Optimal Synthesis of Protein Purification Processes

Elsa Vásquez-Alvarez,[†] Maria E. Lienqueo,[‡] and José M. Pinto*,[†]

Department of Chemical Engineering, University of São Paulo, São Paulo (SP), 05508-900, Brazil, and
Centre for Biochemical Engineering and Biotechnology, Department of Chemical Engineering,
University of Chile, Beauchef 861, Santiago, Chile

There has been an increasing interest in the development of systematic methods for the synthesis of purification steps for biotechnological products, which are often the most difficult and costly stages in a biochemical process. Chromatographic processes are extensively used in the purification of multicomponent biotechnological systems. One of the main challenges in the synthesis of purification processes is the appropriate selection and sequencing of chromatographic steps that are capable of producing the desired product at an acceptable cost and quality. This paper describes mathematical models and solution strategies based on mixed integer linear programming (MILP) for the synthesis of multistep purification processes. First, an optimization model is proposed that uses physicochemical data on a protein mixture, which contains the desired product, to select a sequence of operations with the minimum number of steps from a set of candidate chromatographic techniques that must achieve a specified purity level. Since several sequences that have the minimum number of steps may satisfy the purity level, it is possible to obtain the one that maximizes final purity. Then, a second model that may use the total number of steps obtained in the first model generates a solution with the maximum purity of the product. Whenever the sequence does not affect the final purity or more generally does not impact the objective function, alternative models that are of smaller size are developed for the optimal selection of steps. The models are tested in several examples, containing up to 13 contaminants and a set of 22 candidate high-resolution steps, generating sequences of six operations, and are compared to the current synthesis approaches.

1. Introduction

In every biotechnological process, two subprocesses can be considered after the bioreaction stage. Those are recovery and purification that are part of the so-called downstream processing. The recovery stage includes the preliminary operations such as broth separation, cell rupture, debris removal and refolding. The remaining product is virtually a mixture of proteins, including the target protein, which have different physicochemical properties. After the recovery subprocess the total protein concentration is 60−70 g/L (*1*). In the purification stage, all proteins are subject to a sequence of high-resolution operations (two or more) in order to attain a specified level of purity in terms of the desired product (ideally the yield is 100%). Wheelwright (*2*) and Asenjo (*3*) point to the need for the optimization of the downstream processing, since this stage represents the major manufacturing and investment costs in biochemical processes.

The final quality of biotechnological products is determined at the purification level, which may be regarded as the most important stage in the whole production process. This requirement is even more critical for therapeutic products such as vaccines and antibiotics that require an extremely high purity level.

Among the diverse forms of recovery and purification of a protein mixture, the most important group comprises chromatographic operations, also known as high-resolution steps. The specified purity is attained after several steps, and in each of them the mixture is basically split into two streams, one that contains the target protein and the other that is discarded. It is then fundamental to synthesize an optimal sequence of purification steps such that their total number is as small as possible.

Mathematical models of purification operations are important simulation and design tools that can be used individually or be a part of the whole bioprocess. Luong et al. (*4*) developed a mathematical model to describe the dynamic behavior of the ultrafiltration process for protein purification based on affinity interactions and membrane separations. Mao and Hearn (*5*) studied several alternative approaches for the optimal operation of ion-exchange and affinity chromatographic methods for protein purification. The chromatographic operation has been considered as a four-stage process. On the basis of experimental data and on the application of mathematical models, practical methods for the selection of operating parameters were derived. These have as main objective to simplify the determination of mass transfer parameters required for scale-up and process economic optimization. Computer-aided process design combined with experimental planning was used for determining the optimal conditions in ion-exchange chromatography (*6*). The optimal control problem considering the gradient elution chromatography (GEC) was used in purification of biochemical products. Using the control parametrization technique, the control variables are approximated

---

* Email: jompinto@usp.br.
[†] University of São Paulo.
[‡] University of Chile.

by piecewise linear functions, and consequently a sequence of nonstandard optimal parameter selection problems is obtained. Each of these approximate problems was then shown to be equivalent to a standard min-max optimization problem and solvable by the existing optimization software (*7*).

Optimization models that analyze the retention behavior in nonsuppressed ion chromatography of anions have been compared and assessed for determining the optimal eluent composition for a separation chromatography using phthalate eluents of varying concentration and pH values (*8*). Madden and Haddad (*9*) developed a similar study but focused on the behavior of mixtures of carbonates/hydrogencarbonates in suppressed ion chromatography. In total, seven models were tested ranging from one that is based on the dominant equilibrium approach and competing ion effective charge approach to an empirical model. These were analyzed over an extensive set of experimental retention data (24 anions).

Porter and Ladisch (*10*) developed a model for the purification process of α-galactosidase from soybean seeds in order to determine overall cost. The authors compare two sequences, one that uses affinity chromatography in the final two steps and another that uses affinity chromatography in the final step preceded by ion-exchange chromatography. In conclusion, the predominant costs are equipment, the stationary phase and reagents.

Another modeling study to estimate the extraction and purification costs of recombinant proteins was developed by Evangelista et al. (*11*). The model is based on the production of β-glucuronidase from transgenic corn seed, and the process was divided into three sections. The first is a milling section where corn flour is produced. The second section consists of the aqueous extraction of protein from corn flour, removal of spent solids, concentration of the crude extract and drying of the spent solids. Finally, the third section is the purification that includes three-stage chromatography (two ion exchange and one hydrophobic interaction), diafiltration and freeze-drying.

A very significant approach in several areas in biotechnology, such as synthesis and design of large-scale protein separation synthesis (*12*−*15*), involves the development of artificial intelligence (AI) systems. These operate through the manipulation of heuristic rules, algebraic equations and databases (*16*).

A review of computer techniques based on artificial intelligence that are applied to the analysis of chromatographic methods for protein separation has been recently reported in the literature by Bryant and Rowe (*17*) with emphasis on commercially available tools. Eriksson et al. (*18*) and Forslund (*19*) describe the implementation of a deductive synthesizer for protein purification, which encompasses stages from extraction to high-resolution chromatography techniques. A more detailed planner that is restricted to liquid chromatography of high-resolution was proposed by Eriksson et al. (*20*) and relies on a knowledge-based system. The authors divide the problem in to three levels: selection, sequencing and treatment between steps.

Asenjo and Maugeri (*16*) discuss the development of a rational process design and its potential application to a large-scale process for purifying proteins, with the objectives of obtaining high yield and yet a high-purity product, while minimizing costs. The authors use an expert system that develops and organizes a database.

Leser and Asenjo (*21*) developed a knowledge-based expert system for the selection of protein recovery, separation and purification processes for multicomponent mixtures. Fundamental databases that contain the characteristics of protein molecules, equations that describe heuristics and the behavior of proteins are elements of the expert system. The same system was further extended by Lienqueo et al. (*22*) in order to include an algorithm to model the amount of each protein contaminant eliminated after each step. Lienqueo et al. (*23*) developed two criteria to select the sequence of purification. One uses the selection separation coefficient and the other uses the final level of purity. Results for the synthesis of a model protein mixture as well as for the purification of β-1,3-glucanase were validated experimentally (*24*).

The main objective of this paper is to determine the optimal selection and sequencing of purification steps that are required to separate contaminants from a protein mixture. For that purpose, mathematical models based on mixed integer linear optimization are developed. The main target is to obtain the minimum number of stages of purification, by applying high-resolution chromatography in order to attain a purity level required to the final product and then maximizing the purity of product.

The remaining of the paper is organized as follows. First, the problem of protein purification is described. Then the mathematical models and solution strategies are presented, followed by examples that illustrate the proposed methods, as well as a comparison between the proposed approaches and the solutions generated by the expert system derived by Lienqueo and co-workers (*22*, *23*). Finally, conclusions are drawn for the models and main results.

2. Problem Description

Given a mixture of proteins in different concentration levels and a desired product specification in terms of a minimum purity, the problem is to synthesize a high-resolution purification process, which is usually carried out by liquid chromatography. Selection of these purification operations is based on the efficiency of different chromatographic techniques that may be employed to separate the target protein from the contaminant ones. Therefore the objective is to select chromatographic steps and their sequence in order to minimize the total number of operations and/or maximize the purity.

