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:

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

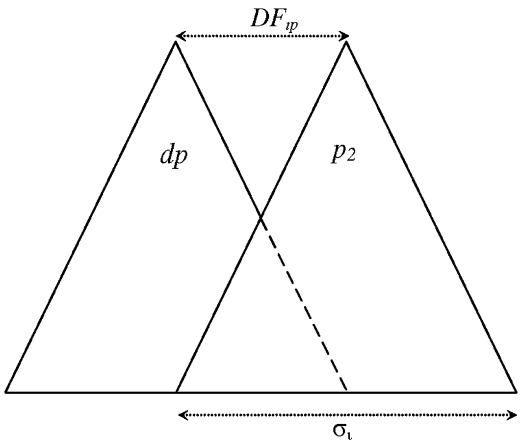


Fig. 1 – Representation of deviation factor, DF_{ip}.

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

ion exchange, the value for the peak width is $\sigma_i = 0.15$ and for hydrophobic interaction $\sigma_i = 0.22$ (Vasquez-Alvarez et al., 2001).

Finally, the efficiency of each chromatographic technique can be quantified by the concentration factor, CF_{ip}. The concentration factor is practically the ratio of the mass before and after each chromatographic technique i. As described in Vasquez-Alvarez and Pinto (2004) the concentration factor, CF_{ip}, is usually a function of DF_{ip} and σ_i. For this model though, some percentage of product losses is allowed. For this to be quantified, two extra variables are introduced. Starting cut-point, $xs_{i,dp}$, is the starting time for collecting the product and finishing cut-point, $xf_{i,dp}$, is the ending time for collecting our product (target protein). In order to calculate CF_{ip}, both $xs_{i,dp}$ and $xf_{i,dp}$ have to be determined first.

The mathematical expressions presented below represent the $CF_{i,dp}$ calculations for the target protein. A graphical representation is illustrated in Fig. 2 where the triangles refer to the target protein and the shaded areas represent the remaining amount of the target protein within the mixture after chromatographic technique i has been applied. It is important to note that three different cases may arise depending on the relative positions of the cut-points.

For the contaminants, depending on $xs_{i,dp}$, $xf_{i,dp}$ and DF_{ip}, new variables called shifted cut-points are introduced and defined below. The concentration factor is calculated based on the methodology shown in Fig. 2, but in this case CF_{ip} is also a function of DF_{ip} because of the shifted cut-points defined in Eqs. (4) and (5).

$$xs_{ip} = xs_{i,dp} - DF_{ip} \quad \forall i, p \neq dp \tag{4}$$

$$xf_{ip} = xf_{i,dp} - DF_{ip} \quad \forall i, p \neq dp \tag{5}$$

Next, the material balances for each protein in the mixture are neccesary. m_{ip} is the mass of each protein p after each chromatographic technique i and is calculated in the following set of constraints where mo_{ip} is the initial mass of each protein p in the mixture and m^1_{ip}, m^2_{ip} denote the masses after selection and no-selection of technique i (Simeonidis et al., 2005; Polykarpou et al., 2009).

$$\begin{aligned}
&m_{ip} = CF_{ip} \cdot mo_p \cdot E_i + mo_p \cdot (1 - E_i) \quad \forall i = 1, p \\
&m_{ip} = CF_{ip} \cdot m^1_{i-1,p} + m^2_{i-1,p} \quad \forall i \geq 2, p \\
&m_{i-1,p} = m^1_{i-1,p} + m^2_{i-1,p} \quad \forall i \geq 2, p \\
&m^1_{i-1,p} \leq U \cdot E_i \quad \forall i \geq 2, p \\
&m^2_{i-1,p} \leq U \cdot (1 - E_i) \quad \forall i \geq 2, p
\end{aligned} \tag{6}$$

where U is an appropriate upper bound.

Finally, the purity and recovery specifications are enforced by constraints (7) and (8).

$$m_{i,dp} \geq sp \cdot \sum_{p'} m_{ip'} \quad \forall i = I \tag{7}$$

$$m_{i,dp} \geq fr \cdot mo_{dp} \quad \forall i = I \tag{8}$$

3.2. Material balance transformation

The material balances shown in Eq. (6) use nonlinear terms given that the concentration factors are variables and depend on the selection of cut-points, $xs_{i,dp}$, $xf_{i,dp}$. In order to linearise this set of constraints, a strategy similar to that proposed by Natali et al. (2009) is followed. The final concentration for each protein in the mixture is given by the following relationship.

$$m_{Ip} = mo_p \cdot \prod_i \overline{CF_{ip}} \quad \forall p \tag{9}$$

where $\overline{CF_{ip}}$ is anew auxiliary variable defined by:

$$\begin{aligned}
&\overline{CF_{ip}} = CF_{ip} \quad \text{if} \quad E_i = 1 \quad \forall i, p \\
&\overline{CF_{ip}} = 1 \quad \text{if} \quad E_i = 0 \quad \forall i, p
\end{aligned} \tag{10}$$

