

MINLP Models for the Synthesis of Optimal Peptide Tags and Downstream Protein Processing

Evangelos Simeonidis,[†] Jose M. Pinto,[‡] M. Elena Lienqueo,[§] Sophia Tsoka,^{||} and Lazaros G. Papageorgiou^{*,†}

Centre for Process Systems Engineering, Department of Chemical Engineering, UCL (University College London), London WC1E 7JE, U.K., Department of Chemical and Biological Sciences and Engineering, Polytechnic University, Brooklyn, New York 11201 and Department of Chemical Engineering, University of São Paulo, São Paulo, SP 05508-900, Brazil, Centre for Biochemical Engineering and Biotechnology, Department of Chemical Engineering, Millennium Institute for Advanced Studies in Cell Biology and Biotechnology, University of Chile, Beauchef 861, Santiago, Chile, and Computational Genomics Group, Research Program, The European Bioinformatics Institute, EMBL Cambridge Outstation, Cambridge CB10 1SD, U.K.

The development of systematic methods for the synthesis of downstream protein processing operations has seen growing interest in recent years, as purification is often the most complex and costly stage in biochemical production plants. The objective of the work presented here is to develop mathematical models based on mixed integer optimization techniques, which integrate the selection of optimal peptide purification tags into an established framework for the synthesis of protein purification processes. Peptide tags are comparatively short sequences of amino acids fused onto the protein product, capable of reducing the required purification steps. The methodology is illustrated through its application on two example protein mixtures involving up to 13 contaminants and a set of 11 candidate chromatographic steps. The results are indicative of the benefits resulting by the appropriate use of peptide tags in purification processes and provide a guideline for both optimal tag design and downstream process synthesis.

1. Introduction

Downstream processing for the purification of a protein product in biochemical production plants largely determines the quality of the product and is also a major component of the manufacturing and investment costs (1). Of all separation methods during the downstream process, chromatographic operations are of major interest in the production of high-value biomolecules. The specified purity level of the target protein product is usually achieved by applying several chromatographic steps. Such flowsheets are usually optimized on a unit per unit basis, thus creating the need for a more systematic synthesis and design procedure for purification sequences that considers the entire process instead of each unit individually.

A number of bioprocess synthesis and design methods have been reported, employing expert knowledge to select unit operations in order to synthesize economically favorable processes (2–4). Recently, methodologies based on optimization techniques have been presented. An implicit enumeration algorithm developed for chemical processes (5) was extended in order to synthesize optimal bioprocesses in a system-wide sense (6). The technique incorporated heuristics based on physical property in-

formation in an implicit enumeration algorithm to solve the synthesis problem, thus reducing the search space. A mixed integer linear programming (MILP) framework using established criteria for modeling chromatographic techniques was presented (7), in which mathematical models for each technique rely on physicochemical data on the protein mixture that contains the desired product and provide information on its potential purification. The latter formulation was further improved by exploiting the advantages of convex hull representations (8, 9) and by also considering the incorporation of product losses and the calculation of the amount of product recovered (10). The above methodologies can improve the production of a biotechnological product by optimizing the purification sequence on the basis of physicochemical data for the product and the contaminant proteins. However, they do not consider whether any benefit is conferred by modifying these physicochemical properties of the product to enhance the separation and thereby reduce the number of required downstream purification steps.

Such physicochemical properties modifications can be accomplished through the use of purification peptide tags that are genetically fused onto the protein product. Purification tags are short sequences of amino acids that utilize a specific structural protein property to facilitate purification: affinity, charge, attraction to metal chelates, solubility or hydrophobicity. It has recently been demonstrated that considerable improvements in yields and costs of downstream purification processes can be achieved with the use of such tags (11).

Sizes of purification tags range from full enzymes, such as β -galactosidase, which can be fused onto protein

* To whom correspondence should be addressed. Tel: +44 20 7679 2563. Fax: +44 20 7383 2348. E-mail: l.papageorgiou@ucl.ac.uk.

[†] University College London.

[‡] Polytechnic University and University of São Paulo.

[§] University of Chile.

^{||} The European Bioinformatics Institute.

products and can be usually purified using a specific affinity interaction, to very short amino acid sequences (e.g., poly-his tag), for which a particular physicochemical property is exploited to accomplish separation. The latter case of small peptide tags presents numerous advantages (12), for example, the need of only minor genetic modifications to the protein product, as these are small molecules. They are assumed to have a minimal effect on tertiary structure and biological activity and may not require cleavage for many applications due to their small size. The most commonly used small peptide tags are the poly-arginine tag or arg tag (5 amino acids), for purification with cation exchange (13); the poly-histidine tag or his tag (6 amino acids), for immobilized metal affinity chromatography, a widely used tagging technique (14); the FLAG tag, a small (8 amino acids) hydrophilic tag (15); the Strep tag (8 amino acids) (16); the c-myc tag (11 amino acids), a tag commonly used as a detection system but rarely applied for purification purposes (17); and the S tag (15 amino acids) (18).

The advantages of purification tags are well-recognized; nevertheless selecting a purification tag that is optimal in a generic sense is a challenging task. Although there is a relative abundance of previous research in the use of recombinant technology to improve separation characteristics of protein products (12, 19–21), this has mainly focused on specific tags, which have advantages in certain situations but are not necessarily optimal. Only one study examines the development of a framework for selecting peptide tags in protein purification (11); however, predictions are based on a single physicochemical property. The need arises for a systematic methodology that selects the most advantageous peptide tag and the appropriate steps to achieve the required purity, while taking into account a multitude of protein product physicochemical properties.

The aim of this paper is to develop a mixed integer non-linear programming (MINLP) framework that considers simultaneously the design of optimal peptide tags for each particular protein product and the synthesis of downstream protein processing. The systematic framework presented herein exploits the advantages of integer optimization, considers the manipulation of two protein properties, and can be expanded to more than two physicochemical features given that these are available. Physicochemical property data are used to specify the amino acid composition of the most advantageous and shortest tag and concurrently select operations among a set of candidate chromatographic techniques that must achieve a specified purity level, while optimizing a suitable performance criterion (e.g., minimization of purification steps).

Next, the problem of optimal tag design and synthesis of downstream processing is defined. In section 3, a mathematical programming formulation is presented. The applicability of the resulting MINLP models is demonstrated in section 4 through two illustrative examples. Finally, some concluding remarks are made in section 5.

2. Problem Statement

Overall, the problem of simultaneous optimal tag design and synthesis of downstream protein processing can be stated as follows:

Given:

- a mixture of proteins (p : 1, ..., P) with known physicochemical properties;

- a set of available chromatographic techniques (i : 1, ..., I) each performing a separation by exploiting a specific physicochemical property (charge or hydrophobicity);
- the properties of the 20 amino acids (k : 1, ..., 20); and
- a specification for the desired product (dp) in terms of a minimum purity level.

Determine:

- the amino acid composition of the shortest and most advantageous peptide tag;
- the physicochemical properties of the tagged protein (desired product + tag); and
- the flowsheet of the high-resolution purification process.

So as to optimize a suitable performance criterion.

