Aldevron accelerates growth using operations research in biomanufacturing

Submitted to *Interfaces*
manuscript

Aldevron Accelerates Growth Using Operations Research in Biomanufacturing

(Authors' names blinded for peer review)

In the biomanufacturing industry, production and planning decisions are often challenging owing to batch-to-batch variability and uncertainty in the production yield, quality, cost, and lead times. To improve biomanufacturing efficiency, a multidisciplinary team of researchers collaborated over five years to develop a portfolio of decision support tools. The developed tools provide a data-driven, operations research-based approach to reduce biomanufacturing costs and lead times. These decision support tools comprise multiple deterministic and stochastic optimization models to optimize production and planning decisions. To optimize production decisions related to fermentation and protein purification, optimization tools were developed to provide a decision support mechanism that links the underlying biological and chemical processes with business risks and financial trade-offs. To optimize planning decisions, interactive scheduling and capacity planning tools were developed to enable efficient use of the expensive and limited resources. Although developed in collaboration with Aldevron, these tools address common industry challenges, and they have been shared with a wider industry community through working group sessions.

Key words: Biomanufacturing, new drug development, engineered proteins, Markov decision processes, scheduling, cost and lead time reduction, innovative applications of OR.

More than 325 million patients worldwide have benefited from next generation drugs (e.g., recombinant proteins and monoclonal antibodies) to treat various types of cancer, diabetes, cardiovascular diseases, and other health issues (The European Biopharmaceutical Enterprises 2015). These drugs are produced using *biomanufacturing* technologies. The biomanufacturing industry is growing rapidly and becoming one of the key drivers of personalized medicine and life sciences. As such, these drugs constitute up to 70% of the global pharmaceutical research and development pipeline (Pharmaceutical Research and Manufacturers of America 2016), and the global biopharmaceutical market is projected to reach $291 billion by 2021 (Mordor Intelligence 2017).

Despite its success, biomanufacturing is a challenging environment. It is labor and cost intensive, and it involves high risk of failures. In contrast to conventional pharmaceutical manufacturing, where medicines are chemically synthesized, biomanufacturing methods use

living systems (i.e., bacteria, virus, or animal cells) to produce these drugs. The use of live systems introduces several manufacturing challenges, such as batch-to-batch variability in terms of quality, production yield, and processing time. Owing to these challenges, the cost of developing a new biopharmaceutical drug can reach up to $1.2 billion, with an average lead time of 10 years (Long and Works 2013).

To hedge against manufacturing challenges and risks, most large pharmaceutical companies work with small and medium-sized enterprises (SMEs) to subcontract various phases of their biomanufacturing research and development operations. Market analysis shows that 70% of biotechnology companies are SMEs (The European Biopharmaceutical Enterprises 2015). There are two main reasons for this. First, many of these SMEs are spinoffs/startups from research labs that have unique expertise in the therapy being developed. They are therefore uniquely positioned to conduct such research and development. Second, subcontracting the work to these SMEs allows large pharma companies to pursue multiple lines of research in parallel, increasing the probability of eventual success while transferring much of the failure risk to individual SMEs. Therefore, for these SMEs, reducing costs and lead times is vital to their business profitability, but their success has a cascading effect on the rest of the supply chain involved in the drug development process. As Tom Foti, the Vice President of Aldevron, one such SME, states "We are producing 50 liter cultures here, but our clients [large pharmaceutical companies] are dealing with $5,000$ liter cultures. If we can build optimization models here and demonstrate the feasibility of how it works, our clients could also do that."

However, the application of operations research (OR) methodologies to the biomanufacturing industry is in its infancy. One of the main reasons is that, to date, the competitive advantage in biomanufacturing has been driven by the scientific advances related to the underlying biological and chemical processes. With rapid industry growth and growing competition, there is an increasing need for a data-driven, OR-based decision-making mechanisms to improve manufacturing efficiency and to reduce costs and timelines.

This paper describes the development of optimization models and decision support tools to reduce costs and lead times in the manufacturing of custom engineered proteins at Aldevron. Aldevron is a provider of plasmid DNA, proteins, enzymes, antibodies, and other biologicals throughout a variety of life science applications at every stage of new drug development and manufacturing. Although this paper focuses on the OR applications at

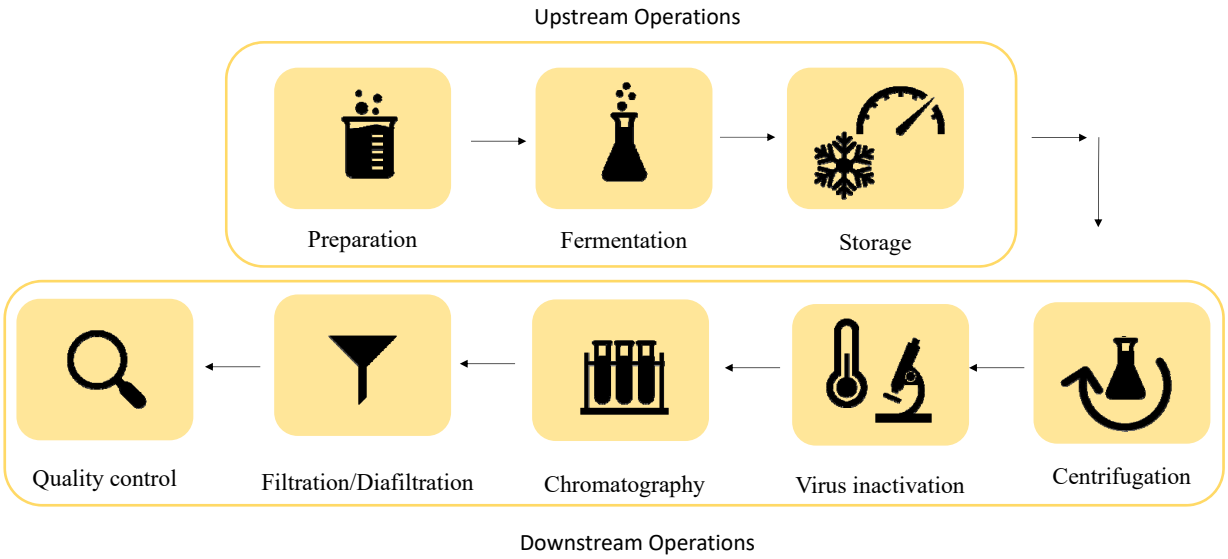


Figure 1 **Protein manufacturing process is classified into two main steps, upstream and downstream operations.**

Aldevron, the true impact of this work extends to similar SMEs in biomanufacturing and other industries. This research has been shared with the broader biomanufacturing industry through working group sessions (BioWGS 2016), and the research outcomes have been recognized by government agencies, media, and professional societies, i.e., the Wisconsin Economic Development Corporation (WEDC 2014), Xconomy (Engel 2014), BioForward (Foti et al. 2016), the Institute for Operations Research and the Management Sciences (INFORMS 2017, 2016), and the Production and Operations Management Society (Martagan et al. 2016a).

Overview of Protein Manufacturing Operations

Figure 1 presents a high-level process map of protein manufacturing operations. The production process can be broadly classified into two main steps, upstream and downstream operations. Upstream operations include culture preparation, fermentation, and storage. Downstream operations comprise a series of purification processes and quality control.

Upstream Operations: The production process starts with preparing raw materials (i.e., plasmid preparation and cell re-engineering). For example, foreign DNA is often introduced into insect or mammalian cells to make them express the protein of interest. Once raw materials are ready, fermentation is carried out in flasks or stainless steel vessels, where the cell culture grows and produces the protein of interest (which is also called the "target protein"). Fermentation is a highly controlled process where physical and chemical

parameters (e.g., temperature, pH, oxygen transfer rate) are closely monitored through online/offline controls. Each production order has a unique manufacturing protocol that provides a "recipe" for the fermentation process. This includes specific instructions on how to conduct the process (e.g., adding cells, adding fresh media, feeding cells, adjusting the process parameters, making corrective actions, setting specification limits for critical parameters). The outcome of the fermentation process is a batch mixture that consists of a limited amount of target protein mixed with numerous unwanted impurities (e.g., host cells, byproducts such as ammonia).

Downstream Operations: The main objective of downstream operations is to eliminate all unwanted impurities from a given batch. Downstream operations are also called protein purification operations, and they often consist of centrifugation (i.e., separating impurities through spinning), virus inactivation and lysis (i.e., disintegrating a cell by rupturing the cell wall or membrane), chromatography (i.e., separation by exploiting physical and chemical properties), and ultrafiltration/diafiltration (i.e., a dilution process) followed by the final quality control. Among these operations, chromatography is critical and is one of the focuses of this paper.

