Optimal Process Synthesis for the Production of Multiple Recombinant Proteins

Oscar A. Iribarren,[†] Jorge M. Montagna,[†] Aldo R. Vecchietti,[†] Barbara Andrews,[‡] Juan A. Asenjo,[‡] and José M. Pinto*[,§]

INGAR - Conicet Institute of Process Design and Development, Avellaneda 3657, 3000 Santa Fe, Argentina, Center for Biochemical Engineering and Biotechnology, Department of Chemical Engineering, Millenium Institute for Advanced Studies in Cell Biology and Biotechnology, University of Chile, Beauchef 861, Santiago, Chile, Department of Chemical Engineering, University of Sao Paulo, Sao Paulo SP 05508-900, Brazil, and Othmer Department of Chemical and Biological Sciences and Engineering, Polytechnic University, Six Metrotech Center, Brooklyn, New York 11201

This paper presents a novel solution strategy for the synthesis of multiproduct and multihost protein production processes. There are several possible hosts that may express each of the products, and different downstream processing separation and purification tasks are needed, which in part depend on the host selection. Moreover, alternative unit operations may be available for some of these separation tasks. Finally, these processing units may be arranged in different configurations. A single mixed-integer optimization model represents the different decisions involved in synthesizing a plant for producing multiple proteins. The mathematical model optimizes the profit of the multiproduct plant and allows the decisions to be made simultaneously, namely, the choice of hosts, downstream operations, the configuration and size of units, as well as their scheduling. An example is solved for a plant that must produce four proteins for which there are alternative hosts for their expression (*Escherichia coli*, Chinese hamster ovary cells, and yeast that, depending on the product, may express it as an extracellular or intracellular protein) that require 15 stages with choices of unit operations as well as in or out of phase operations. Given the very large quantity of novel recombinant proteins for a number of novel therapeutic uses presently being approved or "in the pipeline", multiproduct and multihost recombinant protein production plants have recently been or are being built for the manufacture of these products. The strategy presented in this paper is of crucial value for the optimal utilization of such plants.

Introduction

The main host for recombinant proteins for many years has been *Escherichia coli*; however, developments with yeast and mammalian cells have grown at a very rapid pace, which has resulted in several important industrial processes and commercial products. As many recombinant proteins produced in yeast can be made to be secreted from the cell and the yeast allows for at least partial glycosylation, it has become increasingly attractive as a host.

Optimization-based models for the design of multiproduct batch plants have been previously published (e.g., Grossmann and Sargent, 1979; Knopf et al., 1982; Salomone and Iribarren, 1992; Ravemark and Rippin, 1998). In particular, the design of bioprocessing plants has been addressed by Montagna et al. (2000), who considered constant time and size factors and by Asenjo et al. (2000) and Pinto et al. (2001), in which process performance models that are represented as algebraic equations and result from mass balances and kinetic expressions were developed. Also in the context of biochemical processes, Groep et al. (2000) developed performance models and showed the interactions among the different unit operations for a typical enzyme production process in a plant with fixed topology. Samsatli and Shah (1996a) have addressed a similar problem by developing a design procedure for a biochemical plant that consists of two subproblems. The first subproblem determines the processing conditions of all unit operations using dynamic optimization with manual branch and bound. The only structural decision at this level concerns the number of fermenters in parallel, and scheduling decisions are aggregated. At a second level, a scheduling problem determines the sequence and timing of operations (Samsatli and Shah, 1996b).

Steffens et al. (2000a) developed a hybrid technique for synthesizing downstream purification processes. A heuristic screening procedure is proposed based on physical property information for reducing the size of the process superstructure. Then, an implicit enumeration algorithm is applied. Regarding the purification stage in bioprocesses, Vasquez-Alvarez et al. (2001) developed a mixed-integer programming model for the synthesis of purification steps of a protein mixture. The scope was

* To whom correspondence should be addressed. E-mail: jpinto@poly.edu.
[†] INGAR - Conicet Institute of Process Design and Development.
[‡] University of Chile.
[§] University of Sao Paulo and Polytechnic University.

10.1021/bp034330+ CCC: $27.50
Published on Web 03/26/2004

extended by Steffens et al. (2000b), who incorporated purification tags and developed an expert system for the synthesis of the downstream processes. Bryant and Rowe (1998) present a review on computational techniques based on artificial intelligence methods that extract information from chromatographic analysis data.

Although there is a significant amount of research on the development of expert systems for the synthesis of bioprocesses, much less has been reported in terms of optimization-based approaches. This paper presents a unified representation for the synthesis of bioprocesses that is fully based on mixed-integer optimization and extends the design models previously presented by Montagna et al. (2000) and Pinto et al. (2001). Therefore, the main difference of the previous approaches with respect to the one presented here is that they do not consider synthesis options of alternative hosts, which defines the tasks to be performed at the processing stages, nor within a stage options with respect to the unit operation to be used. In previous work, the stages to be used for each product and the units to be used at each stage are established beforehand. It should be noted that the different alternative units available for any stage may have different performance, i.e., production rates, yields, operating times etc.

In this paper the use of process modeling and optimization tools for solving the synthesis problem of a plant for the production of multiple recombinant proteins that could in principle be expressed in several hosts (e.g., yeast, Chinese hamster ovary (CHO) cells, and *E. coli*) is reported. The decisions to be optimized are the host for synthesizing each product, which unit operations to perform the required downstream separation tasks, the number of units at each stage, and their operation arrangement for reducing idle times and under utilization of installed capacity.

The paper is organized as follows. First we present the general mathematical model for this synthesis problem, followed by the process example and details on the modeling of stages and the tradeoffs involved in this problem. Finally, we report the implementation of the optimization program and analyze the optimal design found using the proposed approach.

Problem Definition and Model Formulation

We must synthesize a multiproduct plant that is able to produce a set of P products. For each product i, $i = 1, ..., P$, there is a known production target Q_i to be produced over time horizon H. The plant consists of a set of stages j, $j = 1, ..., N$, which are shared by the products.

For each product i there is a set of hosts H_i available for expression. In the case that a particular host is selected for producing a product, this sets a sequence of downstream tasks to be performed, which fixes a subset of stages. Let J_h be the set of stages j needed for this product if expressed by host h. For each stage j there may be different units available for performing it, with D_j being the set of units available for performing stage j.

The objective of the model is to minimize the cost of the plant, satisfying the production targets of the P products considered. The objective function considered is as follows:

$$\text{Min Cost} = \sum_j \sum_{d \in D_j} \alpha_{jd} M_{jd} G_{jd} V_{jd}^{\beta_{jd}} \qquad (1)$$

In eq 1, V_{jd} is the size of the units at stage j that correspond to option d. As presented in the previous section, there is a set D_j of optional units, from which one has to be selected. Constraints enforce that only one option of the set D_j is chosen, determining that only the sizes of these units are considered in the cost of the plant, whereas the size of the units belonging to other options are set to zero and have no impact in the objective function.

Also in eq 1, α_{jd} and β_{jd} are cost coefficients distinctive for each kind of equipment, used in the correlations for estimating the cost of the units. Variables M_{jd} and G_{jd} characterize the arrangement of the units at stage j when selecting option d. A number of arrangements are available to reduce the cost of multiproduct plants; the usual one is a parallel unit working either in or out of phase, with all the units of the same size. If M_{jd} parallel units operate out of phase, this reduces the cycle time of the stage, i.e., the time elapsed between successive batches exiting the stage. This also decreases the idle time of downstream stages in the case stage j represents the bottleneck for the production train, thus reducing the size of these stages. If G_{jd} units operating in phase are adopted, these operate simultaneously as if they were the same unit. The batch fed to the stage is split among the available units, while the batches exiting these units are merged before exiting this stage. This arrangement is particularly useful when the batch size surpasses the upper bound capacity of the equipment. The model handles the simultaneous consideration of units working in or out of phase by adopting M_{jd} subsets of units working out of phase, each one consisting of G_{jd} units working in phase.

At each stage and for each product, the size of the units must allow the processing of the incoming batch. This is enforced by the following constraint:

$$V_{jd} \geq \frac{S_{ijdh}\, B_i}{G_{jd}} - M1_{ijdh}\,(1 - z_{ijdh}) \qquad \forall i, j \in J_h, d \in D_j, h \in H_i \quad (2)$$

where z_{ijdh} is a binary variable defined as follows:

$$z_{ijdh} = \begin{cases} 1 & \text{if option } d \text{ is selected at stage } j \text{ for} \\ & \text{producing product } i \text{ using host } h \\ 0 & \text{otherwise} \end{cases}$$

Just one option can be selected, thus:

$$\sum_{d \in D_j} z_{ijdh} \leq 1 \qquad \forall i, h \in H_i, j \in J_h \qquad (3)$$

Constraint 2 is of the Big-M type, where $M1_{ijdh}$ is a large constant that is defined according to the dimensions used. In the case that option d is selected for stage j with host h used for producing i, variable z_{ijdh} is 1 and constraint 2 holds. Otherwise, the binary variable z_{ijdh} is 0 and the constraint is trivially satisfied.

B_i is the batch size of final product i and S_{ijdh} is the size factor for stage j that uses option d obtained from the recipe for product i with host h. This is the minimum capacity required at this stage, for producing one unit mass of product i. Furthermore, in eq 2 the size requirement is affected by the number of duplicate units in phase G_{jd}. Constraint 2 establishes that the size of the unit must be larger than the sizes required by each of the products that are produced at this stage.

