



Process-scale renaturation of recombinant proteins from inclusion bodies

by

Nicholas Kotlarski

Thesis submitted for the degree of

Doctor of Philosophy

in

The University of Adelaide

Department of Chemical Engineering

Faculty of Engineering

May 1998



To the best of my knowledge and belief, this work contains neither material which has been accepted for the award of any other degree or diploma in any university of any other tertiary institution, nor material previously published or written by another person, except where due reference has been made in the text.

I give consent for this copy of my thesis to be available for photocopying and loan once deposited in the University of Adelaide's Library.

Signed

Date : 19.8.98

Acknowledgments

The work presented in this thesis was supported by an Australian Postgraduate Award for Industry. Accordingly, I am deeply indebted to the University of Adelaide's Department of Chemical Engineering, GroPep Pty Ltd, and the Cooperative Research Centre for Tissue Growth and Repair. The consolidated personalities from these three organisations have been unfailing in their assistance and provided continual moral support during the times of self-doubt. I cannot adequately convey my sincere gratitude to the people who have assisted me.

My supervisors Eric Dunlop, Brian O'Neill, Anton Middelberg, and Geoff Francis deserve particular thanks for their guidance, first into and then through the field of Biochemical Engineering. Their passion for this area of research has been a primary influence in my professional life. Other colleagues that I owe most to are Damien Hocking, Chris Mansell, Andrew MacKenzie, Mark Collis, Andrew Kleinig, Chris Colby, Alex Smith, Lene Jorgensen, and Andrew Talbot.

To my mother, father, and Katrina: thank you for your concern, nurturing and for enduring me over the years. Together with other close friends your effort have contributed to this work in more ways than I can possibly describe.

Financial support from Adelaide Brighton Cement Holdings Ltd in the form of the William Schroder Postgraduate Scholarship in Chemical Engineering and from the University of Adelaide in the form of a Research Abroad Scholarship are gratefully acknowledged.



GroPep Pty Ltd



COOPERATIVE RESEARCH CENTRE
for
TISSUE GROWTH and REPAIR

Summary

Scale-up of a biochemical process involving expression of an Insulin-like Growth Factor-I analogue (LongR³IGF-I) as inclusion bodies within the bacterium *Escherichia coli* has been investigated. The principal focus was directed to the operation of refolding wherein the biological potency of the protein is imparted.

Previous experimental studies on the optimisation of refolding have focused on batch-mode. Most employed pure protein solutions with few examining refolding from unfractionated inclusion body mixtures. Highest yield is favoured by low protein concentration and empirical optimisation of refolding-buffer. These conditions incur major cost for large-scale production. This work addresses all of these important refolding issues: high protein concentration, avoiding expensive additives, and presence of cellular contaminants.

Refolding was modelled as a three-state (intermediate, native, and aggregate) competition between correct refolding (first-order) and aggregation (second-order). Numerical simulation of overall large-scale process operating cost for refolding in batch, fed-batch, and continuous modes was conducted. Dimensionless analysis rendered the findings independent of particular rate constants and production capacity. Batch mode proved the most expensive, with the more economical fed-batch and continuous modes performing almost identically. Subsequent experimental work pursued fed-batch refolding as a superior large-scale operation.

Experimental work addressed three issues: mixing, cellular contamination, and validation of fed-batch refolding. The renaturation system that was investigated employed an aqueous 8 M urea solution under reducing conditions to denature the protein expressed within inclusion bodies. Refolding of this denatured protein was subsequently initiated by dilution give a urea concentration of 2 M under oxidising conditions. The total concentration of LongR³IGF-I was maintained below approximately 2 kg m⁻³ in all refolding trials. Reaction rate constants of correct refolding and aggregation provided a measure of performance. These were estimated by regression of transient protein concentrations in batch refolding trials to the three-state model. LongR³IGF-I concentrations were quantified by analytical chromatography of acid-quenched samples.

Preliminary batch refolding experiments were conducted in the presence of high and low concentrations of soluble and insoluble cellular contaminants.

Mixing rate did not significantly alter the rate of correct refolding (rate constant, $k_N = 0.17 \pm 0.05 \text{ min}^{-1}$). However, it did influence the rate of aggregation: the regressed second-order rate constant of aggregation increased from approximately 1.6 (without stirring) to $3.3 \text{ m}^{-3} \text{ kg}^{-1} \text{ min}^{-1}$ (rapidly stirred). High concentrations of contaminants reduced the accuracy of protein quantitation such that definitive results could not be obtained from the contamination trials. Fed-batch refolding gave approximately 5-fold greater maximal yield of renatured protein than control batch refolding trials in concentrated solutions.

Refolding in fed-batch mode has been identified by process simulation to offer superior economic performance than the current practice of operation in batch mode for large-scale production at high protein concentration. Successful renaturation of LongR³IGF-I in small-scale fed-batch refolding trials have validated the applicability of the method. The key issues of mixing and cellular contamination have been addressed. While the three state model adequately describes batch refolding kinetics, an alternative model is proposed to describe refolding with sufficient accuracy for scale-up.

Table of contents

Disclaimer	i
Acknowledgments	ii
Summary	iii
Table of contents	v
Publications arising from this research	ix
Abbreviations	x
1. RENATURATION OF RECOMBINANT PROTEINS	1
1.1 SCOPE	2
1.2 GENERATING RECOMBINANT THERAPEUTIC PROTEINS	3
1.2.1 Selection of host and expression system	4
1.2.2 Fermentation	6
1.2.3 Cell disruption	8
1.2.4 Primary recovery	8
1.2.5 Dissolution (/purification)	10
1.2.6 Refolding	11
1.2.7 Purification and formulation	12
1.3 PROTEIN STRUCTURE	13
1.3.1 Conformational states	13
1.3.2 Molten globule state	16
1.3.3 Nature of inclusion bodies	16
1.4 DISSOLUTION OF PROTEIN INCLUSION BODIES	20
1.5 REFOLDING	23
1.5.1 Refolding mechanism	23
1.5.2 Off-pathway reactions	26
1.5.3 Traditional laboratory-scale methods for refolding	28
1.5.4 Standard methods for improving refolding	31
1.5.5 Recent developments in protein refolding	33
1.5.6 Issues for large-scale refolding by dilution	36
1.5.6.1 Process simulation	36
1.5.6.3 Mass transfer	38
1.5.6.3 Contaminants	39
1.6 CONTAMINANTS AND PROCESS MODELLING	42
1.6.1 Traditional approach to biological variability	42
1.6.2 Key Parameters	44
1.6.3 Compartmentalisation	44
1.6.3.1 Soluble compartment	47
1.6.3.2 Insoluble compartment	47
1.6.3.3 Proposed compartments	48
1.7 PROTEIN OF INTEREST	51
1.8 THESIS STRATEGY	54

2. MATERIALS AND METHODS	57
2.1 MATERIALS	58
2.2 ANALYTICAL METHODS	59
2.2.1 Analytical disc centrifuge	59
2.2.2 Total protein quantitation	59
2.2.3 High performance liquid chromatography	60
2.2.4 Spectrophotometry	62
2.2.5 Electrophoresis	62
2.2.6 Concentration of acid-quenched samples	64
2.2.7 Densitometry	64
2.2.8 General experimental methods	64
2.3 NUMERICAL METHODS	66
2.3.1 Process simulation	66
2.3.2 PeakFit™	67
2.3.3 Mathcad	69
2.3.4 Estimates of variance in rate constants	69
2.4 PROCESSING OF STARTING CELLULAR FRACTIONS	70
2.4.1 Host	71
2.4.2 Vector	71
2.4.3 Inoculum preparation	71
2.4.4 Fermentation	72
2.4.5 Homogenisation	74
2.4.6 Batch centrifugation	74
2.4.7 Continuous centrifugation	75
2.5 CHARACTERISATION OF STARTING CELLULAR FRACTIONS	76
2.6 DISSOLUTION OF INCLUSION BODIES	82
2.6.1 Chemical limitation of dissolution	82
3. MATHEMATICAL FRAMEWORK	83
3.1 KINETICS OF REFOLDING	84
3.1.1 Kinetic models of correct refolding	84
3.1.2 Rate of correct refolding	86
3.1.3 Models that account for the aggregation reaction	87
3.1.4 General kinetic model for reactor design	89
3.1.5 Rate constants	91
3.1.6 Recent kinetic studies	92
3.2 REACTOR DESIGN EQUATIONS	95
3.2.1 Batch refolding	96
3.2.2 Fed-batch refolding	97
3.2.3 Continuous refolding	99
3.3 GENERIC LARGE-SCALE PROCESS DESCRIPTION	100
3.3.1 Batch refolding	100
3.3.2 Fed-batch refolding	101
3.3.3 Continuous refolding	101
3.4 DESIGN BASIS	103

3.5 OBJECTIVE FUNCTION	104
3.6 PROCESS COSTING	105
3.6.1 Unit purchase costs	105
3.6.1.1 Fermenter	105
3.6.1.2 Steriliser	106
3.6.1.3 Compressor	107
3.6.1.4 Homogeniser	107
3.6.1.5 Centrifuge	107
3.6.1.6 Storage tanks	108
3.6.1.7 Dissolution reactor	108
3.6.1.8 Refolding reactors	109
3.6.1.9 Ultrafiltration	110
3.6.1.10 Total Direct-Fixed-Capital-dependent costs (DFC costs)	112
3.6.2 Cost of consumables	112
3.6.2.1 Fermentation medium	112
3.6.2.2 Urea	113
3.6.2.3 Additives for dissolution and refolding	114
3.6.2.4 Depth filter	114
3.6.2.5 Ultrafiltration membrane	114
3.6.2.6 Total cost of process consumables	115
3.6.3 Waste management	115
3.7 SELECTION OF PROCESS PARAMETERS	118
3.7.1 General process parameters	121
3.7.2 Protein-specific rate constants	124
4. SIMULATION: OVERALL PROCESS INTERACTIONS	125
4.1 COMPARISON OF OPERATING MODES - BASE CASE	126
4.1.1 Batch mode	126
4.1.2 Fed-batch mode	128
4.1.3 Continuous mode	130
4.1.4 Transient intermediate concentration	132
4.1.5 Discussion - base case	133
4.2 SENSITIVITY ANALYSIS	137
4.2.1 Variation in equilibrium	137
4.2.2 Sensitivity to consumable cost estimates	139
4.3 IMPLICATIONS FOR EXPERIMENTAL STUDIES	140
4.4 CONCLUSIONS FROM SIMULATION	142
5. REFOLDING KINETICS FOR MODIFIED PROCESSES	143
5.1 DISSOLUTION	144
5.1.1 Experimental protocol - dissolution	144
5.1.1.1 Mixing experiments	144
5.1.1.2 Contaminant trials	145
5.1.1.3 Dissolution for concentrated refolding	147
5.1.1.4 Dissolution for fed-batch refolding	148

5.1.2 Results and discussion - dissolution	148
5.1.2.1 Clean inclusion bodies	148
5.1.2.2 Dirty inclusion bodies	157
5.2 MIXING EFFECTS	161
5.2.1 Experimental protocol - mixing	161
5.2.2 Results and discussion - mixing	163
5.3 SOLUBLE CONTAMINANTS	170
5.3.1 Experimental protocol - soluble contaminants	170
5.3.2 Results and discussion - soluble contaminants	171
5.4 INSOLUBLE CONTAMINANTS	176
5.4.1 Experimental protocol - insoluble contaminants	176
5.4.2 Results and discussion - insoluble contaminants	177
5.5 CONCENTRATED REFOLDING	182
5.5.1 Experimental protocol - concentrated refolding	182
5.5.2 Results and discussion - concentrated refolding	183
5.6 FED-BATCH REFOLDING	189
5.6.1 Experimental protocol - fed-batch refolding	189
5.6.2 Results and discussion - fed-batch refolding	191
5.7 IMPLICATIONS OF EXPERIMENTAL FINDINGS	200
6. GENERAL DISCUSSION	202
6.1 MASS-TRANSFER	203
6.2 STARTING MATERIAL	204
6.3 THREE-STATE REFOLDING KINETICS	208
6.4 LARGE-SCALE REFOLDING	210
6.5 FUTURE WORK	211
7. CONCLUSIONS	213
8. REFERENCES	215
9. APPENDICES	237
9.1 NOMENCLATURE	238
9.2 DERIVATION OF REACTOR DESIGN EQUATIONS	241
9.2.1 Batch reactor	241
9.2.2 Fed-batch reactor	243
9.2.3 Continuous stirred tank reactor (CSTR)	245
9.3 MASS-TRANSFER LIMITED DISSOLUTION	248

Publications arising from this research

Kotlarski, N., O'Neill, B. K., Francis, G. L., Middelberg, A. P. J., and Dunlop, E. H., Improved large-scale renaturation of proteins from inclusion bodies, *Manuscript in preparation*.

Kotlarski, N., O'Neill, B. K., Francis, G. L., and Middelberg, A. P. J., (1997), Design analysis for refolding monomeric protein, *AIChE J.*, **43**: 2123-2132.

Presentations

Kotlarski, N., Francis, G. L., O'Neill, B. K., Middelberg, A. P. J., and Dunlop, E. H., (1998), Improving protein refolding by adopting a fed-batch approach, *1998 AIChE Annual Meeting, November 15-20 1998, Miami Beach, U.S.A., submitted*.

Kotlarski, N., Francis, G. L., O'Neill, B. K., Middelberg, A. P. J., and Dunlop, E. H., (1998), Improved protein refolding at high concentration using a fed-batch approach, *14th Australasian Biotechnology Conference, 19-23 April 1998, Adelaide, Australia*, P 14.

Kotlarski, N., O'Neill, B. K., Francis, G. L., and Middelberg, A. P. J., (1997), Cost minimisation for refolding therapeutic proteins, *Proceedings of CHEMECA '97, 29 September-1 October 1997, Rotorua, New Zealand*, PB4C. (ISBN: 0-0908-22-0).

Kotlarski, N., O'Neill, B. K., Francis, G. L., and Middelberg, A. P. J., (1996), The cost of suboptimal refolding of proteins that aggregate during batch refolding, *10th International Biotechnology Symposium, 25-30 August 1996, Sydney, Australia*, P1-56.

Kotlarski, N., O'Neill, B. K., and Middelberg, A. P. J., (1995), The impact of competing proteolysis during protein refolding on process economics, *Proceedings of the 4th Pacific Rim Biotechnology Conference, 6-9 February 1995, Melbourne, Australia*, pp. 239-340.

Middelberg, A. P. J., **Kotlarski, N., and O'Neill,** (1995), The optimisation of biochemical processes involving protein refolding with competing proteolysis, *The 1995 IChemE Research Event*, **2**: 1160-1162.

Greenwood, M., **Kotlarski, N., O'Neill, B. K., Falconer, R., Francis, G. L., and Middelberg, A. P. J.,** (1994), Dissolution and refolding of Long-R³-IGF-I, *The 1994 IChemE Research Event*, **1**: 250-252.

Abbreviations

BOD	Biological oxygen demand
CAB	Carbonic Anhydrase B
CHO	Chinese Hamster Ovary (cell)
CSTR	Continuous stirred-tank reactor
df	Degrees of freedom
DFC	Direct-fixed-capital
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol ($C_4H_{10}O_2S_2$)
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Sodium ethylenediaminetetraacetate ($C_{10}H_{14}N_2O_8Na_2$)
Gly	Glycine ($C_2H_5NO_2$)
GuHCl	Guanidine hydrochloride ($CH_5N_3 \cdot HCl$)
HEWL	Hen Egg White Lysozyme
HPLC	High performance liquid chromatography
IBs	Inclusion bodies
IGF-I	Insulin-Like Growth Factor-I
IL-2	Interleukin-2
IPTG	Isopropyl β -D-thiogalactopyranoside ($C_9H_{18}O_5S$)
LPS	Lipopolysaccharide
MilliQ™	Type 1 laboratory reagent grade water purified using Millipore™ system
MSE	Mean squared error
OX- β -ME	2-hydroxyethyl disulfide ($(HOC_2H_4)_2S_2$)
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PDI	Peptidyl Disulfide Isomerase
PFR	Plug-flow reactor
PGK	Phosphoglycerate Kinase
PPI	Peptidyl-Prolyl <i>cis-trans</i> Isomerase
Sarcosyl	Sodium-lauryl sarcosinate ($C_{15}H_{29}NO_3Na$)
SDS	Sodium docecyl sulfate (or sodium-lauryl sulfate, $C_{12}H_{25}O_4SNa$)
SSE	Sum of squared error
TFA	Trifluoroacetic acid ($C_2HF_3O_2$)
TPA	Tissue Plasminogen Activator
Tris	Tris[hydroxymethyl]aminomethane ($C_4H_{11}NO_3$)



1. RENATURATION OF RECOMBINANT PROTEINS

Isolated DNA coding for desirable proteins arguably presents an inexhaustible supply of useful biologically active products. Techniques for incorporating DNA into efficient vectors, subsequently transforming microorganisms, and expressing protein are now routine. Product isolation from the microorganisms that produce them can be achieved using standard practices. However, methods for efficient *large-scale* processing have not kept pace with the rapid advances in recombinant manipulation. There is a need to develop processing technology to realise the benefits that recombinantly-produced proteins offer.

The work presented in this thesis is aimed at improving the efficiency of large-scale renaturation of recombinant monomeric proteins expressed as inclusion bodies within the bacterium *Escherichia coli*. The approach employed has been characterisation of unit interactions and application of fundamental chemical engineering theory to the refolding operation which historically has been scaled-up without considering integrated process design.

1.1 SCOPE

Numerous combinations of host organism and expression system produce desired proteins in a biologically inactive form, requiring renaturation before they become useful. This study considers expression of a monomeric analogue of Insulin-Like Growth Factor-I (IGF-I) as inclusion bodies within *Escherichia coli* (*E. coli*). A basis for rational process scale-up of renaturation has been sought by:

- Limited characterisation of the dissolution operation.
- Extensive investigation of the refolding operation.

The findings, while specific for the protein considered, should have relevance to other monomeric proteins expressed as inclusion bodies. Overall, a methodology for scale-up of refolding will be presented that addresses:

- A suitable kinetic model for describing the refolding of monomeric proteins.
- Simulation of overall process performance to compare modes of operation and direct experimental investigations.
- Impact of refolding on overall process economics.
- Whether mass-transfer limits the observed kinetics of refolding.
- Extent of purification required prior to refolding.
- Validation of simulation findings.

This introductory chapter outlines the current methods used to produce useful recombinant proteins in *E. coli*. The fundamental basis of protein refolding is presented and then process-scale renaturation is introduced in the context of an overall production process to establish the relevance of the work presented in this thesis. A structured method to deal with host-cell contaminants during renaturation is presented. The analogue of IGF-I used in all experimental studies is described and the strategy of the approach adopted is outlined.

1.2 GENERATING RECOMBINANT THERAPEUTIC PROTEINS

There is a range of cell types routinely used as hosts for expression of recombinant protein: prokaryotes, fungi, mammalian and insect cells. Each system has unique processing requirements, advantages, and disadvantages.

A typical production process, for a monomeric protein that is expressed in insoluble form within *E. coli*, is presented in Figure 1.1. The protein product is first expressed within *E. coli*, released by high-pressure homogenisation, recovered by differential sedimentation, dissolved, and refolded. The refolding operation is the main focus of this thesis. Each operation is discussed further in sections 1.2.2 to 1.2.7.

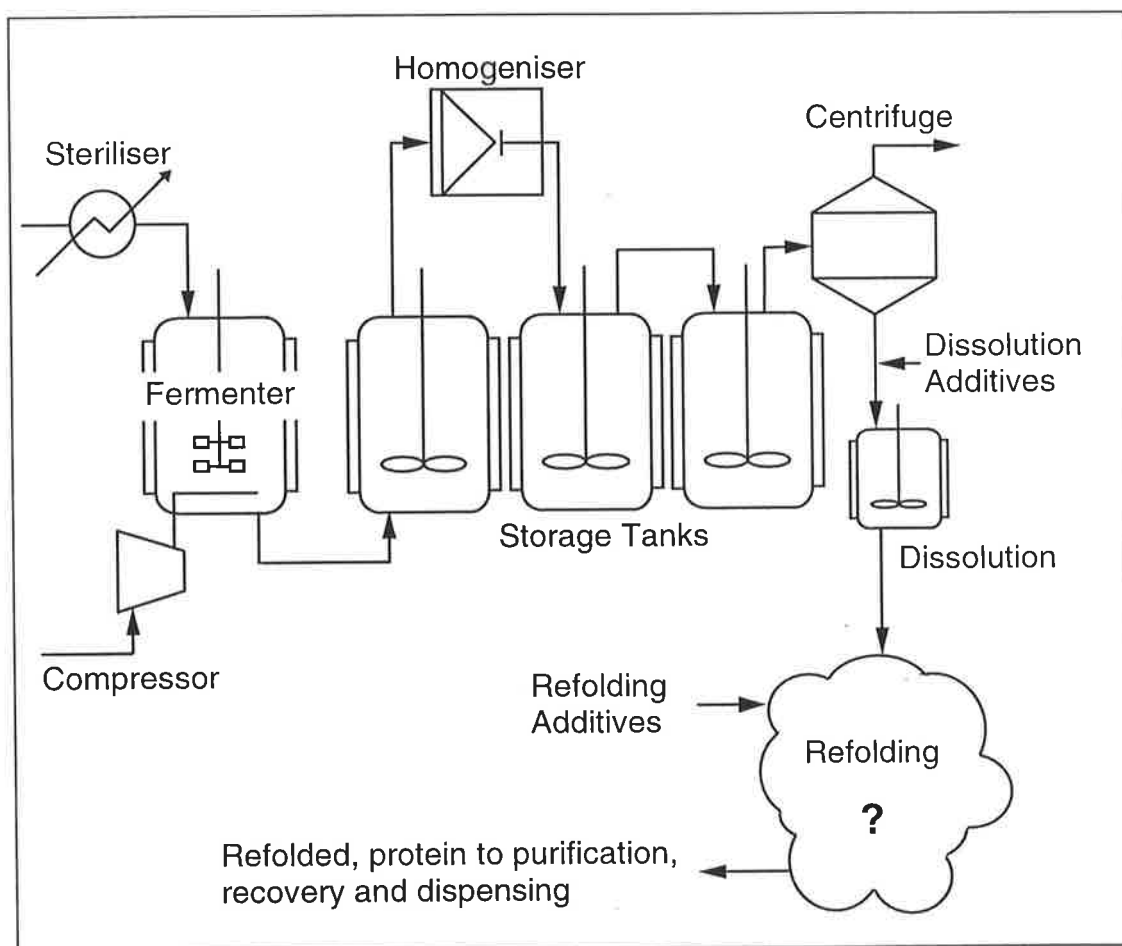


Figure 1.1: Generic flow sheet for a production process to isolate and renature protein expressed as inclusion bodies in *E. coli*.

1.2.1 Selection of host and expression system

The most commonly used host organism is *E. coli* (Georgiou, 1988). It is one of the most simple prokaryotes, capable of rapid growth in simple, defined media. Extensive research into its molecular composition render it the best characterised of any bacteria (Kane and Hartley, 1988; Schaechter and Niedhardt, 1987). When combined with a suitable vector, it can express high concentrations of the desired protein product (Chaudhuri, 1994a; Gold, 1990). Use of *E. coli* as the host has the advantage of reducing the risk of adventitious viral contamination harmful to mammals. These characteristics make *E. coli* an obvious choice for expressing recombinant therapeutic proteins (Georgiou and Baneyx, 1996).

The major disadvantage of *E. coli* as a host is that it is incapable of post-translational modification of non-native (heterologous) proteins *e.g.* glycosylation or multimer assembly. This has generally restricted its use to expression of simple monomeric proteins that do not require glycosylation. For more complex proteins, yeasts (*e.g.* *Saccharomyces cerevisiae* or *Pichia pastoris*) or mammalian cell lines (*e.g.* Chinese Hamster Ovary (CHO) cells) are often selected as superior host organisms (Georgiou, 1988).

E. coli must be genetically engineered to express heterologous proteins. This is achieved by introducing deoxyribonucleic acid (DNA) that codes for the desired protein into these cells. The most effective strategy is to include the fragment of DNA in a plasmid. A plasmid is extra-chromosomal DNA that will be inherited by progeny during culture of a plasmid-containing parent cell. Plasmids inserted into *E. coli* generally contain DNA that confers resistance to an antibiotic to the host cell. Antibiotic resistance can then be used as a selection pressure to ensure that the viable population contains the plasmid. Additionally, the plasmid must include a promoter sequence preceding the structural gene to regulate expression of the protein product. Generally a linker sequence is inserted between the promoter and structural genes.

The switch used to induce expression of the protein product depends on the promoter sequence used in the plasmid. A commonly used promoter system relies on the *trc* gene to induce expression once plasmid-bearing cells are exposed to the chemical, isopropyl β -D-thiogalactopyranoside (IPTG). Alternative expression systems may rely on stimulation of physiological stresses such as oxygen depletion, temperature change, or nutrient starvation as the switch. The appropriate choice of induction system is an exercise in optimisation

where efficiency of expression, cost, and robust performance must be considered. Which ever system is chosen, initiating induction of product expression stresses the host organism which results in a physiological response (Georgiou, 1988; Harcum and Bentley, 1993). Typically, proteolytic activity is increased, particularly due to the La protease, responsible for degradation of misformed proteins (Chung, 1993; Kresze, 1991).

Degradation of the desired protein product by endogenous host proteases is a major problem in bioprocessing. Trace amounts of host proteases are potentially able to cleave valuable product into useless fragments unless carefully removed during the production process. The well characterised genetic structure of *E. coli* allows construction of mutant strains that are deficient in particular proteases, thus eliminating any potential problem (Baneyx and Georgiou, 1990, 1991; Georgiou and Baneyx, 1996; Hoenger *et al.*, 1993). The drawbacks to producing such mutant strains are that the specific protease responsible for degradation must first be identified and the growth rate of mutants may be lower than the wild-type (Baneyx *et al.*, 1991; Meerman and Georgiou, 1994).

High level expression of heterologous proteins within *E. coli* generally results in product accumulation within the cytoplasm as dense insoluble particles called inclusion bodies (Marston, 1986; Rudolph and Lilie, 1996; see section 1.4.3). Strategies for avoiding formation of inclusion bodies have been developed. Lowering the culture temperature during induction of product synthesis has been observed to increase the yield of soluble protein at the expense of reduced growth rate (Mitraki *et al.*, 1991). Expression of eukaryotic proteins as fusion proteins with different leader sequences has been successfully employed to direct the protein from the cytoplasm onto the cell surface (Georgiou *et al.*, 1994) or into the medium (Cheah *et al.*, 1994; Mascarenhas *et al.*, 1994; McCoy and LaVallie, 1994; Wong *et al.*, 1988) in active forms. Alternatively coexpression of molecular chaperones and refolding isomerases have been shown to improve soluble protein expression (Georgiou and Valax, 1996; Humphreys *et al.*, 1996). Coexpression of helper proteins has the disadvantage of placing a greater physiological load on the host, thereby reducing the rate of product expression (Baneyx *et al.*, 1991; Samuelsson *et al.*, 1996).

Failure to exploit the *in vivo* refolding mechanisms of the cell and the need to renature the inactive protein are obvious disadvantages. However, the *in vivo* strategies adopted to express protein product in a soluble form have not provided an ideal alternative (Rudolph

and Lilie, 1996) so production by expression as inclusion bodies is more common than soluble expression (Hevehan and De Bernardez-Clark, 1997). This is because there are significant advantages to expression of protein product as inclusion bodies. The protein sequestered within inclusion bodies is more resistant to cellular proteases, less toxic to the host cell, and can be rapidly concentrated from homogenate by filtration or centrifugation (De Bernardez-Clark and Georgiou, 1991; Enfors, 1992; Kane and Hartley, 1988, 1991; Marston, 1986; Valax and Georgiou, 1993b). Rational protein engineering to expression products as peptide fusions can even increase the production of inclusion bodies (Belagaje, 1995). Termination conditions at the completion of fermentation can also increase the amount of insoluble product that can be recovered (Builder and Ogez, 1986; Dorin *et al.*, 1995). The question of whether inclusion body formation is favourable or a disadvantage is a matter of overall process optimisation with minimum cost as the objective (Ernst *et al.*, 1997; Hamel *et al.*, 1996).

The scope of this thesis has been restricted to expression systems in which the desired product is expressed as inclusion bodies within *E. coli*. In these expression systems, once the product has accumulated within the cell, it must be recovered and renatured by dissolving and refolding as outlined below.

1.2.2 Fermentation

The purpose of fermenting recombinantly engineered organisms is to generate broth containing a high concentration of desired product. For inducible cells, a concentrated culture is first generated without triggering product expression so that a high cell density can be rapidly established without the physiological penalty of synthesising product. Once the culture is sufficiently advanced, product expression is initiated.

Unlike cultures of mammalian cells or other simple eukaryotic systems, *E. coli* does not require complex medium to sustain growth (Cleland and Wang, 1993). Simple sugars provide an adequate carbon source and, together with defined mixtures of inorganic compounds and trace elements, can supply all nutritive requirements. The rapid metabolism of *E. coli* can lead to mass-transfer limitation of the oxygen required for respiration at high cell densities, even in well mixed bioreactors with airflow rates up to

one volume per volume per minute (v.v.m) (Turner *et al.*, 1994). Accumulation of acetate as a by-product of respiration in nutrient-rich conditions can result in significant inhibition of growth rate (Ko *et al.*, 1994; Turner *et al.*, 1994). Therefore, a fed-batch approach is commonly used to control the feed rate of nutrients and reduce acetate production and oxygen limitation for fermentations conducted at high cell densities (Cockshott and Bogel, 1992; Georgiou, 1988; Turner *et al.*, 1994; Yoon *et al.*, 1994a, 1994b).

Fermentation conditions can have a significant impact on the physiological state of the bacteria being cultured (Fishov *et al.*, 1995; Niedhardt, 1987). While some of the obvious effects of varying culture conditions are different growth rate and concentration of product (Kazan *et al.*, 1995), some of the less apparent changes can have a greater impact on the overall process performance.

When growing under normal conditions approximately 1-2% of the total *E. coli* protein is degraded by intracellular proteases each hour to maintain growth. However, under conditions where nutrients are significantly limited, 5-12% of the total cellular protein is degraded per hour, attesting to an increase in general protease activity (Lazdunski, 1989). These observations are as a result of the 'stringent response' which is normally observed whenever microorganisms experience a significant change in environmental conditions (Harcum and Bentley, 1993). Under these conditions *E. coli* alter their cellular constituents to adapt to their environment (Schaechter and Neidhardt, 1987). The relative proportions of other non-protease components also varies. For example, the distribution of outer membrane proteins in *E. coli* have been observed to be significantly affected by culture conditions (Lugtenberg *et al.*, 1976; Manning and Reeves, 1977; Osborn and Wu, 1980).

Since absolute separation of host cell material from inclusion bodies cannot be achieved (see section 1.3.3), varying fermentation conditions will alter the composition of the inclusion body material isolated for renaturation (Valax and Georgiou, 1993a). Increased proteolytic activity in the isolate may result in product degradation during subsequent processing once it is released from the protective state of the inclusion body (Babbitt *et al.*, 1990 & 1991; Sugimura and Nishihara, 1988). If interaction between product and outer membrane proteins occurs during refolding, altering amounts of membrane proteins may decrease the renaturation yield (see section 1.5.6.3). Therefore, the impact of fermentation conditions on subsequent operations must be considered.

1.2.3 Cell disruption

To recover the protein product from fermentation broth it must be extracted from within the host organism. Many methods have been developed to rupture the cell envelope and release intracellular products. The methods may be broadly classified as mechanical or non-mechanical (Middelberg, 1995). Mechanical methods of cell disruption include sonication, wet milling, and high-pressure homogenisation. Non-mechanical methods include physical disruption, such as repeated freeze/thaw cycles (Johnson and Hecht, 1994) and chemical or enzymatic disruption, such as exposure to chaotropic agents or lysozyme, respectively (French *et al.*, 1996; Hart *et al.*, 1994; Kung, 1984).

High-pressure homogenisation is the preferred method for large-scale disruption for *E. coli* (Chaudhuri, 1994a; Middelberg, 1995). This is because it is a mechanical method and does not require use of additives that are potentially detrimental to product recovery, it is robust, and capable of sustaining high capacity (Siddiqi, 1996). Disruption by high-pressure homogenisation relies on the combined effects of pressure change, viscous shear stresses and impingement on a stationary surface (Kleinig and Middelberg, 1997). Multiple passes are normally required to achieve near total disruption and give a slurry with properties suitable for separation of the inclusion bodies from cellular debris (Datar and Rosen, 1987; Middelberg *et al.*, 1992). Physical damage of inclusion bodies by the extreme action of high-pressure homogenisation has been demonstrated to be insignificant (Olbrich, 1989).

1.2.4 Primary recovery

Recovery of insoluble inclusion bodies from the cell homogenate presents an opportunity to increase the concentration of the desired product and remove contaminants. Initial studies into inclusion body processing used low-speed batch centrifugation to recover the insoluble fraction of the homogenate (Marston, 1986). On the process-scale, differential sedimentation in a continuous disc-stack centrifuge, or crossflow filtration are the preferred methods for collection of product.

The most recent development in the isolation of inclusion body protein has been the use of two-phase liquid extraction to preferentially dissolve the protein product in the cell homogenate, thereby eliminating the need to use either centrifugation or filtration (Köhler

et al., 1993). Two-phase extraction has been coupled with chemical disruption to give a production process that does not require mechanical disruption or inclusion body collection (Hart *et al.*, 1994; Swartz *et al.*, 1996). A disadvantage of two-phase extraction is that it requires extensive empirical optimisation to establish suitable conditions (Rudolph and Lilie, 1996) and this explains why such an approach is not generally employed.

Differential sedimentation is still the most commonly employed operation for primary recovery (Hart *et al.*, 1994; Middelberg, 1996a). It offers superior economic performance for large-scale operation (Datar and Rosen, 1993) and has been well characterised (Olbrich, 1989; Wong, 1997). Crossflow filtration promises improved performance, but requires further development (Meagher *et al.*, 1994).

The thermodynamic stability of protein in the form of inclusion bodies is normally exploited to improve the separation of host cell contaminants from the product during centrifugation (see section 1.3.3). Centrifugation is typically conducted after suspending the inclusion bodies in solutions of urea, ethylene diaminetetraacetic acid (EDTA), lysozyme, Triton X-100, or deoxycholate. The inclusion body protein remains stable at the chosen concentration of these reagents, but exogenous host proteins are preferentially solubilised (Bowden *et al.*, 1991; Estapé and Rinas, 1996; Hart *et al.*, 1990; Langely *et al.*, 1987; Marston *et al.*, 1984; Rhee *et al.*, 1990). Alternatively, the viscosity and density of the bulk fluid may be altered to improved selective recovery of the more dense inclusion body particles (Fischer *et al.*, 1993a; Hershenson *et al.*, 1990).

Most reviews of inclusion body processing advocate washing of inclusion bodies before starting renaturation (Chaudhuri, 1994a; De Bernardez-Clark and Georgiou, 1991; Fischer *et al.*, 1993a; Rudolph, 1990). This is reflected in the process patents for production of many products. For example, Genentech, Inc. holds one of the first patents for improving the selectivity of centrifugation, conducting the operation at high ionic strength (Olson, 1985). A more recent patent held by Rhone-Poulenc Rorer Pharmaceuticals Inc. covers use of wash-solutions containing deoxycholate, lithocholate, cholate, anionic-, cationic-, and zwitterionic-detergents, polyoxyethylene sorbitol esters or polyoxyethylene p-t octylphenol (Darling *et al.*, 1996).

The features of inclusion body processing on the microscopic scale are presented in Figure 1.2. *E. coli* are fermented, induced to express product, homogenised, and centrifuged.

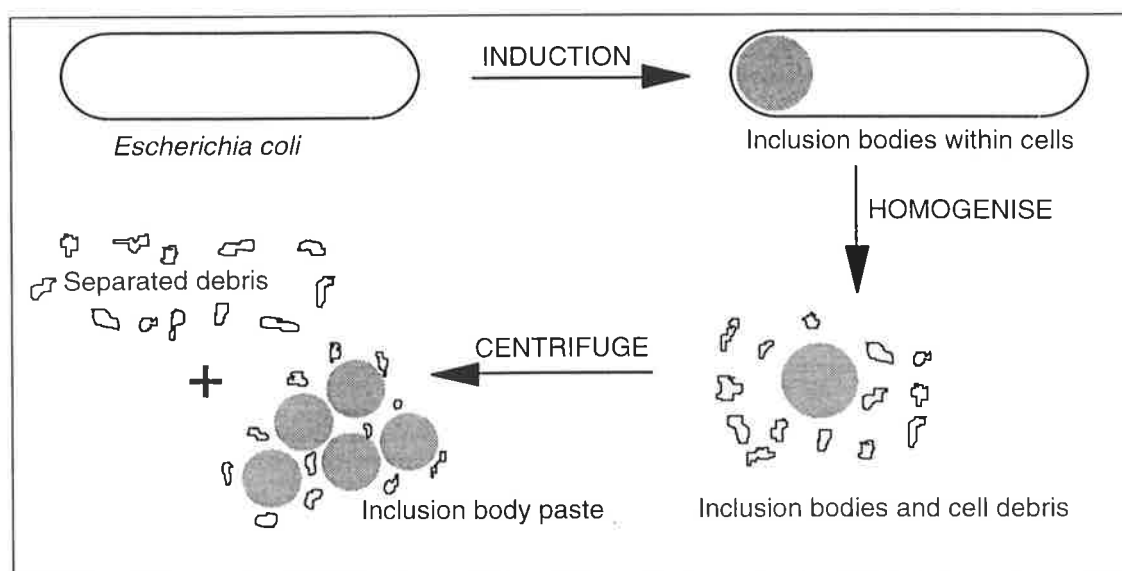


Figure 1.2: Schematic of fermentation, homogenisation and primary recovery of inclusion bodies.

1.2.5 Dissolution(/purification)

The protein trapped within inclusion bodies must be liberated to give monomers before biological potency can be imparted. This process is referred to as dissolution and is illustrated in Figure 1.3.

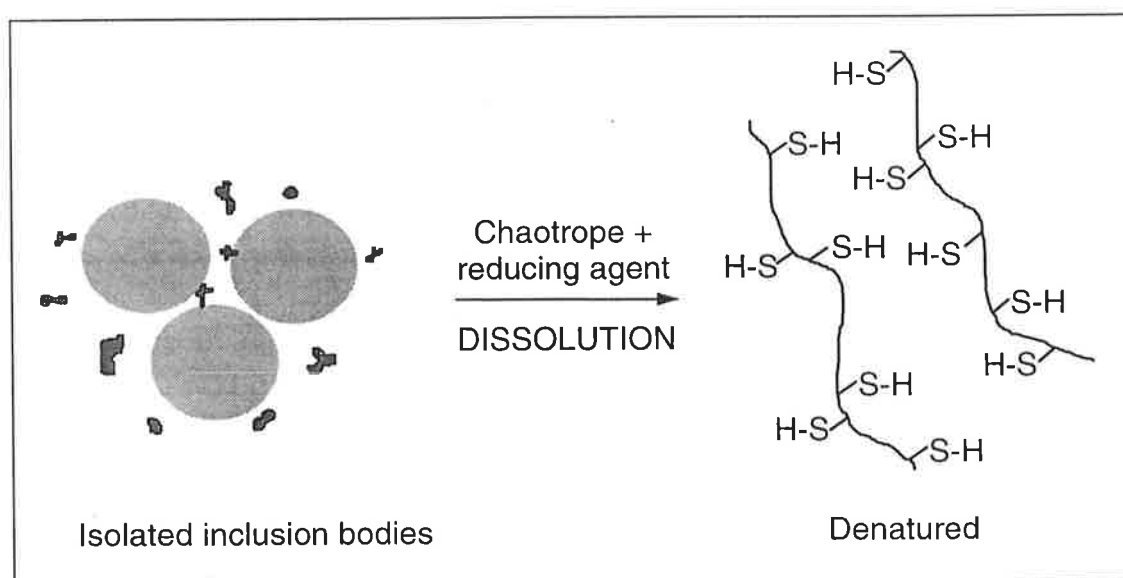


Figure 1.3: Schematic of the dissolution operation: isolated inclusion bodies are incubated under denaturing conditions to give a mixture containing reduced, unfolded monomers of recombinant protein

Dissolution can be achieved using solutions of detergents such as Triton, sodium dodecyl sulfate (SDS) or cetyltrimethylammonium chloride (CTAC) (Cardamone *et al.*, 1995; Olson, 1985). However, the chaotropic agents urea and guanidine hydrochloride (GuHCl) are used more commonly (Dill and Shortle, 1991; Fischer *et al.*, 1993a). Most of the therapeutic proteins that are produced in bacteria have internal disulfide bonds in their native state (Chaudhuri, 1994a). Therefore, a reducing agent is routinely added to break any existing inter- or intra-molecular disulfide bonds and alkaline conditions are preferred. Commonly used reducing agents are 2-mercaptoethanol and dithiothreitol (DTT) (De Bernardez-Clark and Georgiou, 1991; Fischer *et al.*, 1993a). Dissolution is discussed in greater detail in section 1.4.

Solubilising the protein of interest provides an opportunity to separate it from host cell contaminants. For example, if proteases are inactive while denatured, but can subsequently be reactivated, removing them early in the process is most advantageous (Jaenicke and Rudolph, 1986). Centrifugation can be used to remove the material that remains insoluble (Chaudhuri, 1994a; Builder and Ogez, 1986) and chromatographic separation can be used to reduce the concentration of soluble contaminants (Marston and Hartley, 1990). Size-exclusion or ion-exchange chromatography are also commonly used (Babbitt *et al.*, 1991; Knuth and Burgess, 1987; Olson, 1985; Rudolph and Fischer, 1990).

1.2.6 Refolding

Refolding is the *in vitro* operation that imparts biological activity to the denatured protein obtained by dissolution. It is physically achieved by changing the solvent from denaturing to one favouring the native state. Generally, finding conditions that favour the native state is not difficult. Many laboratory-scale methods have been developed to accomplish exchange of the solvents: direct dilution, dialysis, diafiltration, and size-exclusion chromatography. The most commonly used method is direct batch dilution into a solution of appropriate chemical composition (Fischer *et al.*, 1993a; Galliher, 1991; Rudolph, 1990). However, aggregation during the transformation between states, resulting in product loss, is a major problem with all methods (Chaudhuri *et al.*, 1996; Fischer *et al.*, 1993a; Georgiou, 1988; Schein, 1990). Therefore, solvent exchange is normally conducted at very low protein concentration in a single batch containing multiple cosolvents that

reduce interaction between monomers. There are many publications reporting extensive empirical optimisation of chemical additives to reduce aggregation of particular proteins.

Large-scale refolding places high demands on the renaturation protocol as expensive additives and huge processing capacity to deal with highly dilute solutions detract from overall economic viability (Burton *et al.*, 1989). Despite this, it is common for large-scale processes to be scaled directly from laboratory-scale processes (Chaudhuri, 1994b).

Work into development of refolding methods that are capable of achieving high renaturation yield at high overall protein concentration has identified the use of reverse-micelles, two-phase mixtures, affinity ligands, concurrent chromatographic separation, and rational protein engineering as alternatives to empirical optimisation of chemical additives. All of these methods promise more general solutions to increasing the selectivity of refolding in favour of active product, but they are not yet developed to the stage where they are generally applicable. A more elementary approach that has been successfully demonstrated, is to conduct refolding by a series of dilutions, in fed-batch or continuous modes. The optimum chemical environment identified from batch refolding trials can be used. Improved processing methods for large-scale refolding are discussed in greater depth in section 1.5.

1.2.7 Purification and formulation

The cellular contaminants that remain in solution with the desired active protein product after refolding must be removed. Separation processes that use crossflow filtration, chromatography, reverse-micelles, precipitation, and affinity precipitation procedures have been developed (Dorin *et al.*, 1995; Kaul and Mattiasson, 1992; Walker *et al.*, 1994). However, separation is generally achieved using a series of different forms of chromatography operations (Kaul and Mattiasson, 1992). Commonly used systems are: gel filtration to separate on the basis of size, ion-exchange on the basis of charge, reverse-phase on the basis of hydrophobicity, and affinity chromatography on the basis of specific affinity. Freeze drying may be employed to remove volatile solvents that may be used during the chromatographic purification. Once the protein has been purified sufficiently, it is dispensed for distribution.

1.3 PROTEIN STRUCTURE

The highly specific biological activity that proteins exhibit does not arise because of the gross atomic composition of the molecules alone; the specific sequence of amino acid residues must assume a particular stereoisomer. The physical organisation of the secondary structure elements of protein into the three-dimensional tertiary structure of a protein from an unstructured state is referred to as folding. Complex folding systems have evolved to facilitate the *in vivo* transformation of the nascent peptide to a biologically active conformation. The development of recombinant engineering techniques has presented the problem of achieving the same transformation *in vitro*. As described above, protein can generally be expressed to give the mature sequence, but it is not fold correctly. The inactive material expressed within inclusion bodies must be renatured to give the correct conformation by firstly denaturing and then refolding the polypeptide chains.

When one considers the range of possible amino acid sequences and bond angles between neighbouring residues of a protein, it becomes a significant exercise in combinatorial mathematics just to determine the number of potential folding stereoisomers. Yet refolding can occur spontaneously within the time-scale of minutes. This observation lead Levinthal (1968) to propose that refolding must follow a predefined pathway *i.e.* that the native conformation is not achieved by random search of every possible conformer. Soon after, Anfinsen (1973) proposed that the sequence of amino acids of a protein (primary structure) provided all of the information necessary to define the tertiary structure of the molecule; effectively the second interpretation of the genetic code where the first interpretation provides the actual sequence. Since then, there has been extensive research into the thermodynamics and reaction pathways of protein refolding.

1.3.1 Conformational states

The driving force for proteins to change structure is minimisation of conformational energy (Dill, 1990; Jaenicke, 1991). However, the tertiary structure of biologically active proteins in aqueous solution is barely stable at room temperature (see Dill, 1987 and Dill 1990 for reviews of protein thermodynamics). The free energy of stabilisation for a typical protein is approximately $50 \pm 15 \text{ kJ mol}^{-1}$ (Creighton, 1987; Jaenicke, 1991; Jaenicke, 1995). This is equivalent to only one or two hydrogen bonds; a minor energy barrier when considering

the overall sum of electrostatic, steric repulsion, Van Der Waals, hydrogen bonding, and hydrophobic interactions that occur not only between neighbouring residues, but also long-range within a globular molecule. The decrease in entropy associated with folding into a highly structured globular molecule presents the major opposing force (Creighton, 1987; Dill 1985; Pace *et al.*, 1996). Therefore, interaction between protein molecules and their surrounding solvent is one of the critical factors controlling the stable conformation.

When a protein is in an environment in which it is biologically potent, it is referred to as being in the native state. As outlined above, the native protein conformation is delicately balanced. Even proteins that are stabilised by intramolecular disulfide bonds in their native state cannot be considered as having rigid tertiary structure, but rather a time-average of rapidly oscillating structures as competing free energy contributions interact in a dynamic system (Dill, 1987; Kim and Baldwin, 1982). The time-averaged conformation is maintained in such a way that the molecule is able to exhibit its biological activity.

When a protein is exposed to a chemical environment that only slightly alters the thermodynamic conditions to disfavour the native state, the structure possessing native biological activity is generally lost. Only minor changes in the tertiary structure are required to introduce sufficient variation to render the protein no longer active (West *et al.*, 1997). As the chemical environment is altered further from conditions favouring the native state, the balance of competing forces will be tilted in favour of greater entropy and oscillations in the structure of the molecule will increase. Therefore, to refer to a single 'denatured-state' is incorrect; denatured protein is highly polymorphic (Jaenicke, 1987). However, microscopic heterogeneity can be averaged to give a macroscopically observable quantity that may be grouped together and referred to as a denatured-state (Creighton, 1987; Dill and Shortle, 1991). The observable quantity, and hence qualification of the state, will depend on the extent of the difference between the denaturing and native-favouring environments. Classification according to observable quantities leads to division of denatured states into two broad categories: compact and unfolded.

Some elements of secondary structure will be conserved within all denatured protein states (Dill and Shortle, 1991; Jaenicke, 1989; Shortle, 1996). These elements correspond to local sequences adopting the same stereoisomer as that of the native conformation. However, the overall structure of the protein will not be as compact and well packed as when it is in the native state. The extent to which the molecule is expanded determines the

overall observable measurements. For example, a change in the values of ultra-violet absorbance and hydrodynamic radius for a particular protein give an indication that the tertiary structure is being altered (Bastiras and Wallace, 1992; Brems *et al.*, 1987; Creighton and Pain, 1980). Mildly denaturing conditions do not disrupt the structure sufficiently to cause extensive fluctuation and the protein remains quite compact. At the extreme of harshly denaturing conditions, in which intra-molecular disulfide bonds have been reduced, significant amounts of the structure will be disrupted allowing the molecule to unfold and its structure to oscillate extensively.

In the case of *in vivo* refolding within the restricted environment of the cell, cofactors are expressed to assist temporary stabilisation of non-native protein conformations and, consequently, to promote formation of the native state. Non-native conformations may be specifically stabilised to assist in targeting to particular cellular locations and to prevent adventitious inter-molecular association by molecular chaperones such as GroEL and GroES or their equivalents (Bochkareva *et al.* 1988; Buchner, 1996; Gatenby *et al.*, 1993). Catalysts such as protein disulfide isomerase (PDI) and peptidyl-prolyl *cis-trans* isomerase (PPI) are expressed; they increase the rate of these rate-limiting isomerisations, advancing formation of the native state (Jaenicke, 1991; Schmid *et al.*, 1993).

Proteins are generally less stable when in unfolded states than in the native state. They are more susceptible to degradation by proteases when denatured either *in vivo* or *in vitro* (Creighton, 1987; Jaenicke and Rudolph, 1986; Schein, 1990). For example, in *E. coli* abnormal or misfolded proteins may be specifically targeted by the La protease which is under the control of the *lon* gene (Enfors, 1992; Gottesman, 1990). *In vitro*, partially folded proteins generally expose larger hydrophobic surfaces than in the native state, leading to an accelerated rate of aggregation via the mechanism of hydrophobic association (Jaenicke, 1995; Matthiesen *et al.*, 1996; Mulkerin and Wetzel, 1989). In a highly denaturing environment, it is likely that proteases will be denatured along with the protein of interest, thereby reducing the rate of degradation or even preventing it all together (Hart and Bailey, 1992; Kung, 1984). The presence of chaotropic agents such as guanidinium ions can also serve to reduce aggregation by shielding the exposed hydrophobic regions of the unfolded protein (Hagen *et al.*, 1990a). Therefore, it is likely that the greatest stability problems will be experienced during refolding as the protein is transferred to conditions promoting formation of the native state.

1.3.2 Molten globule state

A particular compact denatured state that combines some properties of both native and unfolded proteins has been characterised (Ptitsyn, 1995). The state is intermediate between the native and unfolded states and is referred to as the 'molten globule' (Ohgushi and Wada, 1983).

The most important feature of the molten globule state is that it is almost as compact as the native state. A core region of each protein molecule assumes a packing arrangement approximately as condensed as the native state, but some regions of the peptide chain are more expanded (Ptitsyn, 1992). In the molten state the secondary structure of protein is largely formed, but extensive intramolecular fluctuation is still possible (Kuwajima, 1989). Another property of the molten globule state is that a greater proportion of the hydrophobic surface is exposed than in the native state (Semisotnov *et al.*, 1987).

From a protein refolding perspective, Ptitsyn *et al.* (1990) proposed that the molten globule state is a general kinetic intermediate that occurs early in the transition from the denatured to native state. This proposal was supported by Jaenicke (1991) in his review of protein refolding. In the extensive review of the molten globule state and its role in protein folding compiled by Ptitsyn (1995), it appears a consensus has been reached that hydrophobic collapse and the molten globule state is more the rule than the exception in refolding pathways.

1.3.3 Nature of inclusion bodies

When proteins are overexpressed in *E. coli* using modern genetic engineering methods, they generally accumulate within the cell as dense insoluble particles called inclusion bodies (Marston, 1986; Rudolph and Lilie, 1996). The state(s) of protein within inclusion bodies poses an interesting case for consideration, and not only from the thermodynamic viewpoint.

Inclusion bodies are composed of many individual monomers of the overexpressed protein, tightly bound together with other cellular material (Bowden *et al.*, 1991; Hart *et al.*, 1990; Hartley and Kane, 1988; Marston, 1986; Schoemaker *et al.*, 1985). Evidence indicates that inclusion bodies form as a result of aggregation of partially folded monomers (Hart *et al.*,

1990; Mitraki and King, 1989; Mitraki *et al.*, 1991). That is, most of the inclusion bodies that have been characterised contain protein that is neither biologically active nor fully denatured. However, some examples of inclusion bodies with correct tertiary structure and biological activity apparently do exist (Hoess *et al.*, 1988; Tokatlidis *et al.*, 1991).

Inclusion bodies generally form in the reducing environment of the cytoplasm of *E. coli*, but they can be directed to form in the periplasm (Hammarberg *et al.*, 1990; Langley *et al.*, 1987; Valax and Georgiou, 1993a). A popular explanation for their occurrence is that overexpression of the protein of interest exceeds the capacity of the *in vivo* folding system so that nascent and partially folded protein molecules accumulate in large amounts (Cleland and Wang, 1993). Implicit in this explanation is that the capacity of protease systems, responsible for degradation of foreign/misfolded proteins (particularly the La protease), is also exceeded (Gottesman, 1989; Kane and Hartley, 1991). It should be noted that inclusion body expression actually stimulates the *lon* gene to produce more of the La protease, increasing the potential for degradation (Chung, 1993; Kresze, 1991).

Despite being trapped in an intermediate conformation, the protein within inclusion bodies is quite resistant to proteolytic degradation (Gottesman, 1990). For example, both bovine pancreatic trypsin inhibitor (BPTI) and creatine kinase can be expressed to give inclusion bodies that are stable even though the proteins are rapidly degraded by exogenous cellular proteases during renaturation (Babbitt *et al.*, 1991). However, the degree of resistance to proteolysis seems to be dependent on the expression system. For example, β -lactamase has been observed to be more resistant to proteolysis when expressed as inclusion bodies in the cytoplasm than in the periplasm (Bowden *et al.*, 1991; Valax and Georgiou 1993b).

The protein within inclusion bodies is also more stable to change in the suspending solvent than when isolated as monomers. When treated with ethylene diaminetetraacetic acid (EDTA), lysozyme, Triton X-100, and deoxycholate, the inclusion body protein remains stable, but exogenous host proteins are selectively solubilised (Bowden *et al.*, 1991; Estapé and Rinas, 1996; Hart *et al.*, 1990; Langley *et al.*, 1987; Marston *et al.*, 1984). Rhee *et al.* (1990) demonstrated that a β -galactosidase-IGF-II fusion peptide was stable in 4 M urea as inclusion bodies, but was denatured by a solution of 8 M urea. Similar stability of human Tissue Plasminogen Activator (TPA) in 5M urea/2% Triton X-100 was observed by Sarmientos *et al.* (1989). Marston and Hartley (1990) reported that harsh denaturing conditions were required to dissolve protein from inclusion bodies (8M urea or 6M

guanidine hydrochloride (GuHCl)). This evidence attests to the stability of protein expressed as inclusion bodies, suggesting that intermolecular association significantly influences the state of the protein.

There have been conflicting reports as to whether proteins that contain cysteine residues form intra- and/or inter-molecular disulfide bonds when expressed as cytoplasmic inclusion bodies. The cytoplasm of *E. coli* provides a reducing environment which suggests that disulfide bonds should not form and cysteine residues should be present with reduced thiol groups (Hammarberg *et al.*, 1990; Marston, 1986; Rudolph, 1990). Examples of this are inclusion bodies of recombinant bovine and porcine growth hormone (Bentle *et al.*, 1987; Fischer *et al.*, 1993a; Langley *et al.*, 1987). However, dissolution of some inclusion bodies has been shown to require a reducing agent, indicating that intermolecular disulfide bonds have formed (Fischer *et al.*, 1993a; Schoemaker *et al.*, 1985). Marston *et al.* (1984) concluded that atmospheric oxygen may have oxidised the surface of isolated inclusion bodies of prochymosin, although the protein within the core was still reduced. An alternative mechanism was proposed which postulated that high-level expression gave an oxidising microenvironment within the cytoplasm (Marston, 1986; Schoemaker *et al.*, 1985). However, oxidation of inclusion body proteins was shown to occur rapidly once the cytoplasm had been disrupted (Schoemaker *et al.*, 1985). Langley *et al.* (1987) added further weight to the original conclusion of Marston *et al.* (1984) when they reported that despite the expectation that all of the protein was expressed in reduced form, 20-30% of bovine growth hormone present as inclusion bodies was oxidised after cell lysis. In a review of inclusion body processing, Rudolph (1990) concluded that the oxidation of disulfide bonds in cytoplasmic inclusion bodies is most likely to occur after cell lysis. The example of porcine growth hormone was verified independently by demonstrating that minimising exposure to atmospheric oxygen made the inclusion bodies easier to denature (Bastiras, 1992).

The physical characteristics of inclusion bodies are also quite varied, indicating that the protein they contain may be present in different states. Common features are that the protein product is amorphous, is not surrounded by a cellular membrane, and cannot be totally separated from cellular components of the host organism (Kane and Hartley, 1991; Schoemaker *et al.*, 1985). Typically they are spherical and range in diameter up to approximately 1.3 μm , depending on the protein and expression vector (Taylor *et al.*,

1986). However, even cytoplasmic and periplasmic inclusion bodies of the same protein have been shown to exhibit different morphologies and different endogenous cellular contaminants (Bowden *et al.*, 1991; Hart *et al.*, 1994; Valax and Georgiou, 1993a). Inclusion bodies are porous (Bowden *et al.*, 1991; Dorin *et al.*, 1995; Fish and Hoare, 1988) with the overall voidage dependent on the suspending solvent (Taylor *et al.*, 1986). This indicates that proteins are still able to interact with the chemical environment when trapped in inclusion body form.

Growth conditions have been shown to alter the physical characteristics of inclusion bodies. An extreme example is the demonstration that inclusion body formation can be prevented altogether by lowering the culture temperature (Mitraki *et al.*, 1991). Kane and Hartley (1991) concluded that variation of *in vivo* aggregation during expression was protein-specific, depending on the composition and temperature of the medium. The specific protein content of inclusion bodies of *Vitreoscilla* haemoglobin have been observed to vary when grown under different conditions (Hart *et al.*, 1990). Valax and Georgiou (1993a) observed that the 'stickiness' of inclusion bodies of β -lactamase varied with growth conditions, which significantly influenced post-lysis association of host cell contaminants. It would seem logical to conclude that the altered physical characteristics are a reflection of a change in the state of the aggregated protein.

This brief overview of inclusion bodies serves to emphasise that not only is there uncertainty about the conformational state of the protein within inclusion bodies, but the expression system, culture conditions, and physical handling provide additional dimensions of variability. The expression system and culture conditions can have a dramatic effect on the morphology and endogenous contamination of inclusion bodies. Physical handling can alter the oxidation state and the degree of association with exogenous cellular material. The presence of varying amounts of cellular material can significantly affect the behaviour of the protein product. Essentially, when considering protein expressed as inclusion bodies, the biological diversity of the host organism cannot be decoupled from a description of the state of the protein. Fractions of the host organism are so intimately associated with protein sequestered into inclusion bodies that a practical investigation cannot ignore the unique properties that their combination exhibits.

1.4 DISSOLUTION OF PROTEIN INCLUSION BODIES

The major drawback of expressing proteins as insoluble inclusion bodies is often cited as the need to renature the biologically inactive protein (Hammarberg *et al.*, 1990; Kane and Hartley, 1991). Since the structure of protein in inclusion bodies is dependent on so many factors, it is common for a conservative approach to be adopted when designing renaturation processes. Therefore, the most common first step in recovering protein from inclusion bodies is complete dissolution with strong chemical denaturants (Chaudhuri, 1994a; Marston and Hartley, 1990; Rudolph and Lilie, 1996; Sarmientos *et al.*, 1989).

The most widely used strong chemical denaturants are urea and GuHCl (Creighton, 1987; Dill and Shortle, 1991; Hibbard and Tulinsky, 1978). Urea does not have the additional ionic properties of GuHCl, making it a slightly weaker denaturing agent (Builder and Ogez, 1986; West *et al.*, 1997). However, urea is more frequently used because it is less expensive, ecologically safer, and does not interfere with downstream operations (Bentle *et al.*, 1987; Chaudhuri, 1994a). Urea has the disadvantage that it degrades in aqueous solution to form isocyanate ions which can covalently modify amino or thiol groups (Jaenicke and Rudolph, 1986). Therefore, when using urea as a denaturant, it is necessary to use an isocyanate scavenger to prevent microheterogeneities being introduced into the protein product (Rudolph and Lilie, 1996).

Urea is classified as a chaotrope on the Hofmeister series as it weakens interactions between water molecules (Khurgin and Maksareva, 1993; Schiffer and Dötsch, 1996). Disruption of the normal water structure allows preferential interaction of water molecules with other solutes such as protein molecules, potentially resulting in protein unfolding. Urea also binds directly to the protein and diminishes the magnitude of non-covalent stabilising interactions (Creighton *et al.*, 1996). The dissolving action of urea has been reported to involve penetration of urea molecules into the core of each protein molecule (Hibbard and Tulinsky, 1978). Urea is known to preferentially bind to proteins in the denatured state (Creighton, 1987). Further evidence of its selectivity is that urea is unable to dissolve outer membrane proteins in the absence of other destabilising chemical agents (Moldow *et al.*, 1972).

When dissolved in a concentrated urea solution, intermolecular, tertiary, and secondary structures of protein are largely disrupted. For cysteine-containing proteins, dissolution is normally conducted under reducing conditions to reduce any disulfide bonds to cysteine residues and allow complete unfolding (Chaudhuri, 1994b). Complete unfolding is necessary if incorrect disulfide bonds have been formed (Bentle *et al.*, 1987) or in kinetic studies of refolding where it is important to start with fully unfold molecules to avoid artefacts of residual structures (Jaenicke and Rudolph, 1986).

To date, little work has been done to determine the actual mechanisms involved in inclusion body dissolution. Hart and Bailey (1992) noted that there is no theoretical framework for analysing inclusion body dissolution. Most studies empirically determine conditions that give total dissolution of inclusion body protein before subsequent refolding (Fischer *et al.*, 1993a). Using this approach, surprisingly similar conditions have been selected for a range of proteins. A typical dissolution protocol involves mixing a batch of isolated inclusion bodies in a solution of buffered 8M urea at alkaline pH in the presence of a reducing agent for one to two hours at room temperature (Fisher *et al.*, 1993a; Marston and Hartley, 1990; Rudolph, 1990).

The lack of knowledge about dissolution of inclusion bodies is in stark contrast to the extensive studies on the unfolding of pure proteins from the native state. The denaturing influence of pH and temperature have been investigated as well as the effect of various chemical denaturants. Ptitsyn (1995) concluded that in strong chemical denaturants proteins undergo 'all or none' denaturation. That is, they can be adequately described with a two-state model where protein is altered directly from the native state into an unfolded denatured state (Cardamone *et al.*, 1995; Creighton, 1987; Kuwajima *et al.*, 1976). However, equilibrium studies provide ample evidence of multi-state unfolding pathways (Havel *et al.*, 1986; Miller *et al.*, 1993). Kuwajima (1989) concluded that two-state denaturation was an oversimplification of three-state unfolding where the intermediate was the molten globule state. De Young *et al.* (1996) concluded that nerve growth factor undergoes extensive rapid unfolding followed by continuing extremely slow unfolding over many days due to a cysteine knot in the structure of the protein in the native state. Therefore, it seems that the general multi-state denaturation pathway is only adequately described by the two-state model under harsh conditions when long-term structural changes are not considered.

Since urea must diffuse into individual protein molecules within the structure of an inclusion body to dissolve them, dissolution of inclusion bodies is likely to be limited by mass-transfer. This has implications for the consistency of the state of denatured protein extracted from inclusion bodies. Since the nature of inclusion bodies themselves is quite varied (section 1.3.3), it is likely that batch dissolution will exhibit time-dependent characteristics. Care must be taken to use uniform dissolution procedures when comparing the performance of subsequent operations.

1.5 REFOLDING

Refolding is the operation of changing the thermodynamic environment of denatured protein to conditions that favour adoption of its native state. The process exploits the spontaneous changes in state that can be achieved *in vitro* by controlling the thermodynamic environment of a protein. By altering the solvent composition, the denatured protein can assume its native state. For example, refolding may be achieved by dialysis of a particular protein solution in 8 M urea, which provides denaturing conditions, to lower the concentration to 2 M urea which favours adoption of the native state.

1.5.1 Refolding mechanism

The majority of studies of refolding have used solutions of highly purified proteins at very low concentration. The aims of these studies have been to elucidate the mechanisms and thermodynamics of refolding. This section presents a review of the significant refolding mechanisms which have been identified during transition of proteins to their native state.

The driving force for protein refolding is the minimisation of conformational energy by reducing unfavourable contact between exposed residues of the unfolded protein and the solvent (Dill, 1990; Jaenicke, 1987; Jaenicke, 1989). Refolding is conducted almost exclusively in aqueous solution, and the driving force is hydrophobic interaction (Dill, 1990; Jaenicke, 1995; Shortle, 1996). The native state corresponds to the minimum conformational energy (section 1.3.1) so any denatured protein should assume the native state if the thermodynamic environment is favourable and it does not become trapped in an intermediate state that provides a local minimum energy.

While it is beyond doubt that the transition between denatured and native states is under thermodynamic control (Kim and Baldwin, 1990), the vector nature of the change in structure will result in observable structural changes with time amongst the population of conformations present (Dill, 1987; Jaenicke, 1989). Accordingly, many structural models of protein refolding have been developed to describe the changes in measured structural parameters (Kim and Baldwin, 1982).

The most commonly adopted structural refolding scheme is the framework model (Brems *et al.*, 1987; Kim and Baldwin, 1982; Cardamone *et al.*, 1995). The framework model describes the mechanism of refolding as initial formation of hydrogen-bonded secondary structure after unfolded protein has been transferred into native-favouring conditions. Subsequent tertiary interactions lead to development of the native conformation. Kuwajima (1989) observed that denatured protein diluted into conditions favouring native folding rapidly assumed the compact molten globule state. He concluded that formation of a molten globule intermediate is consistent with a framework model of refolding. This view of the protein refolding mechanism is supported by Ptitsyn (1995). The molten globule has been proposed as a common rapidly-appearing intermediate in refolding pathways (Jaenicke, 1991; Ptitsyn *et al.*, 1990; Ptitsyn, 1995).

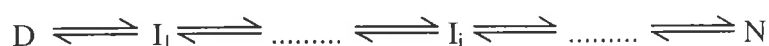
While formation of a molten globule intermediate occurs rapidly, other subsequent intermediates are often observed to form at slower rates before the native state is reached (Grafl *et al.*, 1986; Milner, 1996; Semisotnov *et al.*, 1987). Two particular rate-limiting mechanisms have been identified; namely proline isomerisation and disulfide bond formation (Jaenicke, 1991; Jaenicke, 1995; Schmid *et al.*, 1993).

Peptide bonds along the protein backbone are planar and can assume either *cis* or *trans* conformations. In an unfolded, denatured state 10 to 30% of peptide bonds assume the *cis* conformation (Brandts *et al.*, 1975) and rapidly isomerise to the *trans* conformation upon refolding. The *cis-trans* isomerisation around the prolyl peptide bond is intrinsically slow, requiring high activation energy, because it involves rotation about a partial double bond (Schmid, 1992). In the native state, peptide bonds are virtually exclusively in the *trans* conformation ($\geq 99.9\%$) to reduce steric interactions between adjacent residues (Brandts *et al.*, 1975; Creighton, 1987; Stewart *et al.*, 1990). However, approximately 7% of native-state prolyl peptide bonds are in the *cis* conformation (Schmid, 1992; Stewart *et al.*, 1990). As a result, in proteins that contain proline, the rate of refolding is often observed to be dependent on the rate of proline isomerisation (Borden *et al.*, 1990; Garel, 1992; Kiefhaber *et al.*, 1990a, 1990b).

The extensive publications of Thomas E. Creighton on the role of intramolecular disulfide bond formation has established beyond doubt that these bonds are necessary for structural development during refolding. Early studies into refolding mechanisms used disulfide

bonds as probes to determine reaction pathways by acid-trapping kinetic intermediates (Creighton, 1978), a technique that is still used currently (Milner *et al.*, 1995, Creighton *et al.*, 1996). It has been established that thiol groups need to be approximately 3.8 Å apart before covalent bonding can occur (Creighton and Goldenberg, 1984; Creighton, 1986). Miller *et al.* (1993) concluded that disulfide bond formation was associated with stabilisation of protein structure. The rate of disulfide bond formation has been observed to increase in compact intermediate states because of the proximity of cysteine groups (Miller *et al.*, 1993; Storrs and Przybycien, 1991) and subsequent structural development is more rapid once they have formed (Dill and Shortle, 1991; Lilie *et al.*, 1996). The current view of the role of disulfide bonds in refolding is that they stabilise, but do not specify the pathway to the native state (Creighton, 1995). Disulfide bonds stabilise lower entropy as the state approaches an ordered native state. The bonds provide a simple method for identifying refolding intermediates as they indicate that the structural development needed for the cysteine residues to be close enough to form a covalent bond has occurred (Creighton *et al.*, 1996).

The consensus view of *in vitro* refolding is that it occurs via a pathway of sequential intermediate states (Jaenicke, 1989; Creighton *et al.*, 1996). The protein in a denatured state (D), does not spontaneously assume its native conformation (N) when the denaturing conditions are removed. The structure assumes a number of transition states (I_i) between the denatured and native conformations.



These states arise as kinetic traps in the transition from the initial denatured state due to sequential rate-limiting steps in the refolding mechanism such as proline isomerisation and/or disulfide bond formation. As technology has improved the resolution of the instruments used to measure protein structure, increasing complexity has been ascribed to the range of intermediates which have been characterised (Kim and Baldwin, 1990).

For proteins containing intramolecular disulfide bonds, an oxidising environment is required to facilitate covalent bond formation. This usually occurs in an alkaline solution where the thiolate anion concentration is greatest and the oxidising agent that is added to reform the disulfide bonds of a protein often contains disulfide bonds (Creighton, 1987). Therefore, if the oxidising conditions are removed before refolding is complete, the protein

will be 'trapped' in an intermediate conformation. One common method of removing the oxidising conditions is to lower the pH of the refolding solution to reduce the concentration of the reactive thiolate anions (Creighton, 1978; Weissman and Kim, 1991).

1.5.2 Off-pathway reactions

The review of protein refolding mechanisms presented above outlines a basic pathway for denatured protein to reach the native state via numerous intermediate states. When exposed to an environment favouring correct folding, denatured protein is no longer stable and begins to refold. The native state is reached by reducing unfavourable interactions with both the solvent and within the protein molecule itself. This mechanism assumes that interaction with other solutes is not favourable and all intermediate states are of higher conformational energy than the native state.

However, denatured protein can fold incorrectly and assume a state that has conformational energy similar to that of the native state. Misfolding of this nature during *in vitro* refolding has been observed for IGF-I and some analogues of this molecule (Hober *et al.*, 1992; Miller *et al.*, 1993; Milner *et al.*, 1995). Refolding under oxidising conditions leads to two significantly populated states with different cysteine pairings via their three intramolecular disulfide bonds. The two states are in equilibrium and represent alternative thermodynamic minima (Miller *et al.*, 1993). This is the first example of a protein that apparently does not conform with the concept that formation of a unique and stable tertiary structure is dictated by the primary sequence as proposed by Anfinsen (1973). However, a possible explanation is that IGF-I is naturally expressed as pro-IGF-I and later processed *in vivo* to give the mature sequence by cleavage of the leader sequence. The pro- sequence may lead to a unique native IGF-I structure before the leader sequence is removed (Hober *et al.*, 1992). While the refolding of IGF-I presents an isolated example, many other useful proteins are initially naturally expressed as precursors whereas the mature sequences are expressed directly when recombinant methods are used to produce them. It seems likely that misfolding will be observed when other recombinant proteins are produced.

The primary complication of the basic refolding pathway is caused by interaction between molecules in intermediate states. Typically, irreversible aggregation occurs, resulting in reduced yield. Loss of valuable protein product to off-pathway aggregation is cited as the

major problem with *in vitro* protein refolding (Brems, 1988; Chaudhuri, 1994a; Chaudhuri *et al.*, 1996; Cleland and Wang, 1993; Fischer *et al.*, 1992; Georgiou, 1988; Hevehan and De Bernardez-Clark, 1997; Jaenicke, 1987; Kiefhaber *et al.*, 1991; London, *et al.*, 1974; Rudolph and Lilie, 1996; Schein, 1990; Storrs and Przybycien, 1991; Winther *et al.*, 1994; Zettlmeißl *et al.*, 1979).

Studies have revealed that aggregation is not initiated between monomers in unfolded or native states. Aggregated protein complexes do have some, but not complete, native-like structure (Brems, 1988; Mitraki and King, 1989; Sargent *et al.*, 1988). There is overwhelming evidence that the driving force for this association is hydrophobic interactions (Brems, *et al.*, 1986; Buchner, 1996; Cleland and Wang, 1990b; Cleland and Wang, 1993; Horowitz, 1993; Matthiesen *et al.*, 1996; Mulkerin and Wetzel, 1989). For instance, temperature reduction is known to reduce hydrophobic interactions (Allewell, 1993) and decreased aggregation has been observed at lower temperatures (Gatenby *et al.*, 1993; Mitraki *et al.*, 1987). These observations are consistent with the view that aggregation occurs between monomers that are in a molten globule state due to their exposed hydrophobic surfaces (Kuwajima, 1989; Ptitsyn, 1995). Cleland and Wang (1990a & 1991) have proposed that dimerisation of the molten globule intermediates occurs in the refolding of bovine carbonic anhydrase (CAB) and that this is likely to be the rate-limiting step that leads to aggregation in the case of other globular proteins as well. Kwon *et al.* (1995) observed that aggregation of α_1 -antitrypsin to form oligomers was initiated from dimers of protein in an intermediate state. Speed *et al.* (1996) observed that P22 tailspike and P22 coat proteins form homogenous, not heterogeneous aggregates. They concluded that specific aggregation may be initiated from molten globule intermediates, but larger multimers may occur directly from small multimers. Since the coat protein refolded approximately five times faster than the tailspike protein, hydrophobic association cannot be eliminated as the driving force. Therefore, it is logical to conclude that aggregation is initiated by dimerisation of a molten globule intermediate due to association of hydrophobic surfaces and that the dimers are able to associate further to give larger multimers.

It has been proposed that hydrophobic association is also the primary force in the aggregation which leads to the formation of inclusion bodies (Mitraki and King, 1989). That is, inclusion body formation and aggregation provide the *in vivo* and *in vitro* evidence

of hydrophobic association (King *et al.*, 1996). Reducing the temperature during expression of protein product has been observed to result in soluble protein, instead of inclusion bodies. This is assumed to be due to *in vitro* reduction in hydrophobic interactions at the lower temperature (Mitraki *et al.*, 1991). Also, *in vivo* mechanisms for reducing off-pathway hydrophobic interactions have been identified that involve molecular chaperones such as GroEL binding to hydrophobic intermediates and preventing aggregation (Buchner, 1996; Buchner *et al.*, 1991; Gatenby *et al.*, 1993; Mitraki *et al.*, 1991).

Therefore, minimisation of conformational energy may lead to formation of correctly folded monomers, misfolded monomers, and aggregates. The possible outcomes from *in vitro* refolding of a homogenous solution of reduced, denatured protein are depicted in Figure 1.4.