Thus, variable $\overline{CF_{ip}}$ can be expressed as an exponential form:

$$\overline{CF_{ip}} = e^{(lnCF_{ip}) \cdot E_i} \quad \forall i, p \tag{11}$$

Therefore, by combining (9) and (11) and given that,

$$\overline{lnCF_{ip}} \equiv lnCF_{ip} \cdot E_i \quad \forall i, p \tag{12}$$

the mass of each protein p at the last chromatographic step I can be calculated as shown in Eq. (13).

$$m_{Ip} = mo_p \cdot e^{\sum_i lnCF_{ip} \cdot E_i} \quad \forall p \tag{13}$$

The final mass balance is shown in Eq. (14).

$$m_{Ip} = mo_p \cdot \xi_p, \quad where \quad \xi_p = e^{\sum_i \overline{lnCF_{ip}}} \quad \forall p \tag{14}$$

This is still a nonlinear equation, but now all the nonlinear terms are present in a single term, hence can be linearly approximated. In the next section, various piecewise linear approximations are described in order to remove all nonlinear terms in the model, to represent CF_{ip}, $lnCF_{ip}$ and ξ_p.

3.2.1. Piecewise linear approximations

As mentioned in the previous section, there are two non-linear parts in our model. The first one is relating the cut-points $xs_{i,dp}$, $xf_{i,dp}$ with the areas that lie below them, hence the concentration factors CF_{ip}. The second one relates CF_{ip} with $lnCF_{ip}$ and the last one $lnCF_{ip}$ with ξ_p. In total, three piecewise linear approximations are required.

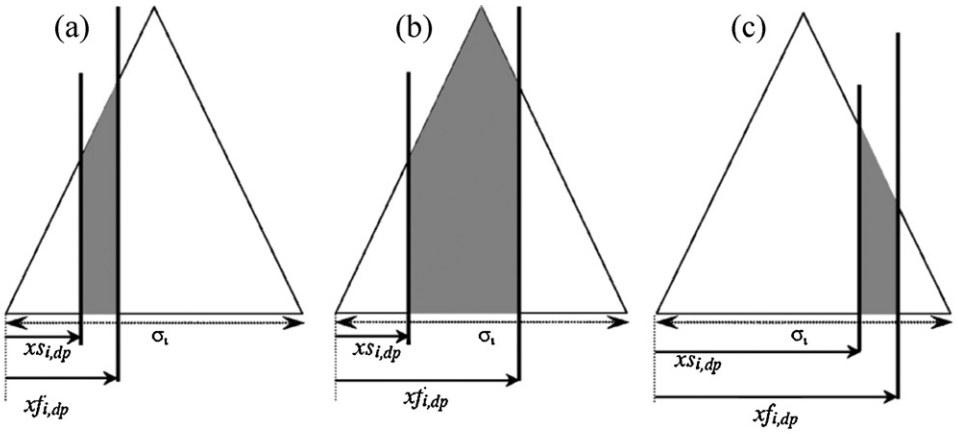
For all required linearisations, the approach presented in Natali and Pinto (2009) was employed in order to obtain the optimal points that approximate the relevant non-linear functions. A summary of the procedure is provided in Appendix A, where a set of points within the non-linear function is given, so that the resulting piecewise linear function is composed of all linear segments between the selected points.

Moving on to the first linearisation, cut-points $xs_{i,dp}$, $xf_{i,dp}$ are related with the areas that lie below them and represent the mass of the protein collected at that specific cut-point. The relevant constraints are shown below.

$$xs_{ip} = \sum_j xl_{ij} \cdot \lambda s_{ipj} \quad \forall i, p \tag{15}$$

$$As_{ip} = \sum_j Al_{ij} \cdot \lambda s_{ipj} \quad \forall i, p \tag{16}$$

$$\sum_j \lambda s_{ipj} = 1 \quad \forall i, p \tag{17}$$



$$(a)\quad xs_{i,dp}, xf_{i,dp} < \frac{\sigma_i}{2} \qquad CF_{i,dp} = \frac{2\cdot\left(xf_{i,dp}^2 - xs_{i,dp}^2\right)}{\sigma_i^2}$$

$$(b)\quad xs_{i,dp} < \frac{\sigma_i}{2}, xf_{i,dp} > \frac{\sigma_i}{2} \qquad CF_{i,dp} = 1 - \frac{2\cdot\left[(xs_{i,dp})^2 + (\sigma_i - xf_{i,dp})^2\right]}{\sigma_i^2}$$

$$(c)\quad xs_{i,dp}, xf_{i,dp} > \frac{\sigma_i}{2} \qquad CF_{i,dp} = \frac{2\cdot\left[(\sigma_i - xs_{i,dp})^2 + (\sigma_i - xf_{i,dp})^2\right]}{\sigma_i^2}$$

Fig. 2 – Representation of chromatographic peaks for the target protein.

$$xf_{ip} = \sum_j xl_{ij}\cdot\lambda f_{ipj} \quad \forall i, p \tag{18}$$

$$Af_{ip} = \sum_j Al_{ij}\cdot\lambda f_{ipj} \quad \forall i, p \tag{19}$$

$$\sum_j \lambda f_{ipj} = 1 \quad \forall i, p \tag{20}$$

In Eqs. (15)–(17) and (18)–(20) the starting and finishing cut-points are calculated along with the areas that lie below them. Parameters xl_{ij} and Al_{ij} define the piecewise linear points used, with xl_{ij} being the abscissa and Al_{ij} the ordinate. Variables λs_{ip}, λf_{ip} are of SOS2 type, so that at most two of them can be non-zero at the same time.