If host h is not selected for producing product i, then all the variables z_{ijdh} associated to this combination must be zero, because its corresponding stages are unnecessary. In this case, presented later, constraint 3 becomes

nonactive. In the case that host h is used for producing product i, then one of the variables z_{ihjd} becomes equal to one and eq 3 is satisfied as an equality because just one option d must be selected for its processing at stage j.

Multiproduct plants that work in overlapping mode operate cyclically producing consecutive batches of product i every cycle time, TL_i. This time is computed as the larger operating time among the stages involved in the processing of product i. At the stages that have units operating out of phase, the cycle time is the original operating time divided by the number of units operating in this mode. Considering that the operating time T_{ijdh} for product i at stage j using option d for host h is given by

$$T_{ijdh} = T^0_{ijdh} + T^1_{ijdh} B_i \quad \forall i, h \in H_i, j \in J_h, d \in D_j \qquad (4)$$

the plant cycle time TL_i is computed from the following expression

$$TL_i \geq \frac{T^0_{ijdh} + T^1_{ijdh} B_i}{M_{jd}} - M2_{ijdh}(1 - z_{ijdh}) \quad \forall i, h \in H_i, j \in J_h, d \in D_j \qquad (5)$$

In eq 5, $M2_{ijdh}$ is a constant large enough when compared with the terms involved in this expression. This is also a Big-M constraint that works in the same way as in eq 2: if option d is used at stage j for product i expressed by host h, then z_{ijdh} is 1 and the constraint applies. Otherwise z_{ijdh} is zero and the constraint is trivially satisfied. As previously noted, the selection of an option d for a stage j depends on whether this stage j is included in the processing route for producing i with host h. To represent this problem we use the binary variable y_{ih} defined as follows:

$$y_{ih} = \begin{cases} 1 & \text{if host } h \text{ is chosen for product } i \\ 0 & \text{otherwise} \end{cases}$$

For each product i only one host h can be chosen from those that are able to express this product. So, the following condition is imposed:

$$\sum_{h \in H_i} y_{ih} = 1 \quad \forall i \qquad (6)$$

The binary variables y_{ih} and z_{ijdh} are related to each other. Obviously, if host h is not selected for some product i ($y_{ih} = 0$), then no option d should be selected for the stages j that correspond to this processing route, i.e., $z_{ijdh} = 0$. Otherwise, necessarily one option d must be selected for each of the stages j included in the processing route. The following constraints establish these conditions:

$$1 - y_{ih} + \sum_{d \in D_j} z_{ijdh} \geq 1 \quad \forall i, h \in h_i, j \in J_h \qquad (7)$$

$$1 - z_{ihjd} + y_{ih} \geq 1 \quad \forall i, h \in h_i, j \in J_h, d \in D_j \qquad (8)$$

The following condition establishes that the production targets of all products must be satisfied within the time horizon H:

$$\sum_i \frac{Q_i \, TL_i}{B_i} \leq H \qquad (9)$$

which requires that the summation of the processing times for producing the P products be no greater than the available time horizon. In eq 9, the ratio between each pair Q_i, which denotes the overall amount of product i to be made within time horizon H, and B_i represents the number of batches of product i that are processed, whereas this ratio multiplied by the cycle time provides the time allocated for producing this product.

Therefore, the model proposed in this work minimizes the objective function (eq 1), subject to constraints of eqs 2, 3, and 5−9. This is a Mixed Integer Non Linear Program (MINLP) in Posynomial form, which is nonconvex. We use the following transformations to convexify the problem (Grossmann and Sargent, 1979):

$$v_{jd} = \ln V_{jd} \qquad (10)$$

$$b_i = \ln B_i \qquad (11)$$

$$tl_i = \ln TL_i \qquad (12)$$

$$m_{jd} = \ln M_{jd} \qquad (13)$$

$$g_{jd} = \ln G_{jd} \qquad (14)$$

Taking into account that variables M_{jd} and G_{jd} must attain integer values, we apply the following transformation to handle continuous and binary variables:

$$m_{jd} = \sum_k c_k \, ym_{jdk} \quad \forall j, d \in D_j \qquad (15)$$

$$g_{jd} = \sum_k c_k \, yg_{jdk} \quad \forall j, d \in D_j \qquad (16)$$

$$\sum_k ym_{jdk} = 1 \quad \forall j, d \in D_j \qquad (17)$$

$$\sum_k yg_{jdk} = 1 \quad \forall j, d \in D_j \qquad (18)$$

where ym_{jdk} and yg_{jdk} are binary variables and $c_k = \ln k$. These transformations modify the objective function (eq 1) and constraints 2, 5, and 9, thus resulting in a convex MINLP Program (Ravemark and Rippin, 1998).

General Models of the Semicontinuous Stages

The general batch process literature considers semicontinuous units to work in series with batch units so that their operating times also include the times for filling or empting the batch units. However, in the process considered, pumps are the only semicontinuous units that transfer batches between units. As the pumping costs do not have a relevant impact on the plant design, they were not explicitly modeled. The times for filling and empting batch items were estimated and included as constant values in the batch cycle times.

On the other hand, the process does have special semicontinuous units with an important economic impact on the cost: centrifuge, microfilter, homogenizer, bead mill, diafiltration, ultrafiltration, and sterile filtration. These stages include tanks that hold the material to be processed by the semicontinuos units (Montagna et al., 2000). These types of aggregated units are illustrated in Figure 1 for the case of a microfilter. Their mathematical model has been introduced by Salomone et al. (1994) and has been slightly modified in this paper to consider duplicate units operating in phase. For the batch items eq 2 holds. The time expression for the stage depends on

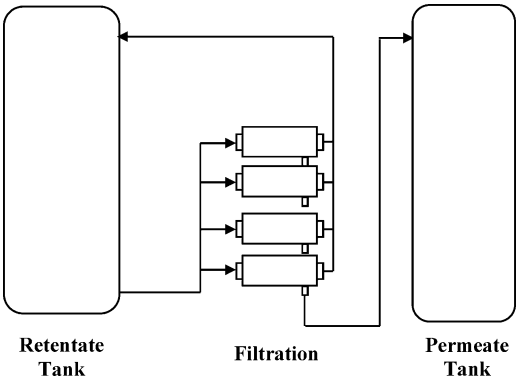

Figure 1. Special semicontinuous unit: microfilter.

both the batch size and the size of the semicontinuous item as follows:

$$T_{ijdh} = T^0_{ijdh} + T^1_{ijdh}\frac{B_i}{G_{jd}R_{jd}} \quad \forall i, h \in H_i, j \in J_h, d \in D_j \tag{19}$$

where R_{jd} is the size of the semicontinuous unit in stage j using option d. Similarly to eq 4, T^0_{ijdh} and T^1_{ijdh} are appropriate time factors that take into account contributions to the total cycle time of the stage that are either fixed amounts of time or proportional to the batch size and inversely proportional to the size of the semicontinuous item.

The sizing equations for semicontinuous items are modeled with a modified version of the expression proposed by Knopf et al. (1982), so to consider duplicated units in phase:

$$G_{jd}R_{jd} \geq \frac{D_{ijdh}\,B_i}{\theta_{ijdh}} \quad \forall i, h \in H_i, j \in J_h, d \in D_j \tag{20}$$

where D_{ijdh} is the duty factor, i.e., the size factor for semicontinuous items, for product i in stage j with option d using host h, and θ_{ijdh} is the operating time that the semicontinuous unit at stage j with option d needs to process a batch of product i produced by host h (Montagna et al., 2000). Comparing eqs 19 and 20 indicates that T^1_{ijdh} is the duty factor of the semicontinuous item.

The batch size B_i trades off the costs of the batch and the continuous items that conform this kind of composite stages: a larger batch size requires a larger batch holding vessel but allows a longer operation time, reducing the deleterious effect of the constant part $T°$ of the cycle time: at the limit, if the batch size were the total annual production, the cycle time would be the time horizon, the constant part $T°$ would be negligible and the semicontinuous unit size would approach its minimum, i.e., the size of a continuous unit.

The objective function eq 1 is adapted to include the cost of both batch and semicontinuous items, with cost expressions for the latter similar to the ones for batch items.

Process Example

The following example corresponds to a plant for producing human insulin, hepatitis B vaccine, tissue plasminogen activator, and super oxide dismutase. The code names used from now on in the paper and the yearly production targets for the four products are presented in Table 1. There are four hosts available for expressing these products: *E. coli*, CHO cells, and yeast (that depending on the product may express it as an extracellular or intracellular protein). The yeast *Saccharomyces cerevisiae* is here treated as two hosts: yeast intracellular and yeast extracellular, because they require different sequences of downstream separation tasks. Table 2 displays which hosts are available for producing each product.

