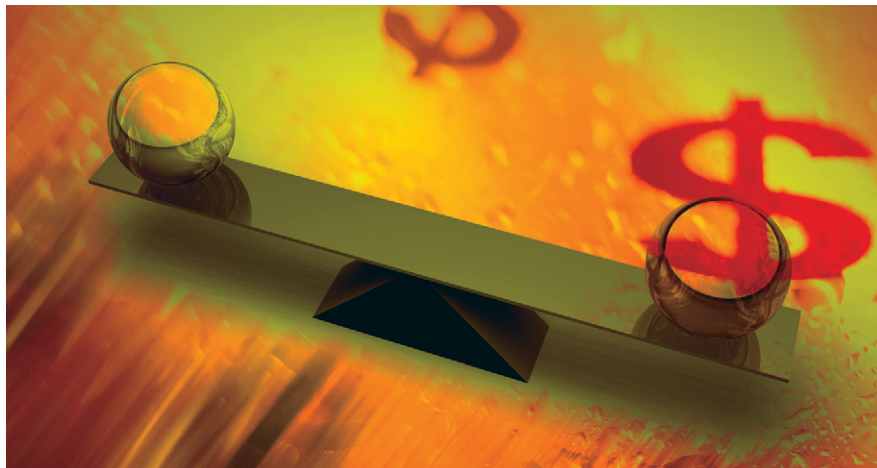


# Process Development's Impact on Cost of Goods Manufactured (COGM)

Chae Han, Philip Nelson, and Amos Tsai

**M**anufacturing throughput (the amount of material a plant can produce per year) is affected by process yield and plant run rate. The higher they are, the more a plant can produce per year, requiring fewer lots to meet annual demand. Although a process development team obviously determines the process yield, the team also determines the impact on the run rate of duration and potential implementation complexity of the entire train of unit operations. Thus, an optimized process maximizes plant throughput by maximizing process yield *and* run rate while assuring product quality and minimizing the cost-per-unit mass of manufactured drug (1). Based on how COGM is defined, we limit ourselves here to considering the impact of plant time and raw material cost. We believe run rate, in addition to yield, should be considered simultaneously during process development to arrive at a process design desirable from a COGM perspective.

Plant cycle time (the time between the start of two consecutive batches) can



KOSTYA KISLEYKO (WWW.SXC.HU) AND WWW.PHOTOS.COM

be affected by process and operating parameters that are often coupled. The longest unit operation tends to define a plant's cycle time. For instance, if a cell culture process takes 10 days to complete and two days to turn around, a plant with a single bioreactor would have a 12-day cycle time. For  $n$  reactors furnished, the theoretical minimum cycle time will decrease to  $12/n$  days until the plant is limited by the next unit operation with a cycle time longer than  $12/n$  days. Operating parameters that can affect plant cycle time are often related to shared use of supporting equipment such as clean-in-place (CIP) skids, media, buffer, and in-process pool hold tanks. The emergence of high-titer monoclonal antibody processes also necessitates using chromatography columns in multiple cycles accompanied by increasing buffer consumption, thus imposing additional operating challenges to existing plants (2, 3).

A good understanding of the interdependence between process requirements and facility capability during development is critical to

maximizing run rate and minimizing COGM in full production. This can be achieved by static or dynamic plant modeling that reflects plant operation and process requirements to a sufficient degree of accuracy.

## CASE 1: RAW MATERIALS AND UNIT OPERATION AFFECT RUN RATE

In general, raw material cost contributes to a higher percentage of COGM as run rate increases. Efforts to reduce that cost, however, must be accompanied by analysis of the entire process to prevent negative impact on run rate (4). This case study shows how selecting a raw material and unit operation with higher unit cost can be justified if the associated run rate is favorable and a comparable step yield can be maintained. Here, the projected annual demand for a drug substance (DS) was up to one metric ton, and one purification step was critical in achieving high yield and throughput for the entire process. However, potentially high raw material cost was associated with that step. After preliminary screening, two options were based on two different