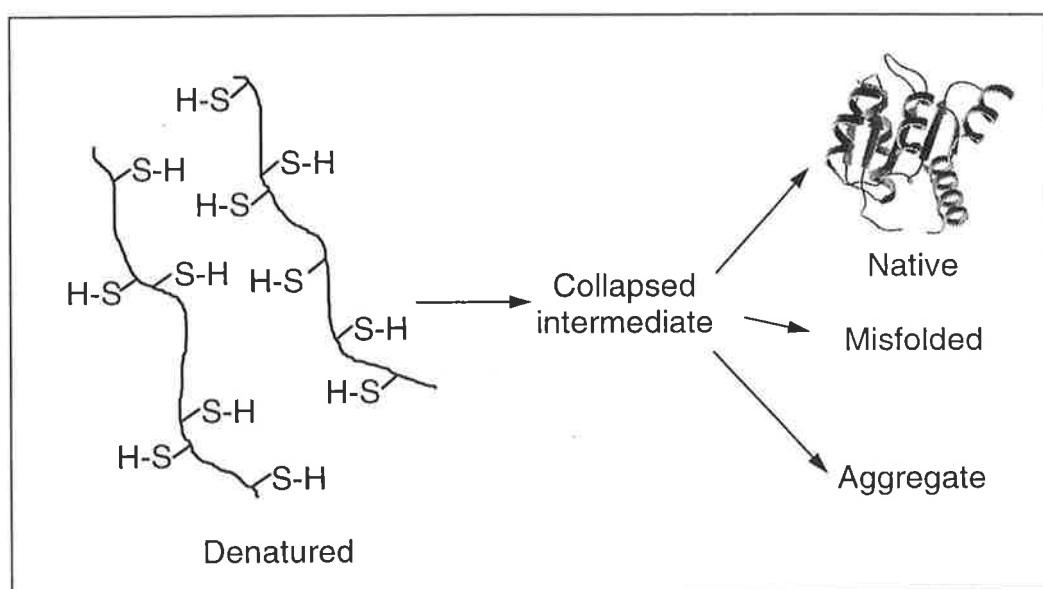


Figure 1.4: Schematic depicting competing reactions during protein refolding

1.5.3 Traditional laboratory-scale methods for refolding

A number of methods have been developed to transfer protein from denaturing conditions into an environment that favours the native state. The most commonly used method is direct batch dilution into a solution of appropriate chemical composition (Fischer *et al.*, 1993a; Galliher, 1991; Rudolph, 1990). For an ideally mixed system, this gives a step change in the thermodynamic environment. To minimise off-pathway aggregation, dilution

to achieve an overall protein concentration in the micromolar range is often recommended (Cleland and Wang, 1993; Fischer *et al.*, 1992; Hagen *et al.*, 1990a; Rudolph, 1990). This is the main disadvantage of the direct batch dilution method: even for only moderate levels of production, very low protein concentration requires large processing volumes and significant costs for equipment and consumables (Chaudhuri, 1994a; Marston and Hartley, 1990). Also, if dilution is too great, non-specific adsorption of the protein to the surfaces of processing equipment can result in significant losses (Horowitz, 1993; Lee and Rhee, 1996).

Dialysis is also a commonly used method for gradually changing solvent composition to conditions that promote native folding (Fischer *et al.*, 1993a; Marston, 1986; Rudolph and Lilie, 1996). There are examples of refolding of proteins where dialysis has given yields superior to those obtained using dilution (Hanemaaijer *et al.*, 1989; Teipel and Koshland, 1971a). Dialysis has also been employed to exchange denaturing agents while maintaining denaturing conditions before initiating refolding (Olson, 1985). However, the slow rate of solvent exchange under these conditions makes it unattractive for large-scale commercial production (Chaudhuri, 1994a).

Diafiltration and size-exclusion chromatography are alternative methods that have been used for solvent exchange to give suitable refolding conditions (Burton *et al.*, 1989; Hejnæs *et al.*, 1992; Hershenson *et al.*, 1990; Vicik and De Bernardez-Clark, 1991). Diafiltration offers superior flux rates and scalability than dialysis and more gradual thermodynamic change than dilution. However, this process is not well characterised (Chaudhuri, 1994a). Refolding using size-exclusion chromatography is also not well characterised. Recent advances that couple concurrent chromatographic separation with refolding are presented in section 1.5.5.

To facilitate formation of internal disulfide bonds, refolding must be conducted under oxidising conditions. The least expensive method of providing an oxidising agent is to rely on atmospheric oxygen. Air can either be sparged through the refolding mixture or merely allowed to contact the surface. Oxidation of the protein disulfide bonds is catalysed by trace concentrations of metal ions such as Cu^{2+} (Hershenson *et al.*, 1990; Fischer *et al.*, 1993a; Rudolph and Lilie, 1996). Nevertheless, it is commonly found that air oxidation requires more than 24 hours (Bentle *et al.*, 1987; Brierley *et al.*, 1997; Stempfer *et al.*, 1996). Air oxidation has been purported to be the superior method as it poses less chance

of residue derivitisation (Weir and Sparks, 1987). The extended duration of the time needed for air oxidation is the main disadvantage of the method. Additional disadvantages are that reproducibility is often poor and method can still result in a low yield of active protein (Chaudhuri, 1994a; Rudolph, 1990; Rudolph and Lilie, 1996). Therefore, it is more common to control disulfide bond formation with chemical additives.

Formation of mixed disulfides prior to refolding is a commonly used approach (Buchner and Rudolph, 1991; Goldberg *et al.*, 1991; Wetzel, 1986). Cysteine residues that are exposed in unfolded, denatured states can react with oxidised glutathione or sodium-sulfite/tetrasulfite to give a solution of mixed disulfides. Relatively brief incubation of the mixture allows cysteine residues to react with the chemical additives, preventing formation of intramolecular disulfide bonds until the additive is displaced (Fisher *et al.*, 1993a). Native intramolecular disulfide bonds form by disulfide shuffling because they are more energetically favourable. Formation of mixed disulfides improves protein solubility by introducing hydrophilic groups to the molecule and controls disulfide bond formation until reduced thiols are added to the solution (Rudolph and Lilie, 1996). Seely and Young (1991) concluded that initial formation of mixed disulfides and reducing conditions for refolding of secretory leucocyte protease inhibitor can have a profoundly favourable impact on the overall yield of renatured protein.

The most commonly used method for controlling disulfide bond formation involves the glutathione system (Chaudhuri, 1994a; Fisher *et al.*, 1993a). By varying the relative concentrations of reduced and oxidised glutathione, and protein-cysteine residues under native-favouring conditions, the yield of correctly folded protein can be maximised. The mixture of oxidised- and reduced-glutathione acts as a redox couple to promote intramolecular disulfide shuffling. Evidence of the effectiveness of the glutathione system is its successful use in renaturing recombinant human TPA, which has 17 internal disulfide bonds in its active state (Rudolph *et al.*, 1995). Unfortunately glutathione adds significant cost to the refolding process (Fischer *et al.*, 1993a), making it a less attractive strategy as production requirements increase. Therefore, other redox couples are used such as oxidised β -mercaptoethanol (OX- β -ME) with dithiothreitol (DTT) or β -mercaptoethanol to provide oxidising and reducing potentials, respectively (Milner *et al.*, 1995; Mitraki *et al.*, 1987; Speed *et al.*, 1997).

1.5.4 Standard methods for improving refolding

Traditionally, the motivation to develop refolding processes has been to identify methods that manage to produce biologically active product. Improvement of methods to increase the yields of correctly refolded protein has been a side-issue; it has been more a question of curiosity about refolding pathways and mechanisms. Processing time, complexity and corresponding cost have not been regarded as important factors. In those investigations where high yield was an objective, empirical optimisation of the chemical environment has been used to identify conditions that maximise yield of native protein (Chaudhuri, 1994a). Marston and Hartley (1990) have presented a extensive list of 'variables at the refolding stage which are critical' as they impact on the yield of correctly refolded protein. The variables that they identified were: pH, incubation temperature, time, ionic components of the solvent, concentration and rate of removal of denaturant, protein concentration, purity of refolding mixture, and presence of redox agents. Other reviews on the topic of protein refolding have generally identified subsets of these variables, possibly with the addition of cosolvents as an important factor. For example, addition of L-arginine (Buchner and Rudolph, 1991; Hevehan and De Bernardez-Clark, 1997), ethanol (Brierley *et al.*, 1997; Chang *et al.*, 1995; Hejnæs *et al.*, 1992; Skriver *et al.*, 1995; Yamada *et al.*, 1996), or organic polymers such as polylysine, polyalanine, and polyethylene glycol (PEG) (Cleland and Wang, 1990c) have been demonstrated to improve refolding yield. A good example is the process patent held by Michaelis *et al.* (1995) for refolding eukaryotic alkaline phosphatase which includes sulfate salts, carbohydrates or polyalcohols having at least two carbon atoms as cosolvents to improve the yield of biologically active protein.

Not surprisingly, the referred values for all of these refolding variables must be determined on a case-by-case basis for different proteins (Rudolph and Lilie, 1996; Speed *et al.*, 1997). However, there are some common features of refolding protocols that result in higher yield of active protein. Most procedures use a low concentration of purified protein, a chemical disulfide agent, and a non-denaturing concentration of chaotrope in an alkaline solution at 10-30°C. The impact of protein purity on refolding yield is discussed further in section 1.5.6.3. Justification for low protein concentration and chemical disulfide agent are outline above, and chaotrope concentration and alkaline pH remain to be considered.

Refolding in the presence of non-denaturing concentrations of chaotrope has been shown to improve yield by reducing aggregation (Hevehan and De Bernardez-Clark, 1997; Rudolph, 1990; Rudolph and Lilie, 1996). It is important to maintain the concentration of chaotrope at non-denaturing levels as otherwise denatured intermediates that aggregate rapidly may be stabilised (particular the molten globule state) (Kwon *et al.*, 1995; Mitraki *et al.*, 1987; Storrs and Przybycien, 1991). For example, Bente *et al.* (1987) concluded that the concentration of urea in the refolding mixture was the most important factor in determining the yield of active bovine growth hormone.

High pH favours correct refolding as most useful proteins contain intramolecular disulfide bonds (Chaudhuri, 1994a) and disulfide shuffling is accelerated under alkaline conditions (Kim and Lee, 1997; Marston, 1986; Rudolph *et al.*, 1991). This is because thiol groups are ionized at alkaline pH (Jocelyn, 1972). For example, the first thiol group of reduced DTT generally ionises at a pH of approximately 8.3 and the second thiol ionises at approximately 9.5 (Creighton, 1978). Ionized thiols are the more reactive species and the rate of interchange of protein disulfide bonds is accelerated with increased ionisation, promoting correct folding.

As described in section 1.3.1, *in vivo* folding generally involves specific interactions with other cellular species of proteins. The role of PPI, PDI and molecular chaperones in protein folding is currently a major area of research. *In vitro* refolding methods do not have the benefit of these cofactors (Buchner and Rudolph, 1991). Therefore, one logical progression for improving refolding methods and minimising *in vitro* aggregation is to mimic the *in vivo* pathways. The methods discussed above consider only chemical optimisation to increase recovery of native protein. This *in vitro* approach does provide a limited *in vivo* equivalent of chaperones, by stabilising unfolded monomers to prevent aggregation. However, it lacks the biological specificity of chaperones which effect total isolation of monomers in intermediate states from one another. *In vitro* optimisation of the redox environment to increase the rate of disulfide bond formation does increase the rate of disulfide bond formation, but it does not utilise the catalytic benefits of PDI that are available *in vivo*. The more recent methods for improved refolding are presented in the next section. Some of them may be viewed as attempts to develop *in vitro* equivalents of *in vivo* mechanisms.

1.5.5 Recent developments in protein refolding.

Production of useful therapeutic proteins by *E. coli*, has placed increased emphasis on the need to improve *in vitro* protein refolding efficiency. Rather than the earlier approach on elucidating refolding pathways, the primary focus of recent studies has been on achieving maximal yield of renatured product as efficiently as possible. The effective approach of refolding highly purified protein solutions at very low concentration is not economically feasible. The crux of increasing refolding efficiency is to do so at high overall protein concentration so that the costs of chemical additives and equipment are reduced (Goldberg *et al.*, 1991). As outlined in section 1.5.2, the major problem encountered when refolding is off-pathway reaction of intermediates to form non-functional aggregates. *In vivo*, cofactors work in concert to prevent protein aggregation and to catalyse transformation to the native state. Therefore, improved *in vitro* refolding is likely to be achieved by the combined action of all of the methods outlined in section 1.5.4, with additional *in vitro* equivalents of chaperones and isomerases.

Recent advances in methods for protein refolding have included the use of reverse micelles, rational protein engineering, concurrent chromatographic separation, affinity-based ligands, and different modes of dilution to reduce aggregation of intermediate species. There has also been progress in optimisation of the chemical environment. The extreme example is the process for IGF-I production that Genentech, Inc. have developed (Chang *et al.*, 1994; Chang *et al.*, 1995). Conditions which result in dissolution and refolding in a single step have been identified. Using approximately 2M urea as the chaotrope, values of temperature, pH, and concentrations of DTT, methanol, NaCl, and protein were defined which gave acceptable recoveries of active protein at relatively high protein concentration ($\approx 5 \text{ g L}^{-1}$). The starting material was an isolate of periplasmic inclusion bodies highly contaminated with various *E. coli* components. The system has been extended to use concurrent two-phase liquid extraction to selectively separate the correctly refolded product in a single step from whole cells containing periplasmic inclusion bodies (Hart *et al.*, 1994; Swartz *et al.*, 1996).

Reverse micelles offer a method of achieving 100% renaturation, by isolating denatured monomers of protein and subsequently refolding them while are physically separated. Hagen *et al.* (1990a) obtained quantitative renaturation of bovine pancreatic ribonuclease A

at a protein concentration of 1 g L^{-1} , but noted that protein transfer into the micelles was slow and denaturant removal was difficult. However, active γ -interferon could not be obtained using reverse micelles (Hagen *et al.*, 1990b) because instead of refolding quantitatively, the denatured protein associated with the surfactant. Chaudhuri and Mall (1994) observed that the chaotropes used to dissolve inclusion bodies, particularly urea, can disrupt micelles. However, they have demonstrated the potential of the system using hen egg-white lysozyme (HEWL) (Mall *et al.*, 1994). The need for careful optimisation of the chemical conditions required for a *particular* protein prevents the application of the reverse micelle method for *in vitro* refolding of *all* proteins (Rudolph and Lilie, 1996).

Fusion peptides of IGF-I have been shown to increase the solubility of aggregated multimers of the native peptide during refolding and they could also be captured for reprocessing (denatured and refolded) to increase the overall recovery (Samuelsson *et al.*, 1991). More recent studies with fusion peptides of IGF-I have demonstrated that this approach can increase yield of active product, most probably because of greater stability of intermediates (Samuelsson and Ulhén, 1996). The fusion proteins also provide a useful affinity tag product recovery by affinity chromatography. The use of hydrophobic fusions at either or both ends of proteins to provide helper sequences that increase renaturation yield has been patented (Ambrosius *et al.*, 1996). The patent covers the use of extension peptides that are from two to twenty amino acids long which contribute to the overall hydrophobicity of any peptide, claiming that this improves the yield of native protein.

Concurrent chromatographic separation has been used in two distinct ways to improve the yield of correctly folded protein. The value of size-exclusion chromatography to improve refolding been demonstrated (Batas and Chaudhuri, 1996). In this method, the chromatographic medium acts as an inert solid that reduces the effective concentration of monomers in unstable intermediate states, thereby reducing aggregation. Speed *et al.* (1997) have observed a similar reduction in aggregation during the refolding process conducted using polyacrylamide gel as the inert solid. In the second method, the protein to be renatured is bound to a matrix to isolate monomers from each other during buffer exchange. Stempfer *et al.* (1996) have immobilised α -glucosidase on an ion-exchange medium and achieved 400% greater yield of native-conformation protein at 1000 times greater concentration than direct batch dilution. Novo Nordisk A/S has patented a method for refolding a range of proteins using an immobilised antibody to bind them during

denaturation and refolding on the matrix (Markussen, 1994). Yoshida *et al.* (1996) hold a more general patent for promoting correct refolding using either an immobilised subunit of chaperonin-60 or fragments of this molecule as the ligand. Both of these techniques offer the additional advantage of purifying the active product at the same time as refolding. However, care must be taken when selecting the properties of the support matrix, and connection of the ligand to avoid irreversible loss of product (Jaenicke, 1995). The use of chromatography may also be limited by fouling of the matrix with contaminants from the host organism.

The patent held by Yoshida *et al.* (1996) also covers use of chaperonin-60 and its fragments in solution to increase refolding selectivity. A similar affinity-based strategy that is less generic has been developed to achieve quantitative renaturation of IGF-I in the presence of the IGF-I binding protein (Samuelsson *et al.*, 1996). A significant disadvantage of these affinity techniques is that the affinity ligand must be isolated in the first instance.

The effectiveness of a simpler approach has also been demonstrated. Conducting refolding by a series of dilutions has been shown to reduce aggregation when compared to a single dilution process. The improvement has also been observed in fed-batch and continuous modes using the chemical environment selected from batch refolding trials. Using these approaches, significant improvement of renaturation performance has been achieved without the need to increase use of expensive additives (Ambrosius and Rudolph, 1997; Dorin *et al.*, 1995; Fischer *et al.*, 1992; Galliher, 1991; Rudolph and Fisher, 1990; Samuelsson and Ulhén, 1996). Rudolph (1990) was the first to review the step-wise addition of denatured protein to refolding mixtures as a useful method to reduce aggregation. The first patent on 'pulse' addition and refolding in fed-batch mode accompanied the publication (Rudolph and Fischer, 1990). Soon afterwards, Galliher (1991) patented the first continuous refolding system. In spite of its simplicity and the observed benefits of this approach, batch refolding remains the more common mode of operation.

1.5.6 Issues for large-scale refolding by dilution

As outlined in section 1.5.4 the traditional approach adopted to optimise refolding has been to consider only the thermodynamic environment and empirically seek conditions that maximise yield of native protein. Large-scale refolding processes have then been based on direct scale-up of the laboratory-scale process (Chaudhuri, 1994b). Seely and Young (1991) have presented one of the few reports of cost implications on scaling up a refolding operation. Studies that use model-based optimisation to determine conditions that give maximal yield, such as that by Vicik and De Bernardez-Clark (1991), are rare. The principles used in traditional chemical processing industries to assist in process design and improve production, have great potential to enhance bioprocessing through their more widespread application (Kossik, 1996).

1.5.6.1 Process simulation

One of the most powerful tools available in process design and development of superior production methods is process simulation. More than a decade ago the bioprocessing group at Massachusetts Institution of Technology addressed the application of process simulation to bioprocessing. It lead to the publication of a thesis on 'computer-aided design of integrated biochemical processes' (Petrides, 1990). This was effectively the first application of overall process simulation in bioprocessing. It enabled limited performance and economic evaluation of different processes to direct optimisation on an economic basis, rather maximisation of the productivity of individual operations in isolation (Petrides *et al.*, 1989). The findings of this simulation work indicated that the greatest reward for effort would be gained by improved methods of inclusion body isolation and renaturation. These finding were in contrast to other early economic assessments that did not utilise overall process modelling (Fulton *et al.*, 1992; Gordon *et al.*, 1990; Lightfoot *et al.*, 1987). Chromatography was identified as the most important area requiring process improvement in all three studies.

Homogeniser-centrifuge interactions were modelled and optimised on an economic basis by Middelberg *et al.* (1992) and a key conclusion was that homogenisation needed to be maximised to assist in primary recovery and improve the cleanliness of inclusion bodies for

subsequent renaturation. Datar *et al.* (1993) examined the production of active TPA from inclusion bodies within *E. coli* and in soluble form from mammalian cells. They concluded that the cost of refolding dominated the bacterial production process, making it unfavourable relative to production in mammalian cells. However, improving refolding methods had the potential to reduce the unit production cost of bacterial production to less than half of that in mammalian cells. Petrides *et al.* (1995) have also identified that in the case of recombinant human insulin expressed as inclusion bodies within *E. coli*, recovery and purification of the product dominate process economics.

More recently, cost estimation and sensitivity analysis for heparinase I expressed within *E. coli* has been conducted to compare soluble versus insoluble expression systems (Ernst *et al.*, 1997). It was concluded that if the expression system produced more than approximately 60% of the desired product in an insoluble form, it is more economically viable to discard the soluble product and only recover the insoluble material by renaturation. Hamel *et al.* (1996) have presented a similar analysis for production of tailspike endorhamnosidase of phage P22 in *Salmonella typhimurium*.

Comparison of operating modes for refolding and optimisation of protein concentration lend themselves to design analysis techniques typically used in more traditional chemical processing industries. However, none of the studies outlined above used a kinetic model of refolding in their simulations, relying instead on published mass-balance data. Clearly, there is a need to conduct a preliminary design analysis to compare the relative advantages of different refolding modes on a cost basis. This provides the impetus for the studies presented here.

Concurrent to the work presented in this thesis, others have applied process simulation to protein refolding. Chaudhuri *et al.* (1996) simulated refolding of HEWL assuming a simple kinetic model for the refolding pathways. They determined that initial protein concentration was the key parameter in batch refolding, and that refolding in continuous mode could increase yield 25-fold. Middelberg (1996b) used a similar kinetic model to simulate impact of different refolding conditions on the overall economic performance of a process to produce an analogue of IGF-I within *E. coli*. The optimum conditions for refolding in a continuous reactor were identified using a cost-based objective function.

1.5.6.2 Mass transfer

Another area of refolding that has not been systematically investigated is the effect of mass-transfer during transition from denaturing to native-favouring conditions. Since aggregation occurs between molecules in intermediate states, it is likely that the mass-transfer required for buffer exchange may be critical in aggregation of intermediates. For batch (or fed-batch and continuous) refolding by dilution, the mixing rate will influence the rate of mass-transfer. However, the degree of mixing is generally not stated in experimental investigations of refolding.

An early comparison of solvent exchange methods noted that dialysis over four hours gave greater yields than refolding by dilution where the denatured proteins were diluted slowly over 30 s (Teipel and Koshland, 1971a). Similar results have been obtained with diafiltration compared to batch dilution (Burton *et al.*, 1989). These results imply that gradual removal of denaturing agent is the better approach. However, Hejnæs *et al.* (1992) conducted extensive empirical optimisation of the refolding environment and then compared dialysis, chromatography, dilution, and hollow-fibre diafiltration methods for solvent exchange using a consistent renaturation environment. The same yield of renatured protein was obtained for all four methods.

Jaenicke and Rudolph (1986) have advocated rapid dilution at low protein concentration to minimise the exposure of high transient concentrations of intermediate species to reduce aggregation. Goldberg *et al.* (1991) reported that renaturation of HEWL critically depends on the efficiency of mixing during dilution. Lee and Rhee (1996) observed that aggregation of *Pseudomonas* lipase was reduced if dilution buffer was added to denatured solution, rather than vice versa. They concluded that refolding was influenced by the dilution mode and controlled by the initial concentration of denatured protein and therefore might be dependent on the rate of mixing. They also concluded that the time required to equilibrate the refolding solvent was not negligible compared to the rate of aggregation. Phosphoglycerate kinase (PGK) fragments have been observed to associate transiently within the mixing time of the experiment (Pecorari *et al.*, 1996). Terashima *et al.* (1996) advocate the use of a packed flow reactor to avoid the mixing inhomogeneities that occur in batch reactors. Hevehan and De Bernardez-Clark (1997) used vigorous agitation to

avoid creating transient areas of protein concentration in which the protein would be more susceptible to aggregation.

In review articles on refolding, the importance of the rate of solvent exchange has been highlighted as a significant variable. Cleland and Wang (1993) have concluded that the rate of denaturant removal affects the yield obtained from the process. Rudolph and Lilie (1996) summarised that some proteins aggregate quantitatively when refolded by dialysis, while others only aggregate upon refolding by rapid exposure to highly native-favouring conditions. This dichotomy is likely to be specific for the actual protein and the chemical environment that is used, but illustrates the need to establish whether the observed behaviour occurs due to mass-transfer limitation.

1.5.6.3 Contaminants

The initial purification steps of protein that originates as inclusion bodies is generally conducted before dissolution and further purification steps may even be carried out using denatured material (see section 1.2). However, one of the first investigations into renaturation in the presence of controlled amounts of protein and an extract of *E. coli*, demonstrated that the presence of these contaminants had no effect on renaturation yield (London *et al.*, 1974). Similarly, Jaenicke (1974) observed that the addition of bovine serum albumin (BSA) had no effect on the rate or yield of renaturation of lactic dehydrogenase (LDH). Weir and Sparks (1987) observed that concentration of crude protein contaminants from *E. coli* (up to 25-times more contaminant than test protein) did not influence the refolding yield of interleukin-2 (IL-2). However, increasing the concentration of IL-2 promoted aggregation, indicating that intermolecular association was highly specific. The renaturation of F_{ab} fragments was not effected by the presence of up to 3-fold greater concentrations of crude *E. coli* proteins and associated cellular components (Buchner and Rudolph, 1991). Even the self-association of bovine growth hormone and PGK in highly purified solutions has been shown to occur only between particular intermediates (Brems, 1988; Havel *et al.*, 1986; Mitraki *et al.*, 1987). These experimental observations indicate that aggregation of protein during refolding is highly specific. Therefore, an obvious question is: why do many processes involve such extensive purification prior to the refolding operation?

In the first case, renaturation yields of proteins that are susceptible to degradation by proteases from the host organism are likely to be improved by removing cellular contaminants prior to refolding (see section 1.3). Many proteins have been observed to display this behaviour. For example, Babbitt *et al.* (1990, 1991) have reported that the yield of active creatine kinase is greatly improved when host-cell protease are removed before initiating refolding by washing the inclusion bodies or purifying the denatured mixture using anion-exchange chromatography. Proteases that cause degradation of some proteins during dissolution have been found to be specifically associated with the cellular debris present in inclusion body isolates (Hatzfeld *et al.*, 1992; Sugimura and Higashi, 1988).

Secondly, other investigations have demonstrated that peptide aggregation is non-specific. In early process patents it was concluded that protein renaturation was improved by removing contaminants from the denatured mixture prior to refolding (*e.g.* Builder and Ogez, 1986). Galliher (1991) assumed that soluble or solubilised phospholipid, lipopolysaccharide, protein, and nucleic acid contaminants must be removed from all denatured mixtures prior to refolding as a matter of course. In the review of protein stability published by Schein (1990), she concluded that every protein contaminant present during refolding increases the total dilution necessary to avoid aggregation. Valax and Georgiou (1993a) noted that the effects of contaminants on refolding efficiency seems to be system specific. In experimental studies, they observed that the aggregation of β -lactamase during refolding in the presence of *E. coli* contaminants was non-specific (Valax and Georgiou, 1993b). The aggregates included essentially all of the contaminating proteins initially present in the inclusion bodies. They also concluded that other types of contaminants such as phospholipids may additionally influence the recovery of active product.

The review presented above attests to the observations of Valax and Georgiou (1993a); it would appear that the effect of contaminants on refolding efficiency is dependent on the protein product and expression system. During laboratory-scale development, extensive purification prior to refolding may be tedious, but reduces uncertainty when dealing with a poorly characterised production system. Since the large-scale processes are often based directly on the laboratory scale processes, it is apparent that some may be developed without consideration of the extent of purification actually required. For example, high

resolution chromatographic techniques are necessary for some products but may be dispensed with for others. Therefore, an important issue that must be addressed when considering economically-competitive protein renaturation procedures is the extent of host-cell contamination that may be tolerated.

Concurrent to the work presented in this thesis several reports on the effects of contaminants during refolding have been published. These will be discussed further in subsequent chapters to place them in the context of the presented work.

1.6 CONTAMINANTS AND PROCESS MODELLING

From a broad perspective there are three essential requirements of the general process outlined in section 1.2: fermentation, protein renaturation, and purification. Fermentation must be the first operation to generate the product, but there is flexibility in the order and selection of unit operations to achieve the remaining two requirements. Section 1.5.6.3 illustrates that the degree of purification prior to renaturation that is necessary to achieve high recovery is protein-specific. Accordingly, inclusion bodies may be dissolved and refolded in the presence of varying amounts of cellular contaminants (soluble and insoluble). To produce a final active product, purification must be sufficiently rigorous to reduce potentially toxic or pathogenic components to acceptable levels and some cost for removing the contaminants is unavoidable. Clearly, in large-scale processing, cost per unit produced is a more important design objective than yield.

Process simulation is the most cost-effective method for identifying an optimum process (Grossman and Biegler, 1995). It relies on process models to provide accurate *a priori* predictions of the performance of a particular unit for specified values of the feed stream composition and operating conditions. However, it can only be applied to restricted cases in bioprocessing because models of refolding that account for product purity have not yet been developed.

In this section an original system for rational identification of the interactions between host-cell contaminants and product during refolding is presented. The method is a useful first step in the development of a general refolding model that accounts for the presence of contaminants.

1.6.1 Traditional approach to biological variability

Modern analytical techniques are capable of uniquely identifying virtually all of the species that comprise bacteria if an adequate sample of sufficient concentration is available. Individual bacteria are too small to provide sufficient material for most assays, so populations must be sampled and overall average composition determined. Given

sufficient material, the logistical difficulties of conducting all of the analytical procedures make their completion impractical. Additional complexity is introduced because cellular composition varies with changing physiological states (see section 1.2.2). In other words, regardless of the number of analytical techniques completed, the results will only be valid for the set of environmental conditions and growth state used to produce the protein.

The convention employed to describe material generated by bacterial fermentation is to describe its history. Rather than uniquely quantify every component of the sample, sufficient information must be provided to describe the conditions under which the material was generated and stored. For example, to adequately describe a sample of *E. coli*, the history of the population must be included with any measured parameter of composition. The history should include the bacterial strain with details of any manipulation, medium composition, agitation and aeration conditions, temperature, growth phase, and growth rate (Fishov, 1995; Niedhardt, 1987). Table 1.1 gives the average composition of *E. coli* (strain B/r) in balanced growth at 37°C in aerobic glucose minimal medium with a mass doubling time of 40 minutes (Niedhardt, 1987).

Table 1.1: Average cellular composition of *E. coli* B/r cell.

Component(s)	% Total dry weight ^a
Protein	55.0
RNA	20.5
Lipid	9.1
Metabolites, cofactors, ions	3.5
Lipopolysaccharide	3.4
DNA	3.1
Peptidoglycan	2.5
Glycogen	2.5
Polyamides	0.4

^a Assuming that 70% (w/w) of the cell is water

With specific reference to dissolution, the feed material is a concentrated mixture of inclusion bodies contaminated with varying amounts of other cellular components (see section 1.2). The degree of contamination is dependent on both the initial culture, subsequent disruption, and primary recovery. One approach to describe the starting material is to extend the traditional method of specification used to describe fermentation and add to the history of the material by including details of the additional post-

fermentation operations. That is, an adequate specification would have to include not only fermentation, but exhaustive details of all equipment used along with the operating conditions for the steps of termination at completion of fermentation, cell lysis, any concentration steps such as filtration or centrifugation, and storage. This method of description is totally process-specific and precludes generic modelling.

1.6.2 Key parameters

Any useful model should predict the properties of the product stream from one unit to adequately estimate the performance of the following operation. As outlined above, prediction of this sophistication has not yet been achieved for even the initial operation of fermentation. Models of subsequent operations such as homogenisation and centrifugation have removed the need for detailed history to some degree by identifying the key parameters that influence those operations (Middelberg, 1992; Olbrich, 1989; Wong 1997). Using this approach is it possible to predict some measure of performance, but because process interactions are so poorly characterised and system-specific, generic models that predict the feed composition for the subsequent steps have not been developed. There is a need to identify key measurable parameters that can be used to quantify sample purity with sufficient detail to predict performance of dissolution and refolding.

1.6.3 Compartmentalisation

The hypothesis address by this work is that the influence of *E. coli* contaminants on the refolding of a heterologous protein from dissolved inclusion bodies can be correlated with the contaminating protein species present in the mixture (for constant environmental conditions). This assumption is based on the observations that self-association to give inactive multimers and proteolysis are cited as the major problems encountered during refolding (see section 1.5).

Theoretically it should be possible to identify the exact proteins that interact with each other. Once this has been established, rate equations that account for the concentration of each of the contaminants could be determined. Because *E. coli* is estimated to contain at

least 1850 populated protein species (Niedhardt, 1987), the practical implications of this strategy are discouraging, especially as identification of the interacting protein(s) may entail a search for an as-yet uncharacterised species.

An alternative approach that offers more widespread application is to group contaminants together into compartments rather than identify the exact species that interact *i.e.* adopt a systems approach. Since protein:protein interactions during refolding are system specific, refolding efficiency may potentially be increased by removing limited groups of proteins instead of all protein contaminants. The actual protein(s) need not be identified, but may be selectively separated from the product by astute process selection. For example, centrifugation or filtration may separate detrimental soluble contaminants from inclusion bodies and increase the yield of correctly refolded protein. In this case, the soluble and insoluble material are considered as two compartments that must be separated before conducting renaturation.

Once a compartment that includes interacting proteins has been identified, a model can be developed that uses a marker species to quantify the total contribution of the contaminants within the particular compartment. That is, no separation process is absolute so varying amounts of the interfering species will always be present. If the gross assumption is made that all species within a compartment vary in the same relative proportion, by quantifying the effect of differing amounts of the overall compartment, a model can be developed that correlates the observed behaviour to the amount of a marker species. The key issue from a processing perspective is to identify useful compartments.

Figure 1.5 illustrates the membrane structures of an *E. coli* cell containing a cytoplasmic inclusion body (summarised from Brock *et al.*, 1984). The inner and outer membranes are approximately 7 nm in thickness and are separated by the periplasm and murein sacculus. The membranes are attached to each other at between 200 and 400 sites per cell (Lugtenberg and van Alphen, 1983). Under normal growth conditions, the periplasmic space is between 20 and 40% of the total cellular volume (Oliver, 1987).

Lysis of the cell envelope during the process of recovering inclusion body protein from within *E. coli* is unavoidable. Therefore, unless particular care is taken to selectively remove the outer membrane to release the contents of the periplasm, such as in the process of making spheroplasts (French *et al.*, 1996), inclusion bodies will be exposed to all of the

cellular constituents. Therefore, division into broad compartments of soluble and insoluble material is a logical start as this is a common initial processing operation.

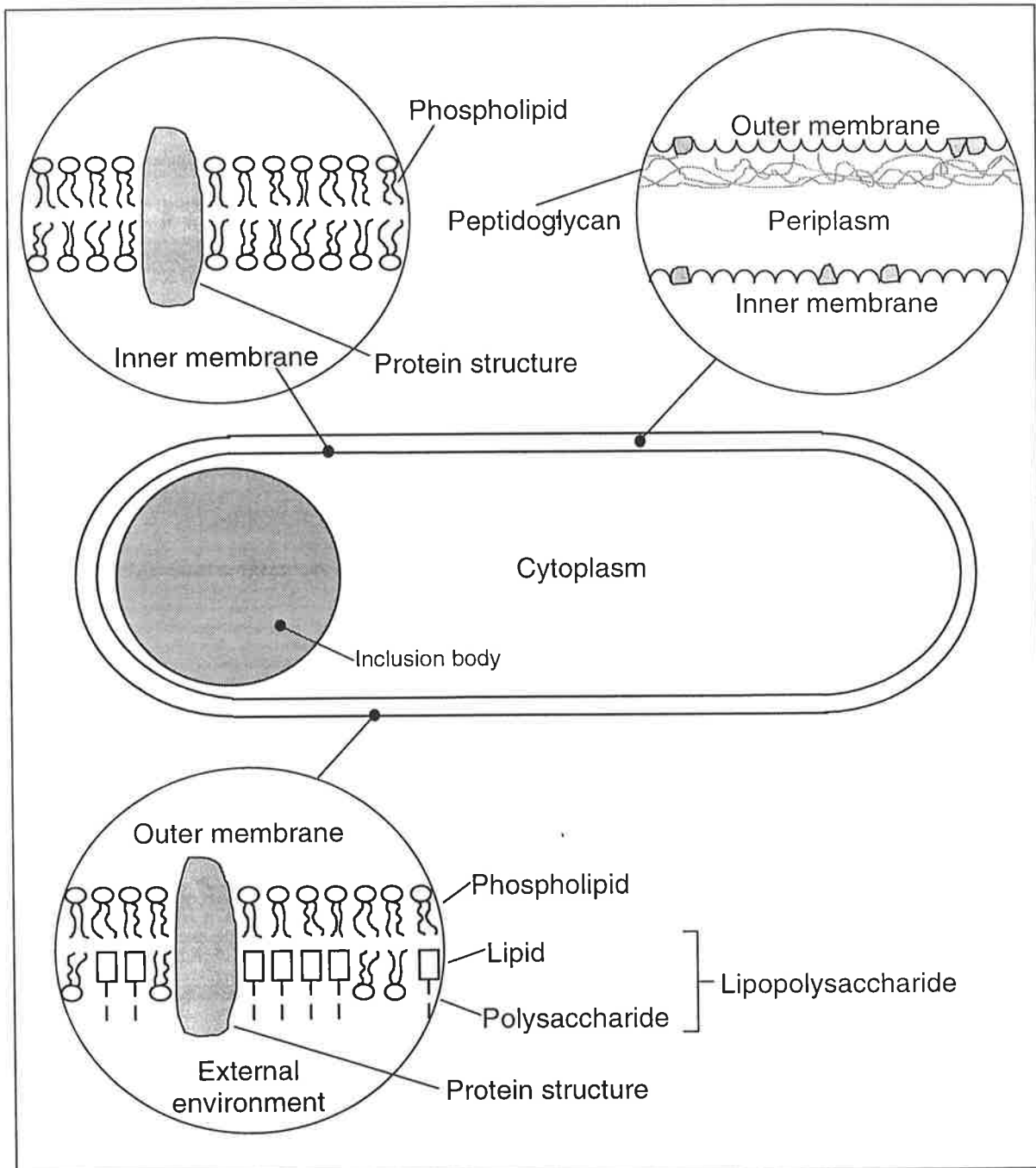


Figure 1.5: Overview of the structure of *E. coli*, illustrating location and subunit composition of inner and outer membranes.

1.6.3.1 Soluble compartment

There is no doubt that soluble proteins will be associated with inclusion body feed material during dissolution as no washing method will ensure their complete removal. Soluble proteins occur in both the cytoplasm and the periplasm. The soluble protease within *E. coli* have been characterised and their cellular location determined (Chung, 1993; Lazdunski, 1989; Swamy and Goldberg, 1982). Therefore, classification of soluble contaminants into periplasmic and cytoplasmic compartments is a logical approach since they may be selectively separated as indicated above.

Suitable enzymatic markers for the periplasmic compartment are acid phosphatase, alkaline phosphatase, and ribonuclease I (Cerny and Teuber, 1971; Lugtenberg and van Alphen, 1983). Enzymatic markers that may be used for the cytoplasmic compartment are aldolase, β -galactosidase, glucose-6-phosphate dehydrogenase, and inorganic pyrophosphatase (Cerny and Teuber, 1971; Marvin *et al.*, 1989, Schnaitman, 1970). However, the use of particular markers will be dependent on the particular strain of *E. coli* as they may express these components in different relative amounts.

1.6.3.2 Insoluble compartment

During *in vitro* processing to isolate the inclusion bodies for subsequent renaturation, it is possible that other contaminants that are insoluble under physiological conditions will be released. These species may interfere with the renaturation process in the same way as any other soluble contaminant of *E. coli*. Insoluble cellular contaminants are localised in the structural membranes and inclusion bodies. Classification of the insoluble compartment into sub-compartments for inner and outer membranes may be useful since they are physically distinct and can be separated. For example, many of the membrane-associated proteases of *E. coli* have been characterised and their location determined (Grodberg and Dunn, 1988; Kaufmann *et al.*, 1994; Lazdunski, 1989; Sugimura and Nishihara, 1988).

The outer membrane is located external to the murein sacculus of *E. coli*. It is comprised of predominantly of phospholipid and lipopolysaccharide (LPS) with protein inserts. Many unique structural outer membrane proteins have been identified (for reviews see Lugtenberg and van Alphen, 1983; Nikaido and Vaara, 1987; Osborn and Wu, 1980).

These species may be used as markers for the outer membrane compartment as they can be identified and quantified by gel electrophoresis or immunoaffinity techniques (Diedrich *et al.*, 1990; Osborn *et al.*, 1972; Schnaitman, 1974). Lipid A is a subunit of lipopolysaccharide and is a major molecular constituent of the outer membrane (Raetz, 1987). Sensitive assays for quantitation of lipid A have been developed as it can produce severe reactions in mammals when adventitiously administered in therapeutic formulations. Therefore, it may be a very useful marker for the outer membrane compartment.

Unlike the outer membrane of *E. coli*, the inner membrane does not contain LPS as an integral component, and is predominantly comprised of phospholipid. The inner membrane also contains many dehydrogenases (Cronan *et al.*, 1987). In particular, the enzymes succinic, lactic, and pyruvic dehydrogenase are known to be associated with the inner membrane. However, succinic dehydrogenase is the most often used as a cytoplasmic membrane marker (Schnaitman, 1970; Wolf-Watz *et al.*, 1973).

1.6.3.3 Proposed compartments

Figure 1.6 outlines the structured breakdown of *E. coli* homogenate into the compartments used in this investigation. The complex cellular components are considered as five major compartments: soluble protein product, other soluble proteins (including proteases), insoluble protein of interest, other insoluble proteins (including proteases and membrane proteins), and the remaining components (those which are measurable and/or unknown) in the final compartment. Four of the compartments consider only protein species, consistent with the hypothesis of the investigation. The fifth compartment contains the balance of cellular constituents. These compartments are more specific than the division into soluble and insoluble, but less specific than division into cellular location.

Division into more specific sub-compartments did not seem warranted for the initial investigation into the effects of gross cellular contamination on renaturation presented subsequently (see sections 5.3 and 5.4). However, to demonstrate the further application of this method, breakdown of the four protein compartments into protease compartments is shown in Figure 1.6 as proteolytic degradation is such a major issue during renaturation.

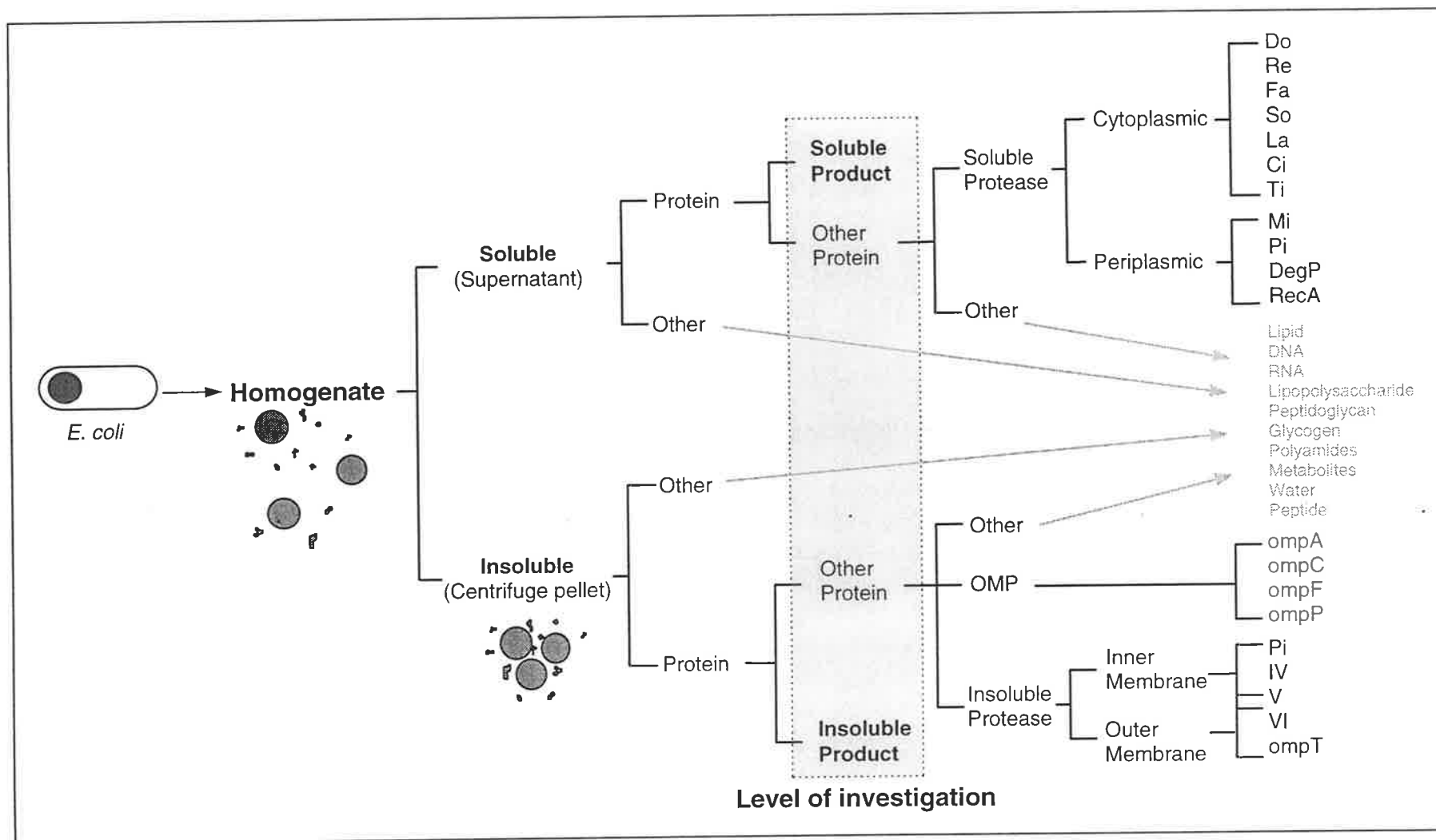


Figure 1.6: Levels of complexity to describe the composition of *E. coli* using compartments. Sample composition depends on quantity and quality of each compartment. The work presented in this thesis is restricted to quantification of only four compartments as the preliminary step in identifying host-cell species that interact with product during renaturation.

Some specific outer membrane proteins are also shown because membrane proteins have been reported to be significant exogenous host contaminants in isolated inclusion body mixtures (Hart *et al.*, 1990; Hartley and Kane, 1988; Marston *et al.*, 1984).

Dissolution must provide the feed material for refolding and the assumption is made that the material that is dissolved and refolded can be described by combination of the four compartments. Various methods can be used to quantify the protein composition of the compartments.

Once the significant interactions have been identified between or within compartments, subsequent investigations can be directed to identify components within sub-compartments (*e.g.* protease localisation). This compartmentalised strategy provides the first structured method for dealing with contaminating species during protein renaturation.

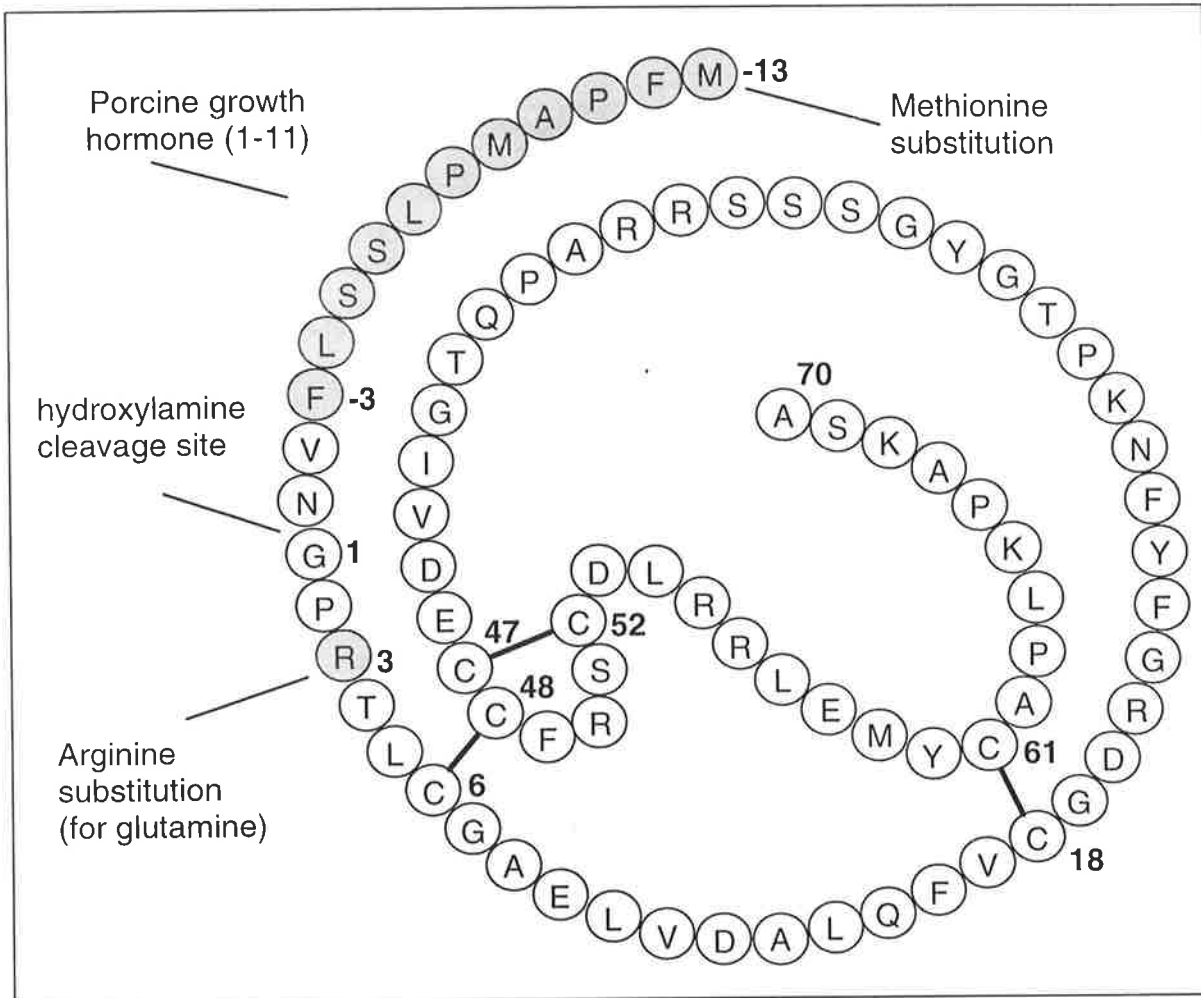
1.7 PROTEIN OF INTEREST

Many useful proteins are produced as inclusion bodies within *E. coli*. They include human insulin, human Tissue-type Plasminogen Activator (TPA), macrophage-colony stimulating factor, human and porcine growth hormones, human growth factors, and their analogues (Belagaje, 1995; Chang *et al.*, 1994; Dorin *et al.*, 1995; Rudolph *et al.*, 1995; Wells *et al.*, 1994).

Insulin-like Growth Factors (IGFs) are members of the insulin-related family of peptides that have a high degree of sequence identity with insulin. They are anabolic hormones that exhibit both endocrine and auto/paracrine activity (Sussenbach, 1989). They stimulate cell proliferation and differentiation, and for this reason have potential clinical applications in aiding post-trauma recovery and in stimulating wound healing and tissue regeneration. Their mitogenic activity makes them effective replacements for serum in cell culture media (Humbel, 1990).

Insulin-like growth factor I (IGF-I) is a protein that contains 70 amino acids, has an isoelectric point (pI) of 8.8 and a molecular mass of 7650 Da. Modern genetic manipulation techniques have made it possible to produce biosynthetic analogs of IGF-I. Some of these derivatives have reduced affinity for the inhibitory IGF-binding proteins, and consequently show increased biological potency (Francis *et al.*, 1992). IGF-I and its analogues are useful test-proteins for developing protein processing strategies (Matthiesen *et al.*, 1996; Samuelsson and Ulhén, 1996) as they have many common features of useful proteins: they are compact and contain multiple intramolecular disulfide bonds. LongR³IGF-I is one of these highly potent analogs (Francis *et al.*, 1992; Wells *et al.*, 1994) and is the protein used for the studies presented in this thesis.

LongR³IGF-I comprises an 11 amino acid sequence derived from porcine somatotropin, a methionine substituted for the first amino acid, a hydroxylamine cleavage site, and the full IGF-I sequence with the third amino acid (glutamate) substituted with arginine ([Met¹]-pGH(1-11)-Val-Asn-[Arg³]IGF-I). The primary structure of LongR³IGF-I is presented in Figure 1.7; the peptide contains 83 amino acids, has a pI of approximately 9.1, and a molecular mass of 9.11 kDa.



conformation ([18-61], [6-47], [48-52]) or aggregates (Hober *et al.*, 1992). Formation of an equilibrium non-native conformation demonstrates that the primary sequence does not contain all of the information necessary to uniquely define the tertiary structure of the molecule as proposed by Anfinsen (1973).

The oxidative refolding mechanism of LongR³IGF-I has been characterised by Milner (1996) using acid-quenching of highly purified protein to trap intermediate species and determine their disulfide arrangement. It was found that disulfide arrangements between IGF-I and the modified fusion peptide form are the same. However, the folding pathway of the analogue was found to differ from that of mature IGF-I. A major difference was that the refolding pathway proceeded through a highly populated transient intermediate with less detectable intermediates than with IGF-I. Also, the relative amount of protein with misfolded conformer was significantly less for the analogue and the rate of formation of native conformation was slightly reduced. It was concluded that the leading extension of 15 amino acids reduces the number of refolding intermediates relative to IGF-I and confers stability to the native folding form, resulting in its increased predominance at equilibrium (Milner *et al.*, 1995). Milner (1996) has also characterised the elution characteristics of LongR³IGF-I in its native, misfolded, and reduced, denatured conformations using analytical HPLC.

The typical production process outlined in section 1.5 has been demonstrated as being suitable for fermentation, and mid-stream processing of inclusion bodies containing LongR³IGF-I (King *et al.*, 1992). Conditions that successfully dissolve the inclusion bodies using 8M urea and subsequently refold the peptide to achieve highest yield of active protein by batch dilution have been selected (Francis *et al.*, 1992). The optimum refolding conditions were determined by systematic variation of pH, oxidising redox couple, and urea concentration. These studies were conducted using inclusion bodies that had been washed prior to dissolution and subsequently purified by size-exclusion chromatography before refolding. Rate constants of first-order refolding and second-order aggregation have been regressed for refolding the material produced using this process (Kotlarski *et al.*, 1995). The work of Francis *et al.* (1992) also describes a chromatographic purification strategy to produce a final fusion peptide of sufficient purity for protein characterisation studies.

1.8 THESIS STRATEGY

The strategy adopted to address process-scale renaturation was to use numerical simulation coupled with experimental investigation. Simulation was used initially as a tool for investigating the impact of different refolding strategies on large-scale process performance. In experimental studies an *E. coli* strain expressing LongR³IGF-I as inclusion bodies was used as the model system. Studies addressed the fundamental issues of mass-transfer limitation and variation in reaction rates with processing strategy. Rate constants regressed to a three-state refolding model were used as a measure of performance for refolding trials. The key findings of the initial simulation work were also tested experimentally.

Models of refolding mechanisms that may have been suitable for reactor design analysis were considered and a simple generic mechanism proposed. Dimensionless reactor design equations for batch, fed-batch and continuous modes of operation were derived using this generic kinetic scheme. A method for quantifying overall process performance in terms of dimensionless operating cost was developed. Simulations were conducted in dimensionless form, not only to allow easy comparison between different flow sheets, but more importantly to make the findings generally applicable to a range of proteins produced via similar processes. The mathematical framework employed for process simulation is presented in Chapter 3. The results of simulations over an extensive range of operating conditions in batch, fed-batch, and continuous modes are presented in Chapter 4.

Experimental investigations into dissolution of inclusion bodies and refolding of LongR³IGF-I under a range of mixing conditions, cellular contamination, and at high concentration are presented in Chapter 5. Consistent dissolution and refolding methods were used throughout to ensure matching thermodynamic conditions between trials.

Dissolution of inclusion body mixtures was studied to establish whether consistent material could be produced for subsequent refolding experiments. The first refolding experiments were designed to determine whether the kinetics of refolding were limited by mass-transfer. A series of refolding trials were conducted under three different mixing rates and tested for equivalence. Subsequent trials focused on quantitating the effects of gross cellular contamination on the kinetics of refolding. Using inclusion body pastes and aqueous

solvents of differing purities, batch refolding studies were conducted. Figure 1.8 illustrates the approach used to study the effect of cellular contaminants on refolding kinetics. Validation of fed-batch refolding concluded the experimental work. A series of batch refolding trials at high protein concentration were conducted as controls. These were followed by repeated trials of fed-batch refolding.

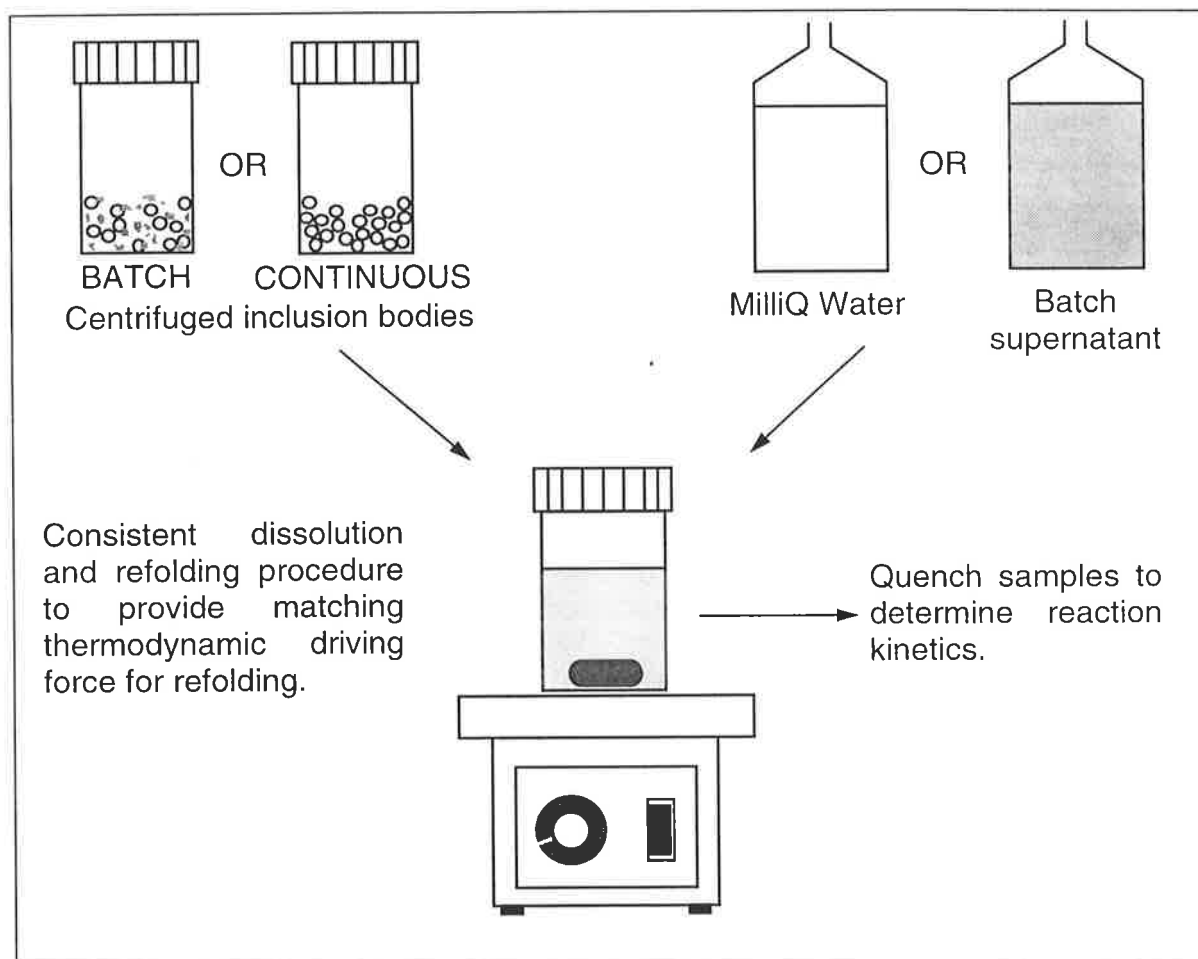


Figure 1.8: Illustration of the overall experimental procedure to quantify the effect of gross cellular contamination on refolding kinetics. Cellular fractions were generated from a single batch fermentation of *E. coli* JM101 containing the expression vector for LongR³IGF-I.

The conceptual structure of this thesis is presented diagrammatically in Figure 1.9. The discussion presented in Chapter 6 draws together the results of process simulation and experimental investigations to highlight the significant findings of this work. Deficiencies in the presented scale-up strategy are outlined and possibilities for future improvements are identified.

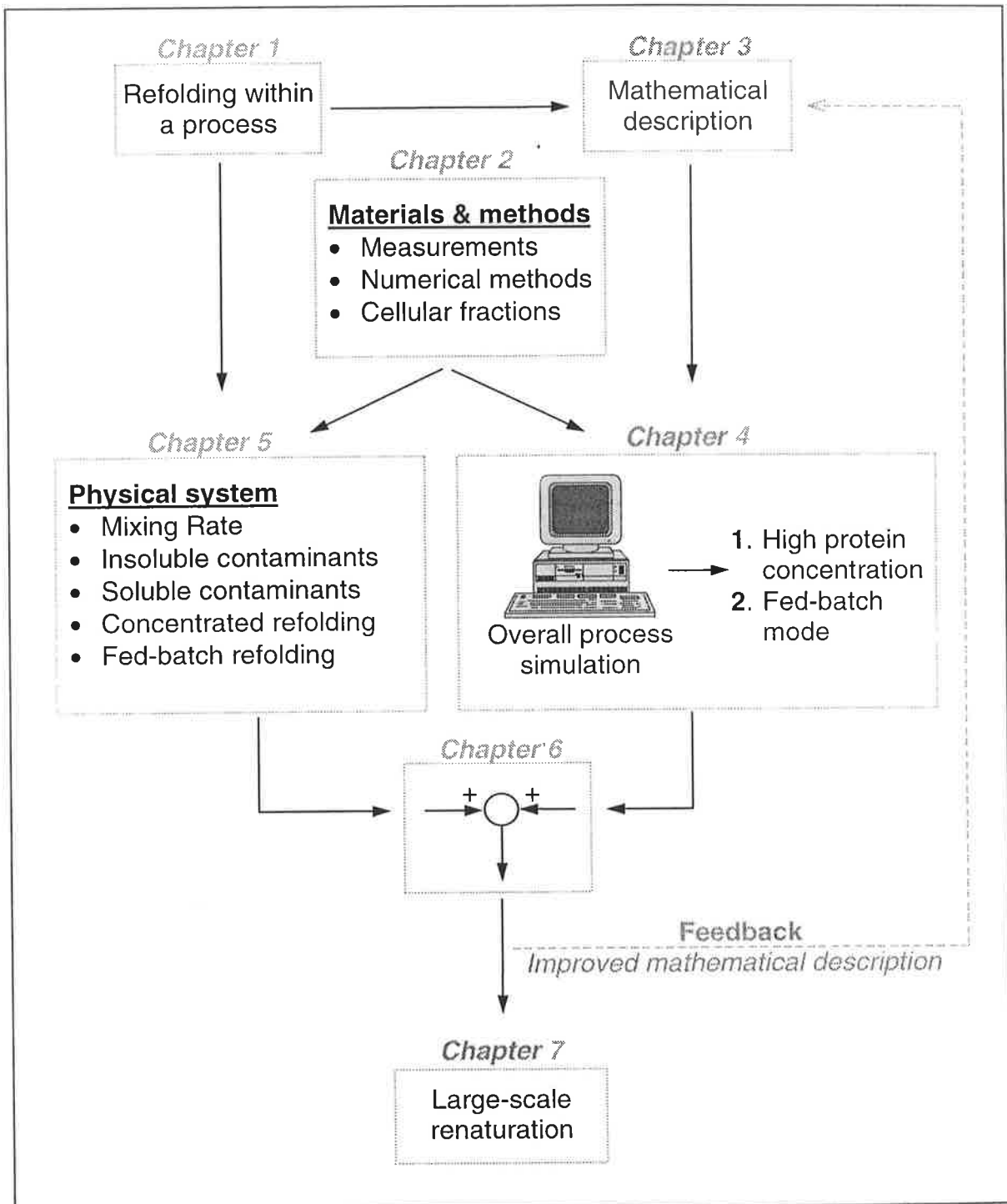


Figure 1.9: The conceptual structure of this thesis and division into chapters.

2. MATERIALS AND METHODS

The procedures used for quantification, numerical methods, and standard experimental protocols are presented in this chapter. Generation of the starting cellular fractions and characterisation by HPLC, SDS-PAGE, and total protein content are described.

2.1 MATERIALS

Unless otherwise stated, the grade and supplier of general chemicals and reagents are as presented in Table 2.1. Unless specified, MilliQ™ reagent grade water was used for all aqueous solutions (Millipore, Bedford, MA, United States of America). Purified receptor grade LongR³IGF-I, for use as control standard, was kindly provided by GroPep Pty Ltd.

Table 2.1 : Grade and suppliers of general chemicals and reagents.

Reagent	Grade	Supplier
Acetic acid	Glacial	BDH Laboratory Supplies, Poole, England.
Bovine Serum Albumin (BSA)	RIA Grade	Sigma Chemical Company, St Louis, MO, United States of America.
Coomassie brilliant blue G		Sigma Chemical Company, St Louis, MO, United States of America
Dithiothreitol (DTT)		Boehringer Mannheim, Mannheim, Germany.
Ethylene Diaminetetraacetic acid (EDTA)	Univar	Ajax Chemicals, Auburn, N.S.W., Australia.
Gluteraldehyde	Unilab	Ajax Chemicals, Auburn, N.S.W., Australia.
Glycerol	Unilab	Ajax Chemicals, Auburn, N.S.W., Australia.
Glycine (Gly)	AnalaR	BDH Merck, Kilsyth, Vic., Australia.
Hydrochloric acid (conc.)		BDH Laboratory Supplies, Poole, England.
2-hydroxyethyl disulfide		Aldrich Chemicals, Milwaukee, WIS, United States of America.
Methanol	HiPerSolv	BDH Laboratory Supplies, Poole, England.
Nitrogen (gas)	High Purity	Airliquide, Melbourne, Australia.
Phenol Red		BDH Chemicals Ltd, Poole, England.
Sodium dodecylsulfate (SDS)	≈ 99%	Sigma Chemical Company, St Louis, MO, United States of America.
Sodium Chloride	AnalaR	BDH Merck, Kilsyth, Vic., Australia.
Tricine SDS running buffer (10×)		Novel Experimental Technology, San Diego, CA, United States of America.
Trifluoroacetic acid (TFA)	HiPerSolv	BDH Laboratory Supplies, Poole, England.
	Biochemika	Fluka Chemie AG, Buchs, Switzerland.
Tris(hydroxymethyl) aminoethane (Tris)	Ultrapure	Aldrich Chemicals, Milwaukee, WIS, United States of America.
Urea	Cryst. extra pure	Merck, Darmstadt, Germany.

2.2 ANALYTICAL METHODS

2.2.1 Analytical disc centrifuge

The size distribution of inclusion bodies released by homogenisation was determined using a DCF4 analytical disc centrifuge (Applied Imaging Ltd, Gateshead, U.K.). The method used was developed by Middelberg (1992) to quantitate cell disruption and later demonstrated as an effective means of measuring the size of LongR³IGF-I inclusion bodies (Middelberg *et al.*, 1993). It has also been used to measure the size distribution of prochymosin, γ -interferon and porcine growth hormone inclusion bodies (Taylor *et al.*, 1986, Middelberg *et al.*, 1990).

Briefly, the method involves using the disc centrifuge to establish a density gradient throughout a spinning fluid. The centrifuge was operated at a speed of 8000 r.p.m. and gain of 6. A volume of 20 mL of water was used as the spin fluid, injection of 1.5 mL of 20% (v/v) ethanol and a three-pulse boost protocol generated the density gradient. Samples were suspended in 20% (v/v) ethanol at a concentration sufficiently dilute to avoid streaming and 0.5 mL samples of each solution were injected and sedimentation of insoluble particles recorded for 15 minutes. The sedimentation time-profile was then transformed into a frequency-diameter distribution using Stokes' Law and an extinction coefficient correction to account for variations in the optical properties of the particles with size (Middelberg *et al.*, 1990). Constant fluid viscosity of 1.14 cP was assumed and inclusion body density was taken as 1260 kg m^{-3} , consistent with the density of porcine growth hormone (1260 kg m^{-3}) (Middelberg, 1996a).

2.2.2 Total protein quantitation

A simple colorimetric assay, based on the method of Bradford (1976), was used to estimate the total protein concentration in refolding mixtures and starting solutions. The crux of the assay is the change in maximum absorbance of an acidic solution of Coomassie Brilliant Blue G-250 dye from 465 to 595 nm when binding to protein occurs. Protein dye concentrate was purchased from Bio-Rad (Bio-Rad Laboratories, Hercules, CA, USA; Cat. no. 500-0006). The microassay procedure of the Bio-Rad assay was used as follows: 200

μL of dye concentrate were added to 800 μL of sample, the mixture was vortexed and incubated at room temperature. After incubating for 15 minutes the absorbance at 595 nm was measured. The increase in absorbance relative to protein-free samples was transformed into an estimate of the protein concentration by calibration of the assay with standard solutions of BSA. Accordingly, the total protein estimates given below correspond to the equivalent concentration of purified BSA standard.

Since acid conditions are required, 0.1M NaCl/0.1% TFA was used to dilute all samples prior to adding the dye reagent. A minimum of a 1:100 dilution of all refolding and dissolution samples was made to ensure that the concentrations of buffering chemicals and urea would be within the range known not to interfere with the assay. Autopipettes were calibrated to deliver their nominal volume ($\pm < 1\%$) and all samples were analysed in duplicate.

The linear range of the assay was observed to be approximately 2 to 12 $\mu\text{g BSA mL}^{-1}$. This compares well with the range of 1.2 to 10 $\mu\text{g mL}^{-1}$ claimed by the supplier.

2.2.3 High Performance Liquid Chromatography

Prepacked reverse-phase high performance liquid chromatography (HPLC) columns were obtained from Applied Biosystems Inc., Foster City, CA, USA (Brownlee™, aquapore butyl, 7 μm particle size, 300 Å pore size, 100 $\times$ 2.1 mm). Columns were stored in methanol when not in use. TFA was used for acidification of all mobile-phase solutions. HiPerSolv grade acetonitrile was purchased from BDH Laboratory Supplies, Poole, England.

Three computer controlled HPLC systems were used for analysis. Two were controlled by GBC DP800 software: ICI DP800 interface, two high-pressure GBC LC1110 HPLC pumps, Laboc Gastor GT-104 in-line degasser, Knauer in-line mixer, GBC LC1610 autosampler, and GBC LC1200 UV/VIS detector set at 214 nm. The third system was a Shimadzu class LC-10 system: CBM-10A communications module, LC-10AD pump, FCV-10AL in-line mixer, GT-102 in-line degasser, Rheodyne 7125 manual injection valve, and SPD-10A UV/VIS detector set at 214 nm.

Stock feed solutions of 0.1% (v/v) TFA and 80% (v/v) acetonitrile/0.08% (v/v) TFA were used to generate linear elution gradients. Flow was maintained at a constant rate of 0.5 mL min⁻¹. The C4 HPLC column was first equilibrated with 0.1% TFA/25% acetonitrile prior to loading samples. Samples were acidified with TFA (pH < 2.5) and centrifuged at $18 \times 10^3 g_{\max}$ for 3 minutes before being loaded. Once the solvent had been eluted and baseline absorbance was established, a linear gradient was applied to the column, increasing the acetonitrile concentration to 50% in 30 minutes. Since most samples analysed were of unfractionated inclusion bodies, each gradient profile included a 72% acetonitrile wash step after the linear elution. Figure 2.1 presents a summary of the gradient in graphical form.

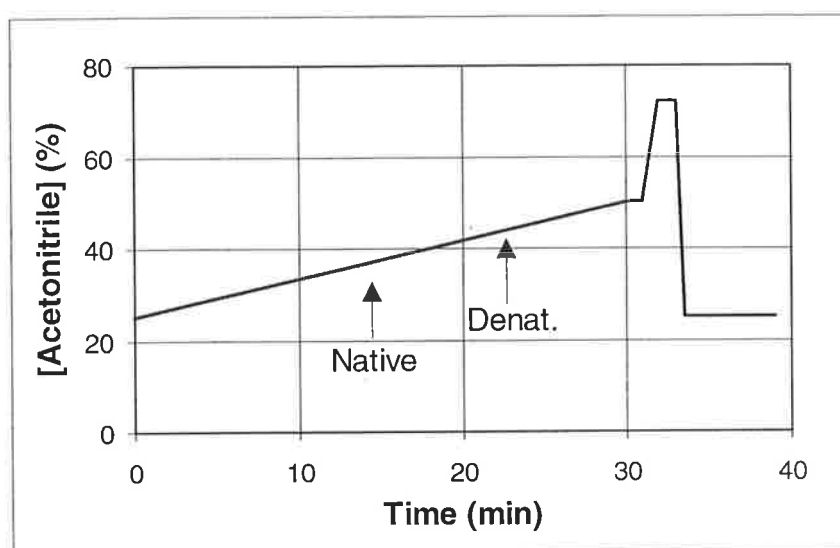


Figure 2.1 : Graph of acetonitrile concentration versus time in the mobile phase used for HPLC. The time at which native- and denatured-conformation protein elutes is shown, taking into account the 'dead-time' of the pipework.

To ensure accurate quantitation, the autosampler injection volume was determined by the difference in sample tube weight post-sampling, assuming a density of 1000 kg m⁻³. A standard sample of receptor grade LongR³IGF-I in either native or reduced, denatured state was analysed approximately every tenth sample. The extinction coefficient of the UV detector was determined daily, based on these standard runs. Blank runs were also conducted routinely to provide current baselines for subtraction and to confirm absence of slow-eluting species. The area of the peak corresponding to native-conformation protein was estimated using PeakFit™ as outlined in section 2.3.2.

To confirm that the active sites on the C4 matrix were not saturated during analysis, giving uncontrolled displacement of the product, a series of control tests were conducted. Receptor grade standard was added to duplicate samples of denatured and refolded material to increase the maximum total amount of protein by approximately 50% relative to the amount routinely loaded. The samples from refolding conducted in the presence of gross cellular contaminants were of prime interest. In all tests, recovery of the spiked protein was within $\pm 5\%$ of the expected amount and no significant shift in the elution time was observed. These results confirmed that the analytical column was loaded below maximum capacity during operation.

The applicability of this analytical technique to resolve LongR³IGF-I in its native conformation from misfolded isomers has been demonstrated by Milner (1996).

2.2.4 Spectrophotometry

Progress of fermentations and dye-binding for assay of the total protein were measured using spectrophotometers. Samples of fermentation broth were transferred into 2.5 mL disposable Kartell cuvettes (Kartell, Milan, Italy; Cat no. 1940, 10 mm path length) and their absorbance measured with a Unicam 8625 UV/VIS spectrophotometer set at 600 nm (Unicam Ltd, Cambridge, U.K.). Cell-free medium was used at the zero-absorbance standard.

While the samples for the total protein quantitation assay were incubating at room temperature, they were transferred into 1.0 mL disposable Kartell cuvettes (Kartell, Milan, Italy; Cat no. 1938, 10 mm path length). At the appropriate time the absorbance was measured using either a Unicam 8625 or Cecil CE 1120 (Cecil Instruments, Cambridge, U.K.) spectrophotometer set at 595 nm. Water was used as the zero-absorbance standard.

2.2.5 Electrophoresis

Non-reduced samples were prepared by addition of at least an equal volume of 2× non-reducing sample buffer (0.9 M Tris-HCl pH 8.45/24 % (v/v) glycerol/8% (w/v) SDS/1.5% (w/v) Coomassie Brilliant Blue/0.5% (w/v) phenol red). To achieve reducing conditions,

1.6 M DTT was added to a final concentration of 50 mM. All samples were heated at 85°C for a minimum of two minutes and centrifuged for three minutes at $18 \times 10^3 g_{\max}$ before loading a volume of 3-15 μL per well.

Precast 10-20% polyacrylamide gels were supplied by Novel Experimental Technology, San Diego, CA, USA (80×80×1 mm, 10 well). A tricine buffering systems was employed as it gives superior resolution of low molecular weight proteins. Stock tricine SDS running buffer was diluted 1:10 immediately before use to give 0.1 M Tris/0.1 M Tricine/0.1% (w/v) SDS. Gels were run using a water-cooled Xcell™ Mini-Cell unit (Novel Experimental Technology, San Diego, CA, USA). The maximum sample volume loaded was 15 μL . Running conditions for the gels were a potential difference of 125 V with a maximum current of 80 mA per gel, supplied by an EC600-90 power-pack (E-C Apparatus Corporation, St Petersburg, Florida, USA).

All gels were run using Novex Mark12™ wide range protein standards in lane 1 (Novel Experimental Technology, San Diego, CA, USA). The proteins (and approximate molecular masses) contained in the standard mixture were: Myosin (200 kDa), β -galactosidase (116.3 kDa), Phosphorylase b (97.4 kDa), Bovine serum albumin (66.3 kDa), Glutamic dehydrogenase (55.4 kDa), Lactate dehydrogenase (36.5 kDa), Carbonic anhydrase (31 kDa), Trypsin inhibitor (21.5 kDa), Lysozyme (14.4 kDa), Aprotinin (6.0 kDa), Insulin B chain (3.5 kDa), and Insulin A chain (2.5 kDa).

Before staining, gels were incubated for 30 minutes in 2.5 (v/v) % glutaraldehyde at room temperature to fix low molecular weight proteins. Gels were stained by incubating in 50 mL of 0.1% (w/v) Coomassie R 350/30% (v/v) methanol/10% (v/v) acetic acid for 30 minutes with gentle agitation (Coomassie R 350 supplied by Pharmacia Biotech AB, Uppsala, Sweden). Destaining was achieved by incubating gels at room temperature in destain solution (30 (v/v) % methanol/10 (v/v) % acetic acid) with gentle agitation. Frequent changes of destain solution over 24 hours were required to reduce background staining satisfactorily. Gels were carefully air-dried between cellophane sheets after equilibration in 20 (v/v) % methanol/2 (v/v) % glycerol.