The separation by chromatographic techniques depends on the protein partition differential between the stationary and mobile phases. Information on physicochemical properties can be used for the target and contaminant proteins and each chromatographic technique is able to perform the separation of the mixture by exploiting a specific physicochemical property. For instance, ion exchange chromatography separates proteins on the basis of their difference in charge. The charge of a protein changes with the pH following the titration curve. Ion exchange can use small differences in charge that yield a very high resolution, and hence it is an extremely efficient operation to separate proteins.

The above information on physicochemical protein properties and chromatographic techniques is then used to generate the best separation steps and their sequence within a set of candidate chromatographic units and operational conditions. For that purpose, first an optimization model that minimizes the total number of chromatographic steps for a given purity level is proposed. The resulting minimum number of total steps may then be used to generate an MILP that maximizes the total purity.

A particular case occurs when the selection of chromatographic techniques in the purification process is

Table 1. Expressions and Parameters for the Chromatographic Steps

chromatographic technique	dimensionless retention times[a] $Kd_{i,p}$		peak width σ_i
anion exchange	$Kd_{AE,p} = \dfrac{7383|(Q_p/MW_p)|}{1 \times 10^{-25} + 15844|(Q_p/MW_p)|}$	if $Q_P < 0$	0.15
	$Kd_{AE,p} = 0$	if $Q_P \geq 0$	
cation exchange	$Kd_{CE,p} = \dfrac{5972|(Q_p/MW_p)|}{1 \times 10^{-25} + 17065|(Q_p/MW_p)|}$	if $Q_P > 0$	0.15
	$Kd_{CE,p} = 0$	if $Q_P \leq 0$	
hydrophobic interaction	$Kd_{HI,p} = 1 - \dfrac{[(NH_4)_2SO_4]_p}{[(NH_4)_2SO_4]_{max}}$, $[(NH_4)_2SO_4]_{max} = 1.5$ M		0.22
gel filtration	$Kd_{GF,p} = -0.4691 \log MWp + 2.3902$		0.46

[a] *MW*: molecular weight in Da.

independent of their sequence. Alternative models can be derived from the previous models both for minimization of the total number of steps and for maximization of final purity.

3. Mathematical Models for Optimal Selection and Sequencing

In this section, we describe optimization models for synthesizing a sequence of purification steps in order to attain specified protein purity.

In the optimization models, the following are the indices:

a protein property
dp desired protein
i chromatographic step ($i = 1, ..., I$)
k order in the sequence ($k = 1, ..., K$)
k^* final order in the sequence ($k^* \leq K$)
p protein (product + contaminants) ($p = 1, ..., P$)

the variables:

$m_{dp,k}$ desired protein mass after chromatographic step in order k
$m_{p,k}$ mass of contaminant p after chromatographic step in order k ($k = 2, ..., K$)
S function to be minimized
T function to be minimized
$y_{i,k}$ binary variable that denotes whether chromatographic step i is used in order k in the sequence
Z_k binary variable that denotes whether order k is the last in the sequence

and the parameters:

A_i set of properties used in technique i
$C_{p,1}$ initial concentration of protein p
$CF_{i,p}$ concentration factor of contaminant p after chromatographic step i (mass reduction of contaminant p after and before chromatographic step i)
$DF_{i,p}$ deviation factor for protein p in chromatographic step i
$Kd_{i,p}$ retention time for protein p in chromatographic step i
$m_{p,1}$ initial mass of protein p
$P_{a,p}$ value of property a for protein p
SP_{dp} specified purity of the desired protein dp
U upper bound on mass
σ_i peak width of chromatographic step i

3.1. Model of a Chromatographic Step. The present work uses established criteria for modeling chromatographic techniques, which are then used in the optimal selection of steps for purification of proteins. These criteria are based on the retention time and on the width of the chromatographic peak. Parameters that were determined experimentally using pure proteins (*22*) are used to predict the chromatographic behavior. This is accomplished through the determination of retention times in different chromatographic solutions and from these retention times the resulting dimensionless retention times ($Kd_{i,p}$) for each one of the used chromatographic techniques. Moreover, with the purpose of characterizing the chromatographic processes the peak width parameter (σ_i), averaged over several proteins, is used in the determination of the reduction of contaminants after applying a chromatographic technique.

These parameters are related to physicochemical properties by means of mathematical correlations, developed for ion-exchange chromatography by Lienqueo et al. (*22*), who had made tests for linear, polynomial and exponential functions. For hydrophobic interaction chromatography, the same authors had determined the best correlation by relating dimensionless retention times to the concentration of ammonia sulfate in which the protein elutes from the chromatographic column and with the maximum concentration of ammonia sulfate in which the elution gradient is initiated. In the case of gel filtration chromatography the relationship is established between the retention time and the logarithm of the molecular weight. Mathematical relationships that predict values for these parameters in the chromatographic techniques are given in Table 1.

Each one of the chromatographic operations associates to it a physicochemical or biochemical property of the protein. Thus for the ion-exchange chromatography (cation and anion) it associates charge property at different pH levels, for hydrophobic interaction chromatography the property of interest is the superficial hydrophobicity, and for gel filtration chromatography it is the molecular weight.

The dimensionless retention time ($Kd_{i,p}$) is a function of the physicochemical properties of protein p and chromatographic step i, while the peak width parameter (σ_i) depends only on the type of chromatographic operations (*23*). Note that both anion and cation exchange steps are functions of charge density (Q_p/MW_p). This charge was determined by eletrophoretic titration curves as indicated by Pharmacia (*25*) and Watanabe et al. (*26*) and presented a very good fit in order to predict the

(a)

(b)

(c)

(d)
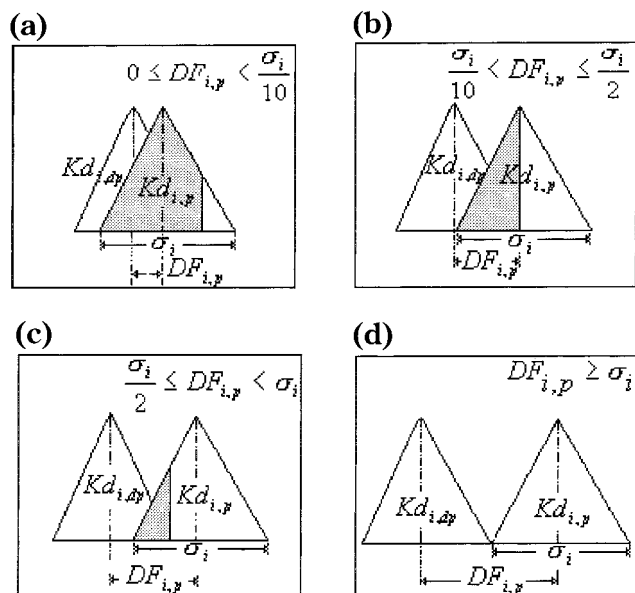
Figure 1. Representation of chromatographic peaks. (The triangle on the left represents the desired protein and the one on the right contaminant p).

orders of protein elution. Consequently, correlations were developed that take into account Q_p/MW_p. Note that it would be possible to use other parameters such as the Z-potential. If the protein charge is negative then the retention time for anion exchange is given by the mathematical relationship shown in Table 1; otherwise the retention time is zero. On the other hand, for cation exchange $Kd_{i,p}$ is given by a mathematical relationship if the charge Q_p is positive; in the opposite case $Kd_{i,p}$ is set to zero. For hydrophobic interaction the retention time is a function of the hydrophobicity and for gel filtration the relevant property is the molecular weight (see Table 1).

Then, in general form every chromatographic technique can be expressed as:

$$Kd_{i,p} = f_i(P_{a,p}) \quad \forall\ i,\ p,\ a \in A_i \tag{1}$$

In eq 1, $P_{a,p}$ denotes the values of each property a of protein p that are relevant for chromatographic step i. The relevant properties for step i are specified in Ai.

