Optimization of a Biotechnological Multiproduct Batch Plant Design for the Manufacture of Four Different Products: A Real Case Scenario

Gabriela Sandoval,[1,2] Daniel Espinoza,[3] Nicolas Figueroa,[4] Juan A. Asenjo[1]

[1] Centre of Biotechnology and Bioengineering, CeBiB, Department of Chemical Engineering and Biotechnology, University of Chile, Beauchef 851, Santiago 8370456, Chile; telephone: +56 22 9784 723; fax: +56 22 6991 084; e-mail: juasenjo@ing.uchile.cl
[2] Facultad de Ingeniería y Tecnología, Universidad San Sebastián, Santiago, Chile
[3] Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería Industrial, Universidad de Chile, Santiago, Chile
[4] Instituto de Economía, Pontificia Universidad Católica de Chile, Santiago, Chile

ABSTRACT: In this work a biotechnological multiproduct batch plant that manufactures four different recombinant proteins for human application is described in some detail. This batch plant design is then optimized with regards to the size of equipment using a mixed-integer linear programming (MILP) formulation recently developed by us in order to find a hypothetical new biotechnological batch plant based on the information of real known processes for the production of the four recombinant protein products. The real plant was divided for practical purposes into two sub-processes or facilities: a fermentation facility and a purification facility. Knowing the specific steps conforming the downstream processing of each product, size, and time factors were computed and used as parameters to solve the aforementioned MILP reformulation. New constraints were included to permit the selection of some equipment—such as centrifuges and membrane filters—in a discrete set of sizes. For equipment that can be built according to customer needs—such as reactors—the original formulation was retained. Computational results show the ability of this optimization methodology to deal with real data giving reliable solutions for a multiproduct batch plant composed of 44 unit operations in a relatively small amount of time showing that in the case studied it is possible to save up to a 66% of the capital investment in equipment given the cost data used.
Biotechnol. Bioeng. 2017;114: 1252–1263.

Introduction

The design of multi-product batch plants using an optimization-based approach has been studied for more than 40 years and different approaches to deal with the complexity of the optimization models that result in non-convex mixed-integer non-linear problems (MINLP) goes from the development of new algorithms that are able solve these type of problems (Borisenko et al., 2011; Kocis and Grossmann, 1989; Li et al., 2012; Viswanathan and Grossmann, 1990) to the reformulation of them into mixed-integer linear problems (MILP) (Moreno and Montagna, 2011; Sandoval et al., 2016; Voudouris and Grossmann, 1992) that can be solved using the known accurate commercial solver CPLEX. In the majority of the cases, the aim of these efforts is to be able to find the optimal design of multi-product batch plants in real scenarios which is, according to Barbosa-Póvoa (2007), still a challenge.

In this work, the methodology developed by Sandoval et al. (2016) is applied to a real case scenario. A biotechnological multiproduct batch plant that manufactures four different recombinant proteins for human application is described in some detail. This batch plant design is then optimized with regards to the size of equipment using a MILP formulation in order to find a hypothetical new biotechnological batch plant based on the information of real known processes for the production of the four recombinant protein products: (P1–P4). (P2) and (P3) are synthesized as part of inclusion bodies in a recombinant *Escherichia coli* while (P1) and (P4) are synthesized by a recombinant *Saccharomyces cerevisiae* as intracellular and extracellular products, respectively. Each production process is composed of 8–21 processing stages of which about 14 can be shared by two or more individual processes. As shown in Figures 1 and 2 and Table I the real plant was divided, for practical purposes, into two sub-processes or facilities: a fermentation facility and a purification facility. As the equipment involved in each stage may be selected from the equipment available offered by different manufacturers or
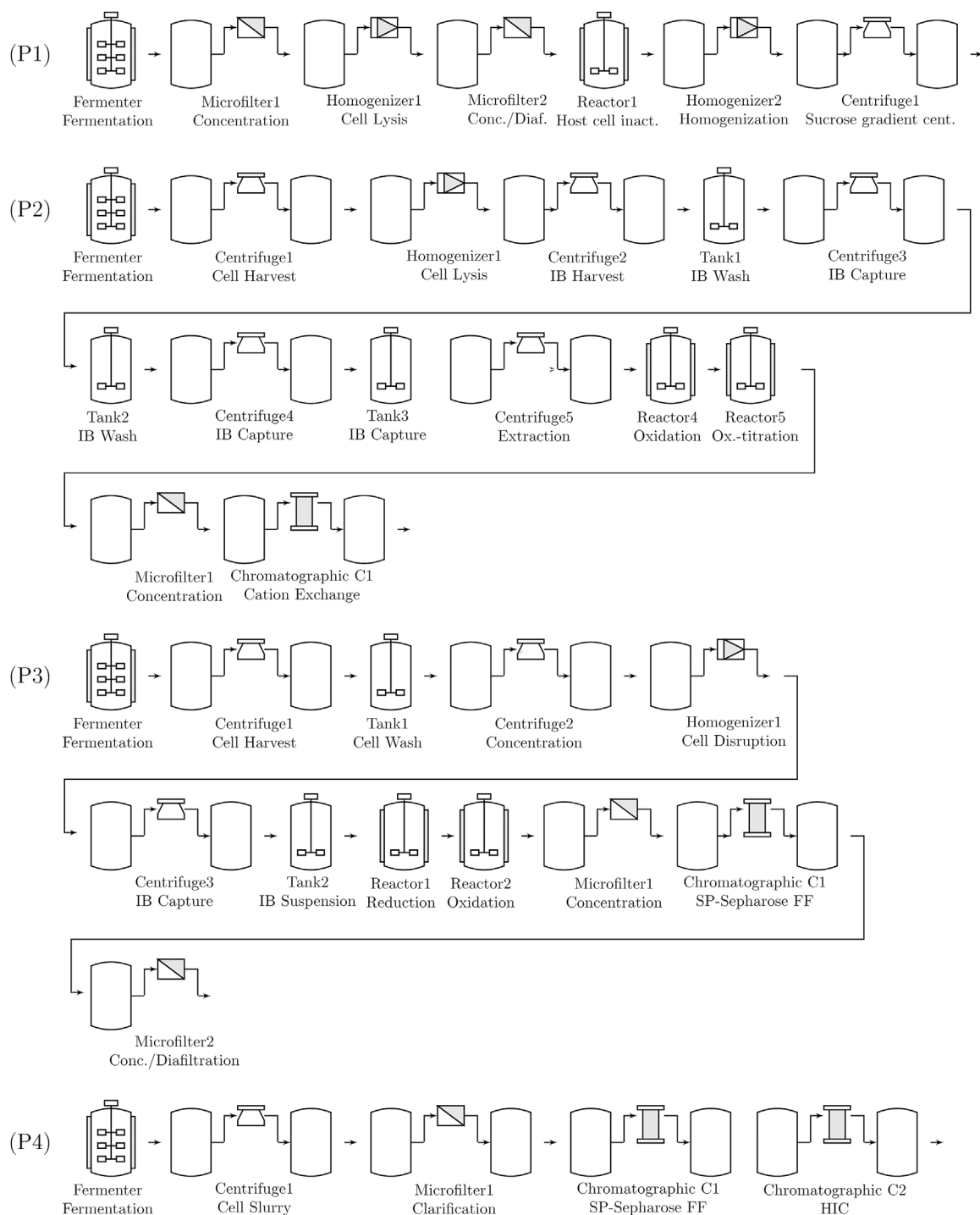
Figure 1. Production processes in the original Fermentation Plant. (P1) and (P4) are intra and extra-cellular products, respectively, synthesized in *S. cerevisiae*. (P2) and (P3) are intracelullar products synthesized in *E. coli*.