2.2.6 Concentration of acid-quenched samples

Since the total protein concentration in samples taken during refolding was low, they were concentrated so that significant bands could be visualised by SDS-PAGE. The general method that was used involved quenching 600 μL of each refolding mixture with 15 μL of neat TFA to halt disulfide shuffling. Then the quenched samples were centrifuged at $18 \times 10^3 \text{ g}_{\text{max}}$ for three minutes and 500 μL of the supernatant was decanted for drying. The samples were then dried under a pressure of less than -95 kPag using a Hetovac system (Heto VR-1 centrifuge and Heto CT-110 condensate trap) evacuated by a Javac DD 150 vacuum pump. Samples were dry after 12-24 hours.

Some of the pellets recovered from acid-quenched HPLC samples were also dried using the Hetovac system. This generally required evaporation of only 20-50 μL per sample and was accomplished in less than 12 hours.

2.2.7 Densitometry

The amount of protein present in bands resolved by SDS-PAGE was quantified by densitometry to allow comparison of relative intensities within the same gel. A Molecular Dynamics 300A densitometer (Molecular Dynamics, California, USA) was used to scan air-dried gels. The scanned images were integrated using ImageQuant™ v3.3 (Molecular Dynamics). Gels were assumed to have destained evenly and the background absorbance was measured for each gel, then automatically subtracted during integration of protein bands. Desired bands were selected manually and integrated using a volume scan. Calculated volumes were copied into EXCEL for subsequent analysis.

Since comparison between different gels was not required and only relative differences were sought, the linearity of the method for quantitation was not validated.

2.2.8 General experimental methods

All HPLC samples were analysed within 48 hours of quenching. Protein quantitation was completed within 24 hours of sampling. To improved the accuracy of quantitation, the

actual volume dispensed by each autopipette was calibrated by measuring the mass of water dispensed for each of 10 pipetting operations. The average and standard deviation for each pipette was calculated and used in evaluation of protein concentration and estimates of error. The error introduced due to pipetting was less than $\pm 2\%$ for all samples.

Dissolution and refolding trials were conducted at room temperature. Samples that had to be stored for more than 12 hours before analysis were refrigerated at 4°C.

2.3 NUMERICAL METHODS

2.3.1 Process simulation

Mathcad 6.0e Plus (Mathsoft Inc., Cambridge, MA, USA) was used to solve the reactor design equations outlined in section 3.2. The design equations for batch refolding were in the form of two dimensionless coupled ordinary differential equations; one first- and the other second-order. They were solved repeatedly at discrete points to determine overall yield using an adaptive step-size fourth-order Runge-Kutta algorithm for various sets of dimensionless groups over a range of conversions. Linear interpolation of the adaptive step-size output gave values for the solved variables at desired values of conversion and dimensionless group. The solution strategy is presented graphically in Figure 2.2. For example, the equilibrium constant, K_n , was held constant at 0.01 and values of y_N and τ determined for a range of X from 0.99 to 0.9999 and $C_{D0}K_A/k_N$ from 4 to 12.

Mathcad was then used to determine the dimensionless DFC, consumable, and waste treatment costs for the operating conditions corresponding to each yield. That is, the capacity, consumable consumption and waste production of each unit operation was calculated for units able to achieve the design basis for each set of yield, conversion and dimensionless groups. The corresponding objective cost function (see section 3.5) was determined from the sum of

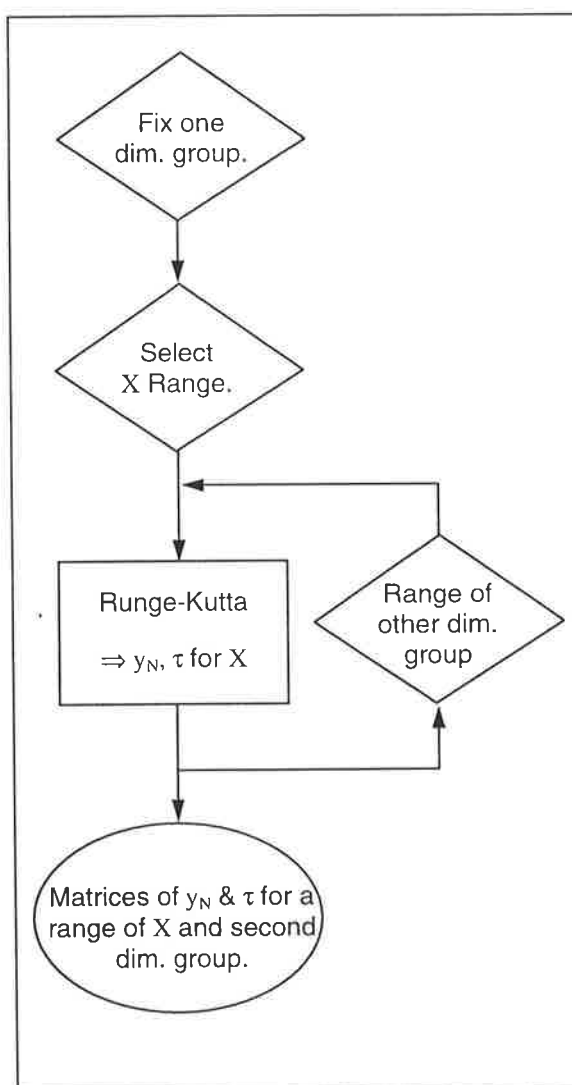


Figure 2.2: Logic diagram of the method used to generate the two matrices of y_N and τ values over ranges of X and a dimensionless group. One dimensionless group must be held constant.

their individual contributions and stored as a two-dimensional matrix. The optimum conditions, giving the minimum value of the objective function, were identified graphically.

The solution methodology for fed-batch refolding was the same as that for batch mode. The design equation for continuous refolding is an explicit quadratic expression for τ and an algebraic solution may be found. Mathcad's symbolic evaluation functions were used to find the two quadratic roots of Equation 3.25. Substitution of a set of X , Da_{BN} , and Da_{AN} values into the roots gave one positive and one negative solution. The positive root was taken as the solution to the design equation. The corresponding yield was determined by substitution into Equation 3.24. As for batch and fed-batch refolding simulations, the values of the objective cost function were calculated for ranges of X and the two dimensionless groups.

2.3.2 Peakfit™

HPLC analysis of solutions containing many protein species results in elution of multiple peaks. To monitor the progress of refolding reactions it is necessary to resolve not only different protein species, but even protein conformations. HPLC has been successfully used to monitor refolding of LongR³IGF-I from receptor-grade starting material (Milner, 1996). The profiles obtained when refolding using unfractionated inclusion bodies as starting material were confounded by coeluting species. To deconvolute profiles and enable accurate quantitation of the native-conformation LongR³IGF-I, PeakFit™ v4.0 (Jandel Scientific, San Rafael, CA, USA) was used.

Raw data from the UV/VIS detectors of the HPLC systems were corrected for baseline variation using the software provided with each system. This baseline-corrected data was then exported into an ASCII file, and saved into a Microsoft™ EXCEL v5.0c worksheet. Data from the EXCEL file was imported into PeakFit for analysis.

To determine which distribution function best described the observed profiles, various alternatives were compared for standard samples of receptor grade protein. Neither pure Gaussian (normal) nor Lorentzian distributions described the single peak well. It was found that a Pearson VII distribution gave the best fit. The Pearson VII function is a

convolution of the Gaussian and Lorentzian distributions. The Savitzky-Golay procedure was employed to smooth the raw data so that peak-centres could be assigned according to turning points in the second-derivative of the profiles. Peaks were regressed to the raw data using least-squares. Individual peak placement and shape was consistent between profiles. An example of a fitted profile is presented in Figure 2.3. The area of the fitted peak at the appropriate elution position was used to determine the concentration of protein species present in the injected sample.

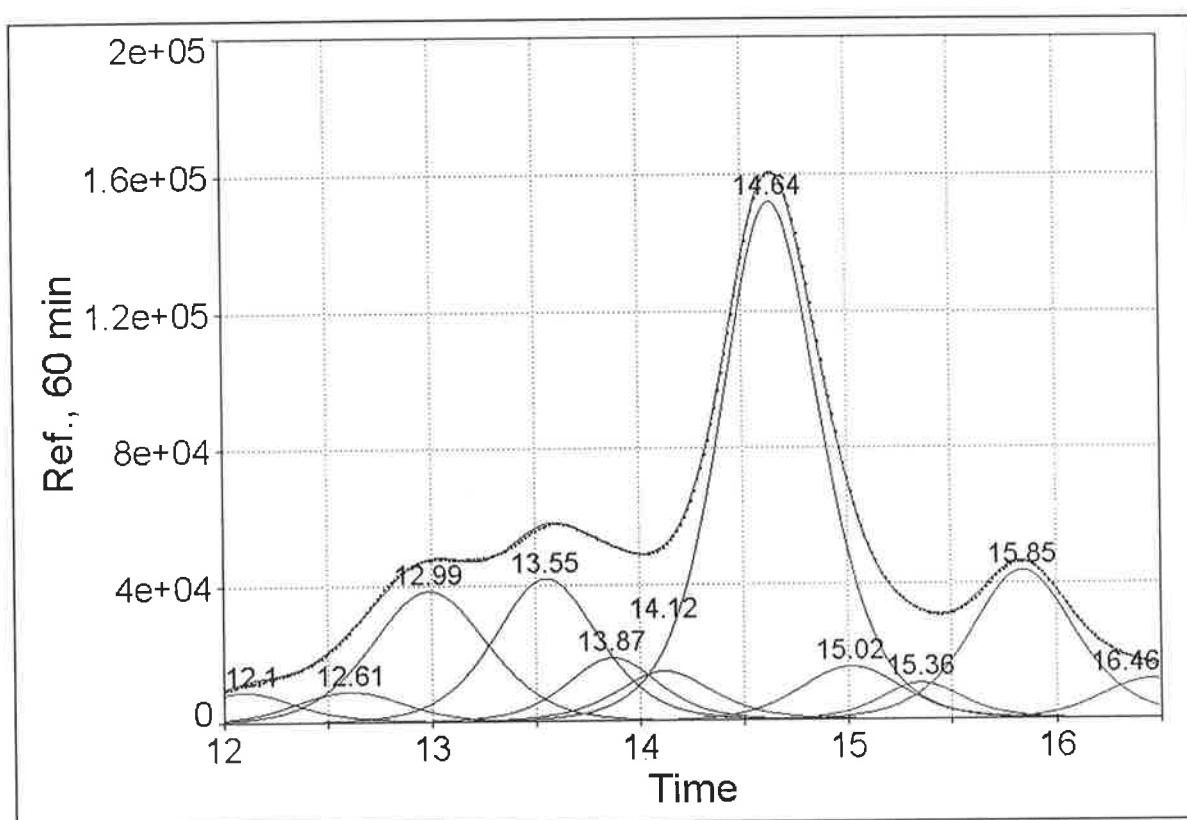


Figure 2.3: An example of the HPLC elution profile for a 60 minute refolding sample of clean inclusion bodies. The PeakFit™ plot presents the raw data as absorbance on the y-axis (μV) versus time on the x-axis. The composite profile, composed of the sum of fitted peaks, is indistinguishable from the raw data. The fitted range was manually selected to include the native species (14.64 minutes).

2.3.3 Mathcad

To determine estimates of the rates constants for different refolding experiments, Mathcad 6.0e Plus (Mathsoft Inc., Cambridge, MA, USA) was used to regress the transient LongR³IGF-I concentration to equations (3.4) and (3.5). An adaptive step-size fourth-order Runge-Kutta algorithm was coded within Mathcad to integrate the two equations inside a minimisation loop. The loop used an in-built Levenberg-Marquardt algorithm to regress the three rate constants to determine the minimum sum of squared errors (SSE).

Mathcad's programming environment offered the advantages of being able to present contour plots of the SSE about the minimum and simple alteration of the rate equations. It was also able to solve the rate equations for fed-batch refolding where a discontinuity is introduced at completion of feeding. However, to provide estimates of variance in the rate constants it was necessary to use a dedicated statistics programme.

2.3.4 Estimates of variance in rate constants

Estimates of variance in the reaction rate constants determined by regression to experimental batch refolding data were calculated using BMDP Release 7 (BMDP Statistical Software, Inc., Los Angeles, CA, USA). The BDMP package uses a modified Runge-Kutta-Fehlberg algorithm to numerically integrate differential equations. It uses a pseudo-Gauss-Newton iterative algorithm for estimating function parameters by least squares fitting.

Regressions for estimation in parameter variance using BMDP were conducted by Mr Philip Leppard in the Department of Statistics, University of Adelaide. Selection of the method used for the statistical analysis presented in section 5 to compare rate constants was suggested by Professor Richard Jarrett from the Department of Statistics, University of Adelaide.

Outlying data points were taken to be those with standardised residual values greater than 2.5σ (Kuehl, 1994). Outliers were not included in regression calculations. Patterning in residual values was tested according to the method of Draper and Smith (1966), by determining the number of same-sign 'runs' in consecutive residuals.

2.4 PROCESSING OF STARTING CELLULAR FRACTIONS

As outlined in section 1.8, various fractions of homogenate were used as the starting material for refolding experiments (see Figure 1.8). The procedure used to prepare the fractions is outlined in Figure 2.4. Broth suspension of whole cells from a 15 L fermentation was homogenised and one third centrifuged in a batch centrifuge. The pellet from batch centrifugation was taken as the 'dirty' inclusion body mixture and the supernatant was stored to provide a supply of soluble contaminating proteins. The remaining two-thirds of the homogenate was diluted to 20 L and differentially sedimented using a continuous centrifuge. The solid fraction was resuspended in 20 L and recovered by two more cycles of differential centrifugation. The solids recovered from the third centrifugation were taken as the 'clean' inclusion body starting material. Specific details of the procedure are presented in the following sub-sections.

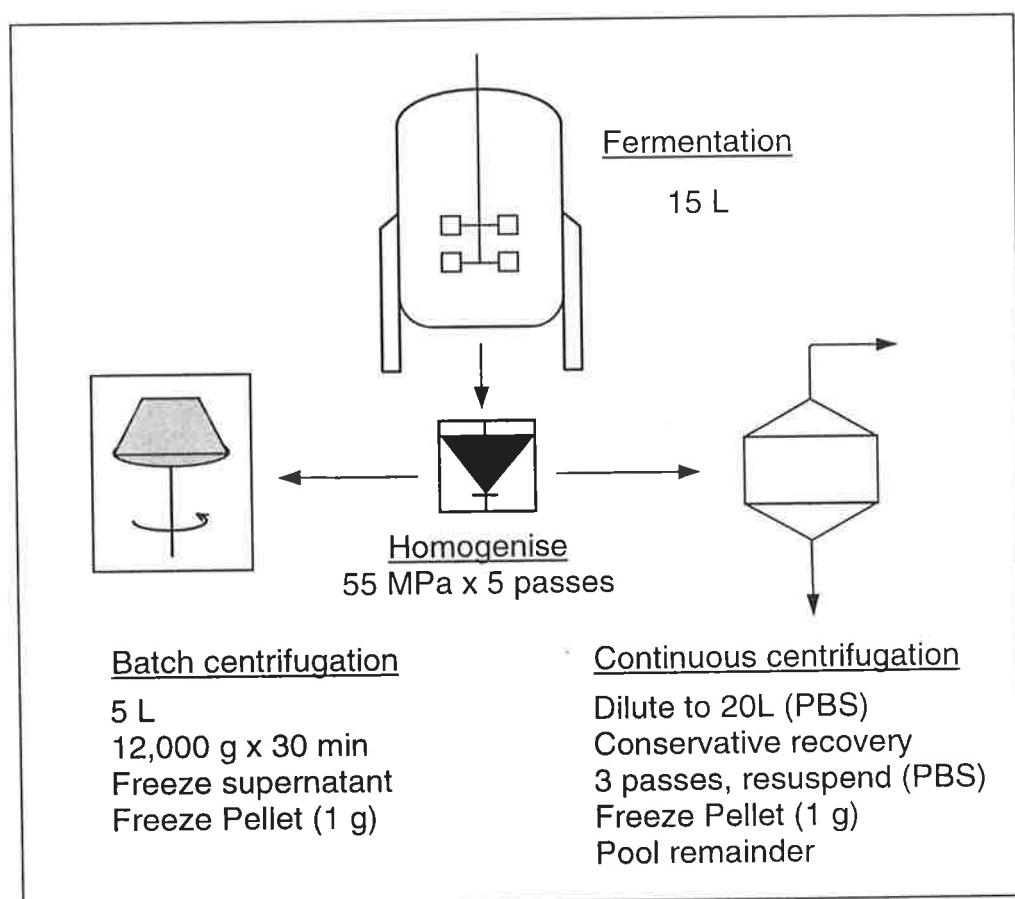


Figure 2.4: Schematic of procedure used to generate starting fractions

2.4.1 Host

E. coli was the bacterial host used for expression of LongR³IGF-I. The particular strain used was JM101 (American type culture collection (ATCC no: 33876)). The genotype for *E. coli* JM101 is F' *traD36 proAB⁺ lacI^q lacZΔM15 Δ(pro-lac) supE mcrA thi λ⁻*. It is an overproducer of the *lac* repressor. A practical feature of this strain is the requirement to supplement minimal medium with thiamine to sustain growth.

2.4.2 Vector

The vector used to express LongR³IGF-I was constructed from the plasmid pGHXSC.4, based on the parent plasmid pKT52 (Vise and Wells, 1987). Construction is described by Francis *et al.* (1992). The vector is illustrated in Figure 2.5. It contains the ampicillin resistance gene, the *trc* promoter, an optimised ribosome binding site and a spacer region together with the modified cDNA coding for the product. Expression of the product within *E. coli* JM101 is induced by addition of IPTG, to give cytoplasmic inclusion bodies. Inclusion of the *amp* resistance gene allows ampicillin to be used as the selection pressure for preparation of the inoculum, thus ensuring that the plasmid has not been rejected from the host during subculturing.

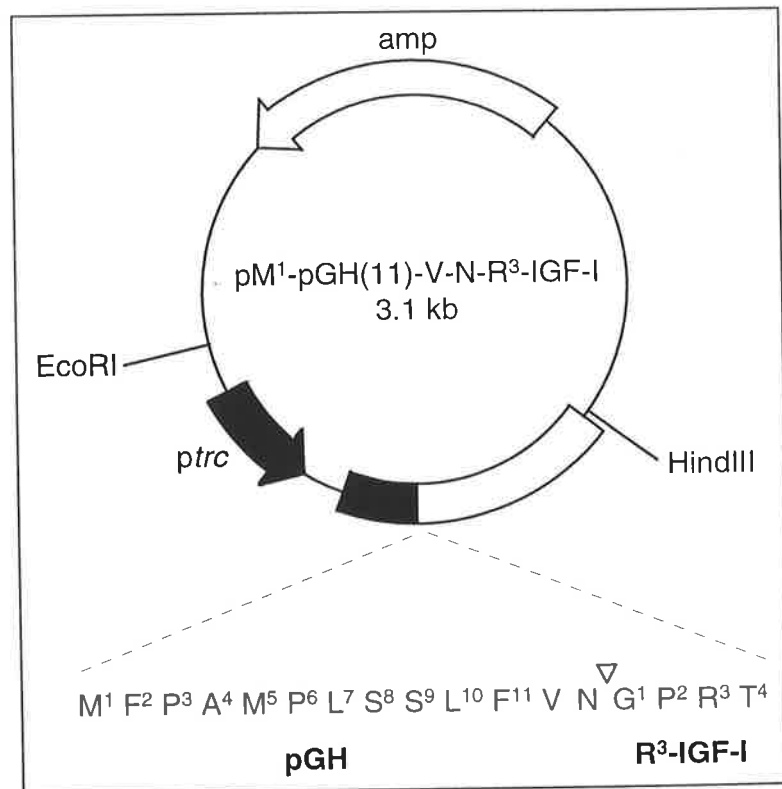


Figure 2.5: Illustration of expression vector. Solid and open areas represent p[Met¹]pGH(1-11) and [Arg³]IGF-I sequences, respectively. Single letter abbreviations for the first 17 amino acids of product are shown.

2.4.3 Inoculum preparation

Plates of *E. coli* JM101 containing the desired plasmid, grown on defined medium in the presence of ampicillin, were kindly donated by GroPep Pty Ltd. Using freshly sterilised stock solutions of trace elements, glucose with magnesium, thiamine, and ampicillin, duplicate 250 mL Erlenmeyer flasks containing 21.1 mL of minimal medium were prepared using aseptic technique. Chemical composition of the medium is presented in Table 2.2. Each flask was conditioned at 37.0°C for approximately 30 minutes before inoculating it with a single colony. Flasks were then incubated at 37.0°C with shaking for 10-12 hours. Immediately prior to inoculating the fermenter, each flask was aseptically sampled and the optical density at 600 nm (OD_{600}) measured to calculate the inoculum volume required. An optical density of less than 1.0 was used to ensure that the inoculum had not become oxygen-limited. The required volume was approximately 2 mL.

2.4.4 Fermentation

Fermentation was conducted using a 20 L Chemap CF-2000 fermenter with a working volume of 15 L (Chemap AG, Volketswil, Switzerland). The fermenter was sterilised *in-situ* after adding the ammonium, sodium, and potassium salts dissolved in 14 L of water purified by reverse osmosis. Trace elements, thiamine, glucose and magnesium salt were added aseptically to give a volume of 15 L with the composition specified in Table 2.2. Prior to inoculation, the medium was conditioned for approximately 30 minutes: pH was adjusted to 6.90, temperature to 37.0°C, and the medium was saturated with oxygen. The fermenter was inoculated in the evening with sufficient inoculum to give an OD_{600} of approximately 5 approximately 14 hours later. During fermentation, broth temperature was controlled to 37.0°C, pH was controlled at 6.9 by addition of 7 (w/v) % NH_4OH , and dissolved oxygen concentration maintained above 40% of saturation by control of stirrer speed at a maximum airflow of 1 v.v.m. Foam was dispersed by the dual action the Fundafoam™ mechanical foam breaker and addition of a chemical antifoam (1 (v/v) % Henkle Foamaster E81-V; Henkle, Vic, Australia.). Expression of LongR³IGF-I was induced by addition of 1.1 g of IPTG (dioxane free; Sigma Chemical Co., St Louis, MO, USA) when the culture had achieved an OD_{600} of approximately 7. Induction was allowed

to proceed for 5.5 hours before termination. The OD₆₀₀ of the broth was approximately 15 at that time.

Fermentation was terminated by reducing the temperature of the broth to below 10.0°C, while maintaining oxygen- and nutrient-rich conditions. Once the temperature of the broth had dropped sufficiently, it was transferred to a jacketed holding vessel for homogenisation.

During all operations involving viable recombinant organisms, care was taken to minimise aerosol generation. In cases where aerosol generation was unavoidable (*e.g.* fermentation and homogenisation), containment was ensured by filtering all off-gas through validated, sterilising 0.22 µm filters.

Table 2.2: Fermentation and shake flask media composition. All reagents for inoculum preparation and fermentation were of laboratory reagent grade or superior.

Reagent	Shake flask medium (g L ⁻¹)	Fermenter medium (g L ⁻¹)
D-glucose	3.2	36.4
NH ₄ Cl	2.47	2.58
KH ₂ PO ₄	2.39	2.54
Na ₂ HPO ₄ ·2H ₂ O	5.13	5.22
K ₂ SO ₄	1.82	1.94
MgSO ₄ ·7H ₂ O	0.62	0.67
FeSO ₄ ·7H ₂ O	0.019	0.020
MnSO ₄ ·H ₂ O	0.0049	0.0051
ZnSO ₄ ·7H ₂ O	0.0082	0.0086
CuSO ₄ ·5H ₂ O	0.00076	0.00080
Na ₃ Citrate	0.0083	0.0880
HCl (conc.)	0.038 (mL L ⁻¹)	0.040 (mL L ⁻¹)
Thiamine	0.038	0.040
Ampicillin	0.046	-

2.4.5 Homogenisation

Cell disruption was achieved by multiple discrete passes through a high-pressure homogeniser. A 15MR APV-Gaulin high-pressure homogeniser (CD valve) was used at a maximum pressure of 55.2 MPa. To ensure complete disruption of whole cells, a total of five discrete passes were conducted. Homogenate feed was cooled to a maximum of 10°C to ensure that the product was not damaged by the heat generated. Between passes, the disruption and inclusion body size were measured using the analytical disc centrifuge assay outlined in section 2.2.1. No reduction in inclusion body size was detected between homogeniser passes. Greater than 95% disruption was achieved after three passes, but additional passes were completed to reduce the size of cell debris and thereby improve the separation efficiency of the subsequent centrifugation operation.

2.4.6 Batch centrifugation

Approximately one third of the homogenate (nominally 5L) was treated with formalin solution to give a final concentration of 0.01% (v/v). The insoluble material was then sedimented by centrifugation in 250 mL tubes at $30 \times 10^3 g_{\max}$ for 30 minutes at 4°C in a Beckman JA-14 rotor using a Beckman J2-21 centrifuge operating at 14×10^3 r.p.m. (Beckman Instruments, Fullerton, CA, USA). Supernatant was recovered from all sample tubes and pooled. Solids were not removed from the centrifuge tubes between cycles, but pooled after completion of the final, fifth run.

The collective solids were mixed thoroughly to give a homogeneous paste, aliquoted in 1.0 g lots into 15 mL screw-capped tubes and frozen at -20°C for subsequent refolding studies. Some smaller aliquots were taken for use in characterisation studies to avoid freeze/thawing of larger amounts. The pooled supernatant was dispensed into 500 mL, 250mL, and 1 mL aliquots and frozen at -20°C.

2.4.7 Continuous centrifugation

Differential centrifugation was carried out with a Veronesi KLE-160 solid-bowl disc-stack centrifuge operating at speed of 8400 r.p.m. ($\Sigma=3775 \text{ m}^2$) (Veronesi Separatori, Bologna, Italy). The rest of the homogenate (nominally 10L) was diluted to 20 L with phosphate buffered saline (PBS) ($1.17 \text{ g L}^{-1} \text{ NaCl}/1.57 \text{ g L}^{-1} \text{ KH}_2\text{PO}_4/2.62 \text{ g L}^{-1} \text{ Na}_2\text{HPO}_4$ pH 6.9 ± 0.1) and centrifuged at a feed rate of 0.4 L min^{-1} . The solids were recovered from the bowl and resuspended in 20 L of PBS and centrifuged at a feed rate of 0.4 L min^{-1} two additional times. After the third centrifuge pass, the solids were recovered, mixed thoroughly give a homogenous paste, aliquoted in 1.0 g lots into 15 mL screw-capped tubes and frozen at -20°C for subsequent refolding studies. Some smaller aliquots were taken for use in characterisation studies to avoid freeze/thawing of larger amounts. The inclusion body paste recovered from the third centrifugation is referred to as 'clean' inclusion bodies.

2.5 CHARACTERISATION OF STARTING CELLULAR FRACTIONS

Particle size distributions of the fermentation broth and suspensions prepared after one, three, and five homogeniser passes are presented in Figure 2.6. Homogenate from the fifth pass appeared to be free of whole cells and to contain inclusion bodies with an average diameter of approximately 0.38 μm . This material was subsequently centrifuged to generate the dirty and clean inclusion body pastes.

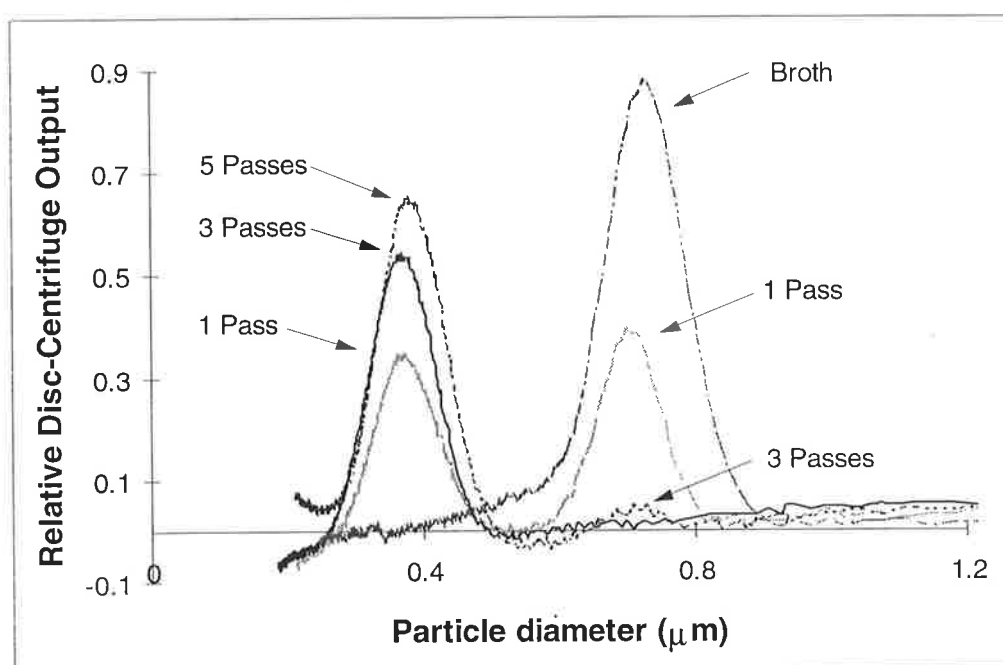


Figure 2.6: Particle size analysis of whole broth and homogenate passes using the DCF4 analytical disc centrifuge. The first peak (approximately 0.38 μm) corresponds to inclusion bodies and the second peak to whole cells. After one pass a large whole-cell peak is present, but is only just detectable after three passes. After 5 passes, no whole cells can be detected by this technique.

Particle size analysis of the supernatants from the three centrifuge passes and batch centrifugation are presented graphically in Figure 2.7. There was a reduction in the concentration of inclusion bodies present in the supernatants from continuous centrifugation with increasing number of passes. However, it can be concluded that at a feed rate of 0.4 L min^{-1} , recovery of inclusion bodies is sacrificed to the supernatant when compared to batch centrifugation. The supernatant from batch centrifugation appeared to

be free from inclusion bodies and to contain reduced amounts of small cellular debris relative to the supernatants from continuous centrifugation (no trough at approximately 0.2 μm). This is consistent with the observation that the batch supernatant was quite clear, while the supernatants for continuous centrifugation were opaque (which was reduced by increasing number of centrifuge passes). Batch supernatant was quite viscous and a clear, dark brown colour. The supernatant from the first continuous centrifuge pass was less viscous and an opaque, dark brown colour. After treating this supernatant with disinfectant and leaving to stand for four hours, significant solid material had sedimented to the bottom of the holding tank. The supernatants from the second and third passed were slightly opaque, but otherwise quite colourless.

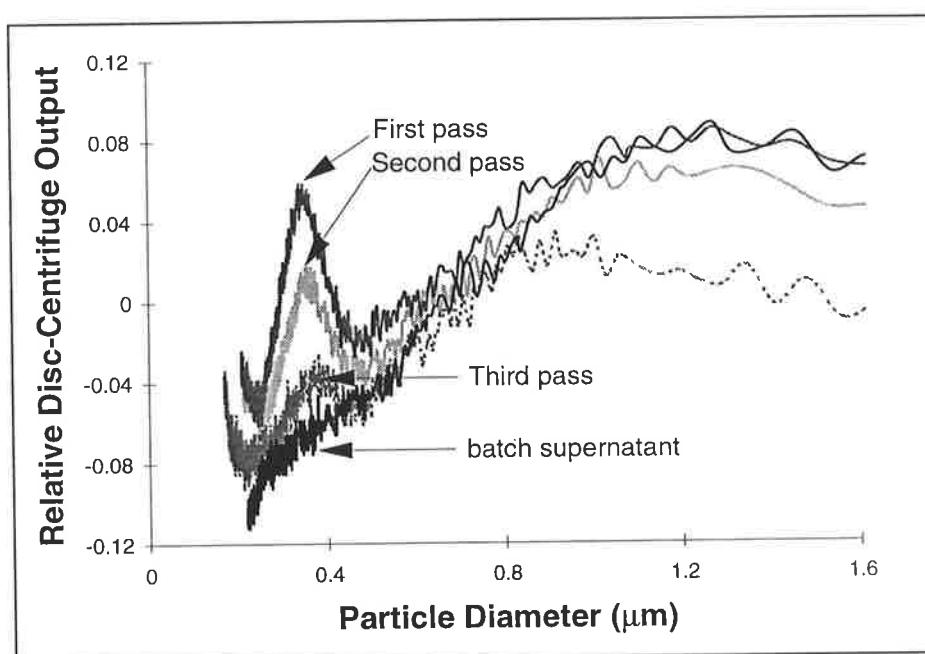


Figure 2.7: Particle size analysis of centrifuge supernatants using the DCF4 analytical disc centrifuge. The output appears very noisy due to the low concentration of solids in the analysed samples. There is a trend towards greater recovery with increasing number of passes for the continuous centrifuge. The supernatant from batch centrifugation appears to be free of inclusion bodies, indicating total recovery.

The ‘clean’ inclusion body paste recovered from continuous centrifugation was quite soft and consistent in texture. When pooled, only moderate mixing was required to achieve a homogeneous paste. The ‘dirty’ inclusion bodies recovered from batch centrifugation required extensive homogenisation to achieve a consistent paste.

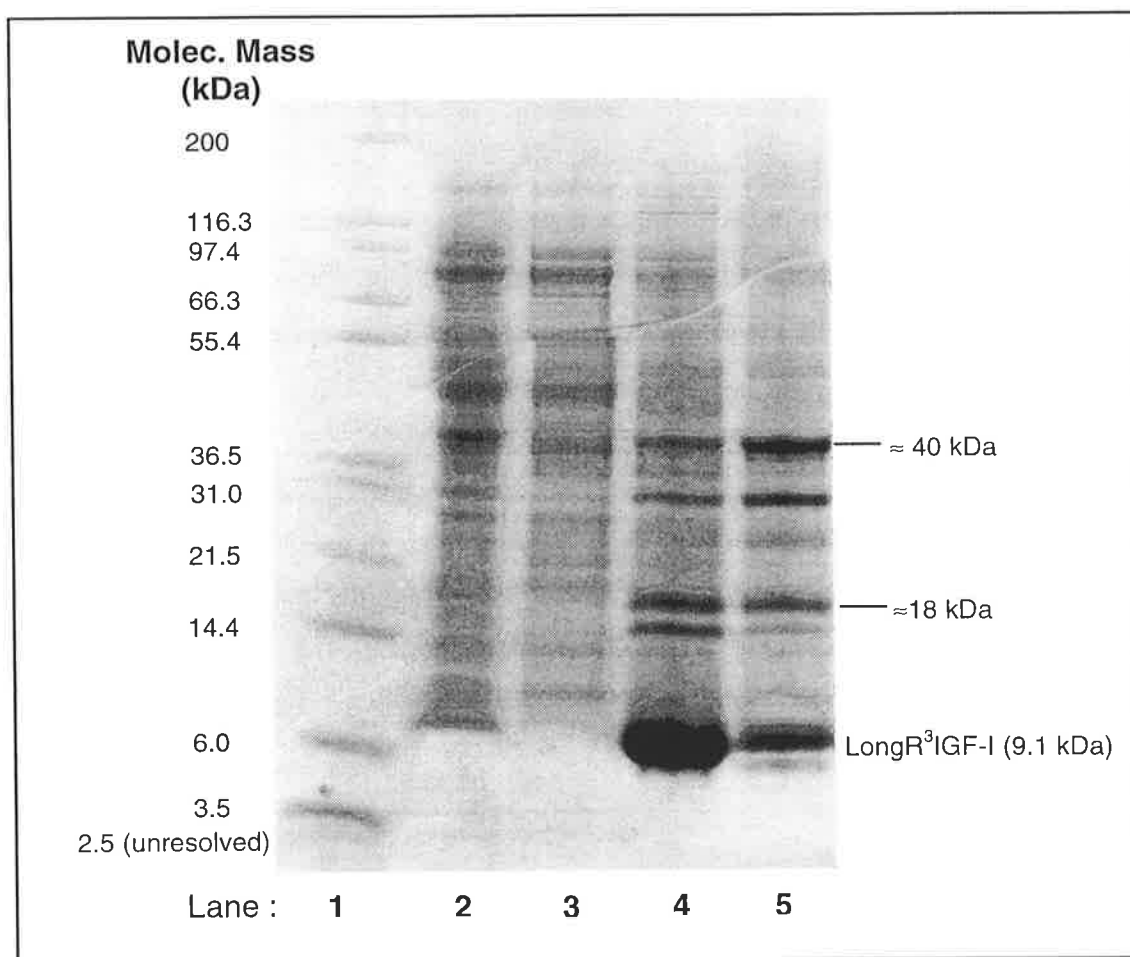


Figure 2.8: Reducing SDS-PAGE of starting material cellular fractions. 10-20% Tricine gel as described in section 2.2.5. *Lane 1:* Mark12™ wide range protein standard, *Lane 2:* 15 μ L whole broth diluted 1:1 with 2 $\times$ load buffer, *Lane 3:* 15 μ L of batch supernatant diluted 1:1 with 2 $\times$ load buffer, *Lane 4:* 3 μ L ‘clean’ inclusion body dissolved 1:20 solids:2 $\times$ load buffer, *Lane 5:* 3 μ L ‘dirty’ inclusion bodies dissolved 1:20 solids:2 $\times$ load buffer

A reducing SDS-PAGE gel of whole broth, batch supernatant, ‘clean’ inclusion bodies, and ‘dirty’ inclusion bodies is presented in Figure 2.8. The distribution of protein species present in the whole broth was quite even. Some major bands at approximately 40 and 45 kDa could be seen. A significant band was present at approximately 9 kDa, corresponding to expressed fusion protein. The batch supernatant was noticeably lacking in low molecular weight species (<20 kDa), but quite abundant in high molecular weight species. A corresponding SDS-PAGE of supernatant from the third centrifuge pass did not contain sufficient protein to visualise any species (results not shown). This would imply that the clean inclusion bodies should be relatively free of soluble protein contaminants.

Lane 4 was overloaded to allow visualisation of contaminating protein species present in the triple-centrifuged inclusion body paste. The distribution of protein species was dominated by the band corresponding to product at 9 kDa. The next most populous species had molecular masses of approximately 16 and 18 kDa. There were significant bands at approximately 31 and 40 kDa. Lane 5 contained the same nominal loading of dirty inclusion bodies as lane 4 did of clean inclusion bodies. The band at 9 kDa was still the dominant species, but the amount of 31 and 40 kDa species was relatively much greater. This indicated that the pellet contained a greater proportion of membrane protein than the triple centrifuged pellet, and significant purification had been achieved by the differential sedimentation.

In all lanes the 40 kDa band was likely to have been due membrane protein A (OmpA). However, the consistent presence of an 18 kDa species may indicate that some protein is expressed as a covalently bound dimer within the inclusion bodies or generated by the analytical process.

Analysis of the batch supernatant and inclusion body pastes, diluted 50% (v/w) with water was conducted to determine their total protein content and whether any of the soluble protein was free Long-R³IGF-I. The total protein concentration in the batch supernatant was $1.4 \pm 0.05 \text{ g L}^{-1}$, but peaks corresponding to Long-R³IGF-I in either the native or denatured conformations could not be conclusively identified in the HPLC profiles (see Figure

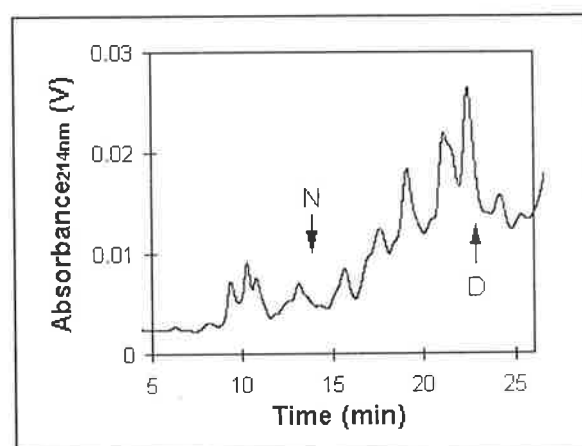


Figure 2.9: HPLC profile of batch supernatant. The expected elution times for native and denatured protein are indicated by an N and D, respectively.

2.9). Surprisingly, quantification of the total soluble protein for both clean and dirty inclusion body pastes revealed that they contained a higher concentration of soluble protein than the batch supernatant. This indicated that either the freeze thawing procedure, or addition of water released additional soluble protein from the inclusion bodies. If no protein were to be released, the maximum possible concentration would be that of batch supernatant (1.4 g L^{-1}). Instead, the total concentrations were measured to be

approximately 10 g L^{-1} and 5 g L^{-1} for clean and dirty inclusion bodies respectively. HPLC analysis revealed that many protein species could be resolved (see Figure 2.10).

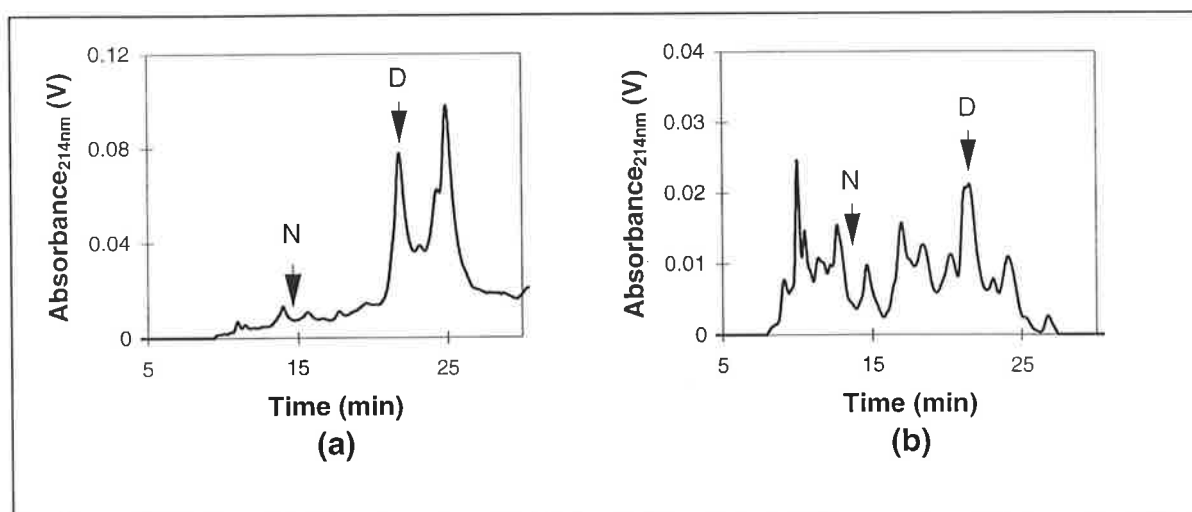


Figure 2.10: HPLC elution profiles for approximately equal loadings of the supernatant recovered from clean (a) and dirty (b) inclusion body pastes when mixed with water. The expected elution times for native and denatured protein are indicated by an N and D, respectively.

The profile of eluting proteins recovered from both inclusion body mixtures appeared to be free from native-conformation protein, but a significant protein peak with an elution time identical to that of reduced, denatured LongR³IGF-I was resolved. The total amount of protein eluted from the clean inclusion bodies was greater than that from the dirty inclusion bodies, but less than 1% of that released when a corresponding HPLC was analysed after exposure to dissolution buffer. Less protein was eluted from the dirty inclusion bodies, with the initial portion of the profile appearing quite similar to that of the batch supernatant, indicating that the inclusion body paste had significant amounts of supernatant associated with it. The total amount of protein eluting from the dirty inclusion body sample was approximately 2% of that eluting from a similar sample taken after long-term (> 1 hour) exposure to dissolution buffer.

Referring to Figure 2.8, many species could be seen that were present in the supernatant, clean pellet, and batch pellet. Presumably, some of these species had become dissociated from the clean inclusion bodies during thawing and incubation. The overall intensity of the sum of bands from cleaned inclusion bodies was greater than for the dirty inclusion bodies, indicating that the concentration of protein was greater in the triple-centrifuged material.

The dominance of late-eluting species in the HPLC profile also indicated that the inclusion bodies had been purified to some extent. The presence of more water-soluble protein in the clean inclusion bodies was surprising, but was confirmed both by total protein quantitation and HPLC analysis. Perhaps the greater concentration of inclusion body product may account for the greater protein release from clean inclusion bodies than from the dirty mixture, if the action of freezing and thawing then incubation in water was sufficient to release some of the misfolded inclusion body protein.

2.6 DISSOLUTION OF INCLUSION BODIES

A standard dissolution procedure was routinely used to provide uniform starting material for refolding experiments. The chemical composition of the dissolution buffer was the same as that used by Milner (1996). Fresh dissolution buffer without reducing agent (8.8 M urea/0.11 M Tris-HCl pH 9.1/ 0.044 M Gly/ 0.011 M EDTA) was prepared within 24 hours of being required. To reduce major aggregates of inclusion bodies in the paste preparations, thus significantly reducing the amount of material that could be solubilised, the thawed pastes were first diluted with water to a concentration of 1.00 mL g⁻¹. The suspensions were then thoroughly mixed for at least 10 minutes using a 1.0 cm flea on a magnetic stirrer to give a homogeneous slurry. Immediately before dissolution, freshly prepared 1.60 M DTT stock was added to the dissolution buffer to a final concentration of 0.016 M. Dissolution was initiated by addition of 10 mL of dissolution buffer per gram of inclusion body paste. This effectively gave a dissolution buffer composition of 8 M urea/0.1 M Tris-HCl pH 9.1/ 0.04 M Gly/ 0.01 M EDTA/0.016 M DTT. Thorough mixing was maintained using a 1.0 cm flea on a magnetic stirrer. Samples for refolding trials were taken 30 minutes after initiating dissolution.

2.6.1 Chemical limitation of dissolution

To determine whether total dissolution was achieved with 10 mL g⁻¹ of dissolution buffer, a 110 µL sub-sample was diluted 1:4 into fresh dissolution buffer. The diluted mixture was incubated at room temperature for a further 60 minutes with stirring. 100 µL was quenched for HPLC analysis by 1:10 dilution into 1.0% TFA and triplicate 50 µL samples were quenched into 10 mL of 0.1 M NaCl/0.1% TFA for total protein analysis

3. MATHEMATICAL FRAMEWORK

Mathematical simulation of large-scale processing is a useful tool for identifying significant unit interactions. Models that truly describe real-world performance are necessary for meaningful findings.

In this chapter models of protein refolding kinetics are presented and selection of a likely generic mechanism discussed. A framework of reactor design equations and cost-capacity equations is presented to allow simulation of a typical production process that involves product expression as inclusion bodies. A method for comparing overall process performance based on a dimensionless annual operating cost function is also outlined. This simulation method removes the need to identify specific unit interactions, while identifying the optimum refolding strategy from their sum.

A summary of the nomenclature is included in Appendix 9.1.

3.1 KINETICS OF REFOLDING

For monomeric proteins, the transition from a denatured state to the correct native conformation *in vitro*, is generally a uni-molecular series of sequential reactions via intermediate species (see section 1.5.1). Rate-limiting steps such as proline *cis-trans* isomerisation or intramolecular disulfide bond formation result in transient accumulation of intermediate species which can be observed using kinetic experiments (Creighton, 1986; Kim and Baldwin, 1982). The observed kinetic rates are dependent on the thermodynamic difference of the protein in denatured and native states (Brems, 1988; Cleland and Wang, 1993; Creighton, 1987; Dill, 1987; Kiefhaber and Schmid, 1992a, 1992b; Kuwajima, 1989; Sharma *et al.*, 1990; Utiyama and Baldwin, 1986). Therefore, while kinetic pathways are reported, the corresponding estimates of rate constants for the rate of transition between states are generally not specified.

3.1.1 Kinetic models of correct refolding

The aim of kinetic modelling is to find the simplest mechanism that fits the observed data without being too specific about the structure of intermediates (Kim and Baldwin, 1982; Utiyama and Baldwin, 1986). Structure of protein in intermediate states has been discussed in section 1.3 and a discussion of refolding mechanisms is presented in section 1.5.

Many models have been proposed to describe the observed kinetic behaviour of correct refolding. The models are restricted to describing correct refolding because they have been based on kinetic studies conducted using extremely dilute solutions of highly purified protein. Under these conditions, intermolecular interactions are minimal.

One of the earliest kinetic models of folding to be dismissed was the exhaustive random search of conformations considered by Levinthal (1968; see section 1.3). However, when a biased random search model is considered, by restriction to conformations that are sterically possible, some plausible thermodynamic support of the model can be justified (Baldwin, 1994).

Two other kinetic models that have been proposed are the nucleation-growth model and diffusion-collision-adhesion model. The basis of the nucleation-growth model is that once a nucleation reaction occurs, refolding proceeds rapidly to the native state. Evidence of populated intermediate states has discounted this model (Kuwayama, 1989). The basis of the diffusion-collision-adhesion model is that small domains of the protein transiently refold, but are not stabilised unless they interact with other similarly structured portions. The rate at which the structured regions interact, and hence the rate of refolding, is proposed to be limited by diffusion. However, the observation that solvent viscosity does not effect the overall rate of refolding for RNase A has discounted this model (Kim and Baldwin, 1982).

The kinetic model that has not been discounted is the sequential folding model which proposes that folding occurs through a defined sequence of intermediate states. Perhaps the reason that this model has not been discounted is that it is so general: it makes no claims to describe the structure of the refolding intermediates. There is ample evidence of populated intermediate states for many proteins to support its application. The structure of these intermediates varies extensively: from pre-molten-globule (Ptitsyn, 1995) to a double-disulfide stabilised state (Hober *et al.*, 1992). While some refolding experiments have not detected intermediates, it is likely that this is because they failed to look for transient states that form rapidly during the initial stages of refolding (Utiyama and Baldwin, 1986). Kim and Baldwin (1990) have proposed that sequential folding is likely to be the general kinetic scheme for all proteins.

For the purposes of designing efficient refolding processes, general understanding of the real-world refolding pathway is required (Cleland and Wang, 1990b). On a process scale, correct refolding must generally compete with off-pathway aggregation (see section 1.5.2). In other words, reactor design requires a kinetic description that accounts for refolding at high concentration where aggregation is a significant factor. The sequential refolding model lends itself to accounting for aggregation by including the possibility of association between intermediate species along the pathway of correct refolding.

3.1.2 Rate of correct refolding

During one of the early investigations of kinetic aspects of protein refolding, Teipel and Koshland (1971a & 1971b) observed that a significant fraction of the total structural transformation during refolding by dilution occurred within the first 30 seconds, but reactivation continued for 50 to 200 minutes. They concluded that the gradual regain of activity was due to minor structural rearrangements following rapid gross structural changes within the first minute of exposure to conditions favouring formation of native structure. Kuwajima *et al.* (1989) observed that the refolding pathway of α -lactalbumin could be described by rapid collapse to a molten globule state within milliseconds, followed by gradual organisation of the specific tertiary structure via a first-order reaction within the timescale of minutes. Similar behaviour of additional proteins has led to the conclusion that the molten globule may be a common, rapidly formed (<0.2 s), transient intermediate in the refolding of globular proteins (Jaenicke, 1991; Kim and Baldwin, 1990; Kuwajima, 1989; Kuwajima *et al.*, 1989; Ptitsyn *et al.*, 1990). Garel (1992) concluded that proteins generally collapse into a molten globule state within approximately one second and then take from minutes to hours to shuffle into the native state. More recently, the refolding by rapid dilution of HEWL has been observed to result in an initial extremely rapid collapse followed by formation of the secondary structure within 50-200 milliseconds and the gradual development of the tertiary structure (Fisher *et al.*, 1993b). The refolding intermediates were shown to be quite hydrophobic relative to the native state, supporting the molten globule intermediate theory.

The rate of accumulation of native-conformation monomeric protein from denatured starting material has been observed to obey first-order kinetics for many proteins (Goldberg *et al.*, 1991; Jaenicke and Rudolph, 1986; Kiefhaber *et al.*, 1990a; Mitraki *et al.*, 1987; Semisotnov *et al.*, 1987; Valax and Georgiou, 1991; Zettlmeißl, 1979). This indicates that there must be a rate-limiting step in the pathway that obeys first-order kinetics following initial formation of a molten globule intermediate. This is consistent with refolding being a unimolecular reaction that requires only intramolecular rearrangements to reach the native state. In proteins where proline isomerisation gives rise to multiple unfolded states, the pathway can be described by parallel first-order reactions, with collapsed molten globule intermediates starting from each of the proline isomers (Grafl *et al.*, 1986; Kiefhaber *et al.*, 1990b; Schmid, 1986; Schmid *et al.*, 1993).

3.1.3 Models that account for the aggregation reaction

In most cases, association of intermediates during refolding has been reported to result in irreversible aggregation which is known to occur between intermediates during the transition to the native state (see section 1.5.2). Kinetic investigations of aggregation have revealed that this unfavourable association occurs during the early events of refolding.

Teipel and Koshland (1971b) proposed a complex equilibrium pathway that included three intermediate states, one of which was the precursor to formation of aggregates. Zettlmeißl *et al.* (1979) concluded that aggregation of lactic dehydrogenase proceeds via a reaction of order 2.5. They were the first to propose a kinetic scheme for competing correct refolding and aggregation reactions from a common starting state (Figure 3.1(a)). Later, Jaenicke and Rudolph (1986) and Jaenicke (1987, 1989) supported the quantitative description of competing first- and higher than first-order refolding reactions. Mitraki *et al.* (1987) placed an upper limit on the order of the aggregation reaction in their investigation of PGK and concluding that the order of the rate-limiting step in the pathway was greater than one, but less than ten. They proposed the kinetic scheme shown in Figure 3.1(b) where aggregation is preceded by a rapid first-order equilibrium to an off-pathway intermediate (I_2).

The refolding mechanism illustrated in Figure 3.1(c) was proposed by Mulkerin and Wetzel (1989) to describe the thermal denaturation of γ -interferon. Cleland and Wang (1990a) identified that the first significant intermediate that appears in the refolding of bovine carbonic anhydrase B (CAB) is protein configured in the molten globule state. The molten globule intermediate associates irreversibly to form oligomers and they proposed that it is likely to be the common mechanism that leads to aggregation of other globular proteins. They determined that the rate of aggregation could be described by a reaction of order 2.6, with a rate constant that was a strong function of denaturant concentration and proposed the reaction pathway illustrated in Figure 3.1(d). Dimerisation was identified as the rate limiting step in the pathway leading to formation of aggregates (Cleland and Wang, 1991).

Hagen *et al.* (1990a) stated that in 1989 the model illustrated in Figure 3.1(e) was generally accepted for all proteins. The order of the aggregation reaction was not discussed, but they concluded that since it was a multimolecular reaction, it was concentration dependent.

Further support of the model was provided when the aggregation of citrate synthase was observed to occur prior to the rate-limiting events in folding (Buchner *et al.*, 1991). Goldberg *et al.* (1991) observed that the aggregation of HEWL occurred within 5 seconds of transfer into renaturation buffer, and suggested that this may be due to irreversible association of intermediates in the molten globule state.

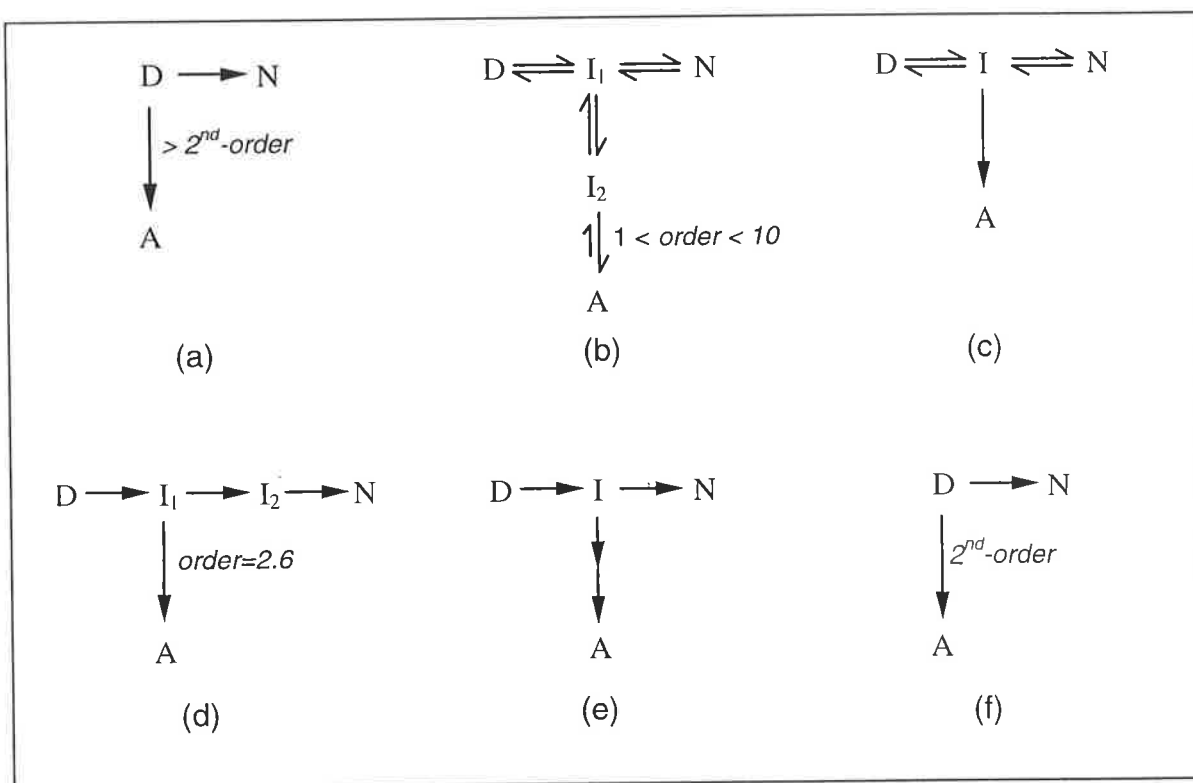


Figure 3.1: D = denatured state, I = intermediate state, N = native state, and A = inactive, aggregated state. (a) competing first- and greater than second-order pathway proposed by Zettlmeißl *et al.* (1979) (b) Kinetic scheme proposed by Mitraki *et al.* (1987) (c) kinetic scheme proposed by Mulkerin and Wetzel (1989) (d) refolding pathway proposed for refolding of CAB (Cleland and Wang, 1990a; 1990b) (e) general refolding model proposed by Hagen *et al.* (1990) (f) simplest model of competing first- and second-order reactions (Kiefhaber *et al.*, 1991; Rudolph 1990). In all models it is either stated or implied that the reactions along the pathway of correct refolding are first-order.

Vicik and De Bernardes-Clark (1991) proposed a model for the refolding of CAB that had the same structure as that depicted in Figure 3.1(b), but assumed that the off-pathway reactions were irreversible. They determined that the order of the aggregation reaction was 2.51 ± 0.28 . That is, in cases where the order of the aggregation reaction was determined, it was found to be approximately 2.5. However, second-order kinetics have been shown to

give an adequate description of the rate of aggregation for LDH that was observed by Zettlmeißl (1979) (Kiefhaber, *et al.*, 1991; Rudolph, 1990). This leads to the simplest useful model of refolding, depicted in Figure 3.1(f). A second-order reaction for aggregation is consistent with dimerisation being the limiting reaction.

Notable exceptions to irreversible aggregation are the transient association of early folding intermediates in the refolding of bovine CAB and growth hormone. Cleland and Wang (1992) have observed that monomers of CAB in a molten globule intermediate state form a stable equilibrium with a soluble dimerised state, and a second-order rate constant can be fitted to describe the rate of dimerisation (in contrast to their earlier investigations above; Cleland and Wang 1990a & 1991). Brems *et al.* (1987) observed that bovine growth hormone can form a transient associated intermediate under partially denaturing conditions (3.8M GuHCl). However, a year later it was reported that the transiently associated conformation was the rate-limiting precursor resulting in irreversible aggregation (Brems, 1988).

3.1.4 General kinetic model for reactor design

The ideal kinetic scheme for process simulation of refolding would be the least complicated reaction pathway that accurately describes the observed behaviour. Following from the discussion of correct refolding models presented in section 3.1.1, the simplest form of kinetic sequential refolding model that describes the typically observed behaviour is a three-state model given in Equation 3.1.



In this model, denatured protein proceeds rapidly to a molten globule state and then to the native conformation via two first-order reactions. The reactions are assumed to be irreversible, which may only be valid under conditions that highly favouring native folding. This model is only a good description of refolding conducted at extremely low protein concentration where association of monomers may be assumed to be negligible. It must be extended to account for aggregation as illustrated in Figure 3.2, to provide a real-world description.

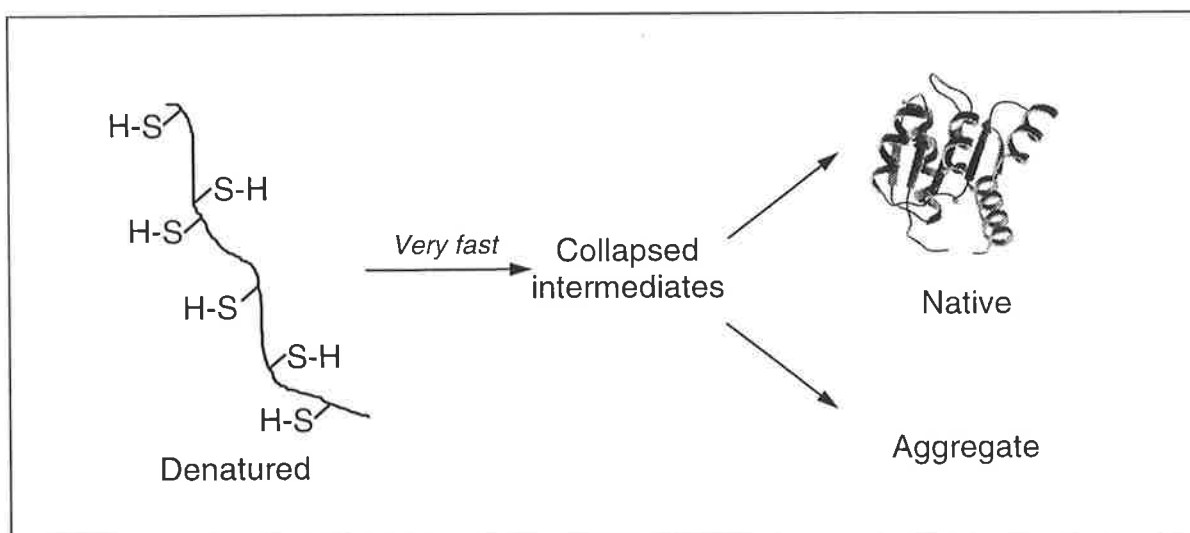


Figure 3.2: Schematic depicting the generally observed behaviour during refolding at moderate to high protein concentration.

The simplest model that allows for an aggregation pathway, yet retains a sound mechanistic basis, adds a competing irreversible aggregation reaction to the scheme represented by Equation 3.1, as shown in Figure 3.3(a). This kinetic scheme is consistent with that proposed by Hagen *et al.* (1990a), illustrated in Figure 3.1(e). However, it does not account for a loss of product once it has correctly refolded, either by proteolysis or equilibrium isomerisation (see section 1.5.2).

One way to provide a product sink is to make the kinetic assumption that including a back-reaction from native to intermediate state can account for losses from the native state. If the rate of aggregation is assumed to be limited by dimerisation, a simple kinetic pathway for refolding can be formulated (see Figure 3.3(b)). This simplified kinetic model has been used throughout this thesis. Since collapse to the molten globule state occurs within the first few milliseconds whereas activity is restored over a period of minutes or hours, the initial collapse to intermediate state is taken to be instantaneous. This makes the model mathematically consistent with that used by Keifhaber *et al.* (1991), with the addition of a pathway for product loss.

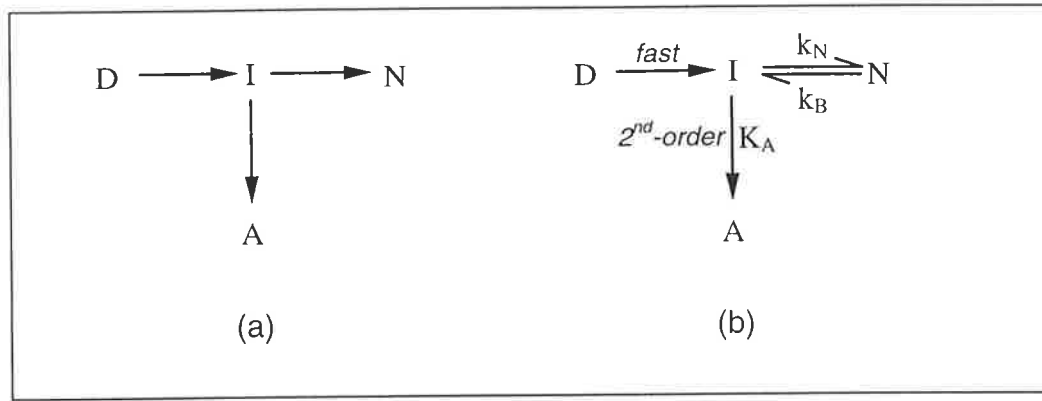


Figure 3.3: (a) The most basic sequential refolding model, modified to include possible aggregation (b) This is the refolding model proposed for use in this study; k_N and k_B are first-order rate constants of equilibrium refolding and K_A is the apparent second-order rate constant of aggregation.

3.1.5 Rate constants

As mentioned above, the rate of refolding is dependent on the thermodynamic difference between the denatured and native states. However, estimates of rate constants are required to use the simple kinetic scheme in simulation studies of refolding.

LongR³IGF-I has been used as the model protein in the present study. Rate constants for refolding fairly impure LongR³IGF-I were determined during preliminary research prior to commencement of this study (Kotlarski *et al.*, 1995; see section 1.7). The observed accumulation of native species during a series of refolding trials was regressed to the simple competing first- and second-order scheme proposed by Kiefhaber *et al.* (1991) and Rudolph (1990) (see Figure 3.1(f)). The rate constants for formation of native species and aggregate species were found to be $k_N = 8.5 \times 10^{-4} \text{ s}^{-1}$ and $K_A = 0.0544 \text{ m}^3 \text{ mol}^{-1} \text{ s}^{-1}$, respectively. These values are in close agreement with reported rate constants of other proteins.

Cleland and Wang (1990a) have reported that for bovine CAB the limiting first-order reaction to achieve the native state had a rate constant of $1.2 \times 10^{-3} \text{ s}^{-1}$ (based on the work of Semisotnov *et al.*, 1987) and aggregation could be described as a second-order reaction with a rate constant in the range 0.03 to $0.3 \text{ m}^3 \text{ mol}^{-1} \text{ s}^{-1}$. It should be noted that they concluded that the range of second-order rate constants could be correlated with the concentration of chaotropic agent present during refolding, consistent with the reaction rate being dependent on the thermodynamic driving force. In later work, Cleland and Wang

(1992) concluded that a transient dimer was generated during the *in vitro* refolding of CAB. They calculated that the limiting first-order rate constant was again $1.2 \times 10^{-3} \text{ s}^{-1}$, but the rate constant to give the equilibrium dimer was approximately $2.3 \text{ m}^3 \text{ mol}^{-1} \text{ s}^{-1}$. In his overview of protein folding Garel (1992) presented first-order rate constants of $4.5 \times 10^{-4} \text{ s}^{-1}$ and $6.0 \times 10^{-4} \text{ s}^{-1}$ for formation of native monomers of UDP-glucose dehydrogenase and AK-homoserine dehydrogenase, respectively. These rate constants compare quite favourably with that for LongR³IGF-I. However, the rate constants presented in the same publication, for correct dimerisation of the two enzymes were 16 and $70 \text{ m}^3 \text{ mol}^{-1} \text{ s}^{-1}$. They are orders of magnitude greater than the regressed constant for Long-R³-IGF-I, but are similar to the value of $100 \text{ m}^3 \text{ mol}^{-1} \text{ s}^{-1}$ for lactic dehydrogenase calculated by Kiefhaber *et al.* (1991), based on the experimental data of Zettlmeißl *et al.* (1979). The corresponding first-order rate constant of correct refolding was calculated to be 0.01 s^{-1} (Kiefhaber *et al.*, 1991). These examples emphasise that the rate constants vary for different proteins and chemical environments and must be determined on a case-by-case basis for a particular refolding scheme.

3.1.6 Recent kinetic studies

Concurrent with the work presented in this thesis, significant support of the generic kinetic scheme proposed in section 3.1.3 has been published. Recently, a number of simplified kinetic schemes have been proposed to describe competing correct refolding and unproductive aggregation reactions during refolding.

Chaudhuri *et al.* (1996) confirmed that the correct refolding of HEWL obeys first-order kinetics and observed that extensive aggregation occurred rapidly after initiating refolding. They assumed a second-order irreversible aggregation reaction to simulate batch and fed-batch refolding as two competing reactions (see Figure 3.4(a)). Matthiesen *et al.* (1996) proposed the simple kinetic scheme presented in Figure 3.4(b) based on the model of Kiefhaber *et al.* (1991). In this model, denatured protein proceeds reversibly along a sequential first-order pathway to the native state, with only one significant intermediate. Aggregation is initiated from the intermediate via a reaction that is usually second-order (or higher). Middelberg (1996b) used the kinetic scheme shown in Figure 3.4(c) to simulate continuous refolding.

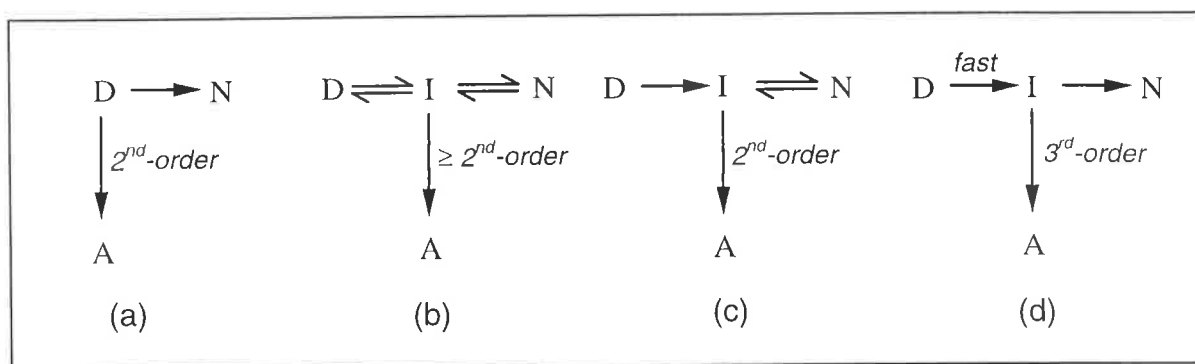


Figure 3.4: (a) Kinetic scheme used by Chaudhuri *et al.* (1996) to simulate the refolding of HEWL (b) kinetic scheme proposed by Matthiesen *et al.* (1996) (c) kinetic scheme used in simulations by Middelberg (1996b) (d) kinetic scheme used by Hevehan and De Bernardez-Clark (1997) and Maachupalli-Reddy *et al.* (1997).

In his review on protein refolding, Chaudhuri (1994a) concluded that refolding obeyed first-order kinetics and that the irreversible aggregation reaction is second-order or higher. Ptitsyn (1995) reiterated the earlier proposition that the molten globule is a common early transient intermediate (< 0.2 s) in the folding of all globular proteins (Ptitsyn *et al.*, 1990; Jaenicke, 1991), but added that a pre-molten globule intermediate (< 20 ms) is also likely to be a common intermediate. Jaenicke (1995) and Rudolph and Lilie (1996) have reported that the basis for aggregation is postulated to be hydrophobic interaction of transient intermediates early in the refolding pathway. They concluded that refolding can be described as a second-order (or higher) reaction, with aggregation competing with first-order refolding. Creighton *et al.* (1996) reiterated that the rate of refolding is generally dependent only on the final refolding conditions. They also concluded that all of the molecules with subpopulations of denatured states (*e.g.* *cis/trans* isomers) refold with the same rate constant even though they have different conformations initially. The review of two-state refolding by Zwanzig (1997) concluded that this most basic model is only applicable when there are no significant thermodynamic barriers between particular unfolded states; behaviour that is not commonly observed.

In experimental investigations, Ballew *et al.* (1996) observed that the earliest steps of folding apomyoglobin into a compact state are complete in less than 20 μ s and occur due to gross structural collapse. During the refolding dead-time of manual mixing, unfolded monomers of PGK were observed to collapse and associate (Pecorari *et al.*, 1996). A slow phase followed in which the oligomers dissociated and produced active protein

($k \approx 4.6 \times 10^{-3} \text{ s}^{-1}$). This behaviour is similar to that of CAB reported by Cleland and Wang (1992). Speed *et al.* (1996) demonstrated that aggregation of P22 tailspike and coat monomers only occurred from an intermediate state. It was also demonstrated that aggregates of P22 tailspike peptides were not in equilibrium with the intermediate that initiated dimerisation and large aggregates can be formed by direct association of multimers, rather than interacting directly with monomers (Speed *et al.*, 1997). Hevehan and De Bernardez-Clark (1997) assumed that collapse of denatured protein into an intermediate state was instantaneous and concluded that the refolding of HEWL was best described by competing first- and third-order reactions. They correlated the rate constants for each reaction with the concentration of denaturant (see Figure 3.4(d)). The same kinetic pathway was used by Maachupalli-Reddy *et al.* (1997) to investigate refolding of HEWL in the presence of controlled amounts of contaminants. The first-order rate constant for refolding was found to be approximately $7.5 \times 10^{-3} \text{ s}^{-1}$, independent of contamination.

3.2 REACTOR DESIGN EQUATIONS

A basic mole balance of species j for a reaction vessel is written:

$$\left\{ \begin{array}{c} \text{Moles } j \\ \text{entering} \end{array} \right\} + \left\{ \begin{array}{c} \text{Moles } j \\ \text{generated} \end{array} \right\} = \left\{ \begin{array}{c} \text{Moles } j \\ \text{exiting} \end{array} \right\} + \left\{ \begin{array}{c} \text{Moles } j \\ \text{accumulating} \end{array} \right\}$$

With reference to Figure 3.5 the general reactor design equation is based on the basic mole balance:

$$Q_{IN} n_j + \int_V r_j = Q_{OUT} n_j + \frac{d}{dt} N_j \quad (3.2)$$

where r_j is the rate of formation of j . In the case of refolding, the species of interest is native-conformation protein, originating from denatured starting material. If it is assumed that the denatured species reacts very quickly to form an intermediate species (see section 3.1.3), this reaction product is the starting material for native protein formation.

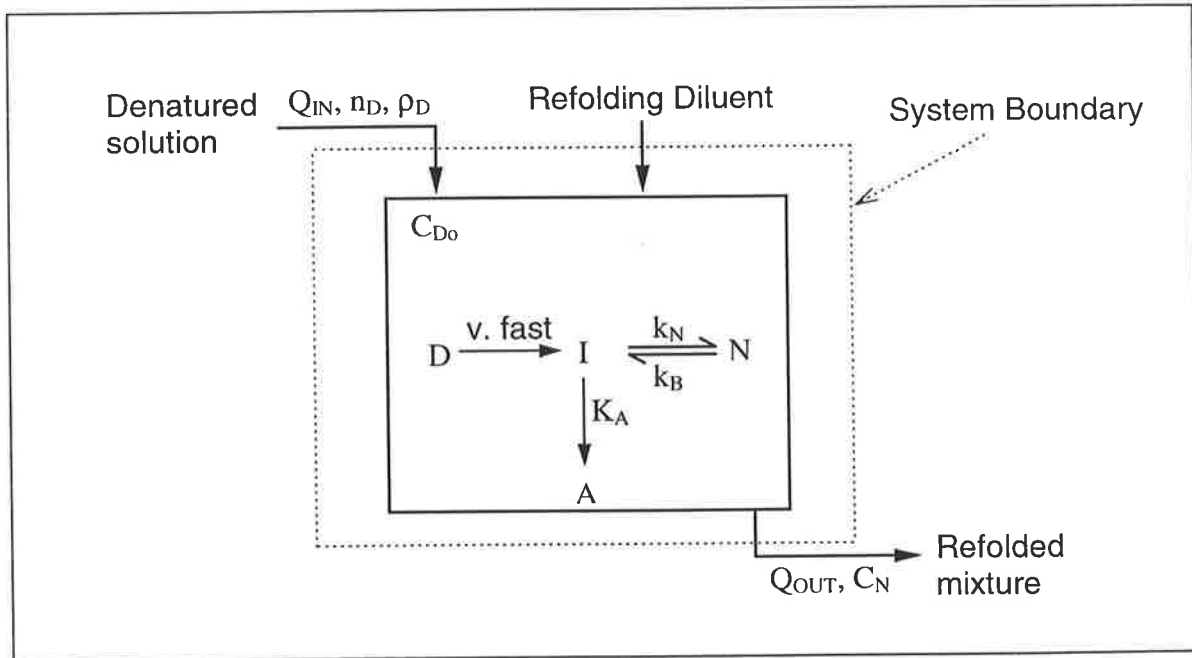


Figure 3.5: General refolding reactor with addition of denatured material, diluent solution and product removal.

