Review Article

Journal of Molecular Recognition

Received: 10 April 2008, Revised: 16 April 2008, Accepted: 18 April 2008, Published online in Wiley InterScience: 10 June 2008

(www.interscience.wiley.com) DOI:10.1002/jmr.898

Protein purification using chromatography: selection of type, modelling and optimization of operating conditions

J. A. Asenjo[a]* and B. A. Andrews[a]

To achieve a high level of purity in the purification of recombinant proteins for therapeutic or analytical application, it is necessary to use several chromatographic steps. There is a range of techniques available including anion and cation exchange, which can be carried out at different pHs, hydrophobic interaction chromatography, gel filtration and affinity chromatography. In the case of a complex mixture of partially unknown proteins or a clarified cell extract, there are many different routes one can take in order to choose the minimum and most efficient number of purification steps to achieve a desired level of purity (e.g. 98%, 99.5% or 99.9%).

This review shows how an initial 'proteomic' characterization of the complex mixture of target protein and protein contaminants can be used to select the most efficient chromatographic separation steps in order to achieve a specific level of purity with a minimum number of steps. The chosen methodology was implemented in a computer- based Expert System. Two algorithms were developed, the first algorithm was used to select the most efficient purification method to separate a protein from its contaminants based on the physicochemical properties of the protein product and the protein contaminants and the second algorithm was used to predict the number and concentration of contaminants after each separation as well as protein product purity.

The application of the Expert System approach was experimentally tested and validated with a mixture of four proteins and the experimental validation was also carried out with a supernatant of *Bacillus subtilis* producing a recombinant β-1,3-glucanase.

Once the type of chromatography is chosen, optimization of the operating conditions is essential. Chromatographic elution curves for a three-protein mixture (α-lactoalbumin, ovalbumin and β-lactoglobulin), carried out under different flow rates and ionic strength conditions, were simulated using two different mathematical models. These models were the Plate Model and the more fundamentally based Rate Model. Simulated elution curves were compared with experimental data not used for parameter identification. Deviation between experimental data and the simulated curves using the Plate Model was less than 0.0189 (absorbance units); a slightly higher deviation [0.0252 (absorbance units)] was obtained when the Rate Model was used. In order to optimize operating conditions, a cost function was built that included the effect of the different production stages, namely fermentation, purification and concentration. This cost function was also successfully used for the determination of the fraction of product to be collected (peak cutting) in chromatography. It can be used for protein products with different characteristics and qualities, such as purity and yield, by choosing the appropriate parameters.

Keywords: protein purification; chromatography; selection; modelling; optimization

INTRODUCTION

Until now, it has been virtually impossible to select separation and purification operations for proteins either for therapeutic or analytical application in a rational manner due to a lack of fundamental knowledge on the molecular properties of the materials to be separated and also the lack of an efficient system to organize such information. A range of techniques is available such as anion and cation exchange, which can be carried out at different pHs, hydrophobic interaction chromatography, gel filtration and affinity chromatography in addition to high performance liquid chromatography (HPLC) and aqueous two-phase partitioning. Evidently, when we are confronted with a complex mixture of partially unknown proteins or a clarified cell extract, there are many different routes one can take in order to choose the minimum and most efficient number of purification steps to achieve a desired level of purity.

Once the type of chromatography is chosen, optimization of the operating conditions is essential. In chromatography, protein adsorption depends on composition and concentration of the mixture and also on operation conditions such as flow rate, ionic strength gradient, sample load, physical properties of the

* *Correspondence to: J. A. Asenjo, Department of Chemical Engineering and Biotechnology, Centre for Biochemical Engineering and Biotechnology, Millennium Institute for Cell Dynamics and Biotechnology (ICDB), University of Chile, Beauchef 861, Santiago, Chile.*
E-mail: juasenjo@ing.uchile.cl

a *J. Asenjo, B. Andrews*
Department of Chemical Engineering and Biotechnology, Centre for Biochemical Engineering and Biotechnology, Millennium Institute for Cell Dynamics and Biotechnology (ICDB), University of Chile, Beauchef 861, Santiago, Chile

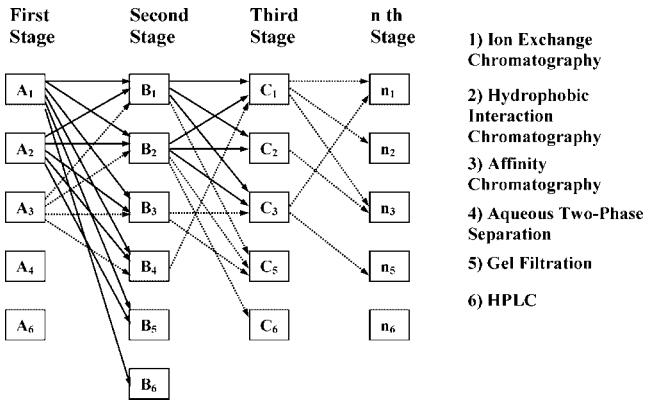

Figure 1. The combinatorial characteristic of choosing the sequence of operations for protein purification.

adsorbent matrix and column dimensions. Composition is usually determined in the production stage (fermentation, sometimes cell disruption and recovery) and thus for a given adsorbent matrix, operational conditions such as flow rate and ionic strength gradient have to be chosen in order to find a function able to represent the performance of the process. Maximization of this performance function can be carried out mathematically if a model able to simulate the chromatography is available. Mathematical models for describing a chromatographic separation can be classified, depending on the assumptions considered in its derivation. Models such as the Plate Model can be used for predicting the retention time and the elution curve. More complex models are those based on thermodynamic and transport phenomena; these models are called Rate Models.

For selecting the sequence of operations for high resolution purification of proteins, there are a large number of options that can be chosen almost in any random order as shown in Figure 1 (Leser and Asenjo, 1994). This review paper describes how to use the physicochemical data of the product protein and other proteins present (contaminants) to select an 'optimal' or suboptimal sequence of operations using the fewest number of operations and once the type of chromatography is chosen on how optimization of operating conditions is carried out using appropriate mathematical models.

SELECTION OF OPERATION

A clear rationale for selection of high resolution purification operations has been developed (Asenjo and Andrews, 2004). It characterizes the ability of the separation operation to separate one protein from another by using the theoretical concept of separation coefficients (Asenjo, 1990; Leser and Asenjo, 1992). It uses a relationship between the separation coefficient ($SC = DF \cdot \eta$) and the variables that show the performance in a separation process: the deviation factor (DF) for differences among physicochemical properties and the efficiency (η) of the process. DF has been defined as the difference in a particular physicochemical property (such as molecular weight, charge or hydrophobicity) between two proteins, which correspond to the target protein and the particular contaminant protein being considered (Asenjo and Andrews, 2004). To include the rule of thumb that reflects the logic of first separating impurities present in higher concentrations, a relative contaminant protein concentration (θ) and the separation selection coefficient (SSC)

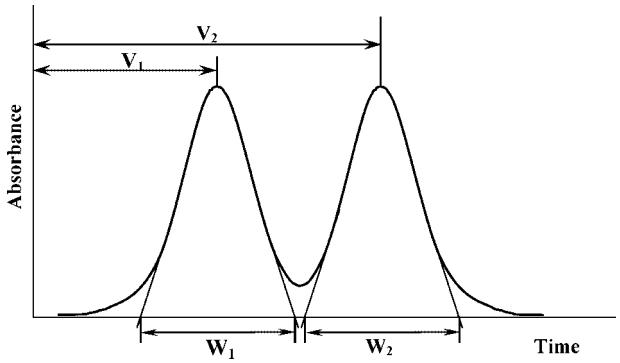

Figure 2. Determination of the Resolution between two peaks.

is defined as the product of the separation coefficient and this relative concentration ($SSC = SC\ \theta = DF \cdot \eta \cdot \theta$).

