Design Analysis for Refolding Monomeric Protein

Nicholas Kotlarski, Brian K. O'Neill, Geoffrey L. Francis, and Anton P. J. Middelberg
Cooperative Research Centre for Tissue Growth and Repair, Dept. of Chemical Engineering,
The University of Adelaide, S. A., 5005, Australia

Renaturation of protein expressed as inclusion bodies within Escherichia coli is a key step in many bioprocesses. Operating conditions for the refolding step dramatically affect the amount of protein product recovered, and hence profoundly influence the process economics. The first systematic comparison of refolding conducted in batch, fed-batch and continuous stirred-tank reactors is provided. Refolding is modeled as kinetic competition between first-order refolding (equilibrium reaction) and irreversible aggregation (second-order). Simulations presented allow direct comparison between different flowsheets and refolding schemes using a dimensionless economic objective. As expected from examination of the reaction kinetics, batch operation is the most inefficient mode. For the base process considered, the overall cost of fed-batch and continuous refolding is virtually identical (less than half that of the batch process). Reactor selection and optimization of refolding using overall economics are demonstrated to be vitally important.

Introduction

High-level expression of heterologous protein within *Escherichia coli* (*E. coli*) generally results in inclusion body formation (Rudolph and Lilie, 1996). The major disadvantage of expression as inclusion bodies is that the protein is insoluble and has minimal biological activity. Attempts to produce active protein *in vivo* by coexpression of molecule chaperones to facilitate refolding have not provided an ideal alternative (Rudolph and Lilie, 1996). Advantageously, expression systems have been developed that generate high levels of a desired protein. Inclusion bodies can be cheaply and rapidly concentrated from bacterial cell homogenates to give a mixture with greater than 50% purity (Kane and Hartley, 1991). The protein they contain is very stable, in particular to proteolysis. The problem of imparting potency is overcome by *in vitro* denaturation and refolding. Intuitively, expression as inclusion bodies is highly undesirable if renaturation is inefficient.

Molecular mechanisms for refolding are not fully elucidated. *In vitro* refolding is usually conducted as a batch operation in both laboratory and industrial processes. Conditions that maximize yield are typically determined heuristically at the laboratory scale on a case-by-case basis, but are often grossly suboptimal for large-scale processes. Most work into increasing *in vitro* refolding yield has been directed at controlling the chemical environment (see Buchner and Rudolph, 1991; Hejnaes et al., 1992). While significant increases in yield may be achieved, the "optimum" conditions occur at relatively low protein concentrations. Despite this, when greater amounts of protein are sought it is common to directly scale up the laboratory process. A preferable approach if the refolding pathway and rates of reaction are known, is to use numerical simulation to determine refolding conditions that maximize yield (Vicik and De Bernardez-Clark, 1991).

Semicontinuous or continuous refolding systems are capable of achieving higher yields than batch operation at high overall protein concentration (Rudolph and Fischer, 1990). Simulations have shown the benefits of adopting a continuous refolding strategy (Chaudhuri et al., 1996; Middelberg, 1996). However, overall process economics must be considered when determining optimum operating conditions for any unit operation. Middelberg (1996) employs numerical simulation to clearly illustrate this for refolding in a continuous stirred-tank reactor (CSTR).

The purpose of this study is to apply fundamental reactor design principles to large-scale refolding of monomeric protein. Refolding in batch, fed-batch, and continuous modes is considered within a generic process. A simple kinetic pathway is assumed that accounts for unproductive aggregation during refolding. Numerical simulation is employed to deter-

Correspondence concerning this article should be addressed to A. P. J. Middelberg.
Current address of G. L. Francis: GroPep Pty Ltd., P.O. Box 10065, Gouger Street, Adelaide, S. A., 5000, Australia.

mine the variation in an objective cost function with changes in refolding conditions for all three modes of operation. Simulations are conducted in dimensionless form to render the findings generally applicable to a range of proteins produced using similar processes and to allow easy comparison between the three flow sheets.

Process Description and Design Basis

Three processes that include operations preceding chromatographic purification of the isolated, biologically active product are considered. They are applicable for industrial-scale production of monomeric proteins expressed as inclusion bodies within *E. coli*. The protein product is assumed to contain intramolecular disulfide bonds. Dissolution, refolding, and ultrafiltration are operated in either batch, fed-batch, or continuous modes to give the three different processes shown in Figure 1.

Operation is 24 h/d for 270 fermentation batches/y. The capacity of the process is sufficient to produce 100×10^{-6} kg $\cdot s^{-1}$ of product. Operation of the fermenter is the critical path of the process, with a cycle time of 24 h (Petrides et al., 1995). The capacity of the processing equipment is assumed to match demand, making allowance for inefficiencies. Losses only occur during centrifugation and refolding operations. Recovery of 75% is assumed to give suitable separation for the centrifugation step. Refolding efficiency is determined by the operating conditions of the reactor. Eight M urea is used as the chaotrophic agent for dissolution. Fourfold dilution from denaturing to refolding conditions is required to provide a suitable chemical environment. Ullage of 25% has been included when sizing tanks and stirred reactors. The cost of seed fermenters has not been included. They are necessary to provide an adequate inoculum to the production fermenter, but their capacity is assumed to be insensitive to variation in fermentation requirements.

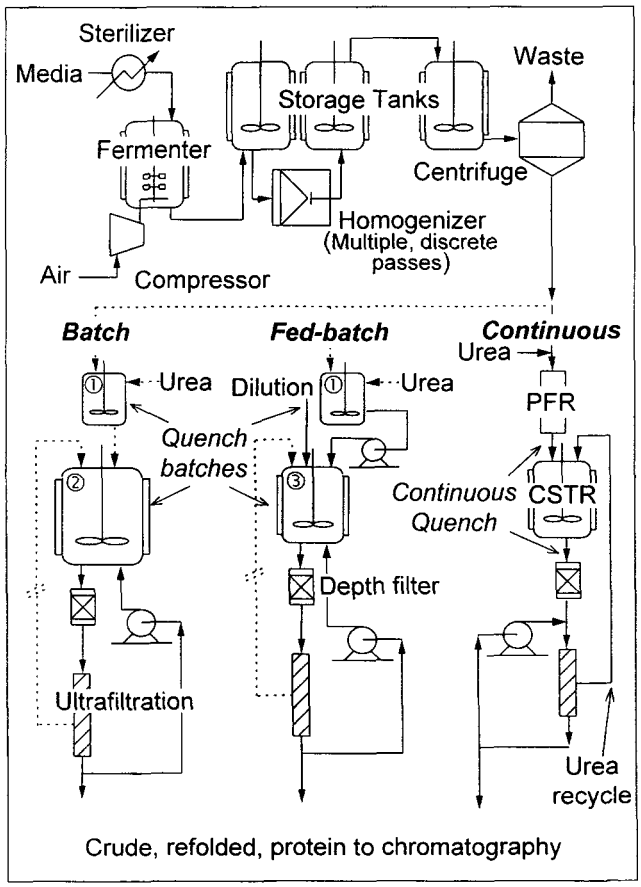


Figure 1. Flowsheets for a production process to isolate and renaturate protein expressed as inclusion bodies in *E. coli*.

