
**Separation Science and Technology**
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713708471

# Experimental Validation of the Predictions of a Mathematical Model for Protein Purification and Tag Selection

M. E. Lienqueo[ab]; C. Shene[c]; A. Quiroga[ab]; O. Salazar[ab]; J. C. Salgado[abd]; J. A. Asenjo[ab]
[a] Centre for Biochemical Engineering and Biotechnology, University of Chile, Santiago, Chile [b] Institute for Cell Dynamics and Biotechnology: a Centre for Systems Biology, University of Chile, Santiago, Chile [c] Department of Chemical Engineering, Universidad de La Frontera, Chile [d] Laboratory of Process Modeling and Distributed Computing, Department of Chemical Engineering and Biotechnology, University of Chile, Santiago, Chile

Online publication date: 24 November 2010

**To cite this Article** Lienqueo, M. E. , Shene, C. , Quiroga, A. , Salazar, O. , Salgado, J. C. and Asenjo, J. A.(2010) 'Experimental Validation of the Predictions of a Mathematical Model for Protein Purification and Tag Selection', Separation Science and Technology, 45: 15, 2153 — 2164
**To link to this Article: DOI:** 10.1080/01496395.2010.507434
**URL:** http://dx.doi.org/10.1080/01496395.2010.507434

*Separation Science and Technology*, 45: 2153–2164, 2010

ISSN: 0149-6395 print/1520-5754 online
DOI: 10.1080/01496395.2010.507434

Taylor & Francis
Taylor & Francis Group

# Experimental Validation of the Predictions of a Mathematical Model for Protein Purification and Tag Selection

**M. E. Lienqueo,[1,2] C. Shene,[3] A. Quiroga,[1,2] O. Salazar,[1,2] J. C. Salgado,[1,2,4] and J. A. Asenjo[1,2]**

[1]*Centre for Biochemical Engineering and Biotechnology, University of Chile, Santiago, Chile*
[2]*Institute for Cell Dynamics and Biotechnology: a Centre for Systems Biology, University of Chile, Santiago, Chile*
[3]*Department of Chemical Engineering, Universidad de La Frontera, Chile*
[4]*Laboratory of Process Modeling and Distributed Computing, Department of Chemical Engineering and Biotechnology, University of Chile, Santiago, Chile*



**To improve downstream processing, protein engineering can be used to modify particular properties of a protein, such as specific affinity, charge, or hydrophobicity. The most common modification involves the peptide tags fusion to the protein. Nevertheless, the selection of both the optimal peptide tag and the right purification system to use is not trivial. The aim of this work is to experimentally validate our previous mathematical model based on MINLP models. This model was modified to find the minimum number of chromatographic operations, and the ideal tag for obtaining a required level of purification. A weighted linear combination of the number of purification steps and the target protein purity obtained after the last step was used as the objective function. The chromatographic steps suggested by the model were carried out using an example based on two mutated cutinases. The results show that average deviations between experimental data and those predicted by the model, for yield and purity, are ca. 15%, confirming that model predictions are reliable and could be used for selecting the best purification processes without experimental tests.**

**Keywords** cutinases; experimental validation; model; protein purification processes; tagged proteins

## INTRODUCTION

It is known that a rational design for protein purification is fundamental in order to fulfil economical and technical constraints. This rational approach is supported by an understanding of the separation phenomena involved in every purification stage as well as heuristic knowledge gathered in the laboratory or industrial plant. In fact, previous work has shown that meta-approaches that combine heuristic and phenomenological knowledge can be successfully used to design downstream purification processes (1–4). In those cases, the general aim is to propose a sequence of purification stages to purify the target protein to specific purity and recovery levels, subject to economical restrictions such as the number of purification steps and their cost.

The performance of a purification step (i.e., purity and recovery achieved by it) depends on exploiting differences of the physicochemical properties such as hydrophobicity, charge, size, and biological affinity, between the target protein and contaminants in the sample. In this sense, the largest difference between the separation property chosen between the target protein and a particular contaminant, will result in the most successful separation. Although, operational conditions could be set to maximize these differences, other methods involving the direct modification of a protein's physicochemical property, have proven to be effective (5). For instance, the modification of superficial hydrophobicity of a protein has been used to change its behavior in a hydrophobic interaction chromatography purification stage or its partition in aqueous two-phase systems. Such modifications can be achieved adding polypeptide tags by genetic engineering (6). These tags are amino acid sequences which are added to the C- or *N*-terminal of a protein to confer or improve a particular feature that increases selectivity in a given purification stage without compromising other biological or physicochemical characteristics of the target protein.

It has been reported that the addition of specific tags (7–10) improves the performance of purification stages. Examples of these tags are: poly-His-tag, poly-Arg-tag, calmodulin-binding peptide (CBP), cellulase-binding domain (CBD), DsbA, c-myc-tag, glutathione S-transferase

Received 17 December 2009; accepted 7 July 2010.
Address correspondence to M. E. Lienqueo, Department of Chemical Engineering and Biotechnology, University of Chile, Beauchef 850, Santiago, Chile. Tel.: 56-2-978 4709; Fax: 56-2-699 1084. E-mail: mlienque@ing.uchile.cl

(GST), FLAG-tag, HAT-tag, maltose-binding protein (MBP), NusA, S-tag, SBP-tag, Strep II-tag, thioredoxin, Biotin acceptor peptide (BAP), and many others. (6,11); short hydrophobic peptide tags, e.g., (WP)2, (WP)4, (Y)3(YP)3, (Y)3(P)2, (Y)4, (YP)4, (Y)6, (Y)6(P)2, (Y)8 have been also used (7,9,10). The use of polypeptide tags has many advantages, such as

a. fewer genetic modifications in the target product;
b. they have a minimum impact on the tertiary structure and biological activity of the fusion protein because of their small size;
c. it is relatively easy to remove the tag (a specific cleavage location may be included).

However, designing an optimal tag that improves the protein's behaviour in a purification stage is not an easy task.

### In Silico Polypeptide Tag Design

Brute force approaches such as assessing multiple tags generated randomly (6) are not practical because of the high computational requirements and also because they do not consider the characteristics of the target protein or the expression system. Regarding this, the development of rational methodologies to design polypeptide tags is of great interest. These methodologies should consider the physicochemical characteristics of the target protein and main contaminants, as well as the expression system. In addition, the design of the tag must be done considering its impact on the whole purification process and not only in a single purification stage.

Designing the polypeptide tag as well as the sequence of operations in the downstream purification process is a complex combinatorial problem. Heuristic methodologies have been used to support the synthesis of optimal bioprocesses (12–18); however, most of them do not consider the possibility of using a polypeptide tag to improve the purification process. Attempts have been made to incorporate the design of a polypeptide tag into the optimal bioprocess design: genetic algorithms (15) and mixed integer non-linear programming (MINLP) (19) have been used to solve design optimization problems which differ in the objective function, variations in the physicochemical properties and unit operation models, as well as in restrictions given by the characteristics of the tag.