Table 1. Product Data

product	code name	production target kg/year
human insulin	INS	1 500
hepatitis B vaccine	HBV	1 000
tissue plasminogen activator	TPA	10
superoxide dismutase	SOD	200

Table 2. Potential Hosts for Each Product

	host			
product	yeast extra	yeast intra	*E. coli*	CHO cells
INS	x		X	
HBV		x		x
TPA			X	x
SOD		x	X	

Table 3. Stages Considered for Each Host

		host			
stage	operation	yeast extra	yeast intra	*E. coli*	CHO cells
1	fermentation	x	x	x	x
2	solid−liquid separation	x	x	x	x
	2.A centrifuge				
	2.B microfilter				
3	cell disruption		x	x	
	3.A homogenizer				
	3.B bead mill				
4	solid−liquid separation			x	
	4.A centrifuge				
	4.B microfilter				
5	IB solubilization			x	
6	diafiltration			x	
7	sulfonation			x	
8	refolding			x	
9.	ultrafiltration			x	x
10	chromatography	x	x	x	x
11.	ultrafiltration diafilter	x	x	x	x
12	chromatography	x	x	x	x
13	ultrafiltration diafilter	x	x	x	x
14	gel chromatography	x	x	x	x
15	sterile filtration	x	x	x	x

The separation stages required by each host are displayed in Table 3. Yeast extracellular and CHO cells require the fewest number of stages: the fermenter, a solid liquid separation for cell harvesting, a series of chromatographic separations with intermediate ultrafiltrations for concentrating the diluted exit stream and washing it from the buffer with the addition of distilled water, and finally a sterile filtration. Yeast intracellular and *E. coli* require additional stages for cell disruption and separating the cell debris, whereas *E. coli* requires additional stages for inclusion body solubilization, oxidation and protein refolding. Figure 2 represents the processing stages involved, which depend on the host selected.

Both Figure 2 and Table 3 show that at stages 2, 3, and 4 there are different optional unit operations that could be used to perform the specified tasks. In the case of stages 2 and 4 that specify a solid liquid separation, this can be accomplished either by centrifugation or microfiltration. In the case of the cell disruption specified at stage 3, the optional units considered are a homogenizer and a bead mill. The problem formulation allows that the two optional units be adopted, e.g., a centrifuge

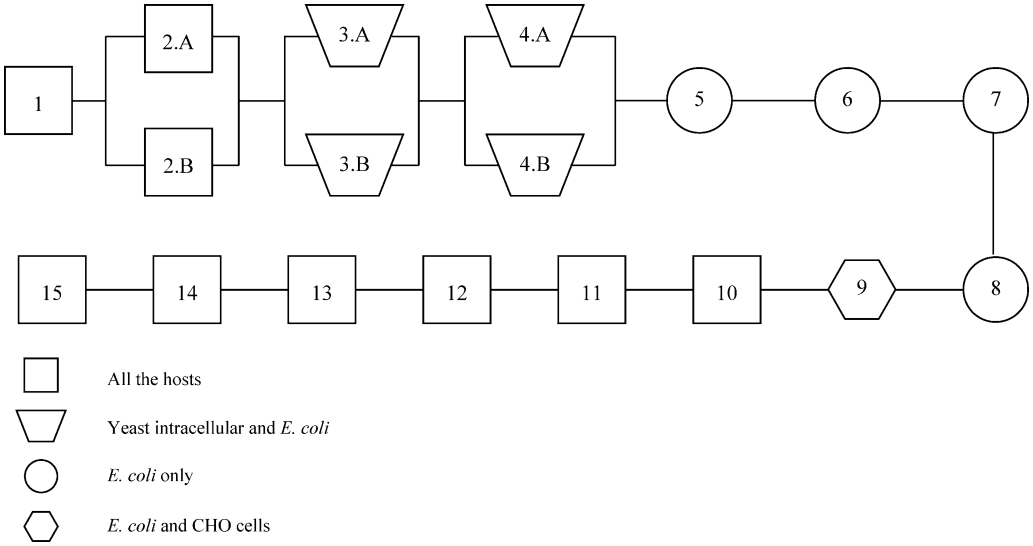

Figure 2. Stages and units used for each host.

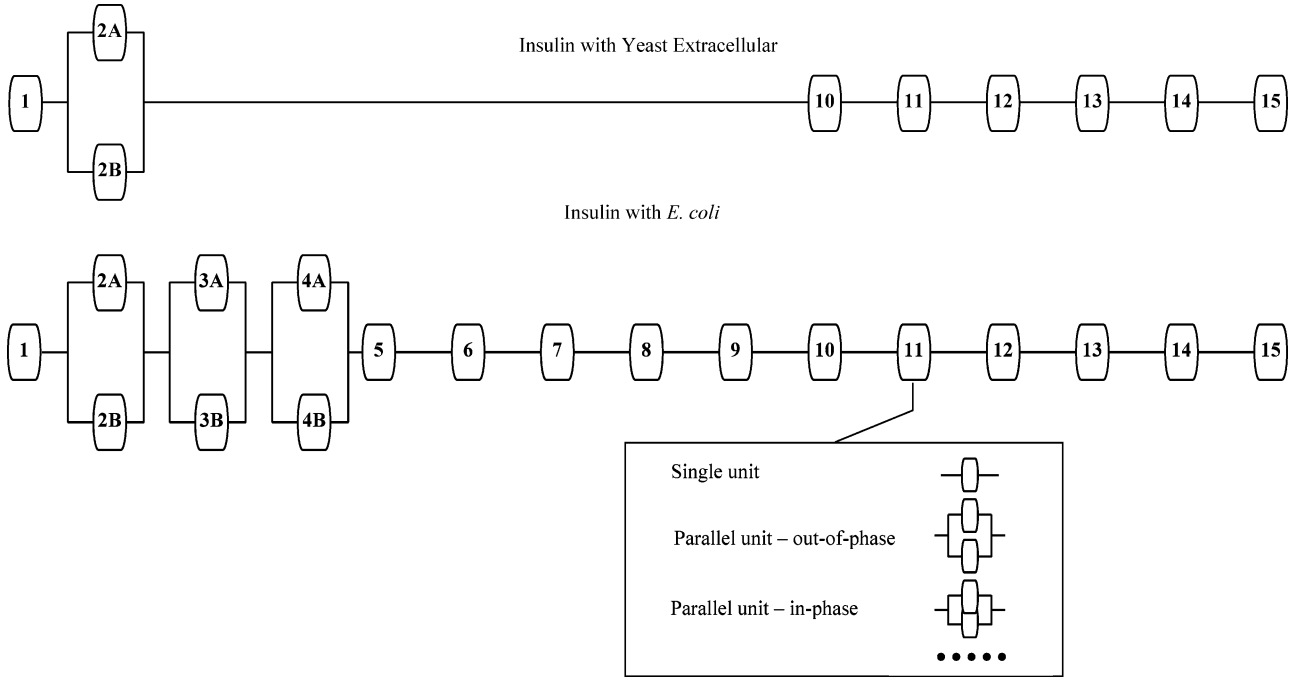

Figure 3. Options for insulin.

for processing some products and a microfiltration unit for the rest, but as these units will be idle when the plant processes the products that do not use this option, the optimal solution would usually select just one option to be shared by all products.

Figure 3 shows the options available in the case of producing insulin, as an illustration of the general problem at hand: one out of the two hosts available should be selected, which fixes which downstream stages are needed. Next, one option must be selected at each stage in which more than one optional unit operation is available. In the case of INS produced by yeast extracellular there are options only at stage 2, whereas if *E. coli* is the selected host, there are options available at stages 2, 3, and 4.

Figure 3 also illustrates the other sort of decisions that the model handles: at any stage there is the possibility of duplicating units, so the number of units in parallel at each stage and whether they will be operated in or out of phase are also decisions to be made. Figure 3 illustrates the alternatives of selecting just one unit, two units operated out of phase and two units in phase indicated by superposing the units; the dotted line denotes other alternatives that correspond to more than two units in parallel. While Figure 3 illustrates the problem for insulin, similar alternatives apply for the other products. As this is a multiproduct plant, the units adopted for an option at a stage are shared among all the products that use the same stage with that option.

Table 4 presents the yields η_{ijh} for product i produced by host h when processed at stage j. At stage 1 the yields are 1 because this is the product generation stage. At stage 2 the yield values are 1 when the product is intracellular because no cells are lost but less than 1 if it is present in the fermentation broth; in this last case the microfiltration option has larger yields because it uses additional distilled water for depleting the retentate from the valuable product. At stage 3 the homogenizer, which is a cheaper technology than bead milling, has smaller

Table 4. Production Yields

		η_{ijh} product/host							
stage	operation	INS yeast	INS *E. coli*	HBV yeast	HBV CHO	TPA *E. coli*	TPA CHO	SOD yeast	SOD *E. coli*
1	fermentation	1	1	1	1	1	1	1	1
2.A	centrifugation	0.75	1	1	0.8	1	0.8	1	1
2.B	microfiltration	0.85	1	1	0.9	1	0.9	1	1
3.A	homogeinization		0.7	0.75		0.7		0.75	0.7
3.B	bead milling		0.8	0.85		0.8		0.85	0.8
4.A	centrifugation		1	0.8		1		0.8	1
4.B	microfiltration		1	0.9		1		0.9	1
5	IB solubilization		0.7			0.7			0.7
6	ultrafiltration		1			1			1
7	sulfonation		0.9			0.9			0.6
8	refolding		0.6			0.6			0.6
9	ultrafiltration		1		1	1	1		1
10	chromatography	0.75	0.75	0.8	0.8	0.9	0.9	0.85	0.85
11	ultrafiltration	1	1	1	1	1	1	1	1
12	chromatography	0.9	0.9	0.85	0.85	0.9	0.9	0.85	0.85
13	ultrafiltration	1	1	1	1	1	1	1	1
14	gel chromatography	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
15	sterile filtration	1	1	1	1	1	1	1	1