Batch, fed-batch, and continuous refolding trains are shown separately, with dashed lines representing batch operation. For batch and fed-batch operation, ultrafiltration is conducted in batch mode, using the refolding reactor as a holding tank. 1 = stirred-tank for dissolution; 2 = stirred batch reactor for refolding; 3 = stirred fed-batch reactor for refolding.

Kinetic Description of Protein Refolding

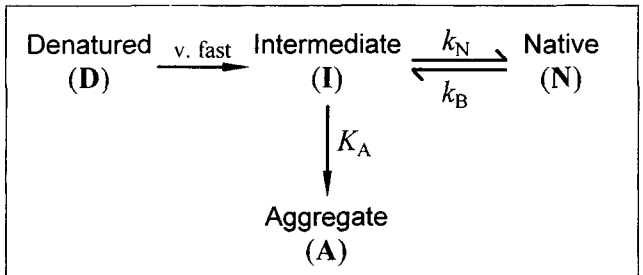
In vitro folding of monomeric proteins is generally a unimolecular reaction via intermediate species (Jaenicke, 1989). Correct refolding has been observed to obey first-order kinetics (Zettlmeissl et al., 1979). This indicates that the rate-limiting reaction that leads to formation of the native conformation is first-order.

Intermediate species may interact in off-pathway reactions to give irreversible aggregation (Brems, 1988). The aggregation reaction is not initiated by monomers in denatured or fully native states, as aggregated protein is partially renatured (Brems, 1988; Mitraki and King, 1989). Cleland and Wang (1990) proposed that dimerization of the first hydrophobic intermediate is the likely rate-limiting step leading to aggregation for most globular proteins. While the measured order of the aggregation reaction slightly exceeds two (Cleland and Wang, 1990; Zettlmeissl et al., 1979), second-order kinetics adequately describe the aggregation reaction (Kiefhaber et al., 1991). A second-order reaction for aggregation is consistent with dimerization limiting the pathway.

Gross structural collapse to an initial intermediate has been observed to occur in less than 0.5 s (Cleland and Wang, 1990; Garel, 1992). In contrast, the refolding time is greater by several orders of magnitude. That is, once exposed to native-favoring conditions, denatured protein effectively assumes an intermediate state, and subsequently reacts slowly to give native protein. In this work it is assumed that refolding results in a single reactive intermediate. The intermediate can either proceed to the native conformation or aggregate irreversibly. It is assumed that once refolding is initiated, all of the fully denatured protein instantaneously form the intermediate (Figure 2). Other groups have proposed simplified models of protein folding pathways that account for the competing aggregation and correct refolding reactions they have observed (Cleland and Wang, 1990; Kiefhaber et al., 1991; Kotlarski et al., 1995; Matthiesen et al., 1996; Zettlmeissl et al., 1979). These models are all consistent with the pathway illustrated in Figure 2.

Batch refolding

For operation at constant volume, the batch-reactor design equation can be used to determine the rates of formation of the key species (I and N) illustrated in Figure 2:



where:

k_N - first-order rate constant of correct refolding

k_B - first-order rate constant of the backward path from native to intermediate

K_A - apparent second-order rate constant of aggregation

Figure 2. Macroscopic kinetic scheme for protein refolding defining pathway and rate constants.

$$\frac{d}{dt}C_I = -k_N C_I - K_A C_I^2 + k_B C_N \tag{1}$$

$$\frac{d}{dt}C_N = k_N C_I - k_B C_N. \tag{2}$$

Equations 1 and 2 can be written in dimensionless form by expressing the concentrations of species I and N in terms of conversion, X, and yield, y_N, and using a dimensionless time. Dimensionless batch refolding time (τ) is simply the product of the first-order rate constant of native protein formation and reaction time (i.e., $\tau = k_N \cdot t$). Hence, Eqs. 2 and 3 become

$$\frac{d}{d\tau}X = 1 - X + \frac{K_A C_{D_o}}{k_N}\cdot(1-X)^2 - \frac{k_B}{k_N}y_N \tag{3}$$

$$\frac{d}{d\tau}y_N = 1 - X - \frac{k_B}{k_N}y_N, \tag{4}$$

where $X = (C_{Do} - C_I)/C_{Do}$ is the conversion of intermediate and $y_N = C_N/C_{Do}$ is the yield of native protein. The initial conditions for Eqs. 3 and 4 are that conversion and yield are both zero. The term $K_A C_{Do}/k_N$ is a dimensionless measure of refolding concentration and k_B/k_N is the equilibrium constant for the formation of native protein.

Fed-batch refolding

The design equation for a fed-batch reactor, of volume V, gives the rates of formation of I and N as

$$\frac{d}{dt}N_I = \frac{\rho_D Q}{M_N} - k_N N_I + k_B N_N - K_A\frac{N_I^2}{V} \tag{5}$$

$$\frac{d}{dt}N_N = k_N N_I - k_B N_N \tag{6}$$

$$\frac{d}{dt}V = 4Q, \tag{7}$$

where ρ_D is the mass concentration of denatured protein in the inlet stream, Q is the inlet volumetric flow rate, and M_N is the molar mass of the protein. The rate of change in volume is due to the fourfold dilution required to give a suitable chemical environment for refolding. The protein dilution is greater than fourfold due to the initial charge in the reactor.