To solve the problem, a few assumptions need to be made. Physicochemical properties of the tagged protein are assumed to be calculated by adding the properties of the tag to the ones of the original protein. The amino acids that comprise the fused tags are assumed to have a fully exposed surface. The possibility of the tag burying itself into the protein is avoided by imposing an upper bound to the number of hydrophobic residues that may be included in the tag. An upper bound is imposed on the number of amino acids that can be present in the peptide tag, so as to avoid interference with the tertiary structure of the protein. At the same time, the formation of a secondary structure (e.g., an α -helix or a β -sheet) from the tag itself should also be avoided; therefore the number of amino acids in the tag should not be larger than 6 or 7 (22). The overall molecular weight of the protein product is assumed to remain constant after the addition of the tag, as the combined molecular weight of a few amino acids is negligible compared to the one of the protein product. The methodologies used herein for the prediction of the physicochemical properties of the fused protein are theoretical estimations, nevertheless they are considered to be sufficiently adequate indications of the alteration of the property in question.

For process synthesis, it is assumed that the protein product is separated completely without any product loss, i.e., no product is left over in the discarded stream after each chromatographic step. Protein–protein interactions in chromatographic steps are assumed to be negligible. Finally, it is usually necessary to introduce membrane steps for buffer exchange and/or protein concentration between chromatographic steps, which could lead to some loss of protein product; however, for the needs of this study these losses are considered insignificant. Additional assumptions for process synthesis, such as the formulation of the models being based solely on physicochemical data and the approximation of the chromatographic peaks by isosceles triangles, can be found in refs 7–10.

3. Mathematical Formulation

The proposed MINLP representations are based on a previously developed MILP formulation (8) for the synthesis of purification bioprocesses. The optimization framework selects a tag that modifies the properties of the protein product in the most beneficial way and concurrently minimizes the number of chromatographic steps in the purification process. Next, the mathematical models are described in detail. The notation of the models is provided as an appendix.

3.1. Peptide Tag Size Constraints. An upper bound is imposed on the number of amino acids in each tag.

$$\sum_k n_k \leq N \quad (1)$$

As already discussed, smaller peptide tags have several practical advantages, including minimal effect on the protein structure, easier separation upon cleavage, and in many cases, no need for cleavage at all (12). Therefore, in this study the bound is usually set to six, which will also help to avoid the possibility of formation of a helix or a β -hairpin from the tag.

A constraint imposed on the amino acid composition of the peptide tag is that *at most* half of the fused amino acids are permitted to have a hydrophobic nature.

$$\sum_{k \in HA} n_k \leq 0.5 \cdot \sum_k n_k \quad (2)$$

This ensures that the tag will not bury itself within the attached protein or form undesirable structures. It is difficult to specifically define the maximum fraction of hydrophobic amino acids in the peptide tag; nevertheless hydrophobic amino acids should be balanced by polar residues in the composition of the tag.

3.2. Physicochemical Property Constraints. The net charge ($Q_{i,dp}$) of the tagged protein is predicted on the basis of the formula suggested by Mosher et al. (23):

$$Q_p = \sum_{k \in BA} \frac{b_k}{\frac{K_k}{[H^+]} + 1} - \sum_{k \in AA} \frac{a_k}{\frac{K_k}{[H^+]} + 1} + z_L \quad (3)$$

where a_k and b_k are the number of acidic and basic amino acids respectively, K_k is the ionization constant, $[H^+]$ is the concentration of hydrogen cations, and z_L is the total charge of ligands bound to the protein (typically metal ions).

According to eq 3, the net charge of a protein is approximated by considering the contribution of amino acids belonging to the basic group, minus the contribution of amino acids belonging to the acidic group. For this study, to estimate the charge $Q_{i,dp}$ of the tagged protein dp in ion exchange operation i , the charge contributions of any basic amino acids that exist in the tag are added to the initial charge of the desired product ($\hat{Q}_{i,dp}$) and, respectively, the charge contributions of any acidic amino acids are subtracted:

$$Q_{i,dp} = \hat{Q}_{i,dp} + \sum_{k \in BA} \frac{n_k}{\frac{K_k}{[H^+]_i} + 1} - \sum_{k \in AA} \frac{n_k}{\frac{K_k}{[H^+]_i} + 1}; \quad \forall i \in IE \quad (4)$$

Values for the ionization constants K_k (23, 24) are presented in Table 1. When proteins with complex structures are encountered, the interactions between various groups may be strong and some charged groups may not be exposed on the protein surface. For these reasons, the method described above may not always be accurate for large molecules, but it is still useful as an indication of how much and in what way the addition of a peptide tag will modify the net charge of the desired product.

Table 1. Ionization Constants for the Basic and Acidic Amino Acid Groups

residue	pK (23)
basic	
Arg	12.50
His	6.50
Lys	10.79
acidic	
Asp	3.91
Cys	8.30
Glu	4.25
Tyr	10.95

Many methods that predict the hydrophobic character of different regions of a protein's amino acid chain on the basis of well-established molecular thermodynamic theories exist (25, 26); however, interaction parameters must be determined experimentally for each polymer system and protein. Generally, significant experimental work is required before a prediction method can be developed. Here the protein's hydrophobicity (H_p) is estimated using a method developed by Lienqueo et al. (27):

$$H_p = \sum_{aa} h_{aa} \cdot r_{aa} \quad (5)$$

$$r_{aa} = \frac{s_{aa}}{\sum_{aa'} s_{aa'}}; \quad \forall aa \quad (6)$$

where h_{aa} is the value of the hydrophobicity assigned to each amino acid aa ($aa = 1, \dots, 20$), r_{aa} is the relative surface area exposed for each amino acid aa , and s_{aa} is the total exposed area of amino acid aa . The denominator in eq 6 represents the total surface of the protein.

According to eqs 5 and 6, the calculation of hydrophobicity is performed by considering the 3-dimensional structure of a protein molecule and the relative contribution of each amino acid on the surface of the protein to its properties (28). There are more than 40 hydrophobicity scales for amino acids in the literature (27), but for this study the normalized values of the scale proposed by Miyazawa and Jernigan (29) are used (presented in Table 2).

The contribution of the original protein molecule to the hydrophobicity of the tagged product is considered to remain constant, and therefore only the contributions of the amino acids in the tag need to be estimated and then added to the initial hydrophobicity of the protein product, \hat{H}_{dp} :

$$H_{dp} = \hat{H}_{dp} + \sum_k h_k \cdot r_k \quad (7)$$

The relative surface area, r_k , for each kind of amino acid k in the tag is given by

$$r_k = \frac{s_k \cdot n_k}{\hat{S}_{dp} + \sum_{k'} s_{k'} \cdot n_{k'}}; \quad \forall k \quad (8)$$

where the total surface of the tagged protein is estimated from the exposed surface of the protein product, \hat{S}_{dp} (considered to remain unchanged), plus the exposed surface of the amino acids in the tag, $\sum_k (s_k \cdot n_k)$. These amino acids are assumed to have a fully exposed surface. In cases where this is not true, selecting to place the peptide tag on the other terminus (the N-terminus

Table 2. Normalized Hydrophobicity and Exposed Surface Area for the 20 Amino Acids

residue	h_k (27)	s_k (30)
Phe ^a	1.000	210
Met ^a	0.987	185
Ile ^a	0.967	175
Leu ^a	0.908	170
Cys ^a	0.819	135
Trp ^a	0.775	255
Val ^a	0.770	155
Tyr	0.484	230
Ala	0.391	115
His	0.354	195
Thr	0.253	140
Gly	0.252	75
Arg	0.202	225
Ser	0.188	115
Gln	0.151	180
Pro	0.151	145
Asn	0.125	160
Glu	0.115	190
Asp	0.105	150
Lys	0.000	200

^a Hydrophobic group.

instead of the C-terminus of the protein product or vice versa) can solve this problem (12). Values for the surface areas of fully exposed of the amino acids (30) are presented in Table 2.