A representation of that function for IEX is shown in Fig. 3. For HIC the only difference is due to the fact that σ_i is equal to 0.22. The actual non-linear function is shown with the solid line, while the piecewise linear approximation is denoted by dotted line connecting to diamond points.

Having calculated the cut-points and the areas that lie below them, the concentration factor has to be calculated as well. But as described above for the mass balance we need $\overline{lnCF_{ip}}$. The function relating CF_{ip} and $lnCF_{ip}$ is graphically shown in Fig. 4 and the mathematical expression is described by Eqs. (21)–(25).

$$\overline{lnCF_{ip}} = \sum_k \beta_{ik}\cdot\mu_{ipk} + sl_{ip} \quad \forall i, p \tag{21}$$

$$\sum_j Al_{ij}\cdot\lambda f_{ipj} - \sum_j Al_{ij}\cdot\lambda s_{ipj} = \sum_k \alpha_{ik}\cdot\mu_{ipk} \quad \forall i, p \tag{22}$$

$$\sum_k \mu_{ipk} = 1 \quad \forall i, p \tag{23}$$

$$sl_{ip} \le -ln(D)\cdot(1 - E_i) \quad \forall i, p \tag{24}$$

$$-ln(D)\cdot E_i \ge \overline{lnCF_{ip}} \ge ln(D)\cdot E_i \quad \forall i, p \tag{25}$$

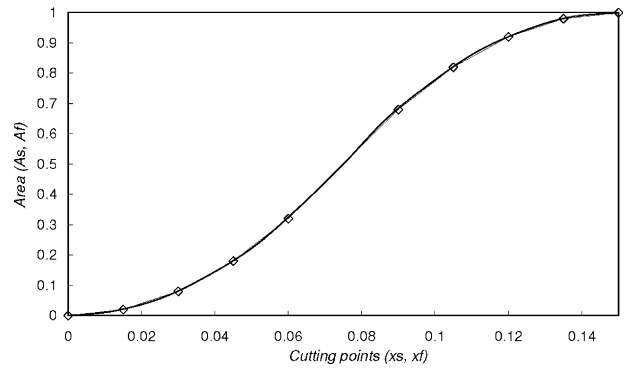


Fig. 3 – Linearisation 1: areas (As_{ip}, Af_{ip}) vs. cutting points (xs_{ip}, xf_{ip}) for IEX.

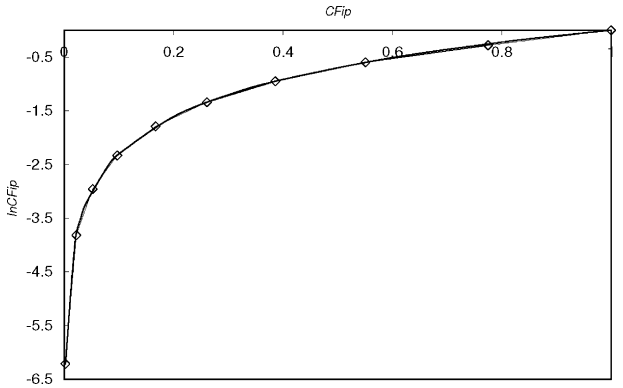


Fig. 4 – Linearisation 2: $lnCF_{ip}$ vs. concentration factor CF_{ip}.

Table 1 – Physicochemical properties of protein mixture (example 1).

Protein	mo_p (mg/ml)	MW_p (Da)	H_p	Q_{ip} (C/molecule) $\times 10^{-17}$				
				pH 4	pH 5	pH 6	pH 7	pH 8
dp	2	67,000	0.86	1.03	−0.14	−1.16	−1.68	−2.05
p1	2	43,800	0.54	1.40	−0.76	−1.65	−2.20	−2.36
p2	2	24,500	0.90	1.22	−0.76	−1.54	−2.17	−2.13
p3	2	22,200	0.89	1.94	1.90	1.98	−1.87	0.91

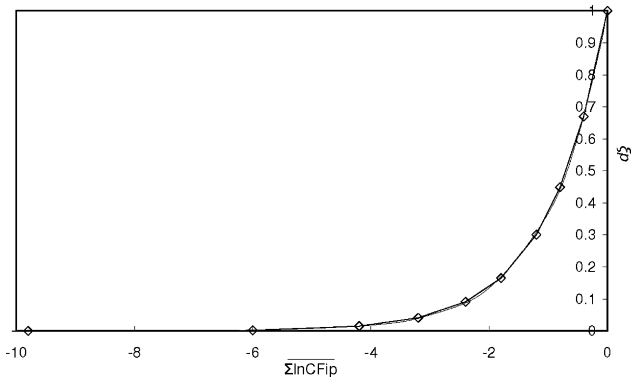


Fig. 5 – Linearisation 3: ξ_p vs. $\sum_i \overline{lnCF_{ip}}$.