sized according to the customer needs, a new selection decision was introduced to the routing model presented by Sandoval et al. (2016), which now allows for the selection of a discrete number of possible sizes when it is necessary.

The computational results obtained show that the methodology developed earlier is capable of solving the optimization problem of a real type biotechnological multi-product batch plant—with 44 operational stages—reliably and in a small amount of time.

The rest of this paper is organized as follows. In Design of a Biotechnological Multiproduct Batch Plant section, the known processes to manufacture the four proteins studied are presented together to the proposed structure of a new multiproduct batch
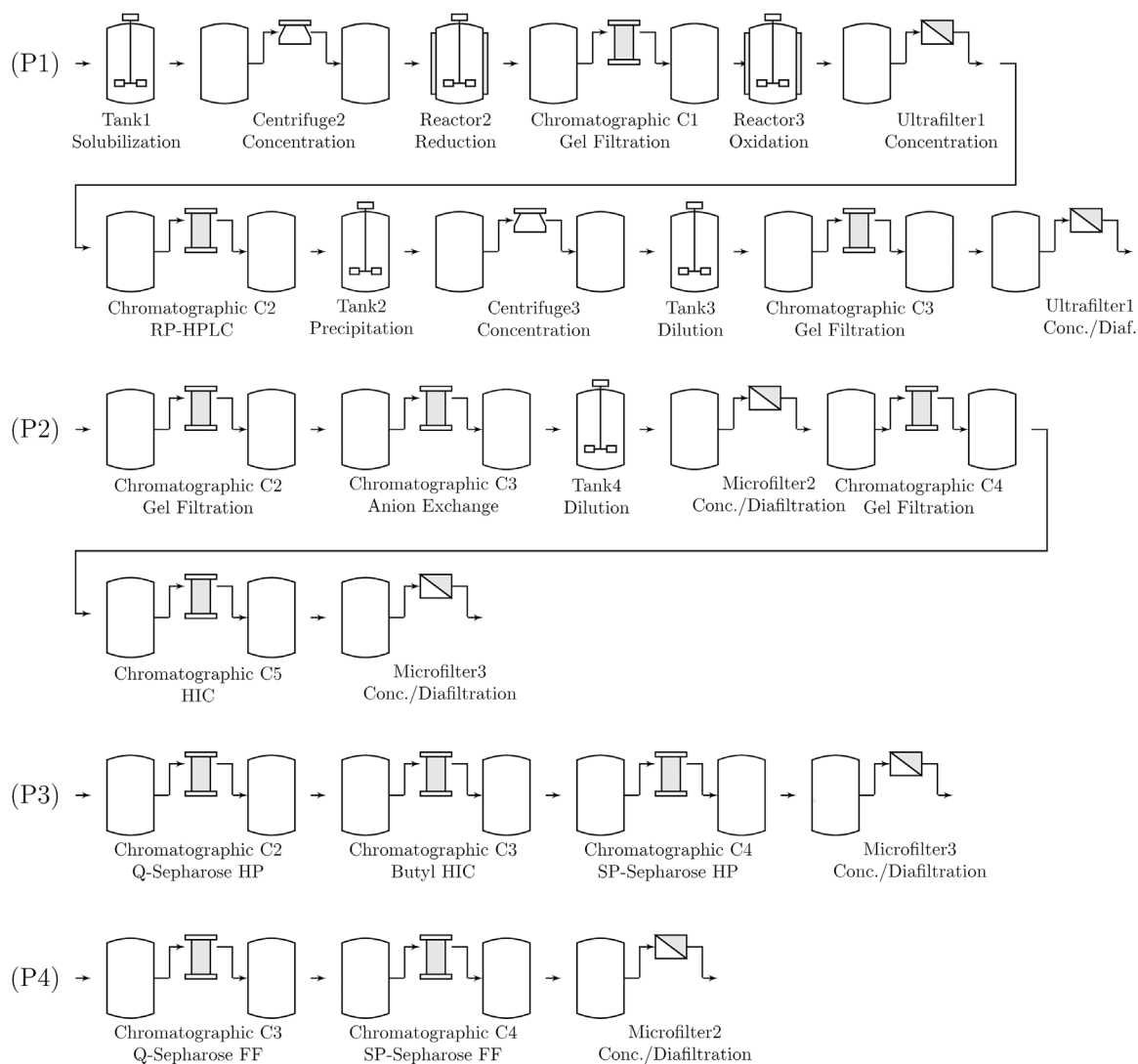
Figure 2. Production processes in the original Purification Plant. (P1) and (P4) are intra and extracel-lular products, respectively, synthesized in *S. cerevisiae*. (P2) and (P3) are intracelullar products synthesized in *E. coli*.

plant that produce them. In Problem Formulation section, the Problem Formulation and the equations used for parameter estimation is presented. Results are discussed in Results and Discussion section and conclusions are presented in Conclusions section.

Design of a Biotechnological Multiproduct Batch Plant

Processes Description

In this section, the downstream processes of the four recombinant proteins studied are described. According to Imperatore and Asenjo (2001), Product 1 is purified in 18 stages, while Products 2 and 3 need 20 and 15 isolation and purification steps, respectively as shown in Figures 1 and 2 and Table I. Product 4 on the other hand,

only needs seven processing stages given its nature of extracellular product.

The original plant is divided into two facilites. In the first, fermentation and primary purification stages are included; in the second, high resolution steps are performed in four different production lines, each processing one or more products per year under single product campaigns. General production data of the original processes and production level are presented in Table II.

