics and erythromycin biosynthesis. *Biochemical Engineering Journal* **16**, 299-310.
- Eriksson,L., Johansson,E., Kettaneh-Wold,N., Wikström,C., and Wold,S. (2000) *Design of Experiments: Principles and Applications*. Umeå / Learnways AB, Stockholm.

- Fernandes,P. and Cabral,J.M.S. (2006) Microlitre/millilitre shaken bioreactors in fermentative and biotransformation processes - a review. *Biocatalysis and Biotransformation* **24**, 237-252.
- Ferreira-Torres,C., Micheletti,M., and Lye,G.J. (2005) Microscale process evaluation of recombinant biocatalyst libraries: application to Baeyer-Villiger monooxygenase catalysed lactone synthesis. *Bioprocess and Biosystems Engineering* **28**, 83-93.
- Ferrer,M., Chernikova,T.N., Yakimov,M.M., Golyshin,P.N., and Timmis,K.N. (2003) Chaperonins govern growth of *Escherichia coli* at low temperatures. *Nature Biotechnology* **21**, 1266-1267.
- Folli,C., Calderone,V., Ottonello,S., Bolchi,A., Zanotti,G., Stoppini,M., and Berni,R. (2001) Identification, retinoid binding, and x-ray analysis of a human retinol-binding protein. *Proceedings of the National Academy of Sciences* **98**, 3710-3715.
- Freyer,S.A., Konig,M., and Kunkel,A. (2004) Validating shaking flasks as representative screening systems. *Biochemical Engineering Journal* **17**, 169-173.
- Galindo,E., Bolivar,F., and Quintero,R. (1990) Maximizing the expression of recombinant proteins in *Escherichia coli* by manipulation of culture conditions. *Journal of Fermentation and Bioengineering* **69**, 159-165.
- Garcia-Arrazola,R., Dawson,P., Buchanan,I., Doyle,B., Fearn,T., Titchener-Hooker,N.J., and Baganz,F. (2005) Evaluation of the effects and interactions of mixing and oxygen transfer on the production of Fab' antibody fragments in *Escherichia coli* fermentation with gas blending. *Bioprocess and Biosystems Engineering* **27**, 365-374.
- Georgiou,G. and Valax,P. (1996) Expression of correctly folded proteins in *Escherichia coli*. *Current Opinion in Biotechnology* **7**, 190-197.
- Gill,A., Cleasby,A., and Jhoti,H. (2005) The discovery of novel protein kinase inhibitors by using fragment-based high-throughput x-ray crystallography. *ChemBioChem* **6**, 506-512.
- Goulding,C.W. and Perry,J.L. (2003) Protein production in *Escherichia coli* for structural studies by X-ray crystallography. *Journal of Structural Biology* **142**, 133-143.
- Graumann,K. and Premstaller,A. (2006) Manufacturing of recombinant therapeutic proteins in microbial systems. *Biotechnology journal* **1**, 164-186.
- Grodberg,J. and Dunn,J.J. (1988) *ompT* encodes the *Escherichia coli* outer-membrane protease that cleaves T7 RNA polymerase during purification. *Journal of Bacteriology* **170**, 1245-1253.
- Hakkila,K., Maksimow,M., Karp,M., and Virta,M. (2002) Reporter genes *lucFF*, *luxCDABE*, *gfp*, and *dsred* have different characteristics in whole-cell bacterial sensors. *Analytical Biochemistry* **301**, 235-242.

- 
- Hammonds,T.R., Maxwell,A., and Jenkins,J.R. (1998) Use of a rapid throughput in vivo screen to investigate inhibitors of eukaryotic topoisomerase II enzymes. *Antimicrobial Agents and Chemotherapy* **42**, 889-894.
- Hanash,S. (2003) Disease proteomics. *Nature* **422**, 226-232.
- Hannig,G. and Makrides,S.C. (1998) Strategies for optimizing heterologous protein expression in *Escherichia coli*. *Trends in Biotechnology* **16**, 54-60.
- Harms,P., Kostov,Y., French,J.A., Soliman,M., Anjanappa,M., Ram,A., and Rao,G. (2006) Design and performance of a 24-station high throughput microbioreactor. *Biotechnology and Bioengineering* **93**, 6-13.
- Hayler,R. and Nichols,M. (2005) Chapter 3: What is the Six Sigma process management? In *What is Six Sigma process management?* McGraw-Hill, New York; London.