Most often, each client order has a *yield requirement* (i.e., the minimum amount of target protein) and *purity requirement* (i.e., the minimum acceptable quality, where *purity* is a measure of batch quality based on the protein and impurity amounts present in a batch). If the final product does not satisfy these production requirements, the biomanufacturer incurs large penalties. For example, clients often do not purchase the final batch if it does not comply with the purity requirement. Other penalties can be associated with the cost of disappointing the clients, changes in the planned manufacturing lead time, the impact on potential future orders, and yield penalty costs per each unit of protein in short. The degree of success and failure often depends on how well the upstream fermentation and downstream chromatography operations were conducted.

In the paragraphs below, we provide a high-level description of the fermentation and chromatography operations and highlight the challenges encountered in practice. This description is intentionally kept at a high level, without going into the specifics of the biology/chemistry and the associated challenges. While there are several scientific publications that elaborate on these challenges (e.g., Ngiam et al. (2003), Gnoth et al. (2007), McNeil and Harvey (2008), GE Healthcare (2010)), our goal here is to highlight the manufacturing

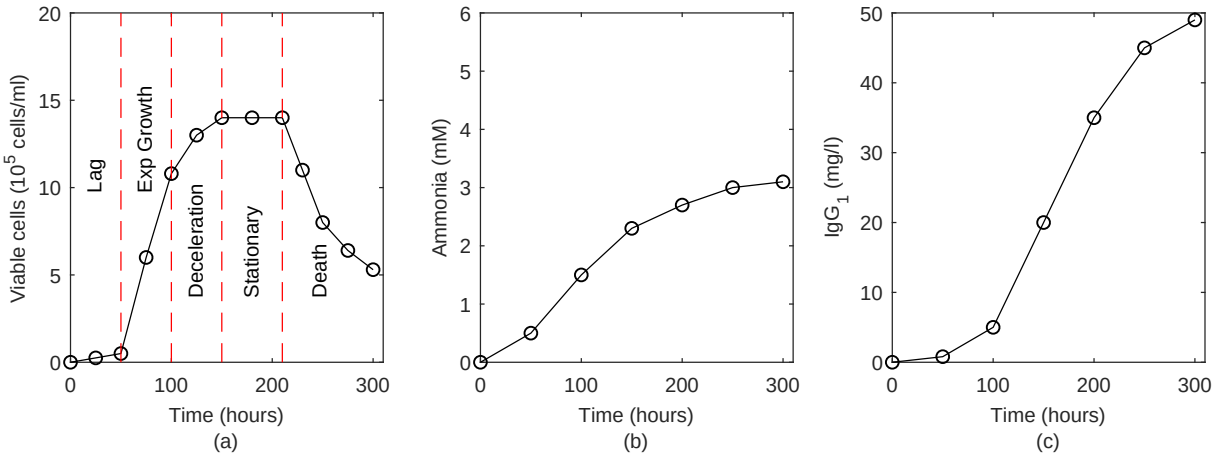


Figure 2 **The cell culture goes through several metabolic phases during fermentation. Target protein (IgG_1) and unwanted impurities (ammonia) accumulate at a non-stationary rate due to these metabolic phases.**

and planning challenges (i.e., those relevant to OR) that are layered on top of the other scientific challenges.

Fermentation Operations

During the fermentation process, the cell culture goes through several metabolic phases, as shown in Figure 2a. First, the cells adjust to their new environment (lag phase). Next, their growth rate steadily increases for a period of time (exponential growth phase), after which it slows down (deceleration phase) and reaches a steady state (stationary phase) followed by cell death (death phase). Owing to these metabolic phases, the target protein and unwanted impurities accumulate at a non-stationary rate. For example, Figure 2b and 2c plot the expected amount of ammonia (unwanted impurity) and IgG_1 (the final product of interest) obtained over time. We see that IgG_1 and ammonia accumulate simultaneously during fermentation. This phenomenon is referred to as the *purity–yield trade-off* in the fermentation process, and it presents an important challenge. Waiting "too long" to stop the process, in anticipation of a higher yield of the target protein, may result in a higher amount of impurities. In turn, this could increase the difficulty of the downstream purification operations.

In addition, the fermentation process is subject to randomness in the process outcome. Although fermentation is highly controlled, the process relies on live cells, such as viruses or bacteria, and this leads to variability in terms of yield (amount of target protein), quality (amount of unwanted impurities), processing time, and operating costs. In addition,

the fermentation process can also be subject to multiple failure processes through random shocks. A random shock could be attributable to machine breakdowns, human error, or the underlying biological and chemical dynamics (e.g., cell mutation). In practice, a random shock is detected through an abrupt change in one or more of the critical process parameters. For example, a random shock could lead to a sharp, unexpected drop in the viable cell concentration. If a random shock occurs, the operator adopts a corrective action prescribed in the manufacturing protocol. However, these random shocks still add another layer of challenge in practice because they can lead to batch failures.

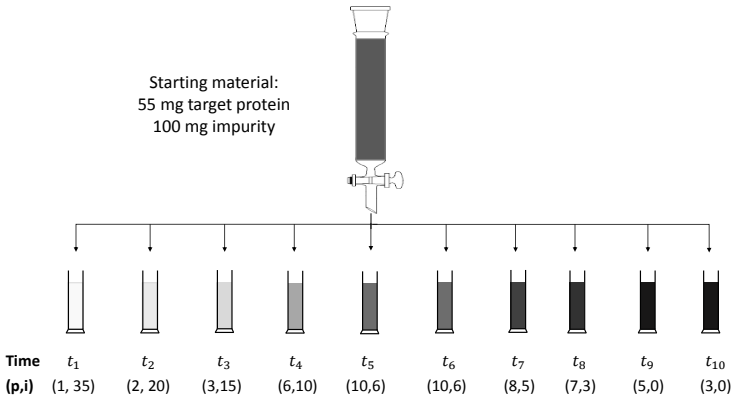
In this work, we focus on two types of batch failure: sudden failure and progressive failure. *Sudden failure* represents a catastrophic failure caused by the underlying biological and chemical dynamics. Examples of sudden failure are bacterial contamination and cell mutation. On the other hand, *progressive failure* happens if the amount of impurities present in the batch exceeds a predetermined specification limit imposed by regulations. As part of natural cell growth, it is expected that impurities will accumulate over time (see Figure 2b). However, a random shock can accelerate the rate at which unwanted impurities accumulate and, in turn, can increase the risk of progressive failure. If a batch fails because of sudden or progressive failure, it is completely discarded.

In this setting, it is important to determine the best harvesting policies (stopping time) within the guidelines of the approved manufacturing protocol. There is an important need for a formal decision-making framework that can simultaneously account for the failure risks, purity–yield trade-offs, and relevant economic parameters. Current practice either ignores such key manufacturing trade-offs or adopts conservative suboptimal policies to hedge against financial risks. Both strategies increase costs and lead times.

Chromatography Operations

Chromatography is a separation technique that relies on the difference in the physical and chemical characteristics between the target protein and impurities. For example, anion-exchange chromatography separates molecules based on their charges, while hydrophobic interaction chromatography separates molecules based on hydrophobicity. Chromatographic separation is carried out in columns packed with special resins binding to either the target protein or impurities. The scientist collects the material flowing through the chromatography column at various time intervals (e.g., one minute) as illustrated in Figure 3a.

(a) Dynamics of chromatography



(b) Example of chromatography data

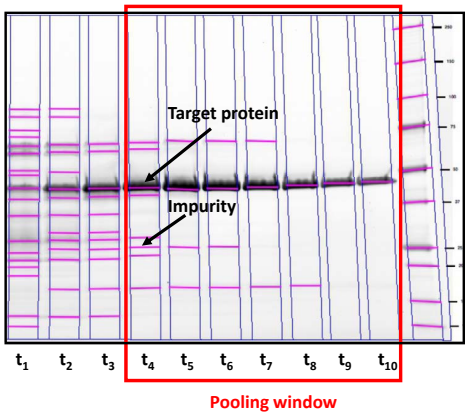


Figure 3 **Collected samples are labeled as (p,i) in Figure 3a where p is the amount of target protein (mg) and i is the amount of impurity (mg) in the sample. Chromatography output is analyzed using gel electrophoresis to determine the amount of target protein and impurities in the collected samples as given in Figure 3b.**

Starting material in the column contains target protein (represented as black) along with many impurities (represented as white). After the completion of chromatography, the scientist uses techniques such as gel electrophoresis to determine the amount of target protein and impurities that flow through the column during each time interval. (In practice, there are several different techniques for collecting and analyzing chromatography data, and gel electrophoresis is only one of them.) Figure 3b shows the output of gel electrophoresis for the collected samples given in Figure 3a. In this figure, each column on the x-axis represents a *lane* and corresponds to the volume flowing through the chromatography equipment during a specified time interval of the process. The material collected in each lane contains a fraction of the target protein and varying amounts of other impurities, as shown by the arrows in Figure 3b. Since gel electrophoresis sorts molecules based on size, the y-axis in Figure 3b describes the composition of the volume collected as a function of the molecular size of the protein and impurity constituents. The size of the black pixels in Figure 3b is correlated with the amount of the target protein present in a lane.