The other parameter is the deviation factor $(DF_{i,p})$ that is also dimensionless and indicates the distance between the product and contaminant peaks. For every protein and chromatographic step i the deviation factor can be calculated with respect to the desired protein dp (see Figure 1), according to:

$$DF_{i,p} = |Kd_{i,dp} - Kd_{i,p}| \quad \forall\ i,\ p \tag{2}$$

It is important to note that the deviation factor $(DF_{i,p})$ could measure any other specific property of product and contaminants.

With the objective of determining the degree of separation of each protein or quantifying the amount of contaminant that remains with the product after applying a chromatographic technique, we take into account two considerations developed by Leser (*27*) and Lienqueo (*28*). First, the peak shapes are approximated by triangles. Second, the mass reduction of contaminant is a function

of the intersection of the contaminant and desired protein triangles. As a consequence of these approximations, a percentage of error Δ is established in the calculation of the concentration factor $(CF_{i,p})$. This parameter is represented by mathematical relationships determined by approximating two chromatographic peaks of equal size and form, one corresponding to the product and another to the contaminant protein.

In Figure 1, admitting that the peaks have constant form and considering that the peak on the left refers to the product and the one on the right to the contaminant protein, the area of the figure formed by the intersection of the two triangles (shaded areas) represents the amount of contaminant p that remains in the mixture (with the product) after applying chromatographic technique i (*22*). Note that 100% of the target protein is recovered.

For any protein p and chromatographic step i the following mathematical relationships for $CF_{i,p}$ hold:

$$\text{If } 0 \le DF_{i,p} < \frac{\sigma_i}{10} \quad \text{then } CF_{i,p} = 1 \tag{3a}$$

$$\text{If } \frac{\sigma_i}{10} \le DF_{i,p} < \frac{\sigma_i}{2} \quad \text{then } CF_{i,p} = (1 + \Delta)\left(\frac{\sigma_i^2 - 2DF_{i,p}^2}{\sigma_i^2}\right) \tag{3b}$$

$$\text{If } \frac{\sigma_i}{2} \le DF_{i,p} < \sigma_i \quad \text{then } CF_{i,p} = 2(1 + \Delta)\frac{(\sigma_i - DF_{i,p})^2}{\sigma_i^2} \tag{3c}$$

$$\text{If } DF_{i,p} \ge \sigma_i \quad \text{then } CF_{i,p} = \Delta \tag{3d}$$

Then as a function of $DF_{i,p}$ calculated in eq 2, as well as of σ_i, it is possible to calculate $CF_{i,p}$ from eqs 3. Note that in Figure 1, $DF_{i,p}$ is observed as a difference between the times corresponding to the peak values, and $CF_{i,p}$ represents the ratio between the mass of contaminant p after and before chromatographic step i. The mathematical relationships expressed in eqs 3 represent graphical approximations of the chromatograms for two different proteins. As a result of these approximations, a fraction Δ of proteins is neglected and admitted not to separate from the product. The value of 0.02, which was determined experimentally, is used for Δ.

With the development of models for each of the chromatographic steps it is then possible to define optimization models for the synthesis of purification processes. These will be described in the next two items.

3.2. Model for Minimizing the Number of Purification Steps. The model M_{1a} for protein purification using high resolution that must be selected to attain a final purity level of the desired product is composed of the following constraints:

(a) assignment constraints

$$\sum_i y_{i,k} \le 1 \quad \forall\ k \tag{4a}$$

$$\sum_k y_{i,k} \le 1 \quad \forall\ i \tag{4b}$$

(b) ordering constraints

$$\sum_i y_{i,k+1} \le \sum_i y_{i,k} \quad k \le K - 1 \tag{5a}$$

$$Z_k \ge \sum_i y_{i,k} - \sum_i y_{i,k+1} \quad k \le K - 1 \tag{5b}$$

$$\sum_i y_{i,k'} \geq Z_k \quad \forall \, k, k' \leq k \tag{5c}$$

$$\sum_i y_{i,k'} + Z_k \leq 1 \quad \forall \, k, k' > k \tag{5d}$$

$$\sum_k Z_k = 1 \tag{5e}$$

(c) contaminant constraints

$$m_{p,k} = \sum_i CF_{i,p} m_{p,k-1} y_{i,k-1} \quad \forall \, p, k \geq 2 \tag{6}$$

(d) purity constraints

$$m_{dp,k+1} \geq SP_{dp} \sum_{p'} m_{p'k+1} - U(1 - Z_k) \quad k < K \tag{7}$$

Constraint 4a states that at most one chromatographic process i is chosen in order k, while 4b enforces that each chromatographic step is selected at most once in the sequence.

To avoid that empty steps are chosen, as well as to reduce the combinatorial search, constraint 5a states that step k are assigned in increasing order, that is, step $k + 1$ can only be chosen if step k is also selected. Constraints 5b–5d define the last step of the sequence, denoted by Z_k. According to 5b, if step k is chosen ($y_{i,k} = 1$ for some i) but not $k + 1$ ($y_{i,k+1} = 0$ for all i) then k is the last step in the sequence. On the other hand, if k is the last step in the sequence ($Z_k = 1$) then 5c enforces that all previous steps are selected ($k' \leq k$) while 5d states that all the steps that succeed k are empty. Equation 5e imposes that there is exactly one final step in the sequence.

Equation 6 relates the mass values of each protein in subsequent steps. Note, however, that they are in non-linear form, since the summation term contains the product of a continuous and a discrete variable. Equation 6 can be linearized as:

$$m_{p,2} = \sum_i CF_{i,p} y_{i,1} m_{p,1} \quad \forall \, p \tag{8a}$$

$$m_{p,k} - CF_{i,p} m_{p,k-1} \geq -U(1 - y_{i,k-1}) \quad \forall \, i, p, k \geq 3 \tag{8b}$$

$$m_{p,k} - CF_{i,p} m_{p,k-1} \leq U(1 - y_{i,k-1}) \quad \forall \, i, p, k \geq 3 \tag{8c}$$

It is important to note that eq 6 may be used directly for the first chromatographic step, since the initial mass of all proteins $m_{p,1}$ is known a priori, as expressed in 8a. Constraints 8b and 8c on subsequent steps are enforced only if any of them is selected. U is a valid upper bound on protein mass.

Constraint 7 enforces the purity specification for the target protein. Note from eq 7 that if Z_{k^*} is set to one, (k^* is the last chromatographic step), then the following condition is imposed:

$$\frac{m_{dp,k^*+1}}{\sum_p m_{p,k^*+1}} \geq SP_{dp} \tag{9}$$

Constraint 9 imposes that the purity of the desired protein after the last step achieves its specified value. If $Z_k = 0$, then constraint 7 is relaxed.

The objective function is the minimization of the total

number of chromatographic steps, given as:

$$\text{Min } S = \sum_k k Z_k \tag{10}$$

Note that eq 10 is equivalent to minimize the total number of binary variables, that is, $\sum_i \sum_k y_{i,k}$.

The resulting MILP (M_{1a}) has the following structure:
Minimize (eq 10)
Subject to:

assignment constraints (eqs 4)

ordering constraints (eqs 5)

contaminant constraints (eqs 8)   (M_{1a})

purity constraints (eq 7)

$y_{i,k} \in \{0,1\}$

$m_{p,k}, Z_k \geq 0$

The solution of model M_{1a} provides a sequence, which contains the minimum number of steps that are defined to satisfy the required purity level. Nevertheless, several sequences may be selected that attain this specified purity. Note that, for any given selection, any sequence would satisfy final purity for the objective function defined in eq 10.

It is important to note that information regarding the high-resolution models in constraints 8 as discussed in section 3.1 is calculated a priori.

3.3. Model for Maximization of Purity. A second model (M_{1b}) is proposed for maximizing the purity level for a fixed number of steps. The number of steps may be the one obtained by model M_{1a} or any arbitrary number. By denoting k^* the fixed number of steps, the objective function for maximizing the purity of the mixture is defined as follows:

$$\text{max } T'' = \frac{m_{dp,k^*+1}}{\sum_p m_{p,k^*+1}} \tag{11}$$

where m_{dp,k^*+1} represents the mass of the desired protein after the mixture goes to step k^*.

