Contents lists available at SciVerse ScienceDirect

Chemical Engineering Research and Design

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

An MILP formulation for the synthesis of protein purification processes

Eleftheria M. Polykarpou[a,b], *Paul A. Dalby*[a], *Lazaros G. Papageorgiou*[b,*]

[a] *Department of Biochemical Engineering, University College London (UCL), Torrington Place, London WC1E 7JE, UK*
[b] *Centre for Process Systems Engineering, Department of Chemical Engineering, University College London (UCL), Torrington Place, London WC1E 7JE, UK*

ABSTRACT

This paper presents a mixed integer linear programming (MILP) model for the optimal synthesis of chromatographic protein purification processes including the time line in which our target protein product is collected. The model is linearised using piecewise linear approximation strategies and tested on three example protein mixtures, containing up to 13 contaminants and selecting from a set of up to 21 candidate steps. The results are also compared with previous literature models attempting to solve the same problem and show that the proposed approach offers significant gains in computational efficiency without compromising the quality of the solution.

Keywords: Downstream processing; Protein purification; Mixed integer linear optimisation; Piecewise linear approximation

1. Introduction

Process chromatography has been the prime tool of the biotechnology industry over the last decades. Its development within the last 20 years resulted in a large rise of revenues of the major healthcare companies (Curling and Gottschalk, 2007). Although alternative bioseperation technologies are making their way in the market, process chromatography will remain the high resolution process for industries for the years to come (Przybycien et al., 2004).

Although chromatography has been around for decades, there is still a need for more efficient design and operation, since it has always been a major bottleneck for industry, because of its complexity and its high capital and operating costs (Ngiam et al., 2003). Downstream processing can account for up to 80% of the total manufacturing cost of the product (Lowe et al., 2001). This emphasises the need for new tools and strategies that can provide solutions for the challenge of downstream processing design (Nfor et al., 2008) which is also encouraged by the Food and Drug Administration (FDA) (FDA, 2009).

One of the major challenges to be addressed is the selection of the chromatographic steps employed in the purification process. In an average biochemical process, several chromatographic steps are required to achieve a product quality within confined specifications. However, biopharmaceutical companies usually operate in suboptimal conditions and for that reason, many efforts have focused on developing systematic approaches for the efficient design of process chromatography.

The first efforts focused on knowledge-based and heuristics (Ostlund, 1986; Asenjo et al., 1989; Wheelwright, 1989; Eriksson et al., 1991). However, these methods inherently hold the drawback of not determining the best solution because of the size of the design space. For this reason, many authors have tried to develop systematic methods in order to predict and optimise the different performance criteria (e.g. chromatographic steps) (Asenjo et al., 1989; Lienqueo et al., 1999; Lienqueo and Asenjo, 2000; Steffens et al., 2000). Later on, several authors developed mathematical models based mainly on mathematical programming. In Vasquez-Alvarez et al. (2001) and Vasquez-Alvarez and Pinto (2004), two MILP models were developed, utilising physicochemical properties of all components in the mixture, in order to synthesise the optimal flowsheet for a specified purity and recovery. More recently, mathematical models based on mixed-integer non-linear

* Corresponding author. Tel.: +44 020 7679 2563; fax: +44 020 7383 2348
E-mail address: l.papageorgiou@ucl.ac.uk (L.G. Papageorgiou).
Received 8 March 2011; Accepted 30 November 2011
0263-8762/$ – see front matter © 2011 The Institution of Chemical Engineers. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.cherd.2011.11.021

programming were developed by Simeonidis et al. (2005) that simultaneously select the optimal sequence of peptide tags and synthesise the purification process and later on linearised it by employing piecewise linear approximation (Natali et al., 2009).

In this work, a linear formulation is proposed based on the MINLP developed by our group (Polykarpou et al., 2009), using the piecewise linear approximation technique presented in Natali and Pinto (2009). In this model, not only the minimum number of chromatographic steps is determined, but also the time line in which the target protein product was collected. This novel linear model, can overcome the inherent drawbacks of its non-linear precursor.

The remainder of this article is structured as follows. In the next section, the problem for the downstream process synthesis is given, followed by the mathematical formulation, where the basics of chromatographic modelling and the piecewise approximations employed are described. Next, the numerical results are presented and analysed and the computational performance of the proposed formulation is evaluated. Finally, the main conclusions of this work are discussed.

2. Problem description

The overall problem for the synthesis of the purification processing can be stated as follows.