3.3. Dimensionless Retention Time Constraints.

For the modeling of various chromatographic techniques, established criteria based on the retention time and on the width of the chromatographic peak were used. The necessary parameters have been determined experimentally using pure proteins, to predict chromatographic behavior (27). Retention times are defined as a function of a physicochemical property and can be calculated using developed mathematical correlations (see eqs 9 and 10), which relate retention times of each particular chromatographic technique to the appropriate physicochemical property (e.g., ion exchange chromatography and charge; hydrophobic interaction and hydrophobicity).

In ion exchange chromatography, proteins adsorb either to exchangers that bind negatively charged molecules (anion exchange) or to exchangers that bind positively charged molecules (cation exchange). Retention times are a function of charge density (Q_{ip}/MW_p), determined by electrophoretic titration curves (31). The dimensionless retention times (KD_{ip}) are estimated on the basis of mathematical expressions and property data for the protein product and the contaminants (32). The proposed correlations for ion exchange chromatography were obtained using bind-and-elution conditions; the elution was obtained with an increasing NaCl gradient between 0.0 and 2.0 M NaCl.

For anion exchange:

$$\text{If } Q_{ip} \geq 0, KD_{ip} = 0$$

$$\text{If } Q_{ip} < 0, KD_{ip} = \frac{8826 \cdot |Q_{ip}/MW_p|}{1 + 18845 \cdot |Q_{ip}/MW_p|}, \quad \forall i \in AE \quad (9)$$

For cation exchange:

$$\text{If } Q_{ip} \leq 0, KD_{ip} = 0$$

$$\text{If } Q_{ip} > 0, KD_{ip} = \frac{7424 \cdot |Q_{ip}/MW_p|}{1 + 20231 \cdot |Q_{ip}/MW_p|}, \quad \forall i \in CE \quad (10)$$

It should be noted that expressions 9 and 10 only need to be modeled where they refer to dp ; dimensionless retention times for the contaminant proteins (KD_{ip}) remain constant and are used as parameters in the model. The modeling is performed using constraints 11–15:

$$Q_{i,dp}^+ - Q_{i,dp}^- = Q_{i,dp}; \quad \forall i \in IE \quad (11)$$

$$Q_{i,dp}^+ \leq M \cdot x_{i,dp}; \quad \forall i \in IE \quad (12)$$

$$Q_{i,dp}^- \leq M \cdot (1 - x_{i,dp}); \quad \forall i \in IE \quad (13)$$

where M is an appropriate upper bound. Binary variables $x_{i,dp}$ express whether the charge of the protein is positive or negative. The absolute value of $Q_{i,dp}$ is assigned to either $Q_{i,dp}^+$ or $Q_{i,dp}^-$ with constraint 11, because either $Q_{i,dp}^+$ or $Q_{i,dp}^-$ always has to be equal to zero due to constraints 12 and 13. For anion exchange chromatography

$$KD_{i,dp} = \frac{8826 \cdot (Q_{i,dp}^-/MW_{dp})}{1 + 18845 \cdot (-Q_{i,dp}/MW_{dp})}; \quad \forall i \in AE \quad (14)$$

If the protein charge is negative, it has to follow that $x_{i,dp} = 0$, and the retention time is given by constraint 14; otherwise, when the protein charge is positive, the retention time for anion exchange is zero, as $x_{i,dp}$ is forced to 1 and $Q_{i,dp}^- = 0$. Similarly, in the case of cation exchange chromatography

$$KD_{i,dp} = \frac{7424 \cdot (Q_{i,dp}^+/MW_{dp})}{1 + 20231 \cdot (Q_{i,dp}/MW_{dp})}; \quad \forall i \in CE \quad (15)$$

For a positive protein charge, $x_{i,dp} = 1$ and the retention time is given by constraint 15; otherwise, the retention time for cation exchange is zero.

Hydrophobic interaction chromatography utilizes the hydrophobic character of the proteins to separate the mixture according to their relative hydrophobicity. Most hydrophobic amino acids are located near the core of the protein structure and away from the surface, but there usually are hydrophobic residues on the protein surface as well. A formula developed by Lienqueo et al. (27) by evaluating a series of experimental and computational data is used here to estimate the dimensionless retention times for hydrophobic interaction ($KD_{HI,dp}$). Elution for hydrophobic interaction chromatography was obtained with a decreasing ammonium sulfate gradient between 2.0 and 0.0 M ammonium sulfate. Retention times are a function of hydrophobicity; the function in this case (on phenyl sepharose) is a quadratic equation:

$$KD_{HI,dp} = -12.14 \cdot H_{dp}^2 + 12.07 \cdot H_{dp} - 1.74 \quad (16)$$

As with ion exchange chromatography, dimensionless retention times for hydrophobic interaction in the case of the contaminant proteins ($KD_{HI,p}$) remain constant.

3.4. Concentration Factor Constraints. The concentration factor, CF_{ip} , represents the ratio between the mass of contaminant p after and before chromatographic step i . It can be calculated through a set of equations that describe the chromatographic peaks of the desired product and one of the contaminants (3). The chromatographic peaks are approximated by two triangles (chro-

matograms), one referring to the product and the other to the contaminant protein.

The chromatograms are approximated with eqs 19 presented below. In this set of equations, σ_i is the peak width parameter, which depends only on the type of chromatographic operations and is calculated by averaging over several proteins. For ion exchange, the value for the peak width is $\sigma_i = 0.15$; for hydrophobic interaction $\sigma_i = 0.22$. Parameter Δ is a percentage of error, for which an experimentally determined value of 0.02 is used. Deviation factors, DF_{ip} , indicate the distance between the desired product's chromatographic peak and the chromatographic peak of one of the contaminants. Deviation factors express the driving force of the separation process; they are defined as the difference between the retention times of the product ($KD_{i,dp}$) and the contaminant in question (KD_{ip}) for each particular chromatographic step.

$$DF_{ip} = |KD_{i,dp} - KD_{ip}|; \quad \forall i, p \neq dp \quad (17)$$

For the estimation of the deviation factor in this work the nondifferential relationship 17 is modeled with the following constraint:

$$DF_{ip} = [(KD_{i,dp} - KD_{ip})^2 + \epsilon^2]^{1/2}; \quad \forall i, p \neq dp \quad (18)$$