**PRODUCT FOCUS:** PROTEINS, ANTIBODIES, PARENTERAL PRODUCTS

**PROCESS FOCUS:** UPSTREAM AND DOWNSTREAM PROCESS DESIGN

**WHO SHOULD READ:** PROCESS DEVELOPMENT AND MANUFACTURING, FACILITY-DESIGN ENGINEERS

**KEYWORDS:** MANUFACTURING THROUGHPUT, STATIC MODELING, COGM, RAW MATERIALS, RUN RATE, UNIT OPERATIONS, CHROMATOGRAPHY

**LEVEL:** INTERMEDIATE

purification principles: raw material A (RM A) and raw material B (RM B). The estimated unit cost of RM B was as much as nine times that of RM A. In addition, the mass-based processing capacity of RM A was superior to RM B by a factor of three. Therefore, RM A was a clear winner based on unit raw material cost and mass-based processing capacity. However, the projected operating flow rate for RM B was three times faster than that of RM A. RM B can potentially decrease the process time and overall plant time that directly affect overall COGM. Plant modeling was performed to calculate the COGM difference between the RM choices (Table 1).

Table 1 shows that RM A requires fewer cycles than RM B to process an entire batch of DS because of its higher mass-based processing capacity (~3×). However, the actual time required to process an entire batch would almost double if RM A is selected. This difference comes from the high flow rate capability with RM B. Not only does that speed up batch time, but it also shortens the overall plant time required to meet annual demand by ~13 weeks (Table 1). This provides significant annual COGM savings and frees up capacity to be used by other products in a multiproduct licensed facility.

Another factor to consider in a COGM analysis was annual raw material cost. Here, the campaign is defined as total time needed to produce enough lots to meet annual demand. As stated earlier, the basic unit cost of RM B (\$/L) is almost 9× higher than that of RM A. However, the total cost difference between them was only a factor of two instead of nine after accounting for annual requirement. That was possible because RM B has a lifetime of 15 lots compared with RM A's 6.6 lots. Furthermore, after accounting for both total plant time and total raw material cost, the analysis showed that overall COGM is ~30% less using RM B as the raw material of choice for the process, translating into significant annual savings.

This case study demonstrates that minimizing raw material cost alone does not always optimize COGM. Instead, the impact of a raw material on COGM

factors, in this case the plant time cost, should be investigated and demonstrated to reach the cost optimum of a process for successful commercialization.

## CASE 2: MEETING PROJECTED DEMAND WITH FIXED THROUGHPUT

In this case study, the challenge lies in designing a plant that can produce up to one metric ton of a DS at the peak of its commercial life cycle. Flexibility for responding to a postlaunch drop in demand also must be designed in because the current demand projection was highly speculative. In addition, ongoing process development could potentially change the parameters (step capacity, duration, and yield) upon which the plant design is based by ±10%.

These uncertainties required thoroughly understanding the interdependence of facility design options, process duration, and operating strategies and their impact on meeting the range of required plant throughput. To predict design and process impact on annual output, plant modeling that incorporated operating and process parameters was used to account for their dependency (5, 6). This plant needed to accommodate a series of unit operations comprising bioproduction, harvest, biochemical reaction (BR × *n*), and purification. Table 2 captures the most updated information in step duration and yield of the DS for each unit operation at the time of conceptual design, with the BR × *n* operation as the rate limiting step. Assuming no duplicate equipment, the maximum theoretical run rate was estimated to be 120 lots per year if the plant operated 360 days a year in 24/7 shifts.

To better understand how process parameters and facility options affect COGM in the face of uncertain demand, the BR × *n* operation became the design focus because it determined the achievable run rate. At current starting concentration and an overall process yield of 11.3%, the entire process could produce ~0.2 kg of DS per 1,000 L of BR × *n* volume. With the peak demand projected to be one metric ton per year, a 42,000-L reaction vessel, or 1,000/(120 × 0.0002) L was required. Because of uncertainty around this projected peak demand, the design team

**Table 1:** Process performance comparison of RM A and RM B

	Cycles*	RM Reuse Rate	Plant Time Needed to Meet Annual Demand
RM A	3 cycles (3.32 day)	6.6 lots	34 weeks
RM B	10 cycles (1.96 day)	15 lots	21 weeks

\* Number of cycles (or days) required per batch to process the entire volume expected for this purification step

**Table 2:** Most updated process information to be used at the time of conceptual design

Unit Operation	Step Yield (%)	Duration (days)
Bioproduction and harvest	100	2–3
Biochemical reaction	30	3
Purification 1	70	1.5
Purification 2	70	1.5
Purification 3	90	1.5
UF/DF	90	1.0
Final bulk	95	1.2
<b>Overall Yield</b>	<b>11.3</b>	

