STRATEGIES FOR SOLVING THE SYNTHESIS PROBLEM IN MULTIPRODUCT BATCH PROTEIN PRODUCTION PROCESSES

Aldo Vecchietti[(1)(2)], Jorge Montagna[(1)(2)] and Oscar Iribarren[(1)]
[(1)] INGAR – Instituto de Desarrollo y Diseño – Santa Fe – Argentina
[(2)] Universidad Tecnológica Nacional – Regional Santa Fe
Avellaneda 3657- Santa Fe – Argentina

Abstract

This paper deals with the synthesis and design of a multiproduct batch protein production processes. Four products and at most 15 batch stages are involved in the plant synthesis and design. Constant size and time factors are used for the design constraints. The host, batch stages and type of unit operating in some of the stages are discrete decisions included in the model. A disjunctive model has been formulated for the problem. The model solution has been performed transforming the disjunctive model into Mixed Integer Nonlinear Programs (MINLP) using the Big-M and convex-hull relaxations of a disjunctive set. The results obtained and the behavior of both MINLĐ models used to solve the problem is shown.

Keywords

Multiproduct protein production plant, Synthesis, Generalized Disjunctive Model, Convex Hull, BigM, MINLP

Introduction

Product development is one of the main concerns for biotechnological industry. New products are generated at the laboratory and then the process to produce and purify the products in large amounts is required. Sometimes the process engineer has several options to choose between different processes and the decision it makes has a big impact on the economy of the manufacturing process. In recent works in the literature the optimal synthesis, design and operation of the protein production plant has been studied (Montagna et al. (2000); Asenjo et al (2000); Alvarez et al, 2001; Steffens et al., 2000). The paper of Steffens et al. (2000) shows the synthesis of a bioprocess using physical property information. They generate a ranked list of flow sheets, which may be analyzed later in more detail with another tool. Alvarez et al. (2000) propose an optimal synthesis methodology based on mixed integer linear (MILP) mathematical programs. The authors deal with the purification section of a downstream process based on chromatographic stages. They also use physicochemical data information of a protein mixture. This paper explores a methodology for the synthesis and design of a multiproduct protein production plant, using constant size and time factors for the batch stages. As a case of study we consider the production of four products: A, B, C, and D.

The synthesis decisions are related to the host to use and unit selection for the batch stages. Four different hosts are used: H1, H2, H3 or H4. Two alternative hosts can be used for each product, the options are: H1 or H2 for product A, H3 or H4 for product B, H1 or H4 for product C, and H1 or H3 for product D. Different stage sets are needed according to the host selected. Table 1 presents the hosts, batch stages and alternative units involved for each of the four products

Table 1: Host and batch stages involved in the production of proteins A, B, C and D

	Host (Products)			
Stages	H1 (A,C,D)	H2 (A)	H3 (B,D)	H4 (B,C)
1	x	x	x	x
2.A or 2.B	x	x	x	x
3.A or 3.B	x	-	x	-
4.A or 4.B	x	-	x	-
5	x	-	-	-
6	x	-	-	-
7	x	-	-	-
8	x	-	-	-
9	x	-	-	x
10	x	x	x	x
11	x	x	x	x
12	x	x	x	x
13	x	x	x	x
14	x	x	x	x
15	x	x	x	x

x : the stage apply to the host
-: stage do not apply to the host

There is an intersection stage set between the different hosts while some stages operate for one of the host exclusively. Once the host is selected some structural decisions must be made in order to select the unit type operating in some stages. In this case the options are: membranes (unit 2.A in Table 1) or centrifugation (unit 2.B) can make the cell harvesting operation. The same option is included for the solid-liquid separation (4.A and 4.B respectively). The options for the operation of the cell rupture stage are homogenization (3.A) or bead milling (3.B). Stages 5 (solubilization), 6 (ultrafiltration), 7 (sulfonation), and 8 (refolding) are included in the batch plant in the case Escherichia Coli is adopted as host. Decisions are hierarchical because once the host is selected then the units operating the multiproduct batch plant must be chosen. Table 1 shows that at most 15 batch stages are involved in the protein production plant but not all of them are mandatory depending on host selected.

The approach used to solve the problem is executed into two steps: first, the problem is formulated as a Generalized Disjunctive Program (GDP) (Lee and Grossmann, 2000), then is transformed as a Mixed Integer NonLinear Program Problem (MINLP) using two relaxations of a disjunctive set: Big-M and convex-hull. The results and experiences obtained with the problem solution are shown.