By setting the system boundary for the refolding reaction to include the dissolution reactor, an expression for conversion of intermediate species can be manipulated to give

$$X = \frac{\tau}{\tau_{\text{Feeding}}} - \frac{N_I}{N_{D_o}}, \tag{8}$$

where τ is the dimensionless time of refolding and τ_{Feeding} is the dimensionless duration of feeding. For fed-batch operation, the yield must be defined in terms of moles of species (i.e., $y_N = N_N/N_{D_o}$) as volume is not constant in the reactor. To allow direct comparison of fed-batch and batch refolding, the final nominal protein concentration should be the same for both processes. To achieve this the initial charge will be the final volume less the volume that is fed during refolding:

$$V_o = \frac{\dot{m}_N t}{C_{D_o} M_N y_{N\,\text{final}}} - \frac{4\dot{m}_N t}{\rho_D y_{N\,\text{final}}}. \tag{9}$$

Using the previous terms for dimensionless time, the expressions for conversion, yield, and initial volume can be used to write Eqs. 5 to 7 as

$$\frac{d}{d\tau}X = -\frac{k_B}{k_N}y_N + \left(\frac{\tau}{\tau_{\text{Feeding}}} - X\right) + \frac{K_A C_{D_o}}{k_N\left(1 - \frac{4C_{D_o}M_N}{\rho_D}\right)\left(\frac{V}{V_o}\right)}\cdot\left(\frac{\tau}{\tau_{\text{Feeding}}} - X\right)^2 \tag{10}$$

$$\frac{d}{d\tau}y_N = \frac{\tau}{\tau_{\text{Feeding}}} - X - \frac{k_B}{k_N}y_N \tag{11}$$

$$\frac{d}{d\tau}\frac{V}{V_o} = \frac{4}{\tau_{\text{Feeding}}\cdot\left(\frac{\rho_D}{C_{D_o}M_N} - 4\right)}. \tag{12}$$

The initial conditions for Eqs. 10, 11, and 12 are that conversion and yield are zero and the initial volume ratio is one (i.e., $V = V_o$).

Continuous refolding

For an ideally mixed CSTR, the design equations are

$$C_I = C_{D_o} - (k_N C_I + K_A C_I^2 - k_B C_N)\cdot t \tag{13}$$

$$C_N = (k_N C_I - k_B C_N)\cdot t, \tag{14}$$

where t is the residence time in the reactor. The dimensionless time is equivalent to a Damköhler number ($\tau = k_N t$ = Damköhler number). Equations 13 and 14 can be used with

conversion, yield, and dimensionless time expressions for batch refolding to give

$$X = \tau\left[(1-X) + \frac{K_A C_{D_o}}{k_N}(1-X)^2 - \frac{k_B}{k_N} y_N\right] \quad (15)$$

$$y_N = \frac{1-X}{\frac{1}{\tau} + \frac{k_B}{k_N}}. \quad (16)$$

Substitution of Eq. 16 into Eq. 15 can be simplified to yield

$$\tau^2\left[Da_{BN} Da_{AN}(1-X)^2\right] + \tau\left[1 - X + Da_{AN}(1-X)^2 - Da_{BN} X\right] - X = 0, \quad (17)$$

where Da_{BN} is the ratio of Damköhler numbers for the reverse and forward folding reactions ($k_B t/k_N t$) and Da_{AN} is the ratio of Damköhler numbers for the aggregation and correct refolding reaction ($K_A C_{D_o} t/k_N t$) (Middelberg, 1996). Da_{BN} is equivalent to the equilibrium constant for native protein formation, and Da_{AN} can be viewed as dimensionless refolding concentration. Equation 17 defines a quadratic expression for τ, soluble for sets of Da_{BN}, Da_{AN}, and X. The solution can be substituted into Eq. 16 to determine the corresponding yield.

Objective Function

The performance indicator used to compare different schemes is a dimensionless operating-cost function (Middelberg, 1996). This function is the sum of three components (annual direct-fixed-capital (DFC)-dependent cost, consumables, and waste treatment) normalized by a characteristic cost. The annual DFC-dependent costs (ϕ_{dfc}) account for maintenance, depreciation, insurance, local taxes, and factory expenses. Hence, the objective function is

$$\phi^* = \frac{\phi_{dfc} + \phi_{\text{con}} + \phi_{wt}}{\phi_{ch}} = \phi_{dfc}{}^* + \phi_{\text{con}}{}^* + \phi_{wt}{}^* \quad (18)$$

where ϕ_{ch} is a characteristic cost and the superscript indicates dimensionless annual cost component. The characteristic cost is the purchase cost of a fermenter with the capacity to produce the mass of a desired product specified by the design basis for 100% recovery.

Costing and Design

Evaluation of the three components was performed following the method outlined by Middelberg (1996). All costs were calculated in 1994 US$. Equipment purchase costs were calculated using continuous cost-capacity functions of the form

$$(\text{Cost})_j = 10^{\alpha_j}(\text{Capacity})^{\beta_j}, \quad (19)$$

where α_j and β_j are exponents and capacity is a characteristic measurement particular to unit j. The DFC-dependent costs were calculated by multiplying the total sum of equipment purchase costs by a factor, ζ. The value of ζ for this type of process is approximately 1.43 (Petrides et al., 1995). Table 1 summarizes the values of the cost-capacity exponents and capacity expressions used for each unit. The cost function for fermenters is not continuous. A set of exponents are presented for fermenters with volume between one and fifty cubic meters ($1 < V_f < 50$ m^3). The second values are used to estimate the cost of fermenters exceeding fifty cubic meters in volume ($V_f > 50$ m^3).

Material usage was calculated for a batch of product, and the total cost determined using conservative estimates of specific costs. Two process streams were identified as contributing virtually all of the waste: supernatant from centrifugation and retenate from ultrafiltration. Volume and biological oxygen demand components were used to calculate the cost of treating these two streams. Table 1 contains values of the specific consumable and waste-treatment costs with the corresponding expressions for total cost. Other process-specific and design parameters are presented in Table 2.

The only protein-specific variables that have been used are reaction-rate constants and protein molar mass. An analog of insulin-like growth factor-1 (Long-R^3-IGF-1, Wells et al., 1994) has been used as the model protein. It is a small molecule with a molar mass of 9.1 kg·mol^{-1} containing three internal disulfide bonds. Kotlarski et al. (1995) have determined that for Long-R^3-IGF-1 the rate constant for formation of native species is $k_N = 8.5 \times 10^{-4}$ s^{-1} and for formation of the aggregate species is $K_A = 0.0544$ m^3·mol^{-1}·s^{-1}. Cleland and Wang (1990) and Garel (1992) have published rate constants that compare favorably. However, a range of values (varying over three orders of magnitude) has been published for different proteins (Garel, 1992; Kiefhaber et al., 1991). Some results are based on the disappearance of denatured species rather than the appearance of product. These are of no use in reactor design.

Simulation Results

Simulations are conducted using Mathcad 6.0 Plus to solve the design equations outlined before and the corresponding dimensionless cost components. For base-case simulation, the rate constant for the back-reaction of native to intermediate was set equal to 1% of the forward reaction (i.e., $k_B/k_N = 0.01$).