Table 5. Cost Data for All Stages

stage	cost (U.S. $)		
1		63,400 $V^{0.6}$ fermenter	
2.A	5,750 $V^{0.6}$ feed tank	28,600 $R^{0.7}$ centrifuge	5,750 $V^{0.6}$ product tank
2.B	5,750 $V^{0.6}$ retentate tank	2,900 $R^{0.85}$ microfilter	5,750 $V^{0.6}$ permeate tank
3.A	5,750 $V^{0.6}$ holding tank	12,100 $R^{0.75}$ homogenizer	
3.B	5,750 $V^{0.6}$ holding tank	27,630 $R^{0.5}$ bead mill	
4.A	5,750 $V^{0.6}$ feed tank	28,600 $R^{0.7}$ centrifuge	5,750 $V^{0.6}$ product tank
4.B	5,750 $V^{0.6}$ retentate tank	2,900 $R^{0.85}$ microfilter	5,750 $V^{0.6}$ permeate tank
5		31,000 $V^{0.5}$ solubilization reactor	
6	5,750 $V^{0.6}$ holding tank	2,900 $R^{0.85}$ diafilter	
7		31,000 $V^{0.5}$ sulfonator	
8		31,000 $V^{0.5}$ refolding	
9	5,750 $V^{0.6}$ holding tank	2,900 $R^{0.85}$ ultrafilter	
10	5,750 $V^{0.6}$ holding tank	310,000 $A^{0.55}$ chromatographic column	5,750 $V^{0.6}$ product tank
11	5,750 $V^{0.6}$ holding tank	2,900 $R^{0.85}$ ultrafilter	
12	5,750 $V^{0.6}$ holding tank	310,000 $A^{0.55}$ chromatographic column	5,750 $V^{0.6}$ product tank
13	5,750 $V^{0.6}$ holding tank	2,900 $R^{0.85}$ ultrafilter	
14	5,750 $V^{0.6}$ holding tank	310,000 $A^{0.55}$ chromatographic column	5,750 $V^{0.6}$ product tank
15	5,750 $V^{0.6}$ feed tank	2,900 $R^{0.85}$ centrifuge	5,750 $V^{0.6}$ permeate tank

yield because its action is harsher and tends to denature the product released. Stage 4 is again a solid liquid separation with yield value equal to 1 if the product is in the solid inclusion bodies but less than 1 if it is in solution, with the yield larger in the case of microfiltration because of the product recovery by diafiltration. Stages 5−8 recover the product from *E. coli* inclusion bodies, the different reaction stages have various reported yields while in the intermediate ultrafiltration stages that concentrate the solution and wash it from the reactants used in the previous step, the yield is 1 because the membrane cutoff is selected to retain the protein. Stage 9 concentrates the feed to the chromatographic steps with η equal to 1; this stage is needed only in the case of very diluted solutions: the ones coming from the refolding step of proteins produced by *E. coli* and the solution of proteins produced by CHO cells at small concentrations. The purification stages 10−15 have various reported values of η for the chromatographic steps and η equal to 1 for the filtration steps that fully retain the proteins. Note that, as will be shown in this paper, yields are included in the size and time factors.

Detailed Models of the Stages

Tables 5−7 present the cost data and the stage models used in this process example, including the different options in case that they exist. Items 1, 5, 7, and 8 are batch reactors and thus represented by conventional batch stages described by a size factor in units [m^3/kg] and a fixed batch cycle time [h]. The remaining stages consist of vessels and semicontinuous units and are described by duty factors [m^3/kg].

Tables 6 and 7 present the data for stage 1, the fermenter, which is the product generation stage. The size factors of this stage are numerical values that are the inverse of the final concentration of the product in the fermentation broth, divided by the multiplication of the yields of all the downstream stages. This term, which is the overall yield of the downstream process, relates the mass of product present in the batch leaving the fermenter with the mass of final product leaving the plant. The size factor of stage 1 then depends on which options are selected at the stages downstream of it. The cycle times of this conventional batch stage are fixed amounts of time, which include the time for charging and discharging plus the operating time that is fixed by the kinetics of the reaction once the extent of reaction has been fixed.

Stage 2 performs a solid liquid separation, which has two optional unit operations, centrifugation and membrane separation, and in both cases the stage consists of two tanks and the semicontinuous unit. The cycle time of this batch stage consists of a fixed amount for charging and discharging, plus the operating time of the semicontinuous unit.

Table 6 presents the data for centrifugation. In this case, one tank contains the feed and the other contains the product of the stage that is either the concentrated