Resolution and efficiency

In chromatography, resolution is a variable used to measure column performance; it represents a means of interpreting column data and provides a basis for comparing results from different operating conditions (Leser *et al.*, 1996). Considering two peaks in a chromatogram, chromatographic resolution is defined as the distance between the peak maxima divided by the mean peak width, as shown in Figure 2 ($Rs = 2\ (V_2 - V_1)/(W_1 + W_2)$). As the separation coefficient is also a measure of the process performance, it can be assumed that it is proportional to the resolution ($SC\ \alpha\ Rs$) (Asenjo and Andrews, 2004). When an equal concentration of all proteins in a mixture is present, it can be shown that the efficiency of the process can be defined ($\eta = SC/DF\ \alpha\ Rs/DF$). This concept is not based on rate or equilibrium analysis but corresponds to a semiempirical analysis of separation of two components. Experimental data have shown a relatively constant behaviour of the efficiency for each particular separation process (Watanabe *et al.*, 1994; Leser *et al.*, 1996).

First algorithm: selection separation coefficients

A DF for each individual property such as charge, molecular weight and hydrophobicity of pairs of proteins was calculated. The efficiency (η) reflects the unequal ability of different separation processes (and/or different materials used) to exploit differences in the deviation factor to separate the proteins. This value is relatively constant for each type of separation and chromatographic material used and was found experimentally (Watanabe *et al.*, 1994; Leser *et al.*, 1996). The property chosen for gel filtration was molecular weight (mw), for hydrophobicity it was either the concentration of $(NH_4)_2SO_4$ (M) at which the protein eluted or the chromatographic KD value using a specific hydrophobic matrix and a decreasing gradient of $(NH_4)_2SO_4$. For ion-exchange chromatography (IEC), the property used to evaluate the deviation factor was charge as a function of pH and charge density. Clearly, charge density (charge/MW) gave a better correlation as a function of retention time in ion-exchange chromatography (Asenjo and Andrews, 2004).

To calculate the SSC, the system will read a database containing the information on the properties of the main contaminant proteins present in the specific expression system used. The data on the 13 predominant proteins present in a

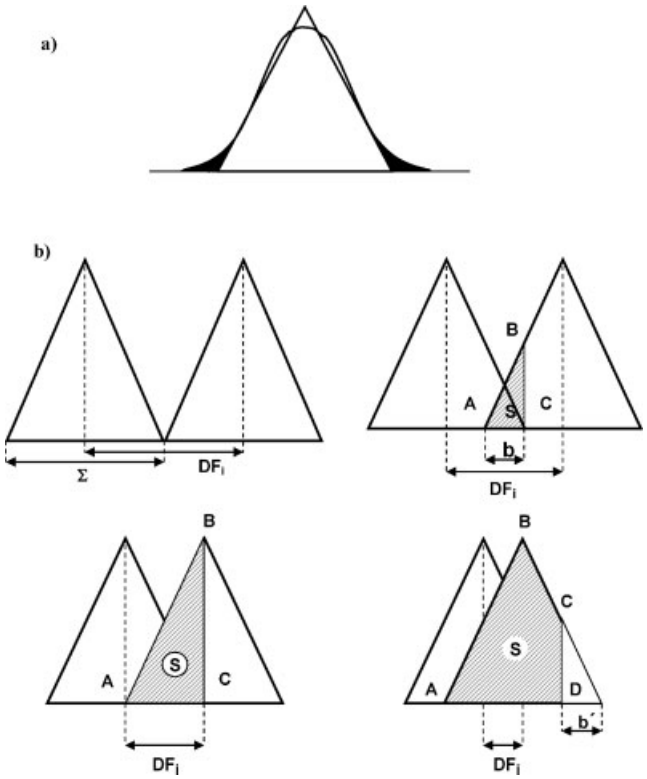

Figure 3. Representation of the peaks of a chromatogram as triangles. (a) Adjusting a triangle to a peak. (b) Variation in DF leads to different amounts of contaminant (triangle on the right) in the protein product (triangle on the left) ($\Sigma = 0.15$ for ion exchange, 0.22 for HIC and 0.46 for GF).

commercial strain of *Escherichia coli* used for producing recombinant proteins that was kindly donated by an industrial source (Chiron) was used in the Expert System (Woolston, 1994; Asenjo and Andrews, 2004). This database was then used to select the first high resolution purification step. Thus, the first algorithm that allows to choose a separation step would calculate the value of the SSC for all main contaminant proteins (e.g. 5, 10, 15) and all properties (e.g. 7–12) and choose the highest value of the SSC to select the best separation process.

Table 1. Values of Σ for the chromatographic processes used in the system

Chromatographic process	Σ
Size exclusion	0.46
Hydrophobic interaction	0.22
Ion exchange	0.15

Second algorithm: elimination of protein contaminants; a dynamic database

After each high resolution step, the concentration of contaminant proteins decreases and the number of steps has to be sufficient to eliminate contaminants until the product reaches the desired level of purity. In order to find the new concentration of all proteins after each separation step, a simple algorithm was developed based on the behaviour of a chromatographic separation that will give a very approximate value of the concentration of each of the contaminants after each separation step.

The amount of a protein contaminant eliminated after a chromatographic step is graphically shown in Figure 3 for three different situations. Figure 3a shows how the representation of the chromatographic peaks was simplified to a triangle. In Figure 3b the protein product corresponds to the triangle on the left and the contaminant to the one on the right. The shaded area (S = ABC or ABCD) corresponds to the amount of contaminant left with the product in each case (Leser *et al*., 1996). The variable Σ, shown in Table 1, corresponds to the peak width and has been experimentally determined (Lienqueo *et al*., 1996). It was also found that under the conditions normally used for protein purification, which are well below saturation, the value of Σ is virtually independent of protein concentration. Table 2 shows

Table 2. Concentration and relative concentration (%) of the main contaminants present in *Escherichia coli* and a model protein, showing how these values evolve during a consultation

	Loading		After the first step		After the second step		After the third step	
	Weight 0 (g/L)	Conc 0 (%)	Weight 1 (g/L)	Conc 1 (%)	Weight 2 (g/L)	Conc 2 (%)	Weight 3 (g/L)	Conc 3 (%)
Cont 1	11.24	14.97	0.22	1.62	0.00	0.00	0.00	0.00
Cont 2	7.06	9.40	0.06	0.44	0.00	0.00	0.00	0.00
Cont 3	4.63	6.17	0.24	1.76	0.01	0.19	0.01	0.20
Cont 4	5.58	7.43	0.11	0.81	0.00	0.00	0.00	0.00
Cont 5	4.83	6.43	0.09	0.66	0.00	0.00	0.00	0.00
Cont 6	2.48	3.30	0.04	0.29	0.00	0.00	0.00	0.00
Cont 7	7.70	10.25	0.02	0.11	0.02	0.39	0.00	0.00
Cont 8	6.80	9.05	0.13	0.95	0.00	0.00	0.00	0.00
Cont 9	7.53	10.03	7.56	55.51	0.15	2.89	0.00	0.00
Cont 10	6.05	8.06	0.12	0.88	0.01	0.19	0.00	0.00
Cont 11	3.89	5.18	0.00	0.00	0.00	0.00	0.00	0.00
Cont 12	1.48	1.97	0.02	0.15	0.00	0.00	0.00	0.00
Cont 13	0.83	1.11	0.01	0.07	0.00	0.00	0.00	0.00
Product	5.00	6.66	5.00	36.71	5.00	96.34	5.00	99.80

Table 3. Sequence suggested by the Expert System to obtain a purity superior to 94% in the purification

SSC Criterion chromatography steps	Purity	Purity criterion chromatography steps	Purity
Cation exchange at pH 6.0	33.1%	Anion exchange at pH 7.0	63.7%
Hydrophobic interaction	49.5%	Hydrophobic interaction	94.5%
Anion exchange at pH 7.0	97.0%		

how the concentration in grams per litre and the relative concentration (%) of the protein product (purity) and of the main contaminants present in *E. coli* change during a consultation with the Expert System (Leser, 1996).