Based on the scouting data, the scientist decides to collect material flowing through only certain lanes. We refer this problem as the *pooling window selection problem*. A pooling window represents a selection of consecutive lanes to be collected during the chromatography process. Figure 3b illustrates one potential pooling window that collects lanes t_4 to t_{10}. In practice, selecting the right pooling window could be challenging owing to several factors:

- *Purity–yield trade-off:* A purity–yield trade-off is often inevitable in chromatography, implying that the scientist stands to lose some fraction of the target protein at each chromatography step to improve the batch purity. For example, if she pools lanes $t_4 - t_{10}$ in Figure 3b, then she collects a large portion of the target protein and impurities. However, if she pools lanes $t_8 - t_{10}$, then she collects lesser of amount of target protein and impurity. This implies that a smaller (larger) pooling window might help to meet the purity (yield) requirement, but there is a risk of not meeting the yield (purity) requirement.
- *Multiple dependent steps:* Protein purification often requires multiple chromatography steps in series. Therefore, suboptimal decisions made at an earlier chromatography step could have a limiting effect on the best possible outcome of subsequent steps.
- *Randomness:* Because of the underlying biological and chemical dynamics, the amount of target protein and impurities obtained at a chromatography step involves randomness.
- *Engineered proteins:* Each project represents an engineer-to-order protein, and hence, operating decisions need to be customized for each order.
- *Starting batch:* The quality of the starting material is one of the critical factors for success. For example, the starting material could involve "too little" target protein or "too much" impurity, such that the final purity and yield requirements can never be achieved even though the biomanufacturer makes the best operational decisions during the chromatography process.

Our objective in this research is to develop a formal, rigorous decision support tool that addresses the manufacturing challenges during the chromatography process.

Capacity Planning and Scheduling Challenges

Protein manufacturing operations are performed by highly skilled scientists using specialized equipment. *Capacity planning* for these limited specialized resources is critical for successful and timely completion of orders. Recognizing the manufacturing challenges described above, for each client order, the amount of labor and equipment needed for each step in fermentation and chromatography is estimated and incorporated into a capacity plan and production schedule that determines the lead time and delivery date to the client. Failing to satisfy delivery dates results in penalty costs and loss of credibility and reputation. Capacity planning in the biomanufacturing setting is particularly challenging because of the custom nature of orders. Effective capacity planning requires extensive data collection and capacity estimation based on the unique requirements of each order. Although

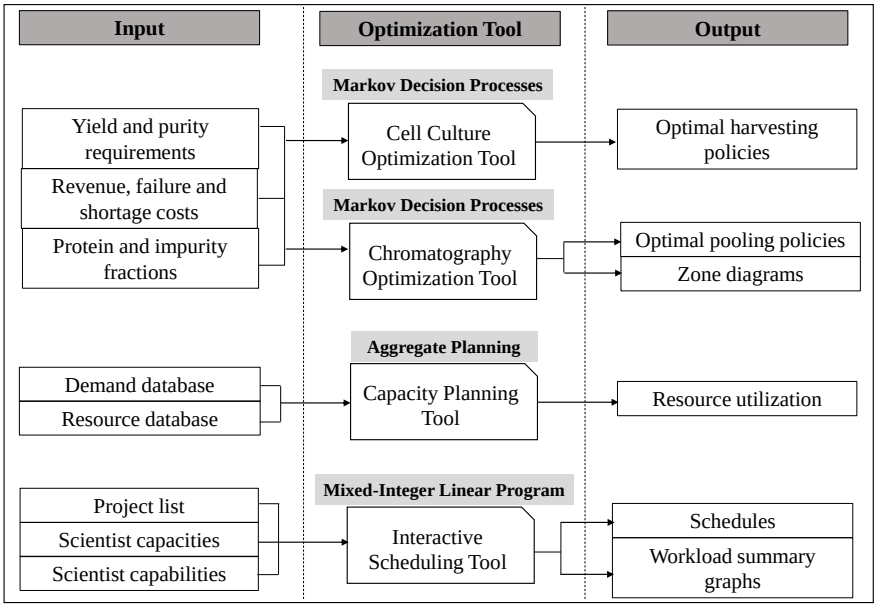


Figure 4 **This diagram summarizes the input and output of the developed tools.**

the biomanufacturer collects detailed data about the operations for regulatory compliance and quality control, the data are typically not structured for use in data-driven capacity planning decisions.

In addition, biomanufacturers also face unique challenges in *operational scheduling*. Each client order requires several tasks to be completed, namely, preparation, fermentation, chromatography, quality control, validation, and testing. The tasks and their durations differ between orders depending on the final product requirements, quality of the starting material, and required quality certifications. The use of live cells often introduces "no-wait" constraints between steps. The engineered nature of these products adds uncertainty at each step and imposes simultaneous requirements on highly skilled labor resources and specialized equipment to guarantee the best outcome.

If the planned schedule fails to achieve the expected outcome in terms of yield and purity, additional steps have to be planned into the schedule to meet the customer requirements. This can impact the schedules for other orders. Investigating different schedules to identify the best schedule that reacts to these dynamics is a challenge.

Operations Research Methods Provide Solutions

A portfolio of OR tools (see Figure 4) have been developed over a period of five years of collaboration with a team of researchers from the University of Wisconsin-Madison and

Aldevron to support various phases of protein manufacturing operations. The cell culture optimization tool addresses the harvesting decisions during fermentation, and the chromatography optimization tool was developed to optimize chromatography operating decisions. The capacity planning and interactive scheduling tools were created to support production planning decisions. These tools are designed so that they can operate independently of each other. This would allow other biomanufacturers to use one or multiple of these tools as needed. Aldevron has been an active participant in the testing and implementation of several of these tools. The resulting research outcomes have also been validated and disseminated to a broader group of biomanufacturers (BioWGS 2016).

Theoretical analysis of these problems has been reported in previous work (Martagan et al. 2016b, 2018). However, these studies focus only on a particular aspect, either fermentation or chromatography operations. In contrast, this work elaborates on the variety of problems that SMEs in the biomanufacturing industry face and describes the implementation of the theory through special decision support tools. It also focuses on practical aspects, such as the development of efficient solution procedures, challenges faced during implementation over the last five years, and benefits realized from implementation. We hope that this study will show how SMEs in related industries can benefit from OR to improve their business.

The Cell Culture Optimization Tool

The cell culture optimization tool helps to determine optimal harvesting policies (i.e., the best time to stop). The cell culture optimization tool uses the theory of Markov decision processes (MDP) and combines the cell-level dynamics (i.e., biology and chemistry of the underlying operations, as shown in Figure 2) with the manufacturing-level dynamics (i.e., purity–yield trade-offs, failure risks, and financial implications) to support decision making.

The states of the MDP model capture the time elapsed since the last shock, amount of target protein, and amount of impurities present in the batch. Possible actions are either to stop the process (i.e., harvest) or to continue to operate based on the predetermined protocol. Transition probabilities are defined based on the probability density distributions that describe the evolution of the amount of target protein and impurity over time. These probability distributions are non-stationary because of the different phases of cell growth. Other inputs to the MDP model are operating costs and rewards, specification limits on

the amount of impurities, and parameters required to model batch failures (e.g., shock arrival rate, probability of surviving a sudden failure). The objective is to identify an optimal harvesting policy that maximizes the value function defined as the expected total discounted profit (see Appendix A and Martagan et al. 2016b for details).

The optimization tool uses the policy iteration algorithm to solve the optimization problem. It is developed using the MATLAB® software and provides a user-friendly interface with Microsoft Excel where the scientists can easily enter the problem parameters and view the outputs.

Table 1 The tool generates a lookup table showing optimal policy and expected profit for each system state.

Time (hr)	State (n,w,m)[a]	Profit	Policy[b]
200	(4, 2, 30)	514	C
200	(4, 2, 35)	522	H
200	(4, 2, 40)	529	H
200	(4, 2, 45)	533	H
200	(4, 2, 50)	541	H
200	(4, 3, 0)	468	C
200	(4, 3, 5)	476	C

[a] System state is (n, w, m), where n denotes the time elapsed since the last shock, w is the amount of impurity present in the batch, and m is the amount of target protein in the batch.

[b] "C" means to continue the fermentation, and "H" is to harvest.

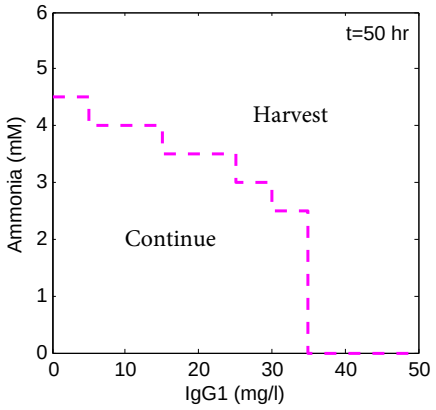


Figure 5 Optimal policies have control-limit structure and provide guidelines that are easy to implement in practice.