Recently we proposed a MINLP model (20) based on the Simeonidis' model (19). The objective function of this model is the minimization of the number of purification stages and tag length subject to the achievement of a desired purity level. However, some of the tags designed by the Simeonidis methodology could produce protein loss in the purification stages, protein-protein interactions (10), or bad exposure to the solvent. Our model solves some of the Simeonidis' model caveats considering a finite number (26) of widely used small polypeptide tags and the loss of product mass in the purification stages. The objective function of this model is the maximization of the profit of the process; that is, to maximize the recovery of the desired protein and to minimize the cost due to the purification steps. The model was tested with experimental data for the protein cutinase, and the results showed that for a wide range of purity levels the tag FLAG, an 8-amino-acid peptide (DYKDDDDK), which slightly increases the charge and the hydrophobicity of the target protein, was the most selected.

In the present work we present an improved version of our previous model where a weighted combination of the number of purification steps and target protein purity after the last step was used as the objective function. Also, a constraint for restricting the final purity was considered in order not to exceed the specified value. For validating model predictions the purification of a cutinase modified through 2 tags ((WP)2 and (Y)3) was evaluated. Results predicted by this model for the yield and the purity of each purification step, are compared with experimental results. Deviations of model predictions and the effect of hydrophobic tags in the proposed purification sequences have been evaluated.

## THEORY

### Problem Description

The general problem can be stated as follows:
Given:

- A mixture of the target protein ($p=0$) and contaminants ($p=1,\ldots,P$) with known physicochemical and biochemical properties;
- A set of 11 available chromatographic techniques [anion exchange at pH 4, 5, 6, 7, 8, ($i=1\ldots5$); cation exchange at pH 4, 5, 6, 7, 8 ($i=6\ldots10$); and hydrophobic interaction chromatography ($i=11$)];
- The physicochemical and biochemical properties of well-known polypeptide tags ($n$: $1,\ldots,N$);
- Correlations for predicting time (KD) and concentration factors (CF) in each chromatographic technique; and
- A specified purity level for the desired product ($sp$).

The aim was to choose the best peptide tag from the set of those most widely used and the sequence of the high-resolution purification process in order to obtain a product with purity ideally equal to the specified value.

### Mathematical Formulation

The mathematical model with decision variables, based on Mixed Integer Non-Linear Programming (MINLP) formulation developed by Lienqueo and coworkers for the

synthesis of purification bioprocesses was used (20). The solution of the model gives the minimum sequence of purification steps needed to reach the specified purity of the particular tagged protein. The number of purification stages and its sequence is selected in order to minimize an objective function that takes into account the target protein purity (revenues) and the purification costs (number of purification steps).

## Objective Function

The objective function defined to find the tag and the minimum sequence of chromatographic steps to reach a final purity of the target protein is:

$$\min\left(-a_1\sum_{n=1}^{N}\mathrm{Purity_{n,KM}}x_n+\frac{a_2}{K_M}\sum_{n=1}^{N}\left(\sum_{k=1}^{K_M}\sum_{i=1}^{11}\mathrm{y_{n,i,k}}\right)x_n\right) \tag{1}$$

where $K_M$ is the maximum number of purification steps that could be selected; in our case this parameter was fixed at 6. Parameters $a_1$ and $a_2$ give a relative weight to the revenues obtained through final purity of the target protein and the purification costs estimated as the number of purification steps needed to accomplish the purification process, respectively. Values of the parameters $a_1$ and $a_2$ depend on the product (industrial or pharmaceutical), and its sum is constrained to one; in our case values were tested between 0.50–0.99 and 0.50–0.01, respectively. Another cost source is that due to the genetic modification of the protein. This is principally due to the primers used and will depend on the number of amino acids on the tag. However, this cost is incurred only once, when the protein is initially cloned, and thus it is considered negligible compared with the purification cost. Also, the cost related to tag cleavage and the re-purification step to remove the cleaved tag is considered the same for any tag.

The objective function is subjected to constraints for tag selection, mass balances, specifications of the final purity of target protein, and those related to the chromatographic sequence. In Eq. (1) $x$ and $y$ are decision variables that define the tag and the purification steps ($i$) for each tag ($n$) at each of the $k$ stages, respectively.

## Constraint Equations for Tag Selection

Within the set of possible tags one with no amino acids is found, which is equivalent to a protein without any tag. By considering this tagless alternative, the model is able to choose a purification operation (chromatography) by using techniques linked to the tagless protein product.

At each iteration the model finds optimal conditions for one of the $n$ tags, and thus,

$$\sum_{n=1}^{N}\mathrm{x}_n=1 \tag{2}$$

For that tag only, different purification stages are evaluated,

$$\sum_{i=1}^{11}y_{n,i,k}+\sum_{i=1}^{11}y_{n',i,k}\le 1\quad n'\neq n \tag{3}$$

An upper limit for the number of stages is imposed according to

$$\sum_{k=1}^{K_M}\sum_{i=1}^{11}y_{n,i,k}\le K_M\cdot x_n\quad\forall n \tag{4}$$

Purification process has to have at least one chromatographic step and this is found at the first stage. Thus,

$$\sum_{i=1}^{11}y_{n,i,1}=1\cdot x_n\quad\forall n \tag{5}$$

Each chromatographic technique is constrained to be used at most once in the sequence

$$\sum_{k=1}^{K_M}y_{n,i,k}\le 1\cdot x_n\quad\forall n,i \tag{6}$$

In order to find the last chromatographic step in the sequence a decision variable $\alpha_{n,k}$ is defined as,

$$\alpha_{n,k}=1-\sum_{i=1}^{11}y_{n,i,k}\quad\forall n,k \tag{7}$$

$\alpha_{n,k}$ is equal to 1 if at stage $k$ no chromatographic step is selected for the target protein labeled with tag $n$. On the other hand, if at stage $k$ the last chromatographic step in the sequence occurs, then the following constraint is satisfied:

$$\sum_{i=1}^{11}y_{n,i,k+1}-\sum_{i=1}^{11}y_{n,i,k}\le 0\quad k<K_M;\forall n \tag{8}$$

A vector z, with $K_M$ elements equal to 0 or 1 is defined; initially all elements in z are set to 0. If $k$ is the last stage at which a chromatographic step is selected for the protein labeled with tag $n$, then $z_{n,k}$ is set to 1. The following Eqs. (9–11) determine the position of the last chromatographic step in the sequence of $K_M$ steps:

$$\sum_{i=1}^{11}y_{n,i,k}-\sum_{i=1}^{11}y_{n,i,k+1}-z_{n,k}\le 0\quad k<K_M;\forall n \tag{9}$$

$$\sum_{i=1}^{11}y'_{n,i,k}-z_{n,k}\ge 0\quad k'\le k;\forall n \tag{10}$$

$$z_{n,k}+\sum_{i=1}^{11}y'_{n,i,k}\le 1\cdot x_n\quad k'>k;\forall n \tag{11}$$

For the specified tag, the sum of z elements must be equal to 1,

$$\sum_{k=1}^{K_M} z_{n,k} = 1 \cdot x_n; \quad \forall n \tag{12}$$

A constraint related to the selection of ionic exchange chromatography steps is imposed according to:

$$y_{n,i+1,k} + y_{n,i,k+1} \leq 1 \quad i < 4 \text{ or } 6 < i < 10; \quad k < K_M; \forall n \tag{13}$$

## Mass Balance Equations

Determination of the remaining amount of protein after each chromatographic step was accomplished using the convex hull representation proposed by Vasquez-Alvarez and Pinto (17). This representation is described by the following equations. Mass of protein $p$ in the mixture after a chromatographic step $i$ at stage $k$, when tag $n$ is used, $m_{n,p,k}$ is obtained from:

$$m_{n,p,k} = \sum_{i=1}^{11} CF_{n,p,i} \cdot m^1_{n,p,i,k} + m^2_{n,p,k} \quad \forall n, p, k \tag{14}$$