The relationships (3) that describe the concentration factor, CF_{ip} , for protein p in chromatographic technique i are:

$$CF_{ip} = 1 \quad \text{if } 0 \leq DF_{ip} < \frac{\sigma_i}{10} \quad (19a)$$

$$CF_{ip} = (1 + \Delta) \cdot \left(\frac{\sigma_i^2 - 2 \cdot DF_{ip}^2}{\sigma_i^2} \right) \quad \text{if } \frac{\sigma_i}{10} \leq DF_{ip} < \frac{\sigma_i}{2} \quad (19b)$$

$$CF_{ip} = 2 \cdot (1 + \Delta) \cdot \frac{(\sigma_i - DF_{ip})^2}{\sigma_i^2} \quad \text{if } \frac{\sigma_i}{2} \leq DF_{ip} < \sigma_i \quad (19c)$$

$$CF_{ip} = \Delta \quad \text{if } DF_{ip} \geq \sigma_i \quad (19d)$$

Here, concentration factors for the various chromatographic steps are modeled with sigmoid functions, which provide an accurate approximation of relationships in eqs 19. Because the peak width parameter, σ_i , is the same for both anion and cation exchange, there is only need for one equation for all ion exchange steps, but a separate one is required for hydrophobic interaction.

$$CF_{ip} = \frac{3.722}{3.727 + 0.579 \cdot e^{(54.410 \cdot DF_{ip} - 2.176)}} + 0.019; \quad \forall i \in IE, p \neq dp \quad (20)$$

$$CF_{HI,p} = \frac{3.937}{3.933 + 0.105 \cdot e^{(36.005 \cdot DF_{HI,p} - 0.299)}} + 0.018; \quad \forall p \neq dp \quad (21)$$

In the case of the desired product (dp), the concentration factor $CF_{i,dp}$ is always equal to 1, since the assumption of no product loss has been made. Therefore

$$CF_{i,dp} = 1; \quad \forall i \quad (22)$$

These concentration factors are introduced in the process synthesis constraints (section 3.5) that address the synthesis problem.

3.5. Process Synthesis Constraints. The convex hull representation applied for contaminant separation is presented in Vasquez-Alvarez and Pinto (8); the formulation for the selection of appropriate chromatographic steps and the indication of the remaining amount of protein after each step i is also applied here.

$$m_{1p} = CF_{1p} \cdot w_1 \cdot m_{0p} + m_{0p} \cdot (1 - w_1) \quad \forall p \quad (23)$$

$$m_{ip} = CF_{ip} \cdot m_{i-1,p}^1 + m_{i-1,p}^2 \quad \forall p, i = 2, \dots, I \quad (24)$$

$$m_{i-1,p} = m_{i-1,p}^1 + m_{i-1,p}^2 \quad \forall p, i = 2, \dots, I \quad (25)$$

$$0 \leq m_{i-1,p}^1 \leq m_{0p} \cdot w_i \quad \forall p, i = 2, \dots, I \quad (26)$$

$$0 \leq m_{i-1,p}^2 \leq m_{0p} \cdot (1 - w_i) \quad \forall p, i = 2, \dots, I \quad (27)$$

The mass of protein that is left after the first chromatographic step is given by constraint 23. In the following steps, constraints 24–27 hold. When chromatographic step i is selected (i.e., $w_i = 1$) the mass of contaminants is reduced (i.e., $m_{ip} = CF_{ip} \cdot m_{i-1,p}$), whereas when step i is not selected the mass remains constant (i.e., $m_{ip} = m_{i-1,p}$).

The mass of the desired protein product must meet the purity demand after the last chromatographic step I .

$$m_{I,dp} \geq SP \cdot \sum_p m_{I,p} \quad (28)$$

4. Solution Approach

Typical performance criteria to be optimized in the problem defined above are the number of purification steps and/or the size of the tag. The overall problem is non-convex; nonlinearities arise in constraint 8, for the estimation of the relative surface area of amino acids in the tag; constraints 14–16, for the evaluation of retention times from the values of the protein product's properties; constraint 18, for the calculation of the deviation factor; sigmoid constraints 20 and 21, for the computation of the concentration factor; and constraints 23 and 24, for estimating the mass of proteins after each chromatographic step. The presented models constitute MINLP formulations. A two-stage solution procedure is proposed to identify the shortest amino acid sequence that can produce the optimal flowsheet for the purification process.

Stage 1. Designing flowsheets with fewer purification steps can significantly reduce costs (33). The overall objective is to minimize the total number of selected chromatographic steps in the purification process, subject to the constraints described above (problem P1). A tag is also selected in this stage, but the selection is further improved at stage 2.

Problem P1

$$\text{minimize } \sum_i w_i \quad (29)$$

subject to:

peptide tag size constraints 1 and 2;
physicochemical property constraints 4, 7, and 8;
dimensionless retention time constraints 11–16;
concentration factor constraints 18 and 20–22;

Table 3. Physicochemical Properties of Protein Mixture in First Example

proteins	m_{op} (mg/mL)	MW_p (Da)	H_p	S	Q_{ip} (C/molecule) $\times 10^{-17}$				
					pH 4	pH 5	pH 6	pH 7	pH 8
dp	2	22200	0.27	9573.15	1.60	1.57	1.64	1.55	0.75
p_1	2	77000	0.23	29287.60	0.93	0.33	-0.12	-0.34	-0.50
p_2	2	23600	0.31	10910.80	2.15	1.46	1.17	0.78	0.38
p_3	2	43800	0.28	15880.90	1.16	-0.63	-1.36	-1.82	-1.95

process synthesis constraints 23–28;

$$x_{i,dp}, w_i \in \{0,1\}; \quad \forall i \in IE \quad (30)$$

$$n_k \in Z^+; \quad \forall k \quad (31)$$

$$CF_{ip}, DF_{ip}, H_{dp}, KD_{i,dp}, m_{ip}, Q_{i,dp}^+, Q_{i,dp}^-, r_k \geq 0; \\ \forall i,p,k \quad (32)$$

Stage 2. The objective here (problem P2) is to minimize the number of amino acids in the tag, subject to the same constraints as before, plus an additional constraint that fixes the number of chromatographic steps as identified in stage 1. Then, the smallest tag that can produce the optimal flowsheet is finally determined.

Problem P2

$$\text{minimize } \sum_k n_k \quad (33)$$

subject to:

constraints 1, 2, 4, 7, 8, 11–16, 18, 20–28, 30–32, and

$$\sum_i w_i \leq i^* \quad (34)$$

where i^* is the number of steps identified in stage 1.

5. Computational Results

The proposed formulation is applied to two example mixtures. Solutions were obtained with the network-enabled, problem-solving environment NEOS Server 4.0 (<http://www-neos.mcs.anl.gov/>; 34–36) using the SBB solver for the solution of the MINLP models and CONOPT3 as the non-linear programming (NLP) solver.