- Henzler,H.J. and Schedel,M. (1991) Suitability of the shaking flask for oxygen-supply to microbiological cultures. *Bioprocess Engineering* **7**, 123-131.
- Hermann,R., Lehmann,M., and Buchs,J. (2003) Characterization of gas-liquid mass transfer phenomena in microtiter plates. *Biotechnology and Bioengineering* **81**, 178-186.
- Holms,H. (1996) Flux analysis and control of the central metabolic pathways in *Escherichia coli*. *Fems Microbiology Reviews* **19**, 85-116.
- Hounsa,C.G., Aubry,J.M., Dubourguier,H.C., and Hornez,J.P. (1996) Application of factorial and Doehlert designs for optimization of pectate lyase production by a recombinant *Escherichia coli*. *Applied Microbiology and Biotechnology* **45**, 764-770.
- Hsu,Y.L. and Wu,W.T. (2002) A novel approach for scaling-up a fermentation system. *Biochemical Engineering Journal* **11**, 123-130.
- Hughmark,G.A. (1980) Power requirements and interfacial area in gas-liquid turbine agitated systems. *Industrial & Engineering Chemistry Process Design and Development* **19**, 638-641.
- Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-945.
- Humphrey,A. (1998) Shake flask to fermentor: What have we learned? *Biotechnol.Prog.* **14**, 3-7.
- Hunt,I. (2005) From gene to protein: a review of new and enabling technologies for multi-parallel protein expression. *Protein Expression and Purification* **40**, 1-22.
- Ingraham,J.L. and Marr,A.G. (1996) Effect of temperature, pressure, pH, and osmotic stress on growth. In *Escherichia coli and Salmonella: Cellular and Molecular Biology* (Edited by Neidhardt,F.C., Curtiss,R., Ingraham,J.L., Lin,E., Low,K., Magasanik,B., Reznikoff,W., Riley,M., Schaechter,M., and Umberger,H.) Pp. 1570-1578. ASM Press, Washington DC.
-

- Jana,S. and Deb,J.K. (2005) Strategies for efficient production of heterologous proteins in *Escherichia coli*. *Applied Microbiology and Biotechnology* **67**, 289-298.
- Janssen,D. Proteins: Selection, Synthesis, and Purification Strategies of Optimizing Drug Discovery. 2004. Cambridge Healthcare Institute.
- Jeong,K.J. and Lee,S.Y. (1999) High-level production of human leptin by fed-batch cultivation of recombinant *Escherichia coli* and its purification. *Applied and Environmental Microbiology* **65**, 3027-3032.
- John,G.T., Goelling,D., Klimant,I., Schneider,H., and Heinzle,E. (2003a) pH-sensing 96-well microtitre plates for the characterization of acid production by dairy starter cultures. *Journal of Dairy Research* **70**, 327-333.
- John,G.T., Klimant,I., Wittmann,C., and Heinzle,E. (2003b) Integrated optical sensing of dissolved oxygen in microtiter plates: A novel tool for microbial cultivation. *Biotechnology and Bioengineering* **81**, 829-836.
- Jonasson,P., Liljeqvist,S., Nygren,P.A., and Stahl,S. (2002) Genetic design for facilitated production and recovery of recombinant proteins in *Escherichia coli*. *Biotechnology and Applied Biochemistry* **35**, 91-105.
- Joyeux,A., Balaguer,P., Germain,P., Boussioux,A.M., Pons,M., and Nicolas,J.C. (1997) Engineered cell lines as a tool for monitoring biological activity of hormone analogs. *Analytical Biochemistry* **249**, 119-130.
- Junker,B.H. (2004) Scale-up methodologies for *Escherichia coli* and yeast fermentation processes. *Journal of Bioscience and Bioengineering* **97**, 347-364.
- Kensy,F., Zimmermann,H.F., Knabben,I., Anderlei,T., Trauthwein,H., Dingerdissen,U., and Büchs,J. (2005) Oxygen transfer phenomena in 48-well microtiter plates: Determination by optical monitoring of sulfite oxidation and verification by real-time measurement during microbial growth. *Biotechnology and Bioengineering* **89**, 698-708.
- Kerns,E.H. (2001) High throughput physicochemical profiling for drug discovery. *Journal of Pharmaceutical Sciences* **90**, 1838-1858.