Table 6. Size and Duty Factors for All Stages

	S_{ih} [m^3/kg] (D_{ih}) product/host							
stage	INS yeast	INS *E. coli*	HBV yeast	HBV CHO	TPA *E. coli*	TPA CHO	SOD yeast	SOD *E. coli*
1 fermentation	$0.05/\prod_{j=2}^{15}\eta_{ijh}$	$0.03/\prod_{j=2}^{15}\eta_{ijh}$	$0.1/\prod_{j=2}^{15}\eta_{ijh}$	$0.5/\prod_{j=2}^{15}\eta_{ijh}$	$1./\prod_{j=2}^{15}\eta_{ijh}$	$10/\prod_{j=2}^{15}\eta_{ijh}$	$0.04/\prod_{j=2}^{15}\eta_{ijh}$	$0.05/\prod_{j=2}^{15}\eta_{ijh}$
2.A centrifuge								
feed tank	$0.05/\prod_{j=2}^{15}\eta_{ijh}$	$0.03/\prod_{j=2}^{15}\eta_{ijh}$	$0.1/\prod_{j=2}^{15}\eta_{ijh}$	$0.5/\prod_{j=2}^{15}\eta_{ijh}$	$1./\prod_{j=2}^{15}\eta_{ijh}$	$10/\prod_{j=2}^{15}\eta_{ijh}$	$0.04/\prod_{j=2}^{15}\eta_{ijh}$	$0.05/\prod_{j=2}^{15}\eta_{ijh}$
centrifuge D_{ih}	$0.0025/\prod_{j=2}^{15}\eta_{ijh}$	$0.15/\prod_{j=2}^{15}\eta_{ijh}$	$0.005/\prod_{j=2}^{15}\eta_{ijh}$	$0.0025/\prod_{j=2}^{15}\eta_{ijh}$	$5./\prod_{j=2}^{15}\eta_{ijh}$	$0.05/\prod_{j=2}^{15}\eta_{ijh}$	$0.002/\prod_{j=2}^{15}\eta_{ijh}$	$0.125/\prod_{j=2}^{15}\eta_{ijh}$
product tank	$0.05/\prod_{j=2}^{15}\eta_{ijh}$	$0.015/\prod_{j=2}^{15}\eta_{ijh}$	$0.025/\prod_{j=2}^{15}\eta_{ijh}$	$0.5/\prod_{j=2}^{15}\eta_{ijh}$	$0.5/\prod_{j=2}^{15}\eta_{ijh}$	$10/\prod_{j=2}^{15}\eta_{ijh}$	$0.01/\prod_{j=2}^{15}\eta_{ijh}$	$0.025/\prod_{j=2}^{15}\eta_{ijh}$
2.B microfilter								
retentate tank	$0.05/\prod_{j=2}^{15}\eta_{ijh}$	$0.03/\prod_{j=2}^{15}\eta_{ijh}$	$0.1/\prod_{j=2}^{15}\eta_{ijh}$	$0.5/\prod_{j=2}^{15}\eta_{ijh}$	$1./\prod_{j=2}^{15}\eta_{ijh}$	$10/\prod_{j=2}^{15}\eta_{ijh}$	$0.04/\prod_{j=2}^{15}\eta_{ijh}$	$0.05/\prod_{j=2}^{15}\eta_{ijh}$
microfilter D_{ih}	$0.5/\prod_{j=2}^{15}\eta_{ijh}$	$0.12/\prod_{j=2}^{15}\eta_{ijh}$	$0.375/\prod_{j=2}^{15}\eta_{ijh}$	$2.2/\prod_{j=2}^{15}\eta_{ijh}$	$4./\prod_{j=2}^{15}\eta_{ijh}$	$44./\prod_{j=2}^{15}\eta_{ijh}$	$0.15/\prod_{j=2}^{15}\eta_{ijh}$	$0.2/\prod_{j=2}^{15}\eta_{ijh}$
permeate tank	$0.1/\prod_{j=2}^{15}\eta_{ijh}$	No	No	$0.7/\prod_{j=2}^{15}\eta_{ijh}$	No	$14/\prod_{j=2}^{15}\eta_{ijh}$	No	No
3.A homogenization								
holding tank		$0.015/\prod_{j=2}^{15}\eta_{ijh}$	$0.025/\prod_{j=2}^{15}\eta_{ijh}$		$0.5/\prod_{j=2}^{15}\eta_{ijh}$		$0.01/\prod_{j=2}^{15}\eta_{ijh}$	$0.025/\prod_{j=2}^{15}\eta_{ijh}$
homogenizer		$0.045/\prod_{j=2}^{15}\eta_{ijh}$	$0.1/\prod_{j=2}^{15}\eta_{ijh}$		$1.5/\prod_{j=2}^{15}\eta_{ijh}$		$0.04/\prod_{j=2}^{15}\eta_{ijh}$	$0.075/\prod_{j=2}^{15}\eta_{ijh}$
3.B bead mill								
holding tank		$0.015/\prod_{j=2}^{15}\eta_{ijh}$	$0.025/\prod_{j=2}^{15}\eta_{ijh}$		$0.5/\prod_{j=2}^{15}\eta_{ijh}$		$0.01/\prod_{j=2}^{15}\eta_{ijh}$	$0.025/\prod_{j=2}^{15}\eta_{ijh}$
bead mill D_{ih}		$0.045/\prod_{j=2}^{15}\eta_{ijh}$	$0.025/\prod_{j=2}^{15}\eta_{ijh}$		$1.5/\prod_{j=2}^{15}\eta_{ijh}$		$0.01/\prod_{j=2}^{15}\eta_{ijh}$	$0.075/\prod_{j=2}^{15}\eta_{ijh}$
4.A centrifuge								
feed tank		$0.015/\prod_{j=2}^{15}\eta_{ijh}$	$0.025/\prod_{j=2}^{15}\eta_{ijh}$		$0.5/\prod_{j=2}^{15}\eta_{ijh}$		$0.01/\prod_{j=2}^{15}\eta_{ijh}$	$0.025/\prod_{j=2}^{15}\eta_{ijh}$
centrifuge D_{ih}		$0.1875/\prod_{j=2}^{15}\eta_{ijh}$	$0.5/\prod_{j=2}^{15}\eta_{ijh}$		$6.25/\prod_{j=2}^{15}\eta_{ijh}$		$0.2/\prod_{j=2}^{15}\eta_{ijh}$	$0.3125/\prod_{j=2}^{15}\eta_{ijh}$
product tank		No	$0.05/\prod_{j=2}^{15}\eta_{ijh}$		No		$0.02/\prod_{j=2}^{15}\eta_{ijh}$	No
4.B centrifuge								
retentate tank	$0.05/\prod_{j=2}^{15}\eta_{ijh}$	$0.03/\prod_{j=2}^{15}\eta_{ijh}$	$0.1/\prod_{j=2}^{15}\eta_{ijh}$	$0.5/\prod_{j=2}^{15}\eta_{ijh}$	$1./\prod_{j=2}^{15}\eta_{ijh}$	$10/\prod_{j=2}^{15}\eta_{ijh}$	$0.04/\prod_{j=2}^{15}\eta_{ijh}$	$0.05/\prod_{j=2}^{15}\eta_{ijh}$
microfilter D_{ih}	$0.5/\prod_{j=2}^{15}\eta_{ijh}$	$0.12/\prod_{j=2}^{15}\eta_{ijh}$	$0.375/\prod_{j=2}^{15}\eta_{ijh}$	$2.2/\prod_{j=2}^{15}\eta_{ijh}$	$4./\prod_{j=2}^{15}\eta_{ijh}$	$44./\prod_{j=2}^{15}\eta_{ijh}$	$0.15/\prod_{j=2}^{15}\eta_{ijh}$	$0.2/\prod_{j=2}^{15}\eta_{ijh}$
permeate tank	$0.1/\prod_{j=2}^{15}\eta_{ijh}$	No	No	$0.7/\prod_{j=2}^{15}\eta_{ijh}$	No	$14/\prod_{j=2}^{15}\eta_{ijh}$	no	no
5. IB solubilization		$0.1/\prod_{j=5}^{15}\eta_{ijh}$			$1./\prod_{j=5}^{15}\eta_{ijh}$			$0.5/\prod_{j=5}^{15}\eta_{ijh}$
6. diafiltration								
holding tank		$0.01/\prod_{j=5}^{15}\eta_{ijh}$			$1./\prod_{j=5}^{15}\eta_{ijh}$			$0.5/\prod_{j=5}^{15}\eta_{ijh}$
diafilter D_{ih}		$7./\prod_{j=5}^{15}\eta_{ijh}$			$70./\prod_{j=5}^{15}\eta_{ijh}$			$35./\prod_{j=5}^{15}\eta_{ijh}$
7. sulfonation		$0.12/\prod_{j=5}^{15}\eta_{ijh}$			$1.2/\prod_{j=5}^{15}\eta_{ijh}$			$0.6/\prod_{j=5}^{15}\eta_{ijh}$
8. refolding		$1./\prod_{j=8}^{15}\eta_{ijh}$			$20./\prod_{j=8}^{15}\eta_{ijh}$			$2./\prod_{j=8}^{15}\eta_{ijh}$
9. ultrafiltration								
holding tank		$1./\prod_{j=8}^{15}\eta_{ijh}$		from 2.A $0.5/\prod_{j=3}^{15}\eta_{ijh}$ from 2.B $0.7/\prod_{j=2}^{15}\eta_{ijh}$	$20./\prod_{j=8}^{15}\eta_{ijh}$	from 2.A $10./\prod_{j=3}^{15}\eta_{ijh}$ from 2.B $14/\prod_{j=2}^{15}\eta_{ijh}$		$2./\prod_{j=8}^{15}\eta_{ijh}$
ultrafilter D_{ih}		$50./\prod_{j=8}^{15}\eta_{ijh}$		from 2.A $25./\prod_{j=3}^{15}\eta_{ijh}$ from 2.B $35./\prod_{j=2}^{15}\eta_{ijh}$	$1000./\prod_{j=8}^{15}\eta_{ijh}$	from 2.A $500./\prod_{j=3}^{15}\eta_{ijh}$ from 2.B $700/\prod_{j=2}^{15}\eta_{ijh}$		$100./\prod_{j=8}^{15}\eta_{ijh}$
10. chromatography								
feed tank	from 2.A $0.05/\prod_{j=3}^{15}\eta_{ijh}$ from 2.B $0.1/\prod_{j=2}^{15}\eta_{ijh}$	$0.4/\prod_{j=10}^{15}\eta_{ijh}$	from 4.A $0.02/\prod_{j=2}^{15}\eta_{ijh}$ from 4.B $0.05/\prod_{j=2}^{15}\eta_{ijh}$	$0.4/\prod_{j=10}^{15}\eta_{ijh}$	$0.4/\prod_{j=10}^{15}\eta_{ijh}$	$0.4/\prod_{j=10}^{15}\eta_{ijh}$	from 4.A $0.008/\prod_{j=3}^{15}\eta_{ijh}$ from 4.B $0.02/\prod_{j=2}^{15}\eta_{ijh}$	$0.4/\prod_{j=10}^{15}\eta_{ijh}$
chromatographic column	$0.5/\prod_{j=10}^{15}\eta_{ijh}$	$0.5/\prod_{j=10}^{15}\eta_{ijh}$	$0.5/\prod_{j=10}^{15}\eta_{ijh}$	$0.8/\prod_{j=10}^{15}\eta_{ijh}$	$0.5/\prod_{j=10}^{15}\eta_{ijh}$	$0.8/\prod_{j=10}^{15}\eta_{ijh}$	$0.8/\prod_{j=10}^{15}\eta_{ijh}$	$0.8/\prod_{j=10}^{15}\eta_{ijh}$
product tank	$0.1/\prod_{j=11}^{15}\eta_{ijh}$	$0.1/\prod_{j=11}^{15}\eta_{ijh}$	$0.1/\prod_{j=11}^{15}\eta_{ijh}$	$2./\prod_{j=11}^{15}\eta_{ijh}$	$0.1/\prod_{j=11}^{15}\eta_{ijh}$	$2./\prod_{j=11}^{15}\eta_{ijh}$	$2./\prod_{j=11}^{15}\eta_{ijh}$	$2./\prod_{j=11}^{15}\eta_{ijh}$
11. ultrafiltration								
holding tank	$0.1/\prod_{j=11}^{15}\eta_{ijh}$	$0.1/\prod_{j=11}^{15}\eta_{ijh}$	$0.1/\prod_{j=11}^{15}\eta_{ijh}$	$2./\prod_{j=11}^{15}\eta_{ijh}$	$0.1/\prod_{j=11}^{15}\eta_{ijh}$	$2./\prod_{j=11}^{15}\eta_{ijh}$	$2./\prod_{j=11}^{15}\eta_{ijh}$	$2./\prod_{j=11}^{15}\eta_{ijh}$
ultrafilter D_{ih}	$5./\prod_{j=11}^{15}\eta_{ijh}$	$5./\prod_{j=11}^{15}\eta_{ijh}$	$5./\prod_{j=11}^{15}\eta_{ijh}$	$100./\prod_{j=11}^{15}\eta_{ijh}$	$5./\prod_{j=11}^{15}\eta_{ijh}$	$100./\prod_{j=11}^{15}\eta_{ijh}$	$100./\prod_{j=11}^{15}\eta_{ijh}$	$100./\prod_{j=11}^{15}\eta_{ijh}$
12. chromatography								
feed tank	$0.05/\prod_{j=12}^{15}\eta_{ijh}$	$0.05/\prod_{j=12}^{15}\eta_{ijh}$	$0.05/\prod_{j=12}^{15}\eta_{ijh}$	$0.05/\prod_{j=12}^{15}\eta_{ijh}$	$0.05/\prod_{j=12}^{15}\eta_{ijh}$	$0.05/\prod_{j=12}^{15}\eta_{ijh}$	$0.05/\prod_{j=12}^{15}\eta_{ijh}$	$0.05/\prod_{j=12}^{15}\eta_{ijh}$
chromatographic column	$0.7/\prod_{j=12}^{15}\eta_{ijh}$	$0.7/\prod_{j=12}^{15}\eta_{ijh}$	$0.8/\prod_{j=12}^{15}\eta_{ijh}$	$0.8/\prod_{j=12}^{15}\eta_{ijh}$	$0.8/\prod_{j=12}^{15}\eta_{ijh}$	$0.8/\prod_{j=12}^{15}\eta_{ijh}$	$0.8/\prod_{j=12}^{15}\eta_{ijh}$	$0.8/\prod_{j=12}^{15}\eta_{ijh}$
product tank	$1./\prod_{j=13}^{15}\eta_{ijh}$	$1./\prod_{j=13}^{15}\eta_{ijh}$	$2./\prod_{j=13}^{15}\eta_{ijh}$	$2./\prod_{j=13}^{15}\eta_{ijh}$	$2./\prod_{j=13}^{15}\eta_{ijh}$	$2./\prod_{j=13}^{15}\eta_{ijh}$	$2./\prod_{j=13}^{15}\eta_{ijh}$	$2./\prod_{j=13}^{15}\eta_{ijh}$
13. ultrafiltration								
holding tank	$1./\prod_{j=13}^{15}\eta_{ijh}$	$1./\prod_{j=13}^{15}\eta_{ijh}$	$2./\prod_{j=13}^{15}\eta_{ijh}$	$2./\prod_{j=13}^{15}\eta_{ijh}$	$2./\prod_{j=13}^{15}\eta_{ijh}$	$2./\prod_{j=13}^{15}\eta_{ijh}$	$2./\prod_{j=13}^{15}\eta_{ijh}$	$2./\prod_{j=13}^{15}\eta_{ijh}$
ultrafilter D_{ih}	$50./\prod_{j=13}^{15}\eta_{ijh}$	$50./\prod_{j=13}^{15}\eta_{ijh}$	$100./\prod_{j=13}^{15}\eta_{ijh}$	$100./\prod_{j=13}^{15}\eta_{ijh}$	$100./\prod_{j=13}^{15}\eta_{ijh}$	$100./\prod_{j=13}^{15}\eta_{ijh}$	$100./\prod_{j=13}^{15}\eta_{ijh}$	$100./\prod_{j=13}^{15}\eta_{ijh}$
14. chromatography								
feed tank	$0.05/\prod_{j=14}^{15}\eta_{ijh}$	$0.05/\prod_{j=14}^{15}\eta_{ijh}$	$0.05/\prod_{j=14}^{15}\eta_{ijh}$	$0.05/\prod_{j=14}^{15}\eta_{ijh}$	$0.05/\prod_{j=14}^{15}\eta_{ijh}$	$0.05/\prod_{j=14}^{15}\eta_{ijh}$	$0.05/\prod_{j=14}^{15}\eta_{ijh}$	$0.05/\prod_{j=14}^{15}\eta_{ijh}$
chromatographic column	$0.4/\prod_{j=14}^{15}\eta_{ijh}$	$0.4/\prod_{j=14}^{15}\eta_{ijh}$	$0.4/\prod_{j=14}^{15}\eta_{ijh}$	$0.4/\prod_{j=14}^{15}\eta_{ijh}$	$0.4/\prod_{j=14}^{15}\eta_{ijh}$	$0.4/\prod_{j=14}^{15}\eta_{ijh}$	$0.4/\prod_{j=14}^{15}\eta_{ijh}$	$0.4/\prod_{j=14}^{15}\eta_{ijh}$
product tank	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$
15. centrifugation								
feed tank	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$
centrifuge D_{ih}	1.	1.	1.	1.	1.	1.	1.	1.
product tank	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$	$0.1/\prod_{j=15}^{15}\eta_{ijh}$