5.1. Example 1. The first example is based on experimental data (27, 32) for a mixture of four proteins: thaumatin (dp), conalbumin (p_1), chymotrypsinogen A (p_2), and ovalbumin (p_3). The physicochemical properties of the mixture are presented in Table 3. The purity level requirement for the desired product (dp) is 98%. Overall, there are 11 candidate chromatographic techniques: anion exchange chromatography (AE) at pH 4, AE at pH 5, AE at pH 6, AE at pH 7, AE at pH 8, cation exchange chromatography (CE) at pH 4, CE at pH 5, CE at pH 6, CE at pH 7, CE at pH 8, and hydrophobic interaction (HI).

To acquire a point of reference, the example was first solved without any tag fused to the protein product, i.e., using the formulation of problem P1, with an upper bound of zero imposed on the number of amino acids in the tag (i.e., $N = 0$). The resulting mathematical model involves 317 constraints, 41 discrete variables, and 275 continuous variables and was solved in 4.14 s. The optimal solution is presented in Figure 1. The model was able to identify a solution that achieves a purity of 98.1% for the desired product, for which four chromatographic steps are needed: CE pH 6, CE pH 7, CE pH 8, and HI.

Note that the product mass remains constant, but its purity increases after each step: from 25.0% in the

original mixture to 51.4% after the first chromatographic step (CE pH 6), 64.6% after CE pH 7, 71.9% after CE pH 8, and finally 98.1% (above the required purity level) after the final chromatographic operation (HI).

An improved solution was suggested when using a peptide tag. The minimum number of steps is identified by solving problem P1 with an upper bound of 6 amino acids per tag (stage 1). The model was solved in 8.49 s. Only three separation steps are needed: cation exchange chromatography at pH 7, cation exchange chromatography at pH 8, and hydrophobic interaction. Next, the number of purification steps was fixed ($i^* = 3$) and the model was solved again using the formulation of problem P2 (stage 2). The CPU time was 20.57 s. The solution is a tag of 2 lysine residues; the purity demand was achieved (98.0%), and the process included the same three purification techniques (CE pH 7, CE pH 8, HI) as in the solution of problem P1. The results are presented in Figure 2 and in Table 4. The selection of a tag exclusively with lysine amino acids suggests that the optimal solution to the problem at hand is to increase the charge of the desired protein. The use of other basic amino acids would have a stronger effect on the charge than lysine, but their presence would also increase hydrophobicity, which remains unchanged with lysine ($h_k = 0$, Table 2). The observation implies that an increase in hydrophobicity is not beneficial. This was tested by forcing the inclusion of hydrophobic amino acids in the tag (e.g., phenylalanines); the experiments showed that a purity of 98% (as required here) is not achievable when hydrophobicity is increased even by a small amount.

Table 4 presents the dimensionless retention times, deviation factors, and concentration factors for each protein in each chromatographic step for the solution of example 1 with the formulation of problem P2 (as shown in Figure 2). All values are calculated from the mathematical model, as they depend on the physicochemical properties of the protein product, which are liable to change. It should be noted that contaminant retention times remain constant and are therefore used as parameters for the mathematical model.

5.2. Example 2. For the purpose of testing the models, a larger and more challenging example was created. The physicochemical properties of this second mixture of 13 proteins are presented in Table 5. The properties for the proteins of example 2 were determined using a random number generator to produce values within certain parameters, for example the charge of each protein was allowed to range between -3×10^{-17} and 3×10^{-17} C per molecule. A similar procedure was followed for the random assignment of hydrophobicities and molecular weights. The exposed surface of the protein product was considered to be proportional to the molecular weight (i.e., the larger the molecular weight, the larger the exposed surface of the molecule), and was set to $\hat{S}_{dp} = 29287.6$ (see constraint 8). The purity level requirement for protein dp was set to 95%. The same 11 candidate chromatographic steps as for example 1 are available: AE pH 4, AE pH 5, AE pH 6, AE pH 7, AE pH 8, CE pH 4, CE pH 5, CE pH 6, CE pH 7, CE pH 8, and HI.

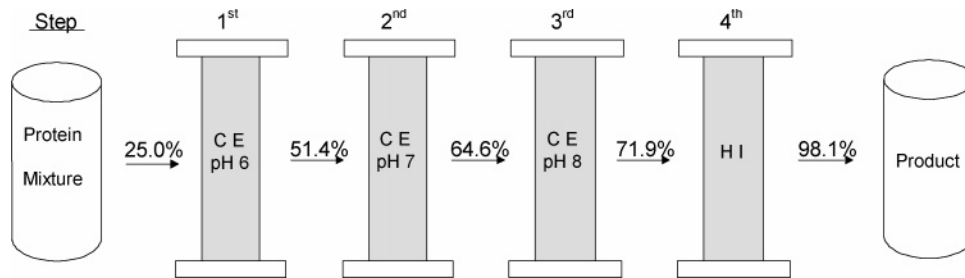


Figure 1. Optimal flowsheet for purification of protein mixture in example 1 without tag.

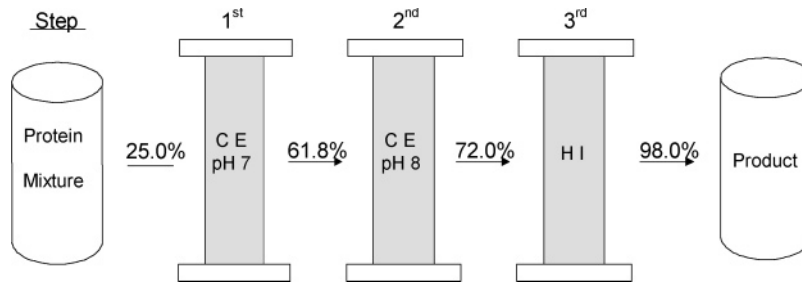


Figure 2. Optimal flowsheet for purification of protein mixture in example 1 with a tag of two lysines.

Table 4. Values of Retention Times, Deviation Factors and Concentration Factors As Estimated for the Solution of Example 1 (98% purity, 3 steps, 2-lys tag)

step ^a	<i>dp</i>			<i>p</i> ₁			<i>p</i> ₂			<i>p</i> ₃		
	<i>KD</i>	<i>DF</i>	<i>CF</i>	<i>KD</i> ^b	<i>DF</i>	<i>CF</i>	<i>KD</i> ^b	<i>DF</i>	<i>CF</i>	<i>KD</i> ^b	<i>DF</i>	<i>CF</i>
AE4	0.000	0.000	1.000	<i>0.000</i>	0.000	1.000	<i>0.000</i>	0.000	1.000	<i>0.000</i>	0.000	1.000
AE5	0.000	0.000	1.000	<i>0.000</i>	0.000	1.000	<i>0.000</i>	0.000	1.000	<i>0.100</i>	0.100	0.216
AE6	0.000	0.000	1.000	<i>0.013</i>	0.013	0.983	<i>0.000</i>	0.000	1.000	<i>0.173</i>	0.173	0.024
AE7	0.000	0.000	1.000	<i>0.036</i>	0.036	0.907	<i>0.000</i>	0.000	1.000	<i>0.206</i>	0.206	0.020
AE8	0.000	0.000	1.000	<i>0.051</i>	0.051	0.797	<i>0.000</i>	0.000	1.000	<i>0.214</i>	0.214	0.019
CE4	0.219	0.000	1.000	<i>0.072</i>	0.147	0.037	<i>0.238</i>	0.019	0.972	<i>0.128</i>	0.091	0.300
CE5	0.218	0.000	1.000	<i>0.029</i>	0.189	0.021	<i>0.204</i>	0.014	0.982	<i>0.000</i>	0.218	0.019
CE6	0.222	0.000	1.000	<i>0.000</i>	0.222	0.019	<i>0.184</i>	0.038	0.898	<i>0.000</i>	0.222	0.019
CE7	0.217	0.000	1.000	<i>0.000</i>	0.217	0.019	<i>0.147</i>	0.070	0.580	<i>0.000</i>	0.217	0.019
CE8	0.153	0.000	1.000	<i>0.000</i>	0.153	0.033	<i>0.090</i>	0.063	0.670	<i>0.000</i>	0.153	0.033
HI	0.629	0.000	1.000	<i>0.413</i>	0.216	0.039	<i>0.832</i>	0.203	0.051	<i>0.701</i>	0.072	0.809