Product 1

Product 1 (P1) is an intracellular protein synthesized by *S. cerevisiae* in a fermentation process that lasts 23 h. After this step a concentration stage follows performed by a microfilter were cells are collected in the retentate stream for a subsequent homogenization step. In order to wash the homogenization buffer a concentration

	Stage		Description	(P1)	(P2)	(P3)	(P4)
Fermentation	1	fer	Fermentation	x	x	x	x
	2	mf1	Concentration	x			
	3	cnt1	Cell harvest		x	x	
	4	tnk1	Cell wash			x	
	5	cnt2	Cell concentration			x	
	6	hom1	Cell lysis	x	x	x	
	7	mf2	Conc./diafiltration	x			
	8	rct1	Cellular inactivation	x			
	9	hom2	Homogenization	x			
	10	cnt3	Concentration	x			
	11	cnt4	IB harvest		x		
	12	tnk2	IB wash		x		
	13	cnt5	IB capture		x		
	14	tnk3	IB wash		x		
	15	cnt6	IB capture		x	x	
	16	tnk4	IB suspension		x	x	
	17	cnt7	Extraction		x		
	18	rct2	Reduction			x	
	19	rct3	Oxidation		x	x	
	20	rct4	Oxidation-titration		x		
	21	mf3	Concentration		x	x	
	22	cnt8	Cell slurry				x
	23	mf4	Clarification				x
	24	chr1	SP-Sepharose FF		x	x	x
	25	mf5	Conc./diafiltration			x	
	26	chr2	HIC				x
Purification	27	tnk5	Solubilization	x			
	28	cnt9	Concentration	x			
	29	rct5	Reduction	x			
	30	chr3	Gel filtration	x	x		
	31	rct6	Oxidation	x			
	32	uf1	Concentration	x			
	33	chr4	RP-HPLC	x			
	34	tnk6	Precipitation	x			
	35	cnt10	Concentration	x			
	36	chr5	Q-Sepharose FF		x		x
	37	chr6	Q-Sepharose HP			x	
	38	tnk7	Dilution	x	x		
	39	mf6	Concentration		x		
	40	chr7	Gel filtration	x	x		
	41	chr8	HIC		x	x	
	42	chr9	SP-Sepharose FF				x
	43	chr10	SP-Sepharose HP			x	
	44	mf7	Conc./diafiltration	x	x	x	x

step to be later fed to a gel filtration chromatographic column. Finally a concentration and diafiltration step is performed.

Product 2

Product 2 (P2) is an intracellular protein synthesized by *E. coli* as inclusion bodies in a fermentation process that lasts 11.5 h. Cell harvest is performed using a centrifugation step followed by the cell lysis and the harvesting of the inclusion bodies. After three consecutive washing and centrifugation steps, the protein is subjected to a an oxidative environment followed by the refolding of the protein. Before a cation exchange chromatography a micro-filtration step is carried out in order to concentrate the mixture and then, in the second facility, two more chromatographic stages are carried out: gel filtration and anion exchange. A dilution and concentration steps followed by a new gel filtration and a hydrophobic interaction chromatography precede a concentration and diafiltration step as the last stage of the process.

Product 3

Product 3 (P3) is also an intracellular protein synthesized by *E. coli* as inclusion bodies but in a fermentation process that lasts 25 h. Cell harvest is performed using a centrifugation step followed by a cell washing, a cell concentration, and followed by cell disruption. A new centrifugation step is carried out in order to capture the inclusion bodies which are then suspended an subjected to a reducing environment. The next stage, in which the protein is refolded, is followed by a concentration, a cation exchange chromatography and a concentration and diafiltration step. In the second facility, three chromatographic stages—anion exchange, hydrophobic interaction, and cation exchange—are performed before the last concentration and diafiltration step.

Product 4

Product 4 (P4) is a recombinant protein synthesized by *S. cerevisiae* as an extracellular product in a fermentation process that lasts 122 h. After fermentation a centrifugation step to discard cells is performed followed by a microfiltration step to clarify the stream before two chromatographic stages: a cationic exchange chroma-tography and an hydrophobic interaction chromatography.

In the second facility, two ion exchange chromatographies follow the process—anion and cation exchange, respectively—ending with a concentration and diafiltration stage.

Estimation of Processes Data and Plant Cost

Equipment Sizes

Based on mass balances and the processes data given by Imperatore and Asenjo (2001), different equipment sizes for each productive process were estimated. A 20% safety factor was considered for tanks in general and a 15% for semi-continuous items different from chromatographic columns that were sized using a relatively low capacity usage.

Estimated equipment sizes are given in Supplementary Material (Tables S1–S4).

together with a diafiltration process follows before cellular deactivation in a stirred tank. After that, a new homogenization step is performed followed by a sucrose gradient centrifugation for the collection of a small fraction containing the protein of interest.

At another facility, the purification process is continued with a solubilization process and a centrifugation to concentrate the product. Then the mixture is subjected to a reducing environment followed by a gel filtration chromatography and a reaction of oxidation to refold the protein. Before the next chromatographic step—a reversed phase chromatography—an ultrafiltration is performed to concentrate the product mixture. The protein of interest is then precipitated and centrifuged before a dissolution

Table II. General production data of the original processes and production level over a time horizon of 15,004 h (Imperatore and Asenjo, 2001).

| Product | Cycle time (h) | Final batch | | Production level (kg) |
		Size (kg)	Volume (L)	
P1	120	6.005	285	288,216
P2	24	3.777	135	120,866
P3	32	7.385	285	886,226
P4	122	14.043	285	533,624

Equipment Costs

Units suchs as tanks, reactors, and fermenters may be built according to customers needs, therefore, the information needed for their sizing corresponds to cost coeficients for functions of the type $c_j V_j^{\gamma_j}$ and lower and upper size bounds. Original data for fermenter and reactor were taken from Iribarren et al. (2004) and for tanks, from Harrison et al. (2015). The corresponding data are presented in Table III.

For the case of the semi-continuous items, costs are determined according to the selected unit offered by the manufacturer. For these cases, cost and equipment sizes are given in discrete sets. Data used in this paper are presented in Table IV.

Multiproduct Batch Plant

Given the description of the four downstream processes it is possible to suggest that some equipment can be shared by two or more processes as is the case of the centrifugation step used for cell harvesting in the isolation processes of Products 2 and 3.

Isolation and purification processes for each individual protein and the identification of the stages than can be shared by 2 or more processes in the proposed multiproduct batch plant is shown in Table I. For practical purposes, stages are organized as "Fermentation" and "Purification" steps.

Problem Formulation

Mathematical Modeling

The routing formulation proposed by Sandoval et al. (2016) permits the sizing of different equipment in a biotechnological multi-product batch plant selecting downstream processes or routes that minimize the equipment costs with single product campaings. In this article,

Table III. Cost coefficients and variable bounds needed to size batch units. Costs can be calculated in U.S.$ with the function $c_j V_j^{\gamma_j}$. Data were updated to year 2015 using CEPCI: Year 2000, 394.1; year 2012, 584.6; year 2015, 573.1.

| Item | Cost coefficients | | Size bounds (L) | |
	c_j	γ_j	Lower	Upper
Fermenter	1462	0.6	20	100,000
Reactor	1425	0.5	20	100,000
Tank	35,238	0.1168	200	5000
	482	0.6217	5000	50,000

Table IV. Available cost and equipment sizes for semi-continuous units. Costs are in 1000 U.S.$. Data were updated to year 2015 using CEPCI: Year 2000, 394.1; year 2012, 584.6; year 2015, 573.1.