Parameters β_{ik} and α_{ik} define the piecewise linear approximations used, with α_{ik} being the abscissa and β_{ik} the ordinate. Variable μ_{ipk} is of a SOS2 type. In Eq. (22), the first term refers to the area that lies below the finishing cut-point, the second term to the area that lies below the starting cut-point and their product is CF_{ip}. Finally, slack variable sl_{ip} is imposed that $\overline{lnCF_{ip}}$ is equal to zero when no separation takes place (i.e. $E_i = 0$) through Eqs. (24) and (25), where D is a small number.

From Eq. (14), the final concentrations of all proteins in the mixture are calculated. This nonlinear equation can be linearised in a similar way as described above. Parameters γ_l and δ_l are the values of the ordinate and abscissa, respectively, and along with SOS2 variable ν_{pl} define the exponential piecewise linear approximation (see Fig. 5) described by:

$$\sum_i \overline{lnCF_{ip}} = \sum_l \gamma_l \cdot \nu_{pl} \quad \forall p \tag{26}$$

$$\xi_p = \sum_l \delta_l \cdot \nu_{pl} \quad \forall p \tag{27}$$

$$\sum_l \nu_{pl} = 1 \quad \forall p \tag{28}$$

4. System definition

Below, a summary of the mathematical proposed model is presented. The objective is to minimise the overall number of chromatographic steps and is subject to the following constraints.

$$Min \quad S = \sum_i E_i$$

Subject to:

- Eqs. (15)–(20), where cut-points xs_{ip}, xf_{ip} along with the areas As_{ip}, Af_{ip} are calculated.
- Eqs. 21–(25), where CF_{ip} against $\overline{lnCF_{ip}}$ is approximated.
- Eqs. (26)–(28), where ξ_p is calculated.
- Eq. (14), where the mass balance is described.
- Eqs. (7) and (8), where the purity and recovery specifications are enforced.

The overall problem is formulated as a mixed integer linear programming (MILP) model. Following the same solution approach as that presented in Polykarpou et al. (2009), first a screening MILP (Vasquez-Alvarez and Pinto, 2004) is solved, in order to determine candidate chromatographic steps, followed by the proposed MILP over the reduced set of alternatives (determined by the first stage).

5. Results and discussion

In this section, the solutions of the proposed model are analysed. The methodology was tested with three examples modeled in the GAMS 22.8 (Brooke et al., 2008). Solutions for the MILP and MINLP models were obtained using the CPLEX and BARON solvers respectively, on a Dell Desktop Core Duo 3.25 GB RAM 3.16 GHz machine.

5.1. *Example 1*

This first example is based on experimental data taken from Vasquez-Alvarez et al. (2001) involving serum from bovine albumin (*dp*), ovalbumin (p_1), soybean trypsin inhibitor (p_2) and thaumatin (p_3). The physicochemical properties as well as the initial protein concentration of the mixture are given in Table 1. In summary, there are 11 candidate chromatographic steps: 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).

The resulting mathematical model involves 661 constraints, 521 continuous variables, and 427 binary variables and was solved in 0.3 s. The optimal solution is presented in Fig. 6, where the value above the arrow is the purity, and below is the recovery achieved. The model was able to identify a solution that achieves purity $sp = 0.981$ and recovery $fr = 0.983$ for the target protein, for which two steps are required: *AE7*, *HI*. The cut-points for *AE7* were: $xs_{AE7,dp} = 0.004$ and $xf_{AE7,dp} = 0.143$ and for *HI*: $xs_{HI,dp} = 0.002$ and $xf_{HI,dp} = 0.220$ and are also presented in Fig. 7.

5.2. *Example 2*

This example utilises data available on Vasquez-Alvarez et al. (2001). The mixture includes target protein β-1,3 glucanase from *Bacillus Subtilis* and 8 contaminants. Physicochemical properties along with initial concentrations of the protein mixture are available in Fig. 2. Overall, there are 21 candidate chromatographic steps. Besides the ones presented in

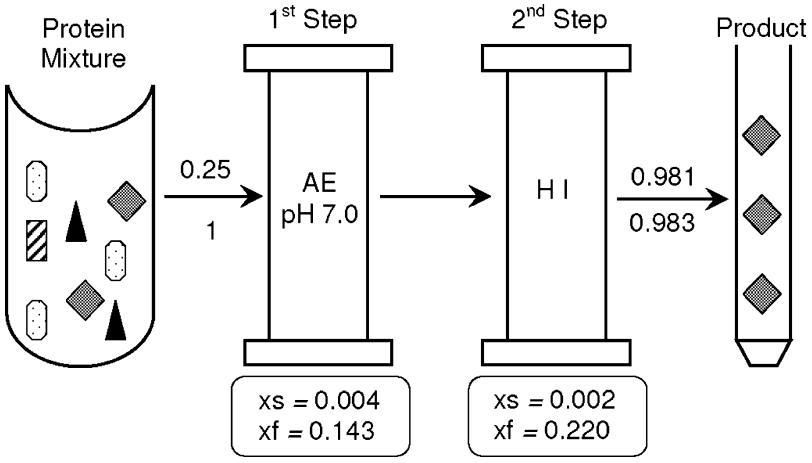


Fig. 6 – Optimal flowsheet for purification of protein mixture (example 1).