^a AE4: anion exchange at pH 4; etc. ^b The values of *KD* for *p*₁, *p*₂, and *p*₃ (denoted in italics) are calculated before the computational experiments and are used as parameters in the model

Table 5. Physicochemical Properties of Protein Mixture in Second Example

proteins	<i>m</i> _{0p} (mg/mL)	<i>MW</i> _p (Da)	<i>H</i> _p	<i>Q</i> _{ip} (C/molecule) × 10 ⁻¹⁷				
				pH 4	pH 5	pH 6	pH 7	pH 8
<i>dp</i>	2	77 000	0.28	2.04	1.06	-0.37	-0.81	-1.13
<i>p</i> ₁	2	22 200	0.27	1.60	1.57	1.56	1.55	0.75
<i>p</i> ₂	2	23 600	0.31	2.15	1.46	1.17	0.78	0.38
<i>p</i> ₃	2	13 500	0.23	1.83	0.65	0.26	-0.20	-0.33
<i>p</i> ₄	2	43 800	0.28	1.16	-0.63	-1.36	-1.82	-1.95
<i>p</i> ₅	2	15 900	0.27	2.89	2.81	2.80	2.64	2.07
<i>p</i> ₆	2	14 400	0.32	-0.46	-0.47	-0.63	-1.21	-1.25
<i>p</i> ₇	2	17 500	0.21	0.45	-0.62	-0.79	-1.26	-1.70
<i>p</i> ₈	2	50 000	0.27	-0.12	-0.32	-0.76	-0.91	-1.04
<i>p</i> ₉	2	12 100	0.18	1.46	0.62	-1.02	-1.33	-1.52
<i>p</i> ₁₀	2	25 500	0.30	1.01	-0.63	-1.27	-1.59	-1.76
<i>p</i> ₁₁	2	26 000	0.28	2.96	1.26	0.92	0.54	0.01
<i>p</i> ₁₂	2	19 900	0.25	0.93	0.33	-0.12	-0.34	-0.50

Example 2 was first solved without tags (i.e., *N* = 0). The formulation of problem P1 was applied; the resulting mathematical model involves 884 constraints, 41 discrete variables, and 752 continuous variables. The optimal solution was identified in 71.78 s and is presented in Figure 3. The product is recovered with 95.3% purity and six chromatographic steps are needed: AE pH 6, AE pH 7, AE pH 8, CE pH 4, CE pH 5, and HI.

Using model P1, an improved solution with only four chromatographic steps (AE pH 6, AE pH 8, CE pH 4, HI) is suggested. The results were produced with an upper

bound of 6 amino acids per tag (i.e., *N* = 6). To test whether this solution can be achieved with a smaller tag, model P2 was applied to identify solutions with the smallest possible number of amino acids in the tag. Using the minimum number of steps for 95% purity (i.e., *i*^{*} = 4), a minimum number of amino acids in the tag was specified. The results are presented in Figure 4. The suggested tag consists of 1 phenylalanine, 1 methionine, and 2 tyrosine amino acids; the process has four chromatographic steps (AE pH 7, AE pH 8, CE pH 4, HI) and a purity of 95.4%. Note that this solution is different from the one produced with model P1; a different set of chromatographic steps was selected for the purification process.

Next, a higher level of purity for the protein product was tested. Example 2 was solved again with a higher demand of 98% for product purity, which leads to an infeasible solution when the model is solved without tags. With an upper bound of 6 amino acids per tag, an improved solution with a purity of 98.4% and five chromatographic steps (AE pH 6, AE pH 8, CE pH 4, CE pH 5, HI) was suggested. The problem was solved again using the formulation of model P2 and an upper bound of 5 for the number of chromatographic steps (i.e., *i*^{*} = 5). The results are presented in Figure 5. The selected tag consists of 1 phenylalanine, 1 tryptophan, and 2

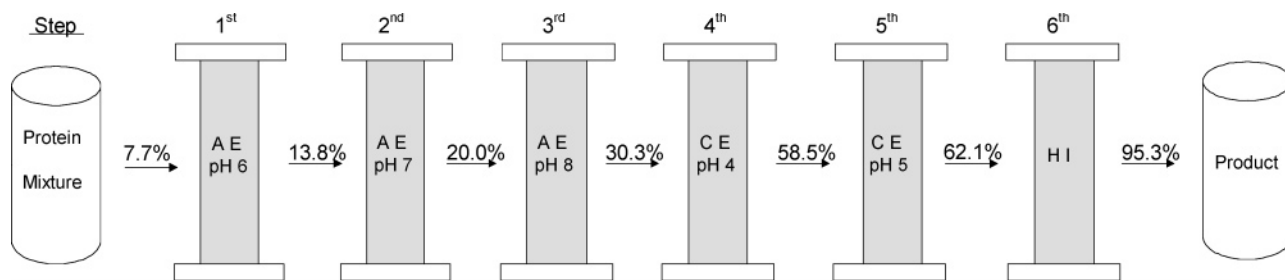


Figure 3. Optimal flowsheet for purification of protein mixture in example 2 without tag.

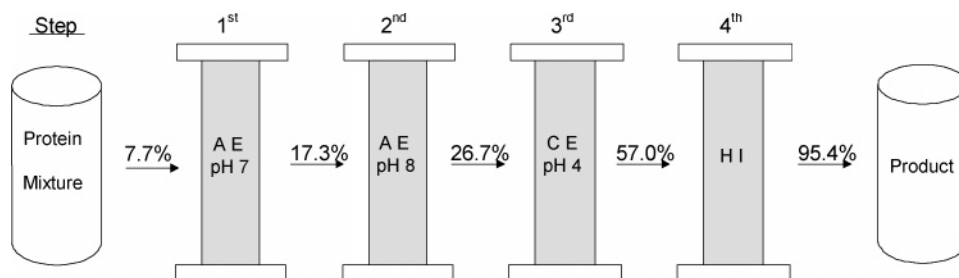


Figure 4. Optimal flowsheet for purification of protein mixture in example 2 with a minimized tag of one phenylalanine, one methionine, and two tyrosine amino acids (purity requirement, 95%).