**Table 3:** Comparing two different designs

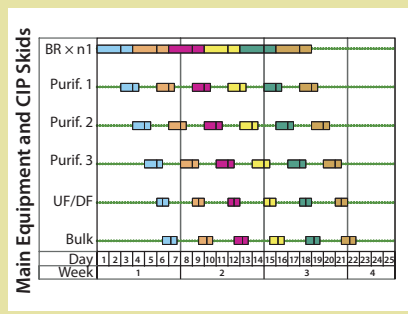
	1 x 42,000 L	4 x 11,000 L
Annual run rate	120 lots	119 lots
Minimal working volume	44,000 L	11,000 L
Size of Purification 1 equipment	2×	1×

**Table 4:** Step duration and run rate for the current process before process improvement

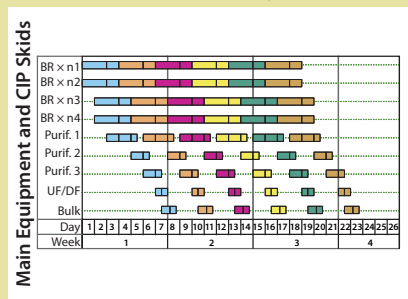
Best Case Purification	Run Rate	Days/Lot
Unit Op 1		2.0
Tank turn around		2.0
Unit Op 2		1.0
Tank turn around		0.5
Unit Op 3		0.5
Unit Op 4		0.5
Unit Op 5		0.5
Unit Op 6		0.5
Bulk filtration		0.5
	<b>Total</b>	<b>8.0</b>

immediately considered it to be impractical to fabricate a single 42,000-L tank that would be likely to produce excess BR × *n* capacity. Instead, the design team investigated using multiple tanks. They explored different options with the process development team, and

**Figure 1:** Equipment occupancy project for six consecutive lots of DS for a plant furnished with a single 40,000-L reactor; annual output = 120 lots



**Figure 2:** Equipment occupancy project for six consecutive lots of DS for a plant furnished with four 11,000L reactor; annual output = 119 lots



chose a plant outfitted with four 11,000-L tanks. That enabled the operation to reach a quarter of the peak commercial demand up to that peak demand without retrofitting, thus resulting in a flexible plant. The teams explored the implications of design options (layouts) on process requirements and operating strategy, as discussed below.

Figure 1 shows how a facility with a single 42,000-L BR × *n* tank can be operated to produce one metric ton of DS annually. In this scenario, the plant produces a lot of DS every three days (120 lots/year). Figure 2 shows how an alternative plant furnished with four 11,000-L BR × *n* tanks can meet the same annual demand. In this scenario, two tanks are used simultaneously within each lot to process half the starting material. At the end of the first BR × *n*, the material is purified immediately. In the meantime, the remaining two 11,000-L BR × *n* tanks would start with a 1.5 day delay, enabling immediate processing of the second set of BR × *n* material by the same purification equipment per lot without intermediate cleaning.

Our scheduling tool indicates that the two plant configurations result in

comparable run rates (Table 3). However, the 4 × 11,000-L design offers the flexibility to respond to actual demand that could be much lower than what is projected by operating a subset of the four tanks. In addition, the equipment sized for the first purification step (Purification 1) is only half the size of that needed in the 1 × 42,000-L design because a batch in this case is split into two cycles.

Before this production strategy could be implemented, data needed to support holding the Purification 1 pool from the first cycle for 1.5 days or longer while maintaining product quality attributes and bioburden control. If such an in-process hold required cold room storage or 0.2-μm sterile filtration, the current plant design, process yield, and operating strategy could all be affected — which would need to be brought to the attention of the design team as soon as data became available.

This case study demonstrated that when the interdependence between process and operating parameters is fully accounted for in a model, it is possible to design a plant to operate at maximum run rate with flexibility that can respond to abrupt demand change by a factor of four.

### CASE 3: OVERCOMING FACILITY LIMITATIONS GUIDED BY COGM

This case study demonstrates how facility design must be taken into account in determining which process improvements will have the most positive impact on COGM. The facility under consideration was originally designed for late-stage clinical production. That called for two independent cell culture and purification trains with three midsize reactors in each train as the maximum production scale. Late in the design phase, two commercial-scale cell culture reactors and harvest vessels (one in each train) were added to enable the facility to support commercialization. Because the purification area was originally sized for clinical instead of commercial-scale production, a number of process changes and facility modifications were needed to minimize the effect of scale misalignment on plant capacity, flexibility, and COGM.