Item	Size	Cost
Micro/ultrafilter (m^2)	5	147
Harrison et al. (2015)	15	172
	30	206
	55	225
Centrifuge (1000 m^2)	10	59
Harrison et al. (2015)	50	83
	100	186
	210	495
Homogenizer (L/h)	55	15
Harrison et al. (2003)	105	37
	290	52
	700	81
	2100	108
	4500	152
Chromatographic column (L)	1	8
Harrison et al. (2015)	2	10
	4	12
	8	17
	15	20
	35	88
	65	186
	150	235
	250	294
	400	392
	580	588
	1000	784
	1600	1078

that formulation was modified to account for the selection of semi-continuous items in a discrete set of sizes and costs.

The model uses an extension of the non-linear formulation proposed by Kocis and Grossmann (1989) for the duplication of units in parallel working in-phase (x_j^1), and out-of-phase (x_j^2). The selection of one route per protein and therefore, the use of just some of the possible stages is carried out by the variables z_{ih}^1 and z_j^2 in *clique* constraints (*clique* constraints are defined in Sandoval et al. (2016)). The former is 1 when for product i synthesis host h is selected and 0 otherwise and the last is 1 when stage j is part of at least one of the routes selected and 0 otherwise. Finally, convex but non-linear inequalities were replaced with sets of linear functions that allows to compute true lower and upper bounds for the objective function. The gap between those bounds can be as small as wanted at the cost of computation time using more cutting points.

Main variations in this work have to do with the use of the formulation proposed by Kocis and Grossmann (1989) applied to the selection of cost and sizes of semi-continuous units and chromatographic columns; and an extension of the use of sets of linear functions applied to the design of stirred tanks. These changes are explained in the following paragraphs.

Objective Function

The objective is to minimize the investment cost of main equipment of the plant given fixed production targets, d_i, over a time horizon δ. As the original cost functions in each stage j are non-linear:

$$c_j \exp\left(x_j^3 + \gamma_j y_j\right) \qquad \forall j \qquad (1)$$

with c_j and γ_j being cost parameters, y_j the logarithm of the equipment size and x_j^3 the duplication of units ($= x_j^1 + x_j^2$), continuous variables v_j are defined to transform former functions into sets of linear functions. The resulting objective function is given by Equation (2):

$$\min \text{ cost} = \sum_{j \in \varepsilon^1} v_j^1 + \sum_{j \in \varepsilon^2 \cup \varepsilon^3} \left[v_j^1 + v_j^2 + v_j^3\right] \qquad (2)$$

where three main types of stages can be identified: batch stages (ε^1), semi-continuous stages (ε^2), and chormatographic stages (ε^3). The former are composed of a batch unit such as a stirred tank or a reactor or fermenter with size $Y_j^1 = \exp(y_j^1)$; and the others have a feed tank (y_j^1), a product tank (y_j^2), and a semi-continuous item or chormatographic column of size $Y_j^3 = \exp(y_j^3)$. Both, feed and product tanks, were considered to be stirred tanks.

For the reactors variable v_j^1 is constrained as in Sandoval et al. (2016). On the other hand since the cost of stirred tanks—in a subset $\bar{\varepsilon}$ —can be modeled as a two piecewise function (see Table III) the former cost (1) can be actually written as Equation (3):

$$\text{cost} = c_j^{1,\text{small}} \cdot \exp\left(x_j^3 + \gamma_j^{1,\text{small}} y_j^1\right) +$$

$$c_j^{1,\text{big}} \cdot \exp\left(x_j^3 + \gamma_j^{1,\text{big}} y_j^1\right) \qquad \forall j \in \bar{\varepsilon} \qquad (3)$$

where $x_j^3 = x_j^1 + x_j^2$ if $j \in \varepsilon^1$ or $x_j^3 = x_j^2$ if $j \in \varepsilon^2 \cup \varepsilon^3$.

Therefore variables v_j^1 and v_j^2 are constrained with Equations as (4a) and (4b).

$$v_j^1 \geq \alpha_k^{1,\text{small},j}\left(x_j^3 + \gamma^{1,\text{small}}_j y_j^1\right) +$$

$$\left(\beta_k^{1,\text{small},j} - \alpha_k^{1,\text{small},j} b_k^{1,\text{small},j}\right) z_j^2 \qquad \forall j \in \bar{\varepsilon}, k$$
$$\in \mathcal{K}^1 \qquad (4a)$$

$$v_j^1 \geq \alpha_k^{1,\text{big},j}\left(x_j^3 + \gamma^{1,\text{big}}_j y_j^1\right) +$$

$$\left(\beta_k^{1,\text{big},j} - \alpha_k^{1,\text{big},j} b_k^{1,\text{big},j}\right) z_j^2 \qquad \forall j \in \bar{\varepsilon}, k \in \mathcal{K}^1 \quad (4b)$$

with $\alpha_k^{1,\text{big},j}$, $\alpha_k^{1,\text{small},j}$, $b_k^{1,\text{big},j}$, $b_k^{1,\text{small},j}$, $\beta_k^{1,\text{big},j}$, and $\beta_k^{1,\text{small},j}$ being the parameters used to build the set of linear constraints and $\mathcal{K}^1$ the set of cutting points used.

Finally, costs of semi-continuous and chromatographic units are constrained using equations as (5) with c_j^{3*} being the cost of the equipment calculated as its size y_j^3 in constraints (6) and (7).

$$v_j^3 \geq \alpha_k^{3j}\left(x_j^1 + x_j^2 + c_j^{3*}\right) + \left(\beta_k^{3j} - \alpha_k^{3j} b_k^{3j}\right) z_j^2 \qquad \forall j$$
$$\in \varepsilon^2 \cup \varepsilon^3, k \in \mathcal{K}^3 \qquad (5)$$

Constraints

Sizing and timing constraints are basically the same as those presented by Sandoval et al. (2016). Differences are given by the introduction of binary variables for the selection of the available equipment sizes and costs.

As was previously stated semi-continuous items including chromatographic columns are only available in a discrete number of sizes. In order to achieve the selection binary variables as z_{jk}^5 were introduced to the model which are restricted by constraints (6) and (7).

$$y_j^3 = \sum_{k \in \mathcal{K}^8} z_{jk}^5 \ln(u_k) \qquad \forall j \in \varepsilon^* \qquad (6)$$

$$\sum_{k \in \mathcal{K}^8} z_{jk}^5 = z_j^2 \qquad \forall j \in \varepsilon^* \qquad (7)$$

where u_k is k-th element of the set of available equipment sizes $\mathcal{U}_{\varepsilon^*}$, defined for each subset of units ε^*: centrifuges, micro/ultrafilters, homogenizers, and chromatographic columns.