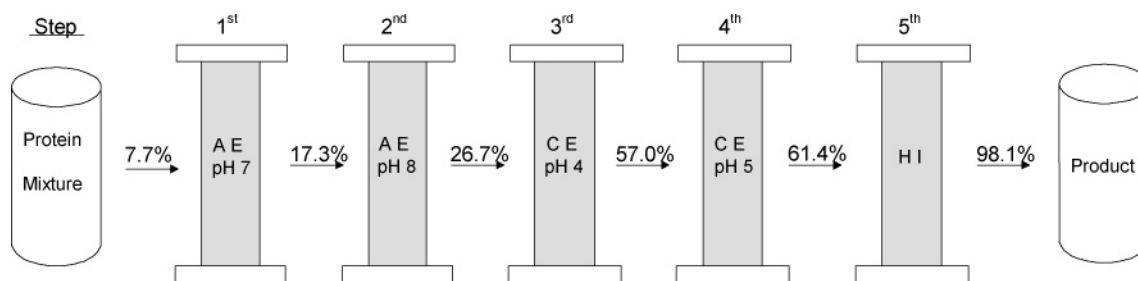


Figure 5. Optimal flowsheet for purification of protein mixture in example 2 with a tag of one phenylalanine, one tryptophan, and two tyrosine amino acids (increased purity requirement, 98%).

Table 6. Summary of Computational Statistics

example	model	max aa's	max steps	discrete/continuous variables	constraints	nodes	CPU time (s)	objective value
1	P1	≤ 0	-	41/275	317	114	4.14	4 ^a
1	P1	≤ 6	-	41/275	317	192	8.49	3 ^a
1	P2	≤ 6	≤ 3	41/275	318	636	20.57	2 ^b
2	P1	≤ 0	-	41/752	884	363	71.78	6 ^a
2	P1	≤ 6	-	41/752	884	346	68.84	4 ^a
2	P2	≤ 6	≤ 4	41/752	885	1777	175.61	4 ^b
2 ^c	P1	≤ 0	-	41/752	884	340	53.62	infeasible ^a
2 ^c	P1	≤ 6	-	41/752	884	514	94.50	5 ^a
2 ^c	P2	≤ 6	≤ 5	41/752	885	2881	320.67	4 ^b

^a Number of chromatographic steps. ^b Number of amino acids in the tag. ^c 98% purity demand.

tyrosines. The process has five chromatographic steps (AE pH 7, AE pH 8, CE pH 4, CE pH 5, HI) and a purity of 98.1%.

Table 6 presents computational statistics for the optimization of the two presented examples with models P1 and P2. Model sizes in terms of discrete and continuous variables and constraints are given. Also shown are CPU times, nodes enumerated and objective values.

6. Conclusions

In this paper, an optimization framework for the simultaneous selection of optimal peptide tags and the synthesis of chromatographic steps for the purification of protein mixtures in downstream processing has been presented. The framework utilizes the advantages of integer optimization and mathematical programming techniques, incorporates recent developments in the synthesis and optimization of downstream purification

processes, and can be extended to consider application to larger examples, use of additional chromatographic techniques, or manipulation of other physicochemical properties. The overall problem has been formulated as an MINLP model and a two-stage solution procedure has been proposed.

Two examples of protein mixtures were tested to demonstrate the efficiency of the optimization-based methodology. In both examples, small peptide tags were used (from two to four residues), and only specific physicochemical properties were modified, especially hydrophobicity and charge, without significant conformational changes or bioactivity. These smaller changes have allowed an important decrease in the number of purification steps. In the first example, only the modification of charge benefited the purification, whereas hydrophobicity was the most influential property in the second example. Results were indicative of the benefits

of the application of optimization-based techniques in the use of purification tags in biotechnological production plants and have provided a useful guideline for both downstream process synthesis and optimal tag design. However, it will be interesting to validate the generated hypotheses by evaluating experimentally the chromatographic behavior of the proteins together with the peptide tags.

Testing the mathematical framework with larger examples and investigating alternative solution strategies is a possible extension to this work. Consideration of additional types of chromatographic steps would also be very interesting, provided that the appropriate correlations become available. For example, ion exchange chromatography with a pH gradient could be considered, which would potentially provide the same degree of purification but using a smaller number of steps. Finally, the modeling of the purification can be extended to incorporate a number of issues, including sequencing of the purification steps, product loss, protein-protein interactions, the use of membrane steps between chromatographic steps, and/or the application of alternative objective functions.

Acknowledgment

The authors thank Paul A. Dalby for offering invaluable insight and for useful discussions and I. David L. Bogle and Eric S. Fraga for comments and discussions. The authors acknowledge financial support from VITAE (Grant B-11487/10B006) and the Royal Academy of Engineering (Grant IJB/LB/ITG 03-330). E.S. was financially supported by EPSRC (Award 00319001) and the Centre for Process Systems Engineering.

Notation

Indices

dp	desired protein product
$i: 1, \dots, I$	chromatographic techniques
k	amino acids
$p: 1, \dots, P$	proteins in the mixture

Sets

AE	anion exchange chromatography
CE	cation exchange chromatography
IE	ion exchange chromatography
AA	acidic amino acid group
BA	basic amino acid group
HA	hydrophobic amino acid group

Parameters

\hat{H}_{dp}	initial product hydrophobicity
h_k	hydrophobicity of amino acid k
K_k	ionization constants
KD_{ip}	retention time of protein p in chromatographic technique i
M	large positive number
m_{0p}	initial mass of protein p
MW_{dp}	molecular weight of product
N	maximum number of amino acids in tag (~6 or 7)
$\hat{Q}_{i,dp}$	initial product charge for chromatographic technique i
\hat{S}_{dp}	initial total surface area of product
s_k	total exposed area for amino acid k
SP	specified product purity

ϵ small number

Integer Variables

n_k number of occurrences of amino acid k in the tag

Binary Variables

$x_{i,dp}$ 1 if product charge is greater than zero; 0 otherwise
 w_i 1 if chromatographic technique i is selected; 0 otherwise

Continuous Variables

$Q_{i,dp}$ product charge for chromatographic technique i

Positive Continuous Variables

CF_{ip} concentration factor of protein p after chromatographic technique i
 DF_{ip} deviation factor of protein p after chromatographic technique i
 H_{dp} product hydrophobicity
 h_k hydrophobicity of amino acid k
 $KD_{i,dp}$ product retention time in chromatographic technique i
 m_{ip} mass of protein p after chromatographic technique i
 r_k relative surface area for amino acid k