Model Constraints

Fixed size and time factors are used to model the multiproduct batch plant design constraints. The basic set of constraints used to model the size and processing time for batch units are:

$$V_j \geq S_{ij} \; B_i / G_j \tag{1}$$

$$TL_i \geq (T_{ij}^0 + T_{ij}^1 \; B_i) / M_j \tag{2}$$

where V_j is the size of stage j [m^3], Bi is the batch size for product i,[kg of product exiting the last stage] and S_{ij} is the size factor at stage j to produce 1 kg of final product i. TL_i is the processing time of product i, it must be the greatest processing time of the stages involved in the production of i (T_{ij}). T_{ij}^0 is a fixed amounts of time to process a batch of product i at stage j, while T_{ij}^1 permits to account the time demands proportional to the batch size of stage j. G_j is the number of parallel units operating in phase, and M_j is the number of parallel units operating out-of–phase. For semicontinuous units:

$$R_j \geq D_{ij} \; \frac{B_i}{\theta_j} \tag{3}$$

where R_j is the size of the semicontinuous item j, usually a processing rate capacity [m^3/h], but for some units, like filters, it corresponds to the size of the filtration area A [m^2]. In any case the sizes are proportional to the batch size B_i [kg] and inversely proportional to the operating time θ_j [h], through a so-called duty factor D_{ij}. In the case of composite stages with a semicontinuous item that processes the material hold in a batch item (e.g.: the case of an Homogenizer) we follow the modeling approach in Salomone et al (1994). The stage is described with equation (1) for the batch item size, but the batch processing time T_{ij} includes the operating time θ_j of the semicontinuous item, so replacing $\theta_{j\ \text{from}}$ equation (3) into equation (2) gives:

$$TL_i \geq (T_{ij}^0 + D_{ij} \; \frac{B_i}{R_j}) / M_j \tag{4}$$

A complete reference about the constraints formulated can be found in Montagna et al. (2000).
Finally the production requirement constraint takes into account the satisfaction of the product demand Q_i for the four products during the horizon time HT:

$$\sum_i Q_i \, TL_i / B_i \leq HT \tag{5}$$

The cost equation associated to the batch stages j is:

$$C_j = a_j G_j M_j V_j^{\alpha_j} \quad (6)$$

where C_j represents the cost of the j stage, a_j and α_j are cost coefficient for stage j, for the case of semicontinuous stages the cost function is the same but replacing V_j by R_j when corresponds.

Constraints 1, 2, 4, 5 and equation 6 are convexified using the following variable transformation:

$b_i = log\ B_i$; $g_j = log\ G_j$; $m_j = log\ M_j$
$v_j = log\ V_j$; $t\ell_i = log\ TL_i$; $r_j = log\ R_j$

Disjunctive Model

Since the synthesis problem for this protein production plant involves several discrete decisions, we have formulated a disjunctive model to pose those decisions. The model is based on the Generalized Disjunctive Programming (GDP) approach proposed by Lee and Grosmann (2000). The basic sets used in the disjunctive model are:

I = { A, B, C, D} Products
H ={H1, H2, H3, H4) Hosts
J = {1, 2, 3,…, 15} stages
d= {a,b} alternative batch units to process stage j

$$Min\ C = \sum_j a_j exp(m_j + n_j + \alpha_j v_j) \quad (7)$$

$$\sum_i Q_i . exp(b_i - t\ell_i) \leq HT \quad (8)$$

$$\bigvee_{h \in H_i} \begin{bmatrix} Y_{ih} \\ v1_j \geq log(S1_{ihj}) + b_i - g_j & , j \in SB1_{hj} \\ r_j \geq log(S2_{ihj}) + b_i - g_j & , j \in SB2_{hj} \\ t\ell_i \geq log(T^0_{ihj}) - m_j & , j \in ST^0{}_{hj} \\ v3_j \geq log(S3_{ihj}) + b_i - g_j & , j \in SB3_{hj} \\ t\ell_i \geq log(D_{ihj}) + b_i - r_j - m_j & , j \in SD_{hj} \end{bmatrix} i \in I \quad (9)$$

$$\bigvee_{h \in H_i} \begin{bmatrix} Y_{ih} \\ \bigvee_{d \in a,b} \begin{bmatrix} Z_{ihjd} \\ v1_j \geq log(SD1_{ihjd}) + b_i - g_j \\ t\ell_i \geq log(D1_{ihjd}) + b_i - r_j - m_j \\ v3_j \geq log(SD3_{ihjd}) + b_i - g_j \end{bmatrix} j \in SB4_{hj} \end{bmatrix} i \in I \quad (10)$$

v1 is the batch size for all batch stages and those semicontinuous stages with more than one unit that involving feed/holding/retentate tank as input, like stages 2.a, 2.b, 3.a, 3.b, 4.a, 4.b, 6, 9, 10, 11, 12, 13, 14 and 15. *v3* is the batch size for semicontinuous stages with more than one unit involving product tank as output, like stages 2.a, 2.b, 4.a, 4.b, 10, 12, 14 and 15. *SB1, SB2, SB3, SB4, ST^0 and SD* are subset of stages whose components depend on the host. Y_{ih} and Z_{ihjd} are Boolean variables to make the disjunctive term true or false.

Observe that (10) corresponds to an embedded disjunction, which is not the form of the GDP formulation of Lee and Grossmann (2000). A transformation proposed by Vecchietti and Grossmann (2000) is used in order to comply with that formulation, which turns (10) into:

$$\bigvee_{d \in a,b} \begin{bmatrix} Z_{ihjd} \\ v1_j \geq log(SD1_{ihjd}) + b_i - g_j \\ t\ell_i \geq log(D1_{ihjd}) + b_i - r_j - m_j \\ v3_j \geq log(SD3_{ihjd}) + b_i - g_j \end{bmatrix} j \in Sb5_{hj} \quad (11)$$

$$Y_{ih} \Leftrightarrow \sum_d Z_{ihjd} \quad (12)$$

The final GDP model is composed by (7), (8), (9), (11), and (12).