Base case

Figure 3 is the dimensionless cost contour for batch refolding as a function of conversion and dimensionless protein concentration. For the presented range, the minimum is 28.6. This occurs at the maximum of the conversion range (0.996) and a dimensionless concentration of 6.1. 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 a yield of approximately 31% at the optimum (Figure 4). The dimensionless time of refolding at this point is 4.9.

Fed-batch operation for the base case gives a minimum of 12.0, or less than 50% of the batch refolding cost (Figure 5). 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

Table 1. Values of Exponents, Specific Costs, and Capacity Expressions for Cost-capacity Evaluation Used in Simulation*

Unit	Parameter	Capacity Expression
Air compressor	$\alpha_a = 5.41,\ \beta_a = 0.98$**,†	$Q_a = \frac{\dot{m}_N t}{ETW\epsilon f y_N} \cdot \frac{\delta}{60}$
Centrifuge	$\alpha_c = 3.68,\ \beta_c = 0.33$**	$\Sigma = \frac{\dot{m}_N t}{ETW f y_N} \cdot \frac{1}{t_c Q_\Sigma}$
Dissolution reactor	$\alpha_d = 4.41,\ \beta_d = 0.59$ (assumed $= \alpha_t$)	$V_d = \frac{\dot{m}_N T'}{\rho_d y_N}$ where $T' = \tau_d$ or t/ϵ
Fermenter	$\alpha_f = 5.04$ or 4.66 $\beta_f = 0.36$ or 0.70††	$V_f = \frac{\dot{m}_N t}{ETW\epsilon} \cdot \frac{1}{f y_N}$
Homogenizer	$\alpha_h = 6.44,\ \beta_h = 0.48$**,††	$Q_h = \frac{\dot{m}_N t}{ETW f y_N} \cdot \frac{N_p}{t_h}$
Sterilizer	$\alpha_s = 7.10,\ \beta_s = 0.60$‡ (assumed)	$Q_s = \frac{\dot{m}_N t}{ETW f y_N t_s}$
Storage tank	$\alpha_t = 4.41,\ \beta_t = 0.59$‡	$V_t = \frac{\dot{m}_N t}{ETW\epsilon} \cdot \frac{1}{f y_N}$
Refolding reactor	$\alpha_r = 4.41,\ \beta_r = 0.59$‡ (assumed $= \alpha_t$)	$V_r = \frac{\dot{m}_N T'}{C_{Do} M_N y_N \epsilon}$ where $T' = \tau_r$ or t
Ultrafiltration unit	$\alpha_u = 3.29,\ \beta_u = 0.87$††	$A_u = \frac{\dot{m}_N t}{C_{Do} M_N y_N J (t_{ru} - t_r)}$ for fedbatch $A_u = \frac{\dot{m}_N}{C_{Do} M_N y_N J}$ for continuous
Dissolution additives	$\gamma_{ad} = 250$‡	$\hat{\phi}_{ad} = \frac{\dot{m}_N t}{\rho_d y_N} \cdot \gamma_{ad}$
Refolding additives	$\gamma_{ar} = 120$‡‡	$\hat{\phi}_{ar} = \frac{\dot{m}_N t}{C_{Do} M_N y_N} \cdot \gamma_{ar}$
Depth filter	$\gamma_{df} = 180$‡,§	$\hat{\phi}_{df} = \frac{\dot{m}_N t}{C_{Do} M_N y_N} \cdot \hat{A}_{df} \cdot \gamma_{df}$
Fermentation media	$\gamma_m = 210$‡‡	$\hat{\phi}_m = \frac{\dot{m}_N t}{ETW f y_N} \cdot \gamma_m$
Ultrafiltration membrane	$\gamma_u = 200$††,‡	$\hat{\phi}_u = \frac{\dot{m}_N}{C_{Do} M_N y_N JL} \cdot \gamma_u$ for continuous $\hat{\phi}_u = \frac{\dot{m}_N t}{C_{Do} M_N y_N JL \cdot (t_{ru} - t_u)} \cdot \gamma_u$ for fedbatch
Urea	$\gamma_{up} = 1.50$‡	$\hat{\phi}_{ud} = \frac{M_u C_{ud} \dot{m}_N t}{\rho_d y_N} \cdot \gamma_{up}$
Urea recycle	$\gamma_{urec} = 0.75$ $(= 0.5\ \gamma_{up})$	$\hat{\phi}_{urec} = \frac{\dot{m}_N t}{y_N} \left(\frac{M_u C_{ur}}{C_{Do} M_N} - \frac{M_u C_{ud}}{\rho_d} \right) \cdot \gamma_{urec}$
Waste treatment	$\gamma_{wtcb} = 1.00$‡ $\gamma_{wtcv} = 0.20$‡	$\hat{\phi}_{wt} = \frac{\dot{m}_N t}{y_N} \left(\frac{\gamma_{wtcv} + (1-E) TW \gamma_{wtcb}}{fETW} + \frac{\gamma_{wtcv}}{\rho_d} + \frac{M_u C_{ud}}{\rho_d} \gamma_{wtcb} \right)$

*Exponents are for equipment fabricated in grade 316 stainless-steel, free on board. Reference to the most relevant source(s) used to estimate each parameter is given.
**Middelberg et al., 1992.
†Peters and Timmerhaus, 1990.
††Kalk and Langlykke, 1986.
‡Petrides et al., 1995.
‡‡Middelberg, 1996.
§Petrides et al., 1989.

Table 2. Values and References for Input Parameters Used for Base-Case Simulation*

Parameter	Value	Reference
$\hat{A}_{df}$	0.04	Petrides et al., 1989
C_{ud}	8 M	Fischer et al., 1993, see text
C_{ur}	2 M	Fischer et al., 1993, see text
E	0.20	Kiefhaber et al., 1991; Middelberg et al., 1992
f	0.75	Design basis
J	5×10^{-6}	Kalk and Langlykke, 1986; Kotlarski et al., 1995
k_B/k_N	0.01	Design basis (base case)
k_N	8.5×10^{-4}	Kotlarski et al., 1995, see text
K_A/k_N	64	Kotlarski et al., 1995, see text
L	100	Middelberg, 1996
$\dot{m}_N$	100×10^{-6}	Design basis
$M_N = M_I$	9.1	Design basis (Long-R^3-IGF-1)
M_u	0.060	—
N_p	4	Middelberg et al., 1992; Petrides et al., 1989
Q_Σ	5×10^{-9}	Middelberg and O'Neill, 1991; Wong et al., 1996
t	86,400	Design basis
t_c, t_h	72,000	Design basis
t_d	3,600	Fischer et al., 1993; Turner et al., 1994; design basis
t_{ru}	68,400	Design basis
t_s	10,800	Petrides et al., 1989; design basis
T	0.55	Niedhardt, 1987; Middelberg et al., 1992
W	30	Kalk and Langlykke, 1986; Petrides et al., 1989, 1995
δ	1.0	Yoon et al., 1994; Turner et al., 1994; design basis
ϵ	0.75	Design basis
ρ_D	50	Design basis
τ_d	3,600	Fischer et al., 1993; design basis
ζ	1.43	Petrides et al., 1989, 1995

*Reference to the most relevant source is given.