Where $m^1{}_{n,p,i,k}$ is the mass of protein $p$ obtained by using a chromatography step $i$ at stage $k$ and $CF_{n,p,i}$ is the concentration factor for protein $p$ labeled with tag $n$ by using the chromatography technique $i$, defined below. When no chromatography step is carried out at stage $k$ then $m_{n,p,k}$ is equal to $m^2_{n,p,k}$. In addition,

$$m_{n,p,k} = \sum_{i=1}^{11} m^1_{n,p,i,k+1} + m^2_{n,p,k+1} \quad k < K_M; \quad \forall n, p \tag{15}$$

For the proteins after the first stage ($k = 1$)

$$m^1_{n,p,i,1} = m_{p,0} \cdot y_{n,i,1} \quad k = 1; \forall n, p, i \tag{16a}$$

$$m^2_{n,p,1} = m_{p,0} \cdot \alpha_{n,1} \quad k = 1; \forall n, p \tag{16b}$$

Where $m_{p,0}$ is the mass of protein $p$ in the mixture subjected to purification. At other stages the following equations have to be satisfied,

$$m^1_{n,p,i,k} = m_{n,p,k-1} \cdot y_{n,i,k} \quad k \geq 2; \forall n, p, i \tag{17a}$$

$$m^2_{n,p,k} = m_{n,p,k-1} \cdot \alpha_{n,k} \quad k \geq 2; \forall n, p \tag{17b}$$

## Constraints for Reaching the Specified Purity

Purity of the target protein ($p0$) in the mixture after each stage is obtained from:

$$Purity_{n,k} = \frac{m_{n,p0,k}}{\sum_{p=1}^{P} m_{n,p,k}} x_n \quad \forall k \tag{18}$$

Purity of the target protein in the mixture after each stage has to increase and thus the following relationship has to be satisfied:

$$Purity_{n,k+1} - Purity_{n,k} \geq 0 \quad k < K_M; \forall n \tag{19}$$

Moreover, purity after the last stage is constrained to values in the following range defined by the specified value, *sp*.

$$0.95 \cdot sp \leq Purity_{n,K_M} \leq 1.05 \cdot sp \quad \forall n \tag{20}$$

By imposing this constraint a value much higher than that specified is avoided. In this way the final product will not be overspecified with the consequent cost associated with more purification efforts. Constraint in relationship (20) also considers that it is very unlikely to reach an exact value for the final purity.Target protein yield was computed from:

$$Yield_{n,p0} = \frac{m_{p0,0}}{m_{n,p0,K_M}} \quad \forall n \tag{21}$$

## Physicochemical Constraints

*Physicochemical Properties of the Tagged Protein*

Physicochemical properties of the tagged protein are calculated by adding the properties of the fused tag to the original protein, which means:

$$P_{dp} = \widehat{P}_{dp} + P_{tag} \tag{22}$$

Where $P_{dp}$, $\widehat{P}_{dp}$, and $P_{tag}$ are the physicochemical properties of the tagged protein, the desired (product) protein, and the polypeptide tag, respectively. This has been shown to be a valid assumption (21).

Table 1 shows how each physicochemical property is computed. Table 2 displays the properties of amino acids used in the calculation of physicochemical properties of the tagged protein.

*Model of Chromatographic Steps*

Four parameters are used for predicting the chromatographic behavior of each technique:

a. dimensionless retention times ($KD_{ip}$),
b. deviation factor ($DF_{i,p}$),
c. width of the chromatographic peak ($\sigma_i$), and
d. concentration factor ($CF_{i,p}$).

These parameters are described in Lienqueo et al. (20) and summarized in Table 3.

TABLE 1
Equations for physicochemical properties of the tagged protein

| Physicochemical properties | Equation | Reference |
|---|---|---|
| Net charge | $Q_{l,i,dp} = \widehat{Q}_{i,dp} + \sum_{k \in BA} \frac{n_{l,k}}{\frac{K_k}{[H^+]_i}+1} - \sum_{k \in AA} \frac{n_{l,k}}{\frac{[H^+]_i}{K_k}+1}$ (23) | (19,28) |

$Q_{l,i,dp}$ is the net charge of the tagged protein with tag $l$, $\widehat{Q}_{l,dp}$ is the net charge of the desired protein, $K_k$ is the ionisation constant (see Table 2), $[H^+]$ is the concentration of hydrogen ions and $n_{l,k}$ is the number of amino acids of class $k$ in each group in tag $l$. BA represents the amino acids belonging to the basic group (ie., Arg, His, Lys). AA represents acids belonging to the acidic group (ie., Asp, Cys, Glu, Tyr).

Hydrophobicity (19,29)

$$H_{l,dp} = \sum_{aa} \left( h_{aa} \cdot \frac{s_{aa}}{\widehat{S}_{dp}} \right) + \sum_{k} \left( h_{l,k} \cdot \left( \frac{s_{l,k} \cdot n_{l,k}}{\widehat{S}_{dp} + \sum_{k} (s_{l,k} \cdot n_{l,k})} \right) \right) \quad (24)$$

$H_{l,dp}$ is the hydrophobicity of the tagged protein with tag, $h_{aa}$ is the value of the hydrophobicity assigned to each amino acid $aa$ (see Table 2), $s_{aa}$ is the total exposed area of class $aa$ amino acids in the desired protein, $s_{l,k}$ is the exposed area of amino acid $k$ in the tag $l$ (these amino acids are assumed to have a fully exposed surface; see Table 2), $\widehat{S}_{dp}$ is the total surface of the original protein (without tag) and $n_{l,k}$ is the number of class k amino acids in the tag $l$.

Molecular weight (19)

$$\mathrm{MW_{dp}} = \widehat{\mathbf{MW}}_{\mathrm{dp}} \quad (25)$$

$MW_{i,dp}$ is the molecular weight of the tagged protein and $\widehat{MW}_{dp}$ is the molecular weight of the desired protein.