Table 2 – Physicochemical properties of protein mixture (example 2).

Protein	mo_p (mg/ml)	MW_p (Da)	H_p	Q_{ip} (C/molecule) $\times 10^{-17}$									
				pH 4	pH 4.5	pH 5	pH 5.5	pH 6	pH 6.5	pH 7	pH 7.5	pH 8	pH 8.5
dp	0.62	31,000	0	1.46	0.09	−0.62	−0.66	−1.02	−1.82	−2.33	−2.52	−2.52	−3.51
p1	0.42	62,500	0	1.46	0.09	−1.06	−0.98	1.17	−1.71	−2.79	−3.52	−3.32	−3.32
p2	0.25	40,600	0	1.46	0.09	−0.55	−0.22	−0.22	−0.26	−0.73	−1.26	−1.82	−3.51
p3	0.25	69,600	0	1.46	0.09	−0.55	−0.22	−0.22	−0.26	−0.73	−1.26	−1.82	−3.51
p4	0.09	40,600	0	1.46	3.14	1.46	0.28	−0.47	−0.89	−1.06	−1.08	−1.04	−1.01
p5	0.09	69,600	0	1.46	3.14	1.46	0.28	−0.47	−0.89	−1.06	−1.08	−1.04	−1.01
p6	2.74	41,000	1.5	1.46	0.93	0.26	−0.35	−0.87	−1.31	−1.65	−1.9	−2.04	−2.06
p7	2.74	32,900	1.5	1.46	0.09	0	−1.7	−2.7	−2.9	−3.51	−3.51	−3.51	−3.51
p8	0.25	35,500	0.2	1.46	0.09	−0.55	−0.22	−0.22	−0.26	−1.26	−1.82	−1.82	−3.51

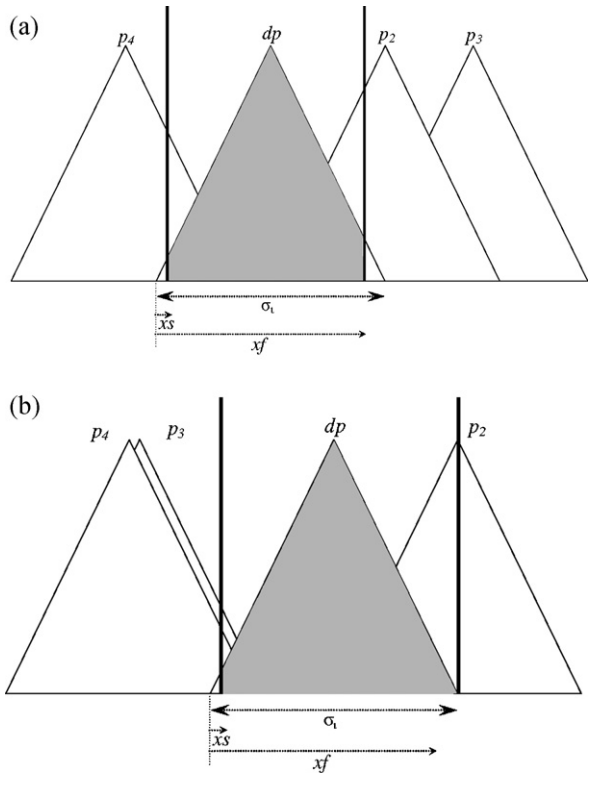


Fig. 7 – Solution of example 1.

example 1, there are additional steps: anion exchange chromatography (AE) at pH 4.5, AE at pH 5.5, AE at pH 6.5, AE at pH 7.5, AE at pH 8.5, cation exchange chromatography (CE) at pH 4, CE at pH 5, CE at pH 6.5, CE at pH 7.5, CE at pH 8.5.

This example involves 2835 constraints, 2224 continuous variables, and 1824 binary variables and was solved in 4.2 s. The optimal solution is presented in Fig. 8, where a purity of $sp = 0.951$ and a recovery of $fr = 0.94$ is achieved after three steps: *AE6.5*, *AE8.5*, *HI*. The cut-points for *AE6.5* were: $xs_{AE6.5,dp} = 0.012$ and $xf_{AE6.5,dp} = 0.147$, for *AE8.5* were: $xs_{AE8.5,dp} = 0.013$ and $xf_{AE8.5,dp} = 0.150$ and for *HI*: $xs_{HI,dp} = 0$ and $xf_{HI,dp} = 0.198$.