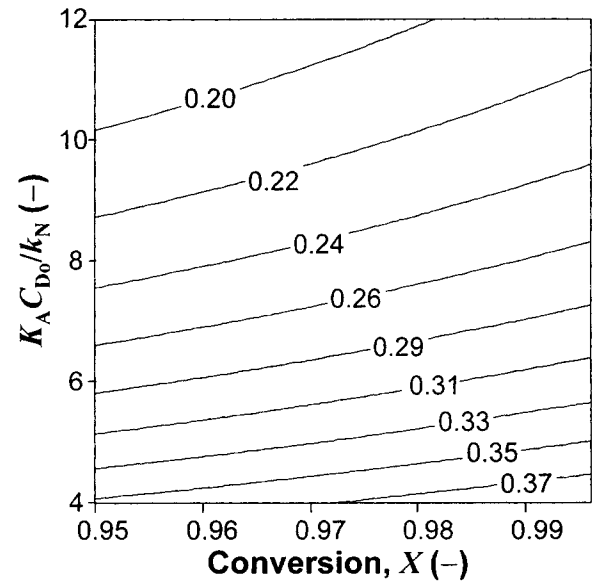


Figure 4. Yield contour (y_N) for batch refolding as a function of conversion and dimensionless protein concentration ($K_A C_{D_0}/k_N$) for k_B/k_N = 0.01.

approximately 65% of the total. Consumable costs contribute approximately 32%, and waste management costs 3%. There is a severe penalty for operation 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 6). A dimensionless time contour reveals that the optimum duration of feeding is 42 (contour not presented).

Optimum operating conditions for continuous refolding occur at a conversion of 0.985 and a dimensionless protein con-

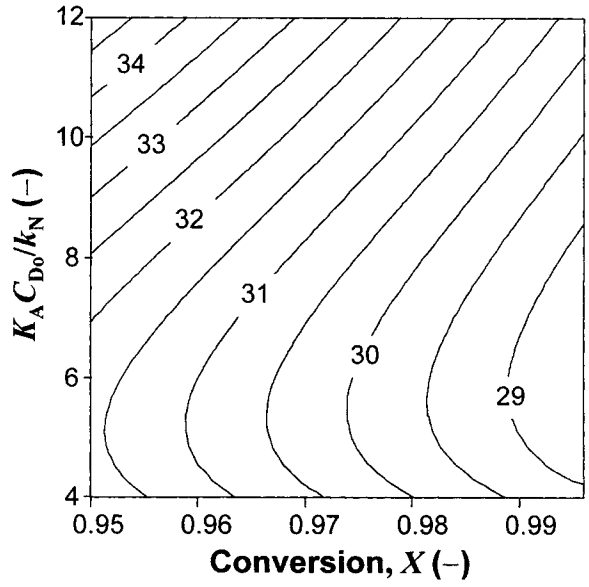


Figure 3. Dimensionless cost contour (ϕ^*) for batch refolding as a function of conversion and dimensionless protein concentration ($K_A C_{D_0}/k_N$) for k_B/k_N = 0.01.

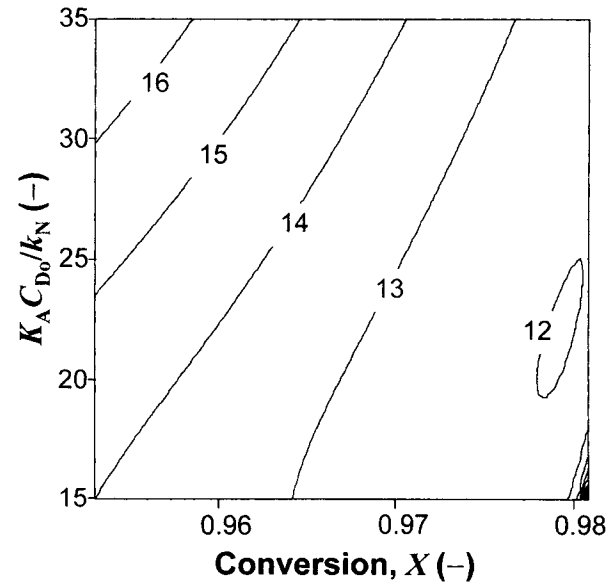


Figure 5. Dimensionless cost contour (ϕ^*) for fed-batch refolding as a function of conversion and dimensionless protein concentration ($K_A C_{D_0}/k_N$) for k_B/k_N = 0.01.

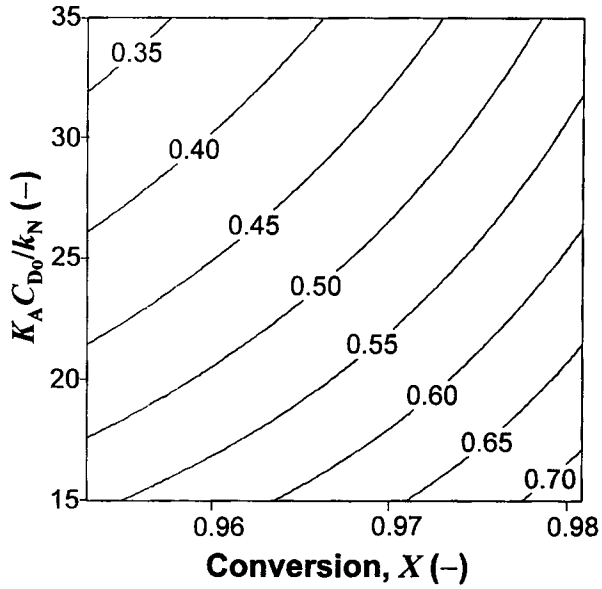


Figure 6. Yield contour (y_N) for fed-batch refolding as a function of conversion and dimensionless protein concentration ($K_A C_{D_o}/k_N$) for $k_B/k_N = 0.01$.