3.2.1 Batch refolding

For a batch reactor the inlet and outlet terms in the general mole balance are zero. For an ideal reactor, the reaction rate is constant through the vessel and the general design equation reduces to:

$$\frac{d}{dt} N_j = V \cdot r_j \quad \text{or} \quad \frac{d}{dt} C_j = r_j \quad (3.3)$$

Using Equation 3.3, the rates of formation of the key species (I and N) in the simplified kinetic scheme described in section 3.1.4 and illustrated in Figure 3.3(b) can be written:

$$\frac{d}{dt} C_I = -k_N C_I - K_A C_I^2 + k_B C_N \quad (3.4)$$

$$\frac{d}{dt} C_N = k_N C_I - k_B C_N \quad (3.5)$$

Equations 3.4 and 3.5 can be written in dimensionless form by dividing by the initial concentration of denatured protein (C_{D0}) and expressing the concentration of species I and N in terms of conversion and yield respectively, and using a dimensionless time. The conversion of protein, X , can be defined as:

$$X = \frac{C_{D0} - C_I}{C_{D0}} \quad (3.6)$$

and the yield of native conformation protein can be defined relative to the amount of initially denatured protein in the batch as:

$$y_N = \frac{C_N}{C_{D0}} \quad (3.7)$$

A suitable dimensionless batch refolding time (τ) can be defined as the product of the first-order rate constant of native protein formation and reaction time. That is,

$$\tau = k_N \cdot t \quad (3.8)$$

Hence Equations 3.4 and 3.5 become

$$\frac{d}{d\tau} X = 1 - X + \frac{K_A C_{D0}}{k_N} \cdot (1 - X)^2 - \frac{k_B}{k_N} y_N \quad (3.9)$$

$$\frac{d}{d\tau} y_N = 1 - X - \frac{k_B}{k_N} y_N \quad (3.10)$$

The initial conditions for Equations 3.9 and 3.10 are that conversion and yield are both zero. Note that the term $K_A C_{D0} / k_N$ can be considered of as a dimensionless measure of refolding concentration and k_B/k_N is effectively an equilibrium constant for the formation of native protein from intermediate species. The two dimensionless differential equations must be solved numerically to obtain a solution. Refer to Appendix 9.2.1 for a detailed derivation of the dimensionless batch design equations.

3.2.2 Fed-batch refolding

For a fed-batch reactor, the inlet stream must be included from the general design equation. For an ideally mixed fed-batch reactor the design equation for species j is written as:

$$\frac{d}{dt} N_j = Q_{IN} \cdot n_j + r_j \cdot V \quad (3.11)$$

where $Q_{IN} \cdot n_j$ is the rate of species j being fed into a reactor of instantaneous volume V . If the same reaction rates as used in Equations 3.4 and 3.5 are used in Equation 3.11, again assuming that initially denatured protein is effectively already an intermediate, two design equations can be written:

$$\frac{d}{dt} N_I = Q_{IN} \cdot n_D - k_N N_I + k_B N_N - K_A \frac{N_I^2}{V} \quad (3.12)$$

$$\frac{d}{dt} N_N = k_N N_I - k_B N_N \quad (3.13)$$

A third design equation is required to account for the increase in volume (V) with time. If it is assumed that the dissolved protein must be *diluted four-fold* with refolding buffer to give a suitable chemical environment for refolding the equation is written:

$$\frac{d}{dt} V = 4Q_{IN} \quad (3.14)$$

Unlike the case of batch operation, the amount of denatured protein present in the reaction vessel will vary with time. Setting the system boundary for the refolding reaction to include the dissolution reactor, the assumption that denatured protein is instantly folded to active intermediate can be used to define a conversion of intermediate species as:

$$X = \frac{\tau}{\tau_{\text{Refolding}}} - \frac{N_I}{N_{D0}} \quad (3.15)$$

where τ is the dimensionless time of refolding and $\tau_{\text{Refolding}}$ is the dimensionless duration of feeding. For fed-batch operation, the yield much be defined as the ratio of amounts of native species to total initial denatured species.

$$y_N = \frac{N_N}{N_{D0}} \quad (3.16)$$

To allow direct comparison of fed-batch and batch refolding the initial charge to the refolding reactor should be such that the nominal protein concentration is the same for both processes. That is, it is necessary to design the final volume of the fed-batch reactor to be the required production capacity divided by the nominal initial concentration of denatured protein. To achieve this the initial charge will be the final volume less the volume that is fed during refolding

$$V_o = \frac{N_{D0} M_N}{C_{D0}} - \frac{4N_{D0}}{n_D} \quad (3.17)$$

where $\dot{m}_N t$ is the mass produced per batch. Dimensionless design equations can be determined using Equations 3.8, 3.15, 3.16, and 3.17 on Equations 3.12 to 3.14 to give:

$$\frac{d}{d\tau} X = -\frac{k_B}{k_N} y_N + \left(\frac{\tau}{\tau_{\text{Refolding}}} - X \right) + \frac{K_A C_{D0}}{k_N \left(1 - \frac{4C_{D0} M_N}{n_D} \right) \left(\frac{V}{V_o} \right)} \cdot \left(\frac{\tau}{\tau_{\text{Refolding}}} - X \right)^2 \quad (3.18)$$

$$\frac{d}{d\tau} y_N = \frac{\tau}{\tau_{\text{Refolding}}} - X - \frac{k_B}{k_N} y_N \quad (3.19)$$

$$\frac{d}{d\tau} \frac{V}{V_o} = \frac{4}{\tau_{\text{Refolding}} \left(\frac{M_N n_D}{C_{D0}} - 4 \right)} \quad (3.20)$$

The initial conditions for Equations 3.18, 3.19, and 3.20 are that conversion and yield are both zero and the initial volume ratio is one. These dimensionless differential equations must be solved numerically to obtain a solution. A detailed derivation of the dimensionless design equations for fed-batch refolding are presented in Appendix 9.2.2.

3.2.3 Continuous refolding

The design equations for continuous refolding are greatly simplified because steady-state operation discounts the need to integrate design equations. If it is assumed that the reaction rate is independent of the position in the reactor (ideally mixed), the design equations are

$$C_I = C_{D0} - (k_N C_I + K_A C_I^2 - k_B C_N) \cdot t \quad (3.21)$$

$$C_N = (k_N C_I - k_B C_N) \cdot t \quad (3.22)$$

where t is the residence time in the refolding CSTR. Equations 3.6, 3.7, and 3.8 can be used to express Equations 3.21 and 3.22 in dimensionless form as:

$$X = \tau \left[(1 - X) + \frac{K_A C_{D0}}{k_N} (1 - X)^2 - \frac{k_B}{k_N} y_N \right] \quad (3.23)$$

$$y_N = \frac{1 - X}{\frac{1}{\tau} + \frac{k_B}{k_N}} \quad (3.24)$$

Substitution of Equation 3.24 into Equation 3.23 can be simplified to give

$$\tau^2 (Da_{BN} Da_{AN} (1 - X)^2) + \tau (1 - X + Da_{AN} (1 - X)^2 - Da_{BN} X) - X = 0 \quad (3.25)$$

where Da_{BN} is the ratio of Damköhler numbers for the reverse and forward reactions to form native conformation protein from intermediate and Da_{AN} is the ratio of Damköhler numbers for the aggregation reaction and the correct refolding reaction. That is,

$$Da_{BN} = \frac{k_B t}{k_N t} \quad Da_{AN} = \frac{K_A C_{D0} t}{k_N t}$$

Da_{BN} is equivalent to an equilibrium constant for native protein formation and Da_{AN} may be viewed as the dimensionless refolding concentration. Equation 3.25 defines a quadratic expression for τ that may be solved for any given set of Da_{BN} , Da_{AN} , and X . The solution can then be substituted into Equation 3.24 to determine the yield for that set of parameters. Derivation of Equations 3.24 and 3.25 are presented in Appendix 9.2.3.

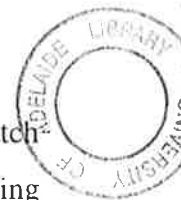
3.3 GENERIC LARGE-SCALE PROCESS DESCRIPTION

The processes considered are intended to be applicable for industrial-scale production of small monomeric proteins expressed as intracellular inclusion bodies within *E. coli* (see Figure 1.1). They include the operations preceding chromatographic purification of the isolated native conformation product. Three processes which have identical fed-batch fermentation, batch disruption and semi-continuous centrifugation operations are considered. The processes differ in the mode of operation of the steps following centrifugation. Dissolution, refolding and ultrafiltration are operated in either batch, fed-batch or continuous modes to give the three slightly different processes (Figure 3.6). Specific process variables (*e.g.* medium composition) are given in the parameter selection sections below (sections 3.7).

Sterile medium and compressed air are fed to a single production fermenter. The production fermenter is inoculated from a seed fermenter with recombinant *E. coli* containing a plasmid coding a specific protein product. At some point during growth, expression of the product is induced. After completion of fermentation, the broth is transferred to one of two jacketed holding vessels for disruption. Multiple discrete passes through a high-pressure homogeniser are employed to achieve the desired level of disruption. The homogenate is cooled between passes to prevent irreversible thermal denaturation of the product. Once sufficiently disrupted, the homogenate is transferred to a third jacketed holding vessel. Differential sedimentation, using a continuous centrifuge, selectively concentrates inclusion bodies from the homogenate. The inclusion body stream is dissolved as either a single batch in a stirred tank or continuously in a plug flow reactor (PFR), depending on the method used for refolding the protein.

3.3.1 Batch refolding

If batch refolding is used, the inclusion bodies are dissolved in a single stirred, batch reactor. The slurry of inclusion bodies is rapidly and thoroughly mixed with a buffered solution of highly concentrated chaotropic agent. The redox potential must be controlled to maintain cysteine residues in reduced form, preventing potential covalent intermolecular



association. Refolding is initiated by rapidly diluting of the dissolved material into a batch reactor containing buffered chaotropic agent at a reduced concentration, under oxidising conditions. The refolding reaction is allowed to proceed for a specific length of time and then quenched. Ultrafiltration follows to reduce the batch volume for subsequent chromatographic purification (see Figure 3.6(a)).

3.3.2 Fed-Batch refolding

The inclusion body stream is collected in a reactor, as for batch refolding, and dissolved in a single batch by mixing with a highly concentrated chaotropic agent under reducing conditions. From the batch, a stream of the dissolved protein is rapidly mixed with a suitable buffer to reduce the concentration of chaotropic agent and change from a reducing to an oxidising environment. The mixture is then diluted further into a stirred tank reactor charged with a buffered chaotropic agent at a moderate concentration. The volume of the initial charge determines the effective protein concentration. The rate of feeding determines the duration of the refolding reaction. The reaction is quenched once the entire batch of dissolved protein has been fed into the reactor. As for batch refolding the solution containing biologically active protein is concentrated by ultrafiltration prior to chromatographic purification (see Figure 3.6 (b)).

3.3.3 Continuous refolding

When continuous operation is employed, the inclusion body rich stream from the centrifuge is continuously mixed with dissolution buffer (same as that used for batch dissolution), but the reaction proceeds in a PFR. The product stream from this reactor is vigorously mixed with a suitable refolding buffer to dilute the protein into a solution with the same chemical composition as that employed for batch and fed-batch refolding. The refold mixture then passes into a CSTR. Product flow from this reactor is concentrated continuously by ultrafiltration. Biologically active product can then be isolated from the retentate using chromatographic separation (see Figure 3.6(c)).

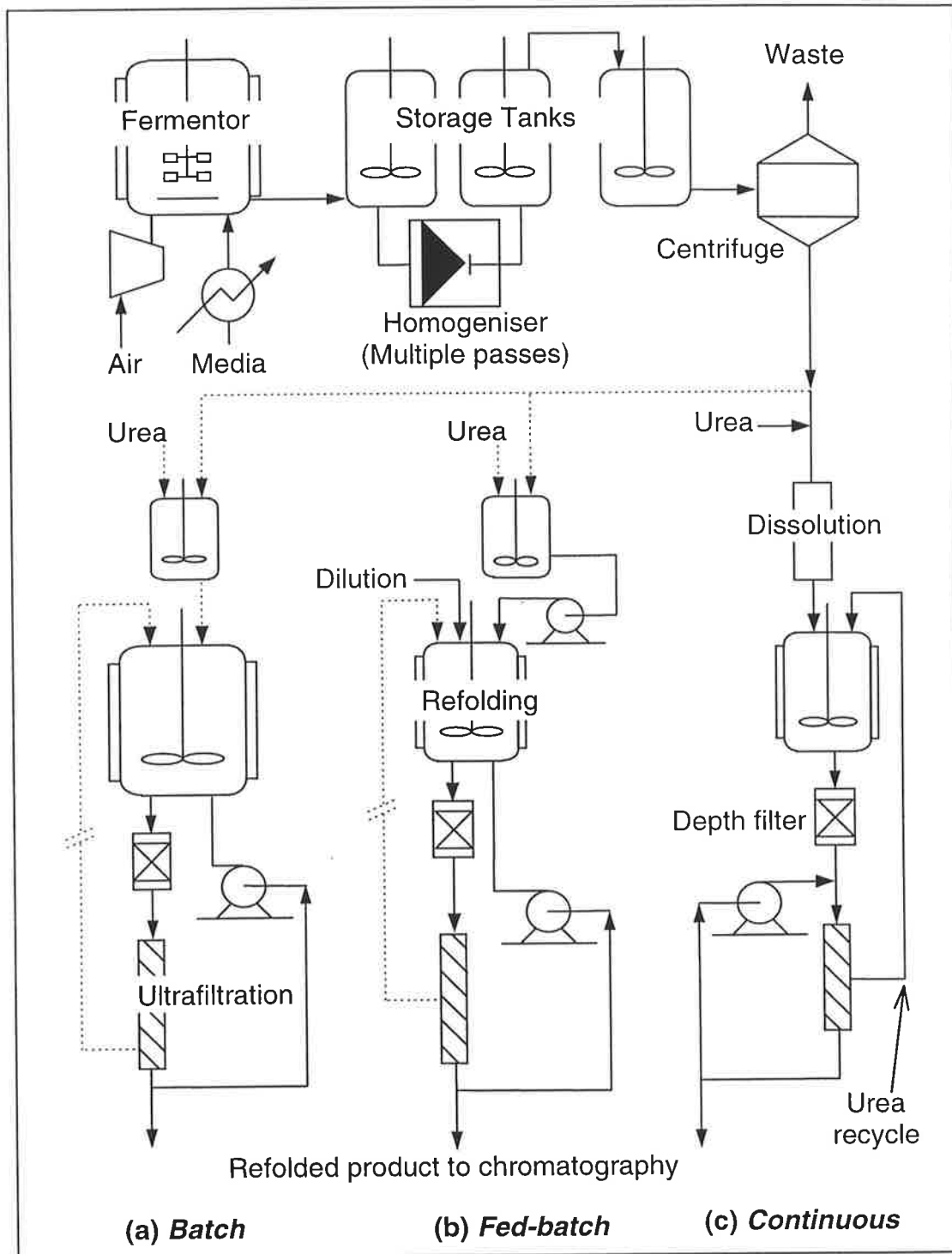


Figure 3.6: Three flowsheets for production processes to isolate and renature protein expressed as inclusion bodies in *E. coli*. (a) Batch operation of dissolution, refolding and ultrafiltration steps. (b) Fed-batch operation of refolding with batch dissolution and ultrafiltration steps. (c) Continuous operation of dissolution, refolding, and ultrafiltration steps. Dashed lines represent batch or fed-batch operation. Ideal mixing is assumed in all stirred tank reactors and plug-flow in the dissolution reactor used for continuous operation.

3.4 DESIGN BASIS

The processes are designed for a recombinant protein product containing intramolecular disulfide bonds, biologically potent as a monomer. It is assumed that operation will be 24 h per day for 270 fermentation batches per year. The capacity of the process is sufficient to produce $100 \times 10^{-6} \text{ kg s}^{-1}$ of crude, biologically active, recombinant protein. This equates to a total annual production of 2300 kg, without allowance for losses during chromatographic purification. Operation of the production fermenter is assumed to be the critical path of the system, with a cycle time of 24 h (Petrides *et al.*, 1989 & 1995). The dissolution, refolding and ultrafiltration operations are designed to be completed within one cycle time. Refolding is assumed to be quenched immediately after completion. In the case of batch refolding the entire batch is quenched at a specific time. For fed-batch refolding, the batch is quenched as soon as the feeding is complete and for continuous refolding, the product stream leaving the reactor is continuously quenched.

The capacity of processing equipment is assumed to exactly meet required demand making allowance for processing inefficiencies. Product losses are assumed to occur only during the differential sedimentation and refolding operations. Ullage of 25% of total vessel volume has been included when sizing tanks and stirred reactors.

Costing of seed ferments has not been included. While they would be necessary to provide an adequate inoculum for the production fermenter, their capacity and hence contribution to overall cost should be insensitive to variation in fermentation requirements. A maximum aeration rate of 1 v.v.m. to the fermenter has been assumed. This is consistent with both published simulation (Middelberg, 1996b; Petrides *et al.*, 1989) and experimental work (Turner *et al.*, 1994; Yoon *et al.*, 1994a & 1994b) involving expression of heterologous protein within *E. coli*.

3.5 OBJECTIVE FUNCTION

The main focus of this work was to examine the impact of different refolding systems on overall process economics. The performance indicator selected to compare between different schemes was a dimensionless operating cost function (Middelberg, 1996b). Annual operating cost was selected, rather than total plant cost, as this gives a more distinct reflection of the effects of process modification. In the initial design stages, calculation of total plant costs have to rely on basic cost-multiplication factors, based on equipment purchase costs, to arrive at a final value (Atkinson and Mavituna, 1991; Kalk and Langlykke, 1986; Perry and Green, 1984; Peters and Timmerhaus, 1990). The effect that this has is to weight the direct-fixed-capital costs relative to operating costs.

The dimensionless operating cost function used is the sum of three components, normalised by a characteristic cost. The annual direct-fixed-capital (DFC) dependent costs such as expenses for maintenance, depreciation, insurance, local taxes and factory expenses are collected as a single term (ϕ_{dfc}). The annual cost of materials consumed during operation is another term (ϕ_{con}). The total annual waste treatment cost (ϕ_{wt}) is the third component. The operating costs for labor, administration, site maintenance, instrumentation, process piping, and insulation have not been calculated. They should be relatively insensitive to the changes in operational capacity considered about the optimum and can therefore be omitted without compromising any observed trends. Since the costs of utilities represent only a small fraction of the annual operating costs (Petrides, 1990, Petrides *et al.*, 1995), they have not been included in the objective function. Hence, the objective function is:

$$\phi^* = \frac{\phi_{dfc} + \phi_{con} + \phi_{wt}}{\phi_{ch}} = \phi_{dfc}^* + \phi_{con}^* + \phi_{wt}^* \quad (3.26)$$

where ϕ_{ch} is a characteristic cost and the superscript indicates the dimensionless annual cost component. For the purpose of the simulations conducted in this study, the characteristic cost has been taken to be the purchase cost of a particular fermenter. The volume of this fermenter is specified by the design basis for mass of product produced per batch, without allowance for loss during processing. That is, the capacity of the fermenter is sufficient to meet the design basis assuming all product is recovered during downstream processing and total renaturation occurs during the refolding operation.

3.6 PROCESS COSTING

Evaluation of the three components of the objective cost function were made using a method presented by Middelberg (1996b). The direct-fixed-capital-dependent costs were calculated by using a simple multiplying factor, ζ , and the total sum of equipment purchase costs, ϕ_{ep} (i.e. $\phi_{dfc} = \zeta \cdot \phi_{ep}$, Petrides *et al.*, 1989, 1995). Equipment purchase costs were calculated using continuous cost-capacity functions. Cost-capacity functions assume that cost can be scaled according to operating capacity. Material usage was calculated for a batch of product and the total cost determined using conservative estimates for the significant consumables. Two process streams were identified as contributing virtually all of the waste: supernatant from centrifugation of the homogenate and permeate from ultrafiltration following refolding. Therefore, the disposal cost of process waste was reduced to that applicable to only two streams.

Note that the method used to calculate the cost components is only suitable for obtaining an appreciation of trends. It is consistent with a *factored estimate* approach (Perry and Green, 1984) which is used for the purpose of feasibility studies to identify the significant trends. The probable error is up to approximately 30 percent.

3.6.1 Unit purchase costs

Cost-capacity equations used to calculate estimates for equipment purchase costs are outlined below. The mass rate of production of native protein, $\dot{m}_N$, is determined by selection of a design basis. Yield of protein successfully refolded to give its native conformation, y_N , is a function of the refolding conditions as outlined in section 1.5.

3.6.1.1 Fermenter

The volume of culture that must be produced to supply starting material sufficient to achieve the design basis depends on the concentration of desired product in the broth at termination of the fermentation and the efficiency of recovery of that protein. The

concentration can be expressed as the product of the dry cell weight at termination, W , the total cell-protein-to-dry-cell-weight ratio, T , and the expression level, E . Therefore, if it is assumed that the efficiency of recovery (excluding refolding) is f , the fermenter volume may be expressed as:

$$V_f = \frac{\dot{m}_N t}{ETW\varepsilon} \cdot \frac{1}{f y_N} \quad (3.27)$$

where t is the batch cycle time and ε is the design margin to allow sufficient ullage. The first term in Equation 3.27 gives the required fermenter volume if total recovery of product is achieved. The product $f y_N$ makes allowance for loss of desired product and is effectively the overall fractional recovery prior to chromatographic purification.

If β_f is the cost-capacity exponent for a fermenter, Equation 3.27 can be used to calculate its cost by:

$$\phi_f = 10^{\alpha_f} V_f^{\beta_f} = 10^{\alpha_f} \left(\frac{\dot{m}_N t}{ETW\varepsilon f y_N} \right)^{\beta_f} \quad (3.28)$$

As discussed above, the characteristic cost selected is that for a fermenter assuming zero product loss during processing. This can be expressed as:

$$\phi_{ch} = 10^{\alpha_f} \left(\frac{\dot{m}_N t}{ETW\varepsilon} \right)^{\beta_f} \quad (3.29)$$

3.6.1.2 Steriliser

Fermentation medium is passed through a continuous steriliser before entering the fermenter itself. The required flow rate, Q_s , will be the volume of broth at termination of fermentation, divided by the allowed cycle time, t_s . At termination, the volume of broth will be the fermenter volume, multiplied by the design margin to allow sufficient ullage. Hence, the cost-capacity equation for a suitable continuous steriliser will be:

$$\phi_s = 10^{\alpha_s} Q_s^{\beta_s} = 10^{\alpha_s} \left(\frac{\dot{m}_N t}{ETW f y_N t_s} \right)^{\beta_s} \quad (3.30)$$

3.6.1.3 Compressor

To estimate the cost of a suitable compressor the air flow rate, Q_a , must be known. For a flow rate of δ air volumes per fermenter-volume per minute (v.v.m.), a cost capacity equation can be expressed as:

$$\phi_a = 10^{\alpha_a} Q_a^{\beta_a} = 10^{\alpha_a} \left(\frac{\dot{m}_N t}{ETW f y_N} \cdot \frac{\delta}{60} \right)^{\beta_a} \quad (3.31)$$

3.6.1.4 Homogeniser

If the time available for homogenisation is t_h and the number of discrete passes of the broth through the homogeniser is N_p , the required capacity, Q_h , can be written as:

$$Q_h = \frac{N_p \varepsilon V_f}{t_h} \quad (3.32)$$

Therefore, a cost-capacity equation for a homogeniser can be written as:

$$\phi_h = 10^{\alpha_h} Q_h^{\beta_h} = 10^{\alpha_h} \left(\frac{\dot{m}_N t}{ETW f y_N} \cdot \frac{N_p}{t_h} \right)^{\beta_h} \quad (3.33)$$

3.6.1.5 Centrifuge

A useful measure of centrifuge capacity is the total settling capacity or sigma factor, Σ . If a normalised centrifuge feed rate, Q_Σ , that achieves a suitable collection efficiency of inclusion bodies and cell debris is known, an acceptable feed rate to the centrifuge, Q_c , can be determined. That is,

$$Q_c = Q_\Sigma \cdot \Sigma \quad (3.34)$$

Inclusion bodies must be recovered from the total broth volume at termination. Therefore, the time available to achieve this separation, t_c , specifies the centrifuge feed rate and the sigma factor can be expressed in terms of known values. Substitution into a cost-capacity equation gives:

$$\phi_c = 10^{\alpha_c} \Sigma^{\beta_c} = 10^{\alpha_c} \left(\frac{\dot{m}_N t}{ETWfy_N} \cdot \frac{1}{t_c Q_\Sigma} \right)^{\beta_c} \quad (3.35)$$

3.6.1.6 Storage tanks

The volume of each storage tank shown in Figure 3.6 must be sufficient to contain the volume of broth at termination. The tanks must be the same volume as the fermenter itself. Equation 3.36 states the cost-capacity equation for such vessels.

$$\phi_t = 10^{\alpha_t} V_f^{\beta_t} = 10^{\alpha_t} \left(\frac{\dot{m}_N t}{ETW\epsilon fy_N} \right)^{\beta_t} \quad (3.36)$$

3.6.1.7 Dissolution reactor

Dissolution is conducted at a constant protein concentration. Therefore, the volume of the dissolution reactor will be directly proportional to the amount of inclusion body concentrate being dissolved. Inclusion body slurry is discharged at a fixed concentration from the centrifuge, specified by the operating conditions. Hence, dissolution reactor volume can be expressed in terms of the selected dissolution concentration, ρ_D . Dissolution is conducted for a fixed time. For a continuous system, a plug flow reactor (PFR) with a fixed residence time, t_d , would be suitable. Volume of a PFR is related to the residence time by the volumetric flow rate, v_d , which can be calculated from the required production rate of product and ρ_D . That is,

$$V_d = v_d t_d = \frac{\dot{m}_N}{y_N M_N n_D} \cdot t_d = \frac{\dot{m}_N}{y_N \rho_D} \cdot t_d \quad (3.37)$$

where M_N is the molecular weight of the native protein. When operation in batch mode is employed, the required dissolution reactor volume can be calculated as:

$$V_d = \frac{\dot{m}_N t}{y_N M_N n_D \epsilon} = \frac{\dot{m}_N t}{y_N \rho_D \epsilon} \quad (3.38)$$

Therefore the cost-capacity equation for the dissolution reactor will be:

$$\phi_d = 10^{\alpha_d} V_d^{\beta_d} = 10^{\alpha_d} \left(\frac{\dot{m}_N T'}{\rho_D y_N} \right)^{\beta_d} \quad (3.39)$$

where T' is t_d for a PFR and t/ϵ for a single batch dissolution.

3.6.1.8 Refolding reactors

Determination of the capacity of the reactor required for refolding depends on which of the three refolding schemes is employed. It is assumed that the cost-capacity parameters are the same for all three reactors.

1. Continuous refolding

For continuous refolding a CSTR is employed. The design parameter of interest is volumetric flow rate, v_o . The inlet concentration of denatured protein after dilution to refolding conditions, C_{D0} , will determine the volumetric flow rate for particular production rate. If the residence time for the refolding reactor is t_r , the volume and hence cost-capacity equations will be given by

$$V_r = \frac{v_o t_r}{\epsilon} = \frac{\dot{m}_N}{y_N M_N C_{D0} \epsilon} \cdot t_r \quad (3.40)$$

and

$$\phi_r = 10^{\alpha_r} V_r^{\beta_r} = 10^{\alpha_r} \left(\frac{\dot{m}_N t_r}{C_{D0} M_N y_N \epsilon} \right)^{\beta_r} \quad (3.41)$$

2. Fed-batch refolding

Refolding in a fed-batch manner will result in the volume and yield varying throughout the period of feeding. Final yield of native conformation protein and effective overall protein concentration will determine the necessary volume for a fed-batch refolding reactor according to:

$$V_r = \frac{\dot{m}_N t}{y_N M_N C_{D0} \epsilon} \quad (3.42)$$

Cost of the reactor can be calculated in the same manner as for the continuous refolding reactor:

$$\phi_r = 10^{\alpha_r} V_r^{\beta_r} = 10^{\alpha_r} \left(\frac{\dot{m}_N t}{C_{D0} M_N y_N \epsilon} \right)^{\beta_r} \quad (3.43)$$

3. Batch refolding

As for refolding conducted in a fed-batch manner, the yield of native conformation protein will be time-dependent. A suitable reactor volume is specified by the protein concentration at initiation and the yield of correctly renatured protein at termination of refolding. Algebraically, calculation of the volume and cost is identical to the fed-batch case:

$$\phi_r = 10^{\alpha_r} V_r^{\beta_r} = 10^{\alpha_r} \left(\frac{\dot{m}_N t}{C_{D0} M_N y_N \epsilon} \right)^{\beta_r} \quad (3.44)$$

Numerically, the final yield and hence reactor volume will vary between the two cases for a particular overall protein concentration.

3.6.1.9 Ultrafiltration

Ultrafiltration is used to concentrate the renatured protein and provide a stream of urea for recycle. The mode and operating conditions employed for refolding will determine the ultrafiltration capacity required. When operating as part of a continuous process, the membrane area must be sufficient to accommodate the flow rate of native protein from the

PFR, v_o . For batch and fed-batch operation, the required capacity will depend on the time available to reach a suitable concentration as well as the volume to be processed.

1. Continuous refolding

To calculate the size of an appropriate ultrafiltration unit, the flux and flow rate through the membrane must be known. It is assumed that the flow rate of concentrate is insignificant compared to the inlet flow rate of dilute native protein. Also, there is a linear relationship between membrane area, A and flux, J . Therefore, the flow rate through the membrane is v_o . The area can be calculated from:

$$A_u = \frac{v_o}{J} \quad (3.45)$$

Using the expression for v_o from Equation 3.40, the cost-capacity equation for an ultrafiltration unit, scaled using area is:

$$\phi_u = 10^{\alpha_u} A_u^{\beta_u} = 10^{\alpha_u} \left(\frac{\dot{m}_N}{C_{Do} M_{NYN} J} \right)^{\beta_u} \quad (3.46)$$

2. Batch and fed-batch refolding

When batch or fed-batch processes are used for refolding, subsequent concentration by ultrafiltration will have to be done in a single batch. Available time and batch volume will determine the required flow rate and hence membrane area. If the time available for refolding and ultrafiltration is t_{ru} , the required membrane area can be calculated by:

$$A_u = \left(\frac{\dot{m}_N t}{C_{Do} M_{NYN} J} \right) \left(\frac{1}{t_{ru} - t_r} \right) \quad (3.47)$$

where t_r is the time taken for the refolding operation.

Substitution of Equation 3.47 into Equation 3.46 gives:

$$\phi_u = 10^{\alpha_u} A_u^{\beta_u} = 10^{\alpha_u} \left(\frac{\dot{m}_N t}{C_{Do} M_{NYN} J (t_{ru} - t_r)} \right)^{\beta_u} \quad (3.48)$$

3.6.1.10 Total Direct-Fixed-Capital-dependent costs (DFC costs)

The total DFC cost for a particular process is the sum of purchase costs for each unit of equipment multiplied by the number of units of that type. The dimensionless DFC cost is this sum divided by the characteristic cost given by Equation 3.29. As stated previously, the total annual DFC-dependent operating cost can be estimated using a multiplying factor with the equipment purchase cost. That is,

$$\phi_{dfc}^* = \zeta \frac{\phi_{ep}}{\phi_{ch}} = \zeta \frac{(\phi_f + \phi_s + \phi_a + \phi_h + \phi_c + n_t \phi_t + \phi_d + \phi_r + \phi_u)}{\phi_{ch}} \quad (3.49)$$

where n_t is the number of storage tanks.

3.6.2 Cost of consumables

Only materials that change in usage with different operating conditions are considered for the purpose of cost estimation. For example, sterilising filters for air supplied to the fermenter have not been considered in the analysis because their capacity should not change significantly irrespective of the conditions used. However, the surface area of depth filter necessary for post-refolding filtration is included as the batch volume will vary. Cost of consumables is calculated for a single batch and the total annual cost then scaled accordingly. Usage is determined from the volume estimates outlined above and cost determined using estimates for the cost per characteristic unit.

3.6.2.1 Fermentation medium

Consumption of fermentation medium can be estimated from the fermenter volume using Equation 3.28 without making allowance for ullage. If the specific cost of the medium is γ_m US dollars per cubic metre, the batch cost will be

$$\hat{\phi}_m = \epsilon V_f \gamma_m = \frac{\dot{m}_N t}{ETW f y_N} \cdot \gamma_m \quad (3.50)$$

3.6.2.2 Urea

The process considered uses urea for dissolution and refolding of the inclusion body protein. Fresh urea is used for the dissolution step, but urea recovered from the ultrafiltration permeate is added for the refolding step. It is assumed that the amount of urea that must be added for refolding is significant compared to the amount required for dissolution. This assumption should be valid as the protein concentration for refolding is typically orders of magnitude less than for dissolution, while the urea concentration required for the two stages differs by less than one order. It is also assumed that the cost of recovering urea is economically attractive. The mass of urea purchased for the dissolution buffer will be

$$m_{ud} = M_u C_{ud} V_{ud} = \frac{M_u C_{ud} \dot{m}_N t}{\rho_D \gamma_N} \quad (3.51)$$

where M_u is the molecular weight of urea and C_{ud} is the urea concentration required for dissolution. Therefore, the cost of urea purchase per batch is

$$\hat{\phi}_{ud} = m_{ud} \gamma_{up} = \frac{M_u C_{ud} \dot{m}_N t}{\rho_D \gamma_N} \cdot \gamma_{up} \quad (3.52)$$

The amount of urea that must be recycled can be determined directly from a material balance.

$$m_{urec} = m_{ur} - m_{ud} \quad (3.53)$$

where m_{ur} is the mass of urea required for refolding during one cycle time (t) or for a single batch.

$$m_{urec} = \frac{M_u C_{ur} \dot{m}_N t}{C_{D0} M_N \gamma_N} - \frac{M_u C_{ud} \dot{m}_N t}{\rho_D \gamma_N} \quad (3.54)$$

where C_{ur} is the concentration of urea used for refolding. Hence, the cost of recycling urea is

$$\hat{\phi}_{urec} = m_{urec} \gamma_{urec} = \frac{\dot{m}_N t}{\gamma_N} \left(\frac{M_u C_{ur}}{C_{D0} M_N} - \frac{M_u C_{ud}}{\rho_D} \right) \cdot \gamma_{urec} \quad (3.55)$$

3.6.2.3 Additives for dissolution and refolding

Additives must be included as part of the dissolution and refolding mixtures to control the chemical environment. Suitable conditions for the additives must be determined during process development. Additives typically include redox agents, such as dithiothreitol, 2-mercaptoethanol or glutathione (Fischer *et al.*, 1993a), chelating agents, amino acids, stabilising organic compounds (e.g. PEG or Triton), buffering salts, and protease inhibitors. The cost for each mixture per cubic metre can be estimated as the sum of costs of the additives. Using these values, the cost of additives per batch can be calculated from:

$$\hat{\phi}_{ad} = V_d \gamma_{ad} = \frac{\dot{m}_N t}{C_{DYN}} \cdot \gamma_{ad} \quad (3.56)$$

$$\hat{\phi}_{ar} = V_r \gamma_{ar} = \frac{\dot{m}_N t}{C_{Do} M_{NYN}} \cdot \gamma_{ar} \quad (3.57)$$

3.6.2.4 Depth filter

Following refolding, any remaining solid material such as cellular debris and undissolved inclusion bodies is typically removed using a dead-end filter. While the capital cost of depth filtration is small and has been ignored, since disposable filters are often employed due to rapid fouling, there is a significant operating cost (Petrides *et al.*, 1989). The approximate surface area of filtration membrane per unit volume of feed solution, $\hat{A}_{df}$, must be known to estimate process requirements. A value should be available from pilot scale trials. Using an estimate for the cost of depth filter per unit area, γ_{df} , the total cost of depth filter can be calculated as:

$$\hat{\phi}_{df} = V_r \hat{A}_{df} \gamma_{df} = \frac{\dot{m}_N t}{C_{Do} M_{NYN}} \cdot \hat{A}_{df} \cdot \gamma_{df} \quad (3.58)$$

3.6.2.5 Ultrafiltration membrane

The total cost of ultrafiltration membrane per batch will depend on the number of times that it can be reused before replacement is required as well as the actual area of membrane

used. If the life of the ultrafiltration membrane is L cycles, the cost per cycle for a continuous process will be given by:

$$\hat{\phi}_u = \frac{A}{L} \gamma_u = \frac{\dot{m}_N}{C_{Do} M_N y_N J L} \cdot \gamma_u \quad (3.59)$$

For a process that employs batch or fed-batch refolding, the cost of ultrafiltration membrane will be given by:

$$\hat{\phi}_u = \frac{A}{L} \gamma_u = \frac{\dot{m}_N t}{C_{Do} M_N y_N J (t_{ru} - t_r) L} \cdot \gamma_u \quad (3.60)$$

3.6.2.6 Total cost of process consumables

The total cost of consumables for a single cycle time or batch is the sum of all seven components. The total annual cost is the sum of components multiplied by the number of batches produced per annum, N_b . To non-dimensionalise the total annual cost for consumables it is divided by a characteristic cost as outlined previously. Hence,

$$\phi_{con}^* = N_b \frac{\hat{\phi}_{con}}{\phi_{ch}} = N_b \frac{(\hat{\phi}_m + \hat{\phi}_{ud} + \hat{\phi}_{urec} + \hat{\phi}_{ad} + \hat{\phi}_{ar} + \hat{\phi}_{df} + \hat{\phi}_u)}{\phi_{ch}} \quad (3.61)$$

3.6.3 Waste management

Volume and biological oxygen demand (BOD) components of waste treatment were also considered. As stated previously two process streams contribute all of the waste for the section of process considered. The supernatant from centrifugation of the homogenate has a significant BOD due to the high level of non-product material produced during fermentation. Permeate from ultrafiltration following refolding typically creates the greatest volume of waste, because of the low protein concentrations required for folding to occur properly. However, the bulk of the permeate is recycled for the refolding step. Therefore, the volume of waste generated is the same as the volume of the dissolution step. Since urea is used during refolding there will be a high BOD associated with this stream which will dominate the volume-associated cost.

The cost of treating the waste supernatant from the centrifugation step expressed in terms of fermentation parameters and specific costs is:

$$\hat{\phi}_{wt,c} = V_f \gamma_{wtcv} + V_f (1 - E) TW \gamma_{wtcb} = \frac{\dot{m}_N t}{fETW y_N} (\gamma_{wtcv} + (1 - E) TW \gamma_{wtcb}) \quad (3.62)$$

where γ_{wtcv} and γ_{wtcb} are the specific volume-associated cost ($\$ m^{-3}$) and the BOD-related cost ($\$ (kg \text{ BOD})^{-1}$) respectively.

The mass of waste urea associated with permeate from the ultrafiltration step equals the mass of fresh urea added for the dissolution, m_{ud} . Using Equation 3.52, the waste treatment cost per batch from the ultrafiltration step is:

$$\hat{\phi}_{wt,u} = V_d \gamma_{wtcv} + m_{ud} \gamma_{wtcb} = \frac{\dot{m}_N t}{\rho_D y_N} \gamma_{wtcv} + \frac{M_N C_{ud} \dot{m}_N t}{\rho_D y_N} \gamma_{wtcb} \quad (3.63)$$

Assuming that there is no other significant waste that must be dealt with, the total cost of waste treatment per batch is given by Equation 3.64 (the sum of Equations 3.62 and 3.63).

$$\hat{\phi}_{wt} = \frac{\dot{m}_N t}{y_N} \left(\frac{\gamma_{wtcv} + (1 - E) TW \gamma_{wtcb}}{fETW} + \frac{\gamma_{wtcv}}{\rho_D} + \frac{M_u C_{ud}}{\rho_D} \gamma_{wtcb} \right) \quad (3.64)$$

In the same manner as for DFC-dependent and consumable cost components, the non-dimensional annual cost for waste management is

$$\phi_{wt}^* = N_b \frac{\hat{\phi}_{wt}}{\phi_{ch}} \quad (3.65)$$

The equations used to estimate the capacity of each unit and the cost per batch of the process consumables and waste treatment are summarised in Table 3.1.

Table 3.1: Summary of the capacity equations for unit sizing and equations of specific cost per batch of consumable usage and waste treatment.

Unit	Capacity Expression
Air compressor	$Q_a = \frac{\dot{m}_N t}{ETW \epsilon y_N} \cdot \frac{\delta}{60}$
Fermenter	$V_f = \frac{\dot{m}_N t}{ETW \epsilon} \cdot \frac{1}{fy_N}$
Sterilizer	$Q_s = \frac{\dot{m}_N t}{ETW fy_N t_s}$
Homogenizer	$Q_h = \frac{\dot{m}_N t}{ETW fy_N} \cdot \frac{N_p}{t_h}$
Storage tank	$V_t = \frac{\dot{m}_N t}{ETW \epsilon} \cdot \frac{1}{fy_N}$
Centrifuge	$\Sigma = \frac{\dot{m}_N t}{ETW fy_N} \cdot \frac{1}{t_c Q_\Sigma}$
Dissolution reactor	$V_d = \frac{\dot{m}_N T'}{\rho_D y_N}$ where $T' = t_d$ or t/ϵ
Refolding reactor	$V_r = \frac{\dot{m}_N T}{C_{Do} M_N y_N \epsilon}$ where $T' = t_r$ or t
Ultrafiltration unit	$A_u = \frac{\dot{m}_N t}{C_{Do} M_N y_N J (t_{ru} - t_r)}$ for fed-/batch $A_u = \frac{\dot{m}_N}{C_{Do} M_N y_N J}$ for continuous
Fermentation media	$\hat{\phi}_m = \frac{\dot{m}_N t}{ETW fy_N} \cdot \gamma_m$
Urea	$\hat{\phi}_{ud} = \frac{M_u C_{ud} \dot{m}_N t}{\rho_D y_N} \cdot \gamma_{up}$
Urea recycle	$\hat{\phi}_{urec} = \frac{\dot{m}_N t}{y_N} \left(\frac{M_u C_{ur}}{C_{Do} M_N} - \frac{M_u C_{ud}}{\rho_D} \right) \cdot \gamma_{urec}$
Dissolution additives	$\hat{\phi}_{ad} = \frac{\dot{m}_N t}{\rho_D y_N} \cdot \gamma_{ad}$
Refolding additives	$\hat{\phi}_{ar} = \frac{\dot{m}_N t}{C_{Do} M_N y_N} \cdot \gamma_{ar}$
Depth filter	$\hat{\phi}_{df} = \frac{\dot{m}_N t}{C_{Do} M_N y_N} \cdot \hat{A}_{df} \cdot \gamma_{df}$
Ultrafiltration membrane	$\hat{\phi}_u = \frac{\dot{m}_N}{C_{Do} M_N y_N J L} \cdot \gamma_u$ for continuous $\hat{\phi}_u = \frac{\dot{m}_N t}{C_{Do} M_N y_N J L \cdot (t_{ru} - t_u)} \cdot \gamma_u$ for fed-/batch
Waste Treatment	$\hat{\phi}_{wt} = \frac{\dot{m}_N t}{y_N} \left(\frac{\gamma_{wtcv} + (1-E)TW\gamma_{wtcb}}{fETW} + \frac{\gamma_{wtcv}}{\rho_D} + \frac{M_u C_{ud}}{\rho_D} \gamma_{wtcb} \right)$

3.7 SELECTION OF PROCESS PARAMETERS

To apply the costing strategy outlined above, values for all cost parameters and process variables must be known. Wherever possible values for cost parameters were collated from multiple literature sources and a suitable estimate selected. All parameters were determined for equipment fabricated from grade 316 stainless steel in terms of 1994 US\$, free on board (FOB). Marshall and Swift equipment cost indices (as published in *Chemical Engineering*) were used to convert costs into 1994 US\$ (Atkinson and Mavituna, 1991; Perry and Green, 1984). Preferential weighting was given to parameters gained directly from industrial sources and those published most recently. The selected values of the cost-capacity exponents and costing parameters are given in Table 3.2.

Two values of the parameters for fermenters are presented. The first values are for fermenters with volume between one and fifty cubic metres ($1 < V_f < 50 \text{ m}^3$). The second values should be used to estimate the cost of fermenters with capacities greater than fifty cubic metres ($V_f > 50 \text{ m}^3$). The discontinuity in cost-capacity equation reflects the need to erect large fermenters ($V_f > 50 \text{ m}^3$) on site rather than to install prefabricated units.

Estimations for specific costs of medium, consumables and waste disposal were based on published values. As for the equipment estimates, costs were evaluated in 1994 US\$. The producer prices index for industrial chemicals, as published in *Chemical Engineering*, was used to convert to the common base year. Aqueous media have a base cost of approximately \$50 per m^3 for high-purity water (Petrides *et al.*, 1995). Estimation of more significant specific costs is discussed below.

The additives for dissolution are a buffering couple and reducing agent (the chaotropic agent considered separately). Tris-HCl at a concentration of 0.1 mol L^{-1} is the most frequently reported buffering couple (Fischer *et al.*, 1993a). Dithiothreitol and 2-mercaptoethanol are commonly used reducing agents (Fischer *et al.*, 1993a), although 2-mercaptoethanol at 0.1 mol L^{-1} is the cheaper alternative. The additional costs of 0.1 mol L^{-1} Tris-HCl and 2-mercaptoethanol are approximately \$180 (Middelberg, 1996b) and \$10 (Petrides *et al.*, 1995) per m^3 of media, respectively. The sum of these gives a value for the dissolution additives of approximately \$250 m^{-3} .

Table 3.2: Summary of values for cost-capacity parameters with most relevant references. The exponents are for equipment fabricated in grade 316 stainless-steel, free on board.

Unit	Parameter	Reference
Air compressor	$\alpha_a=5.41, \beta_a=0.98$	Middelberg <i>et al.</i> , 1992; Peters and Timmerhaus, 1990.
Centrifuge	$\alpha_c=3.68, \beta_c=0.33$	Middelberg <i>et al.</i> , 1992
Dissolution reactor	$\alpha_d=4.41, \beta_d=0.59$ (assumed = α_t)	Design basis.
Fermenter	$\alpha_f=5.04$ or 4.66 $\beta_f=0.36$ or 0.70	Kalk and Langlykke, 1986.
Homogenizer	$\alpha_h=6.44, \beta_h=0.48$	Kalk and Langlykke, 1986; Middelberg <i>et al.</i> , 1992
Sterilizer	$\alpha_s=7.10, \beta_s=0.60$ (assumed)	Middelberg, 1996b.
Storage tank	$\alpha_t=4.41, \beta_t=0.59$	Petrides <i>et al.</i> , 1995.
Refolding reactor	$\alpha_r=4.41, \beta_r=0.59$ (assumed = α_t)	Petrides <i>et al.</i> , 1995.
Ultrafiltration unit	$\alpha_u=3.29, \beta_u=0.87$	Kalk and Langlykke, 1986
Dissolution additives	$\gamma_{ad}=250 \$ m^{-3}$	Petrides <i>et al.</i> , 1995.
Refolding additives	$\gamma_{ar}=120 \$ m^{-3}$	Middelberg, 1996b.
Depth filter	$\gamma_{df}=180 \$ m^{-2}$	Petrides <i>et al.</i> , 1989; Petrides <i>et al.</i> , 1995.
Fermentation medium	$\gamma_m=210 \$ m^{-3}$	Middelberg, 1996b.
Ultrafiltration membrane	$\gamma_u=200 \$ m^{-2}$	Kalk and Langlykke, 1986; Petrides <i>et al.</i> , 1995.
Urea	$\gamma_{up}=1.50 \$ kg^{-1}$	Petrides <i>et al.</i> , 1995.
Urea recycle	$\gamma_{urec}=0.75 \$ kg^{-1}$ ($=0.5\gamma_{up}$)	Design basis.
Waste Treatment	$\gamma_{wtcb}=1.00 \$ BOD^{-1}$ $\gamma_{wtcv}=0.20 \$ m^{-3}$	Petrides <i>et al.</i> , 1995.

It is assumed that the bulk of the medium required for refolding is provided by recycling the permeate from ultrafiltration post refolding. Therefore, the additives that are required for refolding need only provide a suitable oxidising environment. When refolding on a small-scale, it is common to use the method of mixed disulfides, which requires incubation of the dissolved protein in a solution containing oxidised glutathione or sodium sulfite/sodium tetrathionate prior to addition of a reducing agent (Fischer *et al.*, 1993a). While this method tends to give a greater refolding yield, it does have significant costs for reagents associated with it and should be avoided if possible. For example, a dilute oxidised glutathione redox couple at only 0.5 mM: 0.1 mM (reduced:oxidized) increases the cost of refolding additives by approximately $\$300 \text{ m}^{-3}$ (Middelberg, 1996b). Oxidation by atmospheric oxygen is the cheapest option in terms of reducing refolding additives, but the rate of oxidation is slow. The chemical 2-hydroxyethylthiol can provide suitable oxidising conditions, achieving relatively high rates at quite low concentrations (Kotlarski *et al.*, 1995). It has the advantage of moderate cost at approximately $\$120 \text{ m}^{-3}$ (Middelberg, 1996b). Therefore, assuming that a suitable refolding mixture can be obtained by addition of 2-hydroxyethylthiol to the recycled refolding buffer, a base cost of $\gamma_{\text{ur}} = \$120 \text{ m}^{-3}$ has been used for simulations.

A minimal fermentation medium, containing sufficient nutrients to support growth of *E. coli* to a high cell mass, could contain approximately 20 kg m^{-3} of salts, small amounts of additives such as vitamins and amino acids, and 100 kg m^{-3} glucose. Salts and glucose cost approximately $\$1.0 \text{ kg}^{-3}$ and $\$0.6 \text{ kg}^{-3}$ respectively (Petrides *et al.*, 1995). If it is assumed that the cost of additives is equal to the cost of salts, the approximate cost of fermentation medium is $\$150 \text{ m}^{-3}$. This value is in agreement with that published by Middelberg (1996b).

Guanidine hydrochloride and urea are both commonly used chaotropes for dissolution of inclusion bodies at the laboratory scale (see section 1.4). Accordingly, they have been used in processes that have been scaled-up to commercial-scale production (Bentle *et al.*, 1987; Builder *et al.*, 1986; Chang *et al.*, 1994; Datar *et al.*, 1993; Hershenson *et al.*, 1990). For the purpose of this investigation, the chaotropic agent used is urea. Suitable grade urea has a purchase cost of approximately $\$1.5 \text{ kg}^{-1}$ (Petrides *et al.*, 1995). It is assumed that the cost of recycling the urea in the permeate from the ultrafiltration step is half the purchase cost (Middelberg, 1996b) *i.e.* $\gamma_{\text{urec}} = \$0.75 \text{ kg}^{-1}$. This cost includes a component to

account for removal of urea decomposition products that may otherwise chemically modify the protein product (*e.g.* addition of glycine and/or purification by ion-exchange).

The cost of waste treatment, as outlined above, is divided into two components: volume- and BOD-associated costs. Petrides *et al.* (1995) give values for waste treatment or disposal of water as $\$2.0 \times 10^{-4} \text{ kg}^{-1}$ and $\$1.0 \text{ kg}^{-1}$ for most waste materials, including biomass and urea. Assuming that this cost of treatment or disposal of water is based on volume, the specific volume-associated cost will be approximately $\$0.20 \text{ m}^{-3}$. This value is in excellent agreement with that employed by Middelberg (1996b) of 1987 US\$ 0.17 m^{-3} . Therefore, the specific costs for waste treatment and/or disposal are $\gamma_{\text{wtcv}} = \$0.20 \text{ m}^{-3}$ and $\gamma_{\text{wtcb}} = \$1.0 \text{ kg}^{-1}$.

The final costing parameter that requires estimation is the multiplying factor to determine the total annual DFC-dependent operating cost. The value used in this work is $\zeta = 1.43$, as used by Petrides *et al.* (1995). This value agrees well with the value of 1.45 used in the earlier work of Petrides *et al.* (1989), indicating that there is still an extremely high DFC-dependent operating cost for bioprocessing. A summary of the specific cost estimates is presented in Table 3.2.

3.7.1 General process parameters

Values for process variables were either specified by the design basis, process, or again selected from published values. Table 3.3 summarises the typical parameters used for 'base' case simulations and indicates corresponding references. Discussion of selected parameters is given below.

Commercial viability of recombinant bioprocesses is obviously favoured by high expression levels. Cultures should attain at least 30-35 g of dry-cell-weight (d.c.w.) per litre of broth using atmospheric oxygen (Kalk and Langlykke, 1986; Middelberg *et al.*, 1992; Petrides *et al.*, 1989, 1995). The lower value in this range was used for 'base' case simulations (*i.e.* $W = 30 \text{ kg m}^{-3}$). Approximately 55% of the d.c.w. material should be protein (Neidhardt, 1987; Middelberg *et al.*, 1992; Petrides, 1990). The amount of inclusion body protein produced, relative to the amount of total cell protein, varies widely for different expression systems. Typically, the fraction of protein product will be

approximately 20% of total cell protein (Kiefhaber *et al.*, 1991; Middelberg *et al.*, 1992; Petrides *et al.*, 1989; Shire *et al.*, 1984). However, amounts of 10 % (Bowden *et al.*, 1991; Chang *et al.*, 1994), up to 40 % (Buchner and Rudolph, 1991; Kane and Hartley, 1991), have been reported. A value of $E = 20\%$ was assumed as the base case for simulations. High cell densities require a maximum rate of aeration of approximately 1.0 v.v.m. (Turner *et al.*, 1994; Yoon *et al.*, 1994a & 1994b). Therefore, the air compressor required to supply the production fermenter should be designed with $\delta = 1.0$ v.v.m.. This value compares favourably with $\delta = 0.8$ used in the simulations conducted by Petrides *et al.* (1989) and an average aeration rate of $\delta_{\text{average}} = 0.4$ for simulations conducted by Petrides *et al.* (1995).

Table 3.3: Summary of base-case values for design variables used in simulation.

Parameter	Value	References
$\hat{A}_{df}$	$0.04 \text{ m}^3 \text{ m}^{-2}$	Petrides <i>et al.</i> , 1989.
C_{ud}	8 M	Fischer <i>et al.</i> , 1993; see text
C_{ur}	2 M	Fischer <i>et al.</i> , 1993; see text
E	0.20 (–)	Kiefhaber <i>et al.</i> , 1991; Middelberg <i>et al.</i> , 1992.
f	0.75(–)	Design basis.
J	$5 \times 10^{-6} \text{ m}^3 \text{ m}^{-2} \text{ s}^{-1}$	Kalk and Langlykke, 1986; Kotlarski <i>et al.</i> , 1995.
k_N	$8.5 \times 10^{-4} \text{ s}^{-1}$	Kotlarski <i>et al.</i> , 1995 (see text).
K_A / k_N	$64 \text{ m}^3 \text{ kg}^{-1}$	Design basis (base case).
L	100	Middelberg, 1996b.
$\dot{m}_N$	$100 \times 10^{-6} \text{ kg s}^{-1}$ (=2.3 t.p.a.)	Design basis.
M_N	9.11 kg mol^{-1}	Design basis (LongR ³ IGF-I).
M_u	0.06 kg mol^{-1}	–
N_b	270	Design basis.
N_p	4	Middelberg <i>et al.</i> , 1992; Petrides <i>et al.</i> , 1989.
n_t	3	Design basis.
Q_Σ	$5 \times 10^{-9} \text{ m}^3 \text{ m}^{-2} \text{ s}^{-1}$	Middelberg and O'Neill, 1991; Wong <i>et al.</i> , 1996.
t	86400 s	Design basis.
t_c, t_h	72000 s	Design basis.
t_d	3600 s (= 1 h)	Fischer <i>et al.</i> , 1993; Turner <i>et al.</i> , 1994; Design basis.
t_{ru}	68400 s	Design basis.
t_s	10800	Petrides <i>et al.</i> , 1989; Design basis.
T	0.55	Niedhardt, 1987; Middelberg <i>et al.</i> , 1992.
W	30 kg m^{-3}	Kalk and Langlykke, 1986; Petrides <i>et al.</i> , 1989 & 1995.
δ	1.0 v.v.m.	Yoon <i>et al.</i> , 1994a&b; Turner <i>et al.</i> , 1994; Design basis.
ε	0.75	Design basis.
ρ_D	50 kg m^{-3}	Design basis.
ζ	1.43	Petrides <i>et al.</i> , 1989 & 1995

The required number of discrete passes through a high-pressure homogeniser depends on the disruption efficiency of the particular piece of equipment at the specific operating conditions used. Increasing the number of passes also serves to reduce the size of cell debris particles, thereby enhancing the purification obtained by differential sedimentation. Middelberg *et al.* (1992) employed four discrete passes, each with a pressure drop of 62 MPa to adequately disrupt *E. coli*. Three passes with a pressure drop of 80 MPa were used in the process simulated by Petrides *et al.* (1995), while four passes with a pressure drop of 80 MPa were used for the process simulated by Petrides *et al.* (1989). Taking a conservative approach, four passes have been used as the design specification for the required homogeniser capacity ($N_p = 4$). The time available for conducting the four passes must be less than the critical path with allowance for cleaning and sanitation between batches. If the time required for cleaning and sanitation is assumed to be four hours, the time available for homogenisation will be 20 h ($t_h = 72000$ s).

There is a trade-off between the amount and purity of product recovered by differential sedimentation; increasing purity is accompanied by decreasing recovery. For the purpose of simulation it is assumed that a fractional recovery of 75% will achieve a desirable increase in the purity of the inclusion body stream discharged from the centrifuge, while collecting sufficient product to ensure economic viability. For relatively small, dense inclusion bodies, a fractional recovery of 75% is attained using a normalised feed rate (Q_Σ) of approximately $5 \times 10^{-9} \text{ m}^3 \text{ m}^{-2} \text{ s}^{-1}$ (Middelberg and O'Neill, 1991; Wong *et al.*, 1996). For larger, relatively open-structured inclusion bodies, such as prochymosin, the normalised feed rate is greater at approximately $20 \times 10^{-9} \text{ m}^3 \text{ m}^{-2} \text{ s}^{-1}$ (Jin *et al.*, 1994; Mannwieler, 1989). To provide a conservative design, $Q_\Sigma = 5 \times 10^{-9} \text{ m}^3 \text{ m}^{-2} \text{ s}^{-1}$ has been used as the base case for simulations. If the time required to clean and sanitise the centrifuge is the same as that for the homogeniser, $t_c = 72000$ s.

The inclusion body-rich stream that is discharged from a continuous centrifuge typically contains approximately 50 % w/v solids (Datar and Rosen, 1987). Therefore, if it is assumed that dissolution is conducted by mixing this slurry with a denaturing solution to give a 1:10 dilution, the solids content during dissolution will be approximately 50 kg m^{-3} . Since the solids should be predominantly protein product (Builder *et al.*, 1986; Petrides *et al.*, 1995), a product concentration of $\rho_D = 50 \text{ kg m}^{-3}$ has been used. This is comparable

with $\rho_D = 20 - 80 \text{ kg m}^{-3}$ used by Bentle *et al.* (1987) to dissolve refractile bodies of bovine and porcine growth hormone. Strongly denaturing conditions require approximately 8 kmol m^{-3} (8 M) urea. Even using 8 M urea, dissolution takes between 0.5 and 2 hours (Fischer *et al.*, 1993a). For the purpose of this investigation, the time required for dissolution has been taken to be 1 h (*i.e.* $t_d = 3600 \text{ s}$ for $C_{ud} = 8 \text{ M}$).

As outlined in the design basis, dissolution, refolding and ultrafiltration are all conducted within the period of the critical path. If four hours are required for cleaning and sanitation, there will be a total of twenty hours available for operation. One hour is required for dissolution, therefore, nineteen hours are available for dissolution and refolding ($t_{ru} = 68400 \text{ s}$).

The flux through a tangential ultrafiltration membrane depends on the transmembrane pressure applied, design of the membrane, and properties of the fluid itself. Published values range between 10 and $50 \text{ L h}^{-1} \text{ m}^{-2}$ for ultrafiltration membranes with a molecular weight cut-off less than 3×10^5 (Kalk and Langlykke, 1986; Kotlarski *et al.*, 1995). This is equivalent to a flux of between 2.8×10^{-6} and $13.9 \times 10^{-6} \text{ m}^3 \text{ m}^{-2} \text{ s}^{-1}$. For base case simulations, $J = 5.0 \times 10^{-6} \text{ m}^3 \text{ m}^{-2} \text{ s}^{-1}$ has been used.

3.7.2 Protein-specific rate constants

The only protein-specific variables that have been used are for reaction rate constants and protein molecular weight. LongR³IGF-I has been used as the model protein. The molecular mass of the protein is 9.11 kDa. Rate constants for refolding were determined for fairly impure LongR³IGF-I by Kotlarski *et al.* (1995) (see sections 1.7 and 3.1.4). The rate constant for formation of native species in this study was $k_N = 8.5 \times 10^{-4} \text{ s}^{-1}$ and for formation of the aggregate species, $K_A = 0.0544 \text{ m}^3 \text{ mol}^{-1} \text{ s}^{-1}$. For base-base simulations, the rate of the back reaction is assumed to be $k_B = 8.5 \times 10^{-6} \text{ s}^{-1}$ (*i.e.* 1% of the forward reaction rate).

4. SIMULATION: OVERALL PROCESS INTERACTIONS

A systematic comparison of different reactor types and operating conditions for refolding is presented in this chapter. The influence of refolding yield upon all units in the generic production process outlined in section 3.3 is considered. Performance of the overall process is measured using the dimensionless objective function defined in section 3.5.

Optimal refolding conditions were identified by exhaustive search to minimise the dimensionless annual operating cost function. The influences of conversion of denatured protein, protein concentration and reaction rates were extensively investigated. Reactor selection and optimisation of refolding using a global approach were demonstrated to be vitally important.

The key findings presented in this chapter have been published in Kotlarski *et al.* (1997).

4.1 COMPARISON OF OPERATING MODES - BASE CASE

Simulations were conducted for all three refolding schemes using Mathcad 6.0e Plus to solve the design equations and corresponding dimensionless cost components as outlined in the numerical methods section 2.3.1. For the base case simulations, the rate constant for the back reaction of native to intermediate (see section 3.1.3) was taken as one percent of the forward reaction, *i.e.* $k_B/k_N = 0.01$. Optimal conditions were identified as a function of protein concentration and conversion of intermediate species by exhaustive search.

The general effects observed were: increasing overall protein concentration reduced the yield of native protein, while increasing conversion first increased yield to a maximum which then declined as the reverse pathway became significant.

4.1.1 Batch mode

Figure 4.1 is the dimensionless cost contour for batch refolding as a function of conversion and dimensionless protein concentration. The presented range contains the minimum overall operating cost that can be achieved using refolding in batch mode. The value of the objective function at the minimum is 28.6. This occurs at the maximum of the conversion range (0.996) and a dimensionless concentration of 6.2. The component contributions are approximately 44%, 54% and 2% for the DFC-dependent, consumable, and waste treatment costs, respectively. A contour plot of yield over the same range indicates an optimal yield of approximately 31% (Figure 4.2). A contour plot of dimensionless time over the corresponding range of conversion and dimensionless protein concentration is presented in Figure 4.3. The dimensionless time for which the refolding reaction must be allowed to proceed to attain optimal overall performance is 4.9.

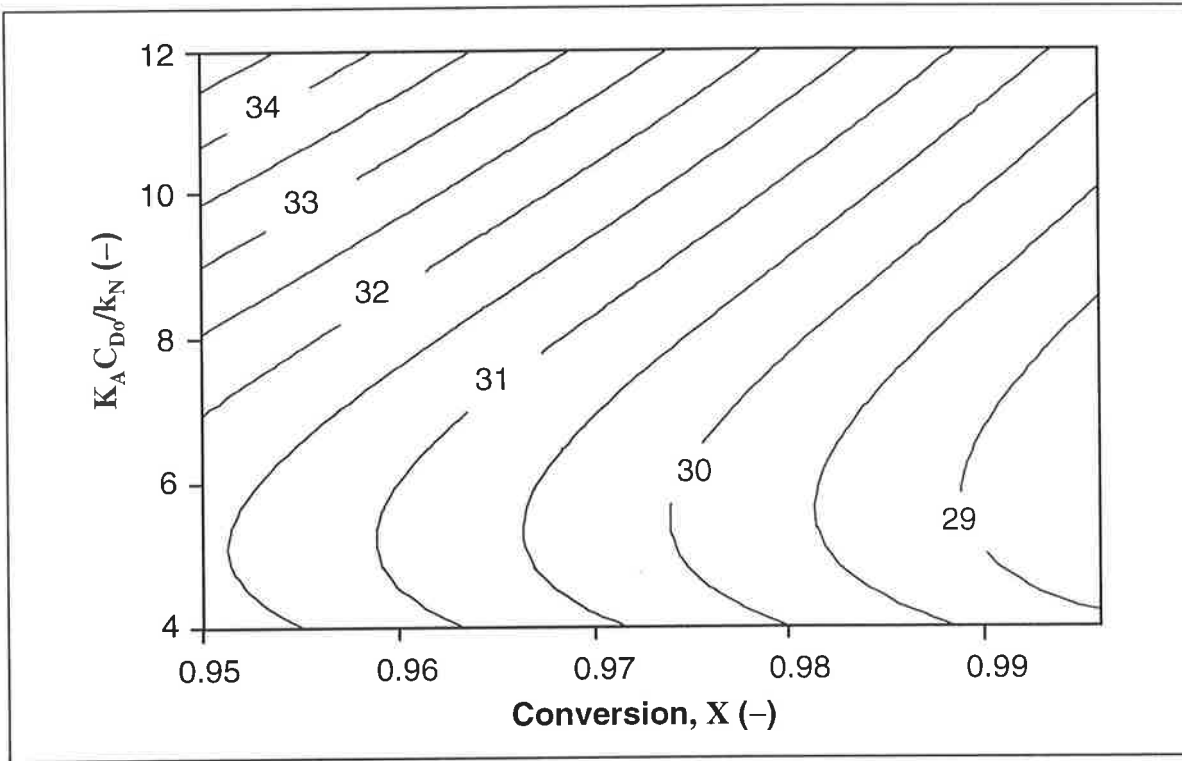


Figure 4.1: Dimensionless cost contour (ϕ^*) for batch refolding as a function of conversion and dimensionless protein concentration ($K_A C_{D0}/k_N$) for $k_B/k_N = 0.01$.

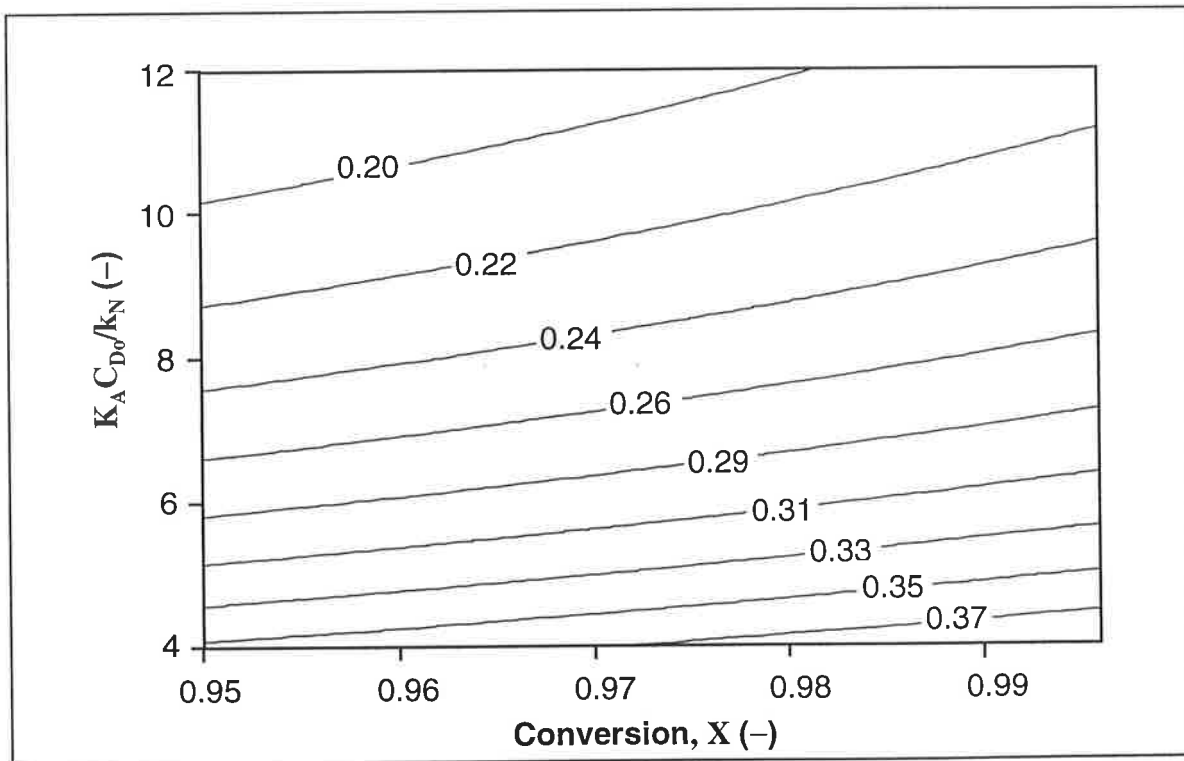


Figure 4.2: Yield contour (y_N) for batch refolding as a function of conversion and dimensionless protein concentration ($K_A C_{D0}/k_N$) for $k_B/k_N = 0.01$.

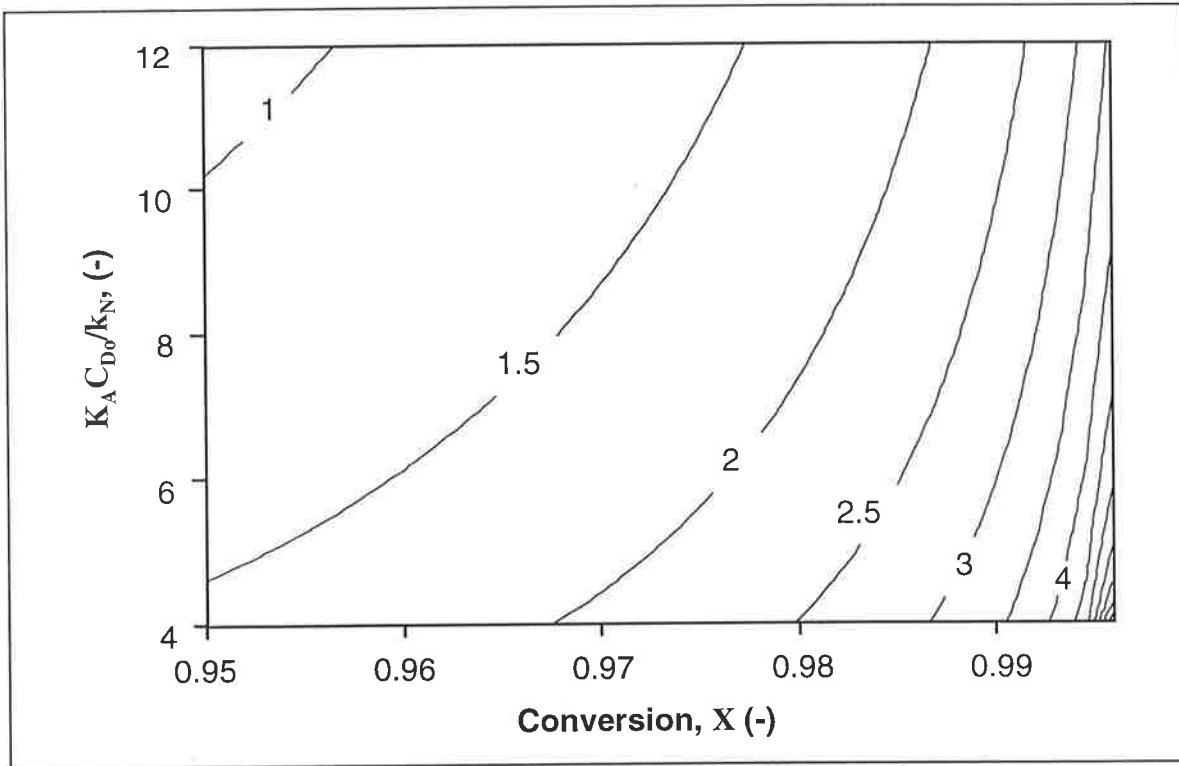


Figure 4.3: Dimensionless time contour (τ) for batch refolding as a function of conversion and dimensionless protein concentration ($K_A C_{D0}/k_N$) for $k_B/k_N = 0.01$.

4.1.2 Fed-batch mode

Fed-batch operation for the base-case gives a minimum value of the objective function of 12.0. This is less than 50% of the optimum batch refolding cost. The dimensionless cost contour for base-case fed-batch refolding is presented in Figure 4.4. A region of optimum operation can be identified at a conversion of approximately 0.98 and an effective dimensionless protein concentration of 22. DFC-dependent costs contribute approximately 65% of the total. Consumable costs contribute approximately 32%, and waste management costs 3%. There is a severe penalty for operating at the lower range of protein concentration and at high conversion. The corresponding contour plot of yield shows that optimum conditions are achieved at a yield of approximately 63% (Figure 4.5). A dimensionless time contour reveals that the optimum duration of feeding is 42 (Figure 4.6).

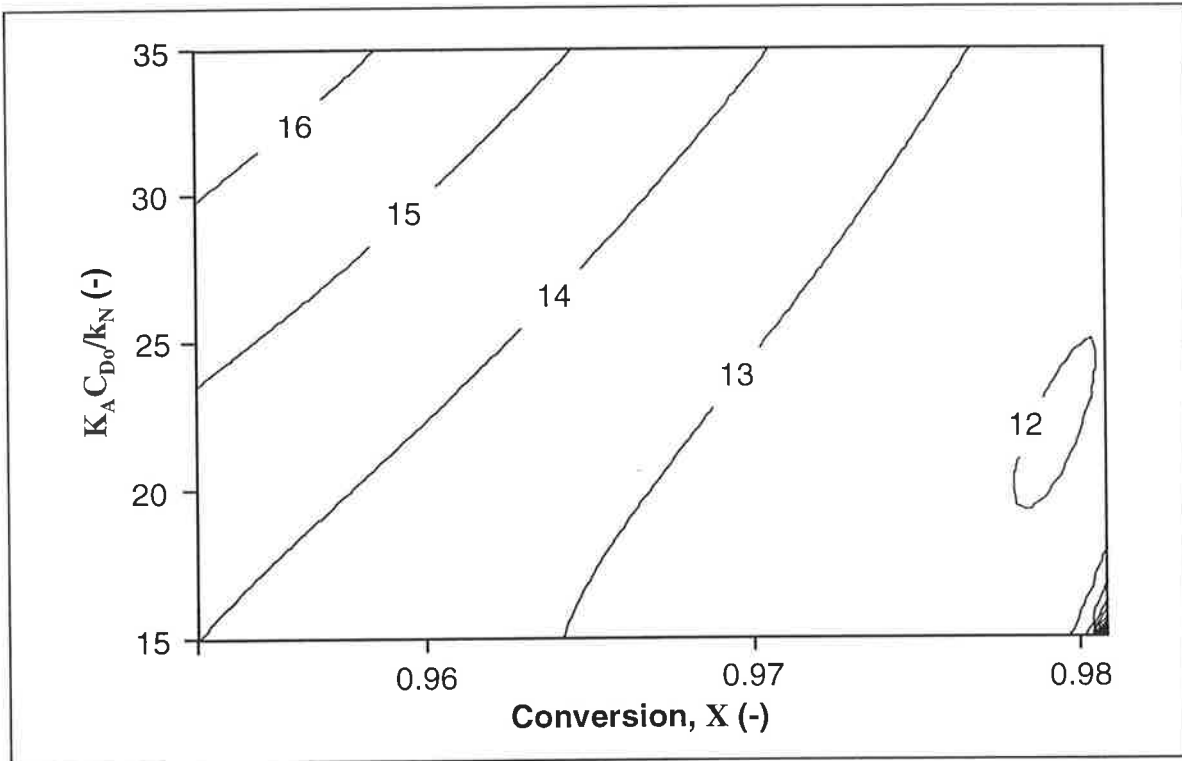


Figure 4.4: Dimensionless cost contour (ϕ^*) for fed-batch refolding as a function of conversion and dimensionless protein concentration ($K_A C_{D0}/k_N$) for $k_B/k_N = 0.01$.