The historical design created a bottleneck in the downstream buffer storage area, largely due to the need for two or even three buffer hold tanks for a single buffer. This practice tied up all of the buffer hold capacity within a unit operation and rendered staggered operation impossible. In addition, the start of the next unit operation was delayed because of the time needed to turn around the same set of tanks used in the previous step. So the increased batch duration and cycle time had a negative impact on the facility's achievable run rate.

Table 4 shows the fastest potential purification run rate without process improvements. At this rate the purification area would be unable to keep up with the commercial-scale production bioreactors. Although purification time was lessened by multiple cycles of undersized chromatography columns, the 2.5 days spent with protein processing idle while buffer storage tanks were turned around was viewed as the highest priority for process improvement. Process changes that could alleviate the bottleneck could increase plant capacity and reduce labor. Four improvements were identified with the potential to address this problem and thus improve plant capacity and process economics.

**In-Line Dilution:** In-line dilution has been developed as a process improvement to debottleneck the current purification process. In-line dilution eliminates the 2.5 days/lot lost to tank turnaround. In addition, it reduces the amount of labor needed by 25% and the amount of WFI/run by almost 20%, we estimate. The labor savings come from eliminating buffer preparation activities, and the WFI savings come from eliminating rinse cycles. With two commercial-scale cell culture reactors feeding the purification train, one harvest can be performed every seven days. In-line dilution would drop the overall run rate from one run every eight days to one every seven days. This somewhat modest improvement in plant productivity (an increase of 12.5%) has significant impact on COGM because fixed costs are the greatest single factor affecting it.

Assuming a hypothetical fixed cost of \$200 million/year for the facility, the

annual run rate increase from 38 to 44 lots/year (including time for an annual shutdown) decreases the cost/run from \$5.2 million to \$4.5 million. Additional fixed annual cost associated with sampling and testing the WFI system is \$776,150, and the system's annual depreciation is \$245,565. By comparison, the variable cost saving in lower WFI demand is <\$224,000. Thus, even at the maximum run rate, fixed WFI costs will always be substantially greater than the variable costs. The greatest reduction in COGM from in-line dilution is the increase in plant capacity by debottlenecking the buffer storage area and reducing WFI consumption.

#### Column Chromatography Capacity:

Using higher dynamic binding capacities for column chromatography steps can potentially reduce COGM through a number of mechanisms. Savings can be had through increased run rate and reductions in labor and WFI, all of which are discussed above. In addition the increase in capacity use can cut raw materials costs with a corresponding direct reduction in COGM. Using the same example a binding capacity increase for Chromatography Op 1 from 20 to

30 g/L would save 33% of the buffer components. Run time for the step would drop by 18%, and resin cost could be reduced by 33%. In this example the resin cost is \$3.50/g of MAb produced at 20-g/L capacity use. That drops by ~\$1.2/g through increased capacity use to 30 g/L, which saves ~\$19,200/run for runs producing 16 kg of product. Once again, the magnitude of change in this facility is magnified by debottlenecking the process by a 33% reduction in the number of buffer preparations and storage tanks required and demand on the CIP system.

**Chromatography Column Sizing:** To minimize the likelihood of downstream processing becoming the rate-determining step that limits plant capacity, chromatography columns must be appropriately sized. What makes column scale-up interesting to a COGM analysis is that, in many respects, larger chromatography operations are much more costly than cycling smaller columns. An optimal size must be neither too large nor too small to handle

**Table 5:** Effect of chromatography column on its cost and operational parameters

Column Diameter	0.6m	1.2m	1.4m	2.0m	2.4m
Column Volume	84.8L	339 L	462 L	942 L	1,356 L
Resin cost	\$676K	\$2.7mil	\$3.7mil	\$7.5mil	\$10.8mil
Flow rate (L/min)	14.1	56.5	76.9	157	226
Cycles	15	4	3	2	1
Run time (hours)	34.35	9.43	7.2	4.57	2.88
Equilibration/wash (L)	8,309	8,309	8,309	8,309	8,309
Elution (L)	5,424	5,424	5,424	5,424	5,424
Strip (L)	4,068	4,068	4,068	4,068	4,068
Regeneration (L)	339	1,356	1,846	3,768	5,424
Storage (L)	254	1,017	1,385	2,826	4,068
Packing interval (at 40 runs/year)	122 days	1.25 years	1.67 years	2.5 years	5 years

expected changes in development and manufacturing scales. Table 5 shows the effect of column scale on several parameters. Scaling up from a 0.6-m to a 1.2-m column has a very significant impact on the time it takes to run this step, reducing it from 34 to nine hours. If this step is rate limiting, the scale-up benefit is likely to outweigh the increased cost of a larger column, a larger skid, an increased use of regeneration and storage buffer, and a larger up-front resin cost.