- Kiefhaber,T., Rudolph,R., Kohler,H.H., and Buchner,J. (1991) Protein aggregation invitro and invivo: A quantitative model of the kinetic competition between folding and aggregation. *Bio-Technology* **9**, 825-829.
- Kim,D.M. and Swartz,J.R. (2001) Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. *Biotechnology and Bioengineering* **74**, 309-316.
- Kleman,G.L. and Strohl,W.R. (1994) Developments in high cell density and high productivity microbial fermentation. *Current Opinion in Biotechnology* **5**, 180-186.
- Kola,I. and Landis,J. (2004) Can the pharmaceutical industry reduce attrition rates? *Nature Reviews Drug Discovery* **3**, 711-716.

- Korz,D.J., Rinas,U., Hellmuth,K., Sanders,E.A., and Deckwer,W.D. (1995) Simple fed-batch technique for high cell-density cultivation of *Escherichia Coli*. *Journal of Biotechnology* **39**, 59-65.
- Kostov,Y., Harms,P., Randers-Eichhorn,L., and Rao,G. (2001) Low-cost microbioreactor for high-throughput bioprocessing. *Biotechnology and Bioengineering* **72**, 346-352.
- Kuhn,P., Wilson,K., Patch,M.G., and Stevens,R.C. (2002) The genesis of high-throughput structure-based drug discovery using protein crystallography. *Current Opinion in Chemical Biology* **6**, 704-710.
- Kumar,S., Wittmann,C., and Heinzle,E. (2004) Minibioreactors. *Biotechnology Letters* **26**, 1-10.
- Lamping,S.R., Zhang,H., Allen,B., and Ayazi Shamlou,P. (2003) Design of a prototype miniature bioreactor for high throughput automated bioprocessing. *Chemical Engineering Science* **58**, 747-758.
- Law,G.H.E., Gandelman,O.A., Tisi,L.C., Lowe,C.R., and Murray,J.A.H. (2006) Mutagenesis of solvent-exposed amino acids in *Photinus pyralis* luciferase improves thermostability and pH-tolerance. *Biochemical Journal* **397**, 305-312.
- Lee,C., Sun,W.J., Burgess,B.W., Junker,B.H., Reddy,J., Buckland,B.C., and Greasham,R.L. (1997) Process optimization for large-scale production of TGF-alpha-PE40 in recombinant *Escherichia coli*: Effect of medium composition and induction timing on protein expression. *Journal of Industrial Microbiology & Biotechnology* **18**, 260-266.
- Lee,S.Y. (1996) High cell-density culture of *Escherichia coli*. *Trends in Biotechnology* **14**, 98-105.
- Li,X.L., Robbins,J.W., and Taylor,K.B. (1992) Effect of the levels of dissolved-oxygen on the expression of recombinant proteins in 4 recombinant *Escherichia coli* strains. *Journal of Industrial Microbiology* **9**, 1-9.
- Lindsay,M.A. (2003) Target discovery. *Nature Reviews Drug Discovery* **2**, 831-838.
- Luli,G.W. and Strohl,W.R. (1990) Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. *Applied and Environmental Microbiology* **56**, 1004-1011.
- Lye,G.J., Ayazi-Shamlou,P., Baganz,F., Dalby,P.A., and Woodley,J.M. (2003) Accelerated design of bioconversion processes using automated microscale processing techniques. *Trends in Biotechnology* **21**, 29-37.
- Maier,U., Losen,M., and Buchs,J. (2004) Advances in understanding and modeling the gas-liquid mass transfer in shake flasks. *Biochemical Engineering Journal* **17**, 155-167.

- Maier,U. and Buchs,J. (2001) Characterisation of the gas-liquid mass transfer in shaking bioreactors. *Biochemical Engineering Journal* **7**, 99-106.
- Maniatis,T., Sambrook,J., and Fritsch,E.F. (1989) *Molecular Cloning : A Laboratory Manual*. Cold Spring Harbor Laboratory.
- Martin,R.M. (2002) Regulating Medicines in Europe: competition, expertise and public health: J Abraham, G Lewis. London: Routledge 2000, pp.237, {pound}17.99. ISBN 0-415-208-785. *Int.J.Epidemiol.* **31**, 272.