In order to consider affinity chromatography as a viable separation, as in many cases suitable affinity ligands for the protein product are well known, it was assumed that if this technique is chosen by the user, all contaminants will be reduced by a fixed percentage (e.g. 90%) in the affinity separation step (Leser, 1996). However, since affinity chromatography will have to be analysed on a case-by-case basis, given the nature of the different ligands that can be used (e.g. metal ions in IMAC, dye or other), it was not included in the Expert System described here.

Robustness and sensitivity

A consultation was carried out using the Expert System to find all the steps necessary to achieve the desired level of purity (e.g. 98%) for the purification of the protein somatotropin produced in *E. coli*. Once a process was found, the original databases were randomly varied at the levels of 10% and 20% to see the effect on the process proposed, in terms of its robustness and sensitivity, by the system (Asenjo and Andrews, 2004).

The sensitivity of the proposed process to random changes in the values determined experimentally was investigated in order to assess the robustness of the system to either variations in the properties of the contaminant proteins present in the *E. coli* cells used or in the experimental measurements. When only the *E. coli* data or both sets of data (*E. coli* and protein product) were randomly varied at the level of 10% the sequence of operations was exactly the same. On the other hand, when the data was varied at the level of 20%, the sequence changed. This clearly shows that the system has the necessary robustness to variations and possible errors in the experimental determination of the data (<10%) but is sensitive enough to larger variations in protein properties (>20%) (Asenjo and Andrews, 2004).

Experimental tests: purity criterion

Considering that a key value in protein purification is the purity level after each step and that now an algorithm has been developed to calculate the purity after a purification step, this criterion was also implemented as a possible selection criterion in addition to the SSC. The purity criterion compares the purity level of the protein product obtained after a particular chromatographic technique has been applied.

Purity is defined as

$$\text{Purity} = \frac{\text{Concentration of the target protein}}{\Sigma\ \text{Concentration of all the proteins present}} \quad (1)$$

After determining which chromatographic technique gives the highest purity level, the system chooses this as the technique to use at this step. It then compares the purity with that required. A sequence of steps is chosen until the required level of final purity is reached. Finally, the system creates a list with the defined sequence of operations.

Two examples have been tested experimentally: a model protein mixture and a recombinant β-1,3-glucanase from *Bacillus subtilis* culture (Lienqueo *et al.*, 1999). The model purification mixture consisted of the purification of BSA (product) from three contaminants (Soybean Trypsin Inhibitor (SBTI), Ovalbumin and Thaumatin). The results obtained for a target of 94% purity are shown in Table 3. The SSC criterion selects a purification sequence based on the elimination of the contaminant and

Table 4. Physicochemical properties and concentration for the main proteins in *Bacillus subtilis* ToC46(pFF1)

Proteins	Initial concentration (mg cm^{-3})	Molecular weight (Da)	Hydrophobicity [$(NH_4)_2SO_4$]	Charge (Coulomb molecule^{-1}) 10^{-25} pH 4.0	pH 5.0	pH 6.0	pH 7.0	pH 8.0
β-1,3-glucanase	0.60	31 000	0.00	1.46	−0.62	−1.02	−2.33	−2.52
Contaminants:								
Low hydrophobic:								
Contaminant 1	2.74	41 000	1.50		0.26	−0.87	−1.65	−2.04
Contaminant 2	2.74	32 000	1.50		0.00	−2.70	−3.51	−3.51
Medium hydrophobic:								
Contaminant 3	0.25	35 500	0.20		−0.55	−0.22	−0.73	−1.82
High hydrophobic:								
Contaminant 4	0.42	62 500	0.00		−1.06	−1.17	−2.79	−3.32
Contaminant 5	0.25	40 600	0.00		−0.55	−0.22	−0.73	−1.82
Contaminant 6	0.25	69 600	0.00		−0.55	−0.22	−0.73	−1.82
Contaminant 7	0.09	40 600	0.00		1.46	−0.47	−1.06	−1.04
Contaminant 8	0.09	69 600	0.00		1.46	−0.47	−1.06	−1.04

Table 5. Sequence suggested by the Expert System for both criteria

Both criteria chromatography steps	Purity	Experimental validation chromatography steps	Purity
Hydrophobic interaction	32.7%	Hydrophobic interaction	33%–38%
Anion exchange at pH 6.5	70.3%	Anion exchange at pH 6.5	65%–70%

property that gives the highest SSC value. On the other hand, the purity criterion chooses the optimum chromatographic step considering all contaminants present. For this reason, in this particular case, the purity criterion gave a better result (only two steps). In this case, when the SSC criterion is used, the first separation step chosen is determined by the protein thaumatin (Asenjo and Andrews, 2004). Since it has a very high pI (>8.0), in this step no other protein is even partially purified. The second step takes care of ovalbumin and the third one of SBTI. On the other hand, when the purity criterion is followed, the first step (anion exchange at pH 7.0) takes care of ovalbumin as well as thaumatin. This discrepancy between both criteria does not happen often. In this particular case, the disagreement was produced because of the extremely high pI of one of the contaminant proteins and partly also because all proteins were present in the same concentration, which does not often correspond to a 'real' situation of contaminants.

Purification of β-1,3-glucanase and experimental investigation

Assessments were done using both the SSC criterion and the purity criterion implemented in the Expert System for purification of a β-1,3-glucanase from *B. subtilis* ToC46 (pFFI) culture. In this case, both criteria gave exactly the same sequence. The data on this system are given in Table 4. The results obtained for 70% purity are shown in Table 5 (Lienqueo *et al.*, 1999).

The chromatograms from the purification sequence are shown in Figure 4a,b and Table 5. Figure 4a shows the separation of 'low hydrophobicity proteins' (contaminants 1 and 2) and part of the 'medium hydrophobicity proteins' (contaminant 3) from the β-1,3-glucanase. In this first step, the main contaminants were eliminated. Figure 4b shows the separation of contaminants 3–4, 5–6 and 7–8 from the β-1,3-glucanase. Figure 4 shows that the scheme for purification suggested by the Expert System is valid for purification of this recombinant β-1,3-glucanase.

MODELLING AND SIMULATION

Once the type of chromatography is chosen, optimization of the operating conditions is essential. In chromatography, protein adsorption depends on composition and concentration of the mixture and also on operation conditions such as flow rate, ionic strength gradient, sample load, physical properties of the adsorbent matrix and column dimensions. Mathematical models for describing a chromatographic separation have been discussed in the Introduction. Two such models are the Plate Model and the more fundamentally based Rate Model.