As an output, the tool generates a lookup table, where optimal actions and the corresponding expected profit are reported for each possible system state (see Table 1 for an illustrative example). In addition, the team observed that the optimal policy can have a control-limit structure under some realistic conditions on costs and state transitions (see Martagan et al. 2016b for details). Figure 5 illustrates the optimal policies at time $t = 50$ hr for a case study on IgG_1 production. In this figure, we observe that it is optimal to continue the process if the amount of protein and impurity is below a certain threshold value. For example, the bottom left of Figure 5 suggests to continue the fermentation process because the amount of protein and impurity in this region is *too little*. However, the optimal policy switches from continue to harvest as the amount of impurity increases at a given protein amount (e.g., top left of Figure 5). This is mainly because of the increased risk of failure

at higher levels of impurity. On the other hand, the optimal policy in the bottom right of Figure 5 suggests to harvest as the amount of target protein is already *high* in this region (i.e., due to limitations in cell culture, quality–yield trade-offs, failure risks, etc.). The tool helps to identify the amount that is just right. We also note that the tool can be adapted to accommodate multiple types of impurities, and the illustration with respect to IgG_1 and ammonia is only an example of the application of this tool.

Discussions with Aldevron and other industry practitioners during working group sessions revealed that the current common industry practice for harvesting decisions is to stop the process during the deceleration or stationary phase in order to get the highest possible yield. When we compared the optimal harvesting policy obtained from the optimization tool with the harvesting strategies typically used in practice, we observed that the optimal policy recommends harvesting earlier than typical practice. At first, this finding was counterintuitive to the practitioners involved in the discussions since the established practice in industry today is used to maximize the protein amount in a batch as a way to increase profit. However, this completely ignores the financial risk arising from failures in the process. Our optimization tool helped to illustrate this trade-off and to convince the practitioners that harvesting earlier could lead to higher expected profit owing to the combined impact of the failure risks involved in operations, purity–yield trade-offs, and randomness in process outcomes.

The Chromatography Optimization Tool

As described earlier, the main challenge in chromatography operations is to carefully select pooling windows to achieve the yield and purity requirements for the customer. The chromatography optimization tool helps to optimize pooling windows at each chromatography step. We formulate this problem as an MDP model with decision epochs corresponding to the beginning of a chromatography step (see Appendix B for the details of the MDP model). The states of the MDP model represent the amount of protein and impurity at the beginning of a chromatography step, and the action space at each epoch comprises a set of candidate pooling windows that can be used for that chromatography step. The model also includes the option of terminating the process at any decision epoch. Chromatography data from the scouting run are used to determine state transitions. The MDP model takes into account chromatography operating costs, penalty costs (e.g., yield shortage and/or

quality failure), and revenue obtained from the final batch as a function of the production requirements (i.e., The client purchases the protein only if the purity requirement is satisfied. The biomanufacturer does not obtain revenue from proteins manufactured in excess of the yield requirement and incurs a yield penalty cost for each unit of protein in short).

The tool uses a backward induction algorithm to determine optimal pooling windows at each chromatography step. The user interface and the underlying model are developed in Java. As the output, the tool generates a lookup table where the optimal pooling windows and optimal profit associated with each state are reported (see Table 2 for an example).

Table 2 The tool generates a lookup table showing optimal pooling windows and expected profit for each system state.

Step	Protein (mg)	Impurity (mg)	Profit	Policy
1	10	0	40	Stop
1	10	1	40	Stop
1	10	2	25	Lanes 4-16
1	10	4	-2	Lanes 4-12
1	10	12	-48	Stop

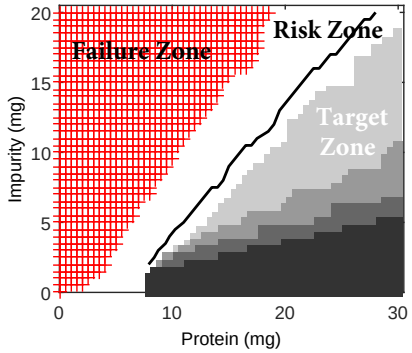


Figure 6 State space is partitioned into decision zones having similar financial characteristics.

We exploited the structural characteristics of the MDP model to derive guidelines on high/low quality starting materials. We showed that the state space can be partitioned into distinct subsets called the *failure zone*, *risk zone*, and *target zone* (see Martagan et al. 2018 for details). Figure 6 illustrates the decision zones generated for a case study at Aldevron. In this figure, if the starting material is an element of the failure zone, this implies that the biomanufacturer has no financial incentives to continue that chromatography step. In contrast, if the starting material is an element of the target zone, then the biomanufacturer can guarantee success in achieving the production requirements of the customer. The risk zone implies that the project might lead to either financial losses or profit, depending on the condition of the starting material and outcomes of individual chromatography steps. The solid line in Figure 6 corresponds to the break-even points. Note that these decision zones are uniquely defined for each order since the target protein is custom engineered.

Prior to the use of the chromatography optimization tool, scientists relied on their experience and domain knowledge to select pooling windows. Decisions related to guaranteeing

a successful outcome or stopping at an intermediate point because of a low chance of eventual success were purely based on experience and intuition. The tool provides an analytic approach for making decisions and effectively incorporating the various financial trade-offs related to chromatography operations. The tool also enables easier communication with clients if additional starting material and/or chromatography steps are needed.

The Capacity Planning Tool

The capacity planning tool helps to develop a data-driven capacity planning process and provides a capacity assessment for managerial decisions. It estimates the utilization of key labor and equipment assets and signals bottlenecks that could produce long lead times and late deliveries. This enables management to take actions to prevent long lead times and meet customer due date commitments.

The tool consists of a resource database on available capacity, a demand database, and an analysis toolpack for capacity assessment. The demand database records the project, labor, and equipment information for each activity in the protein manufacturing process. The resource database stores the capacities and capabilities of different equipment and labor assets.

The capacity analysis toolpack calculates resource utilization over a specified period based on information from the demand and resource databases using an aggregate planning framework and identifies potential bottlenecks. The capacity analysis toolpack also performs what-if analysis to assess managerial decisions such as the need for investments in new capital equipment, requirements for overtime, or the need to adjust due dates for projects. As an illustration, Figure 7 shows the analysis of 60 projects during the development phase of the tool. Our analysis revealed that 14 of 60 projects needed more than 20 days for completion. This was an issue since Aldevron expects their projects to be completed within 2 weeks. To understand the causes for long lead times, we conducted brainstorming sessions and discussed possible root causes such as absenteeism, nature of projects, quality issues, and resource bottlenecks. Using the tool and discussions, the group identified the high utilization of three specific chromatography machines as the root cause for the long lead times on several projects. Several possible solutions, including introducing overtime, hiring new employees, and purchasing new equipment, were evaluated using the capacity analysis toolpack to address this challenge. The tool suggested that purchasing

new chromatography equipment would be the best way to address the challenge. The decision was made, and the new equipment purchased enabled Aldevron to cut the lead times on several projects. Only 8 of the next 60 projects had more than 20 days of lead time, yielding a 50% reduction in the number of projects with long times.

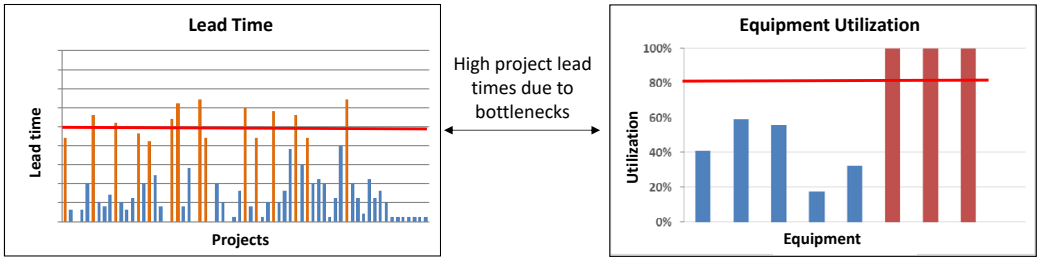


Figure 7 **The tool identified bottlenecks as the root cause for long lead times of the projects completed in a quarter of the year.**

The capacity planning tool generates resource utilization graphs that are actively used at the company. The current implementation shows that the tool has formed a strong basis for effective capacity management at Aldevron. The development of this tool led, in part, to the implementation of Protein Production Management –a work flow management platform with online documentation. This platform enables users to monitor the status of ongoing projects according to their scheduled tasks, due dates, start/end dates of each task, required labor/equipment hours, actual labor/equipment hours, and assigned scientists. Additionally, the details regarding the tasks, including material costs, purification results, quality specifications, quality test results, and client requirements, are documented for future reference. This has had a significant impact on not only planning capacity but also tracking performance variations across projects.