Size and Time Factors

Different equipments are sized to process the incoming batch over a time that is at most equal to a cycle time that depends on each product. These constraints make use of some parameters defined as size and time factors. As in Sandoval et al. (2016), constant values are considered and calculated based on mass balances and data of real processes collected by Imperatore and Asenjo (2001).

Size factors for product i in stage j (in g/L) for items in batch stages, S_{ij}; and for feed and product tanks, S_{ij}^1 and S_{ij}^2, respectively, are calculated using Equation (8):

$$S_{ij} = \frac{V_j}{B_i} \qquad (8)$$

with $V_j[L]$ being the actual tank volume in Tables SI–SIV; and $B_i[g]$, equal to Y_i^4 in the nomenclature used by Sandoval et al. (2016), the batch size of the product i at the end of the process shown in Table II. This last parameter was estimated using the broth volume in the fermenter (80% of the tank volume), the concentration of product in the fermentation broth —2.5 g/L for Product 1; 2 g/L for Products 2 and 3; 3 g/L for Product 4-, and the overall mass yield given by the product of the yields in each stage, η, shown in Tables SV–SVIII.

Operation time for each non-batch stage is modelled with Equation (9):

$$T_{ij} = T_{ij}^0 + T_{ij}^1 \frac{B_i}{R_j} \qquad (9)$$

with $T_{ij} [h]$ being the time needed for the step j to process i; $T_{ij}^0[h]$ the equipment startup time; T_{ij}^1 a constant time factor or duty factor; and R_j, equal to Y_j^3 in previous nomenclature, the semi-continuous equipment size in units according to the equipment type (see Table IV).

Data used to estimate time or duty factors are given in Supplementary Material (Tables SV–SVIII).

Membrane Filtration (Tangential-Flow Filtration)

This step is used for the removal of suspended particles, recovery of cells from fermentation broth, and clarification of homogenates containing cell debris and both of them may be followed by the diafiltration of the retentate (stream with larger particles) (Green and Perry, 2007). This last step is essentially a washing step that can be used either to remove more impurities or to increase yield by recovering more product as permeate in the clarification process. Diafiltration is performed maintaining constant the level of the feed or retentate tank by the addition of a suitable solvent while the permeate is removed through the membrane (Hearn, 2000).

For a batch operation the design equation of the filtration unit is given by Equation (10) (Green and Perry, 2007):

$$A = \frac{V_0}{t} \left(\frac{1 - \frac{1}{X}}{J_{conc}} + \frac{\frac{N}{X}}{J_{diaf}} \right) \qquad (10)$$

where A is the membrane area, V_0 is the initial retentate volume, $X < 40$ is the volume reduction factor given by the ratio between the initial retentate volume and the final retentate volume, N is the ratio between the buffer volumes added and the fixed retentate volume. J_{conc} is the flux in the concentration process (volumetric permeate flow rate/membrane area) and J_{diaf} is the flux in the diafiltration process. Values for J_{conc} went from 0.2 in first stages to 0.03 in the final step (Iribarren et al., 2004). J_{diaf} should be smaller than J_{conc} (Hearn, 2000); their value was considered as $\frac{3}{4} J_{conc}$ when necessary.

Operation time (t in Equation (10)) can be written in the form of Equation (9):

$$T_{i,mf}[h] = T_{i,mf}^0[h] +$$

$$\left\{ \frac{V_0[m^3]}{B_i[kg]} \left(\frac{1 - \frac{1}{X}}{J_{conc} \left[\frac{m^3}{m^2 \cdot h} \right]} + \frac{\frac{N}{X}}{J_{diaf} \left[\frac{m^3}{m^2 \cdot h} \right]} \right) \cdot \frac{1[kg]}{1000[g]} \cdot \frac{1}{0.85} \right\}$$
$$\cdot \frac{B_i[g]}{A_{mf}[m^2]} \qquad (11)$$

where the term 1/0.85 includes the size safety factor. Then, time factor $T_{i,mf}^1$, in $\left[\frac{m^2 \cdot h}{g} \right]$, can be computed using the expression between braces in Equation (11).

Finally, this stage yield is given by Equation (12a) when this step is used for concentration and by Equation (12b) when this step is used for clarification. S_i is the observed solute passage that can be computed as the ratio between the concentration of the protein in the permeate and its concentration in the feed stream. If the solute is fully retained $S_i = 0$ and 1 for a fully passing solute (Green and Perry, 2007).

$$\eta_{i,mf} = e^{-S_i(N + \ln X)} \qquad (12a)$$

$$\eta_{i,mf} = \left(1 - X^{-S_i} \right) \left(1 - e^{-S_i N} \right) \qquad (12b)$$

Centrifugation

Centrifugation utilizes the density difference between the solids and the surrounding fluid (Bell, 1989).

Centrifuges costs are estimated using a Σ factor that is equivalent to a transversal area. Based on the settling velocity of the solid, v_s, and the volumen to be treated, V_0, this factor can be computed using Equation (13):

$$\Sigma[m^2] = \frac{(V_0 [m^3])}{(t [h]) \left(v_s \left[\frac{m}{h} \right] \right)} \qquad (13)$$

Then, time factor $T_{i,cnt}^1$ in $\left[\frac{m^2 \cdot h}{g} \right]$ is computed using Equation (14).

$$T_{i,cnt} [h] = T_{i,cnt}^0 [h] +$$

$$\left\{ \frac{V_0[m^3]}{10^3 \cdot v_s \left[\frac{mm}{h} \right] \cdot B_i[g]} \frac{1}{0.85} \right\} \frac{B_i [g]}{\Sigma [10^3 \, m^2]} \qquad (14)$$

Based on Hatti-Kaul and Mattiasson (2003) the settling velocity was estimated to be 0.2 mm/h for *E. coli* and 7 mm/h for *S. cerevisiae*.

Homogenization

For high pressure homogenizers, the fraction of protein released depends on the operational pressure and the number of passes through the homogenizer (Doran, 2012).