Model solution

The GDP model of the previous section has been transformed into Mixed Integer Nonlinear Programs (MINLP) using the convex-hull and Big-M relaxation of a disjunctive set. A characterization of both relaxations for a disjunctive set can be found in Vecchietti and Grossmann (2001). The two-transformed MINLP problems are solved with a modified version of DICOPT^{++}. MINOS and CPLEX are used as NLP and MILP solvers respectively. The equation (12) was written in the model as a set of integer inequalities representing the equivalence between the boolean variables.

Results

Table 2 shows the results obtained in the discrete decisions. The hosts selected for each product have been: H2 for product A, H3 for product B, H4 for product C and H3 for product D. The table also shows the units selected for batch stages 2, 3 and 4. Stage 1 has 3 units in parallel out-of phase and stage 13 has two units in parallel out-of-phase too, the rest of the stages do not have units in parallel. No units in parallel in phase exist in the plant. Table 3 shows the batch stage size and the operation time for each product. Table 4

compares both MINLP models generated from the GDP. From table 4 it can be seen that the number of equations and constraints for the convex hull is greater than the Big-M but it has a close relaxed optimal value compared with the final solution, as it is expected. The Big-M relaxation performs better than the convex-hull comparing the CPU time needed to reach the solution even when the relaxation gap is worst than the convex-hull case.

Table 2. Host and units selected

Product	Host selected	Stages Selected		
		2.a or 2.b	3.a or 3.b	4.a or 4.b
A	H2	2.b	N/A	N/A
B	H3	2.b	3.a	4.b
C	H4	2.b	N/A	N/A
D	H3	2.b	3.a	4.b

Table 3. Batch size and processing time for each product

Product	Batch size	Processing Time
A	6.67	8.0
B	6.67	15.0
C	0.44	56.0
D	4.46	15.0

Conclusions and future work

The synthesis model of the protein production batch plant presented has several challenges at the formulation level because of the hierarchical decisions of the problem: once the host is selected, for some stages must be determined which unit and number of them operating in parallel in-phase and out-of phase (equation 10). In this sense the main advantage obtained with the approach proposed is that the disjunctive model can be easily generated compared against the traditional mixed integer approach. The disjunctive formulation presents a pattern such that any modeler familiar with the approach can understand it. Besides, the MINLP programs corresponding to the disjunctive model are written in a systematic way by means of the convex hull or Big-M relaxations of a disjunctive set. Several solution algorithms can be then applied to solve the problem. This strategy presents more possibilities to tackle a problem containing discrete decisions than traditional mixed-integer approaches.

From the results obtained for the synthesis of the protein production plant, the convex hull relaxation has a better relaxation gap, but the CPU time to reach the solution is lower in the Big-M relaxation, so no conclusion can be extracted about which relaxation is better.

We plan to work in the future with this model by using different cost functions, variable size and time factors and the inclusion of process information into the batch design.

Acknowledgments

The authors would like to acknowledge financial support received from VITAE Foundation within the Cooperation Program Argentina-Brazil-Chile and the valuable information provided by Dr. Juan A. Asenjo from Chile.

References

Alvarez E. V., Lienqueo M. E. and Pinto J.M (2001). Optimal Synthesis of Multistep Protein Purification Processes. *Latin American Applied Researc*h, **31**

Asenjo J.A., Montagna J.M., Vecchietti A., Iribarren O. and Pinto J.M (2000). Strategies for the simultaneous optimization of the structure ans the process variables of a protein production plant. *Comp. & Chem. Engng*, vol. **24**, 2277.

Lee, S. & Grossmann I. E. (2000). New algorithms for nonlinear generalized disjunctive programming. *Comp. & Chem.Engng*, **24,** 2125.

Montagna, J.; Vecchietti, A.; Iribarren, O.; Pinto J. y Asenjo J (2000). Optimal Design of Protein Production Plants with Time and Size Factor Process Models. *Biotechnology Progress*, **16**, 228.

Steffens M.A., Fraga E. S. and Bogle D.L. (2000). Synthesis of Bioprocesses Using Physical Properties Data. *Biotech. and BioEngng*., **68**, 2, 218.

Vecchietti, A. y Grossmann, I. E. (2000) Modeling Issues and Implementation of Language for Disjunctive Programming, *Comp. & Chem. Engng*., **24**, 2143.

Vecchietti, A. y Grossmann, I. E. (2001). Characterizations and formulations of disjunctions and their relaxations. Proceedings of Enpromer'2001, Santa Fe, Argentina,**I** ,409.

Table 4. Comparison between MINLP models solved (BigM vs. Convex Hull)

Model	Equations	Variables	Discrete Variables	Optimal Solution	Relaxed Solution	Iterations	CPU time (sec.)
Big-M	530	475	194	3,531,944	477,439	5 major	5
Convex Hull	2301	1360	194	3,531,944	3,087,154	4 major	20