5.3. *Example 3*

For our final example, data taken from Simeonidis et al. (2005) was used. This specific example is the largest one of the three and the more complex in terms of separation potential. It involves 13 proteins and all the necessary information are presented in Table 3. There are 11 candidate chromatographic steps as presented in example 1.

It takes 7.3 s to obtain the optimal solution and includes 3451 constraints, 2215 continuous variables, and 2705 binary variables. The optimal solution achieved, is presented in Fig. 9. Two steps are required: *AE7*, *CE4*, *HI* in order to achieve a purity of $sp = 0.93$ and a recovery of $fr = 0.90$ for the target protein. The cut-points for *AE7* were: $xs_{AE7,dp} = 0.013$ and $xf_{AE7,dp} = 0.134$, for *CE4* were: $xs_{CE4,dp} = 0.007$ and $xf_{CE4,dp} = 0.133$ and for *HI*: $xs_{HI,dp} = 0.013$ and $xf_{HI,dp} = 0.212$.

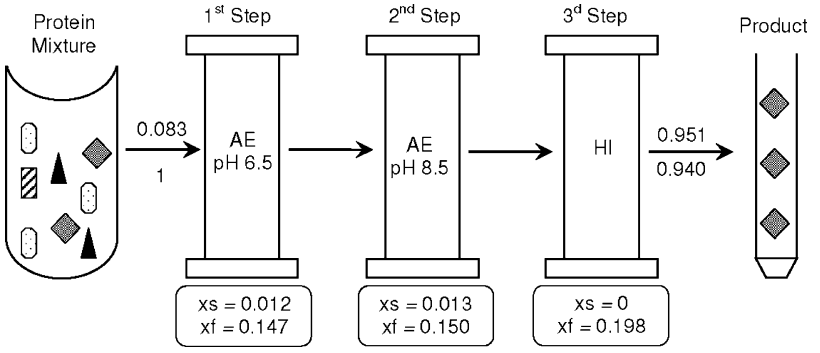


Fig. 8 – Optimal flowsheet for purification of protein mixture (example 2).

Table 3 – Physicochemical properties of protein mixture (example 3).

Protein	mo_p (mg/ml)	MW_p (Da)	H_p	Q_{ip} (C/molecule) $\times 10^{-17}$				
				pH 4	pH 5	pH 6	pH 7	pH 8
dp	2	77,000	0.28	2.04	1.06	−0.37	−0.81	−1.13
p1	2	22,000	0.27	1.60	1.57	1.56	1.55	0.75
p2	2	23,600	0.31	2.15	1.46	1.17	0.78	0.38
p3	2	13,500	0.23	1.83	0.65	0.26	−0.20	−0.33
p4	2	43,800	0.28	1.16	−0.63	−1.36	−1.82	−1.95
p5	2	15,900	0.27	2.89	2.81	2.8	2.64	2.07
p6	2	14,400	0.32	−0.46	−0.47	−0.63	−1.21	−1.25
p7	2	17,500	0.21	0.45	−0.62	−0.79	−1.26	−1.7
p8	2	50,000	0.27	−0.12	−0.32	−0.76	−0.91	−1.04
p9	2	12,100	0.18	1.46	0.62	−1.02	−1.33	−1.52
p10	2	25,500	0.30	1.01	−0.63	−1.27	−1.59	−1.76
p11	2	26,000	0.28	2.96	1.26	0.92	0.54	0.01
p12	2	19,900	0.25	0.93	0.33	−0.12	−0.34	−0.5

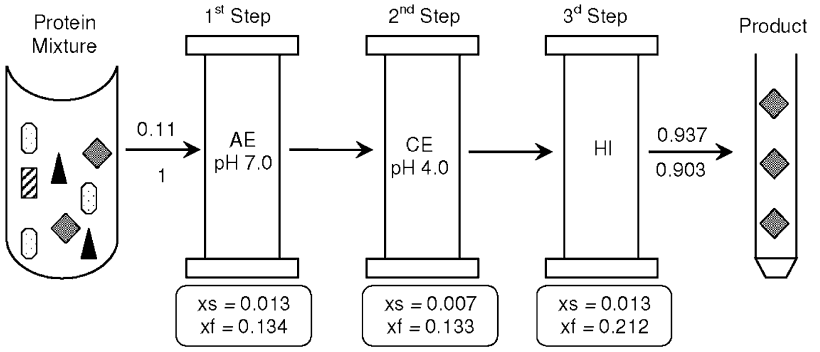


Fig. 9 – Optimal flowsheet for purification of protein mixture (example 3).

5.4. *Comparative results*

In an effort to demonstrate the benefits of the proposed model, a comparison with the MINLP approach introduced in Polykarpou et al. (2009) is undertaken. The MILP model was solved for five, ten and fifteen internal knots for the piecewise linear approximation. All computational results are summarised in Table 4, For all examples ten internal knots were sufficient to obtain the optimal solution. Using five knots was not adequate for the first two examples, since it resulted in sub-optimal solution.