Note that eq 11 is in nonlinear form. However, since the mass of the desired product remains the same along the steps, function 11 is equivalent to:

$$\text{max } T' = \frac{\sum_p m_{p,k^*+1}}{m_{dp,k^*+1}} \tag{12}$$

In eq 12, by removing the (constant) mass of the desired protein, the objective function becomes:

$$\text{min } T = \sum_{p \neq dp} m_{p,k^*+1} \tag{13}$$

The constraints defined for model M_{1b} are as follows:
(e) assignment constraints

$$\sum_i y_{i,k} = 1 \quad k \leq k^* \tag{14}$$

$$\sum_{k=1}^{k^*} y_{i,k} \leq 1 \, \forall \, i \qquad (15)$$

(f) contaminant constraints (linearized)

$$m_{p,2} = \sum_i CF_{i,p} y_{i,1} m_{p,1} \, \forall \, p \qquad (8a)$$

$$m_{p,k} - CF_{i,p} m_{p,k-1} \geq -U(1 - y_{i,k-1})$$
$$\forall \, i,p, 3 \leq k \leq (k^* + 1) \quad (16a)$$

$$m_{p,k} - CF_{i,p}{}^* m_{p,k-1} \leq U(1 - y_{i,k-1})$$
$$\forall \, i,p, 3 \leq k \leq (k^* + 1) \quad (16b)$$

Equation 14 indicates that only one chromatographic process i is chosen in order k, while eq 15 enforces that each chromatographic step is selected at most once in the sequence. Note that eq 14 is an equality constraint as opposed to eq 4a, which is an inequality, as a result of the fact that the minimum number of steps is known for model M_{1b}. The linearized contaminant constraints 16a and 16b are also used in the range of k from 1 to k^*.

In summary, M_{1b} is an MILP composed of objective function 13 and constraints 8a and 14−16 with $y_{i,k} \in \{0,1\}$; $m_{p,k} \geq 0$.

4. Mathematical Models for Optimal Selection

A particular case may occur in which only the selection of steps is performed. This is especially important whenever the sequence does not affect the optimal solution, as is the case when the total number of steps must be minimized in order to attain a final purity level.

A key issue in this formulation is that the decision variables must simply select the subset of chromatographic steps from the candidate techniques. A binary variable w_i that is activated if step i is selected is defined in place of $y_{i,k}$.

In this formulation, steps are selected independently of their sequence; Figure 2 illustrates the difference between this representation and the previous one. Consider a simple example in which there are six candidate steps. Suppose that among these, steps 2, 4, and 5 are chosen in the following sequence: 2 - 5 - 4. In model M_{1a} assignment constraints would enforce that $y_{2,1} = y_{5,2} = y_{4,3} = 1$. In this model we would simply have $w_2 = w_4 = w_5 = 1$.

The optimization model, denoted by M_{2a}, contains the following new variables:

$mo_{i,p}$ mass of protein p after chromatographic step i
R function to be minimized
w_i binary variable that denotes whether chromatographic step i is selected

In this case, contaminant mass can be represented by the following disjunction:

$$\begin{bmatrix} w_i \\ mo_{i,p} = CF_{i,p} mo_{i-1,p} \end{bmatrix} \vee \begin{bmatrix} \neg w_i \\ mo_{i,p} = mo_{i-1,p} \end{bmatrix} \quad \forall \, i \geq 2, p \qquad (17)$$

The first term in disjunction 17 models the case of selection of step i and the second models the opposite situation. Recall that $CF_{i,p}$ is the concentration factor for protein p in step i and denotes the mass reduction of contaminant p after and before chromatographic step i.

(a)

(b)
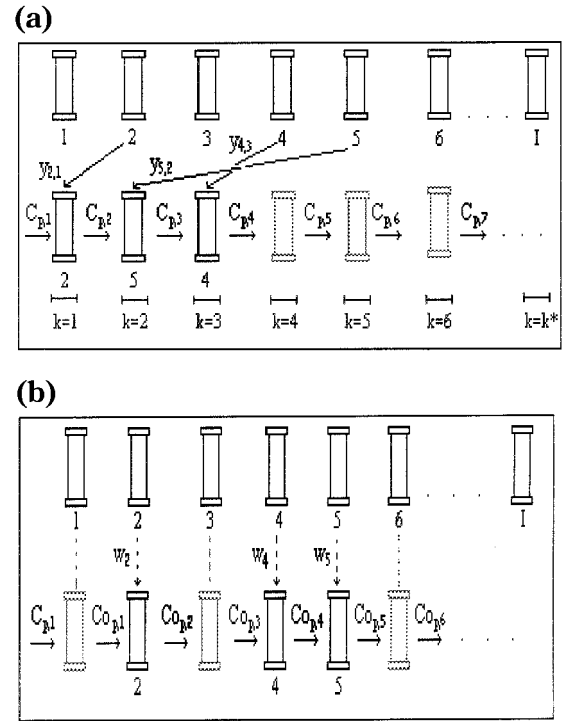
Figure 2. Selection of high-resolution steps: (a) model M_{1a} and (b) model M_{2a}.

Thus, in eq 17 mass of protein p after going through step i is modified if this is selected, otherwise it remains the same.

Disjunction 17 generates the following constraints in mixed-integer representation:

$$-U(1 - w_1) \leq mo_{1,p} - CF_{1,p} m_{p,1} \leq U(1 - w_1) \quad \forall \, p \qquad (18a)$$

$$-U w_1 \leq mo_{1,p} - m_{p,1} \leq U w_1 \quad \forall \, p \qquad (18b)$$

$$-U(1 - w_i) \leq mo_{i,p} - CF_{i,p} mo_{i-1,p} \leq U(1 - w_i)$$
$$\forall \, i \geq 2, p \qquad (19a)$$

$$-U w_i \leq mo_{i-1,p} - mo_{i,p} \leq U w_i \quad \forall \, i \geq 2, p \qquad (19b)$$

As defined in the previous section, $m_{p,1}$ corresponds to the initial mass of protein p in the mixture.

Constraints 18 enforce that if the first chromatographic step is chosen ($w_1 = 1$) then the resulting mass is $mo_{1,p} = CF_{1,p} m_{p,1}$. Otherwise, it remains the same ($mo_{1,p} = m_{p,1}$). The same approach is taken in eqs 19 for the remaining steps.

Finally, there is the purity constraint that imposes that the mass of the desired protein is above the specified purity level SP_{dp} in the last chromatographic step I, which is written as:

$$mo_{I,p} \geq SP_{dp} \sum_{p'} mo_{I,p'} \qquad (20)$$

Note that $mo_{I,p}$ corresponds to the final purity of the mixture, regardless the selection of techniques, as shown in disjunction 17.

For the minimization of the total number of selected steps, the objective function is written as follows:

$$\text{Min } R = \sum_i w_i \qquad (21)$$

The resulting model, denoted as M_{2a}, is then:
Minimize (eq 21)
Subject to:

contaminant constraints (eqs 18 and 19)

purity constraint (eq 20)   (M_{2a})

$$w_i \in \{0,1\}$$

$$mo_{i,p} \geq 0$$

Note that there is a significant reduction in the model size with respect to M_{1a}. The total number of binary variables is I as opposed to $I \cdot K$ variables in M_{1a}, in which K is in general of the same order of I.

In the previous section, model M_{1b} was proposed for the maximization of a given purity level for a fixed number of steps. In a similar way, an alternative model M_{2b} may be formulated. For a given number of steps k^* (which may be given or determined by the solution of M_{1a} or M_{2a}) we define:

$$\sum_i w_i = k^* \qquad (22)$$

The objective function for the maximization of the purity of the desired protein may be defined as follows:

$$\max R = \frac{mo_{I,dp}}{\sum_p mo_{I,p}} \qquad (23)$$

where I corresponds to the last chromatographic step in the set of candidate techniques.

By following the same procedure as the one for generating eq 13, function 23 is equivalent to minimizing the following function:

$$\min R = \sum_{p \neq dp} mo_{I,p} \qquad (24)$$

Finally, model M_{2b} has the following structure:
Minimize (eq 24)
Subject to:

selection constraints (*22*)

contaminant constraints
                    (eqs 18 and 19)(M_{2b})

$$w_i \in \{0,1\}$$

$$mo_{i,p} \geq 0$$

Models M_{2a} and M_{2b} may be solved either in a two-step procedure or individually.