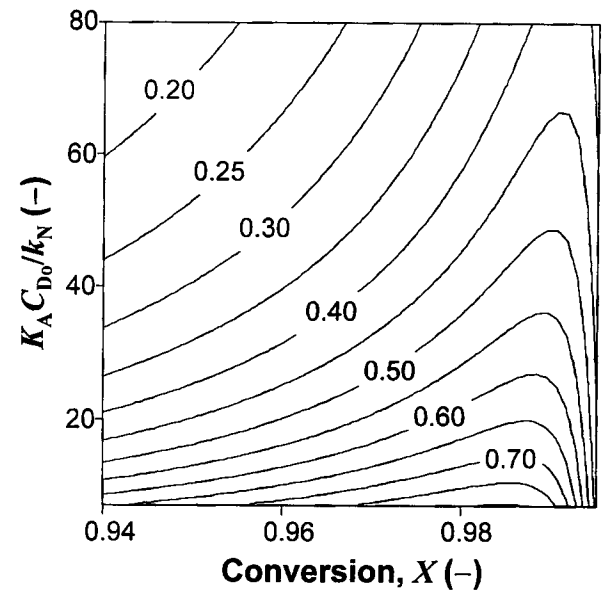


Figure 8. Yield contour (y_N) for continuous refolding as a function of conversion and dimensionless protein concentration ($K_A C_{D_o}/k_N$) for $k_B/k_N = 0.01$.

centration of 21.6. Figure 7 shows the variation in objective function with conversion and dimensionless protein concentration for this refolding mode. The minimum value is approximately 5% less than for fed-batch operation (11.4) and 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 8). This yield is more than twice the yield achieved for overall optimum operation of the batch process. The Damköhler number (dimensionless time) at the optimum is approximately 74. Figure 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.

Variation in equilibrium

Simulations were conducted for two additional values of the equilibrium constant ($k_B/k_N = 0$ and 0.1). For all three

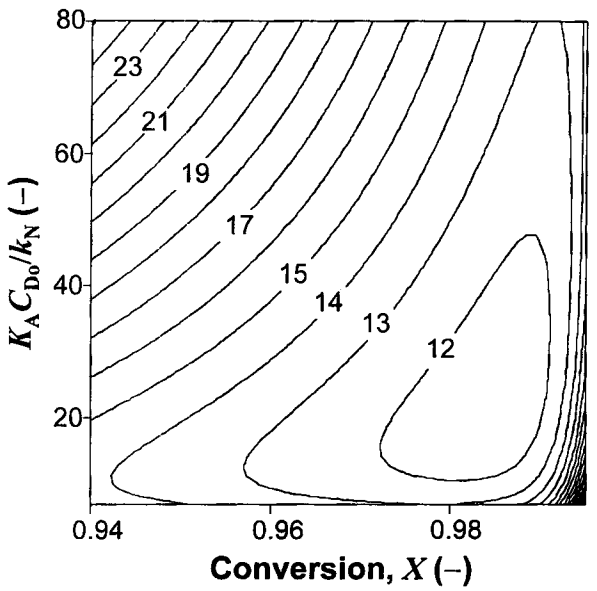


Figure 7. Dimensionless cost contour (ϕ^*) for continuous refolding as a function of conversion and dimensionless protein concentration ($K_A C_{D_o}/k_N$) for $k_B/k_N = 0.01$.

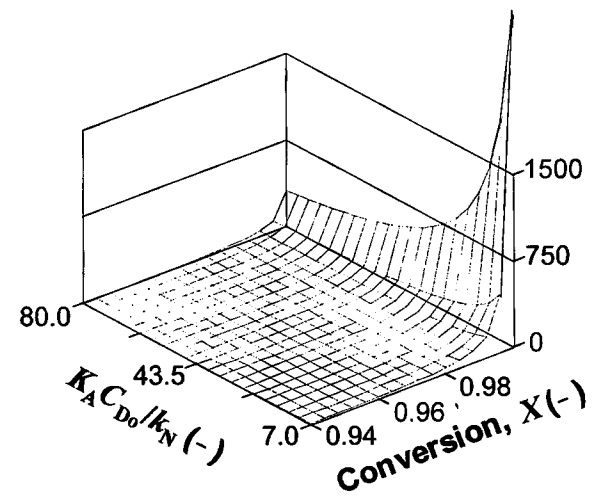


Figure 9. Dimensionless time contour (τ) for continuous refolding as a function of conversion and dimensionless protein concentration ($K_A C_{D_o}/k_N$) for $k_B/k_N = 0.01$; the minimum value of the objective cost function occurs at $X = 0.985$ and $K_A C_{D_o}/k_N = 21.6$, where $\tau = 74$.

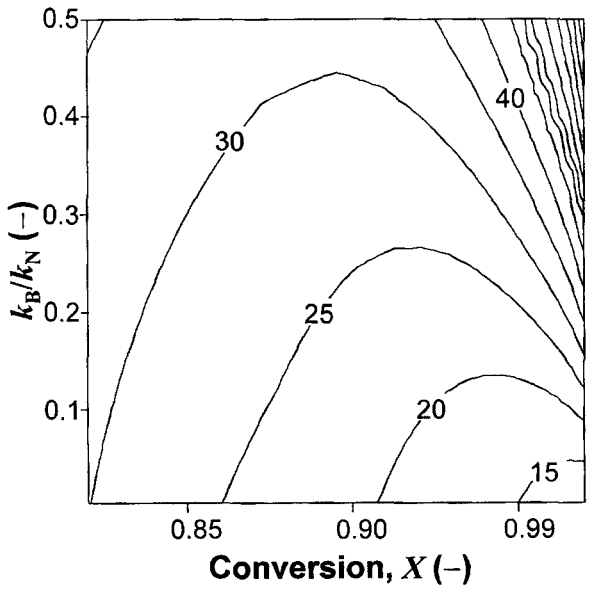


Figure 10. Dimensionless cost contour (ϕ^*) for fed-batch refolding as a function of conversion and equilibrium constant (k_B/k_N); dimensionless protein concentration is 22 ($K_A C_{D_o}/k_N = 22$).

modes of refolding, the optimum operating conversion decreases with increasing equilibrium constant at fixed protein concentration. For pure competing 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 (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 batch operation occurred at dimensionless protein concentrations of approximately 6 for both extreme values of the equilibrium constant. The trend toward 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.

Figure 10 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 an equilibrium constant of 0.01. Batch and continuous operation give contours with similar form. Batch refolding is least sensitive to variation in equilibrium constant, particularly at high concentration. Performance of continuous and fed-batch operation is quite similar over the range of conditions considered (results not shown).