References and Notes

- (1) Datar, R. Economics of Primary Separation Steps in Relation to Fermentation and Genetic Engineering. *Process Biochem.* **1986**, *21*, 19–26.
- (2) Petrides, D. P. BioPro Designer—An Advanced Computing Environment for Modeling and Design of Integrated Biochemical Processes. *Comput. Chem. Eng.* **1994**, *18S*, S621–S625.
- (3) 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**, *20S*, S189–S194.
- (4) Lienqueo, M. E.; Salgado, J. C.; Asenjo, J. A. An Expert System for Selection of Protein Purification Processes: Experimental Validation. *J. Chem. Technol. Biotechnol.* **1999**, *74*, 293–299.
- (5) Fraga, E. S. The Generation and Use of Partial Solutions in Process Synthesis. *Chem. Eng. Res. Des.* **1998**, *76*, 45–54.
- (6) Steffens, M. A.; Fraga, E. S.; Bogle, I. D. L. Synthesis of Bioprocesses Using Physical Properties Data. *Biotechnol. Bioeng.* **2000**, *68*, 218–230.
- (7) Vasquez-Alvarez, E.; Lienqueo, M. E.; Pinto, J. M. Optimal Synthesis of Protein Purification Processes. *Biotechnol. Prog.* **2001**, *17*, 685–696.
- (8) Vasquez-Alvarez, E.; Pinto, J. M. MILP Models for the Synthesis of Protein Purification Processes. *Proc. ESCAPE-11*. **2001**, 579–584.
- (9) Vasquez-Alvarez, E.; Pinto, J. M. Efficient MILP Formulations for the Optimal Synthesis of Chromatographic Protein Purification Processes. *J. Biotechnol.* **2004**, *110*, 295–311.
- (10) Vasquez-Alvarez, E.; Pinto, J. M. A Mixed Integer Linear Programming Model for the Optimal Synthesis of Protein Purification Processes with Product Loss. *Chem. Biochem. Eng. Q.* **2003**, *17*, 77–84.
- (11) Steffens, M. A.; Fraga, E. S.; Bogle, I. D. L. Synthesis of Purification Tags for Optimal Downstream Processing. *Comput. Chem. Eng.* **2000**, *24*, 717–720.
- (12) Terpe, K. Overview of Tag Protein Fusions: From Molecular and Biochemical Fundamentals to Commercial Systems. *Appl. Microbiol. Biotechnol.* **2003**, *60*, 523–533.

- (13) Sassenfeld, H. M.; Brewer S. J. A Polypeptide Fusion Designed for the Purification of Recombinant Proteins. *Bio/Technology* **1984**, *2*, 76–81.
- (14) Hochuli, E.; Döbeli, H.; Schacher, A. New Metal Chelate Adsorbent Selective for Proteins and Peptides Containing Neighboring Histidine-Residues. *J. Chromatogr.* **1987**, *411*, 177–184.
- (15) Hopp, T. P.; Prickett, K. S.; Price, V. L.; Libby, R. T.; March, C. J.; Cerretti, D. P.; Urdal, D. L.; Conlon, P. J. A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification. *Bio/Technology* **1988**, *6*, 1204–1210.
- (16) Schmidt, T. G. M.; Skerra, A. The Random Peptide Library-Assisted Engineering of a C-Terminal Affinity Peptide, Useful for the Detection and Purification of a Functional Ig Fv Fragment. *Protein Eng.* **1993**, *6*, 109–122.
- (17) Evan, G. I.; Lewis, G. K.; Ramsay, G.; Bishop, J. M. Isolation of Monoclonal Antibodies Specific for Human c-myc Proto-Oncogene Product. *Mol. Cell. Biol.* **1985**, *5*, 3610–3616.
- (18) Karpeisky, M. Y.; Senchenko, V. N.; Dianova, M. V.; Kanevsky, V. Y. Formation and Properties of S-Protein Complex with S-Peptide-Containing Fusion Protein. *FEBS Lett.* **1994**, *339*, 209–212.
- (19) Uhlen, M.; Moks, T. Gene Fusions for Purpose of Expression—An Introduction. *Methods Enzymol.* **1990**, *185*, 129–143.
- (20) Sassenfeld, H. M. Engineering Proteins for Purification. *Trends Biotechnol.* **1990**, *8*, 88–93.
- (21) Nygren, P. A.; Stahl, S.; Uhlen, M. Engineering Proteins to Facilitate Bioprocessing. *Trends Biotechnol.* **1994**, *12*, 184–188.
- (22) Creighton, T. E. *Proteins: Structures and Molecular Properties*; W. H. Freeman: New York, 1993.
- (23) Mosher, R. A.; Gebauer, P.; Thormann, W. Computer-Simulation and Experimental Validation of the Electrophoretic Behavior of Proteins: III. Use of Titration Data Predicted by the Protein's Amino Acid Composition. *J. Chromatogr.* **1993**, *638*, 155–164.
- (24) Devereux, J.; Haeblerli, P.; Smithies, O. A Comprehensive Set of Sequence Analysis Programs for the VAX. *Nucleic Acids Res.* **1984**, *12*, 387–395.
- (25) Hopp, T. P.; Woods, K. R. Prediction of Protein Antigenic Determinants from Amino-Acid Sequences. *Proc. Natl. Acad. Sci. Biol.* **1981**, *78*, 3824–3828.
- (26) Kyte, J.; Doolittle, R. F. A Simple Method for Displaying the Hydrophobic Character of a Protein. *J. Mol. Biol.* **1982**, *157*, 105–132.
- (27) Lienqueo, M. E.; Mahn, A.; Asenjo, J. A. Mathematical Correlations for Predicting Protein Retention Times in Hydrophobic Interaction Chromatography. *J. Chromatogr. A* **2002**, *978*, 71–79.
- (28) Berggren, K.; Wolf, A.; Asenjo, J. A.; Andrews, B. A.; Tjerneld, F. The Surface Exposed Amino Acid Residues of Monomeric Proteins Determine the Partitioning in Aqueous Two-Phase Systems. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **2002**, *1596*, 253–268.
- (29) Miyazawa, R.; Jernigan, R. L. Estimation of Effective Interresidue Contact Energies from Protein Crystal Structures—Quasi-Chemical Approximation. *Macromolecules* **1985**, *18*, 534–552.
- (30) Chothia, C. The Nature of the Accessible and Buried Surfaces in Proteins. *J. Mol. Biol.* **1975**, *105*, 1–14.
- (31) Watanabe, E.; Tsoka, S.; Asenjo, J. A. Selection of Chromatographic Protein Purification Operations Based on Physicochemical Properties. *Ann. N. Y. Acad. Sci.* **1994**, *721*, 348–364.
- (32) Lienqueo, M. E. *Development of an Expert System for the Rational Selection of Protein Purification Processes: Optimization of Sequence Selection Criteria*. Ph.D. Thesis (in Spanish), University of Chile: Santiago, Chile, 1999.
- (33) Atkinson, B.; Mavituna, F. *Biochemical Engineering and Biotechnology Handbook*; Macmillan Press: New York, 1991.
- (34) Gropp, W.; Moré, J. Optimization Environments and the NEOS Server. In *Approximation Theory and Optimization*; Buhmann, M. D., Iserles, A., Eds.; Cambridge University Press: Cambridge, 1997; pp 167–182.
- (35) Czyzyk, J.; Mesnier, M. P.; Moré, J. J. The NEOS Server. *IEEE Comput. Sci. Eng.* **1998**, *5*, 68–75.
- (36) Dolan, E. The NEOS Server 4.0 Administrative Guide. In *Technical Memorandum ANL/MCS-TM-250*; Mathematics and Computer Science Division: Argonne National Laboratory, 2001.

Accepted for publication January 26, 2005.

BP049650N