According to Pinto et al. (2001) homogenization time is proportional to the volume fed to the homogenizer, $V_{i,hom} (L) = V_{0,hom} \cdot NP_{i,hom}$, and inversely proportional to the homogenizer capacity, $Cap_{i,hom} (L/min)$, where $V_{0,hom}$ corresponds to the volume received from previous stage plus a 10% extra volume of lysis buffer if needed (Harrison et al., 2003). With this, Equation (9) takes the form of Equation (15) and time factor $T_{i,hom}^1$ can be calculated in $\left[\frac{L}{g} \right]$.

$$T_{i,hom} [h] = T_{i,hom}^0 [h] +$$

$$\left\{ \frac{\left(V_{0,hom}\,[L]\right)\left(NP_{i,hom}\right)}{B_i\,[g]}\frac{1}{0.85} \right\} \frac{B_i\,[g]}{Cap_{i,hom}\left[\frac{L}{h}\right]} \quad (15)$$

Based on data given in Bell (1989) and Clonis (1990), the number of passes NP was considered to be seven in the case of *E. coli* and eight for *S. cerevisiae*. As no especific kinetic data were known, yields were considered to be between 0.8 and 0.9.

Chromatographic Separations

As explained by Doran (2012), the basis of chromatography is the selective retardation of solute molecules during passage through a bed of resin particles.

In the processes studied, two major types of liquid chromatographic separations can be found: gel filtration and adsorption chromatography.

Adsorption chromatographic columns are sized taking into account the amount of protein that can be adsorbed into the column, $B_{i,chr}$, which is in turn related to the final batch size (Pinto et al., 2001):

$$B_{i,chr}[kg] = \left(V_{chr}[m^3]\right)(\pi_i)\left(\beta_{i,chr}\left[\frac{kg}{m^3}\right]\right) = \frac{B_i\,[kg]}{\displaystyle\prod_{n=chr}^{NE}\eta_{in}} \quad (16)$$

with V_{chr} being the column volume; $\beta_{i,chr}$ the column capacity; π_i the fraction of the maximum capacity that is being used by the adsorbed protein; and NE the number of total stages. From this, the column size factor can be computed with Equation (17):

$$S_{i,chr}^3\left[\frac{L}{g}\right] = \frac{1}{\pi_i\beta_{i,chr}\displaystyle\prod_{n=chr}^{NE}\eta_{in}} = \frac{V_{chr}[L]}{B_i[g]} \quad (17)$$

The operation time of the chromatographic step is given by (18):

$$T_{i,chr} = \frac{V_{\text{feed}} + V_{\text{wash}} + V_{\text{elution}} + V_{\text{regeneration}}}{A_{chr}v_{chr}} \quad (18)$$

with V_{feed} being the volume of the mixture to be purified, V_{wash} the buffer wash volume used to eliminate proteins not bound to the resin, V_{elution} the volume of buffer used to recover the protein and $V_{\text{regeneration}}$ the volume of buffer used to regenerate the column. A_{chr} and v_{chr} are the column transversal area and the linear velocity of the mobile phase, respectively.

Considering the time factors, the operation time can be computed using Equation (19):

$$T_{i,chr}\,[h] = T_{i,chr}^0\,[h] + \frac{\left([V_{\text{feed}} + V_{\text{elution}}]\,[m^3]\right)h_{chr}\,[m]}{\left(B_i\,[kg]\right)\left(v_{chr}\left[\frac{m}{h}\right]\right)}$$

$$\frac{B_i\,[kg]}{\left(A_{chr}[m^2]\right)\left(h_{chr}[m]\right)} \quad (19)$$

If the column height (h_{chr}) is set to 0.25 m based on data of Imperatore and Asenjo (2001), and three column volume are used to wash and regenerate the resin, respectively, then the constant time factor can be computed as:

$$T_{i,chr}^0\,[h] = \frac{6 \cdot A_{chr}[m^2] \cdot h_{chr}[m]}{A_{chr}[m^2] \cdot v_{chr}\left[\frac{m}{h}\right]} = \frac{1.5}{v_{chr}} \quad (20)$$

Finally, factor T^1 is calculated using Equation (21).

$$T_{i,chr}^1\left[\frac{L \cdot h}{g}\right] = \frac{\left([V_{\text{feed}} + V_{\text{elution}}]\,[m^3]\right)\left(h_{chr}[m]\right)}{B_i\,[kg]\left(v_{chr}\left[\frac{m}{h}\right]\right)} =$$

$$\frac{\left(T_{i,chr} - T_{i,chr}^0\right)\,[h] \cdot V_{chr}[m^3]}{B_i\,[kg]} \quad (21)$$

In gel filtration, a constant time was considered independent of the feed stream. For adsorption chromatography a velocity of 5.5 m/h was considered for ionic-exchange resins and for hydrophobic resins, 4 m/h. These velocities were defined based on GE Healthcare Life Sciences handbooks that can be downloaded from their web page (http://www.gelifesciences.com/).

Computational Tools/Execution Environment

The MILP model studied was coded using the AMPL modelling language and different instances were solved using the commercial CPLEX solver in its version 12.4.0.0. The execution environment was given by a single thread on a Intel(R) Xeon(R) CPU E5620@ 2.40 GHz with an optimality relative gap of 0.1% and 256 cutting points for an a posteriori gap up to 0.12%.

Results and Discussion

Cost of the Real Plant

The real plant considers both fermentation and purification facilities.

The fermentation facility produces the four products in a single production line; therefore, cost was estimated based on a plant configuration as that shown in Table I. The equipment sizes for each stage were set as the maximum size needed among the four processes; and for semi-continuos items, that size was selected among those available.

Table V. Average execution time and relative a posteriori gaps for an instance of 18 stages solved 100 times.

Cutting points	Execution time (s)		Gap (%)
	Upper approx.	Lower approx.	
2^5	3.78	1.83	0.54
2^6	5.05	6.46	0.30
2^7	13.10	13.79	0.13
2^8	38.65	32.09	0.11

Table VI. Optimized structure of the multiproduct batch plant over a time horizon of 15,004 h. The cost function is equal to U.S.$ 9,716,680 using cost parameters updated to year 2015.

Stage	X^1	X^2	V^1	V^2	V^3/R
1	1	3	4570		
2	1	1	4570		5
3	1	1	3236	324	10
4	1	1	3236		
5	1	1	3236	647	10
6	1	1	718		700
7	1	1	718		5
8	1	1	539		
9	1	1	592		290
10	1	1	6517	261	50
11	1	1	295	200	10
12	1	1	295		
13	1	1	295	200	10
14	1	1	221		
15	1	1	712	200	10
16	1	1	200		
17	1	1	200	200	50
18	1	1	2427		
19	1	3	3641		
20	1	1	3729		
21	1	1	3729		15
22	1	1	4063	451	10
23	1	1	451	378	5
24	1	1	378	341	150
25	1	1	285		15
26	6	1	341	307	15
27	1	1	1097		
28	2	1	1097	200	50
29	1	1	493		
30	1	1	493	296	1 000
31	1	1	592		
32	1	1	592		5
33	2	1	516	309	15
34	1	1	1237		
35	2	1	1237	200	50
36	2	1	307	276	15
37	7	1	200	200	15
38	1	1	275		
39	1	1	200		5
40	1	1	275	200	400
41	8	1	200	200	15
42	5	1	276	248	15
43	4	1	200	200	15
44	1	1	248		5

The purification facility produces the four products in four production lines, each under single product campaigns. Product 1 is purified independently in one production line. Product 2 shares a production line with Products 3 and 4 in two different production lines. In addition another production line only purifies Product 3. Cost was estimated as the addition of the four production lines, each one sharing the same stages as in Table I depending on the products that are being processed. Equipment sizes and their costs were calculated as before.