5. Computational Results

The software General Algebraic Modeling System (GAMS), version 2.25 (*29*) was used to implement the model and its solution method. This program is used for large-scale optimization and has a number of solvers for linear, nonlinear and mixed-integer programming applications.

In this paper the proposed models were solved with the Optimization Subroutine Library OSL (*30*), which is a high performance integrated solver for linear program-

ming (LP) and mixed-integer linear programming (MIP). The LP based branch and bound implementation of OSL used several options in order to achieve better efficiency: dual simplex for LP solution, branch and bound preprocessing and model scaling. Also, several branching strategies were tested.

The models described in the previous sections are solved for three different examples of increasing size in terms of protein types and candidate techniques. The results are compared to solutions generated by expert systems (*23*, *28*) and obtained experimentally (*24*).

5.1. Example 1. In this example (*23*), we consider the purification of a mixture containing four proteins: serum from bovine albumin (*p*1), ovalbumin (*p*2), soybean trypsin inhibitor (*p*3) and thaumatin (*p*4), all in equal concentration. Their physicochemical properties as well as the initial protein concentration of the mixture are shown in Table 2. Note that by considering any given initial volume, the initial mass values are taken as information for the models. The purity level required for protein *p*1 is 98% and may be attained through the appropriate selection from multimodal chromatographic techniques, which exploit differences in molecular charge, size and hydrophobicity. Overall, there are 12 candidate high-resolution steps (see Tables 2 and 3).

The dimensionless retention times ($Kd_{i,p}$) for each protein in every chromatographic step are calculated according to the relations and data shown in Tables 1 and 2. These values, the deviation factor ($DF_{i,p}$) and the concentration factor ($CF_{i,p}$) are calculated from eqs 1−3 and are shown in Table 3 for two representative proteins. These parameters as well as the peak width values (σ_i) are used in the models for determining the optimal synthesis of chromatographic techniques.

The purity obtained is 99.8%, which is attained in three steps and can be seen in Table 4 for all models. When the optimization is performed with model M_{1a}, the final purity and selection of steps are the same as those from model M_{1b} but their sequence is different. In the case of model M_{1a} the result shown is one example of several possible sequences that this model generates, while model M_{1b} not only calculates the maximum purity but also gives the optimal selection. Models M_{2a} and M_{2b} generate the same solution, which selects the same operations as those chosen by the previous models.

Note from Table 2 that the hidrophobicity of *p*2 is much lower than that of the remaining proteins. Therefore by choosing hydrophobic interaction as the first the chromatographic technique, *p*2 is almost completely eliminated, which is the step choice for M_{1b}. The same techniques are also selected for the other models, in a different sequence. Proteins *p*3 and *p*4 are separated by choosing techniques for which their charge is the lowest, that is, anion exchange at pH values 7.0 and 8.0, respectively.

The optimal results can be compared to the solution obtained by the hybrid expert system Prot-Ex (*23*) for 94.5% purity. We obtain the same sequence and final purity from model M_{1a} as that of Prot-Ex, with a small discrepancy in purity for anion exchange at pH 7.0 and hydrophobic interaction (Prot-Ex solution reports 63.7%, whereas M_{1a} obtains 64.1%). However, since the required purity for the MILP models is 98%, an additional step must be added, which is anion exchange at pH 8.0 as seen in Table 4. It is important to note that if model M_{1b} is used for maximizing purity in two steps, a value of 96.3% is obtained in exactly the same sequence as that from Prot-Ex. This sequence was confirmed experimentally (*24*).

Table 2. Physicochemical Properties of the Protein Mixture in Example 1

protein	$C_{p,1}{}^a$ [mg/mL]	MW_p (Da) a_1	hydrophobicity $[(NH_4)_2(SO_4)]_p$ a_2	pH 4.0 a_3	pH 5.0 a_4	pH6.0 a_5	pH7.0 a_6	pH8.0 a_7
						Q_p (charge, C/mol) $\times 10^{-25}$		
$p1$	2.00	67000	0.86	1.03	−0.14	−1.16	−1.68	−2.05
$p2$	2.00	43800	0.54	1.40	−0.76	−1.65	−2.20	−2.36
$p3$	2.00	24500	0.90	1.22	−0.76	−1.54	−2.17	−2.13
$p4$	2.00	22200	0.89	1.94	1.90	1.98	1.87	0.91

^{a} $C_{p,1}$: initial concentration.

Table 3. Parameters for Proteins *p*2 and *p*3 for Each Step in Example 1

number	name	$p2$ Kd	DF	CF	$p3$ Kd	DF	CF
1	anion exchange pH 4.0	0.000	0.000	1.000	0.000	0.000	1.000
2	anion exchange pH 5.0	0.100	0.086	0.377	0.154	0.139	0.012
3	anion exchange pH 6.0	0.174	0.074	0.525	0.233	0.132	0.029
4	anion exchange pH 7.0	0.206	0.074	0.523	0.272	0.140	0.010
5	anion exchange pH 8.0	0.215	0.062	0.666	0.270	0.118	0.094
6	cation exchange pH 4.0	0.124	0.051	0.786	0.161	0.088	0.384
7	cation exchange pH 5.0	0.000	0.000	1.000	0.000	0.000	1.000
8	cation exchange pH 6.0	0.000	0.000	1.000	0.000	0.000	1.000
9	cation exchange pH 7.0	0.000	0.000	1.000	0.000	0.000	1.000
10	cation exchange pH 8.0	0.000	0.000	1.000	0.000	0.000	1.000
11	hydrophobic interaction	0.640	0.213	0.002	0.400	0.027	0.990
12	gel filtration	0.213	0.087	0.615	0.331	0.205	0.615

Table 4. Optimal Values for Each Model in Example 1^a

step proteins	$m_o{}^b$	model M_{1a} AE pH 7.0 m_1	HI m_2	AE pH 8.0 m_3	model M_{1b} HI m_1	AE pH 8.0 m_2	AE pH 7.0 m_3
$p1$	2.00	2.00	2.00	2.00	2.00	2.00	2.00
$p2$	2.00	1.047	0.002	0.001	0.004	0.002	0.001
$p3$	2.00	0.020	0.019	0.002	1.98	0.186	0.002
$p4$	2.00	0.056	0.056	0.001	2.00	0.040	0.001
purity	0.25	0.641	0.963	0.998	0.334	0.898	0.998

step proteins	m_o	model M_{2a} AE pH 7.0 m_1	AE pH 8.0 m_2	HI m_3	model M_{2b} AE pH 7.0 m_1	AE pH 8.0 m_2	HI m_3
$p1$	2.00	2.00	2.00	2.00	2.00	2.00	2.00
$p2$	2.00	1.047	0.697	0.001	1.047	0.697	0.001
$p3$	2.00	0.020	0.002	0.002	0.02	0.002	0.002
p4	2.00	0.056	0.001	0.001	0.056	0.001	0.001
purity	0.25	0.641	0.741	0.998	0.641	0.741	0.998

^{a} HI: hydrophobic interaction; AE: anion exchange; CE: cation exchange. ^{b} On a volumetric basis.

Figure 3 shows the purity range that may be attained as a function of the number of steps for all proposed models. This can be interpreted as follows: if two steps are used, a maximum 96.29% purity is obtained; however, if 99.9% is the final purity required then at least four steps should be used. Note that in this example all models present the same behavior.

5.2. Example 2. In this case we consider the purification of β-1,3 glucanase produced from *Bacillus subtilis*, in which the product must be separated from eight contaminants, denoted by cont_1−8. Their physicochemical properties are shown in Table 5. Contaminants 6 and 7 are the ones with the highest concentration and exhibit low hydrophobicity. Physicochemical data for this example was provided from ref 24.

