Separation and Purification Technology 65 (2009) 86–94

Contents lists available at ScienceDirect

Separation and Purification Technology

journal homepage: www.elsevier.com/locate/seppur

Computer-aided design to select optimal polypeptide tags to assist the purification of recombinant proteins

M.E. Lienqueo*, J.C. Salgado, O. Giaverini, J.A. Asenjo

Centre for Biochemical Engineering and Biotechnology, The Institute for Cell Dynamics and Biotechnology, Department of Chemical Engineering and Biotechnology, University of Chile, Beauchef 861, Santiago, Chile

ARTICLE INFO

Keywords:
Protein purification processes
Tagged proteins
Cutinase

ABSTRACT

The addition of a polypeptide fusion partner, called tag, to facilitate purification and detection of recombinant proteins is well recognized. Many different proteins, domains, or peptides can be fused with the target protein and the advantages of using fusion proteins. Nevertheless, the selection of the optimal peptide tag and the right purification system for a specific target protein is difficult. The objective of this work was to develop a mathematical model with decision binary variables, based on MINLP models, which permits the selection of optimal peptide purification tags and optimizes the protein purification process. This model considers a particular set of well-known peptide tags capable of obtaining the required levels of purification. The objective function of the model is the maximization of the profit of the process; this is maximizing the recovery of the desired protein and to minimize the cost of the purification steps. Additionally, a linear relationship between price of the protein and desired purity level was proposed. The mathematical model was evaluated using an example based on cutinase experimental data. The results compare the differences between the sequences with and without tags. In both cases the number of steps is similar, however the recovery level and profit with tag are bigger than the solution without tag. Additionally, the selected peptide tag, for majority cases studied, was FLAG, an 8-amino-acid peptide (DYKDDDDK), which increments the charge and also slightly the hydrophobicity of the protein. Finally, in this model it is simple to introduce new tags for evaluation, in silico, of its possibilities for developing an optimal purification process. Hence, this model could be useful for optimizing purification processes without experimental tests.

1. Introduction

Protein purification processes are carried out using several chromatographic steps. Each of these techniques exploits physicochemical and biochemical differences between the desired protein and the contaminants of the mixture to be purified. Affinity chromatography exploits specific biochemical interactions between the protein and the matrix ligand; ion exchange chromatography exploits electrostatic interactions at different pH; hydrophobic interaction chromatography is based on hydrophobic interactions, and gel filtration chromatography is based on size differences. In the ideal case that such differences are significant, high purity can be obtained at a high recovery level. Unfortunately, this is not a usual situation, so alternatives have been sought to increase protein affinity to specific chromatographic ligands, modifying the target protein properties [1]. One of these alternatives is the addition or fusion of polypeptide tags by genetic engineering [2]. These polypeptide tags are amino acid sequences added to a protein (called fusion protein) so as to give some particular feature, without a significant alteration of the biological and/or physicochemical features of the proteins. Several works [3–13] have been reported in which these tags have been added, facilitating purification of the target protein. Unfortunately the selection of the optimal polypeptide tag is not easy.

1.1. Use of polypeptide tags for protein purification

There is a range of possibilities to modify proteins to improve purification. Among those alternatives is the modification of superficial properties of the proteins [14], or fusion or addition of affinity tags or polypeptide tags [1,15]. These polypeptide tags are amino acid sequences which are added to a protein to give it a particular feature, e.g., changes in superficial hydrophobicity, charge, attraction by a metallic chelate; with this, selectivity increases in a given purification mode. To achieve this several affinity tags are available, ranging from small peptide sequences to fusion partners

* Corresponding author. Tel.: +56 2 978 4709; fax: +56 2 699 1084.
E-mail address: mlienque@ing.uchile.cl (M.E. Lienqueo).

1383-5866/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.seppur.2008.09.001

with size and complexity similar to a protein. For example, His-tag, Arg-tag, calmodulin-binding peptide (CBP), cellulase-binding domain (CBD), DsbA, c-myc-tag, glutathione S-transferase (GST), FLAG-tag, HAT-tag, maltose-binding protein (MBP), NusA, S-tag, SBP-tag, Strep II-tag, thioredoxin, Biotin acceptor peptide (BAP), etc. [2,16], and short hydrophobic peptide tags, e.g., (TrpPro)2, (TrpPro)4, (Tyr)3 (TyrPro)3, (Tyr)3(Pro)2, (Tyr)4, (TyrPro)4, (Tyr)6, (Tyr)6(Pro)2, (Tyr)8 [4,6,8]. All of these are fused to the C-terminal or N-terminal, modifying one or more properties of the protein, such as affinity, hydrophobicity, charge and solubility [17]. Polypeptide tags show numerous advantages: (a) they need genetically fewer modifications in the target product; (b) as they are small molecules, they have a minimum impact on the tertiary structure and biological activity of the fusion protein; (c) they are relatively easy to remove, a specific cleavage location may be included, so that the tag could be removed at the purification stage using specific proteases such as TEV, 3C, Xa, Entk, Thr and Caspase [16]. The most common and successful process considers placing an affinity tag at the N terminus of the fusion partner, then the protein can be purified, next cleave the tag, and finally re-purify on the same affinity matrix to remove the cleaved tag [18]. The cleavage of fusion tag is required for biochemical studies and therapeutic proteins; for other applications, polypeptide tags may not need to be cleaved; (d) separation techniques are not usually expensive – as they are available at large scale – and they can be applied to a wide range of proteins.