TABLE 2
Properties of amino acids

| Residue | Normalized hydrophobicity $h_{aa}$ (30) | Exposed surface area Gly-X-Gly $s_{l,k}$ [Å$^2$] (31) | pK (30) | $LF_{HI,1}$ (23) | $LF_{HI,2}$ (23) | $LF_{HI,3}$ (23) | $LF_{HI,4}$ (23) | $LF_{HI,5}$ (23) | $LF_{HI,6}$ (23) | $LF_{HI,7}$ (23) | $LF_{HI,8}$ (23) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Phe | 1.000 | 210 | | 9 | 16 | 23 | 30 | 37 | 44 | 51 | 58 |
| Met | 0.987 | 185 | | 9 | 16 | 23 | 30 | 37 | 44 | 51 | 58 |
| Ile | 0.967 | 175 | | 9 | 16 | 23 | 30 | 37 | 44 | 51 | 58 |
| Leu | 0.908 | 170 | | 9 | 16 | 23 | 30 | 37 | 44 | 51 | 58 |
| Cys | 0.819 | 135 | 8.30 | 9 | 16 | 23 | 30 | 37 | 44 | 51 | 58 |
| Trp | 0.775 | 255 | | 9 | 16 | 23 | 30 | 37 | 44 | 51 | 58 |
| Val | 0.770 | 155 | | 9 | 16 | 23 | 30 | 37 | 44 | 51 | 58 |
| Tyr | 0.484 | 230 | 10.95 | 6 | 10 | 14 | 19 | 23 | 27 | 32 | 36 |
| Ala | 0.391 | 115 | | 5 | 8 | 12 | 15 | 19 | 22 | 26 | 29 |
| His | 0.354 | 195 | 6.50 | 4 | 7 | 10 | 14 | 17 | 20 | 23 | 26 |
| Thr | 0.253 | 140 | | 3 | 5 | 8 | 10 | 12 | 14 | 17 | 19 |
| Gly | 0.252 | 75 | | 3 | 5 | 7 | 10 | 12 | 14 | 17 | 19 |
| Arg | 0.202 | 225 | 12.5 | 2 | 4 | 6 | 8 | 10 | 11 | 13 | 15 |
| Ser | 0.188 | 115 | | 2 | 4 | 6 | 7 | 9 | 11 | 12 | 14 |
| Gln | 0.151 | 180 | | 2 | 3 | 4 | 6 | 7 | 9 | 10 | 11 |
| Pro | 0.151 | 145 | | 2 | 3 | 4 | 6 | 7 | 9 | 10 | 11 |
| Asn | 0.125 | 160 | | 1 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Glu | 0.115 | 190 | 4.25 | 1 | 2 | 3 | 4 | 5 | 7 | 8 | 9 |
| Asp | 0.105 | 150 | 3.91 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Lys | 0.000 | 200 | 10.79 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

TABLE 3
Parameters for chromatographic techniques (20)

| Chromatographic technique | Anion exchange chromatography (AEC) |
|---|---|
| Dimensionless Retention Time ($KD_{i,p}$) | $If\ Q_{i,p} \geq 0, \quad KD_{i,p} = 0$<br>$If\ Q_{i,p} < 0, \quad KD_{i,p} = \frac{8826 \cdot \left\lvert \frac{Q_{i,p}}{MW_p} \right\rvert}{1 + 18845 \cdot \left\lvert \frac{Q_{i,p}}{MW_p} \right\rvert}$ (26) |
| Peak width[1] ($\sigma_i$). | $\sigma_i = 0.15$ (27) |
| Concentration factor for contaminant[2] | $CF_{AE/CE,i,p} = \frac{3.722}{3.727 + 0.579 \cdot e^{(54.410 \cdot DF_{i,p} - 2.176)}} + 0.019 \quad \forall i \in IE, p \neq dp$ (28) |
| Concentration factor for target protein | $CF_{AE/CE,dp} = 0.85 \quad \forall i$ (29) |
| Deviation factor[3] | $DF_{i,p} = \lvert KD_{i,dp} - KD_{i,p} \rvert \ \forall i, p \neq dp$ (30) |
| | Cation exchange chromatography (CEC) |
| Dimensionless Retention Time ($KD_{i,p}$) | $If\ Q_{i,p} \leq 0, \quad KD_{i,p} = 0$<br>$If\ Q_{i,p} > 0, \quad KD_{i,p} = \frac{7424 \cdot \left\lvert \frac{Q_{i,p}}{MW_p} \right\rvert}{1 + 20231 \cdot \left\lvert \frac{Q_{i,p}}{MW_p} \right\rvert}$ (31) |
| | These correlations were obtained by using bind-and-elution conditions; elution was obtained with an increasing NaCl gradient between 0.0–2.0 M |
| Peak width[1] ($\sigma_i$). | $\sigma_i = 0.15$ (27) |
| Concentration factor for contaminant[2] | $CF_{AE/CE,i,p} = \frac{3.722}{3.727 + 0.579 \cdot e^{(54.410 \cdot DF_{i,p} - 2.176)}} + 0.019 \quad \forall i \in IE, p \neq dp$ (28) |
| Concentration factor for target protein | $CF_{AE/CE,dp} = 0.85\ \forall i$ (29) |
| | Hydrophobic Interaction Chromatography (HIC) |
| Dimensionless Retention Time ($KD_{i,p}$) | $KD_{HI,dp} = -12.14 \cdot H_{dp}^2 + 12.07 \cdot H_{dp} - 1.74$ (32) |
| | These correlations were obtained by using bind-and-elution conditions; elution was obtained with a decreasing ammonium sulphate gradient between 2.0–0.0 M. |
| Peak width[1] ($\sigma_i$). | $\sigma_i = 0.22$ (33) |
| Concentration factor for contaminant[2] | $CF_{HI,p} = \frac{3.937}{3.933 + 0.105 \cdot e^{(36.005 \cdot DF_{HI,p} - 0.299)}} + 0.018 \quad \forall p \neq dp$ (34) |

(*Continued*)

TABLE 3
Continued

| Chromatographic technique | Anion exchange chromatography (AEC) |
|---|---|
| Concentration factor for desired protein[3] | $CF_{HI,dp} = 1 - \frac{\sum\left(\frac{n_k}{\sum n_k} \cdot LF_{HI,k}\right)}{100}$ (35) |
| | $LF_{HI,k}$ is the loss factor for each amino acid in hydrophobic interaction chromatography, values are shown in Table 2. |

[1]$\sigma_i$ is used for determining reduction of contaminants after applying a chromatographic technique $i$.
[2]CF is used for estimating the mass of contaminant $p$ after chromatographic step $i$ has been applied (20).
[3]DF represents the driving force of the separation process.
[4]It is assumed that the separation occurs with a product loss which depends upon the chromatographic techniques and amino acidic composition of the polypeptide tag.

## MATERIALS AND METHODS

### Materials

We considered the purification of two tagged cutinases expressed in *E. coli*. Cutinase_(WP)2 and cutinase_(Y)3. The product must be separated from seven main contaminants, denoted by cont 1–7. Physicochemical properties of cutinases and contaminants are shown in Table 4 (1,22). The physicochemical properties of cutinase were obtained using the PDB Code 1CEX (23).

### Plasmids and *E. coli* Strain

Plasmid pFCEX1 was derived from pET11a by cloning the cutinase gene (*cut*) from *Fusarium solani* fused to the signal sequence of the alkaline phosphatase (*phoA*) gene at the *N*-terminal of the mature cutinase gene (24) (*phoA-cut*). *E. coli* BL21(DE3) was used as host strain for the expression of the wild type and mutated proteins.