The purification requires 94% for β-1,3 glucanase, and the steps must be synthesized from a set of 22 candidate high-resolution steps. Note that besides the chromatographic techniques from Example 1 there are 10 additional candidate steps. These are anion and cation exchange at 4.5, 5.5, 6.5, 7.5 and 8.5 pH levels. Nevertheless, the same mathematical relationships shown in Table 1 are used, since these are valid for any ion exchange operation despite the pH level.
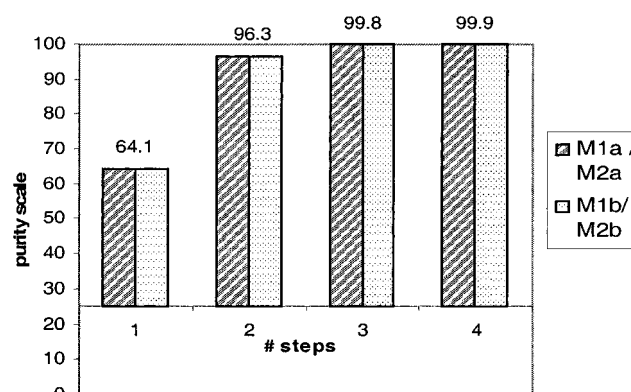
Figure 3. Purity scale for example 1.

The resulting techniques and their sequence are shown in Figure 4 for the four models. Note that this mixture is very difficult to purify. Overall, six steps are required, and a final maximum purity of 94.8% is attained for models M_{1b} and M_{2b}. Note that model M_{1a} selects the hydrophobic interaction as first technique and eliminates the proteins that have low hydrophobicity (contaminants

**Table 5. Physicochemical Properties of the Protein Mixture in Example 2**

proteins	$C_{p,1}$ (g/L)	MW_p (Da) a_1	H^a [NH$_4$(SO$_2$)$_4$]$_p$ a_2	Q_p (charge, C/mol) $\times 10^{-25}$									
				pH 4.0 a_3	pH 4.5 a_4	pH 5.0 a_5	pH 5.5 a_6	pH 6.0 a_7	pH 6.5 a_8	pH 7.0 a_9	pH 7.5 a_{10}	pH 8.0 a_{11}	pH 8.5 a_{12}
target	0.62	31000	0.00	1.46	0.09	−0.62	−0.66	−1.02	−1.82	−2.33	−2.52	−2.52	−3.51
cont 1	0.42	62500	0.00	1.46	0.09	−1.06	−0.98	−1.17	−1.71	−2.79	−3.52	−3.32	−3.32
cont 2	0.25	40600	0.00	1.46	0.09	−0.55	−0.22	−0.22	−0.26	−0.73	−1.26	−1.82	−3.51
cont 3	0.25	69600	0.00	1.46	0.09	−0.55	−0.22	−0.22	−0.26	−0.73	−1.26	−1.82	−3.51
cont 4	0.09	40600	0.00	1.46	3.14	1.46	0.28	−0.47	−0.89	−1.06	−1.08	−1.04	−1.01
cont 5	0.09	69600	0.00	1.46	3.14	1.46	0.28	−0.47	−0.89	−1.06	−1.08	−1.04	−1.01
cont 6	2.74	41000	1.50	1.46	0.93	0.26	−0.35	−0.87	−1.31	−1.65	−1.90	−2.04	−2.06
cont 7	2.74	32900	1.50	1.46	0.09	0.00	−1.70	−2.70	−2.90	−3.51	−3.51	−3.51	−3.51
cont 8	0.25	35500	0.20	1.46	0.09	−0.55	−0.22	−0.22	−0.26	−0.73	−1.26	−1.82	−3.51

^{a} *H*: hydrophobicity.
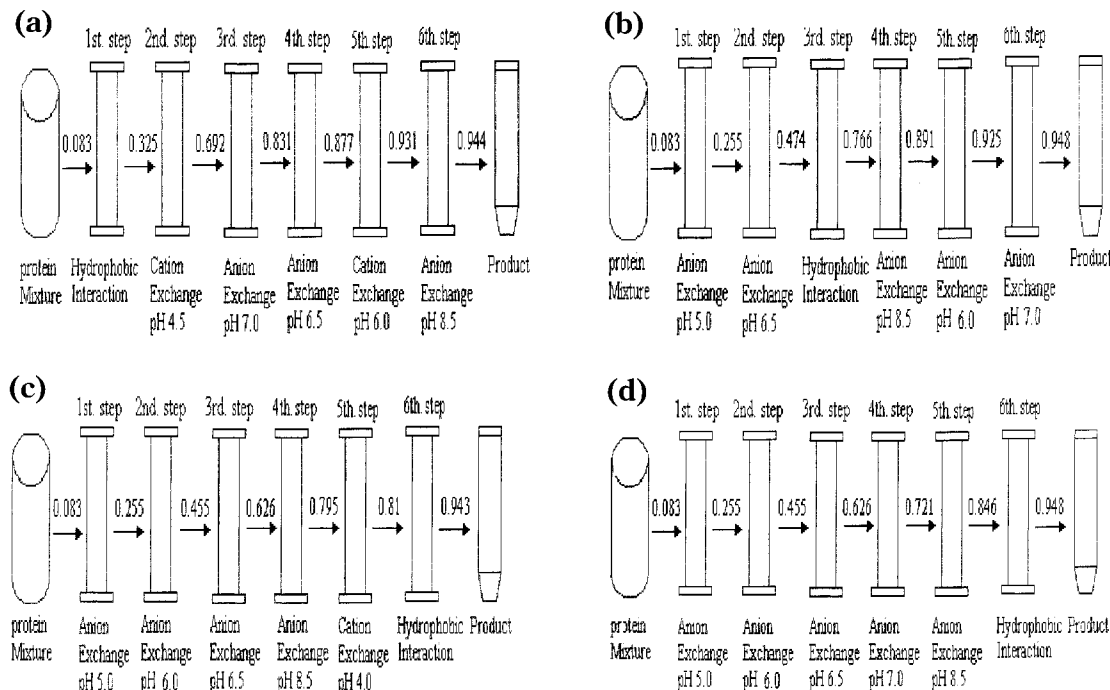
Figure 4. Optimal results for example 2: (a) model M$_{1a}$, (b) model M$_{1b}$, (c) model M$_{2a}$, and (d) model M$_{2b}$.

6 and 7); model M$_{1b}$ determined other techniques that prioritize mass reductions in most of the contaminants. In the particular case when maximizing purity with M$_{1b}$ the model is not solved to global optimality. However, despite this fact, model M$_{1b}$ obtained a higher purity value than the one from M$_{1a}$. Models M$_{1a}$ and M$_{2a}$ generate solutions that minimize the number of steps and still satisfy the required purity.

The solution provided by the ES Prot-Ex Purification (*24*) is limited to a final purity of 70%. The suggested sequence is hydrophobic interaction (reaching 32.7% purity) followed by anion exchange at 6.5 pH level (70.3%). Interestingly, we obtained the same steps with purities of 32.5% and 69.25% respectively by model M$_{1a}$ that are similar results with little discrepancy; this sequence also was confirmed experimentally (*24*). Moreover, we attained final higher purities of 94.4% for model M$_{1a}$ and 94.8% with M$_{1b}$ (see Figure 4). Nonetheless, if the objective was to obtain the optimal sequence and to maximize purity with only two steps through model M$_{1b}$, the result will be exactly as that obtained from Prot-Ex.

Optimal final purities as a function of the number of chromatographic techniques are illustrated in Figure 5. A larger number of steps are required when compared to Example 1. For purities higher than 90%, at least four steps are needed. Also in this example models M$_{1b}$ and
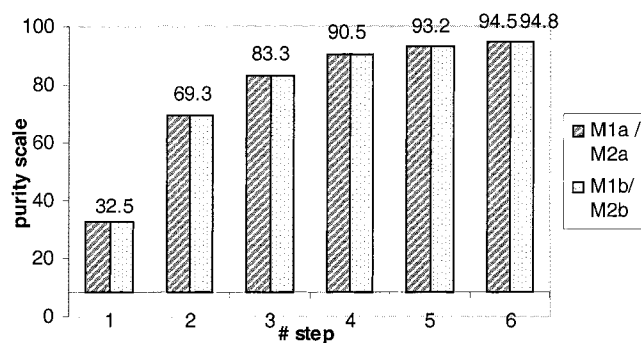
Figure 5. Purity scale for example 2.

M$_{2b}$ generate sequences with higher purity values than those of M$_{1a}$ and M$_{2a}$.