Discussion

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 Long-R^3-IGF-1. While higher protein concentrations have been reported in commercial-scale production (Bentle et al., 1987; Chang et al., 1994), it is 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 (Fisher et al., 1993).

Batch operation is the most expensive mode for refolding. Optimum conditions were identified as relatively low protein concentration, high conversion (> 0.97), and short duration of refolding to give low yield (≈ 30%). Higher yields could have been achieved by using lower protein concentration. However, low concentration requires increased liquid volumes and is penalized 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. In the batch mode, high conversion is rapidly obtained since no fresh reactants are added. Figure 4 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 for operation at optimum conditions. This is intuitive as it has the advantage of maximum equipment utilization. Optimum conditions favor operation at high conversion and high overall protein concentration. However, unlike batch refolding, intermediate species (in the form of denatured species) is 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 Long-R^3-IGF-1, but the intermediate concentration maintained in the CSTR is only 0.04 kg·m^{-3}. Low intermediate concentration (20 times less than batch refolding) produces 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 (1996), given minor differences in costing.

Continuous refolding has the advantage that the duration of refolding can extend beyond the length of the critical path. The dimensionless time of the critical path is 58 (corresponding to 24 h) yet duration of the reaction for optimum base-case operation was 74. Figure 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. The rapid decrease occurs with a dramatic increase in the required holding time, as illustrated in Figure 9. The reason for this is unfolding of the native protein to intermediate conformation. That is, even for a favorable equilibrium (1% back reaction), unfolding of the native protein to intermediate that aggregates is significant if the reaction is allowed to proceed too long. It appears that for the base case, unfolding becomes significant for durations approximately greater than the critical path. The cost

penalty for operation within the range that the back reaction is significant is clearly shown in Figure 7 by the concentration of cost contours to the right of the optimum. The advantage of extending the reaction time beyond the length of the critical path may also be the most significant disadvantage. Unlike batch and fed-batch operation, batch integrity is not maintained throughout the process.

Performance of fed-batch operation is virtually identical to that of continuous refolding over the same range of conversion and protein concentration (Figures 5 and 7). Optimum conditions are a dimensionless protein concentration of 22 (3.1 $kg \cdot m^{-3}$ for Long-R^3-IGF-1), final dimensionless time of refolding of 42.5, yield of 63%, and conversion of 0.98. Investigation of the transient intermediate concentration reveals that it increases rapidly to a maximum of 0.066 $kg \cdot m^{-3}$ and then slowly decays to a final concentration of 0.064 $kg \cdot m^{-3}$. 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 concentration of cost contours at high conversion and low concentration (see Figure 5) is due to this effect and is the major difference between the fed-batch and continuous operation.

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.

Sensitivity of overall performance to changes in consumable costs was assessed by simulation over a fourfold range about the mean estimated value for each specific cost. Batch refolding is most sensitive because consumable costs contribute greater than 50% of the overall cost for near-optimal conditions of operation. For fed-batch and continuous operation the consumable cost contribution is less than 35%. 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. Simulations reveal the order of significance is $\gamma_m > \gamma_{ar} > \gamma_{up} > \gamma_{ad} > \gamma_{df} \cong \gamma_u$, with approximately $+15/-8\%$ variation with media cost and only $+0.5/-0.2\%$ variation with filtration costs.

Conclusion

A method for optimization 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 of this article are a function of the process, method of costing, and refolding pathway. Since a generic process and widely applicable refolding pathway have been used, however, the results should be relevant to the production of many recombinant proteins.

Continuous operation of protein refolding under optimized conditions is the most cost-effective method for renaturation. Fed-batch operation can be optimized to attain almost the same 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%. Batch operation is more sensitive to changes in consumable costs than either fed-batch or continuous refolding because greater 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, insensitivity is due to the high base cost of the overall process for operation at low yield, rather than the robustness of the refolding unit itself.

The importance of considering fed-batch and continuous modes of refolding as alternatives to batch operation has been clearly demonstrated. Overall process optimization must be used to determine operating conditions for individual units of equipment. Economic criteria must be used to identify optimum operating conditions.

Notation

$\hat{A}_{df}$ = specific surface area of depth filter, m^2 $(m^3$ refold$)^{-1}$
A_u = ultrafiltration unit membrane, m^2
C = concentration, $mol \cdot m^{-3}$
E = product expression level, kg product (kg protein)$^{-1}$
f = fractional mass recovery of product (excluding refolding)
J = flux through ultrafiltration membrane, $m^3 \cdot m^{-2} \cdot s^{-1}$
k_N = first-order rate constant of correct refolding, s^{-1}
k_B = first-order rate constant of reverse reaction, s^{-1}
K_A = apparent second-order rate constant of aggregation, $m^3 \cdot mol^{-1} \cdot s^{-1}$
L = life of ultrafiltration membrane, batches
$\dot{m}$ = mass production rate, $kg \cdot s^{-1}$
N = molar amount, mol
N_p = number of discrete homogenizer passes
Q_Σ = normalized centrifuge feed rate (Q/Σ), $m \cdot s^{-1}$
t = fermentation or batch cycle time, s
T = total cell protein to dry-cell weight ratio, kg protein (kg cells)$^{-1}$
W = dry-cell weight concentration, kg cells $\cdot m^{-3}$
γ = specific cost per kg, m^2, or m^3, 1994 U.S. \$ (unit)$^{-1}$
ϵ = design margin to provide sufficient ullage (V_b/V_f)
ζ = annual DFC-dependent cost to equipment-purchase cost ratio
Σ = equivalent centrifuge clarification area, m^2
$\hat{\phi}$ = cost per batch, 1994 U.S. \$
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
I = intermediate conformation protein
h = homogenizer

m = fermentation media
N = native (correctly folded protein)
o = inlet/initial conditions to refolding reactor
r = refolding
ru = refolding and ultrafiltration
s = sterilizer
t = storage tank
u = ultrafiltration or urea
ud = urea for dissolution
up = urea purchase
$urec$ = urea recovered
$wtcb$ = waste treatment (biomass component)
$wtcv$ = waste treatment (volume component)

Literature Cited

Bentle, L. A., J. W. Mitchell, and S. B. Storrs, "Method of Somatotropin Naturation," U.S. Patent No. 4,652,630 (1987).

Brems, D. N., "Solubility of Different Folding Conformers to Bovine Growth Hormone," *Biochemistry*, **27**, 4541 (1988).

Buchner, J., and R. Rudolph, "Renaturation, Purification, and Characterization of Recombinant F_{ab}-Fragments Produced in *Escherichia coli*," *Bio/Technol.*, **9**, 157 (1991).