Cost of the main equipment of the real plant was estimated as U.S.$ 28,230,981, from where U.S.$ 21,056,024 corresponds to the purification facility.

Table VII. Final batch size and cycle time of the four products produced in the multiproduct batch plant optimized over a time horizon of 15,004 h.

Product	Final batch size (kg)	Cycle time (h)
P1	2.195	24
P2	0.668	24
P3	1.902	8.67
P4	6.253	40.67

Accuracy of the Solution Given by the Optimization Model

As established in previous work, (Sandoval et al., 2016) the approach of defining lower and upper approximations for non-linear functions in constraints and in the objective function gives true upper and lower approximations for the actual optimal plant cost. Therefore, a small a posteriori gap between both approximations is expected.

Taking an instance of 18 stages, different number of cutting points were considered to solve the model: 2^5, 2^6, 2^7, and 2^8. Execution times and a posteriori gaps were obtained for a set of 100 runs. Results are shown in Table V. A 2^8 cutting points were selected to solve the instances studied since this gives an a posteriori gap very close to the solver optimality relative gap and still takes little time to solve the instance. At the cost of twice the time, an a posteriori gap of 0% can be obtained using 2^9 cutting points and a gap equal to 0 for the solver. Nevertheless the structure of the plant is basically the same and minor changes in equipment sizes can be observed.

Parameters of the Optimization Model

The model makes use of parameters related to costs, demand, and operating conditions. Costs and cost parameters were considered to be known (Tables III and IV) and for comparison purposes the demand was set as the actual production of the original plant (Table II). On the other hand, parameters related to operating conditions such as time and size factors had to be determined.

For chromatographic stages and homogenization steps expressions given by Pinto et al. (2001) were adapted. For the case of centrifuges and membrane filtration, expressions were obtained from their respective design equations. Most of these equations need an inlet volume that was calculated considering dilution and/or concentration factors (Tables SV–SVIII). In this way, size and time factors limit the equipment size to a maximum reasonable and acceptable concentration of the protein of interest in its small possible size.

Optimization of the Multiproduct Batch Plant of 44 Stages

Based on the cycle times and the batch sizes given in Table II a time horizon of 15,004 h was estimated as the time needed for the real processes to fullfill the required production levels. In the original plant, duplication of units was not considered. This led to a reuse of some equipments and therefore an increase in the cycle time in Products 1 and 3 (see Tables SI and SIII). For example, Product 1

has a cycle time of 120 h even though the longest process lasts only 24 h per cycle (gel filtration steps).

Using constant size and time factors and less than 5 min of computing time, the structure of the plant proposed in Table I was optimized to a minimum cost of U.S.\$ 9,716,680 and the equipment sizes shown in Table VI with V^1 being the size of the feed tank, V^2 the size of the product tank, and V^3/R the size of the semi-continuous item or chromatographic column in each stage. Estimated final batch size and cycle times for each product are presented in Table VII.

Notice that both duplication, in-phase (variable X^1) and out-of-phase (variable X^2), were used in the optimized plant. Duplication out-of-phase decreased in particular the cycle time of Product 4, given by the fermentation step, from 122 h (see Table II) to 40.67 h (Table VII). Duplication in-phase along with the time horizon permited small equipment sizes in general. For instance, if we consider the particular case of stage 41: a Hydrophobic Interaction Chromatography used by (P2) and (P3). In the real plant a cromatographic column of 150 L is used in the stage 20 for Product 2 (Table SII) and in the stage 14 for Product 3 (Table SIII) processing a feed stream of at most 540 L for Product 2 (considering a usage of 80% of the equipment size in Table SII). Since in the optimized facility batch sizes are at least half of that in the real plant (compare Tables II and VII) the feed stream to this stage is smaller and needs less than one third of the real capacity (200 L). In addition to that a duplication in-phase permitted the selection of eight chromatographic columns of only 15 L working simultaneously.

Taking into account that the time needed by the real processes to fulfill the production levels permits the use of smaller equipment it is clear that some of the real plant equipment is oversized. In fact, if instead of using a time horizon of 15,004 h, we use only a one year time horizon (5,904 working hours) for the optimized plant its cost would still be smaller than that of the real plant (U.S.\$ 14,533,800) and the profit could be two and a half times the one obtained in the real case if the annual demand is that shown in Table II.

The comparison done in this work of main equipment costs clearly shows the benefits of using this type of models for plant design. The optimized plant has smaller equipments and underutilization is avoided. This work did not consider operational costs or profits in the optimization presented.

Conclusions

In this article, a biotechnological multiproduct batch plant that manufactures four different recombinant proteins for human application was described in some detail. The batch plant design was then optimized using a modification of the MILP model presented in a former article in order to find a hypothetical new biotechnological batch plant based on the information of real known processes for the production of the four recombinant protein products. The new model includes discrete costs and equipment sizes for semi-continuous items and preserves the selection of costs and sizes in a continuous range for stirred tanks, reactors and fermenters. A major contribution of the present investigation is in determining the explicit equations used to calculate constant size

and time factors for some equipment used in the biopharmaceutical industry.

The application of the model permitted cost savings up to 66% of the cost of the main equipment, showing that this tool is not just useful but necessary in order to design a plant of the optimal and necessary size.

Lower level implementations (in C, C^{++}) could include the effect of variable cost and production target parameters but this is beyond the scope of this article and was left for a future investigation. In addition, as the time required to solve each instance is less than 3 min, there is plenty of space for continuing the addition of new and more complex constraints.