The Interactive Scheduling Tool

While the capacity planning tool provides planning support, the successful delivery of biomanufacturing projects to clients also requires the effective scheduling of projects and lab resources. The scheduling tool helps create schedules for projects with several tasks and gives users the flexibility to create customized schedules in response to the uncertainty related to each task and additional constraints. The objective of the tool is to provide an interactive scheduling system that optimizes resource assignments and that effectively schedules projects to meet delivery commitments. Although the capacity planning tool

and interactive scheduling tool do not interact with each other as they may be used for different time frames, the tools can be used together if needed.

A typical protein manufacturing process starts with plasmid prep, followed by cell culture, purification, and finally quality assessment. Each client project may include some or all of these tasks depending on their requirements, and the labor hours required for each task can vary between projects. As explained earlier, the custom nature of each project, differences in the priority given to various projects, no-wait constraints between tasks, and the preferred allocation of scientists to certain tasks and projects complicate the scheduling process. To find a good and feasible schedule at Aldevron, a team of scientists would often meet as a group for three hours each week. This process was becoming inefficient with the increasing volume and complexity of projects. The time spent for scheduling meetings increased and so did the need for revising schedules to accommodate the uncertainties in protein manufacturing operations.

The interactive scheduling tool models the scheduling problem at Aldevron as a mixed-integer linear program (details in Appendix C). The inputs of the tool are project lists with the associated tasks and scientist capabilities and capacities. The project list given in Table 3 includes the tasks performed for each project in sequence. The duration of each task (in weeks) and the labor hours required each week to complete that task are also given as inputs. The tool recognizes that each task can be assigned to any scientist from a set of scientists who are capable of performing the task. The number of hours that each scientist is available to work each week is stored in an input database. The tool determines scientist assignments for the tasks of each project as an output. It ensures the sequential completion of the required tasks, considers and incorporates scientists availability, and accounts for no-wait and preferred scientist constraints while determining the schedule. It runs the mathematical model with CPLEX in Java through Concert Technology and generates schedules. A simple illustration is given in Table 4, which shows the order and due dates for each project, assigned scientists along with their tasks, and number of hours that they need to spend each week. "Update" and "Priority" columns are filled to revise the schedules in the subsequent runs of the tool.

A key aspect of the tool is that the schedules can be revised by running the model iteratively and interactively using the revision features. The revision features of the tool allow users to specify high-priority projects, preferred scientist and time assignments and

Table 3 This table shows a sample project list used as an input to the interactive scheduling tool.

Project	Task	Order date	Due date	Number of hours in a week	Number of weeks
A1	Purification	12/06/2016	2/27/2017	20	1
A2	Eukaryotic expression	12/28/2016	2/28/2017	10	2
A2	Purification	12/28/2016	2/28/2017	30	1
A2	Quality control	12/28/2016	2/28/2017	8	1
A3	Purification	1/10/2017	3/8/2017	20	1
A3	Quality control	1/10/2017	3/8/2017	16	1

Table 4 This table shows a portion of the schedule generated by the tool.

Project No	Order Date	Due Date	Update	Priority	Status	01/08/2017	01/15/2017	01/22/2017	01/29/2017	02/05/2017	02/12/2017	02/19/2017
A1	12/30/2016	02/05/2017		High	Late		CR-EuE-30	CR-EuE-30	MM-Pur-20			
A2	12/28/2016	02/28/2017	Freeze		On time			KR-EuE-10	KR-EuE-10	JG-Pur-30	JG-QC-8	
A3	12/06/2016	02/27/2017	Freeze		On time						MS-Pur-20	
A4	01/10/2017	03/08/2017	Change		On time				JR-Pur-40			MM-QC-16

to explore different schedules. The user can also freeze portions of the schedule, update scientist availabilities, and optimize the rest of the schedule. In this manner, schedules can be iteratively improved following the steps given in Figure 8.

The interactive scheduling tool captures the needs of several small biomanufacturing companies that face complex scheduling challenges. Incorporating the ability to interactively revise schedules gives managers an opportunity to consider different objectives while generating feasible schedules iteratively. Feedback from users at working group sessions suggests that the flexible structure of the tool will enable efficient scheduling in many biomanufacturing companies.

Implementation

A four-phased approach was adopted for the implementation of each of the tools, cell culture and chromatography optimization, capacity planning, and interactive scheduling. Recognizing that users of these tools will be scientists with several other responsibilities, particular attention was devoted to automatizing the inputs and outputs of these tools and enhancing the tools to provide flexibility in decision making.

- *Phase 1: Data collection.* The team worked on the methods of collecting input data and improving data quality.
- *Phase 2: Tool validation.* Pilot runs were done with chromatography, cell culture, and demand data to test the validity of the tools.
- *Phase 3: Automation of inputs and outputs.* Simple graphical user interfaces were created to enable scientists to run the tools efficiently.

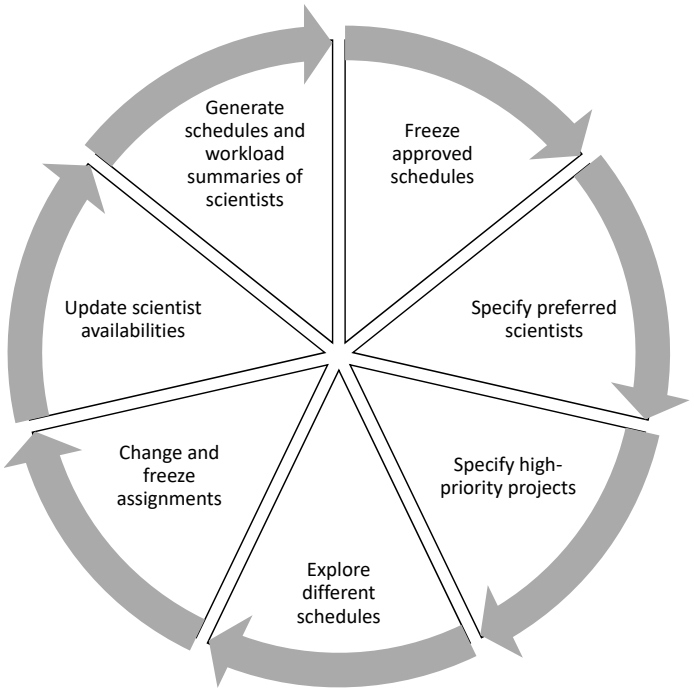


Figure 8 The revision features of the tool provide flexibility for scheduling decisions.

- *Phase 4: Tool enhancement.* New features were introduced to improve the use of the tools, based on the feedback obtained from pilot runs.

Implementation of the Chromatography Optimization Tool

Data collection: Scientists performed initial small-scale experiments called scouting runs to understand the feasibility of the customer requirements after receiving an order. Data collected from these scouting runs provided the inputs regarding protein and impurity amounts for the chromatography optimization tool. Revenue and cost inputs were obtained from management at Aldevron. The cell culture data, including protein and impurity amounts, were obtained in collaboration with scientists at UW-Madison for the development of the cell culture optimization tool.

The chromatography optimization tool in particular required preprocessing the collected data to be used as inputs. The first step in the data preprocessing is to measure the protein and impurity percentages in each lane from the gel images shown in Figure 3. Scientists use gel imaging software to identify the percentage of target protein and impurity in each lane. To calculate the distribution of target protein across the lanes, they measured the total protein in each lane by applying additional techniques/assays. The total protein includes the target protein and impurities. The percent distribution of the target protein

and impurities across the lanes was determined by using the total protein obtained from this assay, and the percentage of target protein in each lane was calculated using gel imaging software.

Tool validation: For the chromatography optimization tool, after the procedures to collect inputs were defined, the team conducted several trial runs where the optimal policies generated by the tools were compared with the actions performed by the scientists. Initial comparisons revealed a few inconsistencies. In these few cases, the team found that the presence of compromised target protein (i.e., misfolded, aggregated) could mislead the inputs of the chromatography optimization tool, as these species could have the same molecular weight as the target protein but are not functionally useful and consequently are considered an impurity. To overcome this challenge, guidelines were established to flag these species as impurities. In a few other cases, the team observed that the chromatography optimization tool suggested pooling lanes in addition to those selected by the scientists, indicating that a higher yield is possible while satisfying the purity requirement.

To test the robustness of the tools, the team conducted sensitivity runs with different cost parameters and probability distributions. For instance, the team studied how an optimal decision related to a chromatography step would change for different probability distributions. These robustness tests were instrumental in building user confidence in the use of these tools.