Moreover, in terms of CPU savings the MILP model was able to solve all examples in less than 10 s as shown in Fig. 10. It is quite interesting, that although the MINLP model has fewer constraints and has even six times fewer binary variables, is even seventy times less efficient than the proposed MILP.

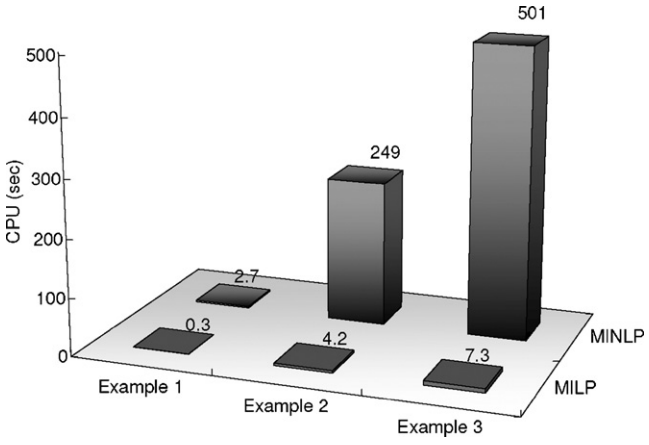


Fig. 10 – Comparison between MINLP (Polykarpou et al., 2009) and proposed MILP.

Table 4 – Computational statistics.

Example	Model	NoC[a]	NoCV/NoBV[b]	CPU (s)	Obj. value
1	MINLP[c]	237	101/63	2.7	2
	MILP[d]	461	321/227	0.1	3
	MILP[e]	661	521/427	0.3	2
	MILP[f]	861	721/627	0.5	2
2	MINLP[c]	1188	499/306	249	3
	MILP[d]	1980	1369/969	0.9	5
	MILP[e]	2835	2224/1824	4.2	3
	MILP[f]	3690	3079/2679	5.2	3
3	MINLP[c]	1454	605/375	501	3
	MILP[d]	2411	1665/1175	2.7	4
	MILP[e]	3451	2705/2215	7.3	3
	MILP[f]	4491	3745/3255	117	3

[a] No of constraints.
[b] No. of continuous variables/no. of binary variables.
[c] Polykarpou et al. (2009).
[d] 5 Intervals.
[e] 10 Intervals.
[f] 15 Intervals.

6. Conclusions

In this paper, a novel MILP model formulation has been presented for tackling the problem of downstream protein processing synthesis. This model simultaneously optimises the process flowsheet composed of distinct chromatographic steps and determines the specific cut-points for product collection by allowing product losses. Further comparisons with previously published models underlined the efficiency of the proposed formulation, which was able to obtain the optimal solutions with significantly less computational time required.

In terms of future work, the modelling of the purification process can be extended in order to incorporate protein–protein interaction and sequencing of the purification steps. Finally, the application of a different objective function that incorporates process economics (product sale price vs. production cost) and evaluates the design variables is a challenge yet to be addressed.

Acknowledgement

EMP acknowledges financial support from Innovative Manufacturing Research Centre. PAD and LGP acknowledge support from the EPSRC Centre for Innovative Manufacturing in Emergent Macromolecular Therapies.

Appendix A. Piecewise linear approximation

Below there is a description of the optimal approximations of single-dimensional nonlinear functions by piecewise linear functions as described by Natali and Pinto (2009). The approach uses a discrete representation of the non-linear function described by pairs (x_i, f_i), $i \in Q = 1, 2, \ldots n_Q$, where Q is the predefined sampling set. Binary variable W_{ij} is equal to 1 if $i \in Q$ and $j \in Q$ are two consecutive points, otherwise is 0. N is the number of knots given *a priori*.

At most one polynomial piece of the approximating function may begin and one piece may end in each of the points in Q.

$$\sum_{\substack{j \in Q \\ j > i}} W_{ij} \leq 1 \quad \forall i \in Q | i > 1 \tag{A.1}$$

$$\sum_{\substack{i \in Q \\ j > i}} W_{ij} \leq 1 \quad \forall j \in Q | j > n_Q \tag{A.2}$$

The first and last points of Q are necessarily part of the knots.

$$\sum_{\substack{j \in Q \\ j > i}} W_{ij} = 1 \quad \forall i = 1 \tag{A.3}$$

$$\sum_{\substack{i \in Q \\ j > i}} W_{ij} = 1 \quad \forall j = n_Q \tag{A.4}$$

Any knot has to be both the start and the end of a polynomial piece of the approximating function and the end of another (except of the first and last ones).

$$\sum_{\substack{i \in Q \\ k > i}} W_{ik} = \sum_{\substack{i \in Q \\ j > k}} W_{kj} \quad \forall k > 2 \ldots n_Q - 1 \tag{A.5}$$

The approximating function is predefined to have N internal knots.

$$\sum_{i \in Q} \sum_{\substack{j \in Q \\ j > i}} W_{ij} = N - 1 \tag{A.6}$$

The values of the approximating function are defined by the following set of constraints.