TABLE 4
Proteomic database for the main proteins in *Escherichia coli*. Adaptation (22)

| | | | | | $Q_{i,p}$[C/molecule $\times 10^{-17}$]* | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | Cutinases | | | | |
| Proteins | $m_{0,p}$ [mg/mL] | MWp[1] [Da] | Hp[2] | $\widehat{S}_{dp}$[3] | pH4 | pH5 | pH6 | pH7 | pH8 |
| Cutinases[4] | | 20625 | 0.29 | 8472 | 0.18 | 0.04 | 0.01 | 0.00 | 0.00 |
| Cutinase_(WP)2 | 29.36 | | | | | | | | |
| Cutinase_(Y)3 | 23.71 | | | | | | | | |
| Contaminant | $m_{0,p}$ [mg/mL] | MWp [Da] | Hp[5] | | pH4 | pH5 | pH6 | pH7 | pH8 |
| C1 | 11.29 | 18370 | 0.71 | – | 1.94 | −0.80 | −1.76 | −2.15 | −2.45 |
| C2 | 7.06 | 85570 | 0.48 | – | 2.35 | −1.17 | −2.83 | −3.50 | −3.68 |
| C3 | 4.63 | 53660 | 0.76 | – | 1.83 | 0.04 | −0.49 | −0.85 | −1.34 |
| C4 | 4.83 | 203000 | 0.36 | – | 4.08 | 0.04 | −1.92 | −3.07 | −4.98 |
| C5 | 2.48 | 69380 | 0.36 | – | 5.22 | 1.02 | −1.90 | −3.05 | −3.90 |
| C6 | 7.70 | 48320 | 0.48 | – | 3.96 | 1.12 | −1.36 | −1.00 | −1.59 |
| C7 | 6.05 | 114450 | 0.63 | – | 10.40 | 3.15 | 0.56 | −0.53 | −1.43 |

[1]Molecular weight was measured by SDS-PAGE with PhastGel media in Phast System.
[2]Hydrophobicity calculated using equation (24).
[3]$\widehat{S}_{dp}$ is the total surface of the original protein (without tag).
[4]Titration curve of cutinase (PDB Code 1CEX) was obtained using the program http://www.iut-arles.up.univ-mrs.fr/w3bb/d_abim/compo-p.html and for the mutants using Eq. (23).
[5]Hydrophobicity was measured by HIC using a phenyl-superose gel in an FPLC and a gradient elution from 2.0 M to 0.0 M $(NH_4)_2SO_4$ in 20 mM Tris buffer.
*Charge was measured by electrophoretic titration curve analysis with PhastGel IEF 3–9 in a Phast System.

### Construction of Fusion Proteins

Modified cutinases were constructed by addition of polypeptide tags to the *C*-terminal of the protein. Oligonucleotides encoding the corresponding tags were fused to the 3′ end of the cutinase gene by PCR. Primer sequences are shown in Table 5. Amplification products were digested with restriction enzymes *Nde*I and *BamH*I and ligated to vector pET11a digested with the same enzymes. Constructions were transformed into *E. coli* BL21(DE3) for the production of the modified cutinases. General DNA handling procedures were as described (25). All constructions were characterized by sequencing of the complete genes.

### Expression of the Tagged Cutinases

The expression of the wild-type and the mutated versions of *F. solani* cutinase gene in *E. coli* BL21(DE3) was performed according to Petersen and co-workers (24), except that cells were grown at 30°C and induction was carried out with 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG), at 18 or 25°C. Since the cutinase gene is cloned behind the signal peptide for the alkaline phosphatase (*phoA*) the gene product is directed to the periplasm of *E. coli*. For protein extraction, the periplasmic fraction was prepared by osmotic shock following the protocol described by Petersen and co-worker (24).

### Purification of the Tagged Cutinases

Cutinase purification was carried out using an *Ä*KTA purifier (GE Healthcare, Uppsala, Sweden). The experiments were performed at room temperature, using a flow-rate equal to 2.0 ml/min. As the first chromatographic step, the proteins were injected in a 8 ml Butyl Sepharose 6FF (100 mm × 10 mm ID) (GE Healthcare, Uppsala, Sweden) column with a decreasing salt gradient, from 1.0 to 0 M ammonium sulphate in 10 column volumes (CVs). For the second chromatographic step, pool fractions were injected in a 8 ml Q-sepharose FF (GE Healthcare, Uppsala, Sweden) column with Tris HCl 20 mM at pH 8.0 with an increasing salt gradient, from 0 to 1 M sodium chloride in 10 CVs. The column eluent was monitored at 280 nm and 2 ml fractions were collected and analyzed for protein concentration, cutinase activity, and sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE, 12%), where the cutinase amount were analyzed using ImageQuant TL (Amersham Biosciences). The column effluent fractions containing cutinase activity were pooled and injected in the next chromatographic step.

All buffers were filtered through 0.22-μm Millipore filters after preparation, and degassed with helium for 10 min.

TABLE 5
Primers used in this work for addition of the hydrophobic tags in *F. solani* cutinase

| Primer | Sequence (5′–3′) |
|---|---|
| Cut_For | ata cat atg aaa caa agt act att gca ctg gca ctc |
| Cut_(WP)2 | tat gga tcc tca cgg cca cgg cca agc aga acc |
| Cut_(Y)3 | tat gga tcc tca ata gta ata agc aga acc gcg |

Underlined triplets correspond to the codons for the hydrophobic amino acids added to the *C*-terminal of *F. solani* cutinase.

### Analytical Methods

Analytical methods for determining protein concentration and cutinase activity assay were carried out as described in (26).

### Software

The software General Algebraic Modelling System (GAMS) was used to implement the model and its solution. This program is used for optimization and has a number of solvers for linear, nonlinear, and mixed-integer programming applications. In this particular case we used the SBB solver, which is based on a combination of the standard branch and bound methods known from mixed integer linear programming and CONOPT as the nonlinear, programming solver. The model was solved in a computer with Intel(R) Core(TM)2 Duo CPU T5800 @ 2.00 GHz and 3069 MB RAM. CPU time for solving the problem is dependent on the number of tags evaluated. For validating the model's prediction the problem was solved 40 times in order to ensure that the optimal case was calculated.

## RESULTS AND DISCUSSION

The aim of this work was the experimental validation of the model proposed, then the deviation between the results predicted (yield and purity) by the model and the experimental results were evaluated.

For validation we studied the purification of two tagged cutinases, expressed in *E. coli*, cutinase_(WP)2 and cutinase_(Y)3. The initial purity level and the final purity level required were different for each tagged cutinase. The initial purity levels for cutinase_(Y)3 and cutinase_(WP)2 were 35% and 40% respectively; and the final purity levels required for cutinase_(Y)3 and cutinase_(WP)2 were 75% and 90%, respectively.

The model allowed to define the sequence of purification stages needed to reach the specified purity in both cases (see Table 6). These purification sequences were experimentally tested. The chromatograms for the purification sequence are illustrated in Figs. 1 and 3, the SDS-PAGE gels are displayed in Figs. 2 and 4 and the level of purity and yield obtained are shown in Table 6.

CPU time for solving the problem is dependent on the number of tags evaluated. For validating the model's prediction two tags were considered; the problem was solved

TABLE 6
Selected purification sequences: Comparison between the predicted and experimental values for yield and purity

| | Predicted | | Experimental | | Difference* | |
|---|---|---|---|---|---|---|
| Purification step | Purity % | Yield % | Purity % | Yield % | Purity % | Yield % |
| Cutinase_(y)3 Final Purity level required: 75% | | | | | | |
| Osmotic stock | | | 35.0 | 100 | | |
| Hydrophobic Interaction | 79.0 | 86.0 | 83.3 | 75.8 | 5.2 | \|−13.5\| |
| Average Cutinase_(y)3** | | | | | 9.3 | |
| Cutinase_(wp)2 Final Purity level required: 90% | | | | | | |
| Osmotic stock | | | 40.0 | 100 | | |
| Hydrophobic Interaction | 72.7 | 90.4 | 95.8 | 76.9 | 24.1 | \|−17.6\| |
| Anion Exchange at pH 8.0 | 94.0 | 76.9 | 97.9 | 60.6 | 4.0 | \|−26.9\| |
| Average | | | | | 14.0 | \|−22.2\| |
| Average Cutinase_(WP)2** | | | | | | 18.1 |
| Average Cutinases*** | | | | | 11.1 | \|−19.3\| |
| Total Average*** | | | | | 15.2 | |

*$\text{Difference} = \left|\frac{\text{Experimental value} - \text{Predicted value}}{\text{Experimental value}}\right| \times 100$ (36).