5.3. Example 3. In this example there is a protein mixture generated from a fermentation process that uses *Escherichia coli*. The objective is to purify the mixture in order to obtain somatotropin that is combined with 13 protein contaminants (*p*1−*p*13). Overall, there are the same 22 steps defined in Example 2. The desired purity level of somatotropin is 98%.

The determination of the dimensionless retention times ($Kd_{i,p}$) requires Tables 1 and 6; the physicochemical properties are taken from ref 22. The calculation of

Table 6. Physicochemical Properties of Mixture for Example 3[a]

				Q_p (charge, C/mol) $\times 10^{-25}$			
	C_p,	MW_p (Da)	$H [NH_4(SO_2)_4]_p$	pH 4.0	pH 4.5	pH 8.0	pH 8.5
proteins	(g/L)	a_1	a_2	a_3	a_4	a_{11}	a_{12}
desired	25.00	22000	0.93	4.77	3.81	−0.03	−0.50
p1	11.29	18370	0.71	1.94	0.25	−2.45	−2.67
p2	7.06	85570	0.48	2.35	0.29	−3.68	−3.64
p3	4.63	53660	0.76	1.83	0.76	−1.34	−1.50
p4	5.58	120000	1.50	3.29	1.38	−2.85	−2.75
p5	4.83	203000	0.36	4.08	1.83	−4.98	−5.65
p6	2.48	69380	0.36	5.22	3.17	−3.90	−4.24
p7	7.70	48320	0.48	3.96	3.16	−1.59	−2.84
p8	6.80	93380	0.93	10.90	5.81	−4.45	−4.31
p9	7.53	69380	0.00	1.09	0.55	−0.32	−0.32
p10	6.05	114450	0.63	10.40	5.94	−1.43	−1.72
p11	3.89	198000	0.06	0.33	0.03	−0.97	−1.57
p12	1.48	30400	0.00	5.17	4.22	0.20	0.08
p13	0.83	94670	0.00	11.70	7.94	0.80	0.51

[a] See complete table in Lienqueo et al. (1996).

Table 7. Summary of Statistical Data for Examples 1−3 (models M_{1a} and M_{1b})

model	example	set K	binary variables	continuous variables	constraints	workspace requested (Mb)	nodes	relaxed objective	CPU time[a] (s)
M_{1a}	1	8	104	29	687	0.94	94	1.2094	6.133
		10	130	37	923	1.22	98	1.2094	10.227
		12	156	45	1167	1.52	96	1.2094	16.320
	2	8	184	64	2502	2.58	332947	1.4948	39898.656
		10	230	82	3338	3.42	304129	1.4948	51074.844
		12	276	100	4182	4.29	331673	1.4948	67959.984
	3	8	184	99	3827	3.62	40	1.0579	47.461
		10	230	127	5103	4.79	47	1.0579	95.922
		12	276	155	6387	6.01	53	1.0579	221.051
M_{1b}	1		36	13	214	0.36	730	0.00	5.836
	2[b]		132	55	2023	1.85	95798	0.00	7747.875
	3		44	29	656	0.75	36	0.00	5.059

[a] Pentium 266 personal workstation. [b] Reported for 0.03 absolute gap.

Table 8. Summary of Statistical Data for Examples 1−3 (models M_{2a} and M_{2b})

model	example	binary variables	continuous variables	constraints	workspace requested (Mb)	nodes	relaxed objective	CPU time (s)
M_{2a}	1	12	49	194	0.33	605	0.4932	4.781
	2	22	199	794	0.81	150709	1.0986	7747.031
	3	22	309	1234	1.16	624	0.2113	69.910
M_{2b}	1	12	49	194	0.33	436	0.00	4.156
	2	22	199	794	0.81	130724	0.00	5905.789
	3	22	309	1234	1.16	456	0.00	77.051

deviation factors ($DF_{i,p}$) and concentration factors ($CF_{i,p}$) is similar to Example 1.

Figure 4 shows the graphical representation of the results obtained from the four models. The sequence obtained from model M_{1a} is one of several possible sequences, which reaches 98.8% purity. It is interesting to note that by solving the problem with model M_{1b} (with the same number of steps) 99.3% purity was obtained, as seen in Figure 4. Model M_{2a} selects the same steps as those from M_{1a}. The same occurs for M_{2b} with respect to M_{1b}. It is not possible to compare the results obtained with the ES solution, since steps other than high-resolution chromatography were employed (*22*).

5.4. Model Statistics. Tables 7 and 8 present statistical data for the optimization models M_{1a}, M_{1b} and M_{2a}, M_{2b}, respectively. These results correspond to cases for which the specified purity is 98% for Examples 1 and 3, and 94% for Example 2. In these tables model sizes in terms of binary variables, continuous variables and constraints are given. In Table 7 there is an additional entry concerning the size of the set K, which defines the total number of possible steps that can be assigned to the sequence.

As expected, for models M_{2a} and M_{2b} there is a significant reduction in the number of binary variables and constraints. However, the number of continuous variables in fact increases.

Also shown in Tables 7 and 8 are overall CPU time, relaxed objective value, workspace required, number of iterations and nodes enumerated. Except for the cases that were explicitly mentioned, all models were solved to global optimality, in a personal workstation Pentium 266 with GAMS/OSL. Note from Table 7 that as the number of total steps K increases (in model M_{1a}) there is a significant increase in model size as well as in computational effort.

Comparison between M_{1a} and M_{1b} shows that CPU times are smaller for Examples 1 and 3 in all cases for the latter model. Moreover, note in Table 8 that CPU times for model M_{2a} are smaller in comparison to those for M_{1a}, as a result of the reduction in model size. Nevertheless, although M_{2b} is smaller in size than M_{1b}, there is an increase in CPU time.

Note that in Example 3 the protein mixture contains more components than in Example 2 and the same set of candidate steps is available in both cases. Interest-
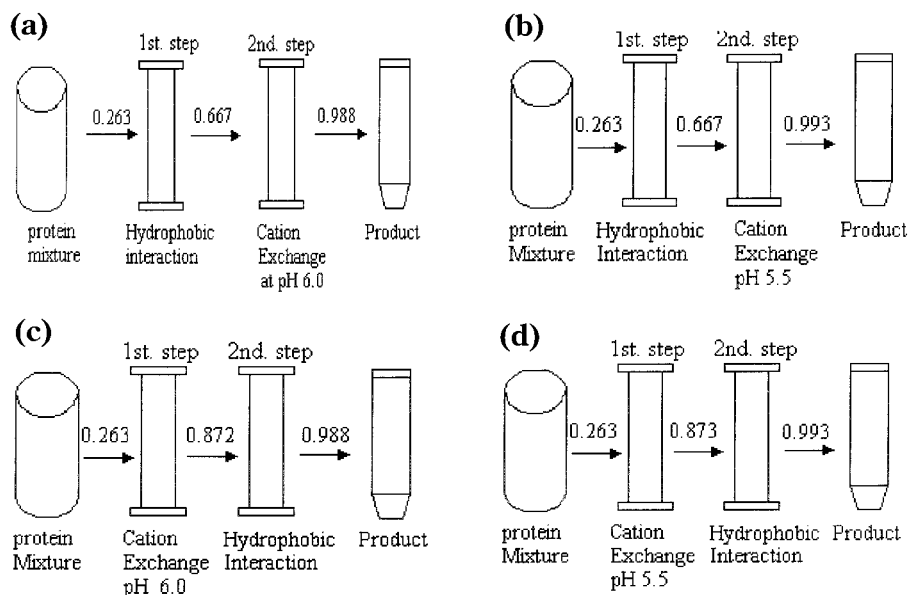
Figure 6. Optimal results for example 3: (a) model M_{1a}, (b) model M_{1b}, (c) model M_{2a}, and (d) model M_{2b}.
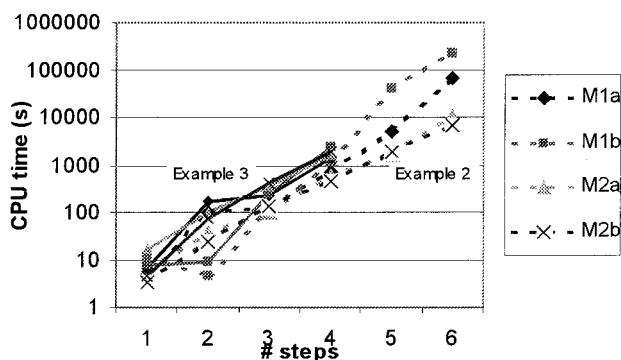
Figure 7. Variation of CPU time for examples 2 and 3.

ingly, the purification in Example 2 is much more difficult despite the smaller model size, which can be shown by the number of selected steps and from the CPU time. Variation of CPU time as a function of each step is better shown in Figure 7 for all models for Examples 2 and 3. It is important to note that while for models M_{1b} and M_{2b} the number of steps is a given parameter, in the remaining models it is directly the objective function value at the optimal solution.