- McDaniel,L.E., Bailey,E.G., and Zimmerli,A. (1965) Effect of oxygen supply on growth of *Escherichia coli*, I. Studies in unbaffled and baffled shake flasks. *Applied Microbiology* **13**, 100-114.
- Mere,L., Bennett,T., Coassin,P., England,P., Hamman,B., Rink,T., Zimmerman,S., and Negulescu,P. (1999) Miniaturized FRET assays and microfluidics: key components for ultra-high-throughput screening. *Drug Discovery Today* **4**, 363-369.
- Micheletti,M., Barrett,T., Doig,S.D., Baganz,F., Levy,M.S., Woodley,J.M., and Lye,G.J. (2006) Fluid mixing in shaken bioreactors: Implications for scale-up predictions from microlitre-scale microbial and mammalian cell cultures. *Chemical Engineering Science* **61**, 2939-2949.
- Mogk,A., Mayer,M.P., and Deuerling,E. (2002) Mechanisms of protein folding: Molecular chaperones and their application in biotechnology. *ChemBioChem* **3**, 807-814.
- Motulsky,H.J. and Christopoulos,A. (2003) *Fitting models to biological data using linear and nonlinear regression. A practical guide to curve fitting*. GraphPad Software Inc., San Diego CA.
- Mundy,C. (2001) The human genome project: A historical perspective. *Pharmacogenomics* **2**, 37-49.
- Naylor,L.H. (1999) Reporter gene technology: the future looks bright. *Biochemical Pharmacology* **58**, 749-757.
- Nicholson,J.K., Connelly,J., Lindon,J.C., and Holmes,E. (2002) Metabonomics: A platform for studying drug toxicity and gene function. *Nature Reviews Drug Discovery* **1**, 153-161.
- Nikerel,I.E., Toksoy,E., Kirdar,B., and Yildirim,R. (2005) Optimizing medium composition for *TaqI* endonuclease production by recombinant *Escherichia coli* cells using response surface methodology. *Process Biochemistry* **40**, 1633-1639.
- Noffke,T. (2007) No Time to Delay. *Pharmaceutical Executive* **27**, 22-27.
- O'Reilly,D.R., Miller,L.K., and Luckow,V.A. (1992) *Baculovirus Expression Vectors: A Laboratory Manual*. W. H. Freeman, New York.

- Ortiz-Ochoa, K., Doig, S.D., Ward, J.M., and Baganz, F. (2005) A novel method for the measurement of oxygen mass transfer rates in small-scale vessels. *Biochemical Engineering Journal* **25**, 63-68.
- Ozawa, K., Dixon, N., and Otting, G. (2005) Cell-free synthesis of <sup>15</sup>N-labeled proteins for NMR studies. *International Union of Biochemistry and Molecular Biology: Life* **57**, 615-622.
- Park, J.K., Moon, J.H., Kim, J.H., and Kim, E.E. (2005) Crystallization and preliminary X-ray crystallographic analysis of peptide deformylase (PDF) from *Bacillus cereus* in ligand-free and actinonin-bound forms. *Acta Crystallographica Section F* **61**, 150-152.
- Peng, L., Xu, Z.N., Fang, X.M., Wang, F., and Cen, P.L. (2004) High-level expression of soluble human beta-defensin-2 in *Escherichia coli*. *Process Biochemistry* **39**, 2199-2205.
- Rai, M. and Padh, H. (2001) Expression systems for production of heterologous proteins. *Current Science* **80**, 1121-1128.
- Rees, D.C., Congreve, M., Murray, C.W., and Carr, R. (2004) Fragment-based lead discovery. *Nature Reviews Drug Discovery* **3**, 660-672.
- Reisman, H.B. (1993) Problems in scale-up of biotechnology production processes. *Critical Reviews in Biotechnology* **13**, 195-253.
- Ren, X., Yu, D., Han, S., and Feng, Y. (2006) Optimization of recombinant hyperthermophilic esterase production from agricultural waste using response surface methodology. *Bioresource Technology* **97**, 2345-2349.
- Reznikoff, W.a.J.A. (1980) The *lac* promoter. In *The Operon* (Edited by Miller, J. and Reznikoff, W.) Pp. 221-224. Cold Spring Harbor Laboratory, New York.
- Riesenber, D. (1991) High-cell-density cultivation of *Escherichia coli*. *Current Opinion in Biotechnology* **2**, 380-384.