**Average between difference of yield and difference of purity.

***Average between Average Cutinase_(wp)2 and Average Cutinase_(y)3.



40 times in order to ensure the optimal case was calculated. In one case 4364 iterations were carried out in 9.12 NLP seconds, for tag selection. In the same problem 3988 iterations were carried out with the tagless protein in 9.01 NPL seconds. Thus the results were obtained in less than 20 secs.

## Purification of Cutinase_(y)3

For the case of cutinase_(y)3 the final purity level required was 75%. The sequence suggested by the model was only one step, Hydrophobic Interaction Chromatography (HIC). In this step, cutinase activity was eluted between 74 and 90 ml (Pool A, see Fig. 1). The amount of contaminants decreased significantly and only one major

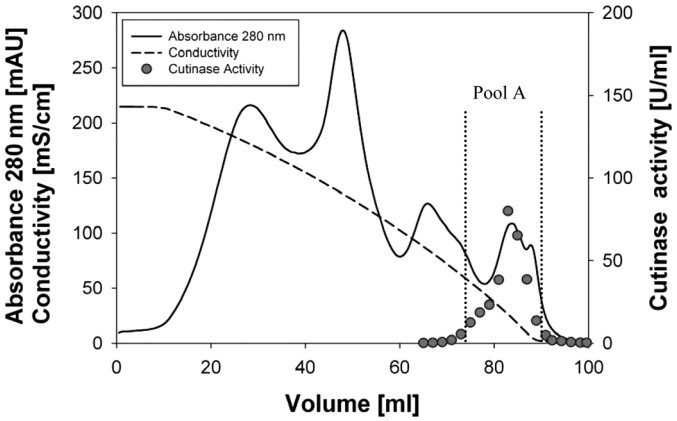



FIG. 1. The elution profile of the sample containing of cutinase_(y)3 from *E. coli* on a 8 ml Butyl Sepharose 6FF (100 mm × 10 mm ID). The flow was 2.0 ml/min. The protein was eluted by a linear gradient of 1.0 to 0 M ammonium sulphate in 10 CVs. Fraction of 2 ml were collected and tested for enzyme activity. (——) Absorbance 280 nm, (· · · · ·) Conductivity (−o−); Cutinase activity.

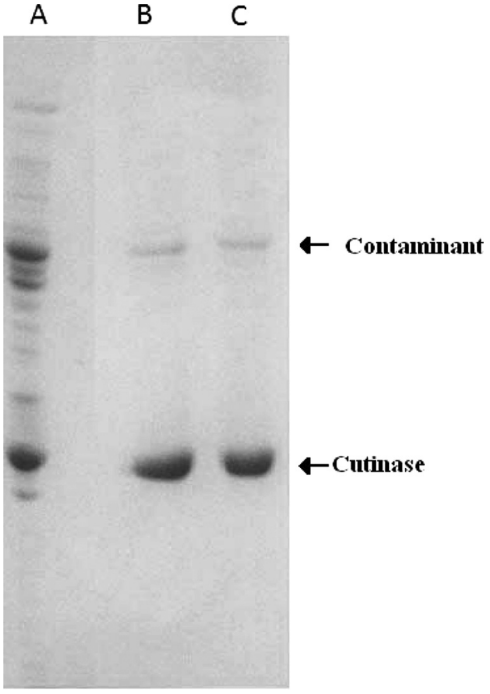
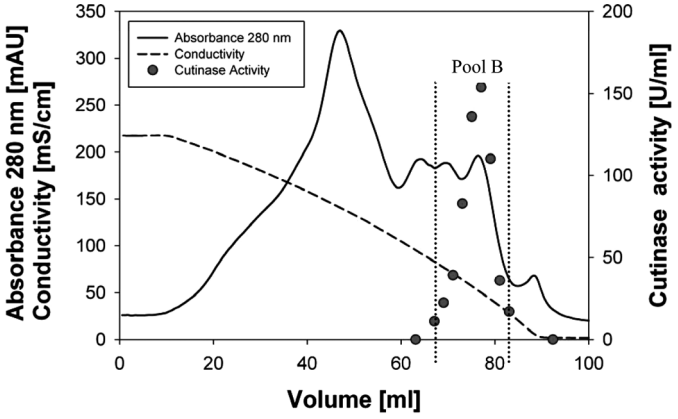
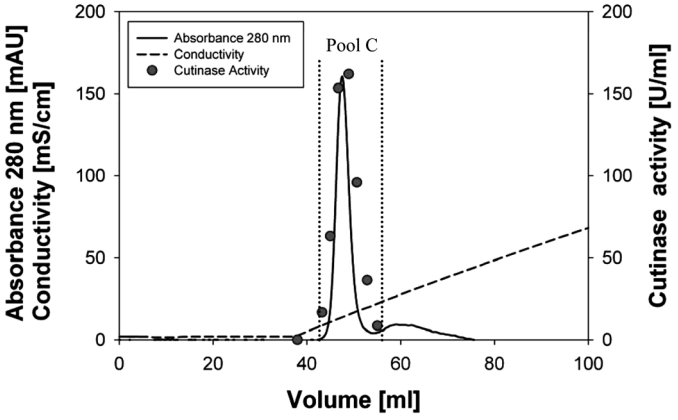



FIG. 2. SDS-PAGE of cutinase_(y)3. **A**: Original fraction of cutinase_(y)3 producing *E. coli* culture; **B and C**: Pool A, original fraction of cutinase_ (y)3 after HIC (see Fig. 1).




(a)




(b)

FIG. 3. (a) The elution profile the sample containing of cutinase_(WP)2 from *E.coli* on a 8 ml Butyl Sepharose 6FF (100 mm × 10 mm ID) (First step suggested by the model). The flow was 2.0 ml/min. The protein was eluted by a linear gradient of 1.0 to 0 M ammonium sulphate in 10 CVs. Fraction of 2 ml were collected and tested for enzyme activity. (—) Absorbance 280 nm; (· · · · ·) Conductivity; (–o–) Cutinase activity; (b) Purification of the pooled fractions containing cutinase_(WP)2 activity. Second steps suggested by the model for the purification: Anion exchange chromatography at pH 8.0 on 8 ml Q-sepharose FF (100 mm × 10 mm ID). The flow was 2.0 ml/min. The protein was eluted by a linear gradient of 0 to 1 M sodium chloride in 10 CVs. Fraction of 2 ml were collected and tested for enzyme activity. (—) Absorbance 280 nm; (· · · · ·) Conductivity; (–o–) Cutinase activity.

contaminant was detected by SDS-PAGE (see Fig. 2). In this gel, a low number of contaminant proteins are observed in the original fraction. An explanation for this is that, given the high yield of cutinase production by the recombinant *E. coli* strain (about 35% of total protein is cutinase), the sample loaded into the gel was very diluted.