Automation of inputs and outputs: After the completion of the validation phase, graphical user interfaces were created for the chromatography optimization tool to enable the scientists to use the tools with minimal knowledge of the underlying MDP. Although these scientists have advanced degrees, their knowledge of OR is minimal. These user interfaces were critical for making the power of OR models accessible to these scientists. Figure 9 shows an overview of the user interface for the chromatography optimization tool. Data are entered in four main modules by using a drop-down menu. More specifically, each sub-figure in Figure 9 represents a data entry module and corresponds to an item in the drop-down menu. The tool dynamically generates the output based on the input entered.

Tool enhancement: During the tool validation runs, the team realized that for some orders, cost estimates related to penalty costs in particular may exhibit significant uncertainty. To overcome this challenge, we introduced a modified version of the chromatography optimization tool in which we added an alternative feature to evaluate the chromatography

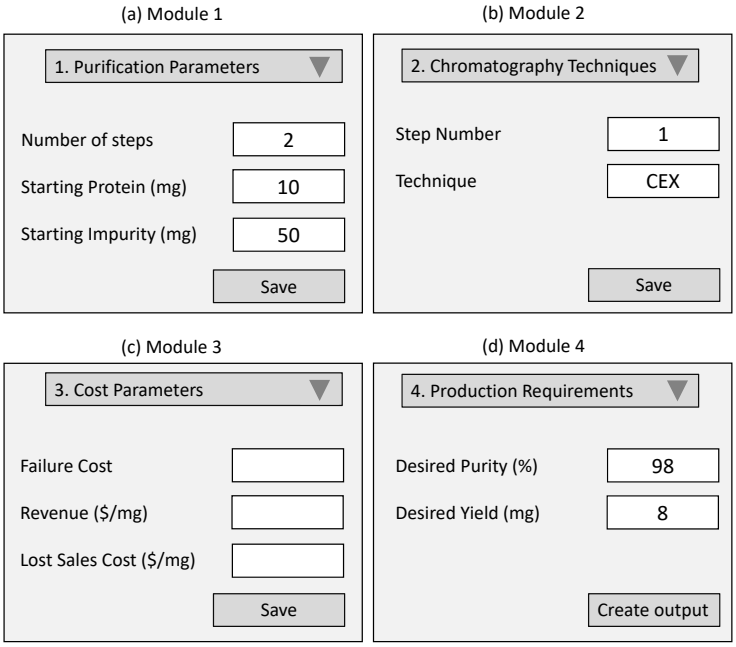


Figure 9 The user enters the required inputs into four main modules in the chromatography optimization tool.

performance without cost parameters. This feature generates optimal policies with the aim of maximizing yield while satisfying purity instead of maximizing profit. This version of the tool provides the maximum protein amount that can be collected. For example, if the starting protein and impurity amounts are in the target zone, the tool lists at least one pooling strategy that satisfies the customer requirements even if the scientist experiences the worst-case scenario with respect to chromatography outcomes. The points in the failure zone shown in this version indicate that it is impossible to satisfy the requirements even with the best-case purification capabilities. Consequently, the tool provides a quick check regarding whether the desired yield and purity are satisfied with the given starting protein and impurity amounts when the cost estimation is difficult. Figure 10 provides an overview of the user interface of this version of the chromatography optimization tool.

Implementation of the Capacity Planning and Interactive Scheduling Tools

Data collection: Prior to the implementation of the capacity planning and interactive scheduling tools, the scientists were recording project information on lab notebooks for traceability purposes. The team initiated electronic data entry to overcome the difficulties of physical documentation and emphasized that this information was also used for capacity planning and scheduling. Our tools raised the importance of good inputs. Aldevron

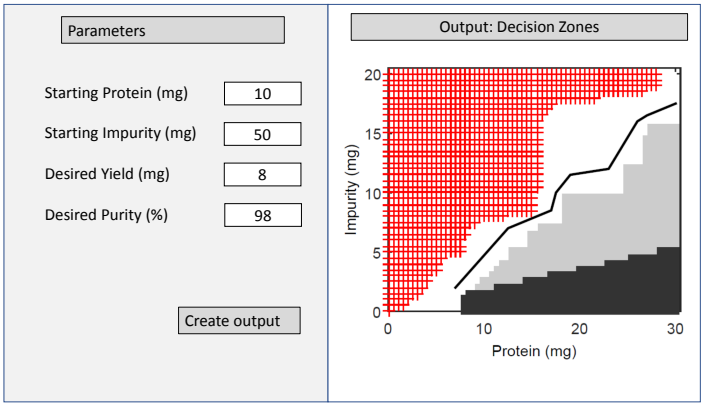


Figure 10 This figure shows the user interface for the chromatography tool lite tool.

launched the Protein Production Management as an offshoot of our collaboration, and this platform provides the required inputs for the capacity planning and interactive scheduling tools.

Tool validation: To validate the capacity planning tool, the team first ensured the accuracy of the input data by comparing labor and equipment hours against values recorded by the scientists. While the equipment hours were tracked with reasonable accuracy, under-reporting of labor hours was common, which impeded their ability to obtain accurate estimates of labor utilization. With the help of continuous monitoring of the recorded data and clear communication in weekly meetings, procedures were established for training the scientists to record the necessary data and prevent under-reporting. After the data validity was ensured, the team discussed the utilization results obtained by the capacity planning tool. For instance, the team investigated whether a machine/scientist presented as highly utilized by the tool was actually a constraint in the lab, as well as the potential solutions for these constraints. These discussions built confidence in the tool outputs.

The team conducted several validation runs for the interactive scheduling tool with different project lists and different planning periods to check whether the schedules presented as an output of the tool satisfied all the constraints. After validating the accuracy of the schedules, the team discussed the effectiveness of the generated schedules and compared them with the actual schedules planned by management at Aldevron. We concluded that the generated schedules were effective and could be implemented. One issue, however, was that the tool could delay the start of a project much later than a manager at Aldevron

would in reality, if doing so did not violate any constraints or induce late deliveries. However, management felt that such schedules are undesirable, as they might limit Aldevron's ability to accept more orders in future weeks. To address this issue, the tool was modified so that projects would be finished as early as possible. In addition, the team discovered that some of the scientists time assignments were not feasible because they were on vacation and had other responsibilities. To address this constraint, the team introduced additional features.

Tool enhancement: After understanding the need for flexible inputs for capacity during the validation runs, the team added scientist capacities as an input to the model in order to specify the available labor hours each week. Moreover, in the case of new hires, changes to scientists responsibilities, or new task definitions, the team introduced scientist capabilities to describe which tasks can be assigned to each scientist. Both scientist capacities and capabilities were recorded in the same Excel file, and users can modify them before running the tool. The flexibility of the interactive scheduling tool has been further enhanced with the revisions that can be made to the resulting schedules. The improved version of the tool was rolled out after adding the options of specifying high priority projects, freezing assignments, and exploring different schedules.

Automation of inputs and outputs: As with the cell culture and chromatography optimization tools, the interactive scheduling tool was set up so that users do not have to interact with the mathematical model. However, the input data were set up so that they can be easily retrieved from the Protein Production Management. The interactive scheduling tool generates schedules and workload summary graphs, and users can run the model again after revising the resulting schedules in the same file.

Benefits of OR Tools in Biomanufacturing Operations

The suite of OR tools has provided a formal framework that enables Aldevron to improve its manufacturing operations, planning, and scheduling. The tools were deployed at Aldevron with the help of regular meetings with Aldevron and necessary IT support. They have also been shared with local biomanufacturing firms (BioWGS 2016).

The suite of OR tools offers the following three major benefits:

Formal assessment of risks and manufacturing capabilities: The capacity planning tool identifies bottleneck resources that may have immediate effects on manufacturing and

scheduling decisions. The capacity planning and interactive scheduling tools help clarify capacity needs and thus the feasibility of client agreements regarding lead time and delivery. The interactive scheduling tool incorporates unique biomanufacturing scheduling constraints into the model and allows users to enter additional constraints and task definitions as required. Both of these tools help Aldevron anticipate risks of late delivery and explore corrective strategies. The chromatography optimization tool captures the uncertainties in the processes and the financial implications of operating strategies. The tool provides a formal assessment of starting material and manufacturing capabilities using a novel zone-based decision-making approach, allowing companies such as Aldevron to both guarantee performance when possible and warn clients to decide to fail early and thus minimize losses.

Increased flexibility in manufacturing and scheduling: The use of OR tools at Aldevron has reduced the risks of failing client requirements. These tools quantify the manufacturing trade-offs and challenges for the clients. Providing performance guarantees through the decision zones has resulted in higher client satisfaction and business growth. From a scheduling point of view, the revision features embedded in the scheduling tool enable managers to investigate different scenarios and find the best schedules that improve on-time deliveries and client satisfaction. The capacity planning and interactive scheduling tools also enable managers to deal with more consistent operations and dynamically respond to resource utilization issues, resource unavailability, and rescheduling needs for tasks in order to prevent late deliveries.