6. Conclusions

This paper presented the development of mixed-integer linear models and optimization strategies for tackling the problem of synthesis of chromatographic steps for the purification of protein mixtures.

First, we proposed a mixed-integer linear programming model for minimizing the total number of steps in order to satisfy a given purity criteria. A second model was then developed in order to maximize the final purity of the desired protein for a given number of steps. Finally, in the cases for which sequencing is not an issue alternative models were developed for the optimal selection of steps that relied on the strategies previously described.

Real-world examples were tested with methodologies that resulted in large-scale MILP models of up to 276 binary variables, 155 continuous variables and 6387 constraints. Models were solved to optimality, and solutions were compared to the ones obtained from expert systems and experimental results. The results of the

proposed method are more reliable than the ones from the literature because it may attain higher levels of purity. These indicate the appropriate selection and sequencing of chromatographic steps, and they can be used as an important guideline for the synthesis and design of purification processes.

References and Notes

(1))Asenjo, J. A. *Profiles on Biotechnology*; Santiago de Compostela: Spain, 1992; pp 505−522.

(2) Wheelwright, S. M. Designing Downstream Processes for Large-Scale Protein Purification. *Bio/Technology* **1987**, *5*(8), 789−&.

(3) Asenjo, J. A. *Separation Processes in Biotechnology*; Marcel Dekker: New York, 1990.

(4) Luong, J. H. T.; Male, K. B.; Nguyen, A. L.; Mulchandani, A. Mathematical Modeling of Affinity Ultrafiltration Process. *Biotechnol. Bioeng.* **1988**, *32*, 451−459.

(5) Mao, Q. M.; Hearn, M. T. W. Optimization of Affinity and Ion Exchange Chromatographic Processes for the Purification of Proteins. *Biotechnol. Bioeng.* **1996**, *52*, 204−222.

(6) Jungbauer, A.; Kaltenbrunner, O. Fundamental Questions in Optimizing Ion-Exchange Chromatography of Proteins using Computer-Aided Process Design. *Biotechnol. Bioeng.* **1996**, *52*, 223−236.

(7) Jennings, L. S.; Teo, K. L.; Wang, F. Y.; Yu, Q. Optimal Protein Separation. *Comput. Chem. Eng.* **1995**, *19*(5), 567−573.

(8) Madden, J. E.; Haddad, P. R. Critical Comparison of Retention Models for Optimization of the Separation of Anions in Ion Chromatography. I. Non-Suppressed Anion Chromatography using Phthalate Eluents and Three Different Stationary Phases. *J. Chromatogr. A* **1998**, *829*, 65−80.

(9) Madden, J. E.; Haddad, P. R. Critical Comparison of Retention Models for Optimization of the Separation of Anions in Ion Chromatography, II. Suppressed Anion Chromatography using Carbonate Eluents. *J. Chromatogr. A* **1999**, *850*, 29−41.

(10) Porter, J. E.; Ladisch, M. R. Ion Exchange and Affinity Chromatography Cost in α-Galactosidase Purification. *Biotechnol. Bioeng.* **1992**, *39*, 717−724.

(11) Evangelista, R. L.; Kusnadi, A. R.; Howard, J. A.; Nikolov, Z. L. Process and Economic Evaluation of the Extraction and Purification of Recombinant β-Glucuronidase from Transgenic Corn. *Biotechnol. Prog.* **1998**, *14*, 607−614.

(12) Siletti, C. A.; Stephanopoulos, G. Computer Aided Design of Protein Recovery Process. *192nd ACS National Meeting*, Anaheim, CA, September, 1986.

(13) Asenjo, J. A. The rational Design of Large-Scale Protein Separation Sequences. *196th ACS National Meeting, MBTD Division*, Los Angeles, September, 1988.

(14) Stephanopoulos, G.; Mavrovouniotis, M. Artificial Intelligence in Chemical Engineering Research and Development. *Comput. Chem. Eng.* **1988**, *12*(*9/10*), R5−R6.

(15) Siletti, C. A. *Computer Aided Design of Protein Recovery Processes*, Ph.D. Thesis, Massachussets Institute of Technology, Cambridge, MA, 1989.

(16) Asenjo, J. A.; Maugeri, F. An Expert System for Selection and Synthesis of Protein Purification Processes. In *Frontiers in Bioprocessing II*; Todd, P., Sikdar, S. K., Bier, M., Eds.; American Chemical Society: Washington, DC, 1992; pp 358−379.

(17) Bryant, C. H.; Rowe, R. C. Knowledge Discovery in Databases: Application to Chromatography. *TrAC, Trends Anal. Chem.* **1998**, *17*, 18−24.

(18) Eriksson, H.; Sandahl, K.; Forslund, G.; Österlund, B. Knowledge-Based Planning for Protein Purification. *Chemom. Intell. Lab. Syst.* **1991**, *13*, 173−184.

(19) Forslund, G. Designing for Flexibility: a Case Study. *Expert Systems* **1995**, *12*(1), 27−37.

(20) Eriksson, H.; Sandahl, K.; Brewer, J.; Österlund, B. Reactive Planning for Chromatography. *Chemom. Intell. Lab. Syst.* **1991**, *13*, 185−194.

(21) Leser, E. W.; Asenjo, J. A. Building an Expert System to Assist the Rational Selection of Large-Scale Protein Purification Processes. In *Separations for Biotechnology III*; Pyle, D. L., Ed.; Royal Society of Chemistry: London, 1994; pp 260−266.

(22) Lienqueo, M. E.; Leser, E. W.; Asenjo, J. A. An Expert System for the Selection and Synthesis of Multistep Protein Separation Processes. *Comput. Chem. Eng.* **1996**, *20* (suppl), S189−S194.

(23) Lienqueo, M. E.; Salgado, J. C.; Asenjo, J. A. Design of an Expert System for Selection of Protein Purification Processes: Comparison Between Different Selection Criteria. *Comput. Appl. Biotechnol. CAB7*, **1998**, 321−324.

(24) Lienqueo, M. E.; Salgado, J. C.; Asenjo, J. A. An Expert System for Selection of Protein Purification Processes: Experimental Validation. *Chem. Technol. Biotechnol.* **1999**, *74*, 293−299.

(25) Pharmacia *IEF and electrophoretic titration curve analysis: Separation Technique File 100 PhastSystem*, Uppsala, 1991.

(26) Watanabe, E.; Tsoka, S. Asenjo, J. A. Selection of chromatographic protein purification operations based on physicochemical properties. *Annals N.Y. Acad. Sci.* **1994**, 721, 348−364.

(27) Leser, E. W. *Prot-Ex: an Expert System for Selecting the Sequence of Processes for the Downstream Purification of Proteins*; Ph.D. Thesis, The University of Reading: Reading, UK, 1996.

(28) Lienqueo, M. E. *Desarrollo de un Sistema Experto para la Seleccion Racional de Procesos de Purificacion de Proteínas: Optimizacion de Criterios de Seleccion de Secuencias*; Ph.D. Thesis, University of Chile: Santiago, Chile, 1999.

(29) Brooke, A.; Kendrick, D.; Meeraus A. GAMS *A User's Guide (release 2.25)*; The Scientific Press: San Francisco, CA, 1992.

(30) IBM *OSL (Optimization Subroutine Library) Guide and Reference (release 2)*; IBM: Kingston, NY, 1991.