Protein concentration and cutinase activity were analyzed in all fractions, and the purity and recovery were calculated after the HIC step, the values are shown in Table 6. Regarding purity, the comparison between the experimental and predicted values gave a deviation of 5.2%, the model underestimated the value of purity. For the yield,

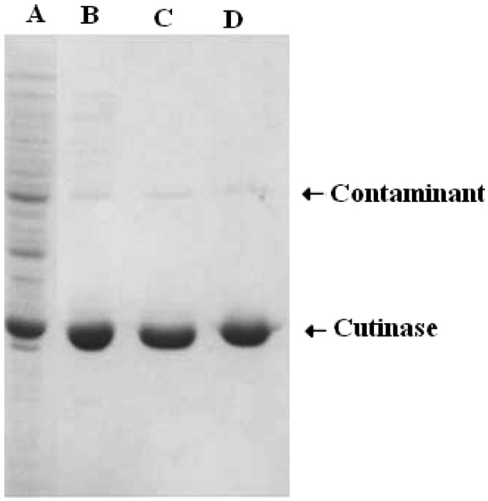



FIG. 4. SDS-PAGE of cutinase_(wp)2. A: Original cutinase_(WP)2 producing *E. coli* culture; B: Pool B, original cutinase_(WP)2 after HIC (see Fig. 3a).; C and D Pool C, pool B after AEC pH 8.0 (see Fig. 3b).

the absolute difference between the experimental and predicted values was 13.5%, the model overestimated the value of yield. Therefore, the average difference (yield and purity) between the predicted and experimental values was c.a. 10.0%.

## Purification of Cutinase_(WP)2

For the case of cutinase_(WP)2 the final purity level required was 90%. For the case of cutinase_(WP)2 the sequence suggested by the model was

a. first step HIC and
b. second step Anion Exchange Chromatography (AEC) at pH 8.0.

The cutinase activity and protein concentration of all fractions were analyzed.

In the case of the first step, HIC, cutinase activity was detected between 68 and 82 ml (Pool B see Fig. 3a), several minor contaminants were detected by SDS-PAGE gel (see Fig. 4). This pool was injected into the second step, AEC at pH 8.0. Here, the cutinase activity was detected between 42 and 55 ml (Pool C, see Fig. 3b); and only one main contaminant was detected by SDS-PAGE gel (see Fig. 4).

Protein concentration and cutinase activity were analyzed in all fractions; purity and recovery were calculated after each step, and the values are shown in Table 6. In the case of purity the comparison between the experimental and predicted values gave a deviation less than 14%, the model underestimated the value of purity. In the case of yield the absolute average deviation between the experimental and predicted values was less than 23%, the model overestimated the value of yield. Then, the average difference between the predicted and experimental values (yield and purity) was c.a. 18%.

Considering both cutinases, in the case of purity the average variation between the predicted and experimental values was 11%. In the case of yield, the absolute average difference between the experimental and predicted values was 19%. Then, the average difference between the predicted and experimental values was ca. 15%.

In both examples the final purity level required was slightly underestimated. However, differences between specified purity level for the desired product (*sp*) and calculated values of final purity are on average less than 4.5%; thus constraint in (20) is active. Because the purification process is not a continuous but a discrete series of operations, it is expected that predicted values for purity and yield do not coincide exactly with the specified values. An improvement in the model's prediction can be obtained by using correlations that consider the effect of operational conditions (i.e., flow and gradient step) in the chromatographic separations.

In both cases calculated yields were overestimated. Yield is a parameter that is affected by operations conditions mainly by the collection criterion defined as the percentage of the peak as to where to start and end the collection (27). In the model yield is calculated from the value of the concentration factor, CF, which for the target protein is fixed at 0.85 in IEC. On the other hand, in HIC the model assumes that the product loss depends on the amino acidic composition of the polypeptide tag; therefore it is necessary to improve the estimation of product loss; i.e., to derive better correlations for calculating the CF for the target protein considering operational conditions of each chromatographic technique (e.g flow, gradient step) and possible inactivation of the enzymes.



### Sensitivity Analysis of Weight Factors ($a_1$ and $a_2$) in the Objective Function

Weight factors in the objective function should be defined depending on the specific purification problem. In our case, the final purity was assumed to be more relevant than the purification costs estimated from the number of purification stages, thus the value of the parameter $a_1$ (0.99) was much higher than that of $a_2$ (0.01). Additionally, the model was run with values for $a_1$ between 0.5 and 0.99, and still the same results were obtained. From these results it can be concluded that for the problem under study (only one tag) the solution is less dependent on the number of chromatographic steps because only one combination is able to give the desired purity by using a minimum number of steps.

## CONCLUSIONS

The new constraints included in the model permitted to constraint the maximum number of purification steps, that in the present study was fixed at 6 ($K_M = 6$), which is normally used in industrial processes; and to minimize deviation to the final purity level required. However, because available correlations do not take into account the effect of operational conditions, it is difficult to obtain a specific value. Future work will focus on the improvement of the model predictions through new correlations that consider different operational conditions for minimizing the overestimation and underestimation of yield and purity, respectively.

The chromatographic sequence suggested by the model was experimentally tested for two mutant cutinases, cutinase_(Y)3 and cutinase_(WP)2. As it was expected because of the addition of hydrophobic tags, the model was able to select for these mutated protein HIC as one of the purification steps. The results show that average deviations between experimental data and those predicted by the model, both for yield and purity, are ca. 15%, which is a very acceptable deviation. More mutant proteins would be needed in order to validate the tag selection which is beyond the scope of the present work. Nevertheless the results obtained so far, and presented in this paper, confirm that our model predictions are reliable and could be used for optimizing purification processes without experimental tests. Additionally, the model could be used for evaluating customer designed tags.

## ACKNOWLEDGEMENTS

This work was supported by FONDECYT project 108143, FONDECYT Research Initiation grant 11080016 and the Millennium Scientific Initiative ICM project P05-001-F. Plasmid pFCEX1 was a generous gift of Dr. Evamaria I. Petersen, Dept. of Physics and Nanotechnology, University of Aalborg, Denmark.

## REFERENCES

1. Leser, E.W.; Lienqueo, M.E.; Asenjo, J.A. (1996) Implementation in an expert system of a selection rationale for purification processes for recombinant proteins. *Ann N Y Acad Sci.*, 782: 441.
2. Lienqueo, M.E.; Salgado, J.C.; Asenjo, J.A. (1999) An expert system for selection of protein purification processes: Experimental validation. *Journal of Chemical Technology and Biotechnology*, 74 (3): 293.
3. Lienqueo, M.E.; Asenjo, J.A. (2000) Use of expert systems for the synthesis of downstream protein processes. *Comput. Chem. Eng.*, 24: 2339.
4. Asenjo, J.A.; Andrews, B.A. (2004) Is there a rational method to purify proteins? From expert systems to proteomics. *J. Molecular Recognition*, 17 (3): 236.
5. Nygren, P.A.; Stahl, S.; Uhlen, M. (1994) Engineering proteins to facilitate bioprocessing. *Trends Biotechnol.*, 12: 184–188.
6. Terpe, K. (2003) Overview of tag protein fusions: From molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.*, 60: 523.
7. Bernaudat, F.; Bulow, L. (2006) Combined hydrophobic-metal binding fusion tags for applications in aqueous two-phase partitioning. *Protein Expression and Purification*, 46: 438.
8. Kepka, C.; Collet, E.; Roos, F.; Tjerneld, F.; Veide, A. (2005) Two-step recovery process for tryptophan tagged cutinase: Interfacing aqueous two-phase extraction and hydrophobic interaction chromatography. *J. Chromatogr. A*, 1075: 33.