Plate and rate models

The Plate Model is based on the plate theory. Briefly, the model assumes that the chromatographic column is formed by a number of plates (*N*p), each of them having the same ratio between the stationary phase volume and the volume of the mobile phase (*H*). For a defined column geometry and if the adsorption kinetics is known, the problem is reduced to solving

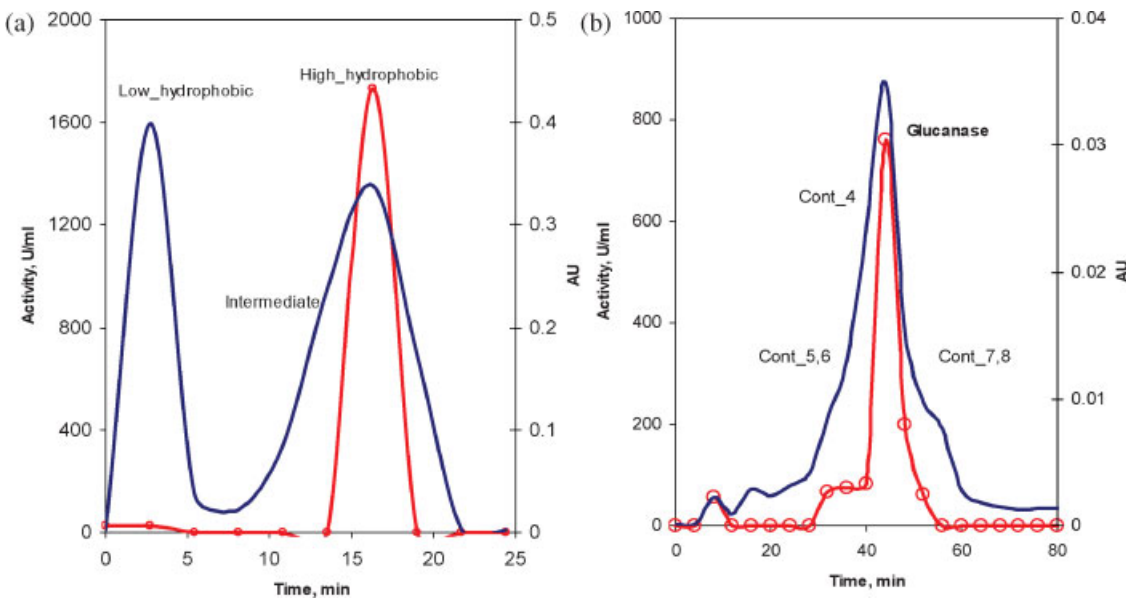

Figure 4. Steps suggested by both criteria for the purification of glucanase (o): glucanase activity. (a) First step suggested: hydrophobic interaction chromatography. (b) Second step suggested: anion-exchange chromatography at pH 6.5.

Table 6. Equations of the Plate Model

Protein at each plate ($i = 1,..N_p$)

$$\frac{dC_i}{d\theta} = \frac{N_p(C_{i-1} - C_i) - C_i H \frac{dK(C_i,I)}{dI}\frac{dI}{d\theta}}{1 + H[K(C_i,I) + C_i \frac{d}{dC_i}K(C_i,I)]}$$

$$\theta = -\frac{V_m}{V_0} \qquad C_0 = C_1 = \ldots = C_{Np} = 0$$

$$-\frac{V_m}{V_0} < \theta \leq 0 \qquad C_0 = 1$$

$$\theta > 0 \qquad C_0 = 0$$

Ionic strength at each plate ($i = 1,..N_p$)

$$I_i = \begin{cases} I_0 & \text{for} \quad \theta \leq (1 + HK_{salt})\frac{i}{N_p} \\ I_0 + G\left[\theta - (1 + HK_{salt})\frac{i}{Np}\right] & \text{for} \quad \theta > (1 + HK_{salt})\frac{i}{N_p} \end{cases}$$

the system of *N*p ordinary differential equations (ODE) shown in Table 6. In order to solve this ODE system, the ionic strength at each plate ($i = 1 \ldots Np$) has to be computed as a function time. Table 6 shows the formula for computing this variable in the case that a constant ionic strength gradient is applied for protein elution (Yamamoto *et al*., 1983).

At low protein concentration, the adsorption kinetics is computed from the value of the distribution coefficient (*K*) that depends on the ionic strength of the mobile phase (*I*). Protein displacement in ion-exchange chromatography is due to changes in the ionic strength of the mobile phase and thus the distribution coefficient and the number of plates cannot be computed from the first and second normalized central moment of the elution curve. However, because during the travelling of the protein through most part of the column, the protein zone is subject to an ionic strength near to the one at which this emerges from the column (Imax), a relationship has been presented for computing the number of plates (Yamamoto *et al*., 1983, Shene *et al*., 2006):

In the more fundamentally based Rate Model the dimensionless elution curves are obtained from the solution of the following partial differential equation:

$$\frac{\partial c_b}{\partial \tau} = -\frac{\partial c_b}{\partial z} + \frac{1}{Pe_L}\frac{\partial^2 c_b}{\partial z^2} - \xi_i(c_b - c_{p,r=1}) \tag{2}$$

subject to the initial and boundary conditions given by

$$\begin{aligned} &\tau = 0 \quad c_b = c_b(0,z) \\ &z = 0 \quad \frac{\partial c_b}{\partial z} = Pe_L\left[c_b(0,\tau) - \frac{C_f(\tau)}{C_o}\right] \\ &z = L \quad \frac{\partial c_b}{\partial z} = 0 \end{aligned} \tag{3}$$

In order to solve the partial differential in Equation 2, the dimensionless concentration profile for each component in the liquid phase contained inside the particles, c_p, has to be computed. These concentration profiles are obtained from the solution of the following partial differential equation:

$$\frac{\partial}{\partial \tau}(\varepsilon_p c_p + (1 - \varepsilon_p)c_p^*) = \eta \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial c_p}{\partial r}\right) \tag{4}$$

subject to the initial and boundary conditions given by

$$\begin{aligned} &\tau = 0 \quad c_p = c_p(0, r, z) \\ &r = 0 \quad \frac{\partial c_p}{\partial r} = 0 \\ &r = 1 \quad \frac{\partial c_p}{\partial r} = Bi[c_b(z,\tau) - c_p(r = 1, z, \tau)] \end{aligned} \tag{5}$$

In relationship (3), $Cf(\tau)$, the time-dependent feeding concentration (for a protein, $Cf(\tau)$, will be different from zero while the sample is loaded into the column; for the displacer, the feeding concentration is often a function of time). Dimensionless variables and parameters in relationships (2)–(5) are shown in Table 7. Since all mass transfer phenomena are taken into account in partial differential Equations 2 and 4, Rate Models can be used for testing different chromatographic conditions (Gu, 1995; Lazo, 1999).

Operational conditions in chromatography such as flow rate and ionic strength gradient are taken into account in both mathematical models and thus predicted elution curves depend on them. However, the quality of the product obtained in chromatography is also dependent on external operational conditions such as the flow rate as well as the size of the fraction of the protein product collected, as shown in Figure 5 (also called peak cutting). From the scheme presented in Figure 5, if the target protein is protein A, the outlet flow can be collected from $t \geq t_i$ until $t \leq t_e$, the period during which the concentration of A in the outlet flow becomes important. However, during this time, a part of the contaminants is also eluted. A way to minimize the contaminant content in the collected volume is by decreasing the collecting interval considering the time elapsed between t_1 and t_2 (Figure 5).

In order to define a performance function for a chromatography that can be used for choosing operational conditions for the separation of a given protein mixture, several parameters should be taken into account:

(a) *Concentration* of the target protein, x_A which is given by:

$$x_A = \frac{\int_{t_1}^{t_2} C_A \cdot F dt}{\int_{t_1}^{t_2} F dt} = \frac{\int_{t_1}^{t_2} C_A dt}{\int_{t_2}^{t_1} dt} \tag{6}$$

As shown in Figure 5 concentration x_A depends on the collecting time and on the resolution of the purification stage

Table 7. Dimensionless variables and parameters of the Rate Model

Concentration of the mobile phase	$c_b = \frac{C_b}{C_b}$
Concentration of the liquid inside the adsorbent particles	$c_p = \frac{C_p}{C_p}$
Concentration of the adsorbed protein	$c_p^* = \frac{C_p^*}{C_o}$
Dimensionless time	$\tau = \frac{vt}{L}$
Dimensionless position in the column	$z = \frac{Z}{L}$
Dimensionless position in the particle	$r = \frac{R}{R_p}$
Peclet number	$Pe_L = \frac{v \cdot L}{D_z}$
Biot number	$Bi = \frac{k \cdot R_p}{\varepsilon_p \cdot D_p}$
	$\eta = \frac{\varepsilon_p \cdot D_p \cdot L}{R_p \cdot 2 \cdot v}$
	$\xi = \frac{3 \cdot Bi \cdot \eta \cdot (1 - \varepsilon_b)}{\varepsilon_b}$

fixed by the flow rate and the ionic strength gradient. Costs involved in the afterward concentration processes (ultrafiltration, lyophilisation) are related to the value of x_A.