Data-driven decision making: The approach taken to develop the OR tools demonstrates the power of data-driven manufacturing optimization in the biomanufacturing industry. The capacity planning and interactive scheduling tools provide solutions for efficient planning and resource utilization. The chromatography optimization tool integrates bioscience with the process economics and financial trade-offs by examining the amounts of target protein and impurities obtained during operations. The outputs of the developed tools not only help improve the cell culture, chromatography, planning, and scheduling processes but also establish better communication with clients.

Conclusions

In the biomanufacturing industry, there is an increasing need for OR-based decision support mechanisms to reduce production costs, lead times, and failure risks. This paper presents a

systematic application of OR methodologies to improve biomanufacturing efficiency. Close collaboration with an interdisciplinary team of researchers at Aldevron resulted in a portfolio of OR tools that rigorously respond to common industry challenges. The portfolio of OR tools proposes solutions for the optimal use of resources in terms of cost and lead time. For example, the cell culture and chromatography optimization tools capture the batch-to-batch variability and uncertainty in the production processes and thus enable the integration of cell-level dynamics (i.e., biological and chemical processes) with higher level manufacturing decisions (e.g., optimal operating policies) to reduce failure risks and costs. The capacity planning and interactive scheduling tools enable effective use of limited capacity with increased flexibility in manufacturing and scheduling. Since 2013, these tools have been validated and enhanced using Aldevron data. The insights obtained from these tools have directly affected business metrics, resulting in a roughly 20% reduction in timelines and costs. These tools have also played an important role as Aldevron has grown 3x since 2013. In addition to the financial impact on the business, during the implementation of the OR tools, Aldevron experienced a systematic change in encouraging increased use of data-driven decision making and use of OR culture within the company. The development of the work flow management platform has also encouraged the adoption of a systematic, OR-based approach to optimize manufacturing and planning decisions.

OR methodologies have not yet been widely used in the biomanufacturing industry. However, we strongly believe that the true impact of this work extends beyond Aldevron's operations. Complementing societal advances in biological research with sound OR methodologies will significantly help the industry accelerate early phase drug development research. The benefits will be significant for the industry as a whole and its customers —patients needing life-saving drugs.

Acknowledgments

We thank the editors and anonymous referees for their constructive comments that helped improve this paper. We also thank the National Science Foundation for supporting this research under Grant No. CMMI 1334933.

Appendix A. MDP Model for the Cell Culture Optimization Tool

• **Decision epochs:** Decision are made at time t, $\mathcal{T} = \{t : 0, \tau, 2\tau, 3\tau, \ldots, T\}$. The length $\tau > 0$ could vary for each bioprocess. Time $t \in \mathcal{T}$ also represents the age of the process. The maximum age is bounded by T because of the biological constraints (e.g., limitations in cell growth and culture lifetime).

• **States:** The system state is represented by (n, w, m) in finite state space $\mathcal{N} \times \mathcal{W} \times \mathcal{M}$. State $n \in \mathcal{N}$ denotes the time elapsed since the last shock, $\mathcal{N} = \{0, \tau, 2\tau, 3\tau, \ldots, T\}$. State $w \in \mathcal{W}$ represents the amount of impurity present in the batch, $\mathcal{W} \equiv \left[0, \bar{W}\right) \cup \{\Delta\}$, where $\bar{W}$ is a specification limit imposed by regulatory requirements to ensure quality and Δ denotes batch failure. State $m \in \mathcal{M}$ represents the amount of target protein present in the batch, $\mathcal{M} \equiv \left[0, \bar{M}\right]$. The highest protein amount that can be achieved is $\bar{M}$ owing to limitations in cell cultures. The process starts at time $t = 0$, with states $n = 0, w = 0$, and $m = 0$.

• **Actions:** The scientist can either continue the fermentation process (C) or harvest (H) at each decision epoch. Hence, $\mathcal{A} = \{C, H\}$. Let $a_t^*(n, w, m)$ denote an optimal action at time t and state (n, w, m). Then, it is optimal to harvest if a batch fails (i.e., $a_t^*(n, \Delta, m) = H$ for $t \in \mathcal{T}$, $n \in \mathcal{N}$, and $m \in \mathcal{M}$) or if it reaches either the maximum age T (i.e., $a_T^*(n, w, m) = H$ for $n \in \mathcal{N}$, $w \in \mathcal{W}$, and $m \in \mathcal{M}$) or the specification limit $\bar{W}$ (i.e., $a_t^*(n, \bar{W}, m) = H$ for $t \in \mathcal{T}$, $n \in \mathcal{N}$, and $m \in \mathcal{M}$).

• **State transitions:** Let X_t be a random variable with realization x_t and probability density function $f_t^x(\cdot)$ (general distribution) in the interval $[x_t^L, x_t^U]$ for $t \in \mathcal{T}$. Note that $f_t^x(\cdot)$ is time dependent. If the process continues at time t, then the amount of protein at time $t+1$ is modeled with the additive function:

$$m_{t+1} = \begin{cases} 0 & \text{if the batch is harvested at time } t, \\ m_t + x_t & \text{if the batch is not harvested at time } t. \end{cases}$$

We define $\zeta(n, \rho)$ as the probability of a shock arrival after n periods of no shocks, where ρ is a vector of parameters of the unspecified distribution of the time between two shocks. In addition, random variable B_t with general distribution $f_t^b(\cdot)$ and realization b_t represents the increase in the impurity amount during period $[t, t+1)$. E_t is a random variable with general distribution $f_t^\epsilon(\cdot)$ and realization e_t, and it represents the increase in the impurity amount because of shocks during period $[t, t+1)$. α_t is the probability of surviving a sudden failure. Note that $f_t^\epsilon(\cdot)$, $f_t^b(\cdot)$, and α_t are time dependent. If a batch continues at time t, the impurity amount at time $t+1$ is given by

$$w_{t+1} = \begin{cases} w_t + b_t + e_t & \text{if there is a shock during } [t, t+1), \\ w_t + b_t & \text{if there are no shocks during } [t, t+1). \end{cases}$$

The reliability function is $R_t(n, w) = \zeta(n, \rho)\alpha_t \int_0^{\bar{W}-w} f_t^{\epsilon+b}(z)\,\mathrm{d}z + (1 - \zeta(n, \rho)) \int_0^{\bar{W}-w} f_t^b(y)\,\mathrm{d}y$, and it represents the batch survival probability at state $(n, w) \in \mathcal{N} \times \mathcal{W}$ during period $[t, t+1)$. The convolution $Z_t = E_t + B_t$ has the density function $f_t^{e+b}(z)\mathrm{d}z = f_t^e * f_t^b$.

• **Costs and rewards:** The cost of operating during one period is $r_c(w, m)$. The operating cost is nondecreasing in w and m. The revenue obtained from harvesting at state (w, m) is $r_h(w, m)$, and it is nondecreasing in m and nonincreasing in w. The penalty cost of failure is $r(\Delta)$.

• **Value function:** $\mathcal{V}_t(n, w, m)$ represents the expected total discounted profit when the batch is in state (n, w, m) at time t. For all $(n, w, m) \in \mathcal{N} \times \mathcal{W} \times \mathcal{M}$, the value function is

$$\mathcal{V}_t(n,w,m) = \begin{cases} -r(\Delta) + \mathcal{V}_0(0,0,0) & \text{if } w = \Delta, \\ \max\{r_h(w,m) + \mathcal{V}_0(0,0,0),\ -r_c(w,m) + \beta\,\mathcal{C}_t(n,w,m)\} & \text{otherwise}, \end{cases} \tag{1}$$

where

$$\begin{aligned} \mathcal{C}_t(n,w,m) \;=\; & [1 - R_t(n,w)]\,[-r(\Delta) + \mathcal{V}_0(0,0,0)] \\ & + \zeta(n,\rho)\,\alpha_t \int_0^{\bar{W}-w} \int_0^{\bar{M}} f_t^{e+b}(z) f_t^x(x) \mathcal{V}_{t+1}(0, w+z, m+x)\,\mathrm{d}x\,\mathrm{d}z \\ & + [1 - \zeta(n,\rho)] \int_0^{\bar{W}-w} \int_0^{\bar{M}} f_t^b(b) f_t^x(x) \mathcal{V}_{t+1}(n+1, w+b, m+x)\,\mathrm{d}x\,\mathrm{d}b. \end{aligned} \tag{2}$$

The discount factor is $0 < \beta < 1$. The function $\mathcal{C}_t(n,w,m)$ denotes the expected rewards obtained from the action of continuing at time t and state (n,w,m). This is an infinite-horizon problem where a harvested or failed batch is immediately replaced with a new one, leading to the expression $\mathcal{V}_0(0,0,0)$ in Equations (1)-(2).