- Roda, A., Pasini, P., Mirasoli, M., Michelini, E., and Guardigli, M. (2004) Biotechnological applications of bioluminescence and chemiluminescence. *Trends in Biotechnology* **22**, 295-303.
- Roebuck, K., Brundin, A., and Johns, M. (1995) Response surface optimization of temperature and pH for the growth of *Pachysolen tannophilus*. *Enzyme and Microbial Technology* **17**, 75-78.
- Russ, A.P. and Lampel, S. (2005) The druggable genome: An update. *Drug Discovery Today* **10**, 1607-1610.
- Ryan, W. and Parulekar, S.J. (1990) Effects of culture conditions on plasmid stability and production of a plasmid-encoded protein in batch and continuous cultures of *Escherichia coli* JM103[Puc8]. *Annals of the New York Academy of Sciences* **589**, 91-110.



- 
- Sang, Y.L. (1996) High cell-density culture of *Escherichia coli*. *Trends in Biotechnology* **14**, 98-105.
- Schein, C.H. (1989) Production of soluble recombinant proteins in bacteria. *Bio-Technology* **7**, 1141-1147.
- Shiloach, J. and Fass, R. (2005) Growing *E. coli* to high cell density - a historical perspective on method development. *Biotechnology Advances* **23**, 345-357.
- Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001) Cell-free translation reconstituted with purified components. *Nature Biotechnology* **19**, 751-755.
- Singleton, P. and Sainsbury, D. (2002) *Dictionary of Microbiology and Molecular Biology*. John Wiley and Sons Ltd, Chichester.
- Sorensen, H.P. and Mortensen, K.K. (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *Journal of Biotechnology* **115**, 113-128.
- Stanbury, P.F. and Whitaker, A. (1993) *Principles of Fermentation Technology*. Pergamon Press Ltd, Oxford.
- Stewart, L., Clark, R., and Behnke, C. (2002) High-throughput crystallization and structure determination in drug discovery. *Drug Discovery Today* **7**, 187-196.
- Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods in Enzymology* **185**, 60-89.
- Sunitha, K., Lee, J.K., and Oh, T.K. (1999) Optimization of medium components for phytase production by *E. coli* using response surface methodology. *Bioprocess and Biosystems Engineering* **21**, 477-481.
- Suto, C.M. and Ignar, D.M. (1997) Selection of an optimal reporter gene for cell-based high throughput screening assays. *Journal of Biomolecular Screening* **2**, 7-9.
- Swalley, S.E., Fulghum, J.R., and Chambers, S.P. (2006) Screening factors effecting a response in soluble protein expression: Formalized approach using design of experiments. *Analytical Biochemistry* **351**, 122-127.
- Swartz, J.R. (2001) Advances in *Escherichia coli* production of therapeutic proteins. *Current Opinion in Biotechnology* **12**, 195-201.
- Tyers, M. and Mann, M. (2003) From genomics to proteomics. *Nature* **422**, 193-197.
- U.S. Food and Drug Administration . Guideline on general principles of process validation. <http://www.fda.gov/cder/guidance/pv.htm> . 1987.
- Urbain, J.L. (2001) Reporter genes and Imagene. *Journal of Nuclear Medicine* **42**, 106-109.
-

- 
- Urban,A., Ansmant,I., and Motorin,Y. (2003) Optimisation of expression and purification of the recombinant Yol066 (Rib2) protein from *Saccharomyces cerevisiae*. *Journal of Chromatography B* **786**, 187-195.
- Van Suijdam,J.C., Kossen,N.W.F., and Joha,A.C. (1978) Model for oxygen transfer in a shake flask. *Biotechnology and Bioengineering* **20**, 1695-1709.
- Van't Riet,K. (1979) Review of measuring methods and results in nonviscous gas-liquid mass transfer in stirred vessels. *Industrial & Engineering Chemistry Process Design and Development* **18**, 357-364.
- van't Riet,K. (1979) Review of measuring methods and results in nonviscous gas-liquid mass transfer in stirred vessels. *Industrial & Engineering Chemistry Process Design and Development* **18**, 357-364.
- Van't Riet,K. and Tramper,J. (1991) *Basic bioreactor design*. Marcel Dekker, New York.
- Vasina,J.A. and Baneyx,F. (1997) Expression of aggregation-prone recombinant proteins at low temperatures: A comparative study of the *Escherichia coli* *cspA* and *tac* promoter systems. *Protein Expression and Purification* **9**, 211-218.