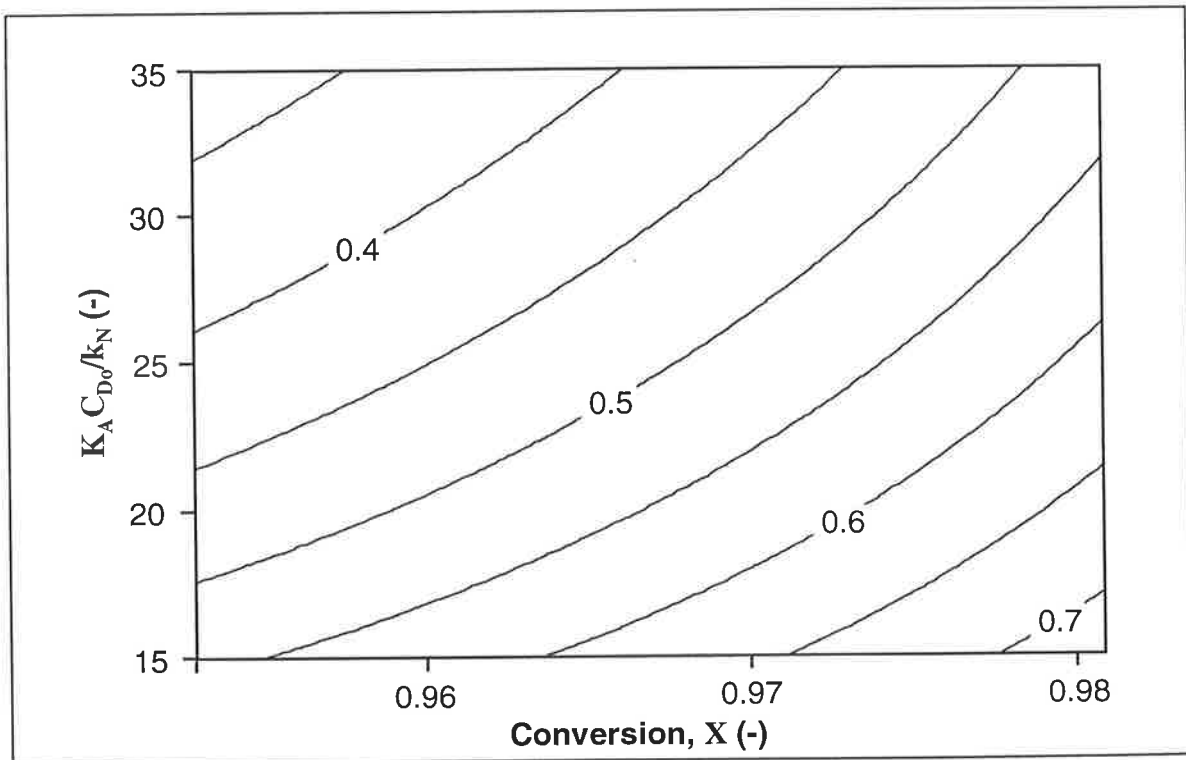


Figure 4.5: Yield contour (y_N) for fed-batch refolding as a function of conversion and dimensionless protein concentration ($K_A C_{D0}/k_N$) for $k_B/k_N = 0.01$.

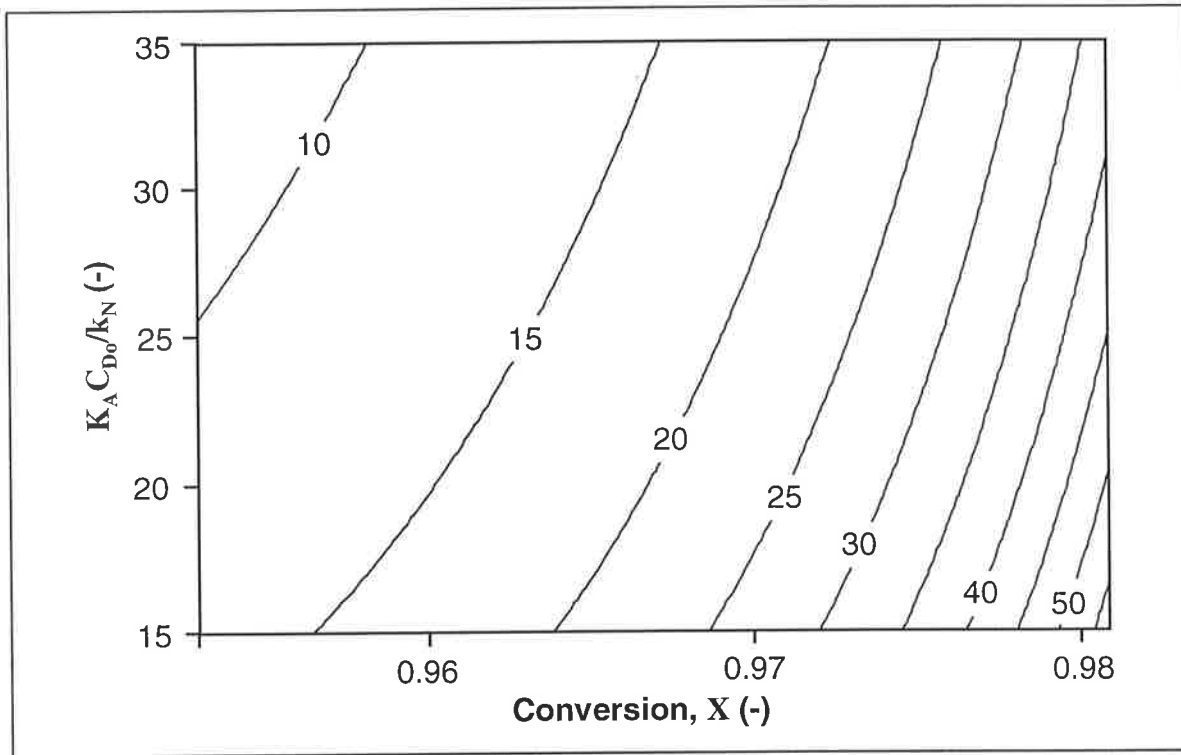


Figure 4.6: Dimensionless time contour (τ) for fed-batch refolding as a function of conversion and dimensionless protein concentration ($K_A C_{D0}/k_N$) for $k_B/k_N = 0.01$.

4.1.3 Continuous mode

Optimum operating conditions for continuous refolding occur at a conversion of 0.985 and a dimensionless protein concentration of 21.6. Figure 4.7 shows the variation in objective function with conversion and dimensionless protein concentration for this refolding mode. The minimum value of the objective function is approximately 11.4, only 5% less than for fed-batch operation and the relative contributions from component costs are virtually identical. The corresponding contour plot for yield shows that minimum cost corresponds to a yield of approximately 63% (Figure 4.8). This yield is more than twice the yield achieved for overall optimum operation of the batch process, but the same as that for fed-batch operation. The Damköhler number (dimensionless time) at the optimum is approximately 74. Figure 4.9 illustrates the variation in Damköhler number with conversion and dimensionless protein concentration. A dramatic increase in Damköhler number is observed for conversions in excess of approximately 0.98, particularly for low values of dimensionless protein concentration.

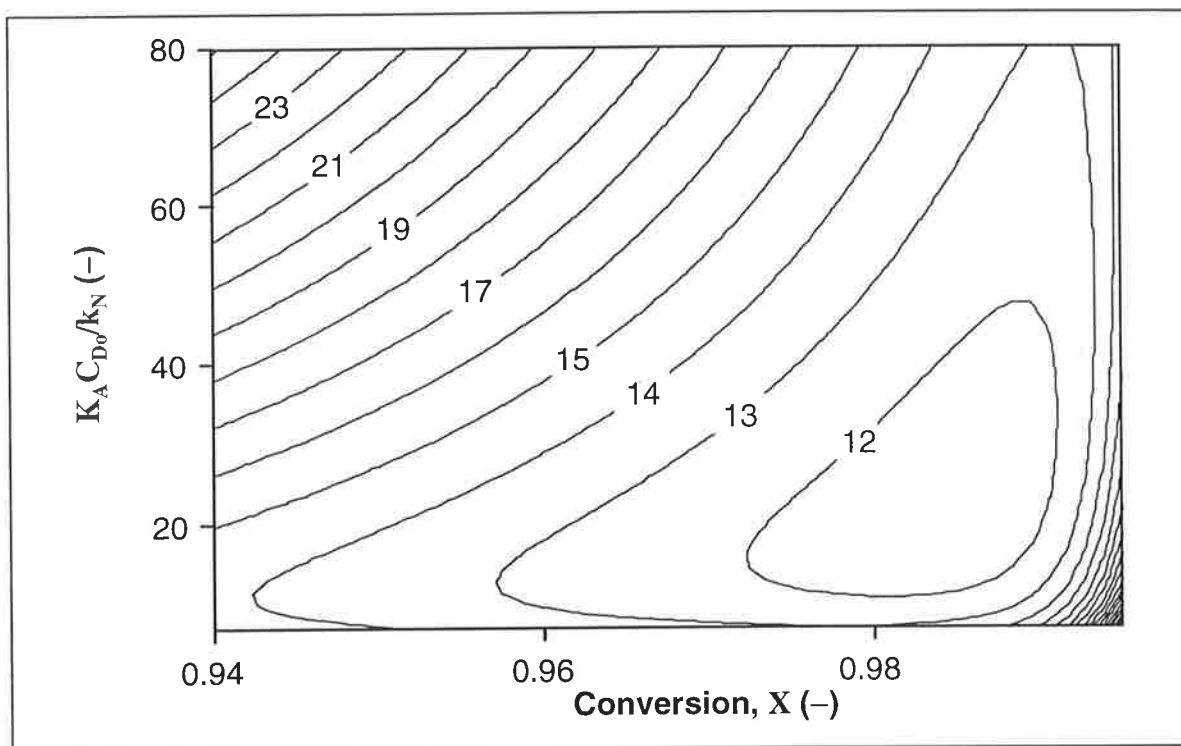


Figure 4.7: Dimensionless cost contour (ϕ^*) for continuous refolding as a function of conversion and dimensionless protein concentration ($K_A C_{D0}/k_N$) for $k_B/k_N = 0.01$.

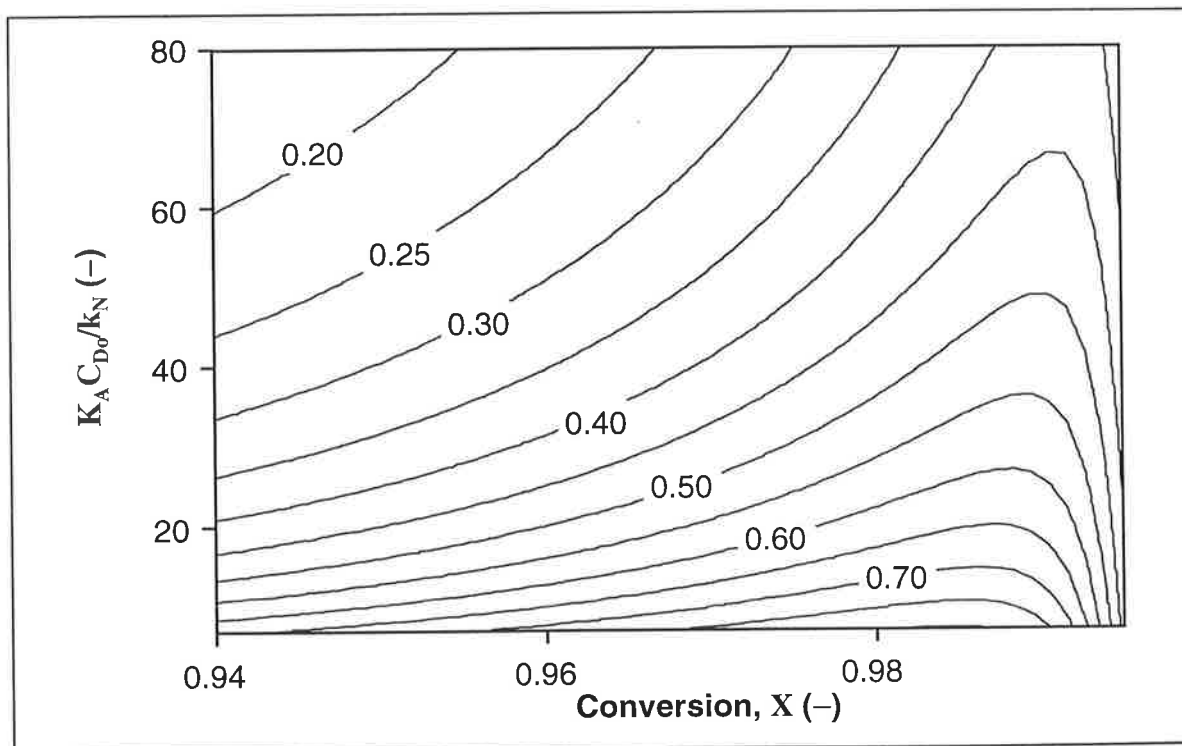


Figure 4.8: Yield contour (y_N) for continuous refolding as a function of conversion and dimensionless protein concentration ($K_A C_{D0}/k_N$) for $k_B/k_N = 0.01$.

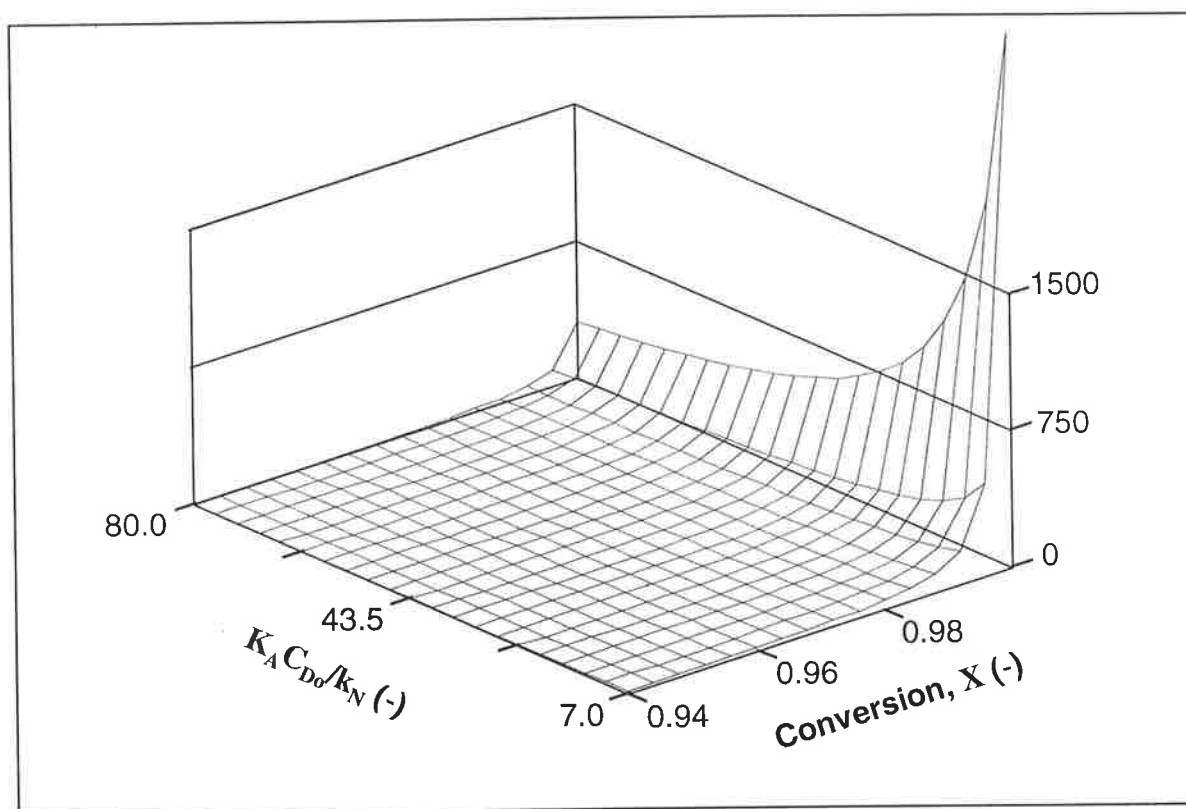


Figure 4.9: Plot of dimensionless time (τ) for continuous refolding as a function of conversion and dimensionless protein concentration ($K_A C_{D0}/k_N$) for $k_B/k_N = 0.01$.

4.1.4 Transient intermediate concentration

Examination of the transient intermediate concentration reveals that the high protein concentrations present during the initial stages of batch operation are avoided in the other two modes. Figure 4.10 illustrates the transient intermediate concentration for the three modes of operation under optimal conditions. Batch refolding starts at a dimensionless concentration of approximately 6.1 and rapidly decreases to reach equilibrium. The final concentration of intermediate is approximately 0.019. For fed-batch refolding, the intermediate concentration is initially zero, then quickly increases to a maximum of 0.47 before gradually declining to a final concentration of 0.45 at the end of feeding. The intermediate concentration is maintained at approximately 0.32 for refolding in continuous mode. It is interesting to note that for most of the reaction time, the intermediate concentration for fed-batch operation is approximately 50% greater than for continuous operation. However, the overall yields of protein in the native conformation are virtually identical.

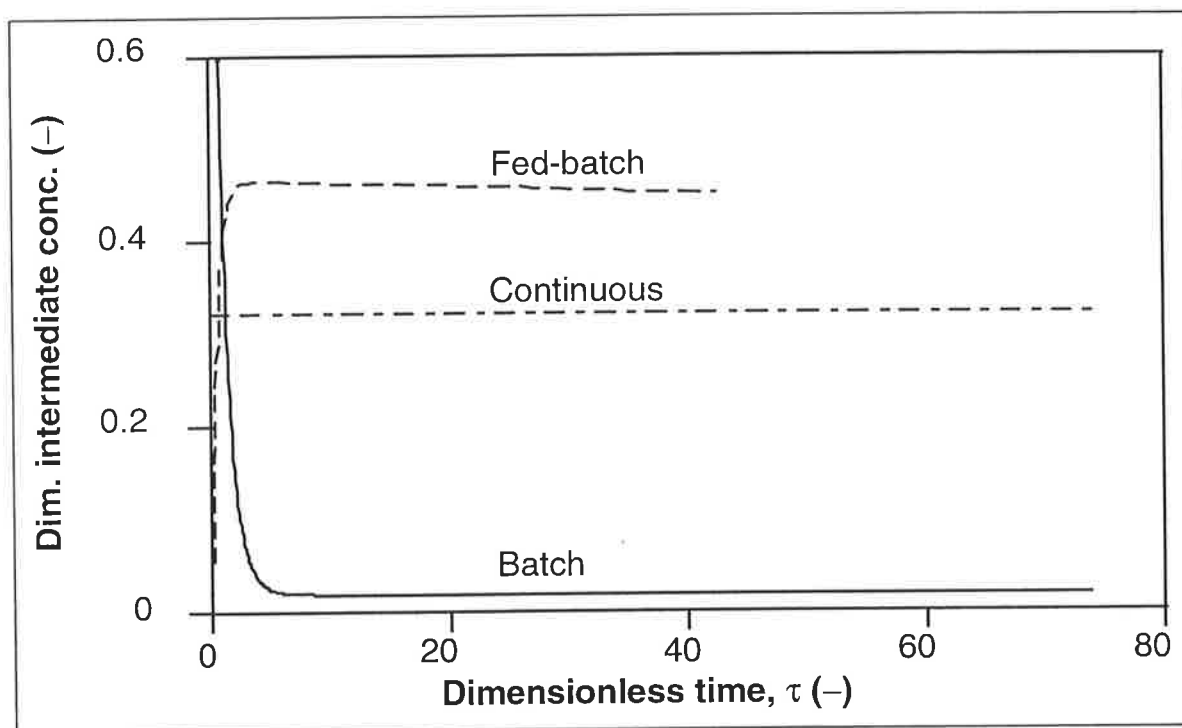


Figure 4.10: Transient dimensionless intermediate protein concentration present in each refolding reactor during refolding in batch, fed-batch and continuous mode under optimum conditions. The duration of the batch reaction extended beyond the optimum (4.9).

4.1.5 Discussion - base case

Optimal dimensionless protein concentrations were not less than 6 for the three modes of operation. This value is equivalent to a concentration of approximately 0.1 mol m^{-3} (0.9 kg m^{-3}) for LongR³IGF-I. While higher protein concentrations have been reported in commercial-scale production (Bentle *et al.*, 1987; Chang *et al.*, 1994), the value identified for LongR³IGF-I was approximately seven times that used in the present small-scale production process (Kotlarski *et al.*, 1995) and several orders of magnitude greater than is typically used for laboratory-scale renaturation (Fischer *et al.*, 1993a).

Batch operation is the most expensive mode for protein refolding. Optimum conditions were identified as relatively low protein concentration, high conversion (>0.97), and short duration of refolding to give low yield ($\approx 30\%$). Higher yields could have been achieved by using lower protein concentration. However, low concentration requires increased liquid

volumes and is penalised by an increase in the cost of refolding/ultrafiltration equipment and consumables. Operation at low yield results in high base cost for the process as the capacity of all upstream units must be increased to compensate for the reduced efficiency of recovery. In the batch mode, high conversion is rapidly obtained since no fresh reactants need to be added. Figure 4.2 shows that refolding yield is relatively insensitive to conversion at high values. Extending reaction duration reduces the time available for concentrating the refolded mixture. Hence, optimum conditions occur at high conversion, but with a relatively short reaction time. The high base cost of the process at low yield makes it less sensitive to variations in concentration and equilibrium constant than either of the alternative modes.

Continuous refolding gives the lowest cost of operation at optimum conditions. This is not unexpected as it has the advantage of maximum equipment utilisation. Also, production in a CSTR is known to improve selectivity of lower-order reaction products relative to that obtained with batch reactors (Rase, 1977). Optimum conditions favour operation at high conversion and high overall protein concentration. However, unlike batch refolding, intermediate species (in the form of denatured species) are continuously added to the reactor. While the overall protein concentration is high, continuous addition maintains a low reactive intermediate concentration. At the optimal conditions identified for the base-case simulation, the inlet concentration of denatured protein is approximately 3.1 kg m^{-3} for LongR³IGF-I, but the intermediate concentration maintained in the CSTR is only 0.04 kg m^{-3} . Low intermediate concentration (20 times less than batch refolding) returns a high yield of correctly refolded protein, reducing the volumetric flow rate throughout the entire process and the size of upstream units (*e.g.* the fermenter). The results for continuous refolding agree with those published by Middelberg (1996b), allowing for minor differences in costing.

Continuous refolding has the advantage that the duration of refolding can be extended beyond the length of the critical path. The dimensionless time of the maximum duration of batch and fed-batch refolding is 58 (corresponding to 19 h) yet duration of the reaction for optimum base-case operation is 74 (approximately 24 h; the critical path of the overall process). Figure 4.8 shows that there is a significant increase in refolding yield with conversion up to a maximum (≈ 0.985) and then a very rapid decrease. That the rapid decrease occurs with a dramatic increase in the required holding time, can be seen in Figure

4.10. The reason for this is unfolding of the native protein to intermediate conformation. In other words, even for a favourable equilibrium (1% back reaction), unfolding of the native protein to intermediate structures that may be lost to aggregation is significant if the reaction is allowed to proceed for too long. It appears that for the base-case, unfolding becomes significant for durations approximately greater than the critical path (design basis = 24 h). The cost penalty for operation within the range that allows the back reaction to become significant is clearly shown in Figure 4.7 by the high density of cost contours to the right of the optimum. It appears that the advantage of extending the reaction time beyond the length of the critical path is not significant because of the increase in refolding volume that must be processed.

Another issue that must be considered at this conceptual design stage is process validation. Unlike batch and fed-batch operation, batch integrity is not maintained throughout the operation in continuous mode, a significant design consideration for production of therapeutic products (Gordon *et al.*, 1990). The added complexity of more stringent process control and validation must detract from the advantage that refolding in continuous mode has to offer.

Performance of fed-batch operation is virtually identical to that of continuous refolding over the same range of conversions and protein concentrations (Figures 4.4 and 4.7). Optimum conditions occur at a dimensionless protein concentration of 22 (3.1 kg m^{-3} for LongR³IGF-I). The corresponding final dimensionless time of refolding is 42.5 and the yield 63%. The plot of transient intermediate concentration reveals that it increases rapidly to a maximum of 0.066 kg m^{-3} and then slowly decays to a final concentration of 0.064 kg m^{-3} . This maximum intermediate concentration is approximately 14 times lower than that for batch refolding and 60% greater than for continuous refolding. The design criteria for refolding and subsequent concentration within the critical cycle time results in massive ultrafiltration capacity as the duration of refolding approaches this upper limit. The high density of cost contours at high conversion and low concentration (see Figure 4.4) is due to this effect and is the major difference between the fed-batch and continuous operation.

Breakdown of the overall dimensionless costs into DFC-dependent and other cost components for optimal operating conditions of the three modes is presented graphically in Figure 4.11(a). An alternative division into fermenter-, refolding-associated, and other costs is presented in Figure 4.11(b).

Figure 4.11(a) shows that the majority of cost-reduction in fed-batch and continuous modes is due to the decrease in non-DFC-dependent costs. In batch mode, the cost of chemical additives, recycling refolding buffer, and ultrafiltration membrane is more than three times that of fed-batch or continuous operation. Yet the difference in refolding yield is only two-fold, disfavouring operation in batch mode. This discrepancy is explained by the difference in overall protein concentration; fed-batch and continuous modes operate optimally at a dimensionless protein concentration approximately four-fold greater than batch mode, resulting in reduced consumable usage and cost.

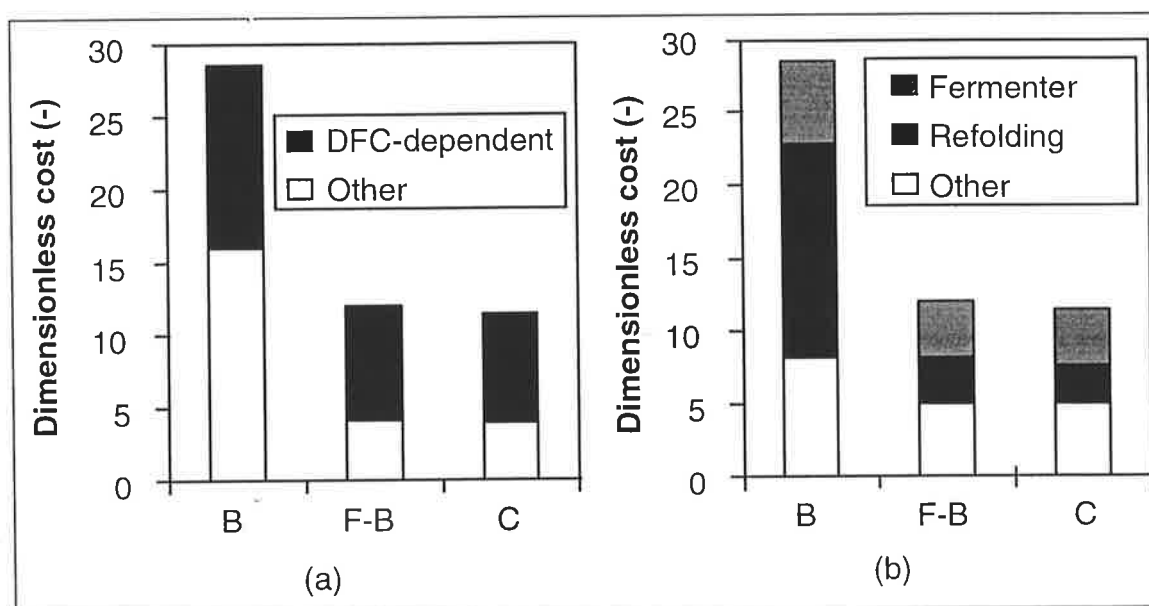


Figure 4.11: Breakdown of dimensionless total annual operating cost into components for each of the three operating modes. (a) Costs subdivided into DFC-dependent costs and other (sum of waste treatment and consumable costs). (b) Costs subdivided into fermenter-associated (media, fermenter, steriliser, and air compressor), refolding-associated (chemical consumables, refolding buffer recycle, dissolution reactor, refolding reactor, ultrafiltration rig and membranes, and depth filters), and other cost components (homogeniser, centrifuge, and holding tanks). B = batch mode, F-B = fed-batch mode, and C = continuous mode.

Figure 4.11(b) illustrates that while the fermentation-associated costs of fed-batch and continuous operation are nearly half that of batch operation (due to the higher yield), the refolding-associated components cost less than one quarter (due to higher yield at lower protein concentration).

4.2 SENSITIVITY ANALYSIS

The simulations conducted in section 4.1 are all carried out using an equilibrium constant of $k_B/k_N=0.01$ and the estimates for specific costs presented in section 3.8. Because, the findings from these simulations are depended on the assumed values, a sensitivity analysis was conducted to determine the relative influence of varying the equilibrium constant and specific cost estimates.

4.2.1 Variation in equilibrium

Simulations were conducted using two additional values of the equilibrium constant and the optimum operating conditions were identified. The two values of k_B/k_N used were 0 and 0.1.

For all three modes of refolding, the optimum operating conversion decreases with increasing equilibrium constant at fixed protein concentration. For pure competing

Table 4.1: Summary of minimum values of the objective function with corresponding conversion and dimensionless protein concentration for the three values of the equilibrium constant.

	$k_B/k_N = 0$	$k_B/k_N = 0.01$	$k_B/k_N = 0.1$
Batch:			
ϕ^*	28.2	28.5	31.7
X	>0.999	> 0.99	0.973
$K_A C_{D0}/k_N$	6.2	6.2	6.3
Fed-batch:			
ϕ^*	11.2	12.0	17.8
X	$\cong 0.98$	0.97 – 0.98	0.935 – 0.94
$K_A C_{D0}/k_N$	20 – 25	20-24	12-15
Continuous:			
ϕ^*	8.2	11.4	19.6
X	>0.999	0.985	$\cong 0.92$
$K_A C_{D0}/k_N$	150 – 250	21.6	$\cong 13$

reactions ($k_B/k_N = 0$) there is only a slight difference in objective cost function curves when compared to the base case. However, an increase in the equilibrium constant from 0.01 to 0.1 gives a significant change in cost contours (not shown). With $k_B/k_N = 0$ and 0.1, the optima are 28.2, 11.2, and 8.3 and 31.7, 17.8, and 19.6 for batch, fed-batch, and continuous refolding, respectively. The increasing values of the optima occurred at lower conversions and with the exception of batch operation, at lower dimensionless protein concentration. The optima for fed-batch operation occurred at dimensionless protein concentrations of approximately 6 for both extreme values of the equilibrium constant. The trend towards lower conversion was least significant for batch operation. For example, at $k_B/k_N = 0.1$ the conversions for optimal operation were 0.97, 0.94 and 0.92 for batch, fed-batch, and continuous reactors. The minimum values of the objective function for the three values of the equilibrium constant are summarised in Table 4.1. The table also includes the corresponding conversion and dimensionless protein concentration.

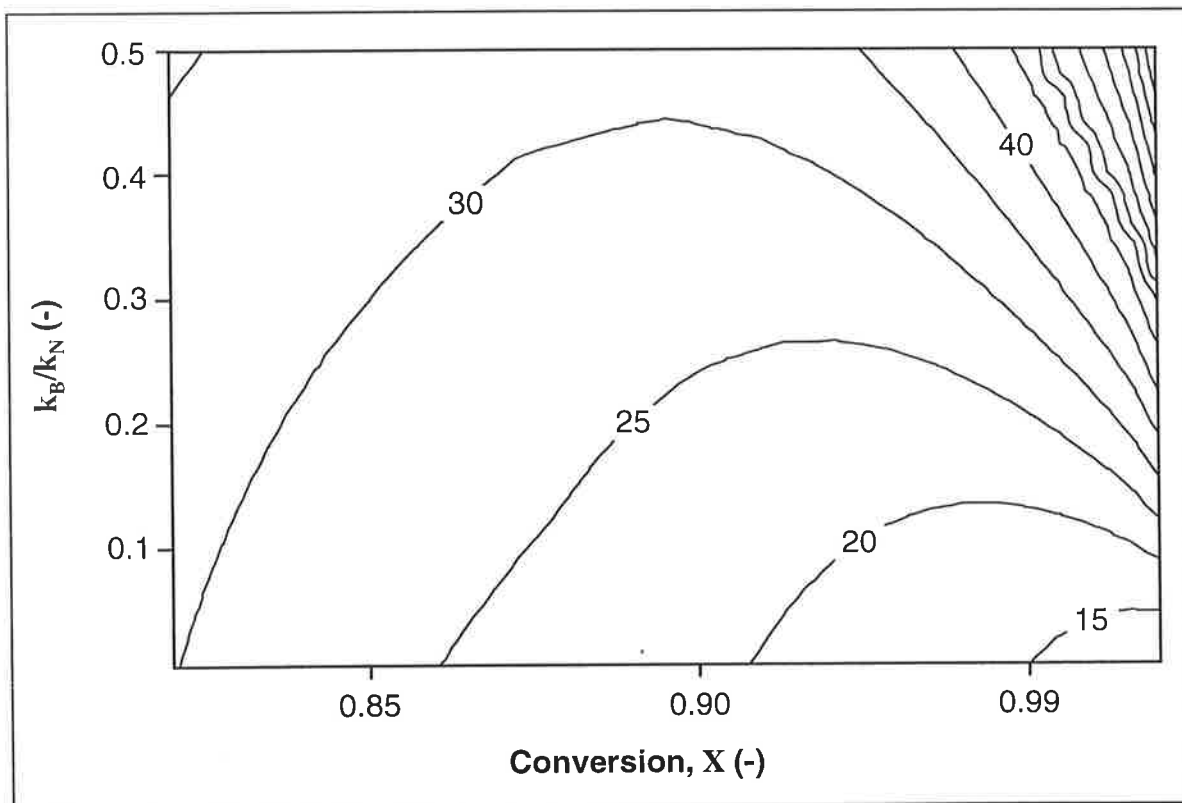


Figure 4.12: Dimensionless cost contour (ϕ^*) for fed-batch refolding as a function of conversion and equilibrium constant (k_B/k_N). Dimensionless protein concentration ($K_A C_{D0}/k_N$) fixed at 22.

Figure 4.12 shows the dimensionless cost contours for fed-batch refolding as a function of conversion and equilibrium constant at a dimensionless protein concentration of 22. This concentration gives optimal operation for the base case simulations in which the equilibrium constant was 0.01. Batch and continuous operation give contours with similar form. Batch refolding is least sensitive to variation in equilibrium constant, particularly at high protein concentrations. Performance of continuous and fed-batch operation is quite similar over the range of conditions considered, and accordingly the results are not shown.

Since the concentration of reactive intermediate is limited in both fed-batch and continuous refolding to approximately the same value, the similar sensitivity to variation in equilibrium constant is understandable. Batch operation has greater base cost which reduces the relative shift in cost with changing rates of reaction, making it least sensitive.

4.2.2 Sensitivity to consumable cost estimates

Sensitivity of overall performance to changes in consumable costs was assessed by simulation over a four-fold range about the mean estimated value for each specific cost. Batch refolding is most sensitive to changing cost estimates because consumable costs contribute greater than 50% of the overall cost for operation about the optimal conditions. For fed-batch and continuous operation the consumable cost contribution is less than 35% of the total cost. For batch refolding the order of significance is $\gamma_{ar} > \gamma_{up} > \gamma_m > \gamma_{df} > \gamma_u > \gamma_{ad}$ with approximately +20/-10% variation with refolding additives and only +0.8/-0.4% variation with dissolution additives. Not surprisingly, fed-batch and continuous refolding display similar sensitivities to one another. Simulations reveal the order of significance is $\gamma_m > \gamma_{ar} > \gamma_{up} > \gamma_{ad} > \gamma_{df} \equiv \gamma_u$ with approximately +15/-8% variation in the minimum value of the objective function with medium cost and only +0.5/-0.2% variation with filtration costs.

4.3 IMPLICATIONS FOR EXPERIMENTAL STUDIES

The simulations of large-scale processing reveal that the current practice of operating refolding in batch mode is likely to be more than twice as expensive as either fed-batch or continuous modes. The refolding-associated operating costs of batch operation constitute more than 50% of the total costs. In contrast, fed-batch and continuous operation have similar refolding-associated component costs which are approximately 25% of the total. The reduction in the proportion of refolding-associated costs is achieved by operation at higher total protein concentrations (approximately four-fold greater than batch mode). The overall reduction in costs is due to higher yields that can be obtained by refolding in fed-batch and continuous modes at the higher overall protein concentrations. The initial high transient concentration of denatured protein that is unavoidable in batch operation is eliminated by gradually adding denatured species to the refolding reaction in fed-batch or continuous modes of operation. An overall yield that is twice that of batch refolding can be obtained at a four-fold greater overall protein concentration using fed-batch or continuous strategies.

In the case-study of process economics presented by Datar *et al.* (1993), a key conclusion was that increasing refolding yield provided a more significant increase in overall process profitability than increasing refolding concentration. This conclusion was based on the assumption that refolding yield was not reduced by increased protein concentration. In the analysis presented above, the improved economic performance of fed-batch and continuous refolding is due to a combination of both increased renaturation yield and protein concentration. These findings take into account the penalty in yield that refolding at high protein concentration incurs. The importance of refolding at high protein concentration, together with maintaining a high renaturation yield has been demonstrated. Additionally, many of the recent developments in protein refolding methods have been aimed at increasing the overall protein concentration during refolding, along with improving renaturation yield (see section 1.5.5).

Refolding in fed-batch mode is likely to be the superior operating strategy for production of therapeutic proteins because batch integrity can be maintained. Secondly, fed-batch operation is likely to be less complex to control than continuous operation and start-up

problems are avoided. These benefits outweigh the arguable reduction in cost that may be achieved by operating in continuous mode. The improved performance of continuous operation is questionable because the factored-estimate method, upon which the results are based, has intrinsic uncertainty that makes the difference between the two modes relatively insignificant.

The potential of continuous modes of operation for improving large-scale refolding will be considered more seriously once the number of reports of successfully employing such techniques have increased. At present, the number of proteins that have been shown to give higher yields of active protein when refolded in continuous mode is small (see section 1.5.5). Therefore, it is important to conduct an experimental comparison of refolding in batch and continuous modes to validate the design analysis presented above. The generic refolding mechanism presented in section 3.1.3 is based on the observations of batch experiments. Its applicability to continuous systems has yet to be shown. Gradual dilution of denatured species is likely to highlight the importance of mixing if mass-transfer is rate-limiting (see section 1.5.6.2).

4.4 CONCLUSIONS FROM SIMULATION

A method for optimisation of refolding monomeric proteins expressed using a simplified kinetic pathway and fundamental reactor design equations has been presented. Numerical analysis in dimensionless form can be used to directly compare different modes of refolding on a cost basis, across a range of operating conditions. The findings are a function of the process, method of costing, and refolding pathway. However, since a generic process and widely applicable refolding pathway have been used, the results should be relevant to the production of many recombinant proteins.

Continuous operation of protein refolding under optimised conditions is the most cost-effective method for renaturation. Fed-batch operation can be optimised to attain a similar performance (within 25%). However, refolding as a single batch results in a process that is at least twice as expensive to operate. Both continuous and fed-batch operation limit the concentration of rapidly formed intermediates present in the refolding reactor. Therefore, despite operation at high overall protein concentration, the reactive intermediate concentration is low and consequently yield is high ($\approx 60\%$).

During batch refolding all of the protein in the reactor is transiently present as reactive intermediate, resulting in a brief period of rapid aggregation. Therefore, optimum operation occurs at extremely high conversion (>0.97) and low overall protein concentration with yields less than 35% of desired product. Batch operation is more sensitive to consumable costs than either fed-batch or continuous refolding because more than 50% of the operating cost is for consumables. Batch refolding does have the advantage of being least sensitive to variation in rates of reaction and operating conditions. However, this insensitivity is due to the high base cost of the overall process for operation at low yield, rather than robustness of the refolding unit itself.

The importance of considering fed-batch and continuous modes for refolding as alternatives to batch operation has been clearly demonstrated. Systematic experimental studies are required to validate the application of continuous modes of operation and the findings of this study. Overall process optimisation must be used to determine operating conditions for individual units of equipment. Economic criteria must be used to identify optimum operating conditions.

5. REFOLDING KINETICS FOR MODIFIED PROCESSES

In this chapter dissolution of the inclusion body starting materials is presented and discussed. Refolding trials are presented that aim to identify whether variation in rate constants occur with:

- different mixing rates
- presence of soluble contaminants from whole homogenate
- presence of insoluble contaminants and whole homogenate

Repeated refolding trials at high concentration are conducted to provide controls for the subsequent fed-batch refolding experiments.

5.1 DISSOLUTION

Before refolding can be examined it is important to ensure that the preceding steps are not the limiting factor in the overall process. To this end, the dissolution step was investigated to determine whether it consistently provided sufficient material for refolding experiments. In this section, the cumulative observations for dissolution of clean and dirty inclusion bodies will be presented and discussed.

5.1.1 Experimental protocol - dissolution

To provide denatured, reduced LongR³IGF-I with various amounts of cellular contaminants, thawed inclusion body paste of this protein was dissolved using procedures based on the method outlined in section 2.6.

5.1.1.1 Mixing experiments

For each trial a batch of clean inclusion bodies was removed from storage at -20°C and allowed to thaw at room temperature. The progress of dissolution was monitored by quenching small samples for total protein analysis and HPLC. Samples for total protein quantitation were quenched in triplicate by dilution of 50 µL of dissolving inclusion body mixture into 10 mL of 0.1 M NaCl/0.1% TFA and vortexing for 15 s. Samples for analysis by HPLC were quenched by dilution of 20 µL of dissolving inclusion body mixture into 1000 µL of 0.1% TFA. After vortexing for 15 s, the samples were allowed to stand at room temperature for at least 5 minutes before being centrifuged at $18 \times 10^3 g_{\max}$ for 3 minutes. The resulting supernatants were carefully decanted and analysed by HPLC to determine LongR³IGF-I content. To test whether maximal dissolution was achieved, the method presented in section 2.6.1 was used.

For one of the dissolution trials, the dissolution buffer was purged with nitrogen gas before adding the reducing agent. During the course of dissolution, the mixture was blanketed with gaseous nitrogen by purging the head space of the reaction vessel after removal of

each sample and then sealing in the gas with laboratory film. The progress of dissolution was monitored for five hours.

5.1.1.2 Contaminant trials

To investigate the effect of soluble contaminants on refolding, two trials were conducted. The first was a control with all buffers prepared in water. The experimental procedure used was the same as that for mixing trial (section 5.1.1.1), except that sampling was continued over a longer period and assays of particle size analysis and SDS-PAGE were included. The second trial used batch supernatant as the solvent instead of water and the same protocol as the one for the first trial.

Dissolution for the first control trial, was conducted using water to prepare all buffers (see section 2.6). For the second trial an aliquot of frozen supernatant was removed from storage at -20°C and allowed to thaw at room temperature. The stock dissolution buffer (8.8 M urea/0.11 M Tris-HCl pH = 9.1/ 0.044 M Gly/ 0.011 M EDTA) and refolding diluent (0.1 M Tris-HCl pH = 9.1/0.01 M EDTA) were prepared using the supernatant. A batch of clean inclusion bodies was removed from storage at -20°C and allowed to thaw at room temperature. The thawed paste was then dissolved using the standard procedure outlined in section 2.6, using the buffers prepared with supernatant.

For both dissolution trials, refolding was initiated after 30 minutes by a 1:25 dilution of a 600 μL subsample to provide the correct refolding environment. Samples for HPLC and total protein quantification were as described for the mixing experiments (section 5.1.1.1) except that for analysis by HPLC, 50 μL instead of 20 μL were quenched into 1000 μL of 0.1% TFA. For SDS-PAGE analysis, a 100 μL sample was centrifuged for three minutes at $18 \times 10^3 g_{\text{max}}$. 50 μL of the supernatant was diluted a minimum of 1:2 into reducing load buffer. For particle size analysis, the solids from 500 μL of dissolving mixture were sedimented by centrifugation at $18 \times 10^3 g_{\text{max}}$ for three minutes. The supernatant was carefully decanted and the recovered solids suspended in 1000 μL of 20% ethanol/0.1% TFA. To achieve homogeneous resuspension, samples were sonicated for 15 minutes in an ultrasonic bath at room temperature. Specimens were stored at 4°C until size analysis was conducted. The sampling protocol is presented diagrammatically in Figure 5.1.

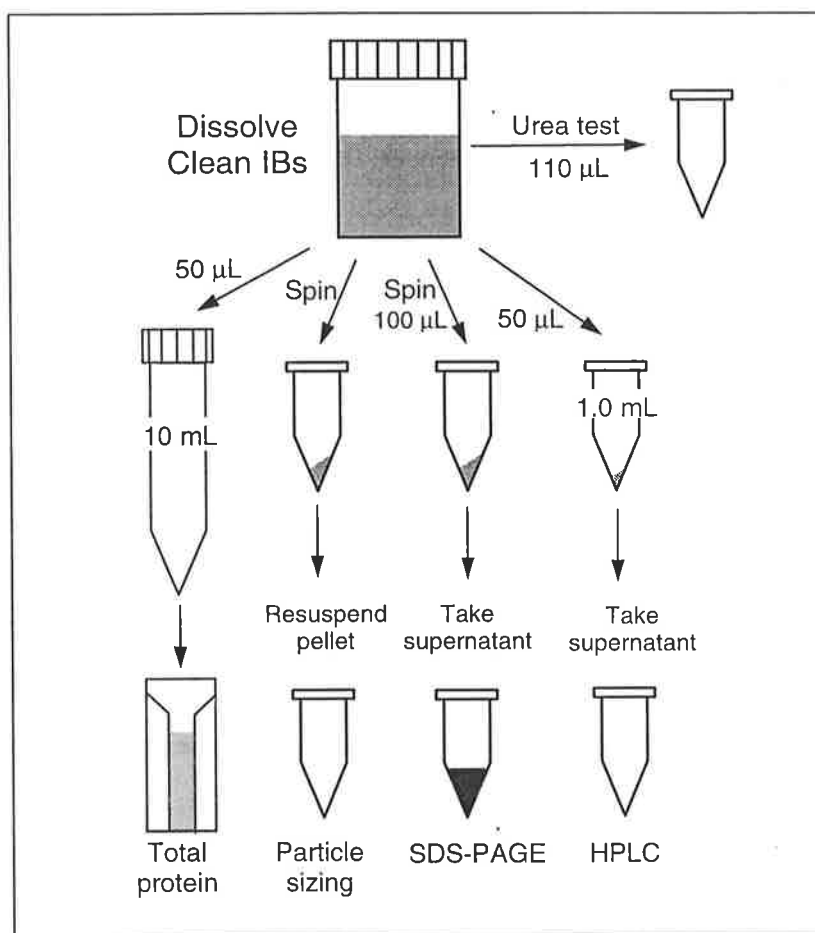


Figure 5.1: Sampling protocol for trials to identify effects of refolding in the presence of soluble cellular contaminants. Where required, samples were centrifuged at $18 \times 10^3 g_{\max}$ for 3 minutes. A 500 μL sample was pelleted and resuspended in 1000 μL of 20% ethanol/0.1% TFA assisted by ultrasonication for particle sizing.

Before refolding trials were attempted using the dirty inclusion bodies, a series of trial dissolutions were conducted to determine the amount of LongR³IGF-I they contained relative to the clean inclusion bodies. Aliquots of the inclusion body pastes that were recovered by batch and continuous centrifugation were removed from storage at -20°C and allowed to thaw at room temperature. The standard dissolution procedure (as described in section 2.6) was used to initiate dissolution and aliquots were quenched into 0.1% TFA 0.5, 1, 7, 11, and 24 hours late for analysis by HPLC. Approximately one hour before the 24 hour sample was quenched, additional DTT was added to each dissolving mixture to increase the overall concentration by 0.016 M. The sample volumes required for refolding trials using dirty inclusion body starting material were adjusted according to the findings.

The dissolution procedure for the two refolding trials using batch-pelleted inclusion bodies was the same as that employed for clean inclusion bodies. The sampling protocol was only slightly different to that for clean inclusion bodies (see Figure 5.1). Sampling for particle size analysis, total protein quantitation, and SDS-PAGE were the same as outlined above. A 1:10 dilution of 100 μL of dissolving mixture into 1.0% TFA was used to quench samples for analysis by HPLC instead of the 1:50 dilution into 0.1% TFA because the dirty inclusion bodies contained a significantly lower fraction of fusion peptide and a significant amount of contaminating material.

For both dissolution trials, additions of small amounts of 1.6 M DTT stock were made to the dissolving mixtures to replenish any reducing agent which might have become oxidised by atmospheric oxygen. Periodic additions were made to increase the overall DTT concentration by approximately 0.020 M, making allowance for the decrease in batch volume due to sampling. To test whether maximal dissolution was achieved, the method presented in section 2.6.1 was used.

5.1.1.3 Dissolution for concentrated refolding

A batch of clean inclusion bodies was removed from storage at -20°C and allowed to thaw at room temperature. The thawed paste was then dissolved using the standard procedure outlined in section 2.6. Samples for analysis by HPLC were quenched by dilution of 20 μL of the dissolving inclusion body suspension into 1000 μL of 0.1% TFA. After vortexing for 15 s, the samples were allowed to stand at room temperature for at least five minutes before being centrifuged at $18 \times 10^3 g_{\text{max}}$ for three minutes. Supernatants were carefully decanted and analysed to determine LongR³IGF-I content. Triplicate samples for total protein quantitation were quenched by dilution of 50 μL aliquots of dissolving mixture into 10 mL of 0.1 M NaCl/0.1% TFA and vortexed for 15 s. Reducing and non-reducing samples were taken for SDS-PAGE by dilution of 20 μL of dissolving mixture into 60 μL of 1.5 $\times$ loading buffer following the procedure described in section 2.2.5. After dissolution had progress for 30 minutes, two 2500 μL samples were taken to initiate refolding.

5.1.1.4 Dissolution for fed-batch refolding

A batch of clean inclusion bodies was removed from storage at -20°C and allowed to thaw at room temperature. Dissolution of the inclusion bodies was initiated using the routine procedure outlined in section 2.6. Sampling of the dissolution mixture was started five minutes after adding the dissolution buffer to the inclusion body slurry. Progress of dissolution was monitored using SDS-PAGE (under both reducing and non-reducing conditions), total soluble protein measurements, and HPLC analysis to determine LongR³IGF-I concentration.

SDS-PAGE samples were quenched by dilution of 20 µL of reacting mixture into 60 µL of 1.5× load buffer. Triplicate 20 µL samples were quenched into 10.0 mL of 0.1 M NaCl/0.1% TFA for total protein quantification. 20 µL samples of dissolving mixture were quenched into 1000 µL of 0.1% TFA for analysis by HPLC.

5.1.2 Results and discussion - dissolution

The observations and results of dissolution trials are presented in two parts: dissolution of clean inclusion bodies and dissolution of dirty inclusion bodies. This was done because, although the materials used were from the same fermentation batch, their dissolution behaviour was quite disparate.

5.1.2.1 Clean inclusion bodies

In all refolding trials, addition of water to the clean inclusion body paste yielded a homogeneous slurry. Addition of the bright, clear, colourless 8.8 M urea dissolution buffer to the slurry initially gave an opaque, brown suspension. Within a few minutes the suspension had changed to a relatively clear solution, dark brown in colour. The appearance had not changed at 30 minutes, when samples were taken to initiate refolding. The appearance did altered somewhat during further incubation, becoming lighter in colour. After incubating for 10 hours it was only slightly opaque and light brown in colour. No further noticeable changes occurred if the preparation was left even longer. Addition of

1.6 M DTT to replenish potentially air-oxidised DTT and blanketing with nitrogen gas did not alter the physical appearance of the dissolving mixture with time.

The rapid initial change in opacity and colour indicated that inclusion bodies dissolved quickly (within the first five minutes). The subsequent lightening in colour and slight increase in opacity presumably indicated that some form of chemical modification had occurred, possibly due to oxidation or aggregation of some of the contaminants.

Measurement of the total soluble protein concentration in the dissolving mixture revealed that the amount of dissolved protein was reproducible between dissolutions and with time. The average overall protein concentration was $13.8 \pm 1.8 \text{ g L}^{-1}$ ($0.90 \geq P$), and the results are presented graphically in Figure 5.2. The consistency of the values obtained for total protein concentrations demonstrated that the standard dissolution method was quite

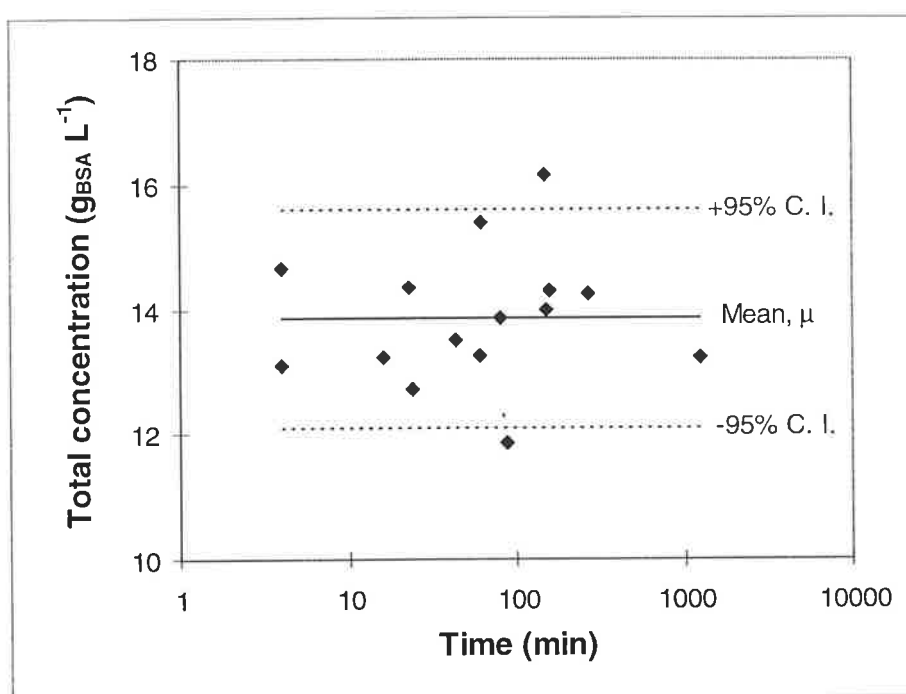


Figure 5.2: Plot of total protein concentration versus time for dissolution of clean inclusion bodies using the standard method. Each point represents the average of triplicate samples analysed in duplicate using the method described in section 2.2.2. Confidence bands for $\pm 95\%$ of the mean are shown, and were calculated by assuming a normal distribution.

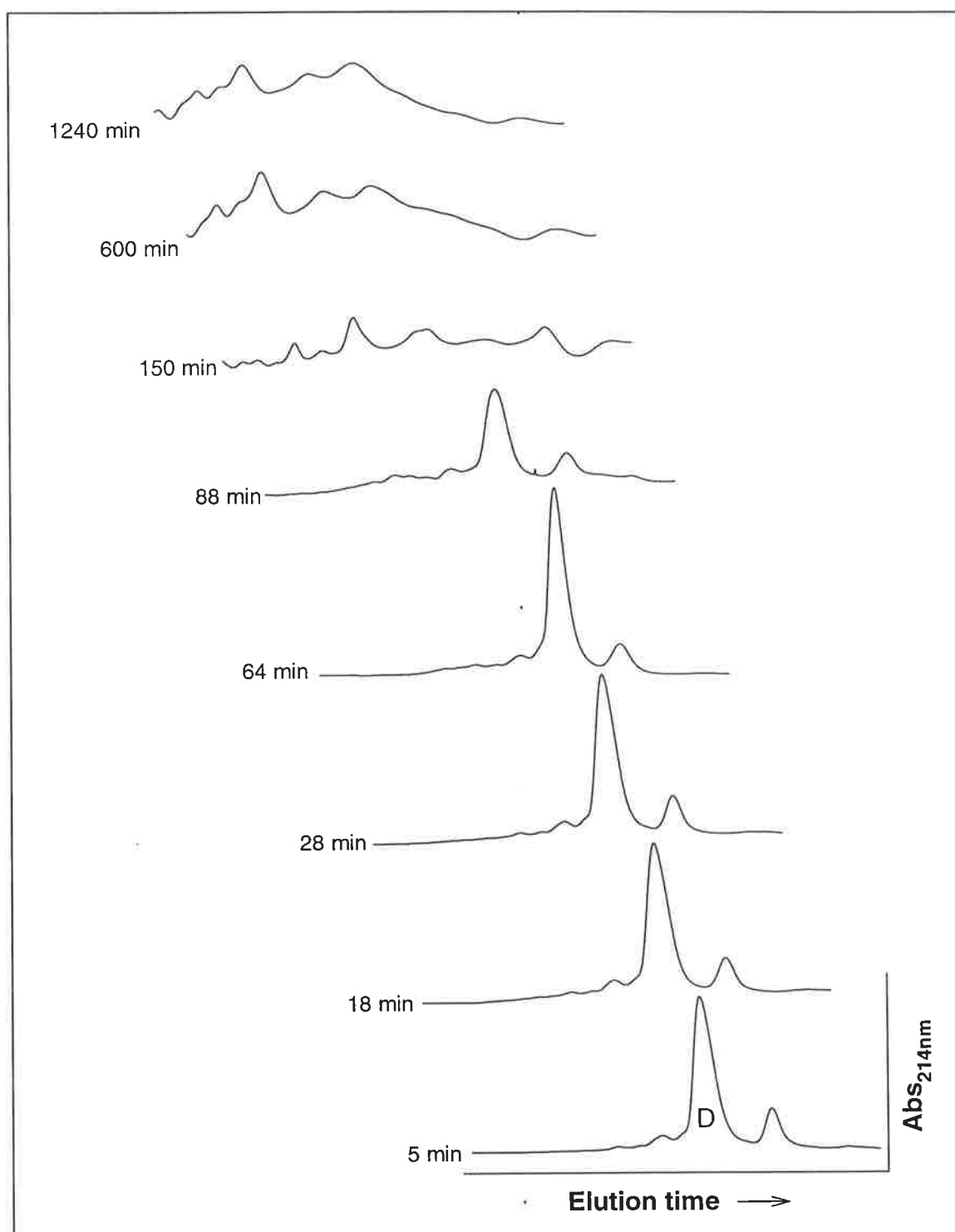


Figure 5.3: A typical series of HPLC profiles for dissolution of homogenous clean inclusion body slurry in 8M urea/0.1M Tris-HCl pH = 9.1/0.04 M Gly/0.01 M EDTA/ 0.016 M DTT. The peak corresponding to elution of reduced, denatured LongR³IGF-I is indicated by the D in profile of the five minute sample. Profiles show the species eluted from 15 to 28 minutes. All profiles have been normalised to account for the volumes and dilutions tested. The absorbance scale for 150, 600 and 1240 minute profiles have been magnified by a factor of three to emphasise the absence of denatured LongR³IGF-I.

reproducible. The relatively constant overall concentration of protein indicated that the change in physical appearance of the dissolving suspensions of inclusion bodies was not caused by a significant reduction in the total amount of soluble protein. This was further supported by the observation that the size of pellets collected from the samples for HPLC analysis after centrifugation did not appear to change with time.

HPLC analysis of samples to quantify LongR³IGF-I concentration revealed that the total amount of dissolved native-conformation protein was reproducible in all trials, but varied dramatically with time. The average amount of denatured LongR³IGF-I present after 30 minutes in all the samples tested was $8.4 \pm 1.7 \text{ g L}^{-1}$ ($0.90 \geq P$). The greater variance in LongR³IGF-I concentration, relative to total protein concentration, may be caused by heterogeneity in the starting material. However, the variance between trials was taken into account when estimating rate constants by using the measured value of the starting concentration in regression calculations.

A typical series of HPLC profiles is presented in Figure 5.3. Near-maximum dissolution was achieved within the first five minutes. After a period of approximately 60 minutes the profiles can be seen to change, with appearance of multiple unresolved leading protein species. There is a gradual decrease in the area of the reduced-denatured peak corresponding to an increase in earlier-eluting proteinaceous species. After approximately 150 minutes, reduced, denatured LongR³IGF-I cannot be clearly resolved. Further incubation results in earlier elution of most proteinaceous species.

Purging the dissolution buffer and blanketing the dissolving suspension with nitrogen gas only slightly delayed the collapse of the denatured LongR³IGF-I into multiple earlier-eluting species occurred (see section 5.1.1.1 for method). However, periodic addition of DTT in the contamination trials (section 5.1.1.2) did maintain LongR³IGF-I in a reduced, denatured conformation. Figure 5.4 shows the HPLC profiles of 150 minute samples quenched from the dissolution of clean inclusion bodies using the standard protocol and with nitrogen blanketing or addition of DTT.

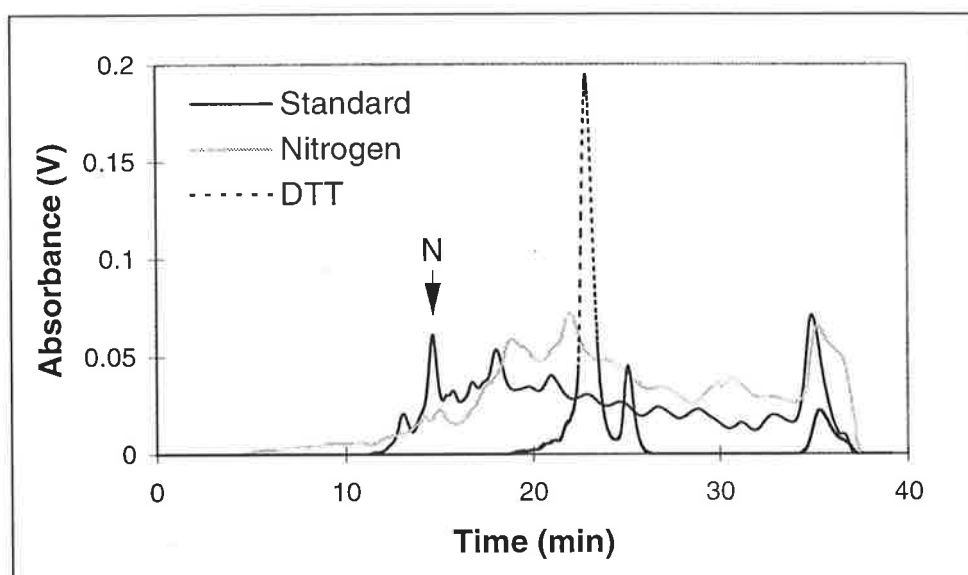


Figure 5.4: HPLC profiles of clean inclusion bodies dissolved for 150 minutes using the standard procedure or blanketed with nitrogen gas or with additional DTT added. Profiles have been corrected to allow for differences in UV-detector extinction coefficients and injected volume. The elution time of native-conformation protein is indicated as N.

For denatured, reduced LongR³IGF-I to assume its native conformation, the internal disulfide bonds must form (Milner 1996), which requires a decrease in the denaturing potential of the chemical environment. The denaturing environment employed in these dissolution trials relied on the presence of 8 M urea and the reducing power of DTT to maintain the thiols groups of the protein in the free state. The trial in which DTT was added indicates that the urea is probably not the limiting factor in the loss of denaturing conditions. The 150 minute sample profile generated using the standard conditions (see Figure 5.4) is typical of the trials with a significant amount of protein eluting at the time which corresponded to native-conformation protein. It is quite likely that this was LongR³IGF-I in its native conformation, indicating that the internal disulfide bonds were able to form. Therefore, the most likely explanation of the observed instability is that the DTT in the dissolution buffer was being depleted by oxidation.

To test this hypothesis, DTT was added to mixtures that were known to have collapsed into multiple folding forms. This experiment was conducted as part of the comparison of clean and dirty inclusion bodies (see section 5.1.1.2). After dissolving under the standard chemical conditions for eleven hours, the solutions of both clean and dirty inclusion bodies gave the typical multiple species profiles as shown in Figure 5.4. After dissolving for 23

hours, DTT was added to a final concentration of 0.016 M and at 24 hours the HPLC profiles of all mixtures appeared virtually identical to the profiles analysed after one hour. This provided conclusive evidence that loss of the reducing environment was the destabilising force that yielded the multiple protein species resolved by HPLC.

The attempt to eliminate oxidation by atmospheric oxygen using a nitrogen gas was not as effective as the addition of extra reducing agent. This may be a result of incomplete exclusion of oxygen using nitrogen gas, or an alternative oxidation mechanism that does not involve exposure to the atmosphere.

For the purposes of providing consistent dissolved material for fed-batch refolding trials, DTT could not be added periodically to the batch of dissolving inclusion bodies. Additional DTT would maintain the desired reduced, denatured state, but the concentration of reduced and oxidised disulfide bond shuffling promoters could not be maintained at a constant value. Since the concentration of reduced thiols in the mixture would be unknown, the correct ratio of oxidising agent could not be determined for reproducible refolding conditions. The use of a redox probe to control a dissolution reactor at a constant reducing potential is questionable as the ratio of reduced:oxidised thiols in the mixture would be altered. Exclusion of atmospheric oxygen did not eliminate instability so an alternative must be found.

Dissolution of inclusion bodies and incubation in a PFR should provide stable dissolved starting material, but the logistical difficulties of this method made it unrealistic to use this approach. Evidence of instability was only observed after the first 60 minutes of dissolution. Therefore, the pragmatic solution employed to provide consistent starting material for fed-batch trials was to dissolve inclusion bodies in a single batch and feed for no longer than 60 minutes. Feeding was initiated after dissolution had progressed for no longer than ten minutes. The HPLC profiles for samples of the dissolved for one of the trials are presented in Figure 5.5. The profiles for the second trial were essentially identical to those obtained for the first trial.

Quantification of the amount of denatured protein using PeakFit™ (section 2.3.2) to integrate each of the profiles is presented in Figure 5.6. The average concentration of reduced, fully denatured LongR³IGF-I during the first 64 minutes was $7.7 \pm 0.8 \text{ g L}^{-1}$ ($0.90 \geq P$). This gradually reduced as the protein became increasingly unstable. At 88

minutes, less than half of the protein remained in the fully reduced, denatured state. In the second trial, the average concentration of reduced, fully denatured LongR³IGF-I was $8.2 \pm 0.8 \text{ g L}^{-1}$ ($P \geq 0.90$) for the first 64 minutes and again decreased thereafter.

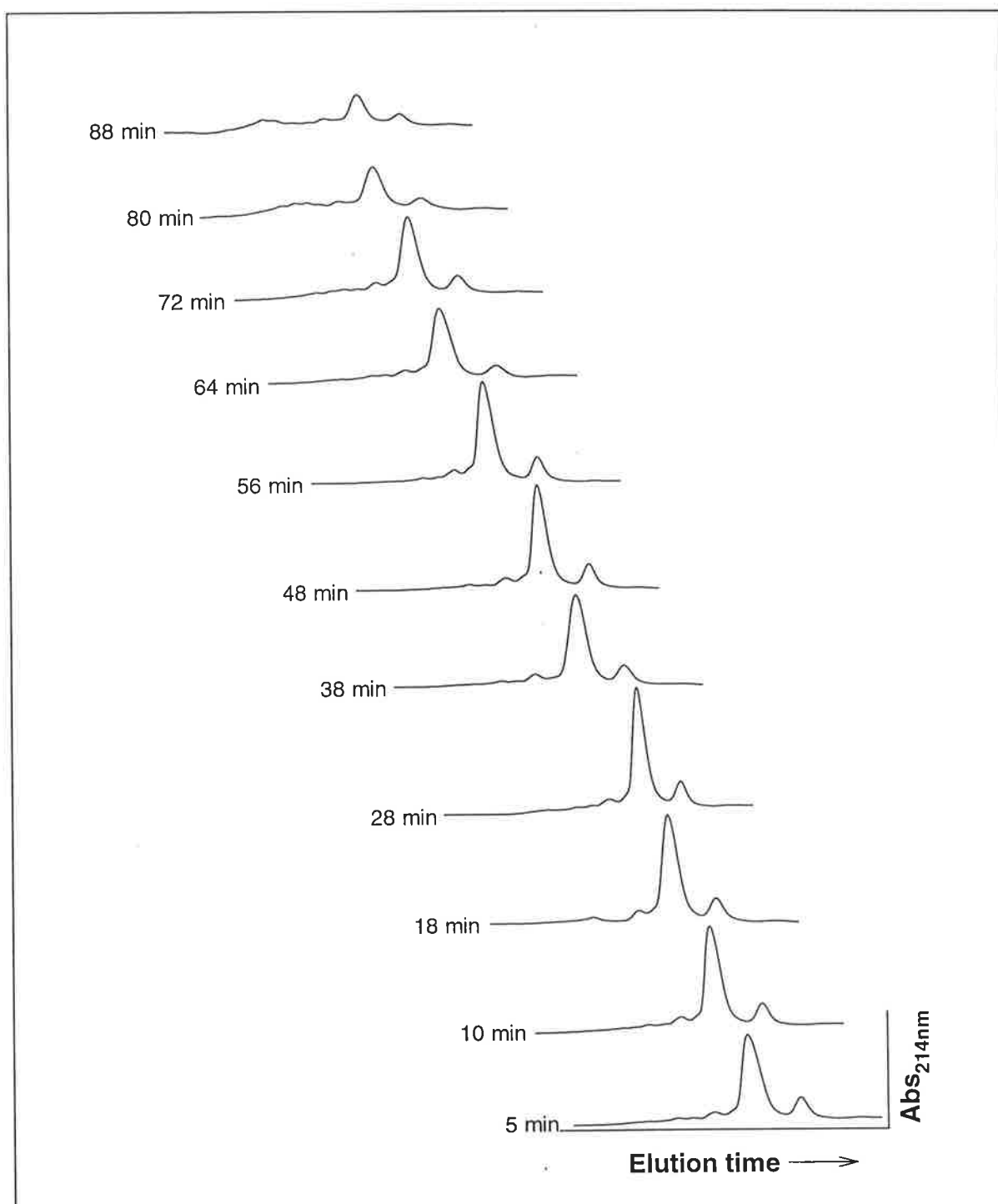


Figure 5.5: HPLC profiles for the dissolving starting material taken during fed-batch refolding. The profiles are corrected to give equivalent loading and scale. The denatured protein shows minor signs of instability after 64 minutes.

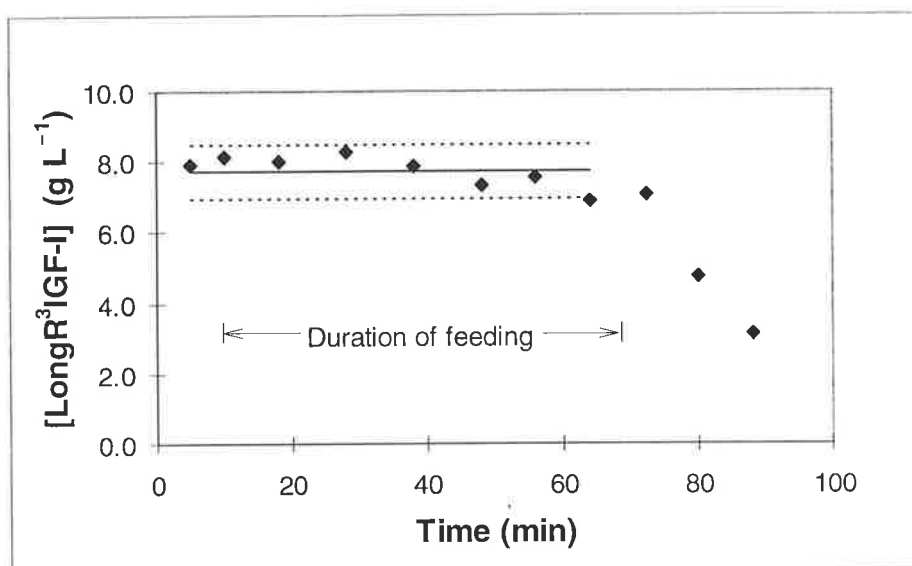


Figure 5.6: Concentration of reduced, denatured LongR³IGF-I with time when dissolved using the standard method. The solid line represents the average concentration during the first 64 minutes and the dashed lines the 95% C.I. bands. Feeding into refolding conditions commenced after 8 minutes of dissolution and was continued for 60 minutes.

For all the refolding trials of clean inclusion bodies where water was used as the solvent, the concentration of protein dissolved was not limited by the addition of dissolution buffer at 10 g L^{-1} . The results of both total protein quantitation and HPLC measurement of LongR³IGF-I concentration were in agreement; a 1:4 dilution into additional dissolution buffer resulted in a 25% reduction in both quantities.

When dissolution was conducted using batch supernatant (see section 2.4) to prepare the solvent for dissolution buffer, both total protein quantitation and HPLC measurement of LongR³IGF-I concentration indicated that dissolution was limited: 1:4 dilution into dissolution buffer resulted in a 32% dilution of the total protein concentration and a 38% dilution of the concentration of denatured protein concentration as measured by HPLC. The total protein concentration was below the average obtained for dissolution samples using water as the solvent ($P < 0.9995$) at $12.1 \pm 1.0 \text{ g L}^{-1}$. These results suggested that the presence of the soluble contaminants in the batch supernatant reduced the amount of urea that was available to dissolve the insoluble material. Furthermore, it is possible that a significant amount of these species were not proteinaceous because the total protein concentration was reduced relative to dissolution using pure water as the solvent. These observations provided circumstantial evidence that urea preferentially dissolved non-inclusion body species.

Unfortunately, the procedure used to quench samples for particle size analysis appeared to cause extensive aggregation of undissolved material, rather than allowing resolution of undissolved or partially-dissolved inclusion bodies. A typical set of particle size analysis results are shown in Figure 5.7, for samples quenched at different times during dissolution of clean inclusion bodies. No peaks at approximately 0.4 μm or lower could be resolved. A single broad peak from 0.5 to 3 μm was obtained instead. This peak was not present in homogenate and centrifugation supernatant samples (see section 2.5).

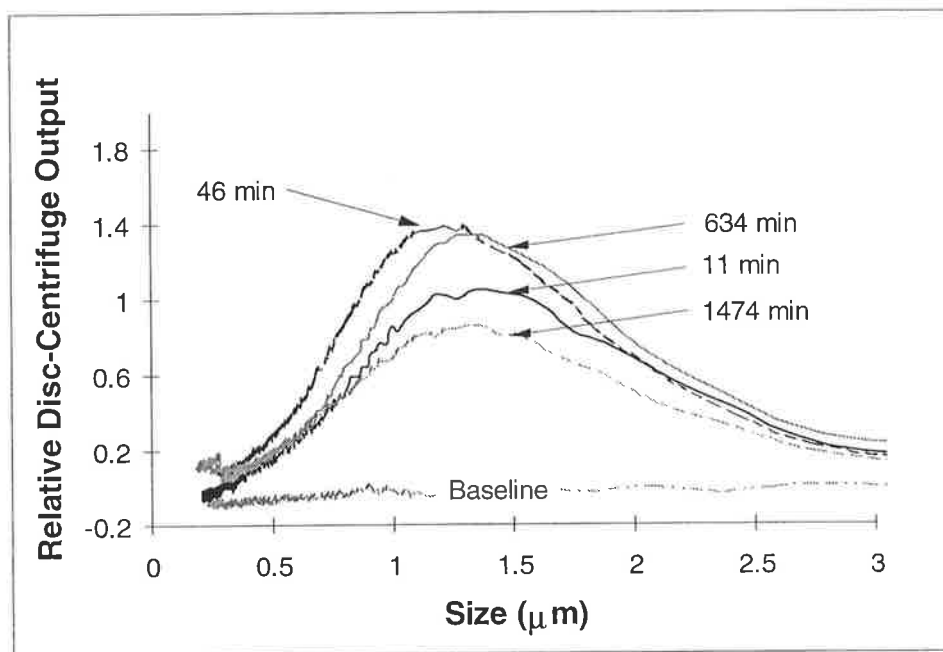


Figure 5.7: Particle size analysis of clean inclusion bodies dissolved for different times using the DCF4 analytical disc centrifuge. All samples are at equivalent concentration. The profiles indicate that the insoluble material had aggregated because no undissolved inclusion bodies are resolved.

SDS-PAGE analysis of the dissolving suspensions did not provide significant evidence of change in protein content with time. The resolved proteins gave a consistent pattern for each trial. A slight difference was observed between corresponding reducing and non-reducing samples, but again independent of the time for which dissolution had been progressing. A typical example of dissolved protein analysed on a 10-20% Novex™ Tricine gel is presented in Figure 5.8. Lanes 2 to 5 and 6 to 9 show samples from two different dissolution trials. Running equivalent samples under reducing conditions decreased the amount of 18 and 27 kDa species and increased the amount of protein corresponding to the molecular weight of LongR³IGF-I.

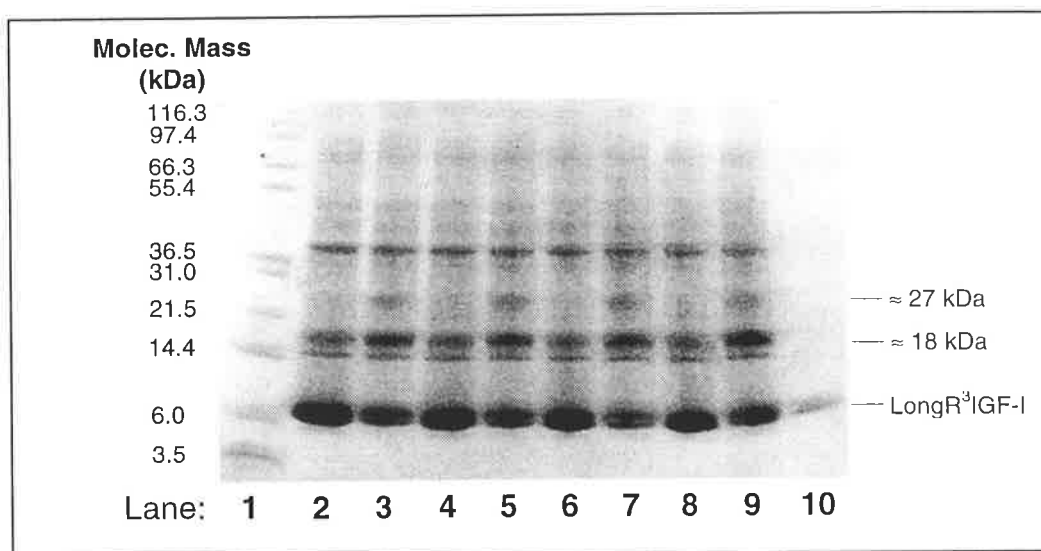


Figure 5.8: SDS-PAGE of dissolving clean inclusion body mixtures (10-20% Novex™ Tricine gel). Lane 1: Mark12 wide range protein standard, Lane 2: 16 min. (reducing), Lane 3: 16 min (non-reducing), Lane 4: 70 min. (reducing), Lane 5: 70 min (non-reducing), Lane 6: 15 min. (reducing), Lane 7: 15 min (non-reducing), Lane 8: 70 min. (reducing), Lane 9: 70 min (non-reducing), Lane 10: LongR³IGF-I standard.