cell suspension in the case of intracellular proteins or the aqueous solution free of cells in the case of extracellular proteins. The size factor of the feed tank is the same as that of the fermenter upstream and the size factor of the product tank depends on the concentration factor adopted. In the case of intracellular proteins the centrifuge concentrates the cells to an extent that is appropriate to processing in the cell disruption step. With extracellular proteins the concentration factor is as large as it can be practically achieved, because the volumetric ratio free cell solution to feed is also the product yield of this stage.

The size of the centrifuge is measured in kW so that the duty factor has units [kW·h/kg] and a value that is the inverse of the feed concentration in m^3/kg times a constant with units kW·h/m^3 that depends on the type of centrifuge and gravitational settling velocity of the

Table 7. Processing Times for All Stages

	T_{ih} [h] product/host							
stage	INS yeast	INS *E. coli*	HBV yeast	HBV CHO	TPA *E. coli*	TPA CHO	SOD yeast	SOD *E. coli*
1	24	24	24	168	24	168	24	24
5		8			12			10
7		12			18			15
8		8			12			10
10	$3.5 + \dfrac{0.6B_i}{\sum_{j=10}^{15} G_j A_j \prod \eta_{ijh}}$	$3.5 + \dfrac{0.6B_i}{\sum_{j=10}^{15} G_j A_j \prod \eta_{ijh}}$	$3.5 + \dfrac{0.6\,B_i}{\sum_{j=10}^{15} G_j A_j \prod \eta_{ijh}}$	$15.0 + \dfrac{0.35B_i}{\sum_{j=10}^{15} G_j A_j \prod \eta_{ijh}}$	$3.5 + \dfrac{0.6B_i}{\sum_{j=10}^{15} G_j A_j \prod \eta_{ijh}}$	$15.0 + \dfrac{0.35B_i}{\sum_{j=10}^{15} G_j A_j \prod \eta_{ijh}}$	$15.0 + \dfrac{0.35B_i}{\sum_{j=10}^{15} G_j A_j \prod \eta_{ijh}}$	$15.0 + \dfrac{0.35B_i}{\sum_{j=10}^{15} G_j A_j \prod \eta_{ijh}}$
12	$8.0 + \dfrac{0.5B_i}{\sum_{j=12}^{15} G_j A_j \prod \eta_{ijh}}$	$8.0 + \dfrac{0.5B_i}{\sum_{j=12}^{15} G_j A_j \prod \eta_{ijh}}$	$15.0 + \dfrac{0.35B_i}{\sum_{j=12}^{15} G_j A_j \prod \eta_{ijh}}$	$15.0 + \dfrac{0.35\,B_i}{\sum_{j=12}^{15} G_j A_j \prod \eta_{ijh}}$	$15.0 + \dfrac{0.35\,B_i}{\sum_{j=12}^{15} G_j A_j \prod \eta_{ijh}}$	$15.0 + \dfrac{0.35B_i}{\sum_{j=12}^{15} G_j A_j \prod \eta_{ijh}}$	$15.0 + \dfrac{0.35B_i}{\sum_{j=12}^{15} G_j A_j \prod \eta_{ijh}}$	$15.0 + \dfrac{0.35B_i}{\sum_{j=12}^{15} G_j A_j \prod \eta_{ijh}}$
14	5	5	5	5	5	5	5	5

solid particles. We took a value of 5 kW·h/m^3 from Petrides et al. (1995) for harvesting *E. coli* in disk stack centrifuges and extrapolated it for the settling velocities of the other cells.

Table 6 presents the data for the second option at stage 2, which is a microfiltration membrane. The feed tank is sized as in the centrifugation option, whereas the product tank is not needed when the protein is intracellular because the first vessel acts as the retentate tank whose cell concentration increases from the initial value of the fermentation broth to the final value corresponding to the concentration factor adopted. When the protein is extracellular the second tank is required and must hold not only the original amount of cell-free broth but also the diafiltration water used to further increase the recovery of valuable product; we took the volume of the outlet batch to be twice the inlet volume in the case of INS produced by yeast (Montagna et al., 2000) and 1.4 times the inlet volume for HBV and TPA produced by CHO cells (Datar et al., 1993).

The size of the membrane unit is measured in m^2 so that the duty factor has units m^2·h/kg and a value that is the ratio volume of permeate to mass of product exiting this stage in m^3/kg times a constant with units m^2·h/m^3 that for diluted feeds is the inverse of the membrane permeability and depends on the type and cutoff of the membrane. Values of permeability in [m^3/h·m^2] of 0.2 for a cutoff of 10 μm to retain yeast and 0.1 for a cutoff of 0.2 μm to retain cell debris (Pinto et al., 2001) were extrapolated to the other cutoff values (100 μm for CHO cells and 1 μm for *E. coli*) with the following correlation:

$$\text{Permeability } [\text{m}^3/\text{h}\cdot\text{m}^2] = 0.125d^{0.2} \qquad (21)$$

which predicts a quite weak dependence on the solid particle diameter d as compared with centrifugation where the settling velocity is proportional to d^2. The tradeoff between selecting centrifugation and membrane separation strongly depends on this issue, with the crossover point in the 1−2 μm range (Asenjo and Patrick, 1990). In the problem formulation presented in this paper the optimal solution may choose any one of them or adopt both of them to be used in different separation tasks.

Stage 3 performs cell disruption that can be performed in two optional units, homogenizer or bead mill. In both cases, the stage consists of a holding tank and the semicontinuous unit that takes the cell suspension held in the tank and discharges the processed stream into the same vessel; the batch is recirculated through the holding vessel as many times as homogenization passes are needed. Thus this stage consists of both a batch and a semicontinuous item and the batch stage processing time consists of the operating time of the semicontinuous item plus a time for filling and emptying the vessel.

Table 6 presents the data for the homogenizer and for the bead mill. In both cases the tank size factor is the inverse of the concentration of the cell suspension received from stage 2. The size of the semicontinuous unit is its volumetric rate capacity measured in m^3/h, and the duty factor in m^3/kg corresponds to the same size factor of the holding tank times the number of disruption passes required. Kula and Schütte (1987) suggest that the yeast requires four homogenization passes but only one bead mill pass with a residence time of 2 min, whereas *E. coli* requires three passes in each unit.

Stages 4−9 do not introduce new unit operations. Thus, for the sake of simplicity we do not provide a detailed description of their size and time factors, but some issues should be noted.