(b) *Purity* of protein (A) is defined as the ratio between the mass of protein A and that of all the proteins in the collected volume:

$$\text{Purity of the } i \text{ component} = \frac{x_i}{\sum_{j=1}^{m} x_j} \tag{7}$$

Purity not only allows to establish the relevance of the chromatography as a purification stage but it can also be used for estimating costs involved in the rest of the purification steps.

(c) *Yield* is defined as the ratio between the mass of the target protein in the collected volume and the mass of the same protein loaded into the column:

$$yield_A = \frac{\int_{t_1}^{t_2} C_A \cdot F dt}{C_{A0} V_0} \tag{8}$$

The yield of a chromatography depends on the collecting time and it can be used to estimate costs of the production stages (fermentation).

(d) *Process time*, is defined as the time at which all the proteins in the mixture loaded into the column are eluted and after which the column can be prepared for the treatment of a new load. Process time can be used as an estimation of the costs involved in the chromatography stage.

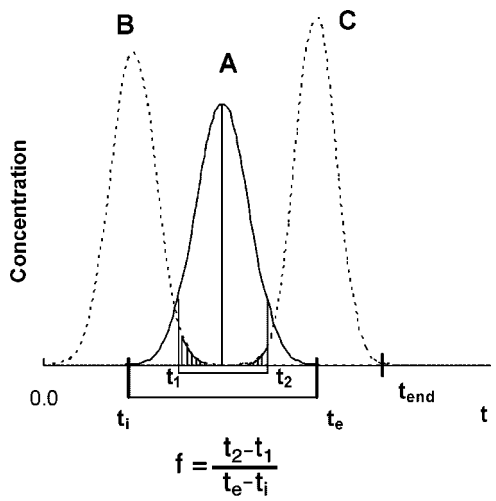

Figure 5. Elution curves of a mixture of proteins. *A* is the target protein product and *f* the fraction of peak *A* collected.

Plate and Rate Models have been used for simulating elution curves of a three-protein mixture in IEC carried out under different operational conditions (flow rates and ionic strength gradients). Q Sepharose FF was used as the adsorbent matrix. A cost function for the protein production process was proposed and flow rate, ionic strength gradient and collection time are selected in order to minimize the cost function for different type of protein products.

Simulation

Elution curves of the three-protein mixture in IEC were experimentally recorded for two values of the ionic strength gradient (*g*) and different flow rates (*F*). Values for the purity and retention times obtained from the IEC elution curves simulated using the Plate and Rate Models are shown in Tables 8 and 9, respectively.

Comparisons between experimental and simulated elution curves and the ionic strength profile computed using the Plate Model for the different flow rates and an ionic strength gradient of 0.055 M/mL is shown in Figure 6. The deviation between the experimental and simulated elution curves is presented in

Table 8. Results of the simulations for the separation of a three-protein mixture in IEC using the Plate Model for different flow rates, *F*, and ionic strength gradients, *g*

			Purity (%)			Retention time (min)			
Run	F (mL/min)	g (M/mL)	P_1	P_2	P_3	P_1	P_2	P_3	Deviation
1-a	0.3	0.100	54.4	49.4	80.6	12.2	13.3	16.5	0.0149
1-b	0.7	0.100	50.9	42.6	54.8	5.3	5.8	7.1	0.0093
1-c	1.0	0.100	50.3	39.8	47.1	3.7	4.1	5.0	0.0109
2-a	0.3	0.055	63.6	54.0	84.6	17.4	19.8	24.9	0.0189
2-b	0.5	0.055	59.9	48.1	71.3	10.5	11.9	15.0	0.0053
2-c	1.0	0.055	55.2	42.0	51.9	5.2	6.0	7.5	0.0064

P_1, α-lactoalbumin; P_2, ovalbumin; P_3, β-lactoglobulin.

Table 9. Results of the simulations for the separation of a three-protein mixture in IEC using the Rate Model for different flow rates, *F*, and ionic strength gradients, *g*

			Purity (%)			Retention time (min)			
Run	F (mL/min)	g (M/mL)	P_1	P_2	P_3	P_1	P_2	P_3	Deviation
1-a	0.3	0.100	64.3	63.1	78.7	12.5	13.8	16.6	0.0192
1-b	0.7	0.100	55.5	55.5	59.6	5.3	5.9	7.1	0.0111
1-c	1.0	0.100	53.0	52.2	74.3	3.7	4.1	5.0	0.0152
2-a	0.3	0.055	78.3	47.5	67.9	17.0	19.6	24.4	0.0252
2-b	0.5	0.055	65.3	55.2	58.0	10.2	11.8	14.6	0.0110
2-c	1.0	0.055	56.8	48.5	60.4	5.0	5.9	7.3	0.0059

P_1, α-lactoalbumin; P_2, ovalbumin; P_3, β-lactoglobulin.

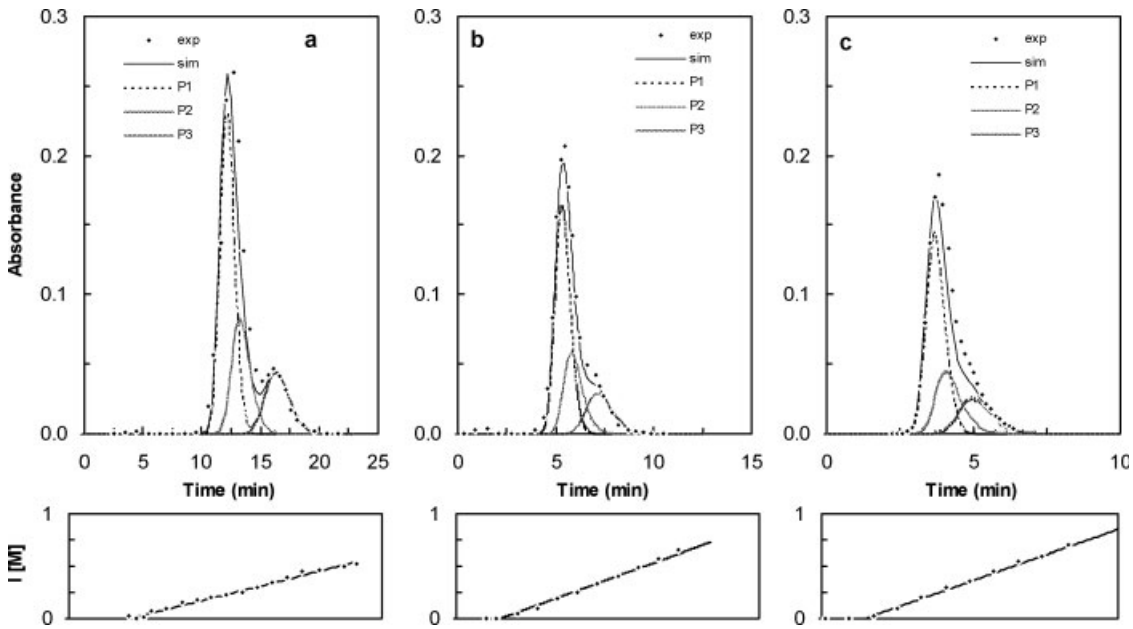

Figure 6. Experimental and simulated elution curves of the three protein mixture using the Plate Model (P_1: α-lactoalbumin; P_2: ovalbumin; P_3: β-lactoglobulin) for an ionic strength gradient of 0.055 M/mL and different flow rates. (a) 0.3 mL/min $N_{P1} = 18$; $N_{P2} = 6$; $N_{P3} = 6$ (b) 0.7 mL/min $N_{P1} = 11$; $N_{P2} = 4$; $N_{P3} = 4$ (c) 1 mL/min $N_{P1} = 6$; $N_{P2} = 2$; $N_{P3} = 2$.