5.1.2.2 Dirty inclusion bodies

As outlined in the previous section, analysis of samples obtained during dissolution of different aliquots of clean inclusion bodies gave good reproducibility. Once initiated, total dissolution was achieved rapidly, was stable for approximately an hour and then became unstable due to changes in the reducing potential of the buffer. Dissolution was not complete when batch supernatant was used as the solvent. The behaviour of dirty inclusion bodies in similar trials was quite different.

The addition of 1.0 mL g⁻¹ of water before addition of dissolution buffer gave a homogeneous slurry only after extensive mixing and homogenisation with an autopipette to break up aggregates. Addition of the dissolution buffer to the slurry gave a medium brown cloudy mixture, free from visible particles. The opaque appearance of the dissolving mixture did not change during incubation at room temperature for up to 24 hours.

Analysis of the dissolving inclusion body mixture for total protein content revealed that the total amount of dissolved protein increased significantly with time. The measured total protein concentrations of samples taken during dissolution of dirty inclusion bodies using water and batch supernatant are presented in Figure 5.9.

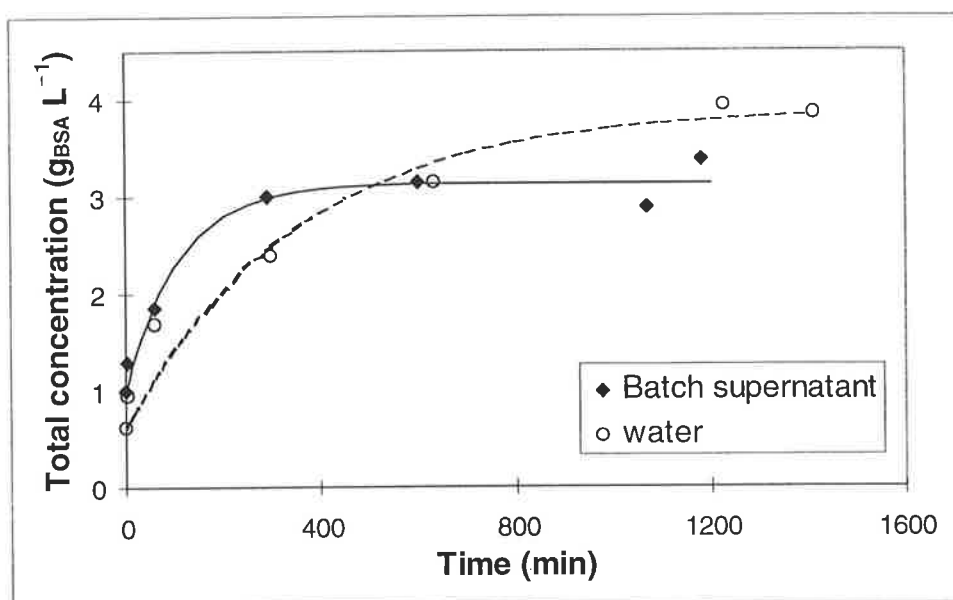


Figure 5.9: Total soluble protein concentration versus time for dissolution using both water and batch supernatant as the solvent. Solid and dashed lines are the regressed curves for trials using batch supernatant and water respectively.

A 1:4 dilution of the dissolving inclusion body suspensions into additional dissolution buffer gave an approximately three-fold increase in the amount of protein dissolved, indicating that the standard 10 mL of buffer per gram of wet-weight of paste was limiting the amount of protein that could be dissolved. Dissolution of clean inclusion bodies using the same method recovered approximately 14 g L^{-1} , as described above. Therefore, it can be concluded that the dissolution buffer was being saturated due to urea interacting with the additional non-protein species present in the batch-centrifuged inclusion bodies. The experimental data were regressed to a first-order model of dissolution of the form:

$$\text{Bulk concentration} = \text{Saturation concentration} \times (1 - e^{-K_c \cdot t}) \quad (5.1)$$

where K_c is an apparent mass-transfer coefficient. This model describes the bulk fluid concentration of dissolving spheres which is limited due to saturation of the fluid (see Appendix 9.3). Regressed curves for each of the data sets are presented with the raw data in Figure 5.9.

HPLC analysis of the dissolving mixtures revealed that the total amount of dissolved native-conformation LongR³IGF-I was similar in the two types of dissolutions and that the total protein concentration increased with time. The average amount of denatured, reduced LongR³IGF-I present after 30 minutes was only approximately 0.2 g L^{-1} (approximately

2% of the amount recovered from clean inclusion bodies). The concentration of LongR³IGF-I present in the dissolving mixtures with time is presented in Figure 5.10, along with corresponding curves regressed to Equation 5.1.

The regressed estimates for apparent mass-transfer coefficient and saturation concentration of the two dissolution trials are presented in Table 5.1. A consistent difference between the two trials can be seen in the saturation concentrations of total protein and LongR³IGF-I. When batch supernatant was used as the solvent, the saturation concentration measured by both assays was lower than when using water. Since the batch supernatant must have contained contaminants which formed complexes with urea, it is not surprising that less LongR³IGF-I dissolved. However, despite having an initial concentration of 1.4 g L⁻¹ in the batch supernatant, the final concentration of protein was lower than when water was used. This indicates that non-protein contaminants were dissolved preferentially, leaving less urea available for dissolving the proteins.

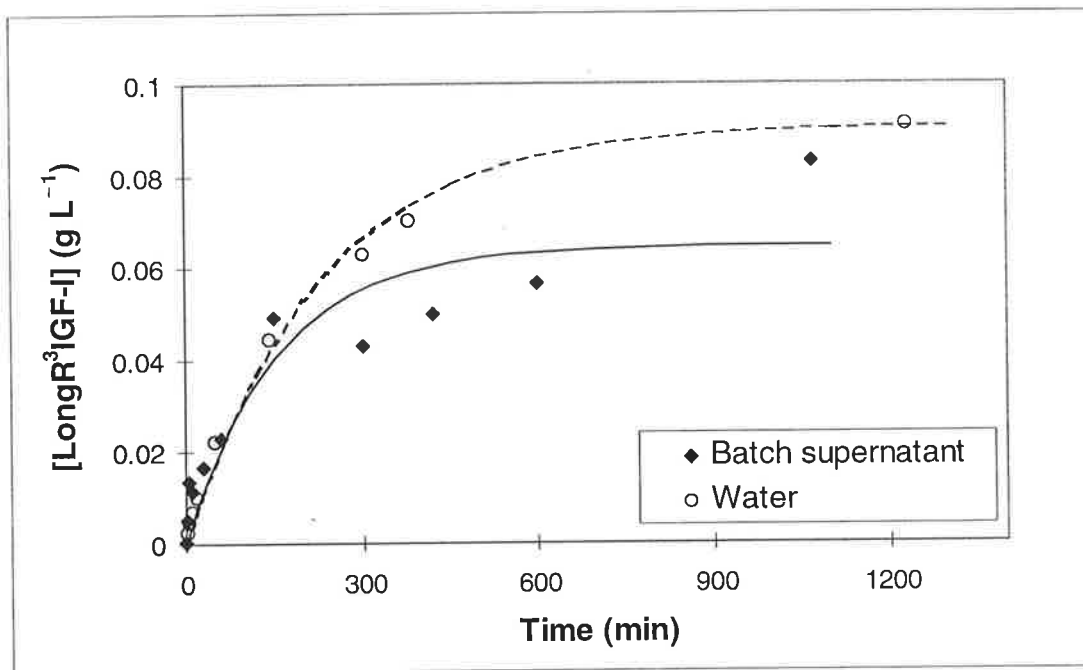


Figure 5.10: Concentration of LongR³IGF-I determined by HPLC during the mass-transfer limited dissolution of dirty inclusion bodies. Solid and dashed lines are the regressed curves for trials using batch supernatant and water respectively.

Table 5.1: Regressed estimates for apparent mass-transfer coefficients and saturation concentrations for dissolution limited by saturation of the bulk fluid.

	Batch supernatant		Water	
Parameter	[LongR ³ IGF-I] by HPLC	Total protein	[LongR ³ IGF-I] by HPLC	Total protein
K_c (min ⁻¹)	6.5×10^{-3}	9.2×10^{-3}	4.3×10^{-3}	2.8×10^{-3}
C_{Dsat} (g L ⁻¹)	0.065	3.1	0.091	3.9

The apparent mass-transfer coefficients determined for each of the dissolution trials indicate that the rate of dissolution was quite similar in magnitude, but slightly slower when water was used as the solvent. The slower rate of dissolution in water was confirmed by both the total soluble protein measurements and the concentration of LongR³IGF-I by HPLC assay.

Analysis of the dissolving mixtures using the method presented in section 2.2.1. to determine the particle size distribution gave similar results to those obtained with clean inclusion bodies. Rather than resolving undissolved or partially dissolved inclusion bodies, the size distribution profiles showed a single broad peak ranging from 0.5 to 3 μ m. A greater dilution was required to give peaks of the same amplitude, indicating that these quenched samples contained approximately ten times more solid material. It can be concluded that sedimentation of the solid material and resuspension in 20% ethanol/0.1% TFA caused aggregation of the material present, preventing any insight into the progress of dissolution of the inclusion bodies.

5.2 MIXING EFFECTS

To determine whether mixing rate had a significant effect on refolding, a series of mixing experiments was conducted. The aim was to determine the rate constants of refolding for three different mixing rates and to test their equivalence statistically.

5.2.1 Experimental protocol - mixing

Six refolding trials were conducted to investigate different mixing conditions, namely three stirring rates with a repeat of each. The method used to dissolve the inclusion bodies is presented in section 5.1.1.1. After dissolution had been allowed to progress for 30 minutes a 600 μL sample was diluted to a final volume of 15.0 mL of 2 M urea/0.1M Tris-HCl pH = 9.1/0.0016 M Gly/10 mM EDTA/ 6.4×10^{-4} M DTT/0.0016 M OX- β -ME with a single rapid injection into a solution of the appropriate composition. The progress of refolding was monitored by quenching small samples at various times and conducting HPLC and total soluble protein analysis on them. Samples for analysis by HPLC were quenched by dilution of 300 μL of refolding mixture into 700 μL of 1.5% TFA. After vortexing for 15 s, they were allowed to stand at room temperature for at least 5 minutes before being centrifuged at 18×10^3 g_{max} for 3 minutes. Supernatants were carefully decanted and analysed to determine LongR³IGF-I content. Samples for total soluble protein quantitation were quenched in triplicate by dilution of 100 μL of refolding mixture into 10 mL of 0.1 M NaCl/0.1% TFA and vortexing for 15 s.

The different mixing rates were established by controlling the frequency of rotation of a 1.0 cm magnetic flea immersed in the solutions. The flea in the test solutions was driven by a magnetic stirrer. Minimum stirring conditions were taken as total absence of mixing. The dissolved protein was injected into the appropriate diluent buffer, gently swirled for 5 s and then allowed to stand on a magnetic stirrer for the duration of refolding. 'Mild mixing' conditions were taken as the minimum speed of the magnetic stirrer (approximately 100 r.p.m.). The 600 μL sample of dissolving mixture was injected into stirred diluent buffer to initiate refolding. For 'maximum mixing' conditions, the speed of the magnetic stirrer was adjusted to provide the maximum stable speed of the flea rotating in the diluent buffer

(approximately 500 r.p.m.). As in the case of mild mixing, refolding was initiated by dilution of the 600 μ L sample of dissolving mixture into the already mixing diluent buffer. An outline of the experimental protocol is presented graphically in Figure 5.11.

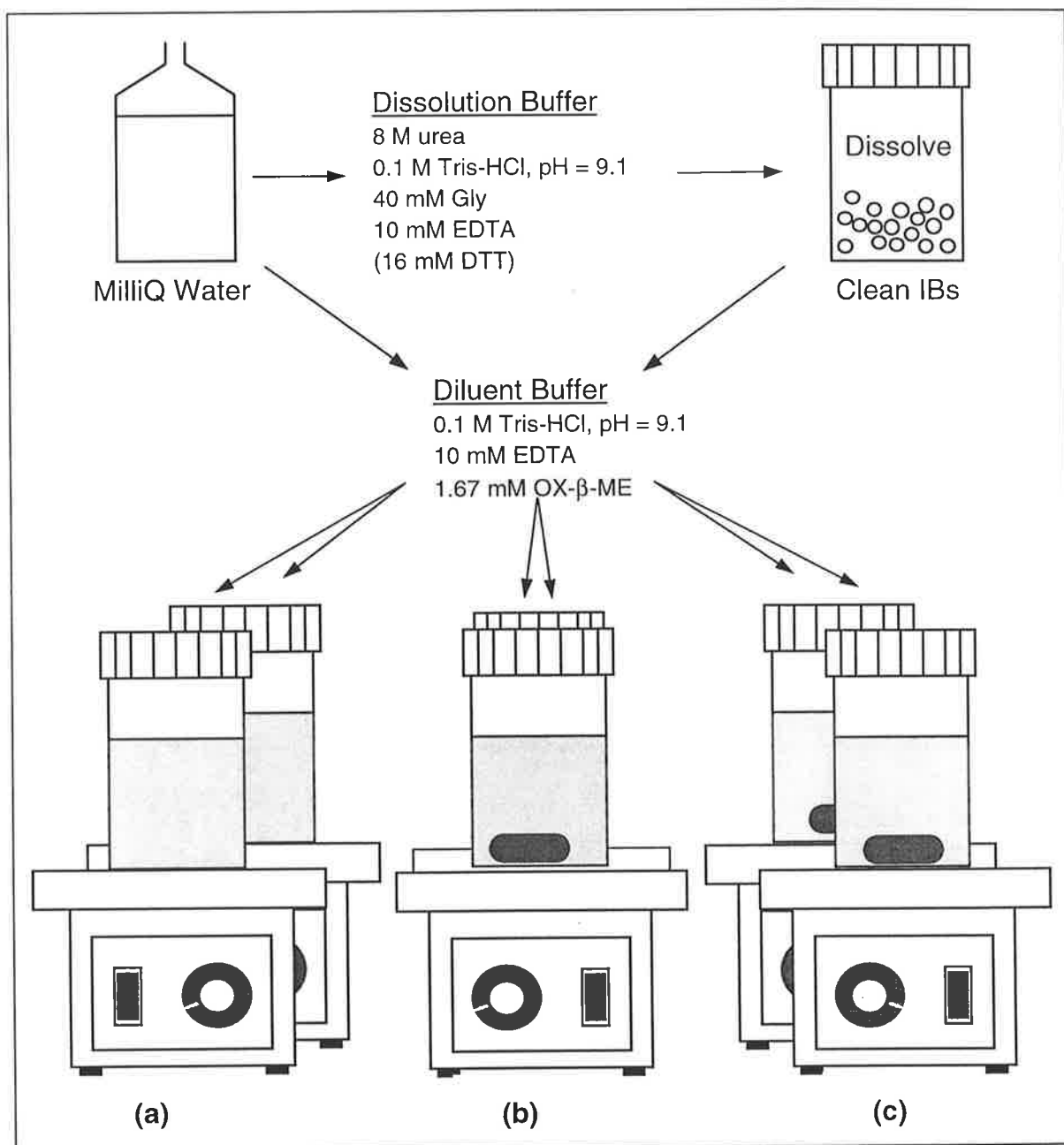


Figure 5.11: Diagram of the protocol used to identify effects of different mixing rates during refolding. Clean inclusion bodies and protein-free water were used for all six trials (a) No mixing (b) stirrer set to minimum (c) stirrer at maximum stable speed. Dissolution and refolding mixtures were assayed for total protein concentration and LongR³IGF-I by HPLC. The refolding mixtures were assayed repeatedly with time.

5.2.2 Results and discussion - mixing

As presented in section 5.1.2, the dissolution procedure used to dissolve clean inclusion body material was reproducible. When 600 μL of the slightly opaque, dark brown dissolving inclusion body mixture were injected into 14.4 mL of the bright, clear colourless refolding buffer, the mixture immediately assumed a slightly opaque, pale straw colour. The appearance of the mixture did not change over time under any of the three stirring conditions. However, the volume of solid material sedimented from samples quenched for analysis by HPLC did show a clear trend with time. The size of the pale, off-white coloured pellets started to reduce over the first hour of refolding and subsequently increased in size as refolding was allowed to proceed further. The minimum pellet was obtained from samples quenched after refolding for approximately one hour under all three mixing conditions. The pellets were dried, resuspended in 20 μL of reducing load buffer, and analysed by SDS-PAGE (see section 2.2.6). An example of the SDS-PAGE separation of the recovered insoluble material for a time-course of refolding samples is presented in Figure 5.12.

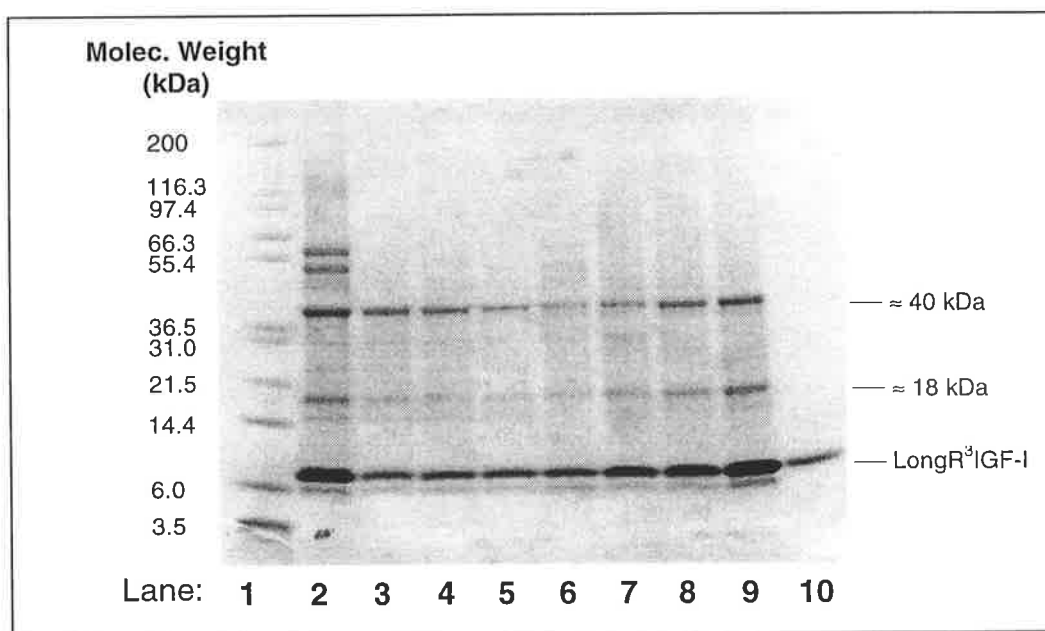


Figure 5.12: Reducing SDS-PAGE of material precipitated during refolding (10-20% Novex™ Tricine gel). *Lane 1:* Mark12 wide range protein standard, *Lanes 2-9:* 10 μL of refolding samples quenched at 0.50, 2.0, 7.0, 40.0, 150.0, 420.0, 600, and 1080 minutes after the start of refolding, *Lane 10:* LongR³IGF-I standard.

The results from SDS-PAGE analysis indicate that the total amount of insoluble protein recovered from the HPLC-quenched samples followed the same trend as the overall physical appearance of the pellets. Since each pellet was resuspended in the same volume of load buffer and the same volume was loaded into each lane of the gel, the relative intensity of stained bands can be compared directly. The samples quenched after only 30 seconds of refolding resulted in extensive precipitation. The amount of insoluble protein recovered in subsequent samples decreased for approximately 150 minutes and then increased. After 18 hours, the amount of insoluble material was approximately the same as in the initial 0.50 min sample.

The SDS-PAGE analysis demonstrated that although many protein species were present in the insoluble material, LongR³IGF-I predominated in the refolding mixtures. However, the approach used did not establish whether the insolubility occurred under the refolding conditions, or as an artefact of the acid-quenching operation. With the exception of the 0.50 min sample, where a significant amount of material was resolved at approximately 50 and 60 kDa, the ratio of product to other contaminants remained relatively constant throughout refolding. This suggested that the insolubility was likely to be due to aggregation under refolding conditions, or at the very least that the proteins were changing state and hence solubility in acidic solution. It cannot be definitively concluded whether the proteins were restricted to self-association or aggregated due to inter-molecular association. The consistent ratio of product, 18 kDa, and 40 kDa species is compatible with non-specific association during refolding as it is most unlikely that three different species would have the same transient acid-insolubility behaviour. The band at 18 kDa may be a dimer of LongR³IGF-I that was not dissociated in reducing SDS-load buffer, and the 40 kDa species is likely to be an outer membrane protein.

The total protein concentration present in the six refolding trials was quite consistent within the range of 0.60 to 0.75 g L⁻¹. This was to be expected given the close agreement between dissolution of batches of clean inclusion bodies using the standard method. There was no significant trend in the overall protein concentration over time for any of the refolding trials. While the concentrations did vary between some consecutive time points, there was no pattern to the differences observed. The variance in the estimates of total protein concentration was constant between trials, with the relative magnitude of the standard deviation approximately 5% of the mean at all points.

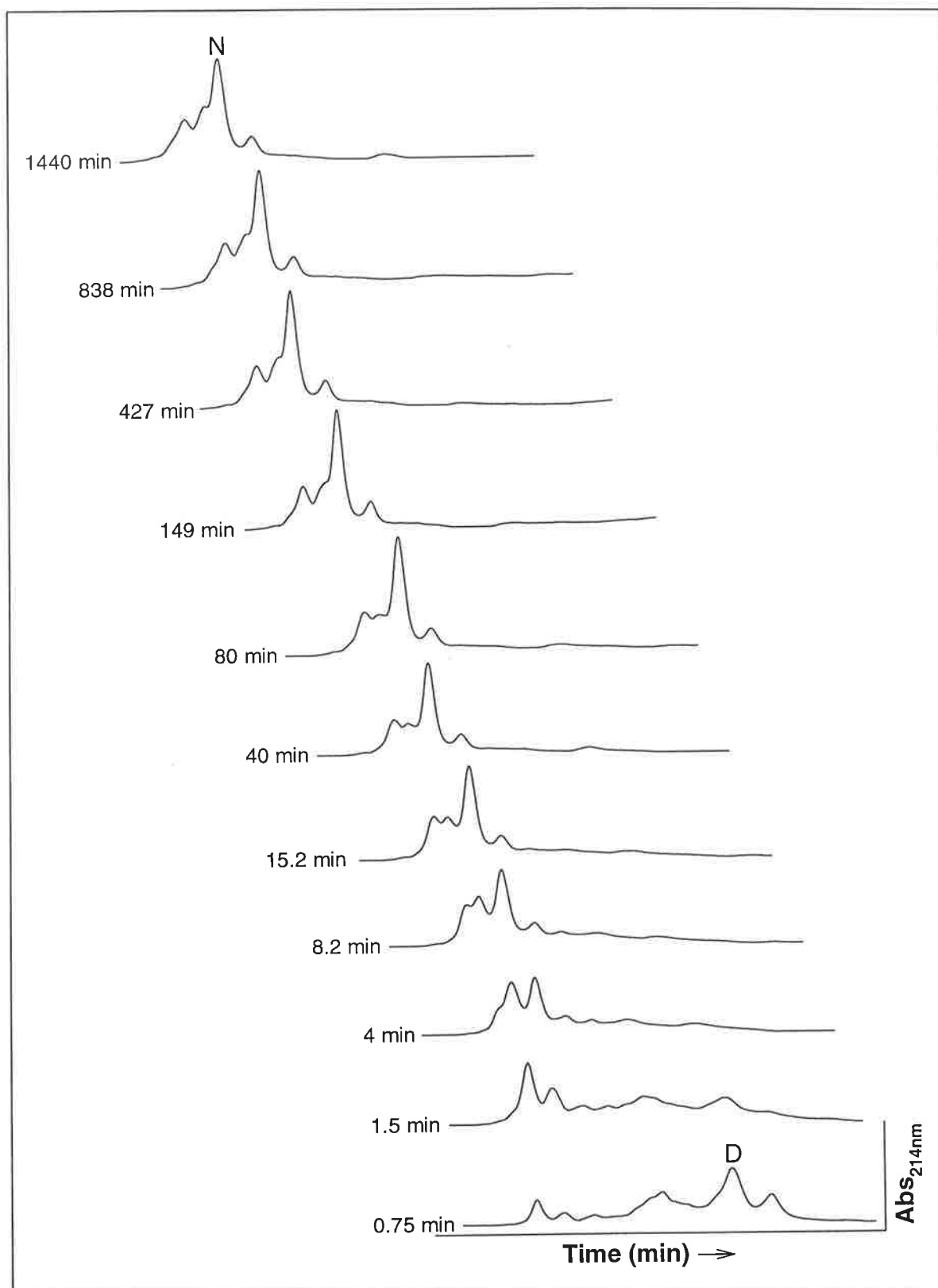


Figure 5.13: A typical series of HPLC profiles for refolding of clean inclusion body starting material without stirring. The peak corresponding to elution of reduced, denatured LongR³IGF-I (D) is indicated in the profile of the 0.75 min sample. Native species (N) is the predominant peak from 8.2 min onwards. Profiles are presented for an elution time of 9 to 28 minutes. All profiles have been normalised to account for the actual volume and dilution analysed.

A decrease in the overall concentration of soluble protein could be readily explained on the basis of the amount of insoluble material recovered from acid-quenched samples. Constant total soluble protein concentration indicated that the amount of protein that aggregated was not significant relative to the overall concentration. Further evidence of that aggregation was insignificant during prolonged incubation was provided by HPLC analysis.

A typical series of HPLC profiles of samples taken during the course of refolding without mixing over 24 hours is presented in Figure 5.13. After 0.75 min, a significant amount of the LongR³IGF-I eluted at the same time as denatured, reduced product, but rapid refolding occurred thereafter since denatured product was not detected at 4 min. The first samples revealed that a rapidly-formed intermediate eluted early in the gradient, followed by many poorly resolved species. As refolding progressed, the amount of rapidly-formed intermediate decreased and native species accumulated. After approximately 80 to 150 minutes, the amount of native protein reached a maximum. During further incubation a leading-shoulder appeared before the native LongR³IGF-I eluted. The appearance of the shoulder was accompanied by a slight decrease in the amount of native species, suggesting that slow structural rearrangement of the native species occurred. The trends observed in the HPLC profiles for the duplicate experiments with minimal, mild, and maximum mixing were consistent with those presented in Figure 5.13, accordingly these data are not shown.

Rapid appearance of a transient intermediate during the refolding of highly purified LongR³IGF-I has been reported by Milner *et al.* (1995). It has been characterised and is known to contain a native-like conformation with two disulfide bonds formed by the [18-61] and [6-48] cysteine pairings (Milner, 1996). The behaviour of LongR³IGF-I during long-term incubation under refolding conditions has not been examined previously. Milner *et al.* (1995) observed that purified LongR³IGF-I reached a final equilibrium distribution of isomers after approximately 100 minutes under refolding conditions, a result consistent with the detection of maximum amount of native protein at approximately 80 to 150 min in the present study.

The concentration of LongR³IGF-I present in each of the refolding mixtures examined in the present study is presented in Figure 5.14. The curves have been regressed to each set of data assuming the three-state refolding model, outlined in section 3.1.4, using BMDP (see section 2.3.4). Rate constants and estimates of variance for the data presented in Figure 5.14 are given in Table 5.2.

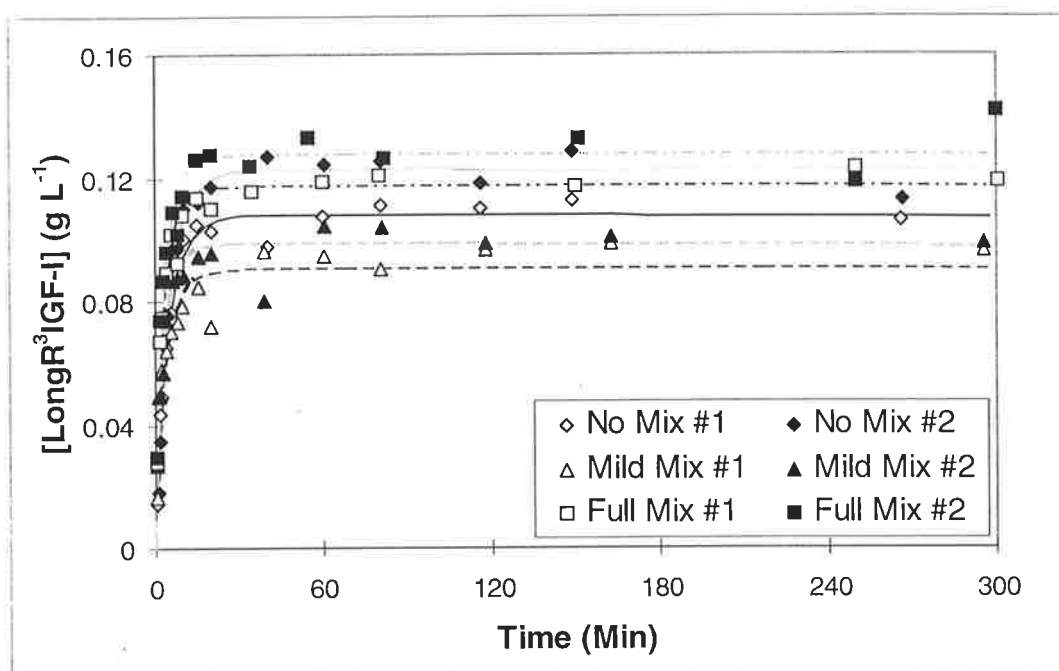


Figure 5.14: Time-course of LongR³IGF-I concentration in six refolding trials at three different mixing conditions. Dashed lines are curves for the unmixed refolding trials, solid lines for the mild mixing trials, and the dot-dash lines for the maximum mixing conditions. The curves have been regressed to each set of data assuming the three-state refolding model using BMDP (see section 2.3.4).

Table 5.2: Rate constants for the appearance of LongR³IGF-I under different mixing conditions. Estimates of rate constants and standard deviations were regressed to the three-state refolding model using BMDP under three mixing conditions.

	First repeat		Second repeat	
	Estimate	Std. Deviation	Estimate	Std. Deviation
No mixing				
k_N (min ⁻¹)	0.146	0.011	0.152	0.007
K_A (mL mg ⁻¹ min ⁻¹)	1.79	0.20	1.40	0.15
k_B (min ⁻¹)	0.0019	8.8×10^{-4}	0.0017	0.0015
Mild mixing				
k_N (min ⁻¹)	0.177	0.018	0.198	0.012
K_A (mL mg ⁻¹ min ⁻¹)	3.55	0.58	3.21	0.24
k_B (min ⁻¹)	-5.4×10^{-4}	0.0039	0.0021	7.6×10^{-4}
Full mixing				
k_N (min ⁻¹)	0.207	0.039	0.194	0.041
K_A (mL mg ⁻¹ min ⁻¹)	3.57	1.48	2.97	1.23
k_B (min ⁻¹)	-3.9×10^{-5}	0.018	-3.9×10^{-4}	0.016

The rate constants for correct refolding and aggregation have a standard deviation of 10-20% of their nominal values. However, the estimate of variance in the first-order back reaction from native to intermediate, indicated that no confidence can be placed on the rate constants themselves. Only one rate constant - k_B - was significantly different from zero.

To determine whether the results of repeat experiments were 'true repeats' and whether mixing had a significant effect on protein refolding, a nested analysis of variance was conducted. The values of transient LongR³IGF-I concentration for the six mixing trials were normalised by dividing by the nominal concentration of denatured LongR³IGF-I initially present in each mixture. The six data sets were each regressed to the three-parameter refolding model, effectively giving an 18-parameter description of the mixing data. Results for duplicate experiments were pooled and regressed to the three-parameter refolding model to give a 9-parameter description of the mixing data. All six data sets were pooled and regressed to give a single 3-parameter description. The rate constants from the overall pooled data were: $k_N = 0.17 \pm 0.05 \text{ min}^{-1}$, $K_A = 0.67 \pm 0.55 \text{ min}^{-1}$, and $k_B = 1.7 \times 10^{-5} \pm 0.035 \text{ min}^{-1}$ ($0.90 \geq P$). While the variance in k_N is acceptable, the range for K_A indicates that there is a broad variation in second-order rate constant. The uncertainty in k_B clearly demonstrates that product loss during extended incubation was negligible. The nested analysis of variance is presented in Table 5.3.

Table 5.3: Nested analysis of variance for mixing experiments.

Source	df	SSE	MSE	F-statistic
Due to mixing	6	0.4153	0.0692	16.3
Between duplicates	9	0.0382	0.0042	10.9
Residual	88	0.0342	0.0004	
Total	103	0.4535		

The analysis of variance indicates that a significant amount of the regression error is due to differences between 'duplicate' experiments. The F-statistic value of 10.9 demonstrates that this is a most significant result ($P > 0.9999$). Comparison of the 3-parameter description to the 9-parameter description gives a measure of the difference between mixing conditions. The F-statistic value of 16.3 conclusively demonstrates that there is a significant difference between the three mixing conditions ($P > 0.999$).

The analysis presented above does not allow correlation of the observed difference in mixing conditions to a particular reaction. Covariance analysis of the reaction rate constants revealed that for all of the data sets, k_N and K_A were highly correlated (correlation coefficients approaching 1.000 in some cases). Clearly, the back reaction from native to intermediate state was not the cause of the observed difference. Since the variance in the estimate of k_N from the overall pooled regression is relatively low, it is probable that the different behaviour can be correlated to the aggregation reaction. To test this hypothesis, a nested analysis of variance was conducted using the normalised data obtained from 0 to 150 min samples with $k_N = 0.17 \text{ min}^{-1}$ and k_B set to zero. Regressions required the estimation of K_A for each of the six data sets, three pooled duplicates, and the overall pooled data. The results of the nested analysis are presented in Table 5.4.

Table 5.4: Nested analysis of variance for mixing experiments with $k_N = 0.17 \text{ min}^{-1}$ and $k_B = 0$ for normalised LongR³IGF-I concentrations collected between 0 and 150 minutes.

Source	df	SSE	MSE	F-statistic
Due to mixing	2	0.2924	0.1462	14.9
Between duplicates	3	0.0294	0.0098	20.2
Residual	72	0.0351	0.0005	
Total	77	0.3219		

The analysis of variance confirmed that a significant amount of the regression error was due to differences between ‘duplicate’ experiments, but could be accounted for by the aggregation rate constant. The F-statistic value of 20.2 demonstrated that this was a most significant result ($P > 0.9999$). Comparison between duplicates and mixing conditions revealed that the rate constants for the aggregation reaction varied significantly with mixing rate. The F-statistic value of 14.9 established that the difference between the three mixing conditions could be described by increasing second-order aggregation with increasing mixing rate ($P > 0.97$).

5.3 SOLUBLE CONTAMINANTS

As outlined in the introductory chapter, interaction of cellular species with heterologous proteins during both *in vivo* and *in vitro* refolding is one of the major uncharacterised aspects of refolding. In an attempt to determine whether any of the soluble material present in the bulk homogenate of *E. coli* JM101 had an effect on the refolding of LongR³IGF-I, two refolding trials were conducted. The rate constants of refolding in the presence of low and high concentrations of soluble protein were determined and statistically tested for equivalence.

5.3.1 Experimental protocol - soluble contaminants

Two refolding trials were conducted; the first was a control trial where all buffers used were prepared with water. The dissolution procedure is presented in section 5.1.1.2. For refolding, the same experimental protocol as at described for the mixing trial using maximum stirring was used but in addition, sampling for SDS-PAGE analysis was included. The overall experimental protocol is presented in Figure 5.15.

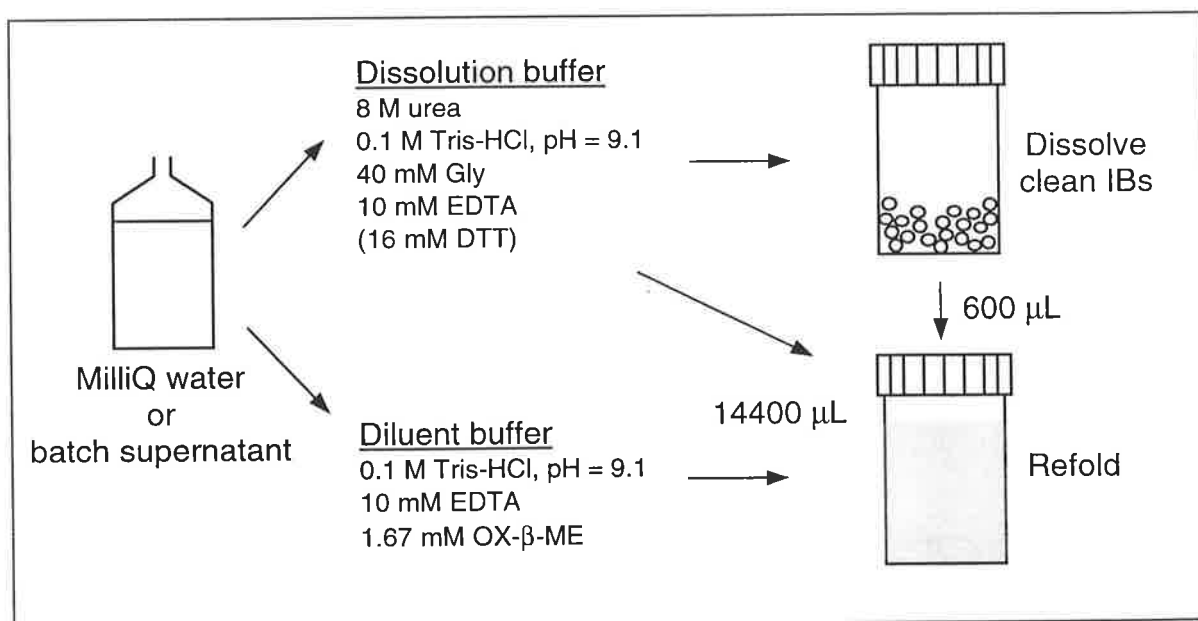


Figure 5.15: Illustration of the overall experimental protocol used for refolding of LongR³IGF-I in the presence of soluble contaminants. All buffers were prepared with either water (protein-free) or batch supernatant.

Samples for HPLC and total protein quantification were prepared in the same way as for the mixing experiments (see section 5.2.1). Analysis by SDS-PAGE of samples taken during refolding were prepared as described in section 2.2.6. The sampling protocol is presented diagrammatically in Figure 5.16.

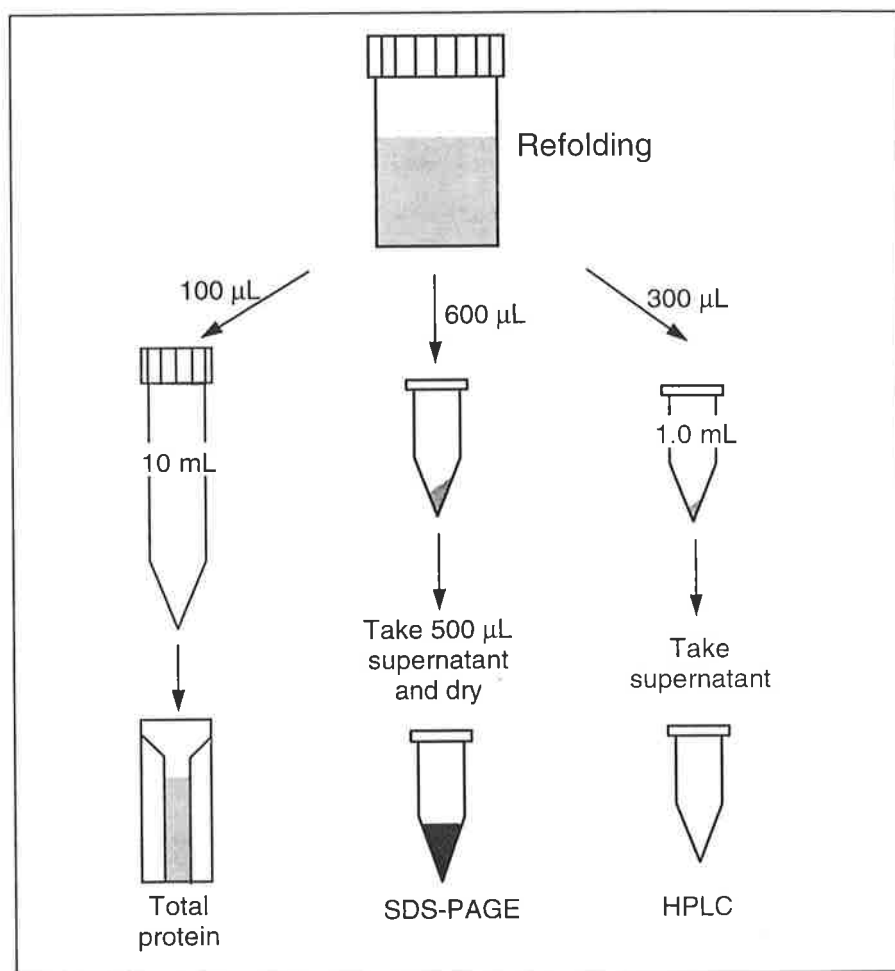


Figure 5.16: Sampling protocol for trials to characterise effects on protein refolding in the presence of soluble cellular contaminants. Where pelleted, samples were centrifuged at $18 \times 10^3 g_{\text{max}}$ for 3 minutes.

5.3.2 Results and discussion - soluble contaminants

The observations made during the control experiment were the same as for the mixing experiments described in section 5.2.2. However, as presented in section 5.1.2.1, dissolution using batch supernatant as the solvent was not as effective as using water. The presence of contaminants in the supernatant apparently (slightly) reduced the amount of urea that was available to dissolve protein present in the isolated inclusion bodies.

Dissolution and diluent buffers prepared with batch supernatant were both clear and had a bright, pale yellow colour. When 600 μL of the slightly opaque, dark brown dissolving inclusion body mixture were injected into 14.4 mL of the prepared refolding buffer, the mixture assumed a slightly opaque, pale straw colour, similar to the control refolding experiment. However, there appeared to be slightly more material precipitated in the presence of soluble contaminants. The appearance of the refolding mixture did not change over time. The pellets of solid material that were sedimented from samples quenched for analysis by HPLC and SDS-PAGE were all of consistent size and appearance.

The total soluble protein concentration during the refolding process in the presence of gross soluble contamination remained constant with an average of $1.69 \pm 0.05 \text{ g L}^{-1}$ ($0.95 \geq P$). This was in contrast to a total soluble protein concentration of approximately 0.6 g L^{-1} when water was used as the solvent. The batch supernatant itself contained approximately 1.4 g L^{-1} of dissolved protein before adding any of the dissolved inclusion bodies. The dissolved inclusion body solution had a total protein concentration of approximately $12.1 \pm 1.0 \text{ g L}^{-1}$ (see section 5.1.2.1). Hence, the nominal protein concentration for refolding was 1.88 g L^{-1} implying that approximately 0.2 g L^{-1} of the soluble protein must have precipitated upon dilution to refolding conditions before the first sample was taken for total protein quantitation.

The HPLC elution profiles of samples quenched during refolding in the presence of batch supernatant are presented in Figure 5.17. The profiles for the control experiment were very similar to those obtained for the mixing experiments (see Figure 5.13). There are two major differences between the profiles presented in Figure 5.17 and the set from the control experiment: many additional species were evident when batch supernatant was used and the relative amount of renatured product was reduced. Co-elution of contaminating species confounded quantitation of the amount of renatured protein. De-convolution of the profiles to account for the additional species using PeakFit™ (see section 2.3.2) removed some of the confusion, but may have introduced systematic a error into the analysis.

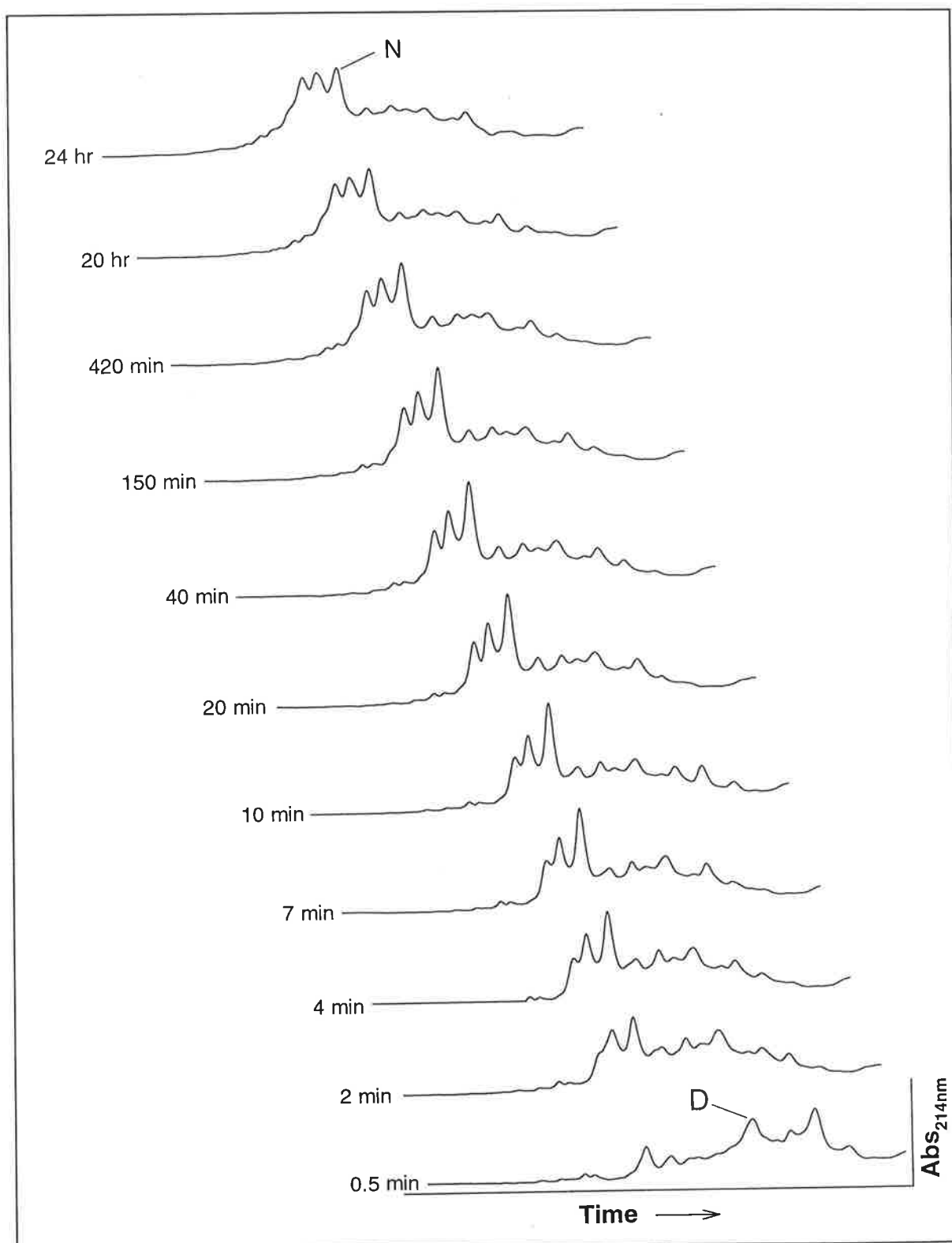


Figure 5.17: HPLC profiles for refolding of clean inclusion body starting material in the presence of batch supernatant with maximum stirring. The peak corresponding to elution of reduced, denatured LongR³IGF-I (D) is indicated in the profile of the 0.5 min sample. Native species (N) is the predominant peak in most samples and the elution time is indicated in the 24 hour sample. Profiles are presented for an elution time of 10 to 35 minutes. All profiles have been normalised to take account of the volume and dilution actually analysed.

Graphs of the relative renaturation of LongR³IGF-I in the presence and absence of batch supernatant are presented in Figure 5.18. Relative renaturation was calculated by dividing the measured concentration of renatured protein by the nominal starting concentration of reduced, denatured LongR³IGF-I (C_{D0}) added to each refolding batch. There was an obvious decrease in the relative amount of native species recovered when renaturation was conducted in the presence of soluble *E. coli* contaminants. In the control experiment approximately 40% of the total protein was LongR³IGF-I (0.25 g L^{-1} from a total of 0.6 g L^{-1}). When batch supernatant was used as the solvent, only 15% of the total protein was LongR³IGF-I (0.21 g L^{-1} from a total of 1.4 g L^{-1}). In other words, approximately three times more soluble protein contamination was present in the samples tested.

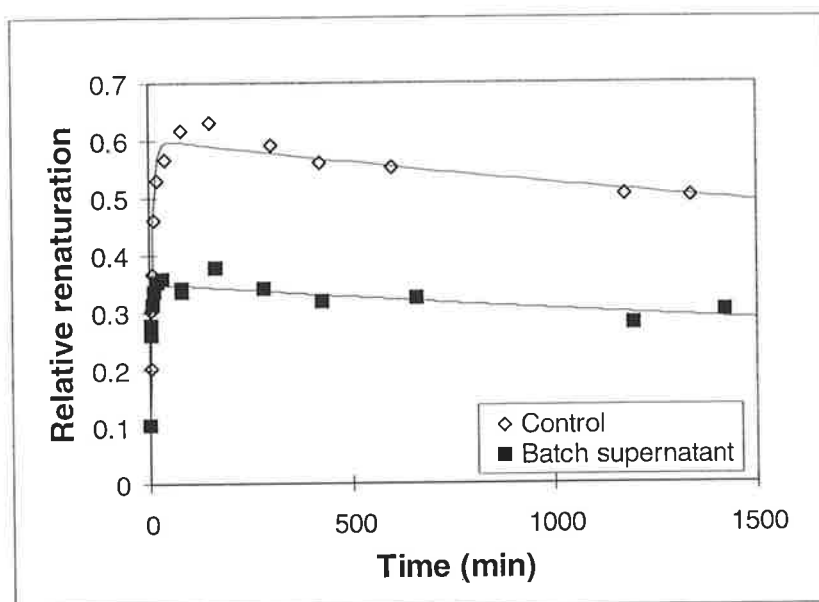


Figure 5.18: Plots of the observed time-dependent renaturation of LongR³IGF-I in the presence of gross soluble contamination ($C_{D0} = 0.214 \text{ g L}^{-1}$) and a control ($C_{D0} = 0.247 \text{ g L}^{-1}$) using water as the solvent. The two curves correspond to the three-state kinetic model of refolding regressed to each data set using BMDP.

Table 5.5: Estimates of rate constants and standard deviations regressed to the three-state refolding model using BMDP for refolding in the presence of batch supernatant and a control.

	Batch supernatant		Control	
	Estimate	Std. deviation	Estimate	Std. deviation
$k_N (\text{min}^{-1})$	0.295	0.038	0.084	0.006
$K_A (\text{mL mg}^{-1} \text{ min}^{-1})$	6.90	1.07	0.49	0.07
$k_B (\text{min}^{-1})$	0.0052	7.3×10^{-4}	0.0035	6.7×10^{-4}

Regressed values for the three rate constants in the three-state model of refolding are presented for both data sets in Table 5.5. The estimates for k_N and K_A in the presence of batch supernatant are highly correlated (asymptotic correlation coefficient = 0.967). For the control experiment, the asymptotic correlation coefficient between k_N and K_A is 0.855. Direct comparison of the corresponding rate constants indicates that there is a most significant difference in the values of k_N and K_A between the two data sets ($0.999 \geq P$). The estimates of k_B are not as significantly different, and can be considered the same if a confidence level of greater than 96% is used. However, since the rate constants are so highly correlated, overall model fits must be compared to affirm a true difference between the two data sets.

An analysis of variance between a single three-parameter description and separate three-parameter regressions is presented in Table 5.6. The F-statistic value of 190 proves that there is an emphatic difference between the three-parameter and six-parameter descriptions. Hence it can be concluded that refolding in the presence of soluble cellular contaminants results in a most significant difference in the refolding behaviour.

Table 5.6: Analysis of variance between refolding in the presence and absence of soluble contaminants.

Source	df	SSE	MSE	F-statistic
Due to contaminants	3	0.2669	0.0890	190
Residual	22	0.0103	0.0005	
Total	25	0.2772		

Analysis of the samples quenched for resolution by SDS-PAGE was unsuccessful because residual TFA could not be adequately removed from the concentrated samples using the Hetovac system. After drying the samples under reduced pressure for more than 24 hours, the amount of residual acid was so great that it exceeded the buffering capacity of the load buffer, even when diluted 1:10, which made the overall protein concentration too low to resolve using Coomassie Blue staining. The relatively high concentration of urea and salts in the refolding samples (> 2 M) was the likely cause of the poor drying behaviour.

5.4 INSOLUBLE CONTAMINANTS

The previous section describes the experiments conducted to identify the effects of soluble cellular contaminants on protein refolding. The complimentary experiments investigating the presence of insoluble contaminants are presented below.

5.4.1 Experimental protocol - insoluble contaminants

The dissolution protocol for refolding trials using batch-pelleted inclusion body paste is presented in section 5.1.1.3.

The experimental protocol and sampling for refolding was essentially the same as for refolding in the presence of batch supernatant, but to account for the lower fraction of LongR³IGF-I in the starting material, refolding was initiated by a 1:4 dilution into refolding buffer instead of a 1:25 dilution. The experimental protocol is illustrated in Figure 5.19. To maintain the same ratio of oxidised to reduced thiol groups, the concentration of OX- β -ME was increased from 0.0016 to 0.010 M. This resulted in refolding occurring in a chemical environment of 2M urea/0.1M Tris-HCl pH = 9.1/0.01 M Gly/0.01 M EDTA/0.004 M DTT/0.01 M OX- β -ME.

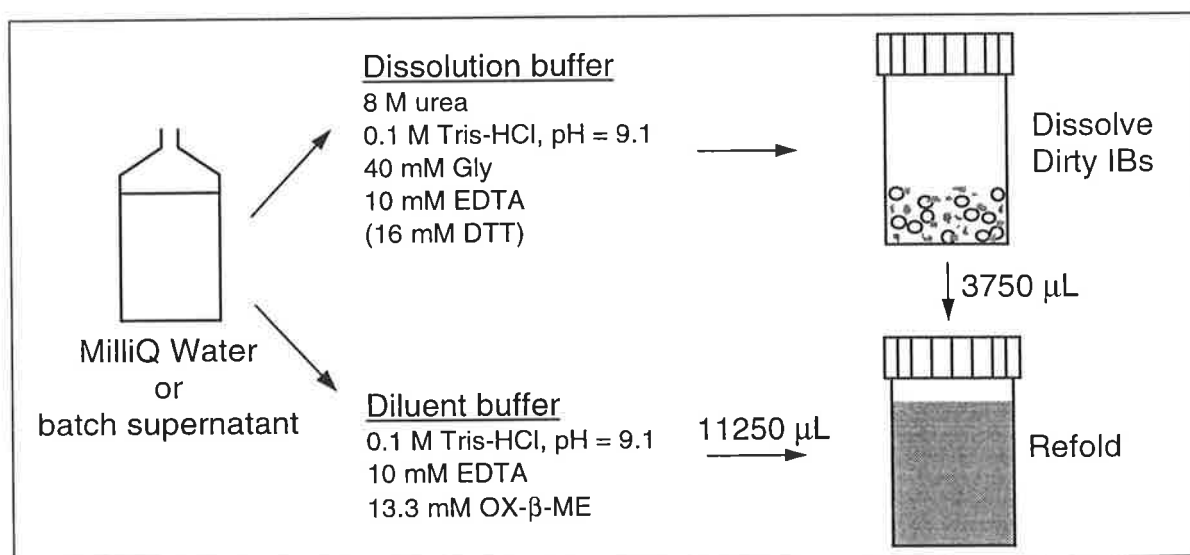


Figure 5.19: Diagram of the experimental protocol used for refolding of LongR³IGF-I obtained from inclusion body paste collected by batch centrifugation. All buffers were prepared with either water (protein-free) or batch supernatant.

5.4.2 Results and discussion - insoluble contaminants

Dissolution of the dirty inclusion bodies in buffers prepared with water and batch supernatant is presented in section 5.1.2.2. When using water as the solvent, rapid 1:4 dilution of the cloudy, brown dissolving inclusion body suspension into the brilliant clear, colourless refolding diluent immediately produced a cloudy, light brown mixture. The appearance of the mixture did not change significantly over the subsequent 24 hours. Similarly, addition of the dissolving inclusion bodies to diluent buffer prepared using the supernatant from batch centrifugation produced a cloudy, light brown mixture only slightly deeper in colour than the one obtained in the absence of soluble cellular contaminants.

The amount of insoluble material sedimented in samples taken for refolding and SDS-PAGE analysis appeared to remain constant for both refolding trials and was noticeably greater than when clean inclusion bodies were used. This observation confirmed that a significant fraction of the starting material remained insoluble during dissolution of unfractionated inclusion bodies using the standard method. As for the trials conducted using clean inclusion bodies, residual TFA could not be successfully removed from the samples taken for analysis by SDS-PAGE. Therefore, electrophoresis was not conducted to identify the significant protein species present in the refolding mixture.

Values of the total soluble protein concentration present in samples from both refolding trials are presented as a function of time in Figure 5.20. When batch supernatant was used as the solvent, the total protein concentration remained relatively constant for the duration of refolding (a 95% confidence interval for the slope of a regressed line of best fit contains zero). The average total protein concentration was $1.68 \pm 0.11 \text{ g L}^{-1}$ ($0.90 \geq P$). Soluble protein present in the batch supernatant would have contributed 1.4 g L^{-1} and at the time that refolding was initiated the dissolving mixture contained approximately 1.9 g L^{-1} of soluble protein. Therefore approximately 0.2 g L^{-1} of protein must have aggregated during the early stages of refolding. Detection of constant total soluble protein concentration in all samples is consistent with the observation that the amount of insoluble material recovered by centrifugation did not alter significantly during refolding.

Use of water as the solvent gave a total protein concentration in the dissolving mixture of approximately 1.6 g L^{-1} when refolding was initiated. However, the initial concentration of soluble protein during the early stages of refolding was approximately 0.7 g L^{-1} . This

indicates that transferring to refolding conditions gave a rapid increase in the concentration of total protein. During refolding, the total soluble protein concentration gradually decreased to approximately 0.5 g L^{-1} after 20 hours. The observed variation in total soluble protein concentration and the evidently unchanging amount of insoluble material recovered appear contradictory. However, since protein typically contributes approximately half of the dry cell weight of *E. coli* (Niedhardt, 1987) the decrease in soluble fraction during refolding may have been masked by other insoluble contaminants.

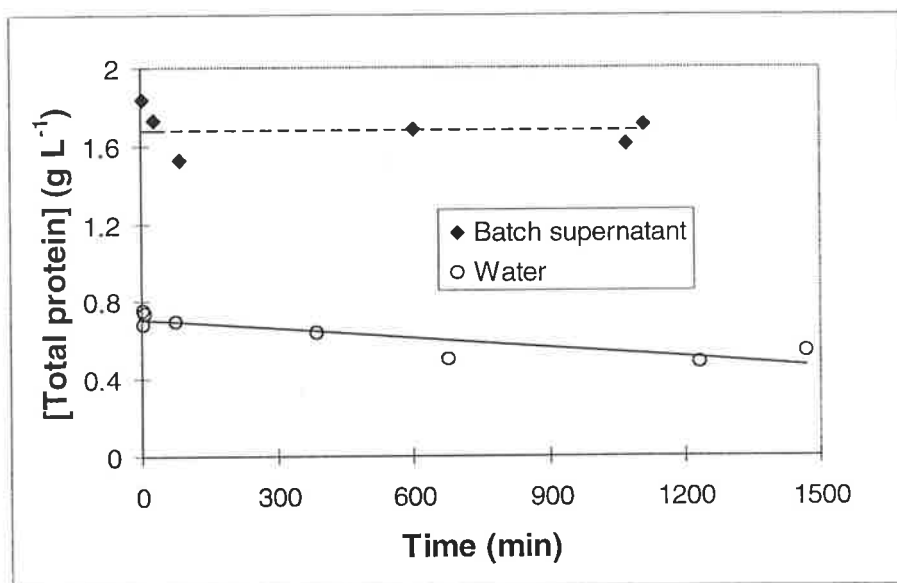


Figure 5.20: Variation in the total soluble protein concentration with time for refolding of dirty inclusion bodies in the presence and absence of batch supernatant. The average concentration when using batch supernatant is shown as a dashed line. The solid line shows the line of best fit for the total protein concentration observed in the absence of batch supernatant.

HPLC profiles for various times during refolding of dirty inclusion LongR³IGF-I bodies in the presence of batch supernatant are presented in Figure 5.21. The significant species observed to elute at approximately 10 minutes are consistent with the early-eluting species present in batch supernatant (see Figure 2.9). A minor peak was seen at the expected elution time of LongR³IGF-I in samples taken after four minutes. However, the size of this peak was relatively insignificant in comparison to the background of co-eluting proteins. This is to be expected since the amount of reduced, denatured LongR³IGF-I added to initiate refolding was less than 1% of the total protein.

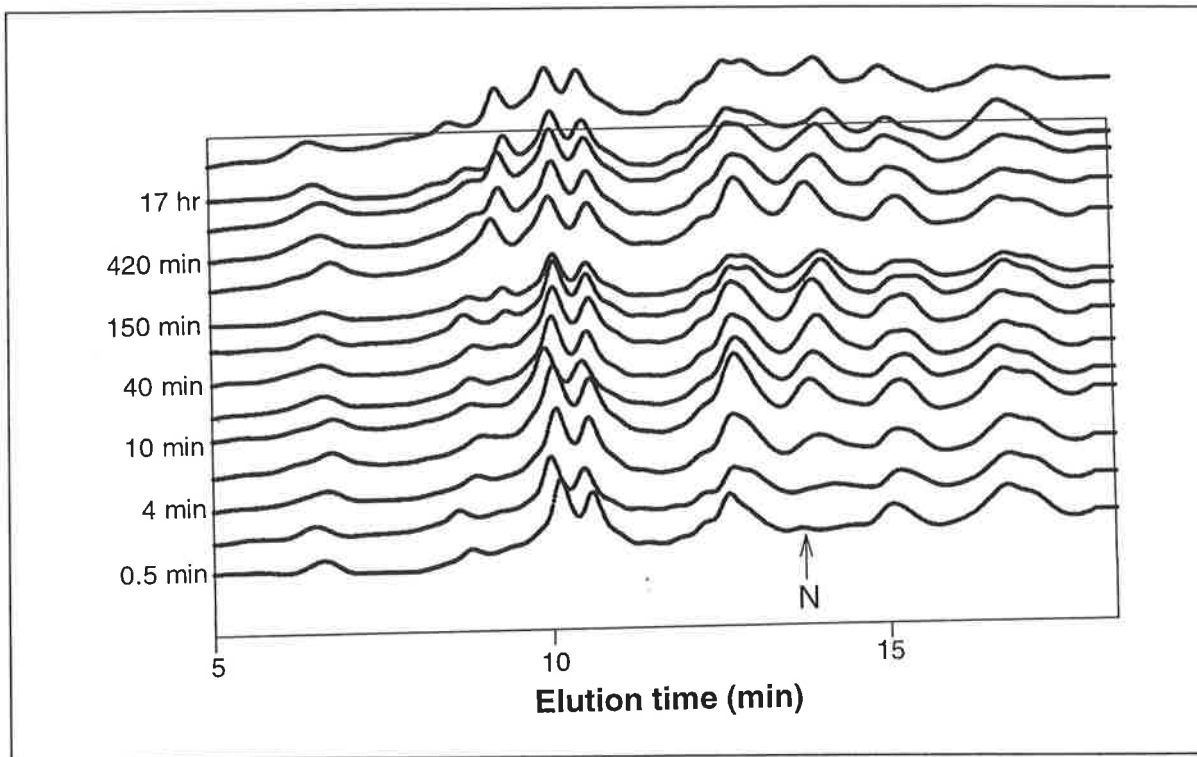


Figure 5:21: HPLC profiles of refolding dirty inclusion bodies in batch supernatant from 0.5 min to 23 hours. The expected elution time of native conformation LongR³IGF-I is indicated. Profiles have been corrected for the actual dilution and injection volume used.

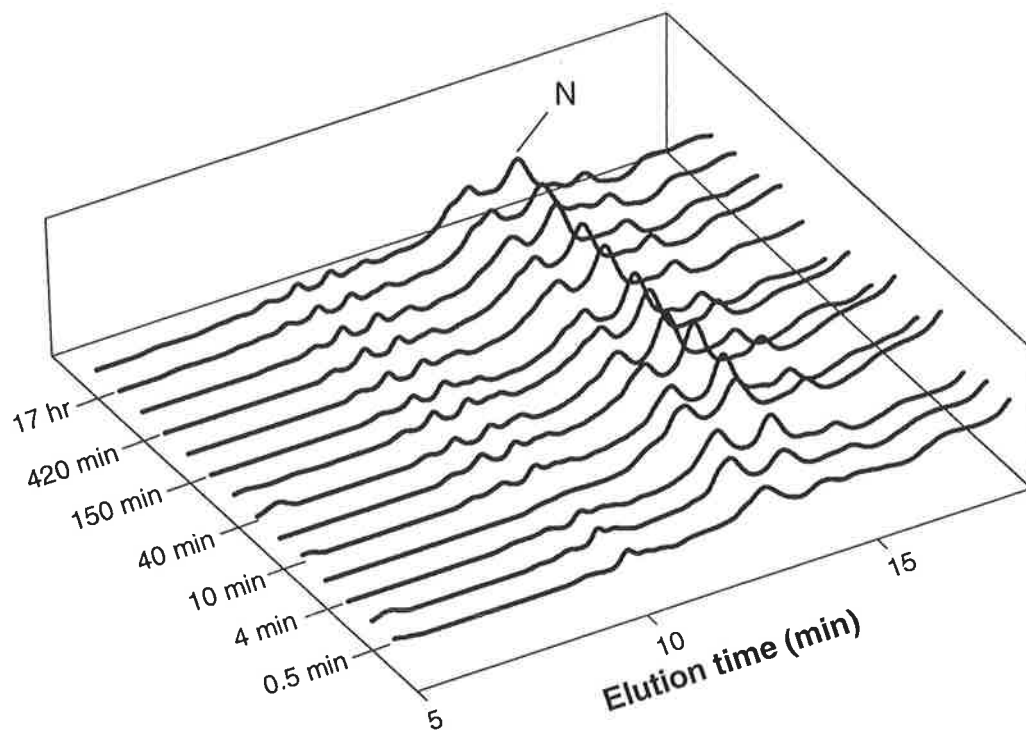


Figure 5:22: An isometric view of the HPLC profiles for refolding of dirty inclusion bodies in water from 0.5 min to 24 hr. The expected elution time of native conformation LongR³IGF-I is indicated. Profiles have been corrected for the actual dilution and injection volume used.

HPLC profiles for refolding dirty inclusion bodies without using batch supernatant are presented in Figure 5.22. In this set of HPLC profiles, the major peak in most samples eluted at a time corresponding to correctly renatured LongR³IGF-I. However the relative amount of this peak was less than when clean inclusion bodies were used (see Figure 5.13). As for the equivalent refolding trial using batch supernatant, the amount of LongR³IGF-I present in the refolding mixture was less than 1% of the total soluble protein as estimated using the method described in section 2.2.2.

An attempt to quantify of the amount of renatured LongR³IGF-I present in each of the HPLC profiles using PeakFit™ is presented in Figure 5.23. Relative renaturation has been calculated by dividing the calculated protein concentrations by the amount of denatured product added to initiate refolding. However, the absolute values are highly suspect, particularly in the presence of batch supernatant, because LongR³IGF-I was not the major species in the peaks eluted at the expected time. The erroneous nature of the quantitation is supported by regression of the values to the three-parameter model of refolding. BMDP was unable to find a converged solution for regression of either data set, so confident estimates of the rate constants could not be made.

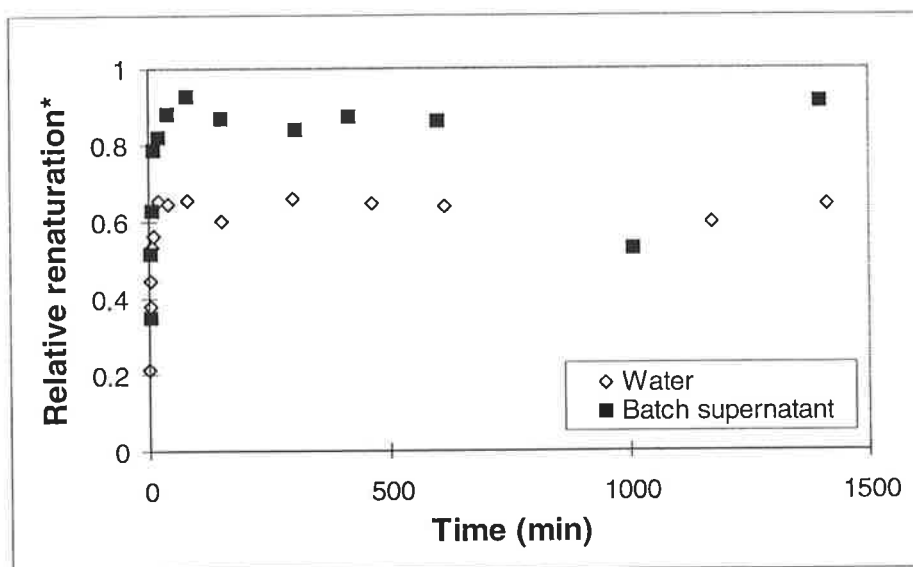


Figure 5:23: Relative renaturation for refolding dirty inclusion bodies ($C_{D0} = 5.7 \text{ mg L}^{-1}$) in the presence and absence ($C_{D0} = 5.1 \text{ mg L}^{-1}$) of batch supernatant.

*Accurate quantitation of the LongR³IGF-I concentration from the data presented in Figures 5.21 and 5.22 was not possible due to co-elution of contaminating species. Therefore, the relative renaturation values are likely to be only approximate.

A more rigorous interrogation of the HPLC profiles was conducted in an attempt to remove the contribution of contaminants. Profiles from consecutive time-points were subtracted from one another, making adjustments for the minor variation in elution time between samples. The areas of the peaks nominally corresponding to LongR³IGF-I were integrated using PeakFit™ and the transient product concentration determined by their cumulative sum. This method proved unsuccessful, with relative renaturation exceeding unity.

In summary, the method used to determine whether cellular components, that are insoluble under physiological conditions, have an effect on refolding was unsuccessful. Contaminants that were solubilised during dissolution of the unfractionated insoluble cellular material predominated in subsequent analytical procedures. Accordingly, accurate quantitation of LongR³IGF-I was beyond the resolution of the analytical technique used and meaningful estimates of rate constants for a three-state refolding mechanism could not be calculated. However, the study did demonstrate that insoluble cellular contaminants can have a dramatic effect on the overall renaturation by reducing the amount of product released from inclusion bodies during the dissolution step.

5.5 CONCENTRATED REFOLDING

The refolding experiments to characterise mixing and contaminant effects were all conducted at LongR³IGF-I concentrations less than 0.4 g L⁻¹. An aim of the experimental work was to employ a fed-batch strategy to refold LongR³IGF-I at high concentration. The highest concentration that can be practically achieved, to give a refolding environment that contains 2 M urea, is a 1:4 dilution of dissolved inclusion body mixture into appropriate diluent buffer. This section presents two refolding trials conducted using a 1:4 dilution of dissolved clean inclusion body starting material for comparison with subsequent fed-batch refolding trials.

5.5.1 Experimental protocol - Concentrated refolding

The method used for dissolution of the clean inclusion bodies, subsequently used for concentrated refolding trials, is presented in section 5.1.1.3. After dissolution had progressed for 30 minutes, two 2500 µL samples were diluted 25% (v/v) to give a chemical environment of composition 2 M urea/0.1 M Tris-HCl pH = 9.1/0.01 M Gly/0.01 M EDTA/4.0 × 10⁻³ M DTT/0.010 M OX-β-ME with a single rapid injection of each sample into well mixed solutions of diluent. Mixing was maintained at the level used as the maximum rate in the studies to identify effects due to mixing (section 5.2). The experimental protocol for the concentrated refolding trials is illustrated in Figure 5.24.

The progress of refolding was monitored by quenching small samples for HPLC, total protein analysis, and SDS-PAGE. Samples for analysis by HPLC were quenched by dilution of 50 µL of refolding mixture into 1000 µL of 0.1% TFA before vortexing for 15 s, and centrifuging to recover the supernatant to be analysed for soluble LongR³IGF-I content. Samples for total soluble protein quantitation were made in triplicate by dilution of 50 µL of refolding mixture into 10 mL of 0.1 M NaCl/0.1% TFA and vortexing for 15 s. For analysis by SDS-PAGE, 20 µL samples were diluted into an equal volume of either reducing or non-reducing loading buffer following the procedure presented in section 2.2.5.

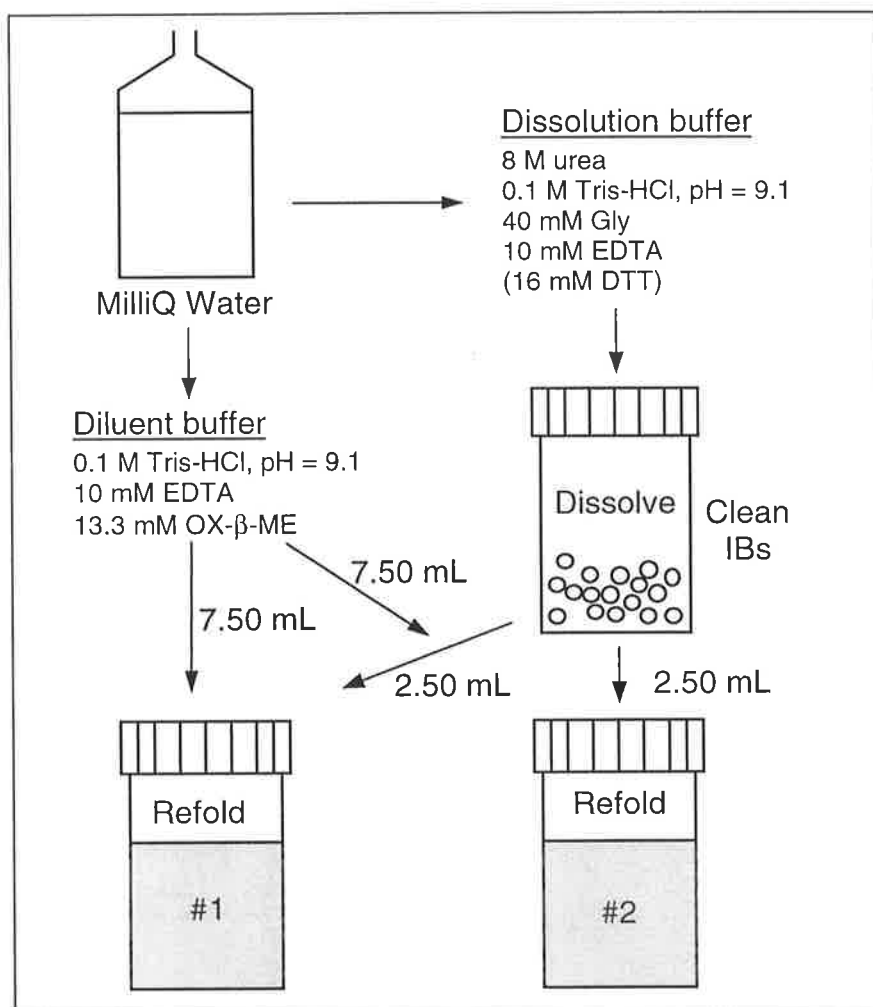


Figure 5.24: Diagram of the procedure used for refolding LongR³IGF-I at high concentration. Dissolution and refolding mixtures were assayed for total soluble protein concentration, LongR³IGF-I by HPLC, and SDS-PAGE (under reducing and non-reducing conditions) with time.