- Walsh,G. (2002) *Proteins Biochemistry and Biotechnology*. John Wiley and Sons Ltd, Chichester.
- Wang,Y.h., Jing,C.f., Yang,B., Mainda,G., Dong,M.l., and Xu,A.l. (2005) Production of a new sea anemone neurotoxin by recombinant *Escherichia coli*: Optimization of culture conditions using response surface methodology. *Process Biochemistry* **40**, 2721-2728.
- Weickert,M.J., Doherty,D.H., Best,E.A., and Olins,P.O. (1996) Optimization of heterologous protein production in *Escherichia coli*. *Current Opinion in Biotechnology* **7**, 494-499.
- Weiss,S., John,G.T., Klimant,I., and Heinzle,E. (2002) Modeling of mixing in 96-well microplates observed with fluorescence indicators. *Biootechnol.Prog.* **18**, 821-830.
- Welsh,S. and Kay,S.A. (1997) Reporter gene expression for monitoring gene transfer. *Current Opinion in Biotechnology* **8**, 617-622.
- Weuster-Botz,D., Puskeiler,R., Kusterer,A., Kaufmann,K., John,G.T., and Arnold,M. (2005) Methods and milliliter scale devices for high-throughput bioprocess design. *Bioprocess and Biosystems Engineering* **28**, 109-119.
- Weuster-Botz,D. (2005) Parallel reactor systems for bioprocess development. In *Advances in Biochemical Engineering/Biotechnology* Pp. 125-143. Springer Berlin, Heidelberg.
- Wittmann,C., Kim,H.M., John,G., and Heinzle,E. (2003) Characterization and application of an optical sensor for quantification of dissolved O<sub>2</sub> in shake-flasks. *Biotechnology Letters* **25**, 377-380.
-

- 
- Wood, K.V. (1995) Marker proteins for gene expression. *Current Opinion in Biotechnology* **6**, 50-58.
- Xie, L., Hall, D., Eiteman, M.A., and Altman, E. (2003) Optimization of recombinant aminolevulinate synthase production in *Escherichia coli* using factorial design. *Applied Microbiology and Biotechnology* **63**, 267-273.
- Yim, S.C., Jeong, K.J., Chang, H.N., and Lee, S.Y. (2001) High-level secretory production of human granulocyte-colony stimulating factor by fed-batch culture of recombinant *Escherichia coli*. *Bioprocess and Biosystems Engineering* **24**, 249-254.
- You, D., Chen, Q., Liang, Y., An, J., Li, R., Gu, X., Luo, M., and Su, X.-D. (2003) Protein preparation, crystallization and preliminary X-ray crystallographic studies of a thermostable hypoxanthine-guanine phosphoribosyltransferase from *Thermoanaerobacter tengcongensis*. *Acta Crystallographica Section D-Biological crystallography* **59**, 1863-1865.
- Young, T.B. (1979) Fermentation scale-up: industrial experience with a total environmental approach. *Annals of the New York Academy of Sciences* **326**, 165-180.
- Zanzotto, A., Szita, N., Boccazzi, P., Lessard, P., Sinskey, A.J., and Jensen, K.F. (2004) Membrane-aerated microbioreactor for high-throughput bioprocessing. *Biotechnology and Bioengineering* **87**, 243-254.
- Zhang, X. (2006) Optimization of 1, 3-propanediol production by novel recombinant *Escherichia coli* using response surface methodology. *Journal of Chemical Technology and Biotechnology* **81**, 1075-1078.
- Zhang, Z., Szita, N., Boccazzi, P., Sinskey, A.J., and Jensen, K.F. (2006) A well-mixed, polymer-based microbioreactor with integrated optical measurements. *Biotechnology and Bioengineering* **93**, 286-296.
- Zheng, C.J., Han, L.Y., Yap, C.W., Ji, Z.L., Cao, Z.W., and Chen, Y.Z. (2006) Therapeutic targets: Progress of their exploration and investigation of their characteristics. *Pharmacological Reviews* **58**, 259-279.
- Zimmermann, H.F., John, G.T., Trauthwein, H., Dingerdissen, U., and Huthmacher, K. (2003) Rapid evaluation of oxygen and water permeation through microplate sealing tapes. *Biotechnol. Prog.* **19**, 1061-1063.