Given

- a mixture of proteins ($p: 1, \ldots, P$) with known physicochemical properties;
- a set of available chromatographic techniques ($i: 1, \ldots, I$), each performing a separation by exploiting a specific physicochemical property (charge or hydrophobicity);
- specifications for the desired protein (dp), in terms of minimum purity and recovery levels.

Determine

- optimal flowsheet of the purification process;
- operating starting and finishing cut-points.

So as to optimise the overall number of chromatographic steps to achieve purity and recovery specifications.

3. Mathematical formulation

In this section, an MILP model is proposed that is based on the MINLP model introduced by our group (Polykarpou et al., 2009). The model comprises two parts. Initially, the chromatographic separation model is presented along with the methodology and the actual equations that are the background for the optimisation model. Finally, the material balance for the selection of the optimum flowsheet are defined.

The objective function is to minimise the overall number of steps from a set of alternatives. Binary variable E_i is activated when a chromatographic step i is selected.

Objective function:

$$Min \quad S = \sum_i E_i \tag{1}$$

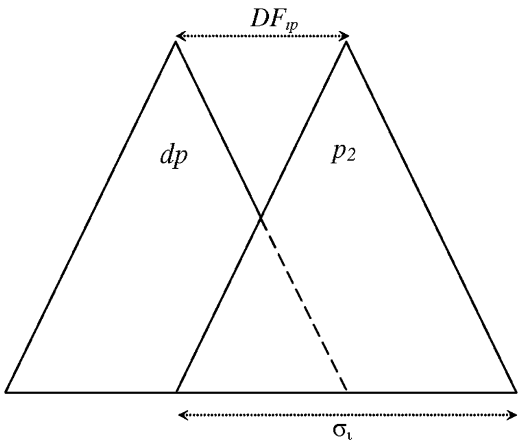

Fig. 1 – Representation of deviation factor, DF_{ip}.

3.1. Chromatographic separation model

As shown by Vasquez-Alvarez et al. (2001) and Lienqueo and Asenjo (2000), the chromatographic peaks are usually approximated by the use of isosceles triangles. The first parameter defined is the dimensionless retention time, KD_{ip}, which was experimentally determined to be a function of a characteristic physicochemical property, P_{ip}. The dimensionless retention time is characteristic for each protein p and each chromatographic technique i. The methodology presented in Lienqueo (1999) was used to estimate the dimensionless retention time for both ion exchange (IEX) and hydrophobic interaction chromatography (HIC). It was observed that the dimensionless retention time for IEX could successfully be described as a function of the charge densities (Q_{ip}/MW_p) for the operating conditions considered, as shown below.

Anion exchange chromatography

$$KD_{ip} = \begin{cases} \dfrac{8826 \cdot |Q_{ip}/MW_p|}{1.10^{-17} + 18875 \cdot |Q_{ip}/MW_p|} & \text{if} \quad Q_{ip} \le 0 \\ 0 & \text{if} \quad Q_{ip} \ge 0. \end{cases}$$

Cation exchange chromatography

$$KD_{ip} = \begin{cases} 0 & \text{if} \quad Q_{ip} \le 0 \\ \dfrac{7424 \cdot |Q_{ip}/MW_p|}{1.10^{-17} + 20231 \cdot |Q_{ip}/MW_p|} & \text{if} \quad Q_{ip} \ge 0. \end{cases}$$

For HIC, the dimensionless retention time can be described through a quadratic function of hydrophobicity based on the methodology proposed by Lienqueo et al. (2002).

$$KD_{ip} = -12.14 \cdot H_p^2 + 12.07 \cdot H_p - 1.74 \quad \forall i \in HI, p \in P \tag{2}$$

Although each protein p needs a different amount of time to elute from a different column/technique i, this information alone is not enough to quantify the efficiency of each chromatographic step. To do that the distance between peaks has to be considered. Deviation factors, DF_{ip}, are defined as the distance between two peaks (Fig. 1), one of them being the target protein's peak as shown in Vasquez-Alvarez et al. (2001).

$$DF_{ip} = KD_{ip} - KD_{i,dp} \quad \forall i, p \ne dp \tag{3}$$

As mentioned earlier the chromatograms are approximated by isosceles triangles. The peak width parameter, σ_i, is assumed to be dependant on the type of chromatographic operation and was calculated by averaging over several proteins (Vasquez-Alvarez et al., 2001; Lienqueo et al., 1996). For