5.5.2 Results and discussion - concentrated refolding

The appearance of the two concentrated refolding mixtures was not discernibly different throughout the duration of the experiments. Rapid dilution of the dissolving mixture of clean inclusion bodies into the stirred, bright, clear, colourless diluent buffer immediately gave a slightly opaque brown solution. The mixture appeared quite similar to those of the dilute refolding trials designed to examine mixing effects. However, within 30 s a significant amount of the soluble material had precipitated to give a milky brown solution. Further incubation did not alter the appearance of the mixture. The amount and appearance of the insoluble material recovered by centrifugation of acid quenched samples for HPLC

did not change with time. The size of the pelleted insoluble material was notably larger than the corresponding samples taken from dilute refolding mixtures.

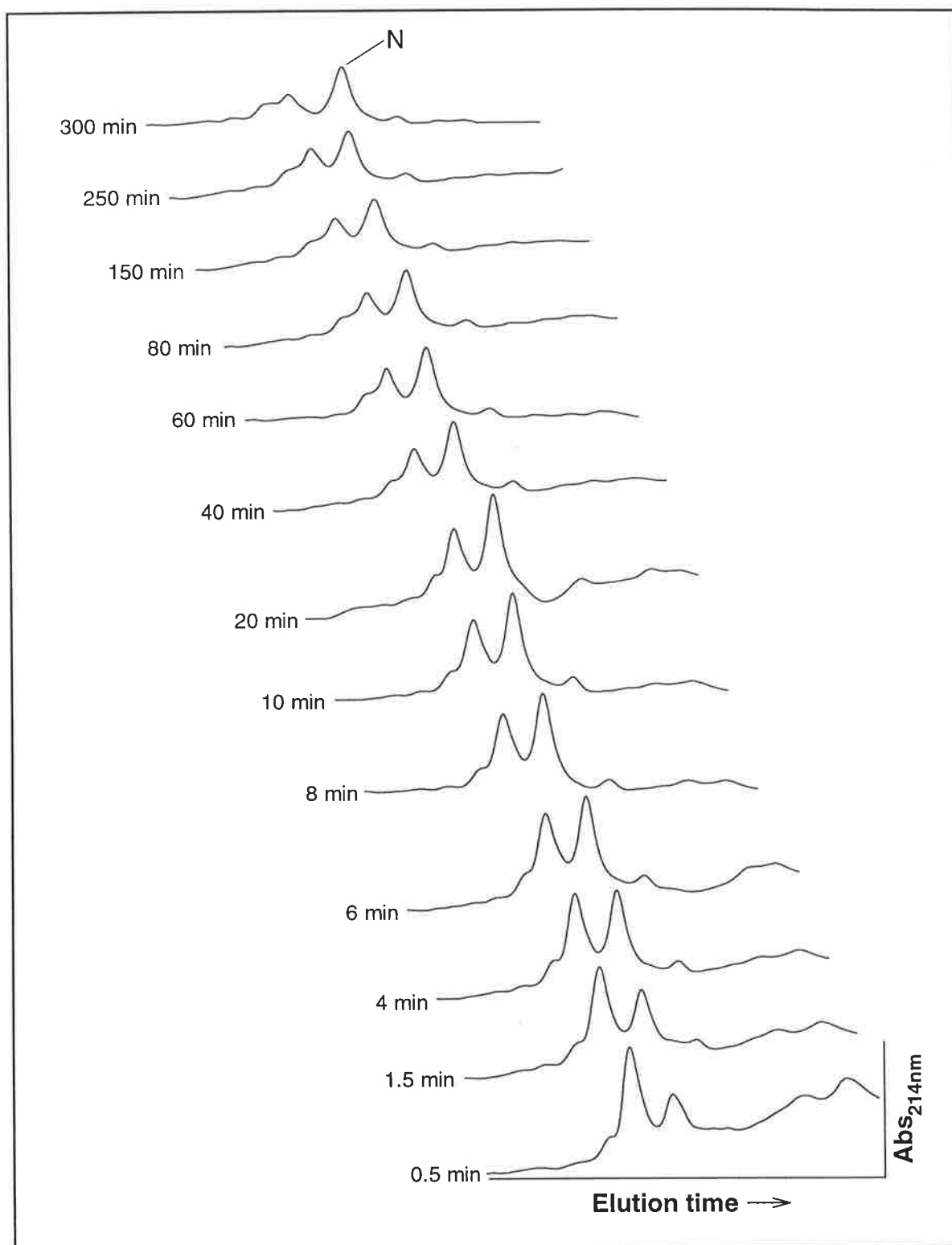


Figure 5.25: A time series of HPLC profiles during concentrated refolding of LongR³IGF-I. The expected elution time of renatured LongR³IGF-I (N) corresponds to the predominant species in most profiles. Profiles have been corrected to show an equivalent loading.

The total concentration of soluble protein in each of the refolding mixtures was estimated to be $3.8 \pm 0.3 \text{ g L}^{-1}$ and $3.4 \pm 0.3 \text{ g L}^{-1}$ ($0.90 \geq P$) using BSA as the standard protein. The amount of reduced, denatured LongR³IGF-I added at the start of refolding corresponded to an overall concentration of $1.87 \pm 0.10 \text{ g L}^{-1}$ (quantification by HPLC). In other words, approximately half of the soluble protein present in the refolding mixtures was denatured fusion peptide. The HPLC profiles for samples taken from each refolding mixture at corresponding times were quite consistent. One set of HPLC profiles is presented in Figure 5.25 as an example.

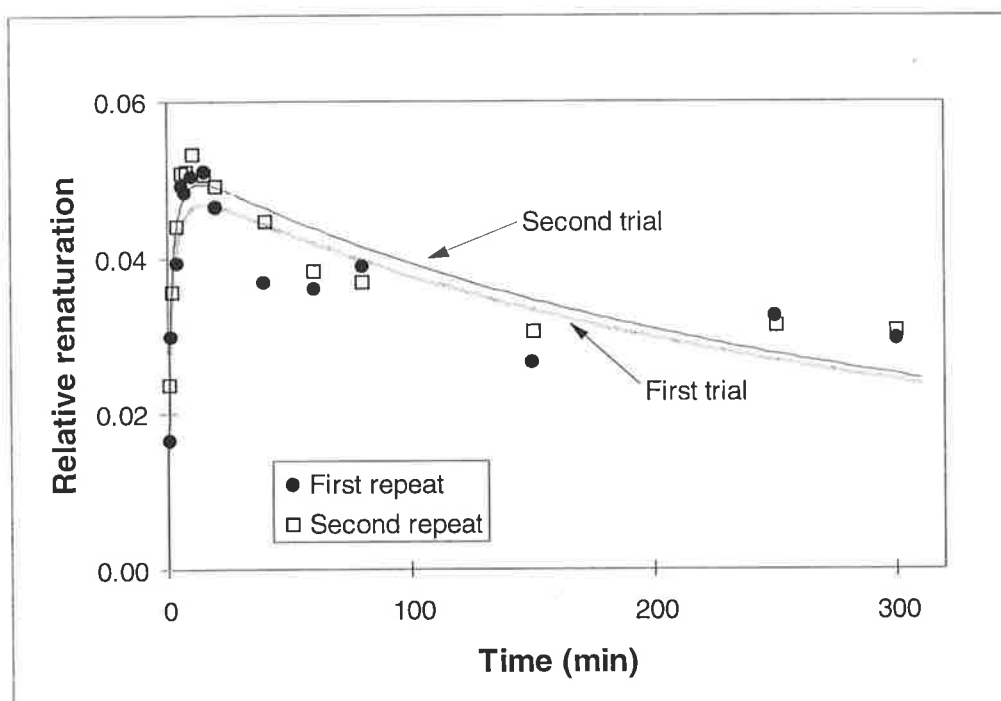


Figure 5.26: Plot of relative LongR³IGF-I renaturation versus time for two concentrated refolding trials. The curves show regressed fits to the three-parameter model of refolding.

Quantification of the amount of correctly renatured LongR³IGF-I, relative to the amount of denatured fusion protein added initially, is presented in Figure 5.26. The results are consistent with the observed trend in the HPLC profiles. The amount of renatured product increased to a maximum within 20 minutes, but by 40 minutes had decreased significantly. Further incubation resulted in only a slight decrease in the amount of native LongR³IGF-I. This behaviour was significantly different to refolding under the relatively dilute conditions used in the mixing experiments. Under dilute conditions, the maximum concentration was reached after incubation for at least two hours and only marginally decreased with time

thereafter (see Figure 5.14). For refolding at high protein concentration, the maximum yield of renatured LongR³IGF-I was approximately 5%. This was about 10-fold less than the yield obtained in the mixing experiments, indicating that the bulk of the product aggregated during refolding.

Regression to the three-state kinetic model for refolding reveals that the simplified pathway does not provide a good description of the observed behaviour. While the two regressed curves in Figure 5.26 follow approximately the correct trend, they fail to describe the rapid decrease in yield observed once the maximum had been reached. The concentration of renatured product appeared to be approaching a stable equilibrium, rather than trending to zero as the model predicts. The rate constants and estimates of variance, regressed to the data using BMDP, are given in Table 5.7. The poor fit of the model is reflected in the relatively large values for the standard deviations in k_N and K_A . Analysis of variance demonstrates that given the uncertainty in the estimates of rate constants, the are true repeats (Table 5.8).

Table 5.7: Estimates of rate constants and standard deviations regressed to the three-state refolding model using BMDP for refolding at high product concentration.

	First repeat		Second repeat	
	Estimate	Std. deviation	Estimate	Std. deviation
k_N (min^{-1})	0.11	0.04	0.12	0.04
K_A ($\text{mL mg}^{-1} \text{min}^{-1}$)	4.8	2.3	4.9	2.0
k_B (min^{-1})	0.013	0.002	0.014	0.002

Table 5.8: Analysis of variance between repeated concentrated refolding trials.

Source	df	SSE	MSE	F-statistic
Between repeats	3	3.59×10^{-5}	1.20×10^{-5}	0.57
Residual	22	4.64×10^{-4}	2.11×10^{-5}	
Total	25	5.00×10^{-4}		

Examination of the residuals by the method of runs (Draper and Smith, 1966) reveals that they are highly patterned, exhibiting sinusoidal behaviour (see Figure 5.27). The model initially over-estimates the rate of accumulation of native LongR³IGF-I, under-estimates the maximum yield, does not account for the rapid decrease in concentration, and tends to

zero rather than a steady state. In theory a good description of the data would give a random distribution of residuals. The probability of obtaining only two positive groups and two negative groups of residuals in each set of data is less than 0.03, demonstrating that the simplified kinetic model is systematically flawed.

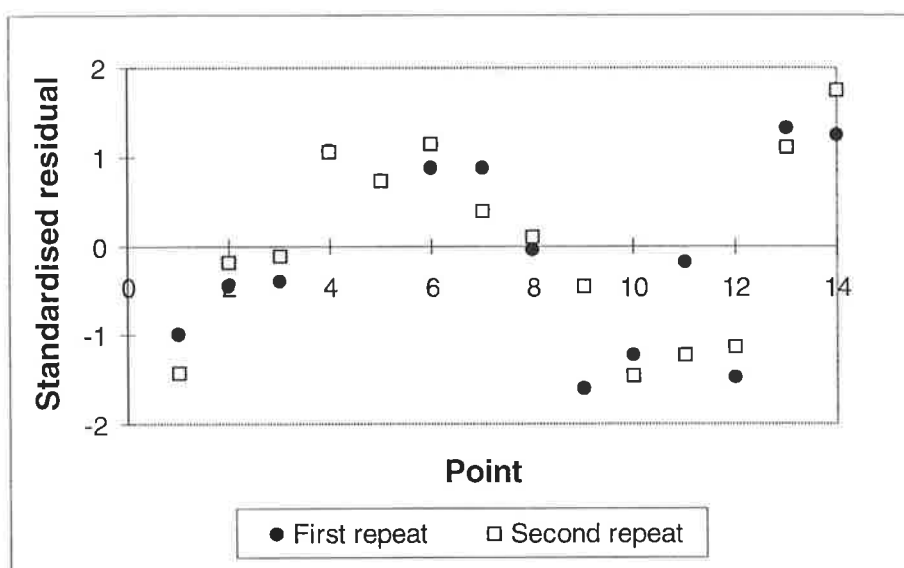


Figure 5.27: Plot of standardised residuals for the two refolding trials conducted at high protein concentration.

SDS-PAGE analysis of samples taken from one of the reacting mixtures is presented in Figure 5.28. Lanes 8 and 9 show the protein content of the dissolving inclusion body mixture under reducing conditions after 20 and 300 minutes, respectively. The overall protein content in the two samples was essentially the same, but the amount of monomer relative to dimer (18 kDa species) was greater in the 300 minute sample. Lanes 2, 4, and 6 correspond to refolding samples quenched into non-reducing buffer after 30, 60 and 300 minutes, respectively. Lanes, 3, 5, and 7 show the same sequence of samples obtained under reducing conditions.

Analysis of the samples run under non-reducing conditions reveals that the protein was extensively aggregated. At 30 and 60 minutes, significant bands were resolved at molecular weights that are multiples of 9 kDa (the molecular weight of LongR³IGF-I), to give a 'ladder' of multimers down the lane. The ladder is apparently absent in the final sample taken at 300 minutes, but the stacking gel contained the bulk of the protein loaded into the lane, indicating that most of the species were too large to enter the running gel. This was the most likely cause of the different resolution in lane 6. A small amount of

aggregated material is also evident in the stacking gel for the 60 minute sample.

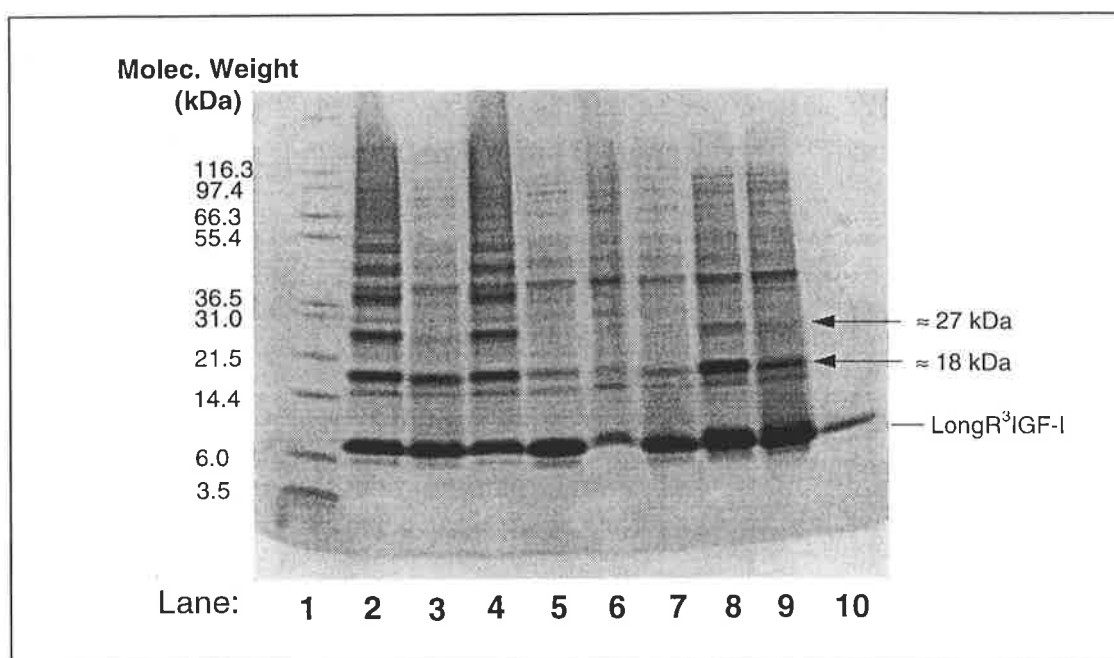


Figure 5.28: SDS-PAGE of refolding at high overall concentration (10-20% Novex™ Tricine gel). *Lane 1:* Mark12 wide range protein standard, *Lane 2:* 30 min. (10 μ L, non-reducing), *Lane 3:* 30 min (6 μ L, reducing), *Lane 4:* 60 min. (10 μ L, non-reducing), *Lane 5:* 60 min (6 μ L, reducing), *Lane 6:* 300 min. (10 μ L, non-reducing), *Lane 7:* 300 min (6 μ L, reducing), *Lane 8:* Dissolved inclusion bodies after 20 min. (7 μ L, reducing), *Lane 9:* Dissolved inclusion bodies after 300 min. (7 μ L, reducing), *Lane 10:* LongR³IGF-I standard.

Analysis of the refolding samples under reducing conditions, showed that the total amount of protein present in the band corresponding to monomeric LongR³IGF-I was approximately the same. However, there was a slight decrease in the amount of protein present in the 18 kDa band in the 60 and 300 minute samples, relative to the 30 minute sample.

The results of SDS-PAGE demonstrated that the aggregates that form during the early stages of refolding were relatively stable. A range of multimeric species, including dimeric forms, were populated during the slow phase of refolding. Prolonged incubation lead to formation of large multimeric species. The HPLC results demonstrate that the amount of monomeric protein appeared to reach a fairly stable level, which implies that the progression to formation of large multimers did not involve association of aggregates with monomers, but occurred by interaction between already associated forms.

5.6 FED-BATCH REFOLDING

Simulation of fed-batch refolding assuming a three-state model for refolding, advocates feeding as slowly as possible to achieve highest yield, provided that the back reaction from native to intermediate state is unfavourable. As discussed in section 5.1.2.1, a practical limitation observed during batch refolding experiments was that the denatured starting material was only stable for approximately one hour. For this reason, repeated fed-batch refolding trials were conducted by feeding batches of dissolving clean inclusion bodies into diluent buffer for 60 minutes. The transient LongR³IGF-I concentrations were compared with corresponding batch experiments and simulations assuming a three-state model of refolding.

5.6.1 Experimental protocol - fed-batch refolding

To check that sampling would not significantly affect the results of fed-batch refolding, the cumulative volume was compared with and without folding for the sampling protocol outlined below. Figure 5.29 presents the predictions in graphical form. Since the largest relative volume loss was 1.4% with an overall decrease of 5.4%, it was assumed to be acceptable.

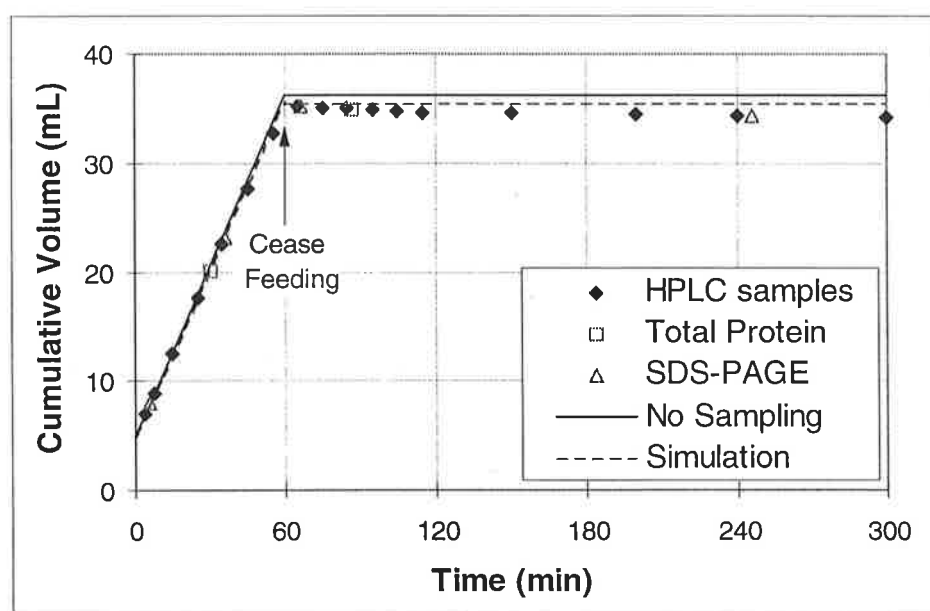


Figure 5.29: Graphs of cumulative volume for fed-batch refolding with and without sampling. The dashed line shows the volume profile used for simulation.

Two GBC LC1110 pumps were used to feed dissolving inclusion body mixture and diluent buffer at a ratio of 1:3. Both pumps were calibrated: one to deliver 8M urea solution at 0.13 mL min^{-1} and the other to deliver diluent buffer at 0.39 mL min^{-1} . Fresh dissolution buffer (8 M urea/0.1 M Tris-HCl pH = 9.1/0.04 M Gly/0.01 M EDTA/0.016 M DTT) and diluent buffer (0.1 M Tris-HCl pH = 9.1/0.01 M EDTA/0.01 M OX- β -ME) were prepared using water as the solvent. The pumps were primed with each of the freshly made buffers and then set to their calibrated flow rates. 1.25 mL of the dissolution buffer and 3.75 mL of the diluent buffer were mixed in a 50 mL beaker and set stirring with a 2.0 cm flea on a magnetic stirrer (Industrial Equipment and Control Pty Ltd, Cat no. CH2 080-001). Dissolution of the inclusion bodies was initiated using the routine procedure outlined in section 2.6. Sampling of the dissolving mixture was started five minutes after adding the dissolution buffer to the inclusion body slurry. After dissolving the inclusion bodies for six minutes, the pump feeding dissolution buffer was primed with the dissolved material and returned to its set-point flow rate (0.13 mL min^{-1}). Eight minutes after initiating dissolution the two feed lines were placed in the refolding mixture and feeding commenced. After 60 minutes both pumps were stopped and refolding continued for a further four hours. Throughout refolding, stirring speed was maintained constant at 50% of maximum. Progress of dissolution and refolding were measured using SDS-PAGE (both reducing and non-reducing), total protein quantification, and HPLC to determine LongR³IGF-I concentration. The experimental protocol is illustrated in Figure 5.30.

For dissolution, SDS-PAGE samples were quenched by dilution of 20 μL into 60 μL of 1.5 $\times$ load buffer. Triplicate 20 μL samples were quenched into 10.0 mL of 0.1 M NaCl/0.1% TFA for total protein quantification. 20 μL samples of dissolving mixture were quenched into 1000 μL of 0.1% TFA for analysis by HPLC.

For refolding, samples for SDS-PAGE were quenched by 1:1 dilution of 20 μL into 2 $\times$ load buffer. Triplicate 20 μL samples were quenched into 10.0 mL of 0.1 M NaCl/0.1% TFA for total protein quantification. Samples for analysis by HPLC were quenched by dilution of 100 μL of refolding mixture into 1000 μL of 1.0% TFA.

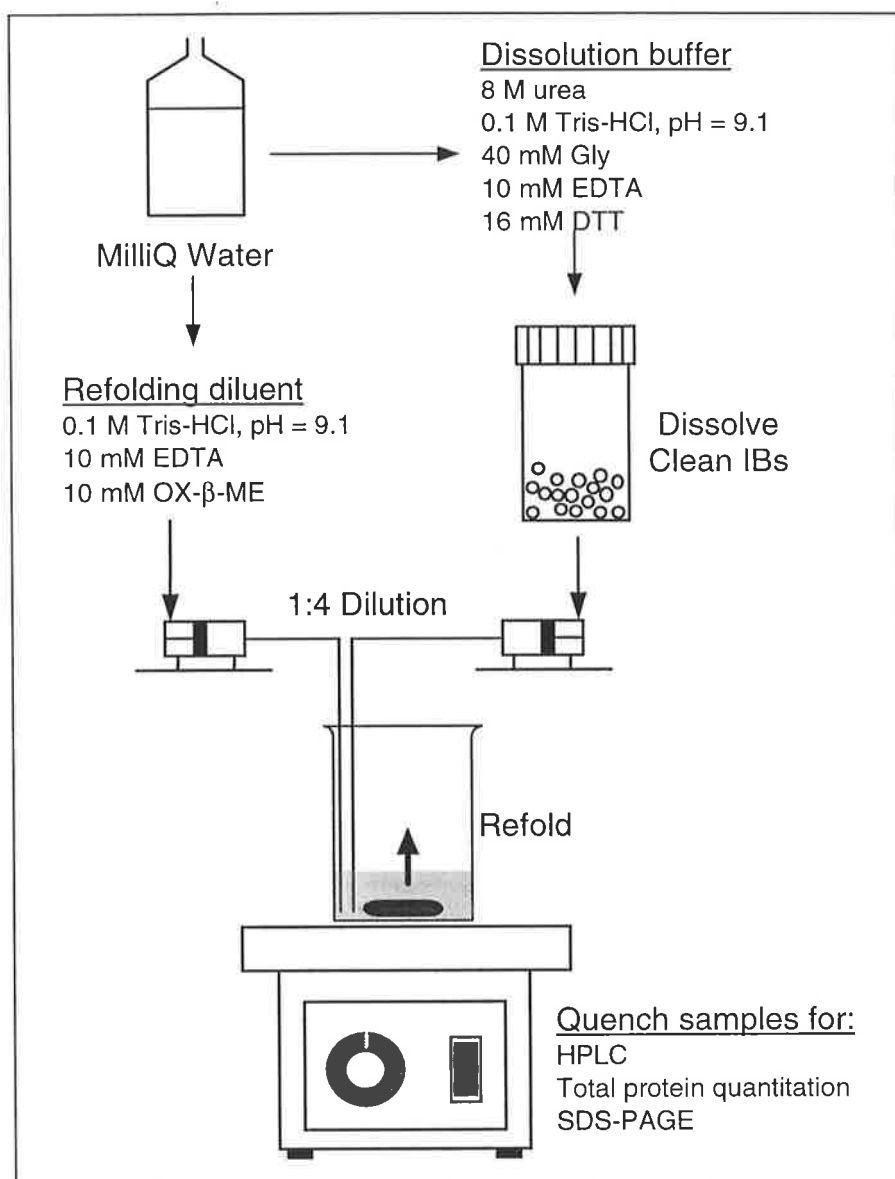


Figure 5.30: Diagram of the protocol used for fed-batch refolding. Flow rate of diluent buffer was maintained at 0.39 mL min^{-1} and dissolving inclusion body mixture at 0.13 mL min^{-1} using GBC LC1110 pumps.

5.6.2 Results and discussion - fed-batch refolding

The appearance of the two mixtures during fed-batch refolding was not noticeably different. The protein-free 5 mL charge had a brilliant clear colourless appearance. Continuous addition of the dissolving mixture resulted in a gradual change in the colour and transparency of the mixture. Initial addition gave only slight opaqueness and a pale straw coloured solution. After ten minutes of feeding the mixture was opaque and yellow in colour. By 30 minutes, the mixture appeared light brown and quite cloudy. At the end

of feeding, the solution was cloudy and light brown in colour. Subsequent incubation without any further addition of denatured material did not alter the appearance of the mixture. The amount of insoluble material sedimented by centrifugation of the acid-quenched samples, to recover particulate-free soluble protein for analysis by HPLC, followed the trend in visual appearance. During feeding, the amount of insoluble material increased with time. After feeding had ceased, the size of the pellets did not change significantly.

The total concentrations of soluble protein in the two batches of dissolved inclusion bodies used for fed-batch refolding were $13.9 \pm 0.1 \text{ g L}^{-1}$ and $11.8 \pm 0.1 \text{ g L}^{-1}$ ($0.90 \geq P$). As presented in section 5.1.2.1, the concentration of denatured LongR³IGF-I (estimated by HPLC) in these mixtures was 8.2 and 7.7 g L^{-1} . Therefore, approximately 40% of the total protein concentration was due to contaminating *E. coli* components. During refolding, the total protein concentration was measured at approximately the midpoint of feeding and half an hour after completion of feeding. The observed and nominal protein concentrations (assuming all protein remains soluble) are presented in Table 5.9. In the first trial, the observed protein concentration was approximately 40% less than the nominal value at both sampling times. In the second trial approximately 30% and 20% of the nominal total soluble protein was missing from the first and second samples, respectively. These values confirm that the cloudy precipitate observed during refolding was likely to contain a substantial amount of aggregated LongR³IGF-I.

Table 5.9: Observed and nominal total protein concentrations during fed-batch refolding trials. Feed concentrations were 13.9 and 11.8 g L^{-1} in the first and second repeats, respectively.

	First repeat		Second repeat	
	Observed	Nominal	Observed	Nominal
Midpoint (g L^{-1})	1.68 ± 0.32	2.63	1.73 ± 0.40	2.36
90 min. (g L^{-1})	1.85 ± 0.31	3.00	2.05 ± 0.40	2.54

The HPLC profiles of samples quenched at corresponding times during the refolding of each batch of clean inclusion bodies were quite consistent. A time series of HPLC profiles for the second refolding trial is presented in Figure 5.31. The first sample was quenched

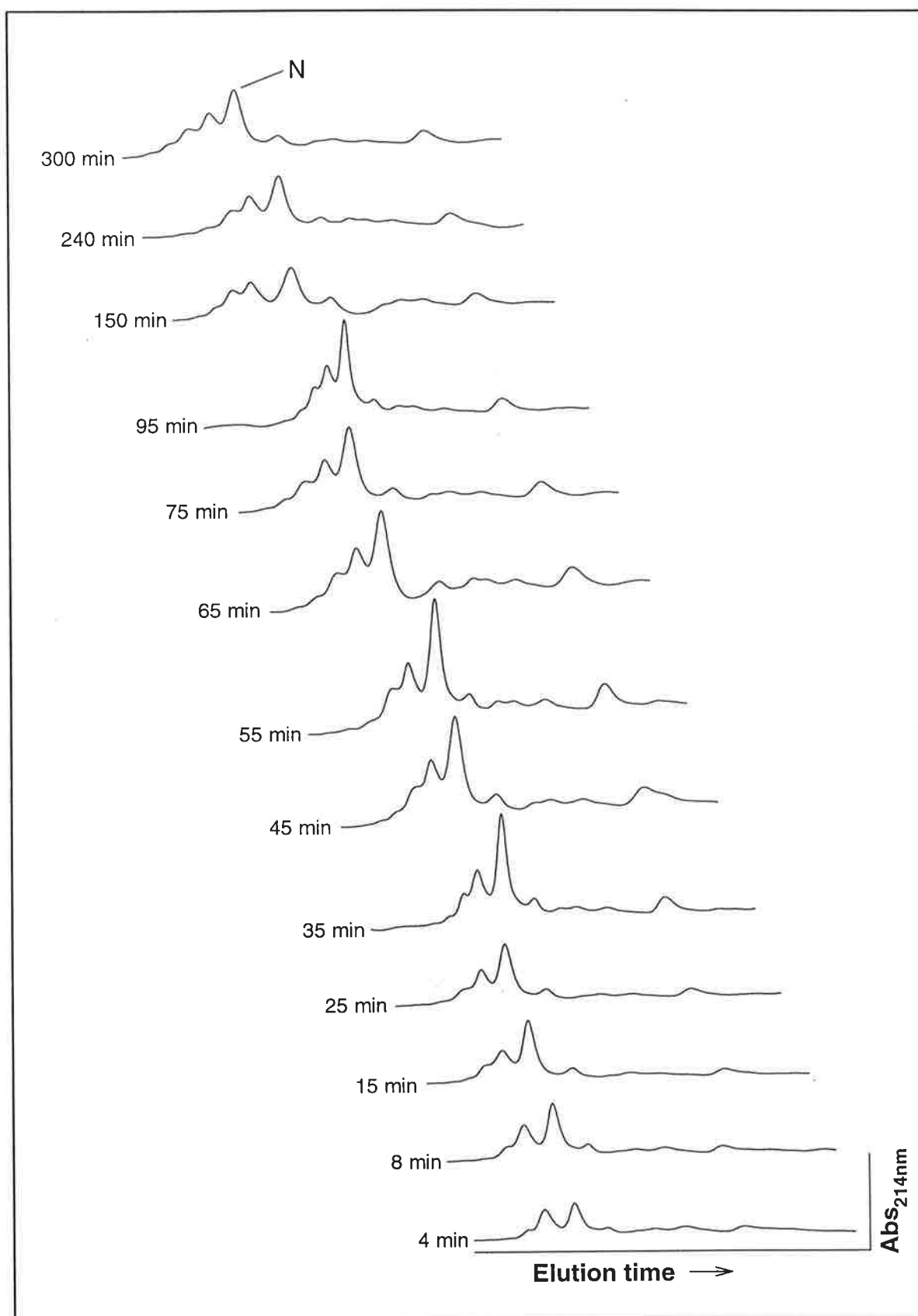


Figure 5.31: A time series of HPLC profiles for fed-batch refolding. The expected elution time of renatured LongR³IGF-I (N) corresponds to the predominant species in all profiles, as indicated in the 300 minute profile. All profiles are have been corrected to show an equivalent loading.

four minutes after feeding of the dissolved inclusion bodies was started. A significant peak was observed eluting at the expected time of native LongR³IGF-I. The relative amount of this peak, and the total area under the curves continued to increase in the profiles up to 55 minutes. Once feeding denatured protein ceased (60 min.), the total area beneath the curves and the relative amount of LongR³IGF-I decreased in the subsequent HPLC profiles.

Values of the concentration of native LongR³IGF-I, determined by integration of the HPLC profiles using PeakFit™ are presented in Figure 5.32. The rapid decrease in the amount of native protein between 60 and 100 minutes was clearly evident in both refolding trials. The amount of native protein appeared to reach a stable equilibrium at approximately 0.2 g L⁻¹. This corresponds to a relative renaturation of approximately 10% at an overall LongR³IGF-I concentration of 1.75 g L⁻¹ (the average of the two trials). In the concentrated batch refolding trials, the overall renaturation yield was approximately 4% at an overall LongR³IGF-I concentration of 1.87 g L⁻¹ (see Figure 5.26). However, if the fed-batch refolding had been quenched at the same time as feeding ceased, the concentration of LongR³IGF-I would have been at a maximum of 0.35-0.4 g L⁻¹ corresponding to a renaturation yield of 20-23%; approximately 5-fold greater renaturation than obtained with batch refolding.

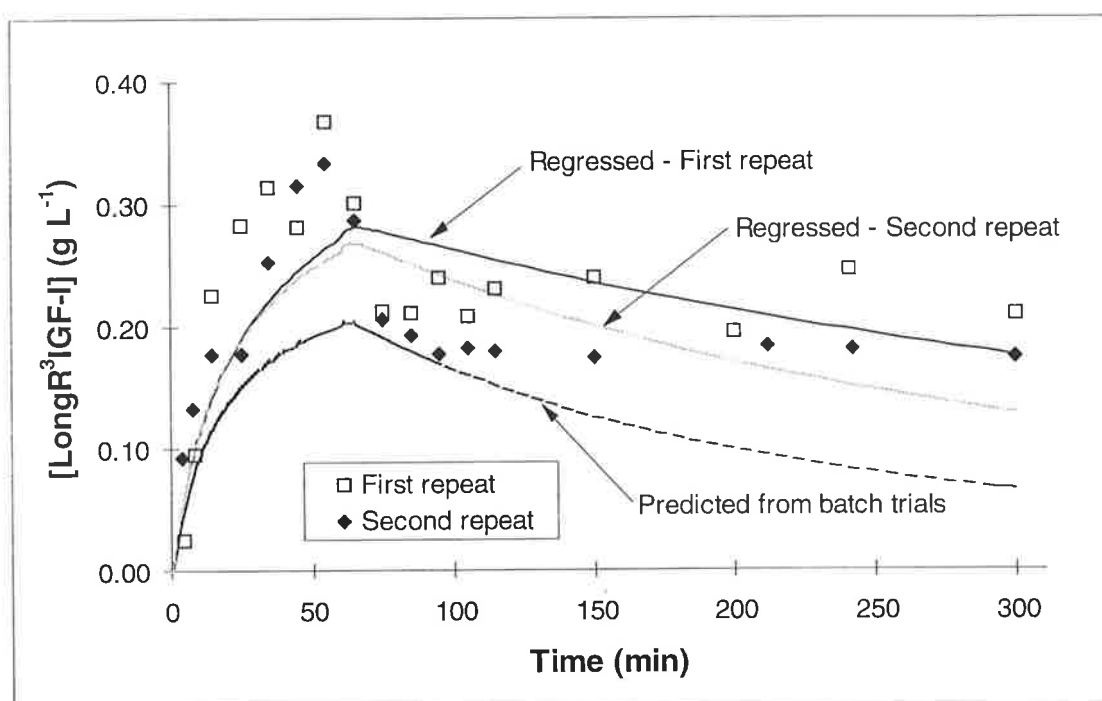


Figure 5.32: Plot of correctly folded LongR³IGF-I concentration versus time for both fed-batch refolding trials. The approximate concentration predicted from batch trials is shown as a dashed line. Regressed curves for each data set are shown for $k_N = 0.12 \text{ min}^{-1}$.

Figure 5.32 shows the predicted LongR³IGF-I concentration during refolding, based on the three-state kinetic model, using rate constants regressed from concentrated refolding trials (the average of values presented in Table 5.6). The shape of the curve during feeding follows the same trend as the experimental data, but at approximately half the protein concentration. Once feeding ceases, the model predicts a gradual decrease in the product concentration, but the observed behaviour was a much more rapid decrease to an apparently constant value. The discrepancy between predicted and observed behaviour suggests that the rate of aggregation during fed-batch refolding was less than detected in the concentrated refolding trials. A possible cause for this altered behaviour is the different mixing conditions used for fed-batch refolding.

Regression of the three-state kinetic model to the observed transient LongR³IGF-I concentration, to determine the three rate constants, proved to be difficult. The additional

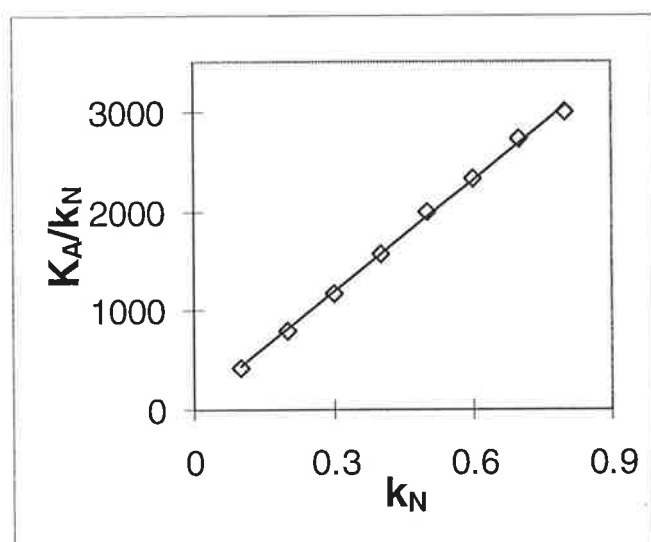


Figure 5.33: Plot of K_A/k_N versus k_N values that approximately describe the same experimental data for fed-batch refolding with $k_B = 0.013 \text{ min}^{-1}$. A line of best fit of the points is shown ($r^2 > 0.999$).

degrees of freedom introduced by the discontinuity in the feeding strategy allow a range of rate constants to describe the data. Figure 5.33 shows a plot of values for K_A/k_N versus k_N that describe the same observed behaviour to give approximately the same minimum SSE with $k_B = 0.013 \text{ min}^{-1}$. The high degree of correlation between K_A and k_N makes regression of transient protein concentration to data obtained from fed-batch refolding experiments difficult to regress. Therefore, to find a sensible converged solution the value of k_N

was fixed at the average rate constant regressed to the concentrated refolding data (*i.e.* 0.12 min^{-1}). Thus, the regressed values of K_A and k_B should account for the difference between the observed and predicted protein concentrations presented in Figure 5.32.

Values for the regressed rate constants K_A and k_B with $k_N = 0.12 \text{ min}^{-1}$ (fixed) are presented in Table 5.10. Even with one of the three parameters fixed, the proportionately

large values for the standard deviation indicate that the model is unable to describe the observed behaviour satisfactorily. The regressed curves for each data set are included in Figure 5.32. Given the relatively large uncertainty in the rate constants, the two fed-batch refolding trials can be regarded as being true repeats with a high degree of confidence (see Table 5.11).

Table 5.10: Estimates of rate constants and standard deviations from fed-batch refolding trials ($k_N = 0.12 \text{ min}^{-1}$).

	First repeat		Second repeat	
	Estimate	Std. deviation	Estimate	Std. deviation
$k_N (\text{min}^{-1})$	0.12	assumed = 0	0.12	assumed = 0
$K_A (\text{mL mg}^{-1} \text{min}^{-1})$	6.5	1.6	5.8	1.6
$k_B (\text{min}^{-1})$	0.0066	0.0024	0.0098	0.0030

Table 5.11: Analysis of variance between repeat refolding trials using fed-batch mode.

Source	df	SSE	MSE	F-statistic
Between repeats	3	0.0014	4.6×10^{-4}	0.63
Residual	28	0.0205	7.3×10^{-4}	
Total	31	0.0218		

Regression using the three-state kinetic model includes similar inadequacies in its description of the fed-batch data as in the description of the concentrated refolding behaviour; underpredicting the maximum concentration, not accounting for the rapid decrease in concentration once the maximum has been reached and trending to zero with subsequent incubation. A plot of the residuals is presented in Figure 5.34. Analysis of the residuals by the method of runs (Draper and Smith, 1966) confirms that the model gives a flawed description. The probability of obtaining this grouping of residuals is less than 7% relative to a normal distribution.

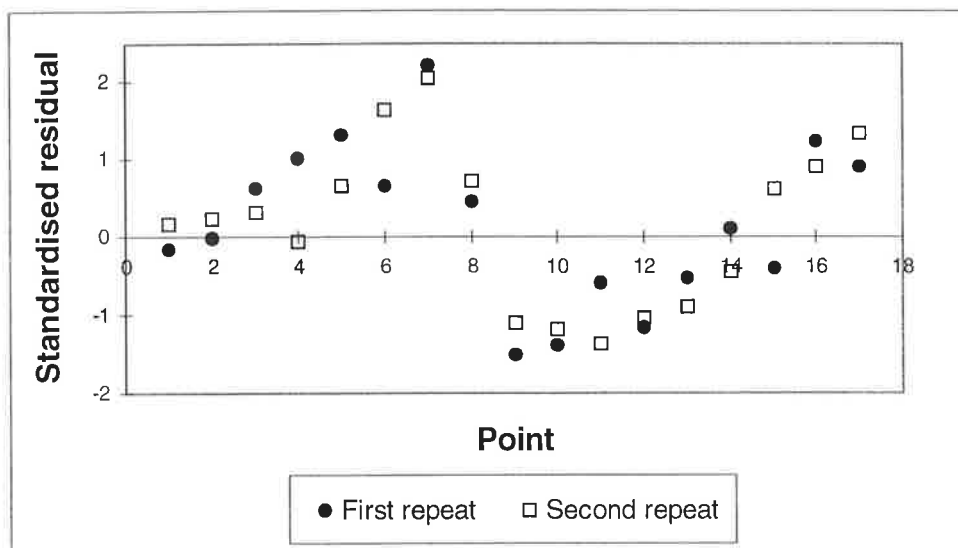


Figure 5.34: Plot of standardised residuals for both fed-batch refolding trials.

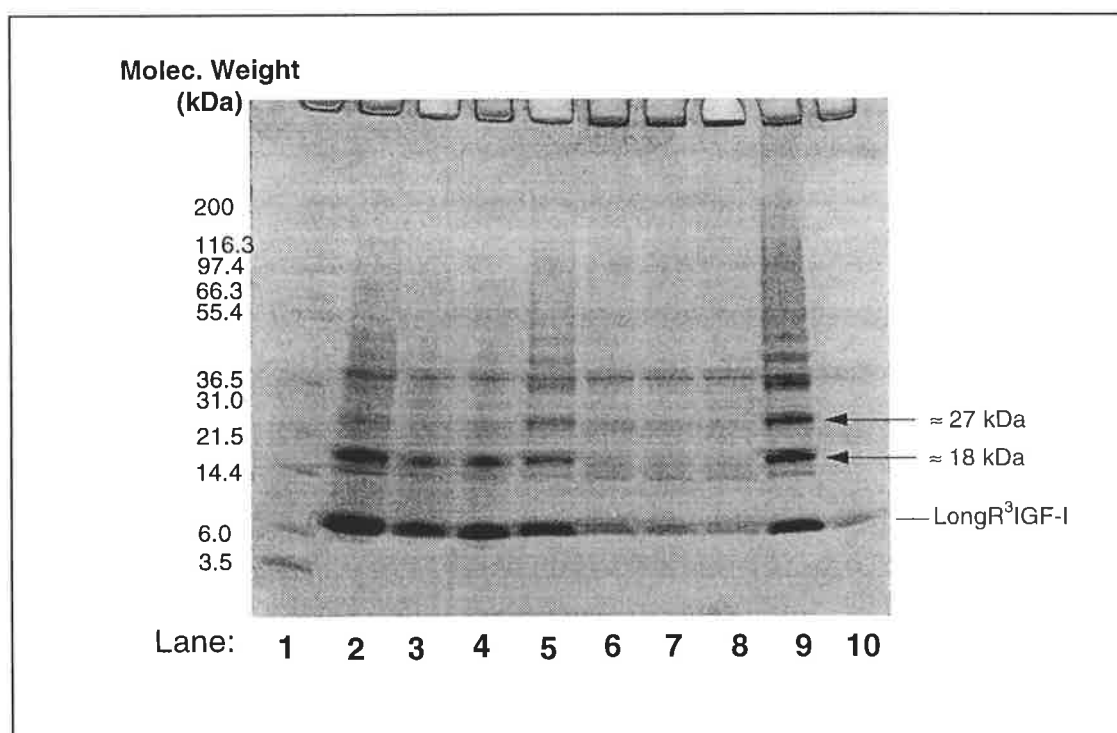


Figure 5.35: SDS-PAGE of refolding in fed-batch mode (10-20% Novex™ Tricine gel). *Lane 1:* Mark12 wide range protein standard, *Lane 2:* Dissolving mixture after 70 min (6 μ L, non-reducing), *Lane 3:* Refolding at termination of feeding (6 μ L, reducing), *Lane 4:* Refolding after 95 min. (6 μ L, reducing), *Lane 5:* Refolding after 6 min. (15 μ L, non-reducing), *Lane 6:* Refolding after 44 min. (15 μ L, non-reducing), *Lane 7:* Refolding after 62 min. (15 μ L, reducing), *Lane 8:* Refolding after 96 min. (15 μ L, reducing), *Lane 9:* Corresponding batch-refolded inclusion bodies after 60 minutes (10 μ L, reducing), *Lane 10:* LongR³IGF-I standard.

Selected results of SDS-PAGE of samples taken from the dissolved starting material and subsequent fed-batch refolding are presented in Figure 5.35. Lane 2 shows a typical separation of the proteins present in the dissolving mixture of inclusion bodies. Under reducing conditions, the notable change in the resolved bands was a decrease in the amount of 18 kDa species and a corresponding increase in the band at the correct position for monomeric LongR³IGF-I. Lanes 3 and 4 show the resolution of protein species present under reducing conditions in the refolding mixture for samples taken at 60 and 95 minutes, respectively. There is was a slight increase in the relative amount of protein present in the band corresponding to dimeric LongR³IGF-I aggregates in the latter sample. Lane 9 shows the distribution of protein species present in concentrated batch refolding under non-reducing conditions for comparison with the results of fed-batch refolding.

Lanes 5 to 8 show a time-course of samples taken from the refolding mixture under non-reducing conditions. After only 6 minutes, the ladder of bands with molecular weights corresponding to multimers of LongR³IGF-I can be seen clearly. Allowing for the greater volume of sample loaded, there is still significantly more protein in the sample taken after 6 minutes than subsequent samples. In the samples taken at 44, 60 and 95 minutes a significant amount of the protein can be seen trapped in the stacking gel. It appears that much of the protein had aggregated extensively and was too large to enter the running gel. By eye, the relative amount of multimers in the samples taken at later times appear to be much less than in the first sample. However, quantification of the ratio of monomeric to dimeric, trimeric, and tetrameric forms by densitometry does not substantiate this conclusion.

The relative amounts of multimeric to monomeric species in material refolded in batch mode was significantly greater than when fed-batch operation is employed. This indicates that aggregation of refolding monomers was reduced by gradual addition of the denatured protein. Quantification of the amount of correctly renatured LongR³IGF-I by HPLC supports this observation. Comparison of the HPLC profiles of batch and fed-batch refolding (Figures 5.25 and 5.31) demonstrates that the relative proportion of correctly renatured LongR³IGF-I was greater when fed-batch refolding was employed. However, comparison of material refolded in fed-batch and batch modes by SDS-PAGE, at approximately the same overall protein concentration, shows that there was much less protein resolved in the fed-batch samples (see lanes 7 and 9 in Figure 5.35). Values of total

soluble protein concentration for batch and fed-batch refolding suggest that a greater fraction of the protein remained soluble during batch refolding. Therefore, it appears that fed-batch refolding resulted in a greater yield of correctly renatured LongR³IGF-I than batch refolding, despite apparently having greater rates of association of monomeric and multimeric species. These observations add further weight to the case which suggests that the generic refolding pathway is not an accurate simplification for the refolding of LongR³IGF-I.

5.7 IMPLICATIONS OF EXPERIMENTAL FINDINGS

The work presented in this chapter has addressed some important issues that must be considered when refolding is conducted by dilution on a large scale. In the absence of a known kinetic refolding pathway, the generic scheme proposed in section 3.1.4 can be regressed to give an adequate description of the refolding of LongR³IGF-I. However, the observed behaviour of refolding at high protein concentration in batch and fed-batch modes demonstrates that the actual refolding pathway is more complex than the assumed simplified scheme.

The rate of mixing has been identified as a significant factor in determining the yield of correctly renatured LongR³IGF-I during refolding. Increased agitation results in an increased rate of off-pathway aggregation and loss of product.

The presence of soluble contaminants that are released from *E. coli* during homogenisation, decreased the yield of correctly renatured LongR³IGF-I. This decrease can be accounted for by an increase in the rate constant of off-pathway aggregation. There also appears to be an increase in the rate constant of correct refolding. However, the significance of this increase may be questioned because the rate constants for correct refolding and aggregation are highly correlated when regressed to the simplified kinetic model.

Renaturation of LongR³IGF-I from an unfractionated mixture of inclusion bodies and cell debris gave very poor recovery, relative to well purified inclusion bodies. Saturation of the chaotrope during dissolution appeared to be the main limiting factor. Non-proteinaceous contaminants were preferentially solubilised, leaving a significant fraction of the product insoluble. This result demonstrates the importance of characterising the effectiveness of the dissolution step.

It was not possible to accurately quantify the rate constants of correct refolding and aggregation using the unfractionated inclusion bodies as a starting material. The relatively low concentration of product in the starting material and poor dissolution behaviour made precise measurement of the amount of renatured product impossible using analytical HPLC. Hence, the application of the compartmentalised strategy, proposed in section 1.6.3, to account for cellular contamination during refolding, has only been partially

demonstrated for the soluble compartment. Insufficient data has been collected to determine conclusively whether the influence of *E. coli* contaminants on the refolding of LongR³IGF-I can be correlated with contaminating protein species.

To provide consistent denatured starting material for fed-batch refolding, it was necessary to use purified inclusion bodies and to conduct refolding before the reducing agent in the dissolution buffer was depleted. For the conditions employed in this study, dissolved LongR³IGF-I was stable for approximately one hour. Even given this restriction, fed-batch refolding at a constant rate gave a greater renaturation yield than batch dilution, when conducted at high concentration. This indicates that fed-batch operation is likely to be the favoured mode of operation for large-scale refolding.

6. GENERAL DISCUSSION

Economic production of recombinant proteins for widespread therapeutic application will benefit from development of efficient large-scale methods for renaturing product expressed as inclusion bodies. As outlined in section 1.8, the strategy adopted in this thesis has been to use the synergy of combined simulation and experimental investigation. A generic three-state kinetic model of refolding and compartmentalised strategy for dealing with host-cell contaminants has been proposed. The three-state kinetic model has been used in simulations and as a measure of performance to compare different refolding conditions.

Simulation has been used to investigate large-scale refolding in batch, fed-batch, and continuous modes. For the first time, a systematic comparison of operating modes has been conducted and optimum conditions identified using minimisation of cost as the objective. A range of preliminary experimental studies have been conducted to address the impact of mass-transfer, cellular contamination, and fed-batch operation during refolding.

In this chapter, the results from simulation and experimental investigation are discussed in conjunction with one another. The validity of the three-state model is questioned in light of the experimental results. Reports of refolding proteins in the presence of contaminants, published concurrently with this research, are presented and compared with the observed refolding of LongR³IGF-I. The deficiencies in this work are outlined and further investigations proposed.

6.1 MASS-TRANSFER

The review of the rate of denaturant removal presented in section 1.5.6.2, is concerned mainly with potential impact on the rate of correct refolding. The basis of the concern is production of unstable intermediates, leading to aggregation and product loss (Rudolph and Lilie, 1996). The primary aim in most reports regarding the rate of denaturant removal by dilution has been to disperse the denaturant and denatured product as quickly as possible to avoid heterogeneity and reduce aggregation (Goldberg *et al.*, 1991; Jaenicke and Rudolph, 1986). Therefore, rapidly mixed conditions are favoured (*e.g.* Hevehan and De Bernardez-Clark, 1997).

The results of this study indicate that for LongR³IGF-I, increasing the mixing rate only slightly increases the rate of correct refolding, but significantly increases the rate of aggregation. The slight increase in the rate of refolding may be an indication that the rate of denaturant removal does limit the measured rate of refolding. However, since the calculated rate constants for correct refolding and aggregation are highly correlated, the uncertainty in their regressed values precludes definitive support of this effect.

Reduced yield of correctly renatured LongR³IGF-I with increased agitation conflicts with the theory that higher shear rates reduce aggregation. A plausible explanation for this behaviour is that given the rate of mixing achieves sufficiently rapid dispersion, greater agitation increases the probability of stochastic association and accelerates the aggregation reaction. Implicit in this explanation is the assumption that at least one folding intermediate of LongR³IGF-I is prone to association. Since it was observed that the native protein was relatively stable yet there was product loss to aggregation, it is reasonable to make this assumption.

A ramification of this mixing behaviour on process simulation is that refolding systems may be designed to employ dual reactors. A turbulently mixed reactor may be used to provide initial dispersion of the denaturing agent and prevent localised heterogeneity under refolding conditions. A separate reactor may then be used to hold the refolding mixture under mildly agitated conditions during the slower events of refolding.

6.2 STARTING MATERIAL

As discussed in section 1.5, most studies into protein refolding have used highly purified protein at very low concentration. To realise the full benefit of process simulation in bioprocessing, models of protein refolding that take into account varying degrees of host-cell contamination and high protein concentration are required. When models of such sophistication have been developed, it should be possible to use the standard methods for optimal process synthesis, which are applied in more traditional chemical process industries. However, development of these models requires greater understanding of the interaction between refolding products and cellular contaminants.

In this thesis, experimental investigations into renaturation in the presence of significant concentrations of soluble and insoluble *E. coli* contaminants have been conducted. The amount of protein product in the purified inclusion body material used in the study accounted for approximately 60% of the total protein content. It has been unequivocally demonstrated that the extent of purification prior to inclusion body renaturation can have a dramatic impact on overall product recovery.

When cellular debris was not removed from the insoluble fraction of whole cell homogenate, dissolution was observed to occur at a relatively slow rate. Dilution to increase the relative amount of urea resulted in greater dissolution of protein from the unfractionated cell debris. Therefore, it can be concluded that the limitation was due to saturation of the dissolution buffer. The standard dissolution protocol used in this study employed a consistent volume to wet-weight ratio of dissolution buffer to inclusion body material. Hence, interaction of urea with non-proteinaceous contaminants must have been the cause of the limitation.

Concurrent to the work presented in this thesis, De Bernardez-Clark's group have investigated the effect of *E. coli* contaminants and select proteins on the refolding of inclusion bodies of macrophage-colony stimulating factor (M-CSF) and HEWL. Investigations were conducted to test the hypothesis that both proteinaceous and non-proteinaceous contaminants may interact with refolding products. Speed *et al.* (1996) have reported that during refolding of P22 tailspike and P22 coat proteins associations which result in aggregation are restricted to homologous interactions. They reported that the

aggregation of P22 tailspike protein could occur between multimers of any size, but may be reduced by step-addition of denatured protein or immobilisation on solid supports (Speed *et al.*, 1997). Dorin *et al.* (1995) reported that the renaturation of M-CSF was improved at least 3-fold by chromatography or diafiltration to remove *E. coli* proteases and nucleic acids. Rudolph and Lilie (1996) concluded that folding processes are not strongly affected by other proteins folding in the same renaturation environment.

Kinetic studies of the refolding of M-CSF from inclusion bodies were conducted using mixtures of crude dissolved inclusion bodies and mixtures of solubilised material that had been purified by reverse-phase chromatography (De Bernardez-Clark *et al.*, 1994). It was found that chromatographic purification did not improve the yield of renatured M-CSF, but did influence the rates of correct refolding and aggregation. HEWL was used as a model protein for investigating refolding kinetics in the presence of controlled amounts of isolated *E. coli* ribosomal RNA, L- α -phosphatidylethanolamine, lipopolysaccharide, plasmid DNA, β -galactosidase, bovine ribonuclease A, and BSA (De Bernardez-Clark *et al.*, 1994; Maachupalli-Reddy *et al.*, 1997). It was found that only the protein contaminants that co-aggregated during refolding (β -galactosidase and BSA) increased aggregation of HEWL. Ribonuclease A had no effect on refolding, nor did any the non-proteinaceous additives used. However, plasmid DNA and LPS increased the rate of aggregation and the phospholipid improved renaturation by reducing the rate of aggregation. These results demonstrate that the interactions between HEWL and contaminants are non-specific in nature, but may arise due to exposed hydrophobic groups in transiently populated molten-globule-like intermediate.

The work presented in this thesis has used a less elegant method to address similar issues using LongR³IGF-I as the model protein. Rather than characterise the effect of individual contaminants, whole-cell homogenate has been used. The findings demonstrate that LongR³IGF-I is quite stable to proteolysis and therefore does not require diafiltration or chromatographic purification to remove proteases before renaturation (unlike Dorin *et al.*, 1995). However, the trial conducted in the presence of all of the soluble contaminants released by homogenisation demonstrated conclusively that it was a significant increase in the rate of aggregation, rather than a reduction in the rate of refolding, that lead to a reduction in the yield of renatured product. This analysis is in agreement with the conclusions of De Bernardez-Clark *et al.* (1994) and Maachupalli-Reddy *et al.* (1997).

It can be concluded that aggregation of native LongR³IGF-I was negligible compared to aggregation of its folding intermediate(s). However, the specificity of association during refolding has not been conclusively determined (as discussed in section 5.2.2). Since the rate of aggregation was increased in the presence of cellular contaminants, at approximately the same product concentration, heterologous association seems most likely. The distinct aggregation of multimers during fed-batch refolding compared with batch refolding indicated that interaction of already associated species lead to the formation of large aggregates. This confirms the conclusions of Speed *et al.* (1997).

To avoid the problems of product quantification experienced in this study when renaturation was conducted in the presence of insoluble cellular contaminants, alternative methods must be employed. One method would be to increase the amount of denaturing agent added and spike the dissolving inclusion bodies to increase the relative amount of product. By conducting experiments in which different amounts of purified product are added, it should be possible to identify whether significant interactions between the product and contaminants are occurring and extrapolate back to the conditions used in this study. The poor dissolution characteristics of unfractionated inclusion bodies observed in this study demonstrated that purification of inclusion bodies prior to renaturation is likely to be advantageous. Hence, an alternative approach to characterise the impact of insoluble contaminants on refolding would be to fractionate the insoluble material to increase the relative amount of inclusion bodies. A range of partially purified inclusion body isolates could be used to characterise the renaturation at different levels of contamination. For example, inclusion bodies recovered from one, two, and three passes through a continuous centrifuge could be used. This approach would be the most pragmatic, as enrichment of the amount of inclusion body material in the insoluble fraction is a likely useful processing strategy to overcome adverse interactions with insoluble contaminants. Alternatively, more sensitive assays that are able to distinguish between low concentrations of different folding isomers in highly contaminated solutions could be used. Immuno-affinity assays may be able to provide this resolution, but they are likely to have significant intrinsic variability themselves.

The application of the compartmentalised strategy outlined in section 1.6.3 has been demonstrated to only a limited extent. It was proposed in order to provide a theoretical framework for analysing inclusion body renaturation. Since the findings of experiments

using unfractionated inclusion bodies were inconclusive and only one trial was conducted with additional soluble material, further work is required to fully establish its utility.

The simulation studies presented as part of this thesis assumed that the concentrated inclusion body slurry supplied from the centrifuge for subsequent renaturation was consistent. Accordingly, the impact of variation in the level of host-cell contamination did not need to be considered. The experimental results illustrate that the renaturation characteristics do vary considerably with the level of host-cell contamination. Therefore, this study reinforces the importance of developing models that account for different levels of host-cell contamination, before simulation of production strategies that use alternative upstream and midstream units is possible.

6.3 THREE-STATE REFOLDING KINETICS

A potential limitation of the kinetic models presented in sections 3.1.3 and 3.1.6 is that they were formulated to account for the observed behaviour of protein refolding in batch mode. It was also assumed that transition between the polymorphic protein states occurs via a single pathway in which the fast events of protein refolding may be ignored and the slower events give rise to the only significant intermediates. As outlined in section 1.5.6.1, it is only recently that reports of simulating refolding in fed-batch and continuous modes using kinetic models have been published. The models used in these simulations and in this thesis have been based on the earlier reported models of batch refolding.

When refolding is conducted by batch dilution to conditions that highly favour the native state, rapidly formed transitory states will be populated at relatively high concentration, but very briefly. If the states are prone to association, a range of aggregation rates may result. Slow events in structural formation will give rise to a single rate of correct refolding. Regression of both the batch and fed-batch refolding data obtained in this study has revealed that the rate constants for correct refolding and aggregation are highly correlated in a model of competing first-order refolding and second-order aggregation. Small changes in the rate of correct refolding can compensate for large variation in the rate of aggregation. Therefore, the generic model proposed in section 3.1.4 (see Figure 6.1(a)) should be expected to describe batch refolding well.

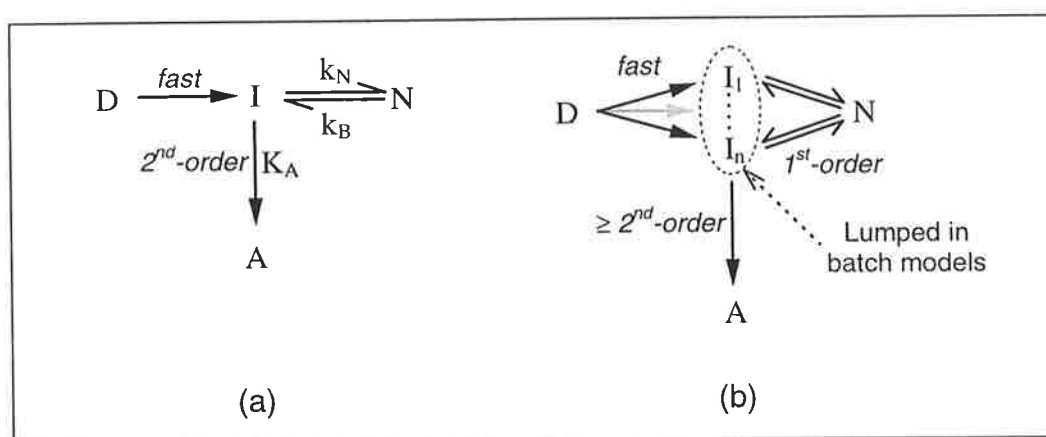


Figure 6.1: (a) the generic kinetic model of refolding proposed for use in this study (b) refolding pathway that includes multiple folding intermediates that may have different rates of correct refolding and aggregation.

In the batch refolding experiments to examine mixing effects, the three-state model provided a good description of native product accumulation. The amount of native protein was not observed to decrease significantly, indicating that the refolding conditions prevented conformational shuffling into states that were prone to aggregation.

In the refolding experiments conducted at high protein concentration, approximately six-fold greater concentrations of oxidised and reduced thiol agents were present. Under these conditions, the amount of correctly folded LongR³IGF-I decreased significantly, indicating that conformational rearrangements occurred more rapidly than in the mixing experiments. The three-state model did not give a good description of native product accumulation. It is likely that these conditions did not sufficiently favour refolding to the native state consequently giving rise to a number of intermediate states. The poor fit provided by the three-state model demonstrated that the state(s) which formed aggregates can not be described by a single on-pathway intermediate.

The same chemical environment as described above was maintained during fed-batch refolding trials. Cessation of feeding during the refolding process in fed-batch mode resulted in a more rapid decrease in the amount of native LongR³IGF-I than during batch refolding. Furthermore, a stable equilibrium concentration of active product was reached more quickly when refolding was conducted in fed-batch mode. The relatively rapid decrease to an apparently stable equilibrium concentration cannot be accounted for by the three-state model of refolding. The observed behaviour suggests that the refolding pathway is quite complex. It is possible that native product can be formed via more than one intermediate state and each intermediate may have a different rate of association. A plausible kinetic pathway to better describe the fed-batch refolding behaviour is illustrated in Figure 6.1(b). In this model multiple transient states are significantly populated because denatured protein is gradually added to the refolding reaction. Each intermediate makes a distinct contribution as it is kinetically accessible to both native and aggregated states.

The experimental results presented in this thesis validate the use of the simplified three-state kinetic model in simulation studies of refolding in batch mode. As discussed above, the three-state model does not describe the observed behaviour of fed-batch refolding well, indicating that an alternative model is required for accurate prediction of transient product concentration. However, in the absence of a more complex protein- and condition-specific model, it provides a useful approximation.

6.4 LARGE-SCALE REFOLDING

The work presented in this thesis, describes the first systematic comparison of refolding in batch, fed-batch, and continuous modes using overall process simulation with a kinetic model to optimise the concentration and the duration of refolding. The simulations assumed that the same kinetic model could be applied to all three refolding modes. They identified that refolding in fed-batch and continuous modes offered significant reduction in the cost of large-scale production. Fed-batch operation is likely to be the favoured mode for production of therapeutic proteins as batch integrity can be maintained. Use of dimensionless analysis would make the findings applicable to any protein that can be described by the three-state kinetic model of refolding.

The experimental studies have demonstrated conclusively that refolding in fed-batch mode is able to achieve greater yield of correctly refolded LongR³IGF-I than batch dilution. This improvement was achieved using exactly the same chemical additives (and hence consumable costs) as batch refolding. In fact, the observed recovery was greater than predicted using rate constants regressed to the three-state description of batch experiments at high protein concentration. It has also been shown that high mixing rates do not need to be maintained during refolding. Potentially, adequate dispersion of large volumes of denatured product to give suitable refolding conditions may be achieved at the onset of refolding, followed by incubation with gentle mixing.

Refolding by dilution in fed-batch mode does not promise the outstanding improvements in yield expected from other refolding methods. However, unlike many of these methods, fed-batch refolding does not require expensive additives or protein-specific cosolvents such as affinity ligands. Simulation studies have demonstrated that fed-batch operation has the potential for broad applicability. The method is relatively simple and has been shown to achieve significantly greater yields at high protein concentration than the generally used method of batch dilution. Therefore, refolding in fed-batch mode is likely to be a suitable operation for process-scale renaturation of recombinant proteins from inclusion bodies.

6.5 FUTURE WORK

Dissolution of inclusion bodies at the two different levels of purity used in this study has demonstrated the importance of the dissolution operation in renaturation. While conditions have generally been found to give total dissolution within an acceptable time, they have been selected by trial and error, rather than complete understanding of the fundamental mechanisms. The dissolution of spherical inclusion bodies lends itself to an engineering investigation. Further investigation of the mechanisms of dissolution may identify more efficient methods, and provide useful models for simulation.

The LongR³IGF-I concentration in the starting material used in the experimental investigations outlined in this thesis, was no more than 60% of the total protein content. Using this impure material, several significant observations were made that concern the mechanism of protein refolding.

The completed mixing studies have confirmed that agitation is an important parameter in determining the yield of correctly refolded product. Further investigations are required to adequately characterise the influence of agitation on aggregation. A protocol similar to that described here should be used to repeat experiments with a range of purified proteins. The results should then be correlated to the engineering groups used to describe mixing, *e.g.* Power or Kolmogoroff numbers.

Some indication of refolding limitation by the rate of denaturant removal was observed. Stop-flow experiments under different mixing conditions to investigate the early events in refolding may confirm whether this behaviour is a significant factor, or an artefact of the method of analysis. Subsequent studies should consider the influence of contaminants on the refolding process.

Fed-batch refolding gave a different distribution of LongR³IGF-I folding products than batch refolding. This indicates that the altered transient concentration of folding products, achieved by gradual addition of denatured protein, may provide insight into the pathway of refolding. The presence of *E. coli* contaminants reduced the accuracy of the method used to determine LongR³IGF-I concentration. Studies using a purified test protein should be able to provide more accurate data. The transient product accumulation observed in this investigation indicates that an alternative kinetic model is required to give improved

description of fed-batch refolding. Additionally, the optimal chemical environment for fed-batch refolding is likely to differ from that for dilute batch refolding because of the higher overall protein concentrations that may be used. The data obtained from a study using relatively contaminant-free protein would be most informative.

The compartmentalised strategy proposed to account for the interaction of cellular contaminants with the protein of interest during renaturation, may be pursued as a useful theoretical framework for identification of interacting species. Further work is required to determine whether it provides the basis for a model able to account for the presence of host-cell contaminants.

7. CONCLUSIONS

This study set out to investigate some of the key issues concerning renaturation of recombinant proteins from inclusion bodies. The intention was to apply fundamental chemical engineering theory to protein refolding. The approach taken was to select a model of protein refolding that was broadly applicable to a range of proteins. This model was then used in process simulation to compare different dilution modes for refolding. Experimental studies were conducted using LongR³IGF-I, expressed as inclusion bodies within *E. coli*, as a model system. Trials were completed to validate the findings of the simulation work, and characterise the impact of process modifications on the sensitivity of refolding rate constants.

The simulation studies revealed that, for large-scale operation, refolding using fed-batch mode to dilute denatured protein is potentially superior to the typically employed process of batch dilution. This simulation study is the first to systematically compare refolding in different modes and to use overall operating cost as a basis for optimising the duration and concentration of protein for refolding.

Most previous experimental studies on refolding have employed pure protein solutions, with few examining refolding from unfractionated inclusion body mixtures. In this study, relatively impure materials have been used, representative of typical process streams. It was found that for LongR³IGF-I, some degree of separation of cell debris from inclusion bodies is advantageous because the presence of contaminants limited the amount of product released during dissolution. Increasing mixing rate did not significantly alter the rate of correct refolding, but did result in a greater rate of aggregation and loss of product. Refolding in the presence of the soluble fraction of whole-cell homogenate reduced the recovery of active protein by increasing the rate of aggregation. Refolding by dilution in fed-batch mode did achieve significantly greater yield of active product than batch dilution.

The proposed three-state kinetic model does provide a good description of batch refolding by dilution to highly native-favouring conditions. However, only an approximate description is possible when less favourable conditions are used or fed-batch refolding is employed. The model can be used to describe refolding in the presence of host-cell contaminants. Future studies should aim to develop an accurate model of fed-batch refolding and correlate reaction rates with cellular contamination.

8. REFERENCES

- Allewell, N. M., (1993), Untangling the energetics of protein folding, *TIBS*, **18**: 1-2.
- Ambrosius, D. and Rudolph, R., (1997), Process for the reactivation of denatured protein, US patent **5,618,927**.
- Ambrosius, D., Dony, C., and Rudolph, R., (1996), Activation of recombinant proteins, US patent **5,578,710**.
- Anfinsen, C. B., (1973), Principles that govern the folding of protein chains, *Science*, **181**: 223-230.
- Atkinson, B. and Mavituna, F., (1991), *Biochemical engineering and biotechnology handbook*, 2nd Edition., Stockton Press, New York, pp. 1059-1109.
- Babbitt, P. C., West, B. L., Buechter, D. D., Kuntz, I. D., and Kenyon, G. L., (1990), Removal of a proteolytic activity associated with aggregates formed from expression of creatine kinase in *Escherichia coli* leads to improved recovery of active enzyme, *Bio/Technol.*, **8**: 945-949.
- Babbitt, P. C., West, B. L., Buechter, D. D., Chen, L. H., Kuntz, I. D., and Kenyon, G. L., (1991), Active creatine kinase refolded from inclusion bodies in *E. coli*. Improved recovery by removal of contaminating proteases, in *Protein Refolding*, Georgiou, G. and De Bernardez-Clark, E. (eds), *ACS symposium series*, **470**: 153-168.
- Baldwin, R. L., (1994), Matching speed and stability, *Nature*, **369**: 183-184.
- Ballew, R. M., Sabelko, J., and Gruebele, M., (1996), The initial collapse of apomyoglobin, *Proc. Natl. Acad. Sci. USA*, **93**: 5759-5764.
- Baneyx, F., Ayling, A., Palumbo, T., Thomas, D., and Georgiou, G., (1991), Optimisation of growth conditions for the production of proteolytically-sensitive proteins in the periplasmic space of *Escherichia coli*, *Appl. Microbiol. Biotechnol.*, **36**: 14-20.
- Baneyx, F. and Georgiou, G., (1990), *In vivo* degradation of secreted fusion proteins by the *Escherichia coli* outer membrane protease ompT, *J. Bacteriol.*, **172**: 491-494.
- Baneyx, F. and Georgiou, G., (1991), Construction and characterisation of *Escherichia coli* strains deficient in multiple secreted proteases: protease III degrades high molecular weight substrates *in vivo*, *J. Bacteriol.*, **173**: 2696-2703.
- Bastiras, S., (1992), Folding and conformational stability of porcine growth hormone, *Ph. D. dissertation, University of Adelaide*.
- Bastiras, S. and Wallace, J. C., (1992), Equilibrium denaturation of recombinant porcine growth hormone, *Biochemistry*, **31**: 9304-9309.
- Batas., B. and Chaudhuri, J. B., (1996), Protein refolding at high concentration using size-exclusion chromatography, *Biotechnol. Bioeng.*, **50**: 16-23.
- Belagaje, R. M., (1995), Method for increased expression of low molecular weight recombinant polypeptides, US patent **5,378,613**.

- Bentle, L. A., Mitchell, J. W., and Storrs, S. B., (1987), Method of somatotropin naturation,US patent **4,652,630**.
- Bird, R. B., Stewart, W. E., and Lightfoot, E. N., (1960), *Transport Phenomena*, John Wiley & Sons, Singapore, pp. 522.
- Bochkareva, E. S., Lissin. N.M.,and Girshovich, A. S., (1988), Transient association of newly synthesized unfolded proteins with the heat-shock GroEL protein, *Nature*, **336**: 254-257.
- Borden, K. L. and Richards, F. M., (1990), Folding kinetics of phage T4 thioredoxin, *Biochemistry*, **29**: 3071-3077.
- Bowden, G. A., Paredes, A. M., and Georgiou, G., (1991), Structure and morphology of protein inclusion bodies in *Escherichia coli*, *Bio/Technology*, **9**: 725-730.
- Bradford, M. M., (1976), A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding., *Anal. Biochem.*, **72**: 248-252.
- Brandts, J. F., Halvorson, H. R., and Brennan, M., (1975), Consideration of the possibility that the slow step in protein denaturation is due to *cis-trans* isomerism of proline residues, *Biochemistry*, **14**: 4953-4963.
- Brems, D. N., (1988), Solubility of different folding conformers of bovine growth hormone, *Biochemistry*, **27**: 4541-4546.
- Brems, D. N., Plaisted, S. M., Kauffman, E. W., and Havel, H. A., (1986), Characterization of an associated equilibrium folding intermediate of bovine growth hormone, *Biochemistry*, **25**: 6539-6543.
- Brems, D. N., Plaisted, S. M., Dougherty, Jr., J. J., and Holzman, T. F., (1987), The kinetics of bovine growth hormone folding are consistent with a framework model, *J. Biol. Chem.*, **262**: 2590-2596.
- Brierley, R. A., Abrams, J. N., Hanson, J. M., and Maslanka, F. C., (1997), IGF-I purification process, US patent **5,650,496**.
- Brock, T. D., Smith, D. W., and Madigan, M. T., (1984), *Biology of microorganisms*, 4th Edition, Prentice/Hall International, New Jersey, pp. 14-56.
- Buchner, J., (1996), Supervising the fold: functional principles of molecular chaperones, *FASEB*, **10**: 10-19.
- Buchner, J. and Rudolph, R., (1991), Renaturation, purification, and characterization of recombinant F_{ab}-fragments produced in *Escherichia coli*, *Bio/Technol.*, **9**: 157-162.
- Buchner, J., Schmidt, M., Fuchs, M., Jaenicke, R., Rudolph, R., Schmid, F. X., and Kiefhaber, T, (1991), GroE facilitates refolding of citrate synthase by supressing aggregation, *Biochemistry*, **30**: 1586-1591.