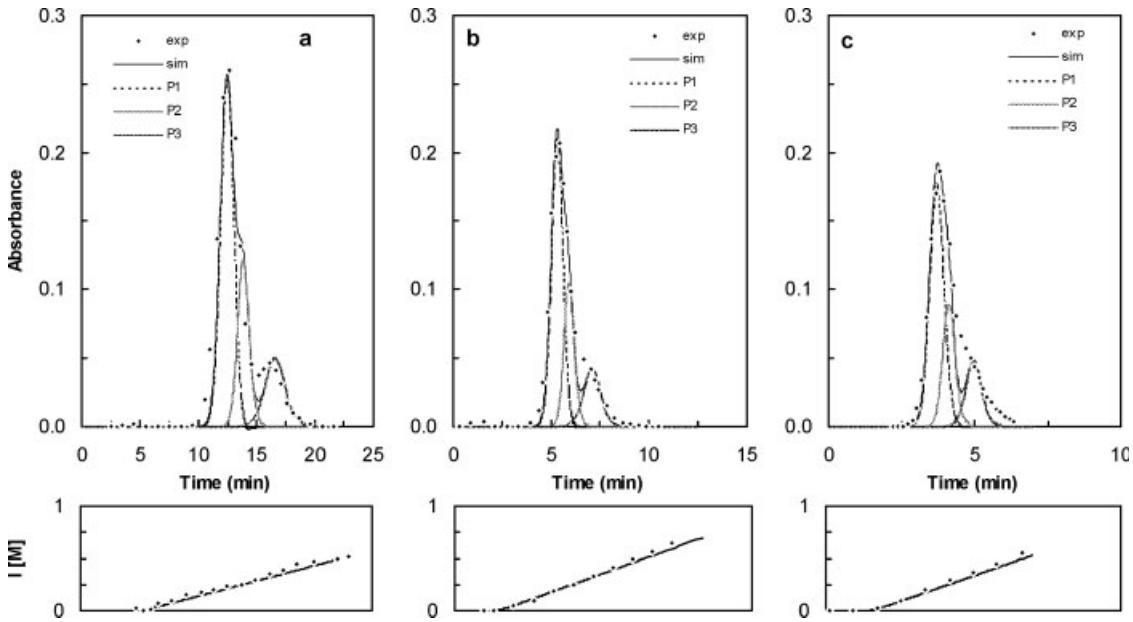

Figure 7. Experimental and simulated elution curves of the three-protein mixture using the Rate Model (P_1: α-lactoalbumin; P_2: ovalbumin; P_3: β-lactoglobulin) obtained for an ionic strength gradient of 0.1 M/mL and different flow rates. (a) 0.3 mL/min $\eta_{P1} = 10$; $\eta_{P2} = 9.5$; $\eta_{P3} = 4.5$ (b) 0.7 mL/min $\eta_{P1} = 6$; $\eta_{P2} = 6$; $\eta_{P3} = 3$ (c) 1 mL/min $\eta_{P1} = 4$; $\eta_{P2} = 4$; $\eta_{P3} = 4$.

Table 8. Maximum value for the deviation was 0.0189 (absorbance units).

Results in Figure 6 and Table 8 indicate that the Plate Model can be used for simulating the elution curve of a protein mixture in IEC. It is important to note that all parameters in the model such as those needed for computing the number of plates and the distribution coefficients for each protein were obtained form independent experiments.

The comparison of the experimental and simulated elution curves and ionic strength profile computed using the Rate Model is shown in Figure 7. In these simulations the adsorption kinetics and parameters were used (Shene *et al*., 2006). Table 9 shows the deviation between simulated and computed values; the maximum deviation was 0.0252 (absorbance units). From the comparison of the results presented in Tables 8 and 9, the average prediction deviation obtained with the Rate Model was slightly higher than that obtained using the Plate Model.

The Rate Model has several advantages over the Plate Model, the most important being that it can be extended for simulating elution curves of more concentrated protein mixtures, where protein interaction effects could be significant and more complex relationships for the adsorption kinetics must be used. Nevertheless, for the case under study the Plate Model is easier to implement computationally and also has a lower CPU demand due to the small size of the ODE system that has to be solved.

OPTIMAL SELECTION OF OPERATING CONDITIONS

Results in Tables 8 and 9 show that flow rate and the ionic strength gradient affect the purity and retention time of the different proteins in the mixture. A higher purity is obtained by using a small ionic strength gradient for a given flow rate. However, when a small value of the ionic strength gradient is chosen, the peak width increases, the maximum protein concentration decreases and the retention time increases. Since IEC is in many cases one step in a protein purification process, its output will affect other steps. The best way of relating how the results obtained in one chromatography step (process time, concentration of the target protein in the collected volume and purity and yield of the target protein) affect other stages in the purification process is through a cost function since in many cases the value of the product is fixed by the market and thus an important way to increase the profit is through the reduction of the processing costs. A cost function for a protein production process and how the chromatography performance affects it, similar to that proposed previously (Huenupi *et al*., 1999) for a protein extraction process, can be defined as follows:

$$\text{Cost} = a_1 \frac{B_1}{\text{Yield}} + a_2 \left(1 - \frac{\text{Purity}}{B_2}\right) + a_3 \frac{B_3}{C_A} + a_4 \frac{t_{\text{end}}}{B_4} \qquad (9)$$

The first term in Equation 9 takes into account costs involved in the fermentation in such a way that the chromatography's step yield decreases and more protein mixture will be needed to fulfill the required production level. The second term represents costs involved in further purification steps, for instance, hydrophobic interaction chromatography (Asenjo and Andrews, 2004). As the purity of the product eluted from the chromatography increases, these costs decrease and they become equal to zero in the case where the separation is accomplished in this stage only. Costs related to concentration processes, such as dialysis, ultrafiltration or freeze–drying, are inversely related to the concentration of the product obtained in the chromatography, which corresponds to the third term in Equation 9. The last term in Equation 9 takes into account the costs of the chromatography determined by the processing time. A longer processing time may result in higher resolution but this will determine the use of a larger unit or more than one unit in parallel, thus increasing the cost for the specified

production level. Values of coefficients a_1, a_2, a_3 and a_4 in relationship (9) give the relative weights to the different terms in the cost function; the sum of these coefficients is constrained to be 1. Parameters B_1–B_4 in relationship (9) are introduced in order to scale the different variables. Values for these parameters will depend on the system geometry and the range of operational conditions that can be used in a given system. Relationship (9) states that costs of the different stages, given by the different terms, are linearly related to the variables. However, scaling indexes similar to those used for equipment scale-up (exponents in the different terms) can be introduced in order to build a more rigorous model (Huenupi *et al*., 1999).

The cost function in relationship (9) was evaluated considering the case in which the target protein is ovalbumin, a protein whose retention time was found to be between those of the other two proteins in the mixture, as a way to consider the worst case in a given protein purification process. It was assumed that flow rate and ionic strength gradients are constrained to take values between 0.3 and 1.0 mL/min and 0.055 and 0.105 M/mL, respectively. From the scheme shown in Figure 5, the fraction (f) of the peak collected is given by

$$f = \frac{t2 - t1}{te - ti} \tag{10}$$

For the range of operational conditions, tested values for parameters B_1, B_2, B_3 and B_4, in relationship (9) were chosen so that each term in expression (9) could reach a maximum value of 1. Hence B_1, B_2, B_3 and B_4 were 0.72, 77%, 0.0345 g/L and 28.54 min, respectively.