However, additional scale up yields much smaller time savings and much higher costs associated with very large systems and much greater buffer volumes. It is very unlikely that a column >1.4 m would be cost effective in this situation. It is likely, however, that a column <1.2 m would be most economical during clinical production. For a few runs of a product that may not gain approval, a long step-run time would be more cost effective than a larger-scale operation. So strategic decisions regarding likelihood of approval, anticipated demand, and the ability to make process changes later may override cost considerations.

**Membrane Chromatography:** In a facility constrained by buffer tank and CIP availability, use of membrane chromatography in flow-through mode may greatly reduce buffer consumption and is thus presents another debottlenecking option.

In each case the product was very pure before this step, which removes trace contaminants and provides additional log removal of viruses. Conventional chromatography columns of sufficient size to handle the product

feed volume in a reasonable amount of time would have offered a huge excess of unused capacity. Filters used to replace these columns offer faster kinetics within their membrane pores than do chromatography media.

Use of membrane chromatography at this step reduced buffer consumption by >10x. This technology provides one more tool with which to address facility throughput constraints caused by limited buffer tank availability. As stated, increasing plant capacity will have the greatest potential for lowering COGM, and addressing bottlenecks in the process will have the greatest impact on increasing plant capacity.


#### A HOLISTIC APPROACH

We have highlighted the impact of process development on manufacturing throughput and COGM in three different scenarios. In the first case, we demonstrated that lowest raw material cost did not necessarily lead to lowest COGM. In fact, when manufacturing run rate was accounted for, the more costly raw material and its associated unit operation were preferable. In the second case study, plant modeling was used to guide simultaneous process development and new facility design, resulting in a facility flexible enough to meet volatile projected demand at maximum run rate to minimize COGM. In the third case, COGM was used as a guide to determine multiple retrofitting options for a fully commercialized facility.

Based on our observations, coupled with the increasing cost pressure on biopharmaceuticals, we recommend that plant modeling to reasonably estimate

the manufacturing run rate become an integral part of early — and particularly late — process development. This holistic approach enables companies to maximize the likelihood of commercial success of their costly development programs.

## REFERENCES

- 1 Chadd HE, Chamow SC. Therapeutic Antibody Expression Technology. *Curr. Opin. Biotechnol.* 12(2), 2001: 188–194.
- 2 Sommerfeld S, Strube J, J. Challenges in Biotechnology Production: Generic Processes and Process Optimization for Monoclonal Antibodies. *Chem Eng Proc.* 44(10) 2005: 1123–1137.
- 3 Dowd C, van Reis R. Cost Reduction Strategies in Recovery Process Design. *Proceedings of IBC's 4th International Conference on Production and Economics of Biopharmaceuticals*; 14–15 November 2001, San Diego, CA.
- 4 SP Fulton, et al. Large-Scale Processing and High-Throughput Perfusion Chromatography. *Bio/Technology* 10(6) 1992: 635–639.
- 5 RG Harrison, et al. *Bioseparations Science and Engineering*. Oxford University Press, New York, NY, 2003; 319–371.
- 6 Petrides DP, Siletti CA. The Role of Process Simulation and Scheduling Tools in the Development and Manufacturing of Biopharmaceuticals. *Proceedings of the 2004 Winter Simulation Conference*, Ingalls R, et al. (Eds.); [www.inform-sim.org/wsc04papers/274.pdf](http://www.inform-sim.org/wsc04papers/274.pdf) (accessed 1/22/10). 

**Chae Han** is a principal scientist in global process engineering at Amgen, One Amgen Center Drive, Thousand Oaks, CA 91320; **Philip Nelson** is principal engineer of a process engineering group for MAb production at Amgen, 6701 Kaiser Drive, Fremont, CA 94555; and corresponding author **Amos Tsai** is a principal scientist at Amgen's Thousand Oaks facility, 1-805-447-9847; [almost@amgen.com](mailto:almost@amgen.com).