- Builder, S. E. and Ogez, J. R., (1986), Purification and activity assurance of precipitated heterologous proteins, US patent **4,620,948**.
- Builder, S. E., Belmont, J. R., and Ogez, S. M., (1986), Purification and activity assurance of precipitated heterologous proteins, U. S. patent **4,620,948**.
- Burton, S. J., Quirk, A. V., and Wood, P. C., (1989), Refolding human serum albumin at relatively high protein concentration *Eur. J. Biochem.*, **179**: 379-387.
- Cardamone, M., Puri, N., and Brandon, M. R., (1995), Comparing the refolding and reoxidation of recombinant porcine growth hormone for a urea denatured state and from *Escherichia coli* inclusion bodies, *Biochemistry*, **34**: 5773-5794.
- Cerny, G. and Teuber, M., (1971), Differential release of periplasmic *versus* cytoplasmic enzymes from *Escherichia coli* B by polymyxin B, *Arch. Mikrobiol.*, **78**: 166-179.
- Chang, J. Y., McFarland, N. C., and Swartz, J. R., (1995), Method for refolding insoluble, misfolded Insulin-like Growth Factor-I into an active conformation, US patent **5,410,026**.
- Chang, J. Y., McFarland, N. C., and Schwartz, J. R., (1994), Method for refolding insoluble, misfolded Insulin-Like Growth Factor-I into an active conformation, US patent **5,288,931**.
- Chaudhuri, J. B., (1994a), Refolding recombinant proteins: process strategies and novel approaches, *Ann. N. Y. Acad. Sci.*, **721**: 374-385.
- Chaudhuri, J. B., Refolding recombinant proteins - Protein chemistry vs process engineering, (1994b), in *ECB6: Proceedings of the 6th European Congress on Biotechnology, Florence, Italy, June 13-15, 1993*, Alberghina, L., Frontali, L., and Sensi, P. (eds), Elsevier Science, pp. 107-110.
- Chaudhuri, J. B. and Mall, S., (1994), Refolding proteins in reverse micelles: choice of system composition and transfer of denatured protein, *presented at Recovery of Biological Products VII, September 25-30, 1994, San Diego, California*.
- Chaudhuri, J. B., Batas, B., and Guise, A. D., (1996), Improving protein refolding yields by minimizing aggregation, *Ann. N. Y. Acad. Sci.*, **782**: 495-505.
- Cheah, K.-C., Harrison, S., King, R., Crocker, R., Wells, J. R. E., and Robins, A., (1994), Secretion of eukaryotic growth hormones in *Escherichia coli* is influenced by the sequence of the mature proteins, *Gene*, **138**: 9-15.
- Chung, C. H., (1993), Proteases in *Escherichia coli*, *Science*, **262**: 372-374.
- Cleland, J. L. and Wang, D. I. C., (1993), *In vitro* protein refolding, in *Biotechnology*, 2nd Edition, vol. 3, Bioprocessing, Stephanopoulos, G., (ed.), VHC Publishers Inc., New York, NY, pp. 527-555.
- Cleland, J. L. and Wang, D. I. C., (1992), Transient association of the first intermediate during the refolding of carbonic anhydrase B, *Biotechnol. Prog.*, **8**: 97-103.

- Cleland, J. L. and Wang, D. I. C., (1991), Equilibrium association of a molten globule intermediate in the refolding of bovine carbonic anhydrase, in *Protein Refolding*, Georgiou, G. and De Bernadez-Clark, E. (eds), *ACS symposium series*, **470**: 169-179.
- Cleland, J. L. and Wang, D. I. C., (1990a), Refolding and aggregation of bovine carbonic anhydrase B: quasi-elastic light scattering analysis, *Biochemistry*, **29**: 11072-11078.
- Cleland, J. L. and Wang, D. I. C., (1990b), Cosolvent assisted protein refolding, *Bio/Technol.*, **8**: 1274-1278.
- Cleland, J. L. and Wang, D. I. C., (1990c), Cosolvent effects on refolding and aggregation, in *Biocatalyst design for stability and specificity*, Himenl, M. E. and Georgiou, G. (eds), *ACS symposium series*, **516**: 151-166.
- Cockshott, A. R. and Bogel, I. D. L., (1992), Modelling a recombinant *E. coli* fermentation producing bovine somatotropin, in *Modeling and control of biotechnical processes 1992, (2nd IFAC Symposium) and Computer applications in fermentation technology (5th International Conference) : selected papers from the IFAC and ICCAFT 5 meetings, Keystone, Colorado, USA, 29 March - 2 April 1992*, Nazrul Kamin, M. and Stephanoloulos, G., (eds), Pergamon Press, N. Y., pp. 219-222.
- Creighton, T. E., (1995), Disulfide-coupled protein folding pathways, *Phil. Trans. R. Soc. Lond.*, **348**: 5-10.
- Creighton, T. E., (1987), Protein folding, in *Protein Engineering*, Oxender, D. L. and Fox, C. F. (eds), Alan R. Liss, NY, pp. 83-90.
- Creighton, T. E., (1986), Disulfide bonds as probes of protein folding pathways, *Method. Enzymol.*, **131**: 83-106.
- Creighton, T. E., (1978), Experimental studies of protein folding and unfolding, *Prog. molec. Biol.*, **33**: 231-297.
- Creighton, T. E. and Goldenberg, D. P., (1984), Kinetic role of a meta-stable native-like two-disulfide species in the folding transition of bovine pancreatic trypsin inhibitor, *J. Mol. Biol.*, **179**: 497-526.
- Creighton, T. E. and Pain, R. H., (1980), Unfolding and refolding of *Staphylococcus aureus* penicillinase by urea-gradient electrophoresis, *J. Mol. Biol.*, **137**: 431-436.
- Creighton, T. E., Darby, N. J., and Kemmink, J., (1996), The roles of partly folded intermediates in protein folding, *FASEB*, **10**: 110-118.
- Cronan, J. E., Gennis, R. B., and Maloy, S. R., (1987), Cytoplasmic membrane, in *Escherichia coli and Salmonella typhimurium, Cellular and Molecular Biology*, Vol. 1, Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M., and Umberger (eds), American Society for Microbiology, Washington, D. C., pp. 31-55.
- Darling, T. L. J., Akhnana, L. Y., Mitschelen, J. L., and Hrinda, M. E., (1996), Methods for purification of recombinantly produced proteins, US patent **5,530,100**.

- Datar, R. V., Cartwright, T., and Rosen, C.-G., (1993), Process economics of animal cell and bacterial fermentations: a case study analysis of Tissue Plasminogen Activator, *Bio/Technology*, **11**: 349-357.
- Datar, R., and Rosen, C.-G., (1987), Centrifugal separation in the recovery of intracellular protein from *E. coli*, *The Chemical Engineering Journal*, **34**: B49-B56.
- De Bernardez-Clark, E. and Georgiou, G., (1991), Inclusion bodies and recovery of proteins from the aggregated state, in *Protein Refolding*, Georgiou, G. and De Bernardez-Clark, E. (eds), *ACS symposium series*, **470**: 1-20.
- De Bernardez-Clark, E., Maachupalli, J., Kelley, B., and Schauer, N., (1994), Effect of impurities on inclusion body protein refolding, presented at *Recovery of Biological Products VII, September 25-30, 1994, San Diego, California*.
- De Young, L. R., Burton, L. E., Liu, J., Powell, M. F., Schmelzer, C. H., and Skelton, N. J., (1996), RhNGF slow unfolding is not due to proline isomerisation: Possibility of a cystine knot loop-threading mechanism, *Protein Science*, **5**: 1554-1566.
- Diedrich, D. L., Stein, M. A., and Schnaitman, C. A., (1990), Association of *Escherichia coli* K-12 ompF trimers with rough and smooth lipopolysaccharides, *J. Bacteriol.*, **172**: 5307-6311.
- Dill, K. A., (1990), Dominant forces in protein folding, *Biochemistry*, **29**: 7133-7155.
- Dill, K. A., (1987), The Stabilities of Globular Proteins, in *Protein Engineering*, Alan R. Liss Inc, NY, pp. 187-192.
- Dill, K. A., (1985), Theory for the folding and stability of globular proteins, *Biochemistry*, **24**: 1501-1509.
- Dill, K. A. and Shortle, D., (1991), Denatured states of proteins, *Ann. Rev. Biochem.*, **60**: 795-825.
- Dorin, G., Gray, D. R., Chang, B. S., Cowgill, C. A., and Milley, R. J., (1995), Process for purifying bacterially produced M-CSF, US patent **5,466,781**.
- Draper, N. R. and Smith, H., (1966), Applied regression analysis, John Wiley & Sons, New York, pp. 95-98.
- Enfors, S.-O., (1992), Control of *in vivo* proteolysis in the production of recombinant proteins, *Trends in Biotechnology*, **10**: 310-315.
- Ernst, S, Garro, O. A., Winkler, S., Venkataraman, G., Langer, R., Cooney, C. L., and Sasisekharan, R., (1997), Process simulation for recombinant protein production: cost estimation and sensitivity analysis for heparinase I expressed in *Escherichia coli*, *Biotechnol. Bioeng.*, **53**: 575-582.
- Estapé, D. and Rinas, U., (1996), Optimized procedures for purification and solubilization of basic Fibroblast Growth Factor inclusion bodies, *Biotechnology Techniques*, **10**: 481-484.

- Fischer, B., Perry, B., Sumner, I., and Goodenough, P., (1992), A novel sequential procedure to enhance the renaturation of recombinant protein from *Escherichia coli* inclusion bodies, *Prot. Eng.*, **5**: 593-596.
- Fischer, B., Sumner, I., and Goodenough, P., (1993a), Isolation, renaturation, and formation of disulfide bonds of eukaryotic proteins expressed in *Escherichia coli* as inclusion bodies, *Biotechnol. Bioeng.*, **41**: 3-13.
- Fischer, B., Sumner, I., and Goodenough, P., (1993b), Renaturation of lysozyme - temperature dependence of renaturation rate, renaturation yield, and aggregation: identification of hydrophobic folding intermediates, *Arch. Biochem. Biophys.*, **306**: 183-187.
- Fish, N. M. and Hoare, M., (1988), Recovery of protein inclusion bodies, *Biochem. Soc. Trans.*, **16**: 102-104.
- Fishov, I., Zaritsky, A., and Grover, N. B., (1995), On microbial states of growth, *Molecular Microbiology*, **15**: 789-794.
- Francis, G. L., Ross, M., Ballard, F. J., Milner, S. J., Senn, C., McNeil, K. A., Wallace, J. C., King, R., and Wells, J. R. E., (1992), Novel recombinant fusion protein analogues of Insulin-like Growth Factor (IGF)-I indicate the relative importance of IGF-binding proteins and receptor binding for enhanced biological potency, *J. Mol. Endo.*, **8**: 213-223.
- French, C., Keshavarz-Moore, E., and Ward, J. M., (1996), Development of a simple method for the recovery of recombinant proteins from the *Escherichia coli* periplasm, *Enzyme Microb. Technol.*, **19**: 332-338.
- Fulton, S. P., Shahidi, A. J., Gordon, N. F., and Afeyan, N. B., (1992), Large-scale processing and high-throughput perfusion chromatography, *Bio/Technology*, **10**: 635-639.
- Gallihier, P. M., (1991), Continuous method of refolding proteins, US patent **4,999,422**.
- Garel, J.-R., (1992), Folding of large proteins: Multidomain and multisubunit proteins, in *Protein Folding*, Creighton, T. E. (ed.), W. H. Freeman and Company, New York, pp. 405-454.
- Gatenby, A. A., Donaldson, G. K., Baneyx, F., Lorimer, G. H., Viitanen, P. V., and van der Vies, S. M., (1993), Participation of GroE heat shock proteins in polypeptide folding, in *Biocatalyst design for stability and specificity*, Himenl, M. E. and Georgiou, G. (eds), *ACS symposium series*, **516**: 140-150.
- Georgiou, G., (1988), Optimizing the production of recombinant proteins in microorganisms, *AIChE Journal*, **34**: 1233-1248.
- Georgiou, G. and Baneyx, F., (1996), Bacterial host strains for producing proteolytically sensitive polypeptides, US patent **5,508,192**.

- Georgiou, G. and Valax, P., (1996), Expression of correctly folded proteins in *Escherichia coli*, *Current Opinion in Biotechnology*, **7**: 190-197.
- Georgiou, G., Francisco, J. A., and Earhart, C. F., (1994), Expression of proteins on bacterial surface, US patent **5,348,867**.
- Gold, L., (1990), Expression of heterologous proteins in *Escherichia coli*, *Meth. Enzymol.*, **185**: 11-14.
- Goldberg, M. E., Rudolph, R., and Jaenicke, R., (1991), A kinetic study of the competition between renaturation and aggregation during the refolding of denatured-reduced egg white lysozyme, *Biochemistry*, **30**: 2790-2797.
- Gordon, N. F., Moore, C. M. V., and Cooney, C. L., (1990), An overview of continuous protein purification processes, *Biotech. Adv.*, **8**: 741-762.
- Gottesman, S. (1989), Genetics of proteolysis in *Escherichia coli*, *Annu. Rev. Genet.*, **23**: 163-198.
- Gottesman, S. (1990), Minimizing proteolysis in *Escherichia coli*: genetic solutions, *Meth. Enzymol.*, **185**: 119-129.
- Grafl, R., Lang, K., Wiba, A., and Schmid, F. X., (1986), Folding mechanisms of porcine ribonuclease, *J. Mol. Biol.*, **191**: 281-293.
- Grodberg, J. and Dunn, J. J., (1988), OmpT encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification, *J. Bacteriol.*, **170**: 1245-1253.
- Grossman, I. E. and Biegler, L. T., (1995), Optimizing chemical processes, *Chemtech*, **25**: 27-35.
- Hagen, A. J., Hatton, T. A., and Wang, D. I. C., (1990a), Protein refolding in reverse micelles: interactions of the protein with micelle components , *Biotechnol. Bioeng.*, **35**: 966-975.
- Hagen, A. J., Hatton, T. A., and Wang, D. I. C., (1990b), Protein refolding in reverse micelles, *Biotechnol. Bioeng.*, **35**: 955-965.
- Hamel, J.-F. P., King, J., Cooney, C. L., and Dervakos, G., (1996), Production and recovery of soluble and insoluble protein: an economic perspective in bioprocessing, *The 1996 IChemE Research Event*, **1**: 163-165.
- Hammarberg, B., Moks, T., Tally, M., Elmlblad, A., Holmgren, E., Murby, M., Nilsson, B., Josephson, S., and Uhlén, M., (1990), Differential stability of recombinant human Insulin-like Growth Factor II in *Escherischia coli* and *Staphylococcus aureus*, *J. Biotechnol.*, **14**: 423-438.
- Harcum, S. W. and Bentley, W. E., (1993), Response dynamics of 26-, 34-, 39-, 54-, and 80- kDa proteases in induced cultures of recombinant *E. coli*, *Biotechnol. Bioeng.*, **42**: 675-685.

- Hart, R. A. and Bailey, J. E., (1992), Solubilisation and regeneration of *Vitreoscilla* hemoglobin isolated from protein inclusion bodies, *Biotechnol. Bioeng.*, **39**: 1112-1120.
- Hart, R. A., Lester, P. M., Reifsnyder, D. H., Ogez, J. R., and Builder, S. E., (1994), Large scale, *in situ* isolation of periplasmic IGF-I from *E. coli*, *Bio/Technology*, **12**: 1113-1117.
- Hart, R. A., Rinas, U., and Bailey, J. E., (1990), Protein composition of *Vitreoscilla* hemoglobin inclusion bodies produced in *Escherichia coli*, *J. Biol. Chem.*, **265**: 12728-12733.
- Hartley, D. L. and Kane, J. F., (1988), Recovery and reactivation of recombinant proteins, *Biochem. Soc. Trans.*, **16**: 101-102.
- Hatzfeld, M., Dodemont, H., Plessmann, U., and Weber, K., (1992), Truncation of recombinant vimentin by OmpT, *FEBS Lett.*, **302**: 239-242.
- Hanemaaijer, R., Westphal, A. H., Berg, A., Van Dongen, W., de Kok, A., Veeger, C., (1989), The gene encoding dihydrolipoyl transacetylase from *Azobacter vinelandii*. Expression in *Escherichia coli* and activation and isolation of the protein, *Eur J. Biochem.*, **181**: 47-53.
- Havel, H. A., Kauffman, E. W., Plaisted, S. M., and Brems, D. N., (1986), Reversible self-association of bovine growth hormone during equilibrium unfolding, *Biochemistry*, **25**: 6533-6538.
- Hejnæs, K.R., Bayne, S., Nørskov, L., Holmegarrd Sørensen, H., Thomsen, J., Schäffer, L., Wollmer, A., and Skriver, L., (1992), Development of an optimized refolding process for recombinant Ala-Glu-IGF-1, *Prot. Eng.*, **5**: 597-806.
- Hershenson, S., Shaked, Z., and Thomson, J., (1990), Process for recovering microbially produced interferon- β , US patent **4,961,969**.
- Hevehan, D. L. and De Bernardez Clark, E., (1997), Oxidative renaturation of lysozyme at high concentrations, *Biotechnol. Bioengng*, **54**: 221-230.
- Hibbard, L. S. and Tulinsky, A., (1978), Expression of functionality of α -chymotrypsin. Effects of guanidine hydrochloride and urea on the onset of denaturation., *Biochemistry*, **17**: 5460-5468.
- Hober, S., Forsberg, G., Palm, G., Hartmanis, M., and Nilsson, B., (1992), Disulphide exchange folding of Insulin-like Growth Factor I, *Biochemistry*, **31**: 1749-1756.
- Hoenger, A., Ghosh, R., Schoenenberger, C.-A., Aepli, U., and Engel, A., (1993), Direct *in situ* structural analysis of recombinant outer membrane porins expressed in an ompA-deficient mutant *Escherichia coli* strain, *J. Struct. Biol.*, **111**: 212-222.
- Hoess, A., Arthur, A. K., Wanner, G., and Fanning, E., (1988), Recovery of soluble, biologically active recombinant proteins from total bacterial lysates using ion exchange resin, *Bio/Technology*, **6**: 1214-1217.

- Horowitz, P. M., (1993), Facilitation of protein folding and the reversibility of denaturation in *Biocatalyst design for stability and specificity*, Himenl, M. E. and Georgiou, G. (eds), *ACS symposium series*, **516**: 167-172.
- Hua, Q.-X., Narhi, L., Jia, W., Arakawa, T., Rosenfeld, R., Hawkins, N., Miller, J. A., and Weiss, M. A., (1996), Native and non-native structure in a protein-folding intermediate: spectroscopic studies of partially reduced IGF-I and an engineered alanine model, *J. Mol. Biol.*, **259**: 297-313.
- Humbel, R. E., (1990), Insulin-like growth factors I and II, *Eur. J. Biochem.*, **190**: 445-462.
- Humphreys, D. P., Weir, N., Lawson, A., Mountain, A., and Lund, P. A., (1996), Co-expression of human protein disulphide isomerase (PDI) can increase the yield of an antibody F_{ab}-fragment expressed in *Escherichia coli*, *FEBS Letters*, **380**: 194-197.
- Jaenicke, R., (1995), Folding and association versus misfolding and aggregation of proteins, *Phil. Trans. R. Soc. Lond.*, **348**: 97-105.
- Jaenicke, R., (1991), Protein folding: local structures, domains, subunits, and assemblies, *Biochemistry*, **30**: 3147-3161.
- Jaenicke, R., (1989), Thermodynamic and kinetic aspects of the refolding and selfassembly of proteins, *GBF Monogr.*, **12**: 3-11.
- Jaenicke, R., (1987), Folding and association of proteins, *Prog. Biophy. Mol. Biol.*, **49**: 117-237.
- Jaenicke, R., (1974), Reassociation and reactivation of lactic dehydrogenase from the unfolded subunits, *Eur. J. Biochem.*, **46**: 149-155.
- Jaenicke, R. and Rudolph, R., (1986), Refolding and Association of Oligomeric Proteins, *Methods in Enzymology*, **131**: 218-250.
- Jin, K., Thomas, O. R. T., and Dunnill, P., (1994), Monitoring recombinant inclusion body recovery in an industrial disc stack centrifuge, *Biotech. Bioeng.*, **43**: 455-460.
- Jocelyn, P. C., (1972), *Biochemistry of the SH group*, Academic Press, London, pp. 51-55.
- Johnson, B. H. and Hecht, M. H., (1994), Recombinant proteins can be isolated from *E. coli* cells by repeated cycles of freezing and thawing., *Bio/Technology*, **12**: 1357-1360.
- Kalk, J. P. and Langlykke, A. F., (1986), in *Manual of Industrial Microbiology and Biotechnology*, Demain, A. L. and Solomon, N. A. (ed.), American Society for Microbiology, Washington, DC, pp. 363-385.
- Kane, J. F. and Hartley, D. L., (1988), Formation of recombinant protein inclusion bodies in *Escherichia coli*, *Tibtech*, **6**: 95-101.
- Kane, J. F. and Hartley, D. L., (1991), Properties of recombinant protein containing inclusion bodies in *Escherichia coli*, in *Purification and analysis of recombinant*

- proteins*, Seetharam, R. and Sharma, S. K. (eds), Marcel Dekker Inc., NY, pp. 121-145.
- Kaufmann, A., Stierhof, Y.-D., and Henning, U., (1994), New outer membrane-associated protease of *Escherichia coli* K12, *J. Bacteriol.*, **176**: 359-367.
- Kaul, R. and Mattiasson, B., (1992), Secondary purification, *Bioseparation*, **3**: 1-26.
- Kazan, D., Camurdan, A., and Hortacsu, A., (1995), The effect of glucose concentration on the growth rate and some of the intracellular components of a recombinant *E. coli* culture., *Process Biochemistry*, **30**: 269-273.
- Khurgin, Y. and Maksareva, E., (1993), Urea-generated free rotating water molecules are active in the protein unfolding process, *FEBS Letters*, **315**: 149-152.
- Kiefhaber, T., and Schmid, F. X., (1992a), Kinetic coupling between protein folding and prolyl isomerisation. I. Theoretical models, *J. Mol. Biol.*, **224**: 217-229.
- Kiefhaber, T., and Schmid, F. X., (1992b), Kinetic coupling between protein folding and prolyl isomerisation. II. Folding of ribonuclease A and ribonuclease T₁, *J. Mol. Biol.*, **224**: 231-240.
- Kiefhaber, T., Quaas, R., Hahn, U., and Schmid, F. X., (1990a), Folding of ribonuclease T₁. 1. Existence of multiple unfolded states created by proline isomerization, *Biochemistry*, **29**: 3053-3061.
- Kiefhaber, T., Quaas, R., Hahn, U., and Schmid, F. X., (1990b), Folding of ribonuclease T₁. 2. Kinetic models for the folding and unfolding reactions, *Biochemistry*, **29**: 3061-3070.
- Kiefhaber, T., Rudolph, R., Kohler, H.-H., and Buchner, J., (1991), Protein aggregation *in vitro* and *in vivo*: A quantitative model for the kinetic competition between folding and aggregation, *BioTechnology*, **9**: 825-829.
- Kim, P. S. and Baldwin, R. L., (1990), Intermediates in the folding reactions of small proteins, *Ann. Rev. Biochem.*, **59**: 631-660.
- Kim, P. S. and Baldwin, R. L., (1982), Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding, *Ann. Rev. Biochem.*, **51**: 459-489.
- Kim., S.-O. and Lee, Y. I., (1997), Enhanced *in vitro* refolding selectivity of the recombinant human Insulin-like Growth Factor I, *Biotechnology Techniques*, **11**: 85-89.
- King, J., Haase-Pettingell, C., Robinson, A. S., Speed, M., and Mitraki, A., (1996), Thermolabile folding intermediates: inclusion body precursors and chaperonin substrates, *FASEB.*, **10**: 57-66.
- King, R., Wells, J. R. E., Kreig, P., Snoswell, M., Brazier, J., Bagley, C. J., Wallace, J. C., Ballard, F. J., Ross, M., and Francis, G. L., (1992), Production and characterisation

of recombinant Insulin-like Growth Factor-I (IGF-I) and potent analogues of IGF-I, with Gly of Arg substituted for Glu³, following their expression in *Escherichia coli* as fusion proteins, *J. Mol. Endo.*, **8**: 29-41.

Kleinig, A. R. and Middelberg, A. P. J., (1997), Numerical and experimental study of a homogenizer impinging Jet, *AIChE J.*, **43**: 1100-1107.

Knuth, M. W. and Burgess, R. R., (1987), in *Protein purification: micro to macro.*, Burgess, R. (ed.), Alan R. Liss, NY, pp. 279-305.

Ko, Y.-F., Bentley, W. E., and Weigand, W. A., (1994), A metabolic model of cellular energetics and carbon flux during aerobic *Escherichia coli* fermentation, *Biotechnol. Bioengng*, **43**: 847-855.

Köhler, K., Nilsson, B., and Veide, A., (1993), Recovery of extracellular human Insulin-like Growth Factor-I and II as a fusion protein from *Escherichia coli* culture broth by aqueous two-phase extraction, *Bioseparation*, **3**: 241-250.

Kotlarski, N., O'Neill, B. K., Francis, G. L., and Middelberg, A. P. J., (1997), Design analysis for refolding of monomeric protein, *AIChE J.*, **43**: 2123-2132.

Kotlarski, N., Yeates, R. S., Milner, S. J., Francis, G. L., O'Neill, B. K., and Middelberg, A. P. J., (1995), Studies into the scale-up of a process to produce a biosynthetic Insulin-like Growth Factor, *Trans. Inst. Chem. Eng. C*, **73**: 27-32.

Kossik, J., (1996), Transferring unit operations from chemical process industries to biotech, *GEN*, **16(18)**: 15.

Kresze, G.-B., (1991), Proteases during purification, in *Purification and analysis of recombinant proteins*, Seetharam, R. and Sharma, S. K. (eds), Marcel Dekker, NY, pp. 85-120.

Kuehl, R. O., (1994), Statistical principles of research design and analysis, Wadsworth Publishing Company, Belmont, California, pp. 116.

Kung, H.-F., (1984), Method for the extraction of immune interferon, US patent **4,476,049**.

Kuwajima, K. (1989), The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure, *Proteins: Struct., Funct., Genet.*, **6**: 87-103.

Kuwajima, K., Mitani, M., and Sugai, S., (1989), Characterisation of the critical state in protein folding, *J. Mol. Biol.*, **206**: 547-561.

Kuwajima, K., Nitta, K., Yoneyama, M., and Sugai, S., (1976), Three-state denaturation of α -lactalbumin by guanidine hydrochloride, *J. Mol. Biol.*, **106**: 359-373.

Kwon, K.-S., Lee, S., and Yu, M.-H., (1995), Refolding of α_1 -antitrypsin expressed as inclusion bodies in *Escherichia coli*: characterisation of aggregation, *Biochim. Biophys. Acta*, **1247**: 179-184.

- Langley, K. E., Berg, T.F., Strickland, T. W., Fenton, D. M., Boone, T. C., and Wypych, J., (1987), Recombinant-DNA-derived bovine growth hormone from *Escherichia coli*. 1. Demonstration that the hormone is expressed in reduced form, and isolation of the hormone in oxidized, native form, *Eur. J. Biochem.*, **163**: 313-321.
- Lazdunski, A. M., (1989), Peptidases and proteases of *Escherichia coli* and *Salmonella typhimurium*, *FEMS Microbiology Reviews*, **63**: 265-276.
- Lee, Y.-P. and Rhee, J. S., (1996), Protein aggregation and adsorption upon *in vitro* refolding of recombinant *pseudomonas* lipase, *J. Microbiol. Biotechnol.*, **6**: 456-460.
- Levinthal, C., (1968), Are there pathways for protein folding?, *J. Chim. Phys.*, **65**: 44-45.
- Lightfoot, E. N., Gibbs, S. J., Cockrem, M. C. M., and Athalye, A. M., (1987), Scaling up protein purification, in *Protein Purification: Micro to Macro.*, Burgess, R. (ed.), Alan R. Liss, NY, pp. 475-489.
- Lilie, H., Rudolph, R., and Buchner, J., (1996), Association of antibody chains at different stages of folding: prolyl isomerisation occurs after formation of quaternary structure, *J. Mol. Biol.*, **248**: 190-201.
- London, J., Skrzynia, C., and Goldberg, M. E., (1974), Renaturation of *E. coli* tryptophanase after exposure to 8M urea. Evidence for the existence of nucleation centres, *Eur. J. Biochem.*, **47**: 409-415.
- Lugtenberg, B., and van Alphen, L., (1983), Molecular architecture and functioning of the outer membrane of *Escherichia coli* and other gram-negative bacteria, *Biochim. Biophys. Acta*, **737**: 51-115.
- Lugtenberg, B., Peters, R., Bernheimer, H., and Berendsen, W., (1976), Influence of cultural conditions and mutations on the composition of the outer membrane proteins of *Escherichia coli*, *Mol. gen. Genet.*, **147**: 251-262.
- Maachupalli-Reddy, J., Kelley, B., and De Bernardez Clark, E., (1997), Effect of inclusion body-contaminants on the oxidative renaturation of hen-egg white lysozyme, *Biotech. Prog.*, **13**: 144-150.
- Mall, S., West, D. M., Chaudhuri, J. B., and Thatcher, D. R., (1994) Lysozyme refolding in reverse micelles: initial studies, in *Separations for Biotechnology 3*, Pyle, D. L. (ed.), The Royal Society for Chemistry, pp. 287-293.
- Manning, P. A. and Reeves, P., (1977), Outer membrane protein 3b of *Escherichia coli* K-12: effects of growth temperature on the amount of the protein and further characterization on acrylamide Gels, *FEMS Micro. Lett.*, **1**: 275-278.
- Mannweiler, K., (1989), The recovery of biological particles in high-speed continuous centrifuges with special reference to feed-zone break-up effects, *Ph. D. dissertation, University College London*.
- Markussen, J., (1994), Protein purification method, US patent **5,317,092**.

- Marston, F. A. O., (1986), The purification of eukaryotic polypeptides synthesized in *E. coli.*, *Biochem. J.*, **240**: 1-12.
- Marston, F. A. O. and Hartley, D. L., (1990), Solubilization of protein aggregates, *Meth. Enzymol.*, **182**: 264-276.
- Marston, F. A. O., Lowe, P. A., Doel, M. T., Shoemaker, J. M., White, S., and Angal, S., (1984), Purification of calf prochymosin (prorennin) synthesised in *Escherichia coli.*, *Bio/Technology*, **2**: 800-807.
- Marvin, H. J. P., ter Beest, M. B. A., and Witholt, B., (1989), Release of outer membrane fragments from wild-type *Escherichia coli* and for several *E. coli* lipopolysaccharide mutants by EDTA and heat shock treatments, *J. Bacteriol.*, **171**: 5262-5267.
- Mascarenhas, D., Zhang, Y., Olson, P. S., Olsen, D. R., and Carrillo, P. A., (1994), Fusion polypeptides and proteins, US patent **5,563,046**.
- Matthiesen, F., Hejnæs, K. R., and Skriver, L., (1996), Stabilization of recombinantly expressed proteins, *Ann. N. Y. Acad. Sci.*, **782**: 413-421.
- McCoy, J. and LaVallie, E. R., (1994), Peptide and protein fusions to thioredoxin and thioredoxin-like molecules, US patent **5,202,646**.
- Meagher, M. M., Bartlett, R. T., Rai, V. R., and Khan, F. R., (1994), The extraction of rIL-2 inclusion bodies from *E. coli* using cross-flow filtration , *Biotechnol. Bioeng.*, **43**: 969-977.
- Meerman, H. J. and Georgiou, G., (1994), High-level production of proteolytically sensitive secreted proteins in *Escherichia coli* strains impaired in the heat-shock response, *Ann. N. Y. Acad. Sci.*, **721**: 292-302.
- Michaelis, U., Rudolph, R., Jarsch, J., Kopetzki, E., Burtschar, H., Schumacher, G., (1995), Process for the production and renaturation of recombinant, biologically active, eukaryotic alkaline posphatase, US patent **5,434,067**.
- Middelberg, A. P. J., (1996a), Large-scale recovery of recombinant protein inclusion bodies in *Escherichia coli*, *J. Microbiol. Biotechnol.*, **6**: 225-231.
- Middelberg, A. P. J., (1996b), The influence of protein refolding strategy on cost for competing reactions, *Chem. Eng. J.*, **61**: 41-52.
- Middelberg, A. P. J., (1995), Process-scale disruption of microorganisms, *Biotech. Advances*, **13**: 491-551.
- Middelberg, A. P. J., Thomas, J. C., and Collis, M. A., (1993), Photosedimentation: A technique for high-resolution particle analysis, *Chemical Engineering in Australia*, **CHE 18**: 18-21.

- Middelberg, A. P. J., (1992), A model for the disruption of *Escherichia coli* by high-pressure homogenisation, *Ph. D. dissertation, University of Adelaide*.
- Middelberg, A. P. J., O'Neill, B. K., and Bogel, I. D. L., (1992), Modelling bioprocess interactions for optimal design and operating strategies, *Trans. Inst. Chem. Eng. C*, **70**: 8-12.
- Middelberg, A. P. J. and O'Neill, (1991), Monitoring the centrifugal recovery of recombinant protein inclusion bodies, *Aust. J. Biotechnol.*, **5**: 87-92.
- Middelberg, A. P. J., Bogle, I. D. L., and Snoswell, M., (1990), Sizing biological samples by photosedimentation techniques, *Biotechno. Prog.*, **6**: 255-261.
- Miller, J. A., Narhi, L. O., Hua, Q.-X., Resenfeld, R., Arakawa, T., Rohde, M., Prestrelski, S., Lauren, S., Stoney, K. S., Tsai, L., and Weiss, M. A., (1993), Oxidative refolding of Insulin-like Growth Factor-1 yields two products of similar thermodynamic stability: a bifurcating protein-folding pathway, *Biochemistry*, **32**: 5203-5213.
- Milner, S. J., (1996), The oxidative folding of Insulin-like Growth Factor-I analogues, *Ph.D. dissertation, The University of Adelaide*.
- Milner, S. J., Francis, G. L., Wallace, J. C., Magee, B. A., and Ballard, F. J., (1995), Mutations in the B-domain of Insulin-like Growth Factor-I influence the oxidative folding to yield products with modified biological properties., *Biochem. J.*, **308**: 865-871.
- Mitraki, A. and King, J., (1989), Protein folding intermediates and inclusion body formation, *Bio/Technol.*, **7**: 690-697.
- Mitraki, A., Betton, J.-M., Desmadril, M., and Yon, J. M., (1987), Quasi-irreversibility in the unfolding-folding transition of phosphoglycerate kinase induced by guanidine hydrochloride, *Eur. J. Biochem.*, **163**: 29-34.
- Mitraki, A., Haase-Pettingell, C., and King, J., (1991), Mechanisms of inclusion body formation, in *Protein Refolding*, Georgiou, G. and De Bernardez-Clark, E. (eds), *ACS symposium series*, **470**: 35-49.
- Moldow, C., Robertson, J., and Rothfield, L., (1972), Purification of bacterial membrane proteins. The use of guanidinium thiocyanate and urea, *J. Membrane Biol.*, **10**: 137-152.
- Mulkerrin, M. G. and Wetzel, R., (1989), pH dependence of the reversible and irreversible thermal denaturation of γ interferons, *Biochemistry*, **28**: 6556-6561.
- Niedhardt, F. C., (1987), Chemical composition of *Escherichia coli*, in *Escherichia coli and Salmonella typhimurium*, *Cellular and Molecular Biology*, Vol. 1, Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M., and Umberger (eds), American Society for Microbiology, Washington, D. C., pp. 3-6.
- Nikaido, H. and Vaara, M., (1987), Outer Membrane, in *Escherichia coli and Salmonella typhimurium*, *Cellular and Molecular Biology*, Vol. 1, Neidhardt, F. C., Ingraham, J.

- L., Low, K. B., Magasanik, B., Schaechter, M., and Umbarger (eds), American Society for Microbiology, Washington, D. C., pp. 7-22.
- Ohgushi, M. and Wada, A., (1983), 'Molten globule state': a compact form of globular proteins with mobile side chains *FEBS Lett.*, **164**: 21-24.
- Olbrich, R., (1989), The characterisation and recovery of protein inclusion bodies from recombinant *Escherichia coli*, *Ph. D. dissertation, University of London*.
- Oliver, D. B., (1987), Periplasm and protein secretion, in *Escherichia coli and Salmonella typhimurium, Cellular and Molecular Biology*, Vol. 1, Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M., and Umbarger (eds), American Society for Microbiology, Washington, D. C., pp. 56-69.
- Olson, K. C., (1985), Purification and activity assurance of precipitated heterologous proteins, US patent **4,518,526**.
- Osborn, M. J. and Wu, H. C. P., (1980), Proteins of the outer membrane of gram-negative bacteria, *Annu. Rev. Microbiol.*, **34**: 369-422.
- Osborn, M. J., Gander, J. E., Parisi, E., and Carson, J., (1972), Mechanism of assembly of the outer membrane of *Salmonella typhimurium*. Isolation and characterization of cytoplasmic and outer membrane, *J. Biol. Chem.*, **247**: 3962-3972.
- Owers-Narhi, L., Hua, Q.-X., Arakawa, T., Fox, G. M., Tsai, L., Rosenfeld, R., Holst, P., Miller, J. A., and Weiss, M. A., (1993), Role of native disulfide bonds in the structure and activity of Insulin-like Growth Factor 1: genetic models of protein-folding intermediates, *Biochemistry*, **32**: 5214-5221.
- Pace, C. N., Shirley, B. A., McNutt, M., Gajiwala, K., (1996), Forces contributing to the conformational stability of proteins, *FASEB*, **10**: 75-83.
- Pecorari, F., Minard, P., Desmadril, M., and Yon, J. M., (1996), Occurrence of transient multimeric species during the refolding of a monomeric protein, *J. Biol. Chem.*, **271**: 5270-5276.
- Perry, R. H. and Green, D. W., (1984), Fixed-capital-cost estimation, in *Perry's Chemical Engineers' Handbook*, 6th Edition, McGraw-Hill, Japan, pp. 25-64 - 25-75.
- Peters, M. S. and Timmerhaus, K. D., (1990), *Plant design and economics for chemical engineers*, 4th Edition., McGraw-Hill, New York.
- Petrides, D., (1990), Computer-aided design of integrated biochemical processes; Development of BioDesigner, *Ph. D. dissertation, University of Massachusetts*, pp. 115-150.
- Petrides, D., Cooney, C. L., Evans, L. B., Field, R. P., and Snoswell, M., (1989), Bioprocess simulation: an integrated approach to process development, *Computers chem. Engng.*, **13**: 553-561.

- Petrides, D., Sapidou, E., and Calandranis, J., (1995), Computer-aided process analysis and economic evaluation for biosynthetic human insulin production - a case study, *Biotech. Bioeng.*, **48**: 529-541.
- Ptitsyn, O. B., (1995), Molten globule and protein refolding, *Advances in Protein Chemistry*, **47**: 83-229.
- Ptitsyn, O. B., (1992), The molten globule state, in *Protein Folding*, Creighton, T. E. (ed.), W. H. Freeman and Company, NY, pp. 243-300.
- Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E., and Razgulyaev, O. L., (1990), Evidence for a molten globule state as a general intermediate in protein folding, *FEBS Letts.*, **262**: 20-24.
- Raetz, C. R., (1987), Structure and Biosynthesis of Lipid A in *Escherichia coli*, in *Escherichia coli and Salmonella typhimurium, Cellular and Molecular Biology*, Vol. 1, Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M., and Umberger (eds), American Society for Microbiology, Washington, D. C., pp. 498-503.
- Rase, H. F., (1977), Chemical Reactor Design for Process Plants, Vol. 1, John Wiley & Sons, New York.
- Rhee, H. J., Lee, Y. I., and Yang, K. H., (1990), High-level expression of human Insulin-like Growth Factor II in *Escherichia coli*, *J. Biotechnol.*, **13**: 293-304.
- Rudolph, R., (1990), in *Modern Methods on Protein and Nucleic Acid Analysis*, Tsechesche, H. (ed.), Walter de Gruyter, Berlin, pp. 149-171.
- Rudolph, R. and Fischer, S., (1990), Process for obtaining renatured proteins, US patent **4,933,434**.
- Rudolph, R. and Lilie, H., (1996), *In vitro* folding of inclusion body proteins, *FASEB J.*, **10**: 49-56.
- Rudolph, R., Buchner, J., and Lenz, H., (1991), Process for activation of recombinant proteins produced by prokaryotes, US patent **5,077,392**.
- Rudolph, R., Fischer, S., and Mattes, R., (1995), Process for the activation of t-PA or Ing after genetic expression in prokaryotes, US patent **5,453,363**.
- Samuelsson, E. and Ulhén, M., (1996), Chaperone-like effect during *in vitro* refolding of Insulin-Like Growth Factor-I using a solubilizing fusion partner, *Ann. N. Y. Acad. Sci.*, **782**: 486-494.
- Samuelsson, E., Jonasson, P., Viklund, F., Nilsson, B., and Ulhén, M., (1996), Affinity-assisted *in vivo* folding of a secreted human peptide hormone in *Escherichia coli*, *Nature Biotechnology*, **14**: 751-755.

- Samuelsson, E., Wadensten, H., Hartmanis, M., Moks, T., and Ulhén, M., (1991), Facilitated *in vitro* refolding of human recombinant Insulin-like Growth Factor I using a solubilizing fusion partner, *Bio/Technology*, **9**: 363-366.
- Sargent, D., Benevides, J. M., Yu, M.-H., King, J., Thomas Jr, G. J., (1988), Secondary Structure and Thermostability of the Phage P22 Tailspike, *J. Mol. Biol.*, **199**: 491-502.
- Sarmientos, P., Duchesne, M., Denefle, P., Boiziau, J., Fromage, N., Delporte, N., Parker, F., Lelievre, Y., Mayaux, J.-F., and Cartwright, T., (1989), Synthesis and purification of active human tissue plasminogen activator from *Escherichia coli*, *Bio/Technology*, **7**: 495-501.
- Schaechter, M. and Neidhardt, F. C., (1987), Introduction, in *Escherichia coli and Salmonella typhimurium, Cellular and Molecular Biology*, Vol. 1, Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M., and Umberger (eds), American Society for Microbiology, Washington, D. C., pp. 1-2.
- Schein, C. H., (1990), Solubility as a function of protein structure and solvent components, *Bio/Technology*, **8**: 308-317.
- Schiffer, C. A and Dötsch, V., (1996), The role of protein-solvent interactions in protein unfolding, *Curr. Opin. Biotech.*, **7**: 428-432.
- Schmid, F. X., (1986), Fast-folding and slow-folding forms of unfolded proteins, *Method Enzymol.*, **131**: 70- 82.
- Schmid, F. X., (1992), Kinetics of unfolding and refolding of single-domain proteins, in *Protein Refolding*, Creighton, T. E. (ed.) W. H. Freeman and Company, NY, pp. 197-241.
- Schmid, F. X., Mayr, L. M., Mücke, M., Schönbrunner, E. R., (1993), Prolyl isomerases: role in protein folding, *Adv. Prot. Chem.*, **44**: 25-66.
- Schnaitman, C. A., (1974), Outer membrane proteins of *Escherichia coli*. IV. Differences in outer membrane proteins due to strain and cultural differences, *J. Bacteriol.*, **118**: 454-464.
- Schnaitman, C. A., (1970), Examination of the protein composition of the cell envelope of *Escherichia coli* by polyacrylamide gel electrophoresis, *J. Bacteriol.*, **104**: 882-889.
- Schoemaker, J. M., Brasnett, A. H., and Marston F. A. O., (1985), Examination of calf prochymosin accumulation in *Escherichia coli*: disulfide linkages are a structural component of prochymosin-containing inclusion bodies, *The EMBO journal*, **4**: 775-780.
- Seely, R. J. and Young, M. D., (1991). Large-scale refolding of secretory leukocyte protease inhibitor, in *Protein Refolding*, Georgiou, G. and De Bernadez-Clark, E. (eds), *ACS symposium series*, **470**: 180-186.

- Semisotnov, G. V., Rodinova, N. A., Kutyschenko, V. P., Ebert, B., Blanck, J., and Pititsyn, O. B., (1987), Sequential mechanism of refolding of carbonic anhydrase B, *FEBS lett.*, **224**: 9-13.
- Sharma, A. K., Minke-Gogl, V., Gohl, P., Siebendritt, R., Jaenicke, R., and Rudolph, R., (1990), Limited proteolysis of γ II-crytallin from calf eye lens. Physiochemical studies on the N-terminal domain and the intact two-domain protein, *Eur. J. Biochem.*, **194**: 603-609.
- Shire, S. J., Bock, L., Ogez, J., Builder, S., Kleid, D., Moore, D. M., (1984), Purification and immunogenicity of fusion VP1 protein of foot and mouth disease virus, *Biochemistry*, **23**: 6474-6480.
- Shortle, D., (1996), The denatured state (the other half of the folding equation) and its role in protein stability, *FASEB*, **28**: 27-34.
- Siddiqi, S. F., Titchner-Hooker, N. J., and Ayazi Shamlou, P., (1996), Simulation of Particle Size Distribution Changes Occurring During High-Pressure Disruption of Bakers' Yeast, *Biotechnol. Bioengng*, **50**: 145-150.
- Skriver, L., Hejnæs., K. R., and Dalboge, H., (1995), Method of producing IGF-1,US patent **5,459,052**.
- Speed, M. A., Wang, D. I. C., and King, J., (1996), Specific aggregation of partially folded polypeptide chains: The molecular basis for inclusion body composition, *Nature Biotechnology*, **14**: 1283-1287.
- Speed, M. A., King, J., and Wang, D. I. C., (1997), Polymerization mechanism of polypeptide chain aggregation, *Biotech. Bioeng*, **54**: 333-343.
- Stempfer, G., Höll-Neugebauer, B., and Rudolph, R., (1996), Improved refolding of an immobilized fusion protein, *Nature Biotechnology*, **14**: 329-334.
- Stewart, D. E., Sarkar, A., and Wampler, J. E, (1990), Occurrence and role of *cis* peptide bonds and protein structures, *J. Mol. Biol.*, **214**: 253-260.
- Storrs, S. B., and Przybycien, T. M., (1991), Commercial-scale refolding of recombinant methionyl bovine somatotropin, in *Protein Refolding*, Georgiou, G. and De Bernadez-Clark, E. (eds), *ACS symposium series*, **470**: 197-205.
- Sugimura, K. and Higashi, W., (1988), A novel outer-membrane-associated protease in *E. coli*, *J. Bacteriol.*, **170**: 3650-3654.
- Sugimura, K. and Nishihara, T., (1988), Purification, characterisation, and primary structure of *Escherichia coli* protease VII with specificity for paired basic residues: identity of protease VII and ompT, *J. Bacteriol.*, **170**: 5625-5632.
- Sussenbach, J. S., (1989), The gene structure of the Insulin-Like Growth Factor family, *Progress in Growth Factor Research*, **1**: 33-48.

- Swamy, K. H. S. and Goldberg, A. L., (1982), Subcellular distribution of various proteases in *Escherichia coli*, *J. Bacteriol.*, **149**: 1027-1033.
- Swartz, J., Builder, S., Hart, R., Reifsnyder, D., and Ogez, J., (1996), Evolution of a process for IGF-1, in *Proceedings of the 10th International Biotechnology Symposium, 25-30 August 1996, Sydney, Australia*, S5-3.
- Taylor, G., Hoare, M., Gray, D. R., and Marston, F. A. O., (1986), Size and density of protein inclusion bodies, *Bio/Technology*, **4**: 553-557.
- Teipel, J. W. and Koshland, Jr, D. E., (1971a), Kinetic aspects of conformational changes in proteins. I. Rate of regain of enzyme activity from denatured proteins, *Biochemistry*, **10**: 792-798.
- Teipel, J. W. and Koshland, Jr, D. E., (1971b), Kinetic aspects of conformational changes in proteins. II. Structural changes in renaturation of denatured proteins, *Biochemistry*, **10**: 798-805.
- Terashima, M., Suzuki, K., and Katoh, S., (1996), Effective refolding of fully reduced lysozyme with a flow-type reactor, *Proc. Biochem.*, **31**: 341-345.
- Tokatlidis, K., Dhurjati, P., Mittet, J., Beguin, P., and Aubert, J.-P, (1991), High activity of inclusion bodies formed in *E. coli* overproducing *colstridium thermocellum* endoglucanase D., *FEBS letters*, **282**: 205-208.
- Turner, C., Gregory, M. E., and Thornhill, N. F., (1994), Closed-loop control of fed-batch cultures of recombinant *Escherichia coli* using on-line HPLC, *Biotechnol. Bioeng.*, **44**: 819-829.
- Utiyama, H. and Baldwin, R. L., (1986), Kinetic mechanisms of protein folding, *Methods in Enzymology*, **131**: 51-70.
- Valax, P. and Georgiou, G., (1991), Folding and aggregation of RTEM β -lactamase, in *Protein Refolding*, Georgiou, G. and De Bernardez-Clark, E. eds, *ACS symposium series*, **470**: 97-109.
- Valax, P. and Georgiou, G., (1993a), Molecular characterisation of β -lactamase inclusion bodies produced in *E. coli*. 1 Composition., *Biotechnol. progr.*, **9**: 539-547.
- Valax, P. and Georgiou, G., (1993b), Characterisation and refolding β -lactamase inclusion bodies in *Escherichia coli*, in *Biocatalyst design for stability and specificity*, Himenl, M. E. and Georgiou, G. (eds), *ACS symposium series*, **516**: 126-139.
- Vicik, S. and De Bernadez-Clark, E., (1991), An engineering approach to achieving high-protein refolding yields, in *Protein Refolding*, Georgiou, G. and De Bernadez-Clark, E. (eds), *ACS symposium series*, **470**: 180-186.
- Vise, P. D. and Wells, J. R. E., (1987), Spacer alterations which increase the expression of porcine growth hormone in *E. coli*, *FEBS Letters*, **213**: 155-158.

- Walker, S., Huddleston, J., Viede, A., and Lyddiatt, A., (1994), Preliminary studies of the partitioning in PEG-phosphate aqueous two phase systems of defined fusion proteins expressed intracellularly in *E. coli*, *The 1994 IChemE Research Event*, 156-158.
- Weir, M. P., Sparks, J., (1987), Purification and renaturation of recombinant human interleukin-2., *Biochem. J.*, **245**: 85-91.
- Weissman, J. S. and Kim, P. S., (1991), Reexamination of the folding of BPTI: predominance of native intermediates, *Science*, **253**: 1386-1393.
- Wells, J. R. E., King, R. M., and Francis, G. L., (1994), Growth hormone fusion proteins, methods of production, and methods of treatment, US patent **5,330,971**.
- West, S. M., Guise, A. D., and Chaudhuri, J. B., (1997), A comparison of the denaturants urea and guanidine hydrochloride on protein refolding, *Trans. IChemE. C.*, **75**: 50-56.
- Wetzel, R. B., (1986), Purification and activity assurance of precipitated heterologous proteins, US patent **4,599,197**.
- Winther, J. R., Sørensen, P., and Kielland-Brandt M. C., (1994), Refolding of a carboxypeptidase Y folding intermediate *in vitro* by low-affinity binding of the proregion, *J. Biol. Chem.*, **269**: 22007-22013.
- Wolf-Watz, H., Normark, S., and Bloom, G., (1973), Rapid method for isolation of large quantities of outer membrane from *Escherichia coli* K-12 and its application to the study of envelope mutants, *J. Bacteriol.*, **115**: 1191-1197.
- Wong, E. Y., Seetharam, R., Kotts, C. E., Heeren, R. A., Klein, B. K., Braford, S. R., Mathis, K. J., Bishop, B. F., Siegel, N. R., Smith, C. E., and Tacon, W. C., (1988), Expression of secreted Insulin-like Growth Factor-1 in *Escherichia coli*, *Gene*, **68**: 193-203.
- Wong, H. H., (1997), Modelling studies of the interaction between homogenisation, centrifugation, and inclusion body dissolution, *Ph. D. dissertation, University of Adelaide*.
- Wong, H. H., O'Neill, B. K., and Middelberg, A. P. J., (1996), Centrifugal recovery and dissolution of recombinant Gly-IGF-II inclusion-bodies: The impact of feedrate and re-centrifugation on protein yield, *Bioseparation*, **6**: 185-192.
- Yamada, H., Saito, Y., Fujimoto, T., Noguchi, Y., Mori, T., Miura, T., Kobayashi, M., and Shimomura, K., (1996), Large scale purification of recombinant Insulin-like Growth Factor I (IGF-I, mercasermin) from a fused protein produced in *Escherichia coli*, *Journal of Fermentation and Bioengineering*, **82**: 134-139.
- Yoon, S. K., Kang, W. K., and Park, T. H., (1994a), Fed-batch operation of recombinant *Escherichia coli* containing *trp* promoter with controlled specific growth rate, *Biotechnol. Bioeng.*, **43**: 995-999.

- Yoon, S. K., Kwon, S. H., Park, M. G, Kang, W. K., and Park, T. H., (1994b), Optimization of recombinant *Escherichia coli* fed-batch fermentation for bovine somatotropin., *Biotechnol. Lett.*, **16**: 1119-1124.
- Yoshida, M., Taguchi, H., and Kunishi, J., (1996), Methods and compositions for promoting protein folding, US patent **5,561,221**.
- Zettlmeißl, G., Rudolph, R., and Jaenicke, R., (1979), Reconstitution of lactic dehydrogenase. Noncovalent aggregation vs. reactivation. 1. Physical properties and kinetics of aggregation, *Biochemistry*, **18**: 5567-5571.
- Zwanzig, R., (1997), Two-state models of protein folding kinetics., *Proc. Natl. Acad. Sci. USA*, **94**: 148-150.

9. APPENDICES

9.1 NOMENCLATURE.....	238
9.2 DERIVATION OF REACTOR DESIGN EQUATIONS	241
9.2.1 Batch reactor	241
9.2.2 Fed-batch reactor	243
9.2.3 Continuous stirred tank reactor (CSTR)	245
9.3 MASS-TRANSFER LIMITED DISSOLUTION.....	248

9.1 NOMENCLATURE

- A = ultrafiltration unit membrane, m^2
 $\hat{A}_{df}$ = specific surface of depth filter, m^2 filter $(m^3 \text{ broth})^{-1}$
 C = concentration, mol m^{-3}
 Da_N = Damköhler number $(k_N \cdot t)$, $(-)$
 Da_{AN} = dimensionless refolding concentration $(K_A C_{D0} / k_N)$, $(-)$
 Da_{BN} = equilibrium concentration (k_B / k_N) , $(-)$
 E = product expression level, $\text{kg product (kg protein)}^{-1}$
 f = fractional mass recovery of product (excluding refolding), $(-)$
 J = flux through ultrafiltration membrane, $m^3 m^{-2} s^{-1}$
 k_B = first-order rate constant of reverse reaction, s^{-1}
 k_c = surface mass-transfer coefficient, $m^3 m^{-2} s^{-1}$
 K_c = apparent mass-transfer coefficient, s^{-1}
 k_N = first-order rate constant of correct refolding, s^{-1}
 K_n = equilibrium constant (k_B / k_N) , $(-)$
 K_A = apparent second-order rate constant of aggregation, $m^3 \text{ mol}^{-1} s^{-1}$
 L = life of ultrafiltration membrane, batches
 M = molecular weight, kg mol^{-1}
 m = mass, kg
 $\dot{m}$ = mass production rate, kg s^{-1}
 n = molar concentration, mol m^{-3}
 N = molar concentration, mol m^{-3}
 N_b = number of batches per annum, $(-)$
 N_p = number of discrete homogeniser passes, $(-)$
 n_t = number of storage tanks, $(-)$
 Q = volumetric flow rate, $m^3 s^{-1}$
 Q_Σ = normalised centrifuge feed rate (Q / Σ) , $m s^{-1}$
 r_j = rate of reaction of species j , $\text{mol m}^{-3} s^{-1}$ or $\text{kg m}^{-3} s^{-1}$
 t = fermentation or batch cycle time, s
 T = total cell protein to dry cell weight ratio, $\text{kg protein (kg cells)}^{-1}$
 V = volume, m^3
 W = dry cell weight concentration, kg cells m^{-3}
 X = conversion of denatured protein, $(-)$
 y_N = yield of correctly refolded protein (C_N / C_{D0}) , $(-)$

Greek letters

- α = cost-capacity constant, (–)
- β = cost-capacity exponent, (–)
- γ = specific cost per kg, m^2 or m^3 , 1994 US\$ (unit)^{–1}
- ε = design margin to provide sufficient ullage (V_b / V_f), (–)
- ζ = annual DFC-dependent cost to equipment-purchase cost ratio, (–)
- v = volumetric flow rate in reactor, $\text{m}^3 \text{s}^{-1}$
- ρ_D = mass concentration of crude product in dissolution, kg m^{-3}
- Σ = equivalent centrifuge clarification area, m^2
- τ = dimensionless time, (–)
- ϕ = cost, 1994 US\$
- $\hat{\phi}$ = cost per batch, 1994 US\$
- ϕ^* = dimensionless cost (ϕ / ϕ_{ch}), (–)

Subscripts

- a = air compressor
- ad = additives to the dissolution mixture
- ar = additives to the refolding mixture
- b = batch
- c = centrifuge
- con = consumables
- d = dissolution
- df = depth filter
- dfc = direct fixed capital
- D = denatured protein
- Df = bulk fluid concentration during dissolution
- Ds = at surface boundary
- Dsat = saturated concentration of dissolved material
- ep = equipment purchase cost
- f = fermenter
- h = homogeniser
- m = fermentation media
- N = native (correctly folded protein)
- o = initial/inlet conditions to refolding reactor
- p = crude protein product in the fermenter
- r = refolding

Subscripts (continued)

s = steriliser

t = storage tank

T = total

u = ultrafiltration

ud = urea for dissolution

up = urea purchase

ur = urea for refolding

urec = urea recovered

wt = waste treatment

wtd = waste treatment disposal of urea

9.2 DERIVATION OF REACTOR DESIGN EQUATIONS

The overall derivation of the dimensionless reactor design equations used for process simulation is presented in section 3.2. This appendix presents the working omitted from the overall derivations.

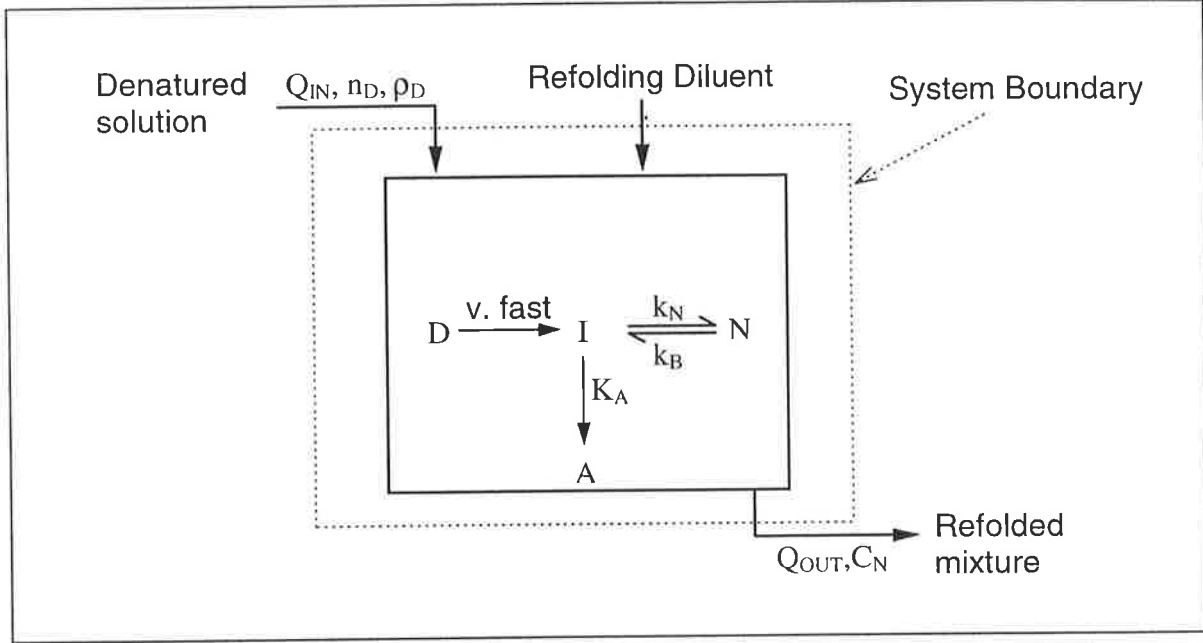


Figure 9.1 : General refolding reactor with addition of denatured material, diluent solution and product removal.

9.2.1 Batch reactor

For an ideally mixed batch reactor, the reaction rate is constant through the vessel and the general design equation (Equation 3.2) reduces to:

$$\frac{d}{dt} C_I = -k_N C_I - K_A C_I^2 + k_B C_N \quad (3.4)$$

$$\frac{d}{dt} C_N = k_N C_I - k_B C_N \quad (3.5)$$

A suitable dimensionless batch refolding time (τ) is the product of the first-order rate constant of native protein formation and reaction time. That is

$$\tau = k_N \cdot t = Da_N \quad (3.8)$$

Substitution of Equation 3.8 into design Equations 3.4 and 3.5 gives:

$$\frac{d}{d\tau} C_I = -\frac{k_N}{k_N} C_I - \frac{K_A}{k_N} C_I^2 + \frac{k_B}{k_N} C_N \quad (9.1)$$

$$\frac{d}{d\tau} C_N = \frac{k_N}{k_N} C_I - \frac{k_B}{k_N} C_N \quad (9.2)$$

Dividing Equations 9.1 and 9.2 by C_{D0} gives:

$$\frac{d}{d\tau} \frac{C_I}{C_{D0}} = -\frac{k_N}{k_N C_{D0}} C_I - \frac{K_A}{k_N C_{D0}} C_I^2 + \frac{k_B}{k_N C_{D0}} C_N \quad (9.3)$$

$$\frac{d}{d\tau} \frac{C_N}{C_{D0}} = \frac{k_N}{k_N C_{D0}} C_I - \frac{k_B}{k_N C_{D0}} C_N \quad (9.4)$$

The conversion of protein, X , can be defined relative to the concentration of denatured protein initially within the reactor, C_{D0} , as

$$X = \frac{C_{D0} - C_I}{C_{D0}} \quad (3.6)$$

The yield of native conformation protein can be defined:

$$y_N = \frac{C_N}{C_{D0}} \quad (3.7)$$

Equation 3.6 can be written:

$$X = \frac{C_{D0} - C_I}{C_{D0}} = 1 - \frac{C_I}{C_{D0}} \quad (9.5)$$

$$\therefore C_I = C_{D0}(1 - X) \quad (9.6)$$

Substitution of equation 9.6 into Equations 9.3 and 9.4 for C_I gives:

$$\frac{d}{d\tau} (1 - X) = -(1 - X) - \frac{K_A C_{D0}}{k_N} (1 - X)^2 + K_n y_N$$

$$\therefore \frac{d}{d\tau} X = (1 - X) + \frac{K_A C_{D0}}{k_N} (1 - X)^2 - K_n y_N \quad (9.7)$$

$$\frac{d}{d\tau} y_N = 1 - X - K_n y_N \quad (9.8)$$

where $K_n = k_B/k_N$ is an equilibrium constant between intermediate and native conformation protein. Equations 9.7 and 9.8 define a pair of coupled, ordinary differential equations with maximum order of two. They must be solved numerically to determine a solution.

9.2.2 Fed-batch reactor

The dimensional design equations for an ideally mixed fed-batch refolding reactor are written in terms of amounts as:

$$\frac{d}{dt} N_I = Q_{IN} \cdot n_D - k_N N_I + k_B N_N - K_A \frac{N_I^2}{V} \quad (3.12)$$

$$\frac{d}{dt} N_N = k_N N_I - k_B N_N \quad (3.13)$$

Assuming that the dissolved protein must be diluted four-fold with refolding buffer give a suitable chemical environment for refolding, a third design equation is:

$$\frac{d}{dt} V = 4Q_{IN} \quad (3.14)$$

Unlike the batch case, the amount of denatured protein present in the reaction vessel will vary with time. Conversion of intermediate species can be defined in terms of the total amount of denatured protein fed into the reactor, N_{Do} , as:

$$X = \frac{\tau}{\tau_{\text{Refolding}}} - \frac{N_I}{N_{Do}} \quad (3.15)$$

$$\therefore \frac{N_I}{N_{Do}} = \frac{\tau}{\tau_{\text{Refolding}}} - X \quad (9.9)$$

where τ is the dimensionless time of refolding and $\tau_{\text{Refolding}}$ is the dimensionless duration of feeding. For fed-batch operation, the yield much be defined as the ratio of amounts of native species to total initial denatured species

$$y_N = \frac{N_N}{N_{Do}} \quad (3.16)$$

To allow direct comparison of fed-batch and batch refolding the initial charge to the refolding reactor should be such that at the end of refolding, the nominal protein concentration is the same for both processes (C_{D0}). That is, design the final volume of the fed-batch refold to be the required production capacity divided by the nominal initial concentration of denatured protein. To achieve this, the initial charge will be the final volume less the volume that is fed during refolding

$$V_o = \frac{N_{D0}M_N}{C_{D0}} - \frac{4N_{D0}}{n_D} \quad (3.17)$$

The flow rate of denatured protein into the refolding reactor, Q_{IN} , must feed the total amount of denatured protein within the duration of feeding. Therefore, an expression for Q_{IN} can be written:

$$Q_{IN} = \frac{N_{D0}k_N}{n_D\tau_{\text{Refolding}}} \quad (9.10)$$

Dividing Equation 3.12 by N_{D0} and substitution of Equations 3.8 and 9.10 for t and Q_{IN} , gives:

$$\begin{aligned} \frac{d}{d\tau} \frac{V}{V_o} &= 4 \frac{N_{D0}k_N}{n_D\tau_{\text{Refolding}}k_N} \cdot \left(\frac{1}{\frac{N_{D0}M_N}{C_{D0}} - \frac{4N_{D0}}{n_D}} \right) \\ \therefore \frac{d}{d\tau} \frac{V}{V_o} &= \frac{4}{\tau_{\text{Refolding}} \left(\frac{M_N n_D}{C_{D0}} - 4 \right)} \end{aligned} \quad (9.11)$$

Substitution of Equation 3.8 into Equations 3.12 and 3.13, and dividing through by N_{D0} gives:

$$\frac{d}{d\tau} \frac{N_I}{N_{D0}} = \frac{Q_{IN} \cdot n_D}{N_{D0} \cdot k_N} - \frac{N_I}{N_{D0}} + \frac{k_B \cdot N_N}{k_N \cdot N_{D0}} - \frac{K_A \cdot N_{D0}}{k_N \cdot V} \left(\frac{N_I}{N_{D0}} \right)^2 \quad (9.12)$$

$$\frac{d}{d\tau} \frac{N_N}{N_{D0}} = \frac{N_I}{N_{D0}} - \frac{k_B \cdot N_N}{k_N \cdot N_{D0}} \quad (9.13)$$

Substitution of Equations 9.9, 3.17, and 9.10 into equation 9.12 gives:

$$\frac{d}{d\tau} \left(\frac{\tau}{\tau_{\text{Refolding}}} - X \right) = \frac{1}{\tau_{\text{Refolding}}} - \frac{\tau}{\tau_{\text{Refolding}}} + X + K_n y_N - \frac{K_A \cdot N_{D0}}{k_N \cdot V} \left(\frac{\tau}{\tau_{\text{Refolding}}} - X \right)^2 \quad (9.14)$$

$$\therefore \frac{d}{d\tau} (X) = \left(\frac{\tau}{\tau_{\text{Refolding}}} - X \right) - K_n y_N + \frac{K_A \cdot N_{D0}}{k_N \cdot V} \left(\frac{\tau}{\tau_{\text{Refolding}}} - X \right)^2 \quad (9.15)$$

Introducing V/V_0 to make the equations dimensionless

$$\therefore \frac{d}{d\tau} (X) = \frac{\tau}{\tau_{\text{Refolding}}} - X - K_n y_N + \frac{K_A C_{D0}}{k_N \cdot \left(1 - \frac{4C_{D0} M_N}{n_D} \right) \left(\frac{V}{V_0} \right)} \left(\frac{\tau}{\tau_{\text{Refolding}}} - X \right)^2 \quad (3.18)$$

Substitution of Equations 9.9 and 3.15 into Equation 9.13 gives:

$$\frac{d}{d\tau} y_N = \frac{\tau}{\tau_{\text{Refolding}}} - X - K_n y_N \quad (3.19)$$

The initial conditions for Equations 3.18, 3.19, and 3.20 are that conversion and yield are both zero and the initial volume ratio is one. This set of coupled dimensionless second-order ordinary differential must be solved numerically to obtain a solution.

9.2.3 Continuous stirred tank reactor (CSTR).

Solution of design equations for continuous refolding are at steady-state and do not require integration. If it is assumed that the reaction rate is independent of the position in the reactor (ideally mixed), the design equations are

$$C_I = C_{D0} - (k_N C_I + K_A C_I^2 - k_B C_N) \cdot t \quad (3.21)$$

$$C_N = (k_N C_I - k_B C_N) \cdot t \quad (3.22)$$

As for batch reactors a suitable dimensionless refolding time (τ) is the product of the first-order rate constant of native protein formation and reaction time. In a CSTR, this dimensionless group is a Damköhler number. That is

$$\tau = k_N \cdot t = Da_N \quad (3.8)$$

Ratios of Damköhler numbers for other rate constants can be defined:

$$Da_{BN} = \frac{Da_B}{Da_N} = \frac{k_B t}{k_N t} \quad (9.16)$$

$$Da_{AN} = \frac{Da_A}{Da_N} = \frac{K_A C_{D0} t}{k_N t} \quad (9.17)$$

$$X = 1 - \frac{C_{D0} - (k_N C_I + K_A C_I^2 - k_B C_N) \cdot t}{C_{D0}} \quad (\text{substituting 3.21 into 9.5})$$

$$\therefore X = (k_N C_I + K_A C_I^2 - k_B C_N) \cdot \frac{t}{C_{D0}} \quad (9.18)$$

$$\therefore X = (k_N (C_{D0} (1 - X)) + K_A (C_{D0} (1 - X))^2 - k_B C_N) \cdot \frac{t}{C_{D0}} \quad (\text{substituting 9.6 into 9.18})$$

$$\therefore X = k_N \cdot t \left[(1 - X) + \frac{K_A C_{D0}}{k_N} (1 - X)^2 - \frac{k_B C_N}{k_N C_{D0}} \right]$$

$$\therefore X = Da_N \left[(1 - X) + Da_{AN} (1 - X)^2 - Da_{BN} y_N \right] \quad (9.19)$$

Substituting Equation 3.22 into Equation 3.7 gives:

$$y_N = \frac{(k_N C_I - k_B C_N) \cdot t}{C_{D0}}$$

$$\therefore y_N = \frac{(k_N C_{D0} (1 - X) - k_B C_N) \cdot t}{C_{D0}}$$

$$\therefore y_N = k_N t (1 - X) - \frac{k_B}{k_N} \frac{C_N}{C_{D0}} \cdot k_N t$$

$$\therefore y_N = \frac{1 - X}{\frac{1}{Da_N} + Da_{BN}} \quad (3.24)$$

Substitution of Equation 3.24 into Equation 9.19 gives:

$$\begin{aligned}
 X &= Da_N \left[(1-X) + Da_{AN}(1-X)^2 - Da_{BN} \cdot \left(\frac{1-X}{\frac{1}{Da_N} + Da_{BN}} \right) \right] \\
 \therefore X \cdot \left(\frac{1}{Da_N} + Da_{BN} \right) &= Da_N \left[\frac{1}{Da_N} + Da_{BN} - X \cdot \left(\frac{1}{Da_N} + Da_{BN} \right) + \right. \\
 &\quad \left. Da_{AN}(1-X)^2 \left(\frac{1}{Da_N} + Da_{BN} \right) - Da_{BN}(1-X) \right] \\
 \therefore \frac{X}{Da_N} + XD_{aN}Da_{BN} &= \left(\frac{1 + Da_N Da_{BN} - X \cdot (1 + Da_N Da_{BN})}{Da_{AN}(1-X)^2 (1 + Da_N Da_{BN}) - Da_N Da_{BN}(1-X)} \right) \\
 \therefore 0 &= \left(\frac{Da_N^2 \cdot (Da_{BN} - XD_{aBN} + Da_{BN}Da_{AN}(1-X)^2 - Da_{BN}(1-X)) +}{Da_N \cdot (1-X + Da_{AN}(1-X)^2 - Da_{BN}X) - X} \right) \\
 0 &= Da_N^2 (Da_{BN}Da_{AN}(1-X)^2) + Da_N (1-X + Da_{AN}(1-X)^2 - Da_{BN}X) - X \quad (3.25)
 \end{aligned}$$

Equation 3.25 defines a quadratic expression for Da_N (τ) that may be solved for any given set of Da_{BN} , Da_{AN} , and X . The solution can then be substituted into Equation 3.24 to determine the yield for that set of parameters.

9.3 MASS-TRANSFER LIMITED DISSOLUTION

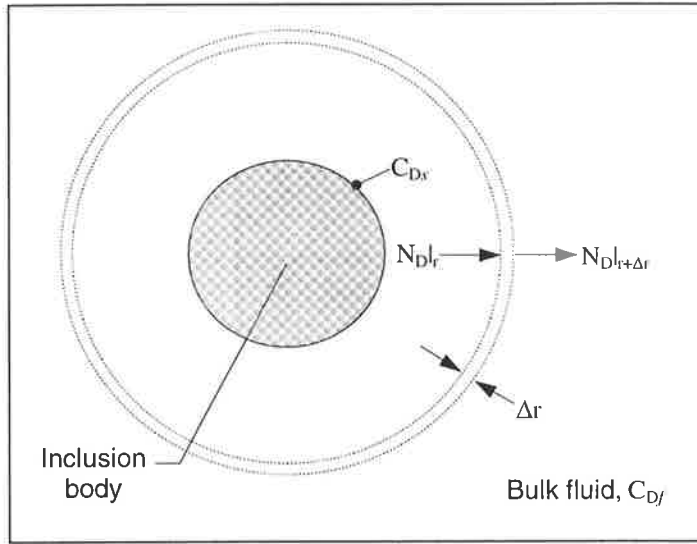


Figure 9.2 : Illustration of the shell balance around a spherical particle (inclusion body) with loss of material (D) to the bulk fluid. Where C_{Ds} is the surface concentration, C_{Df} is the bulk fluid concentration, and $N_{D|r}$ is the mass flux at radius, r , of species D.

The general shell mass balance can be written:

$$\left\{ \begin{array}{l} \text{rate of} \\ \text{mass of} \\ \text{D in} \end{array} \right\} - \left\{ \begin{array}{l} \text{rate of} \\ \text{mass of} \\ \text{D out} \end{array} \right\} + \left\{ \begin{array}{l} \text{rate of production} \\ \text{of mass of D by} \\ \text{chemical reaction} \end{array} \right\} = 0 \quad (9.20)$$

For loss of solid at a surface into a surrounding bulk fluid, the mass flux at the surface, m_{Ds} , can be written (Bird *et al.*, 1960):

$$m_{Ds} = k_c (C_{Ds} - C_{Df}) \quad (9.21)$$

Therefore, the rate of mass entering at the surface will be the product of the mass flux and the surface area, A . If it is assumed that the concentration of species D at the surface of the particle is at the saturation concentration, C_{Dsat} , the rate of mass entering the bulk fluid can be written:

$$\frac{d}{dt} \text{Mass}_D = k_c A (C_{Dsat} - C_{Df}) \quad (9.22)$$

where Mass_D is the mass of species D in the fluid over the area, A. If it is assumed that the volume of solution remains constant and is well mixed, Equation 9.22 can be written:

$$\frac{d}{dt}C_D = K_c(C_{Dsat} - C_{Df}) \quad (9.23)$$

where $K_c = k_c A / V$ is an apparent mass-transfer coefficient with the units of time^{-1} .

Integrating the shell balance equation over the volume of the fluid with time gives:

$$\int_{C_{Df}(0)}^{C_{Df}(t)} \left(\frac{1}{C_{Dsat} - C_{Df}} \right) dC_{Df} = K_c t \quad (9.24)$$

when $C_{Df}(0) = 0$, Equation 9.24 becomes:

$$-\ln(C_{Dsat} - C_{Df}(t)) + \ln(C_{Dsat}) = K_c t$$

$$\ln(C_{Dsat} - C_{Df}(t)) = \ln(C_{Dsat}) - K_c t$$

$$C_{Dsat} - C_{Df}(t) = C_{Dsat} e^{-K_c t}$$

$$C_{Df}(t) = C_{Dsat} (1 - e^{-K_c t}) \quad (9.25)$$

when $C_{Df}(0) \neq 0 = C_{D0}$, Equation 9.25 becomes:

$$-\ln(C_{Dsat} - C_{Df}(t)) + \ln(C_{Dsat} - C_{D0}) = K_c t$$

$$\ln(C_{Dsat} - C_{Df}(t)) = \ln(C_{Dsat} - C_{D0}) - K_c t$$

$$C_{Dsat} - C_{Df}(t) = (C_{Dsat} - C_{D0}) \cdot e^{-K_c t}$$

$$C_{Df}(t) = C_{Dsat} (1 - e^{-K_c t}) + C_{D0} \cdot e^{-K_c t} \quad (9.26)$$

Equations 9.25 and 9.26 were used to regress values of the apparent rate constant and the saturation concentration against the experimentally measured values of total protein concentration and LongR³IGF-I concentration. Regression was performed using the solver function in MicrosoftTM EXCEL v 5.0c to minimise the sum of squared error.