Table 8. Bounds on Equipment Sizes

unit	lower bound	upper bound
fermenter (*V*)	0.2 m^3	100. m^3
microfilter (*R*)	0.1 m^3	50 m^3
homogenizer (*R*)	0.1 m^3/h	20 m^3/h
bead mill (*R*)	0.05 m^3/h	10. m^3/h
centrifuge (*R*)	0.1 kW	20. kW
reactors (*V*)	0.2 m^3	100. m^3
chromatography (*A*)	0.0001 m^2	0.75 m^2
tanks (*V*)	0.2 m^3	100. m^3

The size factors are computed by tracking the amount of product and the batch volume changes that occur at each stage. Consequently, the size factor of a stage always depends on the yields of the downstream stages but may also depend on the yields or volume changes of its upstream stages. For example, in Table 6 for the size factors of the microfilter option at stage 4, the multiplication of yields in the denominator starts from stage 2 because it keeps track of the amount of product starting from the product generation in stage 1.

Similarly in Table 6, that corresponds to the ultrafiltration at stage 9 for concentrating the feed to the first chromatography up to 2.5 kg/m^3 (Datar el al., 1993) the computation of the batch volume that arrives at this stage needs to discriminate whether the option selected at stage 2 is centrifugation or microfiltration.

The batch cycle times of reaction stages 5, 7, and 8 are composed of fixed amounts of charging and discharging times and the operation time, which is also a constant value because these operations are governed by kinetics; once the extent of reaction has been fixed, it sets the reaction time.

From Tables 3 and 4 it can be seen that the only new unit operation that appears in the rest of the process is chromatography. Note that the cross sectional area of the column, A_{jd}, is selected as the characteristic size of this unit. We assume that the height is large enough to accommodate the heights required by each combination of chromatographic packing and feed to be processed. The cost of these packings has not been taken into account because we assume that the sequence of chromatographic steps required by each product is set, i.e., we do not include this decision as an additional degree of freedom. With respect to the unit operating time, it consists of a fixed amount plus the following term that is typical of semicontinuous items:

$$T_{ij} = \frac{V_{aux}\,[m^3]}{A\,[m^2]\cdot v\,[m/s]} + \frac{V_{feed}\,[m^3]}{A\,[m^2]\cdot v\,[m/s]} \qquad (22)$$

In eq 22, V_{aux} is the volume of washing, equilibration, regeneration and eluant solutions, V_{feed} is the volume of the batch fed to the column, *A* is the cross sectional area of the column, and *v* is a design linear velocity of the mobile phase. As V_{aux} is given as a fixed number of packing volumes, when divided by *A* this gives the height of packing, which is fixed for each separation. As *v* is also a fixed design value, the first part of the right-hand side of eq 22 is a fixed, and usually large, value. Otherwise as V_{feed} is related through a concentration term with the batch size B_i, we can recognize that this part of the operation time has the form of the last term at the right-hand side of eq 19.

Finally, Table 8 presents the upper and lower bounds for the size of all equipment involved in the process.

Optimization Results

As previously mentioned, the mathematical model is a MINLP with 507 nonlinear constraints involving 348 continuous and 225 binary variables. This program was solved with the software DICOPT++ included in the optimization package GAMS (Brooke et al., 1992).

DICOPT++ resorts to the Outer Approximation/Equality Relaxation/Augmented Penalty (OA/ER/AP) method proposed by Viswanathan and Grossmann (1990). The model implementation was done in a PC Intel Celeron 650 MHz where it demanded 52 s.

The process example was solved assuming a time horizon *H* of 6000 h, obtaining an optimal solution with a value for the objective function of $6,308,313.70. Table 9 presents the results for each product, reporting in each case the selected host, the batch size and the cycle time. Table 10 reports the results for each stage that are the size of its units and the number of units working in or out of phase. Figure 4 shows the optimal configuration of the plant, indicating that the units work in phase by superposing them or out of phase otherwise.

Two hosts were selected out of four available, with two being the minimum feasible number because none of them was able to produce all products. This outcome has a rationale: the multiproduct plant is best suited for implementing processes that are similar to each other, which minimizes equipment under utilization.

E. coli was selected as one of the hosts, even if it was the one that required the largest number of downstream stages. CHO cell host was discarded apparently because of its much larger fermentation cycle time: as the model did not allow extra storage for reducing the idle time of downstream stages but only to unit duplication, this would have required a large number of fermenters operating out of phase, and duplication of units has an economic penalty because of the 0.6 size exponent in the unit cost correlation.

Once CHO cells were discarded, both *E. coli* and Yeast intracellular had to be selected because they were the only other hosts for producing TPA and HBV, respectively. SOD could be produced by any of them and Yeast intracellular was selected. INS could have incorporated a third host, namely Yeast extracellular which was the host that required the fewer number of downstream stages, but this option was not selected. However, once the stages for cell disruption and cell debris separation had to exist because they were required by other prod-

Table 9. Product Results

product	host used	B_i (kg)	TL_i (h)
INS	*E. coli*	3.86	5.00
HBV	yeast intracellular	3.19	7.72
TPA	*E. coli*	0.24	12.00
SOD	yeast intracellular	2.71	15.44

Table 10. Equipment Results

stage	1st batch item size	semicontinuous item size	2nd batch item size	units in phase	units out of phase
1	1.239 m^3			1	5
2.B	1.239 m^3	0.567 m^2	0.200 m^3	1	1
3.B	0.619 m^3	0.213 m^3/h		1	1
4.B	0.619 m^3	0.886 m^2	0.383 m^3	1	1
5	1.890 m^3			1	2
6	1.890 m^3	26.455 m^2		1	1
7	2.268 m^3			1	3
8	12.485 m^3			1	3
9	4.162 m^3	39.683 m^2		3	1
10	0.571 m^3	0.750 m^2	1.594 m^3	5	1
11	7.969 m^3	25.810 m^2		1	1
12	0.200 m^3	0.750 m^2	1.594 m^3	5	2
13	3.984 m^3	12.905 m^2		2	1
14	0.200 m^3	0.643 m^2	0.200 m^3	3	1
15	0.386 m^3	0.771 m^2	0.386 m^3	1	1

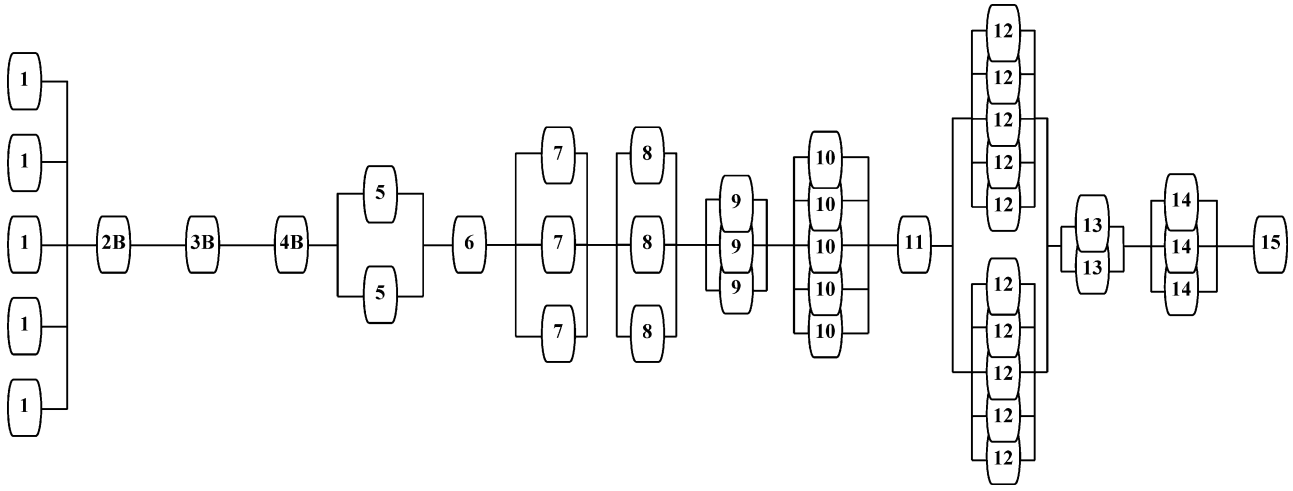

Figure 4. Optimal configuration of the plant.

ucts, the fact that Yeast extracellular does not require them was not a bonus.

Stages 2−4 had optional units for performing the assigned task, and the optimal solution selected the same option for all four products. This could intuitively be expected a priori given that if different products required different options, this would increase under utilization of the equipment of this multiproduct plant.

Microfiltration was preferred over centrifugation in both stages 2 and 4 but the outcome could also have been centrifugation for stage 2, which separates larger solid particles than in stage 4. Probably the larger yields of microfiltration (because it recovers product by diafiltration) decided this issue. Even if the recovery of product with additional distilled water is also possible with centrifugation, this would require successive centrifugations diluting the concentrated solids suspension from the previous stage.

At stage 3, bead mill was preferred over homogenization even if this is a more expensive technology. This occurred because yeast is more effectively disrupted by bead milling than by homogenization, requiring 1 and 4 passes, respectively, through the cell disruption unit. If the multiproduct plant handled only *E. coli*, which requires the same number of passes through both units, then homogenization would have been the preferred choice.

It is interesting to note that the possibility of duplicating units, either in phase or out of phase, has been used at several stages, even combining both options at stage 12. Duplication out of phase was used to reduce the cycle time of the stage, while duplication in phase was used to permit processing of batches that are larger than the upper bound of the units at this stage. At stage 12 in the case of HBV produced by yeast intracellular, the operating time is 15.4 h. Working with two groups of units out of phase results in half the cycle time of the stage, processing a batch each other 7.7 h, which coincides with the plant cycle time for this product, i.e., with this stage still being the bottleneck.