Appendix B. MDP Model for the Chromatography Optimization Tool

- **Decision Epochs:** Decisions are made at the beginning of chromatography step t, $\mathcal{T} = \{t : 1, \ldots, T-1\}$, where $T-1$ denotes the last chromatography step. The end of the planning horizon is T.

- **States:** The system state is represented by (p_t, i_t) in finite state space $\mathcal{P} \times \mathcal{I} \cup \Delta$. State $p_t \in \mathcal{P}$ denotes the amount of protein available in the batch at the beginning of the t^{th} chromatographic step. State $i_t \in \mathcal{I}$ is the amount of impurity at the beginning of the t^{th} chromatographic step. Stopping state Δ represents a batch that is either shipped or scrapped. Δ is an absorbing state with no rewards.

- **Actions:** $w_t \in \mathcal{W}_t$ denotes the choice to use pooling window w_t to run chromatography step $t \in \mathcal{T}$. Action S denotes the choice to stop the chromatography process. Hence, $\mathcal{A}_t = \mathcal{W}_t \cup S$ at $t \in \mathcal{T}$. At the end of planning horizon T, the only available action is to stop.

- **State transitions for the protein:** At each chromatography step $t \in \mathcal{T}$, random fraction $\Theta_t | w_t$ of protein p_t is carried over to the next chromatography step $t+1$, i.e., $p_{t+1} = (\theta_t | w_t) p_t$. Random fraction Θ_t has general distribution $f_t(\cdot | w_t)$ with finite support $[\theta_t^\ell | w_t, \theta_t^u | w_t]$ for $w_t \in \mathcal{W}_t$, $t \in \mathcal{T}$.

- **State transitions for the impurity:** At each chromatography step $t \in \mathcal{T}$, random fraction $\Psi_t | w_t$ of impurity i_t is carried over to the next chromatography step $t+1$, i.e., $i_{t+1} = (\psi_t | w_t) i_t$. Random fraction Ψ_t has general distribution $g_t(\cdot | w_t)$ with finite support $[\psi_t^\ell | w_t, \psi_t^u | w_t]$ for $w_t \in \mathcal{W}_t$, $t \in \mathcal{T}$.

- **Production requirements:** p_d denotes the yield requirement, and γ_d is the purity requirement. Batch purity γ_t at step $t \in \mathcal{T} \cup T$ is defined as $\gamma_t = \frac{p_t}{p_t + i_t}$.

- **Operating and penalty costs:** The operating cost of chromatography step t is denoted by c_t. The penalty cost of failure is c_f. The yield penalty cost at $t \in \mathcal{T}$ is $c_\ell(p_d - p_t)$ if $p_d > p_t$ and zero otherwise.

- **Stopping costs and rewards:** Revenue obtained per unit of protein p_t is $r(p_t)$. Let $r_S(p_t, i_t)$ be the reward obtained from stopping the process at state $(p_t, i_t) \in \mathcal{P} \times \mathcal{I}$ and chromatography step $t \in \mathcal{T}$. Then,

$$r_S(p_t, i_t) = \begin{cases} -c_f & \text{if } \gamma_t < \gamma_d, \\ r(p_d) & \text{if } \gamma_t \geq \gamma_d \text{ and } p_t \geq p_d, \\ r(p_t) - c_\ell(p_d - p_t) & \text{if } \gamma_t \geq \gamma_d \text{ and } p_t < p_d, \end{cases}$$

for $t \in \mathcal{T}$ when $a_t(p_t, i_t) = S$, and for $t = T$.

- **Value function:** $\mathcal{V}_t(p_t, i_t)$ represents the expected total profit when the batch is in state $(p_t, i_t) \in \mathcal{P} \times \mathcal{I}$ at chromatography step $t \in \mathcal{T}$. For all $(p_t, i_t) \in \mathcal{P} \times \mathcal{I}$, the value function is defined as follows:

$$\mathcal{V}_t(p_t, i_t) = \max_{w_t \in \mathcal{W}_t} \left\{ r_S(p_t, i_t),\ -c_t + \mathop{\mathbb{E}}_{\theta_t, \psi_t | w_t} \mathcal{V}_{t+1}(\theta_t p_t, \psi_t i_t) \right\}, \text{ for } t = \{1, \ldots, T-1\}, \text{ and}$$

$$\mathcal{V}_T(p_T, i_T) = r_S(p_T, i_T),$$

where $\mathop{\mathbb{E}}_{\theta_t, \psi_t | w_t} \mathcal{V}_{t+1}(p_t\theta_t, \psi_t i_t) = \int_{\psi_t^\ell | w_t}^{\psi_t^u | w_t} \int_{\theta_t^\ell | w_t}^{\theta_t^u | w_t} f_t(\theta_t | w_t) g_t(\psi_t | w_t) \mathcal{V}_{t+1}(\theta_t p_t, \psi_t i_t) \mathrm{d}\theta \mathrm{d}\psi$. Note that $\mathcal{V}_t(\Delta) = 0$ for $t \in \mathcal{T} \cup T$.

Appendix C. Mathematical Modeling of Scheduling Problems

We solved a mixed-integer linear problem to find schedules using project lists and scientist capacities and capabilities. To simplify notation, we provide only the pseudo-code description of the problem formulation.

Minimize Total tardiness

s.t.

Capable scientist assignment to each task in project $= \mathbb{1}\{$(Task is required for project)$\}$

Weekly labor hours assigned to scientist $\leq$ Scientist capacity

Completion time of task $\leq$ Start time of task's successor

Start time of task + Duration of task = Start time of task's successor

(Completion time of project) $\geq$ (Start time of project's last task) + (Duration of project's last task)

References

BioWGS. 2016. Biomanufacturing Working Group Sessions organized by the University of Wisconsin-Madison and Aldevron. URL http://www.bioforward.org/event/id/392756/Challenges-and-Opportunities-in-Managing-Biomanufacturing-Operations.htm.

Engel, J. 2014. Madison biotech supplier goes lean to stave off foreign competitors. Xconomy. URL http://www.xconomy.com/wisconsin/2014/02/19/madison-biotech-supplier-goes-lean-to-stave-off-foreign-competitors/.

Foti, T., A. Krishnamurthy, P. Leland. 2016. Aldevron collaborates with the University of Wisconsin to develop decision support tools for biomanufacturing. *BioForward Speaker Series* .

GE Healthcare. 2010. Strategies for protein purification: Handbook by GE Healthcare. Tech. rep.

Gnoth, S., M. Jenzsch, R. Simutis, A. Lubbert. 2007. Process Analytical Technology: Batch-to-batch reproducibility of fermentation process by robust process operational design and control. *Journal of Biotechnology* **132** 180–186.

INFORMS. 2016. George B. Dantzig dissertation award, honorable mention. URL https://www.informs.org/Recognizing-Excellence/INFORMS-Prizes/George-B.-Dantzig-Dissertation-Award.

INFORMS. 2017. The research project received the first prize in the 2017 INFORMS TIMES best dissertation award. URL https://higherlogicdownload.s3.amazonaws.com/INFORMS/4d4e9ecf-9747-4303-8889-8c1d08557f6b/UploadedImages/TIMES_2017_Newsletter.pdf.

Long, G., J. Works. 2013. Innovation in the biopharmaceutical pipeline: A multidimensional view. *Analysis Group White Reports*.

Martagan, T., A.Krishnamurthy, P. A. Leland, C. T. Maravelias. 2018. Performance guarantees and optimal purification decisions for engineered proteins. *Operations Research* **66**(1) 18–41.

Martagan, T., A. Krishnamurthy, P.A. Leland, C. T. Maravelias. 2016a. Optimal purification decisions for engineer-to-order proteins at Aldevron. *Production and Operations Management* **25**(12) 2002–2013.

Martagan, T., A. Krishnamurthy, C. T. Maravelias. 2016b. Optimal condition-based harvesting policies for biomanufacturing operations with failure risks. *IIE Transactions* **48**(5) 440–461.

McNeil, B., L. M. Harvey. 2008. *Practical Fermentation Technology*. John Wiley & Sons, New York.

Mordor Intelligence. 2017. Global biopharmaceuticals market growth, trends and forecasts. URL https://www.mordorintelligence.com/industry-reports/global-biopharmaceuticals-market-industry.

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

Pharmaceutical Research and Manufacturers of America. 2016. Biopharmaceutical research industry 2016 profile. URL http://phrma-docs.phrma.org/sites/default/files/pdf/biopharmaceutical-industry-profile.pdf.

The European Biopharmaceutical Enterprises. 2015. Medical innovation that helps millions of patients around the world. URL http://www.ebe-biopharma.eu/uploads/documents/EBE%20Little%20Book%20-%20FINAL.pdf.

WEDC. 2014. Wisconsin Economic Development Corporation (WEDC), Company Feature: Aldevron. URL http://inwisconsin.com/entrepreneurs-and-innovators/launch_blog/company-feature-aldevron/.