Values of coefficients a_1, a_2, a_3 and a_4 depend on the characteristics of the target protein such as the required final purity and its synthesis during the fermentation. Costs involved in a fermentation process for protein synthesis can be assumed to represent between 30 and 70% of total production cost (Huenupi *et al*., 1999). Purification costs ($a_2 + a_4$) can represent between 10–50% of a protein production process. Costs for the concentration stages (a_3) can be considered lower than those involved in the purification stages (between 10 and 30%).

Selection of peak size (peak cutting)

Simulations were carried out using different values of the flow rate and ionic strength gradients and the cost function was evaluated for different values of f [fraction of product peak collected (Eq. 10)]. Three different combinations for the a_i ($i = 1,..,4$) coefficients were considered in order to simulate conditions for the production of different types of target proteins. Table 10 shows the flow rate, ionic strength gradient and the value for f found, for which the minimum value of each of the cost functions was obtained.

In Case 1, the operational conditions for the chromatography in a process in which costs for the fermentation process, subsequent purification steps, and the chromatography contribute in the same degree to the total production costs ($a_i = a_2 = a_4 = 0.30$) are presented (0.30). This could be the case of an enzyme required with a low purity and for which purification is carried out in order to eliminate contaminants that decrease its activity, for instance an industrial enzyme. For this case flow rate, ionic strength gradient and fraction collected were equal to 0.6 L/min, 0.105 M/mL and 0.5, respectively. As expected the final purity is relatively low, 53.1%, and a high yield is obtained.

Table 10. Operational conditions, protein yield and purity in a chromatography separation for minimum production cost based on the cost function given by Equation 9 and different relative weights (a_i) for the different production stages

	Case 1	Case 2	Case 3
a_1	0.30	0.55	0.30
a_2	0.30	0.20	0.50
a_3	0.10	0.05	0.05
a_4	0.30	0.20	0.15
Min Cost	0.593	0.651	0.510
	Operational conditions*		
F (mL/min)	0.6	0.6	0.3
g (M/mL)	0.105	0.105	0.065
f (—)	0.50	0.75	0.45
	Results		
Yield	0.8195	0.9455	0.7605
Purity (%)	53.10	47.32	73.67
C_A (mg/mL)	0.0260	0.0200	0.0243
t_{end} (min)	9.82	9.82	25.34

*Operational conditions were constrained to take values of 0.3–1.0 mL/min for F, 0.055 –0.105 for g and 0.0–1.0 for f (with increments of 0.1, 0.005 and 0.05, respectively).

Case 2 (Table 10) shows the operational conditions in the chromatography for the production of a target protein having very high fermentation costs ($a_i = 0.55$). Subsequent purification stages ($a_2 = 0.2$) and those involved in the IEC ($a_4 = 0.2$) are of the same magnitude and lower than in the previous case. This could be the case of an intracellular target protein with low substrate to target protein yield. In this case costs for cell disruption and separation from cell debris are assumed to be included in those for the fermentation. Results indicate that during the IEC, a higher fraction of the volume should be collected ($f = 0.75$). This results in an even lower purity and higher yield than in the previous case.

Chromatography operational conditions for the production of a target protein required with a high final purity, for instance, a pharmaceutical product, are shown in Case 3 (Table 10). Costs involved in the subsequent purification stages are set as 50% of the total production costs ($a_2 = 0.5$). Operational conditions in the chromatography stage for this case are those of a product with a high purity (73.67%). In order to achieve the high purity, IEC must be carried out at low flow rate (0.3 mL/min) and with small ionic strength gradient (0.065 M/mL). As fermentation costs are relatively low, collection of the eluted protein corresponds to a smaller fraction of total elution time ($f = 0.45$) in order to obtain a high purity protein fraction. This results in a higher purity and lower yield than in the previous two cases and a longer processing time that results in a higher resolution. The actual fractions collected in all cases are clearly shown in Figure 8.

Use of mathematical models for process validation

A recent paper (Kaltenbrunner *et al*., 2007) shows how the Plate Model has been used for an industrial practical application of chromatographic theory for process characterization towards validation of an ion-exchange operation. When a chromatographic operation utilized to purify a human therapeutic protein

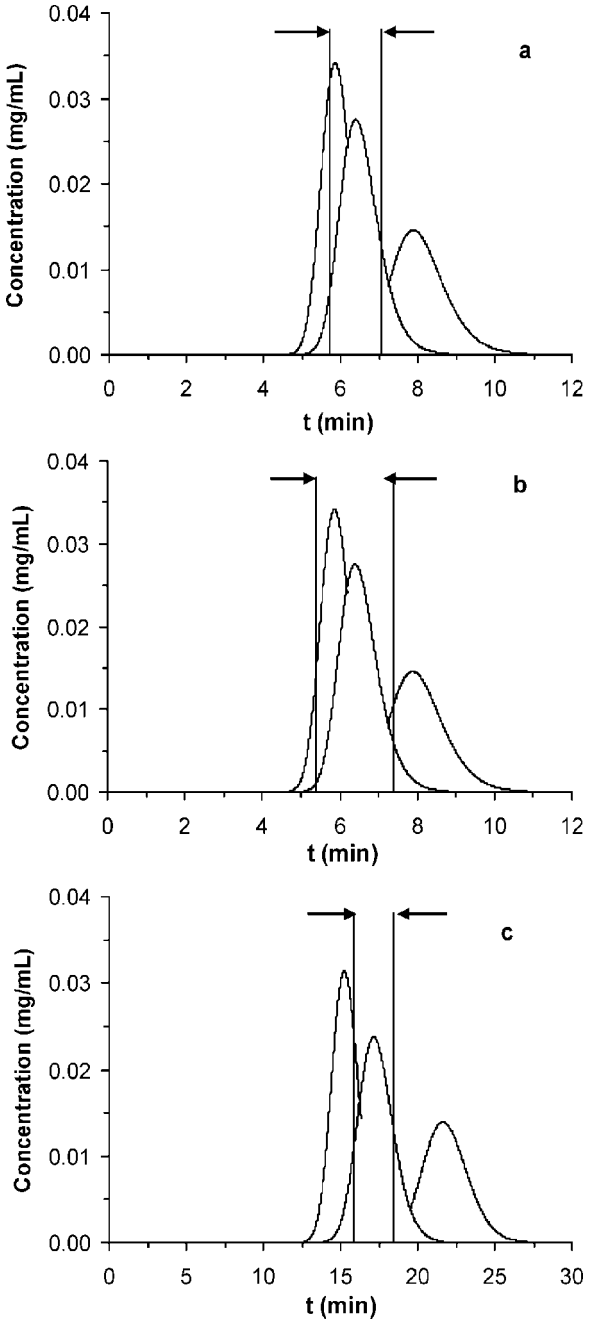

Figure 8. Elution curves of the mixture of proteins in a chromatography separation for the minimum production cost given in Table 8. Between the arrows, the fraction of protein product collected for (a) Case 1, (b) Case 2 and (c) Case 3. The three cases correspond to protein products with different characteristics as described in the text.

is prepared for validation before commercial production, numerous tests have to be performed to establish the relative importance of each operating parameter to define its future role and importance in the framework on in-process controls. This prioritization process is usually performed using an entirely empirical approach. The process flow chart from a decision to commercialize a biologic to product launch is shown schematically as a process flow-sheet in Figure 9. This paper demonstrates the application of a rational approach based on