Considering that the size factor for this chromatographic stage is 1.18 m^2/kg and that the batch size for this product is 3.19 kg, we compute the overall size requirement of 3.75 m^2. Taking into account that the upper bound for chromatographic columns is 0.75 m^2, the model selects 5 units in phase to process the product batch.

An interesting comparison can be made by imposing CHO cells to be included in the set of hosts. In this case, the optimal solution value is 9.79×10^6. It corresponds to an increase of 55% with respect to the optimal solution, which used only *E. coli* and intracellular yeast as hosts. The corresponding product-host assignment, batch sizes, and cycle times are shown in Table 11, whereas the sizes and number of units at each stage are given in Table 12. The main reason for the increase in cost when CHO cells are selected as a host is due to the large number of fermenters that must be added in the first stage.

Table 11. Product Results for CHO Cell Selection

product	host used	B_i (kg)	TL_i (h)
INS	*E. coli*	6.94	8.71
HBV	CHO cells	5.10	15.44
TPA	CHO cells	0.30	15.27
SOD	*E. coli*	5.24	15.42

Table 12. Equipment Results for CHO Selection

stage	1st batch item size	semicontinuous item size	2nd batch item size	units in phase	units out of phase
1	5.216			1	11
2.B	5.216	1.503	7.302	1	1
3.A	0.857	0.251		1	1
4.B	0.857	1.046	0.200	1	1
5	12.001			1	1
6	6.000	27.233		2	1
7	14.401			1	2
8	30.242			1	2
9	10.081	40.984		3	1
10	1.028	0.750	1.542	10	1
11	15.423	50.00		1	1
12	0.200	0.750	1.457	9	1
13	13.110	49.795		1	1
14	0.200	0.694	0.200	5	1
15	0.694	0.797	0.694	1	1

Conclusions

The contribution of this paper to the literature on bioprocess synthesis is to add the decision of selecting the host, solving a problem formulation that also simultaneously optimizes the selection of optional unit operations for the same task and the structure of the production plant considering its multiproduct nature.

The selection of hosts could be made by comparing case studies of single product processes as in Datar et al. (1993), without optimizing the selection of unit operations or structure of the plant. Furthermore, the synthesis of bioprocesses for single products could be made more rigorous by allowing many more unit operations to compete for each separation task as in Steffens et al. (2000a).

However, as shown in this paper, the multiproduct nature of the plant does strongly influence the decision making, e.g., the optimal set of hosts and optimal choice of unit operations are not independent. It was also illustrated how the structure of the plant could influence them, e.g., allowing or not intermediate storage could change the optimal set of hosts.

On the other hand, solving rigorously all the issues simultaneously may prove to be a formidable task, if feasible at all. So in our opinion, valuable approaches such as the above-mentioned ones should be used to bound the optimal options available for each product; nevertheless, the final decision of selecting among them should be arrived at by solving the simultaneous optimization as proposed here.

Finally, given the very large quantity of novel recombinant proteins for a number of novel therapeutic uses presently being approved or "in the pipeline", multiproduct and multihost recombinant protein production plants have recently been or are being built for the manufacture of these products. The strategy presented in this paper is of crucial value for the optimal utilization of such plants.

Acknowledgment

The authors would like to acknowledge financial support received from VITAE within the Cooperation Program Argentina-Brazil-Chile under grant B-11487/10B006. The Millenium Scientific Initiative (ICM 99-031-F) is also gratefully acknowledged.

Notation

Indices

d	unit
h	host
i	product
j	stage

Sets

D_j	set of units available for performing stage j
H_i	set of hosts available for expression of product i
J_h	set of stages j needed if host h is selected
N	set of stages
P	set of products

Parameters

D_{ijdh}	duty factor for product i in stage j with option d using host h
H	time horizon
$M1_{ijdh}$, $M2_{ijdh}$	Big-M constants for mixed-integer constraints
Q_i	production target for product i
S_{ijdh}	size factor for stage j that uses option d obtained for product i with host h
T^0_{ijdh}, T^1_{ijdh}	operating time constants for product i at stage j under option d for host h
α_{jd},β_{jd}	cost coefficient for unit j in stage d
θ_{ijdh}	operating time that the semicontinuous unit at stage j with option d needs to process a batch of i produced by host h

Variables

B_i	batch size for producing final product i
G_{jd}	number of units operating in phase at stage j when selecting option d
M_{jd}	number of units operating out of phase at stage j when selecting option d
R_{jd}	size of the semicontinuous unit in stage j using option d
T_{ijdh}	operating time for product i at stage j under option d for host h
TL_i	cycle time for product i
V_{jd}	size of units at stage j that correspond to option d
y_{ih}	1 if host h is chosen for product i
z_{ijdh}	1 if option d selected at stage j for producing i with host h; 0 otherwise.

References and Notes

Asenjo, J. A.; Patrick, I. *Protein Purification Applications: A Practical Approach*; Harris, E., Angal, S., Eds.; IRL Press: Oxford, 1990; Chapter 1.6.

Asenjo, J. A.; Montagna, J. M.; Vecchietti, A. R.; Iribarren, O. A.; Pinto, J. M. Strategies for the Simultaneous Optimization of the Structure and the Process Variables of a Protein Production Plant. *Comput. Chem. Eng.* **2000**, *24*, 2277−2290.

Brooke, A.; Kendrick, D.; Meeraus, A. *GAMS Release 2.25. User's Guide*; The Scientific Press: San Francisco, 1992.

Bryant, C. H.; Rowe, R. C. Knowledge Discovery in Databases: Application to Chromatography. *Trends Anal. Chem.* **1998**, *17*, 18−24.

Datar, R. V.; Cartwright, T.; Rosen, C. G. Process Economics of Animal Cell and Bacterial Fermentations: A Case Study Analysis of Tissue Plasminogen Activator. *Bio/Technology* **1993**, *11*, 349−357.

Groep, M. E.; Gregory, M. E.; Kershenbaum, L. S.; Bogle, I. D. L. Performance Modeling and Simulation of Biochemical Process Sequences with Interacting Unit Operations. *Biotechnol. Bioengng.* **2000**, *67* (3), 301−311.

Grossmann, I. E.; Sargent, R. W. H. Optimal Design of Multipurpose Chemical Plants. *Ind. Eng. Chem. Process Des. Dev.* **1979**, *18*, 343−348.

Knopf, F. C.; Okos, M. R.; Reklaitis, G. V. Optimal Design of Batch/Semicontinuous Processes. *Ind. Eng. Chem. Process Des. Dev.* **1982**, *21*, 79−86.

Kula, M. R.; Schütte, H. Purification of Proteins and the Disruption of Microbial Cells. *Biotechnol. Prog.* **1987**, *3*, 31−42.

Montagna, J. M.; Vecchietti, A. R.; Iribarren, O. A.; Pinto, J. M.; Asenjo, J. A. Optimal Design of Protein Production Plants with Time and Size Factor Process Models. *Biotechnol. Prog.* **2000**, *16*, 228−237.

Petrides, D.; Sapidou, E.; Calandranis, J. Computer Aided Process Analysis and Economic Evaluation for Biosynthetic Human Insulin Production−A Case Study. *Biotechnol. Bioeng.* **1995**, *48*, 529−541.

Pinto, J. M.; Montagna, J. M.; Vecchietti, A. R.; Iribarren, O. A.; Asenjo, J. A. Process Performance Models in the Optimization of Multiproduct Protein Production Plants. *Biotechnol. Bioeng.* **2001**, *74*, 451−465.

Ravemark, D. E.; Rippin, D. W. T. Optimal Design of Multiproduct Batch Plants. *Comput. Chem. Eng.* **1998**, *22*, 177−183.

Salomone, H. E.; Iribarren, O. A. Posynomial Modeling of Batch Plants. *Comput. Chem. Eng.* **1992**, *16*, 173−184.

Salomone, H. E.; Montagna, J. M.; Iribarren, O. A. Dynamic Simulations in the Design of Batch Processes. *Comput. Chem. Eng.* **1994**, *18*, 191−204.

Samsatli, N. J.; Shah, N. An Optimization Based Design Procedure for Biochemical Processes. Part I: Preliminary Design and Operation. *Trans. Inst. Chem. Eng., Part C* **1996a**, *74*, 221−231.

Samsatli, N. J.; Shah, N. An Optimization Based Design Procedure for Biochemical Processes. Part II: Detailed Scheduling. *Trans. Inst. Chem. Eng., Part C* **1996b**, *74*, 232−242.

Steffens, M. A.; Fraga, E. S.; Bogle, I. D. L. Synthesis of Bioprocesses using Physical Properties Data. *Biotechnol. Bioeng.* **2000a**, *68*, 218−230.

Steffens M. A.; Fraga, E. S.; Bogle, I. D. L. Synthesis of Purification Tags for Optimal Downstream Processing. *Comput. Chem. Eng.* **2000b**, *24* (2), 717−720.

Vásquez-Alvarez, E.; Lienqueo, M. E.; Pinto, J. M. Optimal Synthesis of Protein Purification Processes. *Biotechnol. Prog.* **2001**, *17*, 685−696.

Viswanathan, J. V.; Grossmann, I. E. A Combined Penalty Function and Outer Approximation Method for MINLP Optimization. *Comput. Chem. Eng.* **1990**, *14*, 769−782.

Accepted for publication February 20, 2004.

BP034330+