9. Nilsson, A.; Mannesse, M.; Egmond, M.R.; Tjerneld, F. (2002) Cutinase-peptide fusions in thermoseparating aqueous two-phase systems. Prediction of partitioning and enhanced tag efficiency by detergent addition. *J Chromatogr A*, 946 (1–2): 141.
10. Fexby, S.; Ihre, H.; Van Alstine, J.; Bulow, L. (2004) N-Terminal tagged lactate dehydrogenase proteins: Evaluation of relative hydrophobicity by hydrophobic interaction chromatography and aqueous two-phase system partition. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.*, 807: 25.
11. Esposito, D.; Chatterjee, D.K. (2006) Enhancement of soluble protein expression through the use of fusion tags. *Curr Opin Biotechnol.*, 17 (4): 353.
12. Lakhdar, K.; Farid, S.S.; Titchener-Hooker, N.J.; Papageorgiou, L.G. (2006) Medium term planning of biopharmaceutical manufacture with uncertain fermentation titers. *Biotechnol. Prog.*, 22: 1630.
13. Lakhdar, K.; Zhou, Y.; Savery, J.; Titchener-Hooker, N.J. Papageorgiou, L.G. (2005) Medium term planning of biopharmaceutical manufacture using mathematical programming. *Biotechnol. Prog.*, 21: 1478.
14. Leser, E.W.; Asenjo, J.A. (1994) The rational selection of purification processes for proteins: An expert system for downstream processing design. *Ann N Y Acad Sci.*, 721: 337.
15. Steffens, M.A.; Fraga, E.S.; Bogle, I.D.L. (2000) Synthesis of purification tags for optimal downstream processing. *Comput. Chem. Eng.*, 24: 717.
16. Vásquez-Alvarez, E.; Lienqueo, M.E.; Pinto, J.M. (2001) Optimal synthesis of protein purification processes. *Biotechnol Prog.*, 17 (4): 685.
17. Vasquez-Alvarez, E.; Pinto, J.M. (2004) Efficient MILP formulations for the optimal synthesis of chromatographic protein purification processes. *J. Biotechnol.*, 110: 295.
18. Groep, M.E.; Gregory, M.E.; Kershenbaum, L.S.; Bogle, I.D.L. (2000) Performance modeling and simulation of biochemical process sequences with interacting unit operations. *Biotechnol. Bioeng*, 67: 300.
19. Simeonidis, E.; Pinto, J.M.; Lienqueo, M.E.; Tsoka, S.; Papageorgiou, L.G. (2005) MINLP models for the synthesis of optimal peptide tags and downstream protein processing. *Biotechnol. Prog.*, 21: 875.
20. Lienqueo, M.E.; Salgado, J.C.; Giaverini, O.; Asenjo, J.A. (2009) Computer-aided design to select optimal polypeptide tags to assist the purification of recombinant proteins. *Sep. Purif. Technol.*, 65: 86.
21. Lienqueo, M.E.; Salazar, O.; Henriquez, K.; Calado, C.R.; Fonseca, L.P.; Cabral, J.M. (2007) Prediction of retention time of cutinases tagged with hydrophobic peptides in hydrophobic interaction chromatography. *J Chromatogr A.*, 1154 (1–2): 460.
22. Woolston, P.W. (1994) Database for an expert system for the selection of recombinant protein purification processes. PhD Thesis, The University of Reading, Reading, England.
23. Giaverini, O. (2005) Optimization of polypeptide tags for protein purification, Engineer Thesis, University of Chile, Santiago, Chile.
24. Petersen, S.B.; Jonson, P.H.; Fojan, P.; Petersen, E.I.; Petersen, M.T.N.; Hansen, S.; Ishak, R.J.; Hough, E. (1998) Protein engineering the surface of enzymes. *Journal of Biotechnology*, 66: 11.
25. Sambrook, J.; Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd Ed.; Cold Spring Harbor Laboratory Press.
26. Calado, C.R.C.; Taipa, M.A.; Cabral, J.M.S.; Fonseca, L.P. (2002) Optimisation of culture conditions and characterisation of cutinase produced by recombinant Saccharomyces cerevisiae. *Enzyme Microb. Technol.*, 31: 161.
27. Orellana, C.A.; Shene, C.; Asenjo, J.A. (2009) Mathematical modeling of elution curves for a protein mixture in ion exchange chromatography applied to high protein concentration. *Biotechnol. Bioeng.*, 107: 572.
28. Mosher, R.A.; Gebauer, P.; Thormann, W. (1993) Computer-simulation and experimental validation of the electrophoretic behavior of proteins .3. Use of titration data predicted by the proteins amino-acid-composition. *J. Chromatogr.*, 638: 155.
29. Lienqueo, M.E.; Mahn, A.; Asenjo, J.A. (2002) Mathematical correlations for predicting protein retention times in hydrophobic interaction chromatography. *J Chromatogr A.*, 978 (1–2): 71.
30. Miyazawa, S.; Jernigan, R. (1985) Estimation of effective inter residue contact energies from protein crystal structures. Quasi-chemical approximation. *Macromolecules*, 18: 534.
31. Miller, S.; Janin, J.; Lesk, A.M.; Chothia, C. (1987) Interior and surface of monomeric proteins. *J. Mol. Biol.*, 196 (3): 641.

## APPENDIX

$a_1$ : Relative weight for revenues.
$a_2$ : Relative weight for purification cost.
$K_M$ : Maximum number of purification steps that could be selected (6).
P : Number of contaminants.
N : Number of tags.
sp : Desired purity
$x_n$ : Decision variable equal to 1 if the tag $n$ is selected.
$y_{n,i,k}$ : Decision variable (matrix) equal to 1 if the purification step $i$ for tag n at stage $k$ is selected.
Z : Decision variable (size $K_M$), If $k$ is the last stage at which a chromatographic step is selected for the protein labeled with tag $n$, then $z_{n,k}$ is set to 1.
$\alpha_{n,k}$ : Decision variable equal to 1 if at stage $k$ no chromatographic step is selected for the target protein labeled with tag $n$.
$Purity_{n,k}$ : Purity of the target protein labeled with tag n in the mixture after stage $k$.
$m_{n,p,k}$ : Mass of protein $p$ in the mixture after a chromatographic step $i$ at stage $k$ when tag $n$ is used.
$m_{p,0}$ : Mass of protein $p$ in the mixture subjected to purification
$CF_{n,p,i}$ : Concentration factor for protein $p$ labeled with tag $n$ by using chromatography technique $i$.
$KD_{i,p}$ : Dimensionless retention time for protein $p$ by using chromatography technique $i$.
$DF_{i,p}$ : Deviation factor for protein $p$ by using chromatography technique $i$.
$\sigma_i$ : Width of the chromatographic peak chromatography technique $i$.
$P_{dp}$ : Physicochemical properties of tagged protein.
$\widehat{P}_{dp}$ : Physicochemical properties of desired (product) protein.
$P_{tag}$ : Physicochemical properties of polypeptide tag.

### Indexes

$P$ : Protein, p = 0/contaminant, p = 1, 2, . . . , P.
$i$ : Chromatographic technique $I = 1, 2, \ldots, 11$.
$N$ : Tag, n = 1,2, . . . , N.
$K$ : Purification stage k = 1, 2, . . . , $K_M$.