$$f_k^P = \sum_{\substack{i \in Q \\ i > k}} \sum_{\substack{j \in Q \\ j > k}} \frac{[(x_k - x_i) \cdot f_j + (x_j - x_k) \cdot f_i]}{(x_j - x_i)} \cdot W_{ij} \quad \forall k \in Q \qquad (A.7)$$

To measure the quality of the approximation, the 1-norm of distance between the vectors describing the original function and the piecewise linear approximation. The objective function is to minimise this norm and is given by the following constraints.

$$Z = \sum_{i \in Q} z_i \qquad (A.8)$$

$$z_i \geq (f_i - f_i^P) \quad \forall i \in Q \qquad (A.9)$$

$$z_i \geq -(f_i - f_i^P) \quad \forall i \in Q \qquad (A.10)$$

References

Asenjo, J.A., Herrera, L., Byrne, B., 1989. Development of an expert system for selection and synthesis of protein purification processes. Journal of Biotechnology 11, 275–298.

Brooke, A., Kendrick, D., Meeraus, A., Raman, 2008. GAMS—A User's Guide. GAMS Development Corporation.

Curling, S., Gottschalk, U., 2007. Process chromatography: five decades of innovation. Biopharm International 20, 10–19.

Eriksson, H., Sandahl, K., Forslund, G., Gsterlund, B., 1991. Knowledge-based planning for protein purification. Chemometrics and Intelligent Laboratory Systems: Laboratory Information Management 13, 173–184.

FDA, 2009. International Committee on Harmonization Q8: Pharmaceutical Development FDA.

Lienqueo, M.E., 1999. Desarollo de un sistema experto para la selection racional de processos de purification de proteinas:optimizacion de criterios se seleccion de sequenciase, Ph.D. Thesis, Universidad De Chile Facultad de Ciencias Fisicas y matematicas departamento de Ingeniera Quimica.

Lienqueo, M.E., Asenjo, J.A., 2000. Use of expert systems for the synthesis of downstream protein processes. Computers and Chemical Engineering 24, 2339–2350.

Lienqueo, M.E., Leser, E.W., Asenjo, J.A., 1996. An expert system for the selection and synthesis of multi-step protein separation processes. Computers Chemical Engineering 20, 189–194.

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, 293–299.

Lienqueo, M.E., Mahn, A., Asenjo, J.A., 2002. Mathematical correlations for predicting protein retention times in hydrophobic interaction chromatography. Journal of Chromatography A 978, 71–79.

Lowe, C.R., Lowe, A.R., Gupta, G., 2001. New developments in affinity chromatography with potential application in the production of biopharmaceuticals. Journal of Biochemical and Biophysical Methods 49, 561–574.

Natali, J.M., Pinto, J.M., 2009. Piecewise polynomial interpolations and approximations of one-dimensional functions through mixed integer linear programming. Optimization Methods and Software 24, 783–803.

Natali, J.M., Pinto, J.M., Papageorgiou, L.G., 2009. Efficient MILP formulations for the simultaneous optimal peptide tag design and downstream processing synthesis. AIChE Journal 55, 2303–2317.

Nfor, B.K., Ahamed, T., van Dedem, G.W.K., van der Wielen, L.A.M., van de Sandt, E.J.A.X., Eppink, M.H.M., Ottens, M., 2008. Design strategies for integrated protein purification processes: challenges, progress and outlook. Journal of Chemical Technology and Biotechnology 83, 124–132.

Ngiam, S.H., Bracewell, D.G., Zhou, Y., Titchener-Hooker, N.J., 2003. Quantifying process tradeoffs in the operation of chromatographic sequences. Biotechnology Progress 19, 1315–1322.

Ostlund, C., 1986. Large-scale purification of monoclonal antibodies. Trends Biotechnology 4, 288–293.

Polykarpou, E.M., Dalby, P.A., Papageorgiou, L.G., 2009. An MINLP formulation for the synthesis of chromatographic protein purification processes with product loss. In: European Symposium on Computer Aided Process Engineering-19, Vol. 26 of Computer-Aided Chemical Engineering, pp. 1057–1062.

Przybycien, T.M., Pujar, N.S., Steele, L.M., 2004. Alternative bioseparation operations: life beyond packed-bed chromatography. Current Opinion in Biotechnology 15, 469–478.

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. Biotechnology Progress 21, 875–884.

Steffens, M.A., Fraga, E.S., Bogle, I.D.L., 2000. Synthesis of purification tags for optimal downstream processing. Computers and Chemical Engineering 24, 717–720.

Vasquez-Alvarez, E., Pinto, J.M., 2004. Efficient MILP formulation for the optimal synthesis of chromatographic protein purification processes. Journal of Biotechnology 110, 295–311.

Vasquez-Alvarez, E., Lienqueo, M.E., Pinto, J.M., 2001. Optimal synthesis of protein purification processes. Biotechnology Progress 17, 685–696.

Wheelwright, S.M., 1989. The design of downstream processes for large-scale protein purification. Journal of Biotechnology 11, 89–102.