The most widely used small polypeptide tags are: (a) polyarginine (5–6 arginines) for purification by cation exchange [19–21]; (b) polyhistidine (between 2 and 10, generally 6 histidines, His6) for purification of the immobilized metal ion affinity chromatography type (IMAC), [9,10,12]; (c) hydrophilic sequence FLAG (DYKDDDK) for purification with anti-FLAG monoclonal antibody matrix [22], (d) Strep II-tag sequence (WSHPQFEK) [23,24] for purification with Strep-Tactin (modified streptavidin) matrix; (e) c-myc sequence (EQKLISEEDL) for purification with monoclonal antibody matrices [25]; (f) hydrophobic sequences (4–8 amino acids of the (WP)n or (YP)n) type [3,5,7,11,13], among others. Most of these tags are summarized in Table 1. Nevertheless, the selection of the optimal peptide tag and the right purification process for a specific target protein is not easy [26].

1.2. Selection of an optimal tag

One way of selecting the optimal tag is by assessing multiple tags generated randomly [2], which is a very expensive alternative in terms of computational resources and time. Another alternative is a systematic design, which considers the most widely used tags, the characteristics and purpose of the protein to be purified, and the expression system to be used. Tag selection will also depend upon the physicochemical properties of contaminants; for example, if most of contaminants are hydrophilic, it is convenient to have a hydrophobic tag allowing purification by hydrophobic interaction chromatography. Protein production, recovery and purification processes have usually been optimized on a unit-by-unit basis; for this reason, it would be very attractive to have a procedure to determine the purification sequence considering the global process instead of each unit on a separate basis. Methodologies based on the optimization of chemical processes have been extended to the synthesis of optimal bioprocesses [27–33]. These methodologies include heuristics based on information of physicochemical properties, so as to solve the synthesis process by reducing the search scope. Those methodologies have not considered the advantage of including modifications of the physiochemical properties of the product, such as polypeptide tags, in order to facilitate purification and diminish the number of stages in the global process. Recent work on selection of an optimal purification sequence has included the design of polypeptide tags [17,26]. Steffens et al. [17] proposed the use of combined methods to generate the best tag to be fused to a protein, showing "in silico" that processes having a few units, high recovery levels and low costs can be obtained. In order to obtain the optimal sequence, a cost function is minimized by using genetic algorithm software, considering the net charge and hydrophobicity of the amino acids making up the tag as the main properties to exploit. For the prediction of retention times, models for net charge [34] and hydrophobicity [35] were used. In Steffens's work [17], purification of Bovine Somatotropine (BST) was simulated and purification process diagrams of the fused protein were obtained, which have higher recovery levels and lower costs than with the original protein. However, relatively long tags were considered in this design (12–15 amino acids), and the possible interactions that can exist among those tags were not taken into account; nor were the needs to keep those amino acids exposed, or a possible loss in recovery. On the other hand, Simeonidis et al. [26] uses a whole non-linear programming model (MINLP), in which the target function is to minimize the number of chromatographic stages and length of the tag (with a maximum of eight amino acids). At the same time, they force themselves to get a specific purity level. For this purpose, it exploits the properties of the target proteins, the possible tags, and contaminants (hydrophobicity, charge, molecular weight). To carry out such optimization, it uses various models that allow prediction of behaviour of both the target protein and the system's contaminants. In general terms, the algorithm determines the composition of the shortest and most advantageous tag for the process, which implies a minimal number of stages. The main assumptions this model considers were (a) the tag is completely exposed on the surface; (b) the tag does not make secondary structures (α-helices, β sheets), and it does not interfere with the tertiary structure of the protein; (c) it is thought that no loss exists in any stage; (d) protein–protein interactions are negligible. These last two assumptions, depending on the tag type selected, such as hydrophobic tags, may not be valid, due to the multiple interactions that can be generated between these tags [6]. Particularly, the tags suggested by this model mainly present hydrophobic amino acids, but do not secure their exposure to the surface, conditions that should be validated by experimental studies; loss of product, protein–protein interaction and inclusion of buffer change and concentration stages should also be studied.

In the present work we developed a model based on Simeonidis's model [26]. However, our approach considers a finite number (26) of widely used small polypeptide tags and the objective function was the maximization of the purification process profit that takes into account number of stages and protein product recovery.

2. Theory

2.1. Problem description

Given a mixture of proteins and a desired product specification in terms of a minimum level of purity required, the problem is to select simultaneously the optimal polypeptide tag, from a set of widely used polypeptide tags, and the chromatographic steps that purify the desired protein from the contaminant ones, in order to maximize the purification process profit.

Then, the problem can be stated as follows:

Given:

- A mixture of target protein and contaminants ($p = 1,\ldots,P$) with known physicochemical and biochemical properties;
- A set of 11 available chromatographic techniques ($i = 1\ldots11$; anion exchange at pH 4, 5, 6, 7, 8, cation exchange at pH 4, 5, 6, 7, 8 and hydrophobic interaction);