Nomenclature

Indices

h	host
i	product
j	stage
n	type of unit: 1 for feed tank, 2 for product tank, and 3 for chromatographic colum or semi-continuous item

Sets

$\bar{\varepsilon}$	set of stirred tanks
ε^*	set of semi-continuous items, for example, centrifuges
ε^1	set of batch stages
ε^2	set of semi-continuous stages
ε^3	set of chromatographic stages
$\mathcal{K}^8$	set of number of sizes of semi-continuous items
$\mathcal{K}^n$	set of cutting points for the linear transformation. $n \in \{1, 2, 3\}$
$\mathcal{U}_{\varepsilon^*}$	set of available equipment sizes in subset ε^*

Parameters in the MILP Formulation

$\alpha_k^{n,j}, \beta_k^{n,j}, b_k^{n,j}$	parameters related to linear transformation for item n in stage j defined in k-th cutting point
δ	time horizon
c_j^n, γ_j^n	cost coefficients related to Y_j^n with $n \in \{1, 2, 3\}$
c_j^{3*}	cost coefficient of selected semi-continuous item in stage j
d_i	production target for product i
S_{ij}	size factor for product i in stage j
T_{ij}	Time factor for product i in stage j
T_{ij}^0	time for equipment startup for product i in stage j. $t_{ij}^0 = \ln\left(T_{ij}^0\right)$

T_{ij}^1 constant time or duty factor for product i in stage j. $t_{ij}^1 = \ln\left(T_{ij}^1\right)$

Parameters in Design Equations

η_{ij} yield for product i in stage j

$\beta_{i,chr}$ parameter for chromatography: column capacity for product i

π_i parameter for chromatography: fraction of the capacity used by product i

Σ parameter for centrifugation: design parameter equivalent to a transversal area

A parameter for membrane filtration: membrane area

A_{chr} parameter for chromatography: column transversal area

$\text{Cap}_{i,hom}$ parameter for homogenization: homogenizer capacity for product i

h_{chr} parameter for chromatography: column height

J_{conc} parameter for membrane filtration: flux in the concentration process

J_{diaf} parameter for membrane filtration: flux in the diafiltration process

N parameter for membrane filtration: ratio between the buffer volumes added and the fixed retentate volume

$NP_{i,hom}$ parameter for homogenization: number of passes for product i

S_i parameter for membrane filtration: observed solute passage

V_0 parameter for semi-continuous items: feed volume

V_{chr} parameter for chromatography: column volume

v_{chr} parameter for chromatography: linear velocity of the mobile phase

$V_{elution}$ parameter for chromatography: volume of buffer used to recover the protein

V_{feed} parameter for chromatography: volume of mixture feed to the column

$V_{i,hom}$ parameter for homogenization: volume fed to the homogenizer for product i

$V_{regeneration}$ parameter for chromatography: volume of buffer used to regenerate the resin

V_{wash} parameter for chromatography: volume of buffer used to wash the sample

v_s parameter for centrifugation: settling velocity

X parameter for membrane filtration: volume reduction factor

Variables

B_i continuous variable: final batch size of product i. $B_{i,j}$ is the batch size of product i after stage j

R_j continuous variable: size of semi-continuous item in stage j. Equivalent to Y_j^3 in nomenclature of Sandoval et al. (2016)

v_j^n continuous variable for linear transformation. $n \in \{1, 2, 3\}$

V_j^n continuous variable: volume of tank in stage j with $n \in \{1, 2, 3\}$. Equivalent to Y_j^n in nomenclature of Sandoval et al. (2016)

X_j^n continuous variable: number of units operating in parallel in-phase (X_j^1) and out-of-phase (X_j^2) in stage j. $x_j^n = \ln\left(X_j^n\right)$

x_j^3 continuous variable: $= x_j^1 + x_j^2$

Y_j^1 continuous variable: volumetric capacity for batch items or stirred feed tanks. $y_j^1 = \ln\left(Y_j^1\right)$

Y_j^2 continuous variable: volumetric capacity of stirred product tanks. $y_j^2 = \ln\left(Y_j^2\right)$

Y_j^3 continuous variable: capacity of semi-continuous items. $y_j^3 = \ln\left(Y_j^3\right)$

z_{ih}^1 binary variable for the selection of host. 1 if host h is selected to synthesize product i

z_j^2 binary variable for the selection of stage. 1 if at least one product uses stage j

z_{jk}^5 binary variable for the selection of equipment. 1 if elemnt k-th is selected in stage j

References

Barbosa-Póvoa AP. 2007. A critical review on the design and retrofit of batch plants. Comput Mater Eng 31(7):833–855.

Bell G. 1989. Bioseparations: Downstream processing for biotechnology, volume 11. New York, NY: John Wiley and Sons Inc.

Borisenko A, Kegel P, Gorlatch S. 2011. Optimal design of multi-product batch plants using a parallel branch-and-bound method. In: Malyshkin V, editor. Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), volume 6873 LNCS of lecture notes in computer science. Berlin, Heidelberg: Springer. p 417–430.

Clonis YD. 1990. Separation processes in biotechnology. Process affinity chromatography, volume 9. Boca Raton, Florida: CRC Press.

Doran PM. 2012. Bioprocess engineering principles, 2nd edn. London: Academic Press.

Green D, Perry R. 2007. Perry's chemical engineers' handbook, 8th edn. New York: McGraw Hill professional. McGraw-Hill Education.

Harrison RG, Todd P, Rudge SR, Petrides DP. 2003. Bioseparations science and engineering. Oxford: Oxford University Press.

Harrison RG, Todd PW, Rudge SR, Petrides DP. 2015. Bioseparations science and engineering, 2nd edn. Oxford University Press.

Hatti-Kaul R, Mattiasson B. 2003. Isolation and purification of proteins. Boca Raton, Florida: CRC Press.

Hearn MT. 2000. Handbook of bioseparations, volume 2. London: Academic Press.

Imperatore C, Asenjo JA. 2001. Case study in a multiproduct multifunctional plant. Santiago, Chile: Technical report, Chiron Corporation and University of Chile.

Iribarren OA, Montagna JM, Vecchietti AR, Andrews B, Asenjo JA, Pinto JM. 2004. Optimal process synthesis for the production of multiple recombinant proteins. Comput Aided Chem Eng 18(C):427–432.

Kocis G, Grossmann I. 1989. Computational experience with dicopt solving MINLP problems in process systems engineering. Comput Chem Eng 13(3): 307–315.

Li X, Chen Y, Barton PI. 2012. Nonconvex generalized benders decomposition with piecewise convex relaxations for global optimization of integrated process design and operation problems. Ind Eng Chem Res 51(21): 7287–7299.

Moreno MS, Montagna JM. 2011. Multiproduct batch plants design using linear process performance models. AIChE J 57(1):122–135.

Pinto JM, Montagna JM, Vecchietti AR, Iribarren OA, Asenjo JA. 2001. Process performance models in the optimization of multiproduct protein production plants. Biotechnol Bioeng 74(6):451–465.

Sandoval G, Espinoza D, Figueroa N, Asenjo JA. 2016. MILP reformulations for the design of biotechnological multi-product batch plants using continuous equipment sizes and discrete host selection. Comput Chem Eng 84:1–11.

Viswanathan J, Grossmann I. 1990. A combined penalty function and outer-approximation method for MINLP optimization. Comput Chem Eng 14(7): 769–782.

Voudouris TV, Grossmann IE. 1992. Mixed-integer linear programming reformulations for batch process design with discrete equipment sizes. Ind Eng Chem Res 31:1315–1325.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site.