Chang, J. Y., N. C. McFarland, and J. R. Schwartz, "Method for Refolding Insoluble, Misfolded Insulin-like Growth Factor-I into an Active Conformation," U.S. Patent No. 5,288,931 (1994).

Chaudhuri, J. B., B. Batas, and A. D. Guise, "Improving Protein Refolding Yields by Minimizing Aggregation," *Ann. N.Y. Acad. Sci.*, **782**, 495 (1996).

Cleland, J. L., and D. I. C. Wang, "Refolding and Aggregation of Bovine Carbonic Anhydrase B: Quasi-Elastic Light Scattering Analysis," *Biochemistry*, **29**, 11072 (1990).

Fischer, B., I. Sumner, and P. Goodenough, "Isolation, Renaturation, and Formation of Disulfide Bonds of Eukaryotic Proteins Expressed in *Escherichia coli* as Inclusion Bodies," *Biotechnol. Bioeng.*, **41**, 3 (1993).

Garel, J.-R., "Folding of Large Proteins: Multidomain and Multisubunit Proteins," *Protein Folding*, T. E. Creighton, ed., Freeman, New York, p. 405 (1992).

Hejnaes, K. R., S. Bayne, L. Nørskov, H. H. Sørensen, J. Thomsen, L. Schäffer, A. Wollmer, and L. Skriver, "Development of an Optimized Refolding Process for Recombinant Ala-Glu-IGF-1," *Prot. Eng.*, **5**, 597 (1992).

Jaenicke, R., "Thermodynamic and Kinetic Aspects of the Refolding and Selfassembly of Proteins," *GBF Monogr.*, **12**, 3 (1989).

Kalk, J. P., and A. F. Langlykke, *Manual of Industrial Microbiology and Biotechnology*, A. L. Demain and N. A. Solomon, eds., ACS, p. 363 (1986).

Kane, J. F., and D. L. Hartley, "Properties of Recombinant Protein Containing Inclusion Bodies in *Escherichia coli*," *Purification and Analysis of Recombinant Proteins*, R. Seetharam and S. K. Sharma, eds., Dekker, New York, p. 121 (1991).

Kiefhaber, T., R. Rudolph, H.-H. Kohler, and J. Buchner, "Protein Aggregation *in vitro* and *in vivo*: A Quantitative Model for the Kinetic Competition Between Folding and Aggregation," *Bio/Technol.*, **9**, 825 (1991).

Kotlarski, N., R. S. Yeates, S. J. Milner, G. L. Francis, B. K. O'Neill, and A. P. J. Middelberg, "Studies into the Scale-up of a Process to Produce a Biosynthetic Insulin-like Growth Factor," *Trans. Inst. Chem. Eng. C*, **73**, 27 (1995).

Matthiesen, F., K. R. Hejnaes, and L. Shriver, "Stabilization of Recombinantly Expressed Proteins," *Ann. N.Y. Acad. Sci.*, **782**, 413 (1996).

Middelberg, A. P. J., "The Influence of Protein Refolding Strategy on Cost for Competing Reactions," *Chem. Eng. J.*, **61**, 41 (1996).

Middelberg, A. P. J., and B. K. O'Neill, "Monitoring the Centrifugal Recovery of Recombinant Protein Inclusion Bodies," *Aust. J. Biotechnol.*, **5**, 87 (1991).

Middelberg, A. P. J., B. K. O'Neill, and I. D. L. Bogle, "Modelling Bioprocess Interactions for Optimal Design and Operating Strategies," *Trans. Inst. Chem. Eng. C*, **70**, 8 (1992).

Mitraki, A., and J. King, "Protein Folding Intermediates and Inclusion Body Formation," *Bio/Technol.*, **7**, 690 (1989).

Niedhardt, F. C., "Chemical Composition of *Escherichia coli*, Escherichia coli *and* Salmonella typhimurium," *Cellular and Molecular Biology*, Vol. 1, F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger, eds., ACS, p. 3 (1987).

Peters, M. S., and K. D. Timmerhaus, *Plant Design and Economics for Chemical Engineers*, 4th ed., McGraw-Hill, New York, p. 529 (1990).

Petrides, D., C. L. Cooney, L. B. Evans, R. P. Field, and M. Snoswell, "Bioprocess Simulation: An Integrated Approach to Process Development," *Comput. Chem. Eng.*, **13**, 553 (1989).

Petrides, D., E. Sapidou, and J. Calandranis, "Computed-Aided Process Analysis and Economic Evaluation for Biosynthetic Human Insulin Production—A Case Study," *Biotech. Bioeng.*, **48**, 529 (1995).

Rudolph, R., and S. Fischer, "Process for Obtaining Renatured Proteins," U.S. Patent No. 4,933,434 (1990).

Rudolph, R., and H. Lilie, "*In vitro* Folding of Inclusion Body Proteins," *FASEB J.*, **10**, 49 (1996).

Turner, C., M. E. Gregory, and N. F. Thornhill, "Closed-Loop Control of Fed-Batch Cultures of Recombinant *Escherichia coli* Using On-line HPLC," *Biotechnol. Bioeng.*, **44**, 819 (1994).

Vicik, S., and E. De Bernadez-Clark, "An Engineering Approach to Achieving High-Protein Refolding Yields," *Protein Refolding*, G. Georgiou and E. De Bernardez-Clark, eds., *ACS Symp. Ser.*, **470**, 180 (1991).

Yoon, S. K., S. H. Kwon, M. G. Park, W. K. Kang, and T. H. Park, "Optimization of Recombinant *Escherichia coli* Fed-Batch Fermentation for Bovine Somatotropin," *Biotechnol. Lett.*, **16**, 1119 (1994).

Wells, J. R. E., R. M. King, and G. L. Francis, "Growth Hormone Fusion Proteins, Methods of Production, and Methods of Treatment," U.S. Patent No. 5,330,971 (1994).

Wong, H. H., B. K. O'Neill, and A. P. J. Middelberg, "Centrifugal Recovery and Dissolution of Recombinant Gly-IGF-II Inclusion-Bodies: The Impact of Feedrate and Recentrifugation on Protein Yield," *Bioseparation*, **6**, 185 (1996).

Zettlmeissl, G., R. Rudolph, and R. Jaenicke, "Reconstitution of Lactic Dehydroganase. Noncovalent Aggregation vs. Reactivation: 1. Physical Properties and Kinetics of Aggregation," *Biochemistry*, **18**, 5567 (1979).

Manuscript received Oct. 29, 1996, and revision received Feb. 24, 1997.