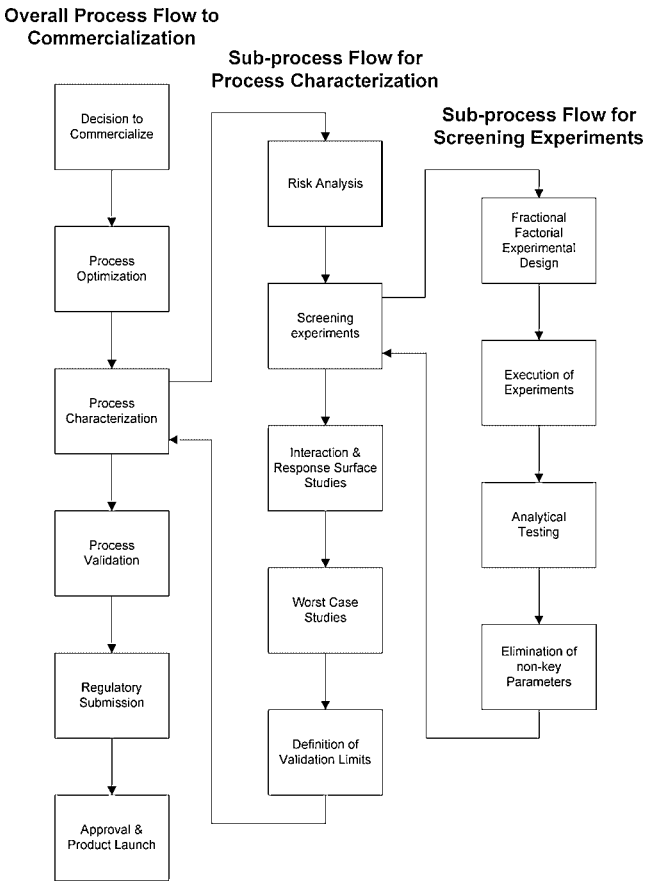

Figure 9. Process flow chart from a decision to commercialize a biologic to product launch. The sub-processes of process characterization and screening experiments are outlined (Kaltenbrunner *et al.*, 2007).

chromatographic theory to prioritize operating parameters. Both methodologies, empirical and rational, were performed to evaluate a specific ion-exchange chromatography operation for the preparative separation of closely related protein species. The paper shows that the application of the rational approach has the potential to accelerate the evaluation and significantly reduce the amount of analytical testing needed.

CONCLUSIONS

In this paper we have reviewed and discussed the proteomic approach to select a purification process for proteins based on physicochemical properties as well as the use of mathematical modelling for the optimization of performance (actual operating conditions) in a chromatographic step. The methodology described constitutes a rational proteomic procedure to separate the main contaminant proteins with a minimum number of steps and the optimization of such steps.

An algorithm to calculate the SSC, a parameter used to select the actual purification at each step was developed, and the translation of physicochemical data of the proteins to chromatographic behaviour was also carried out for ion-exchange chromatography, hydrophobic-interaction chromatography and gel filtration.

Another algorithm used to estimate concentration of each protein contaminant after a chromatographic process is performed was also developed. The methodology described, which was handled by a computer-based Expert System, was tested with recombinant proteins produced in *E. coli*, with a good database for the main protein contaminants and purification of a recombinant protein product with good results.

The system was robust to errors $<10\%$, which is the range that can be found in the experimental determination of the properties in the database of product and contaminants. On the other hand, the system was sensitive, both to larger variations ($>20\%$) in the properties of the contaminant database and the protein product.

The purification strategy proposed was experimentally tested and validated with a mixture of four proteins and the experimental validation was also carried out with a supernatant of *B. subtilis* producing a recombinant β-1,3-glucanase.

In addition to SSC, final purity can also be used as a selection criteria, given the fact that it is also calculated after each separation step is performed, to give the new protein contaminant concentrations in the database. Although both criteria, SSC and purity, will in most cases give similar results, purity may give fewer steps (and thus a better process) when concentrations of contaminant proteins are similar in the crude starting material.

Once the type of chromatography is chosen, optimization of the operating conditions is essential. Chromatographic elution curves for a three-protein mixture (α-lactoalbumin, ovalbumin and β-lactoglobulin), carried out under different flow rates and ionic strength conditions, were simulated using two different mathematical models. These models were the Plate Model and the more fundamentally based Rate Model. Simulated elution curves were compared with experimental data not used for parameter identification. Deviation between experimental data and the simulated curves using the Plate Model was less than 0.0189 (absorbance units); a slightly higher deviation [0.0252 (absorbance units)] was obtained when the Rate Model was used.

Simulation of IEC for protein purification can be used as a tool for choosing operational conditions such as flow rate, ionic strength gradient and the externally fixed operational condition that in this work was termed the collecting time (fraction collected, *f*). In order to do this, a performance function for ion-exchange chromatography has to be defined. However, since a purification stage such as ion-exchange chromatography is integrated into the protein production process, its performance is affected by previous and possibly subsequent processing and purification stages. In this work a cost function for the whole protein production process that can be used for the selection of the operational conditions as well as the fraction of the product to be collected (peak cutting) in chromatography was built and tested. This function can be used for protein products with different characteristics and qualities such as purity and yield by choosing the appropriate parameters.

Acknowledgements

The authors acknowledge support of the Millennium Scientific Initiative (project P05-001-F) and Fondef (project D04I1374).

REFERENCES

Asenjo JA. 1990. Selection of operations in separation processes. In *Separation Processes in Biotechnology*. Asenjo JA (ed.) Marcel Dekker. New York; 3–16.

Asenjo JA, Andrews BA. 2004. Is there a rational method to purify proteins? From expert systems to proteomics. *J. Mol. Recognit.* **17**: 236–247.

Gu T. 1995. Mathematical Modeling and Scale-Up of Liquid Chromatography. Springer: Berlin.

Huenupi E, Gómez A, Andrews BA, Asenjo JA. 1999. Optimization and design considerations of two-phase continuous protein separation. *J. Chem. Technol. Biotechnol.* **74**: 256–263.

Kaltenbrunner O, Giaverinni O, Woehle D, Asenjo JA. 2007. Application of chromatographic theory for process characterization towards validation of an ion-exchange operation. *Biotechnol. Bioeng.* **98**: 201–210.

Lazo C. 1999. Simulation of Liquid Chromatography and Simulated Moving Bed (SMB) Systems. Sudienarbeit Technische Universität: Hamburg-Harburg.

Leser EW. 1996. Prot_Ex: an expert system for selecting the sequence of processes for the downstream purification of proteins. Ph.D. Thesis, University of Reading.

Leser EW, Asenjo JA. 1992. The rational design of purification processes for recombinant proteins. *J. Chromatogr.* **584**: 35–42.

Leser EW, Asenjo JA. 1994. The rational selection of purification processes for proteins: an expert system for downstream processing design. *Ann. N.Y. Acad. Sci.* **721**: 337–347.

Leser EW, Lienqueo ME, Asenjo JA. 1996. Implementation in an expert system of selection rationale for purification processes for recombinant proteins. *Ann. N.Y. Acad. Sci.* **782**: 441–455.

Lienqueo ME, Leser EW, Asenjo JA. 1996. An expert system for the selection and synthesis of multistep protein separation processes. *Comput. Chem. Eng.* **20**: S189–S194.

Lienqueo ME, Salgado JC, Asenjo JA. 1999. An expert system for selection of protein purification processes: experimental validation. *J. Chem. Technol. Biotechnol.* **74**: 293–299.

Shene C, Andrews BA, Lucero A, Asenjo JA. 2006. Mathematical modelling of elution curves for a protein mixture in ion exchange chromatography and for the optimal selection of operational conditions. *Biotechnol. Bioeng.* **95**: 704–713.

Watanabe E, Tsoka S, Asenjo JA. 1994. Selection of chromatographic protein purification operations based on physicochemical properties. *Ann N.Y. Acad. Sci.* **721**: 348–364.

Woolston PW. 1994. A physicochemical database for an expert system for the selection of recombinant protein purification processes. Ph.D Thesis, University of Reading.

Yamamoto S, Nakanishi K, Matsuno R. 1983. Ion exchange chromatography of proteins predictions of elution curves and operational conditions. I. theoretical considerations. *Biotechnol. Bieng.* **25**: 1465–1483.