Leveraging Data for Predictive and Real-time Bioprocess Performance Improvement

Examining the roles of digital twin platforms and PAT-enabled, advanced-control bioreactors in bioprocess improvement for accelerated product development.

Brett Lepore
President
Yokogawa Insilico Biotechnology
North America
Yokogawa Electric Corporation

Paul E. Cizdziel, Ph.D.
Global Market Development
Bioprocess
Yokogawa Electric Corporation

Sponsored by

YOKOGAWA

INTRODUCTION

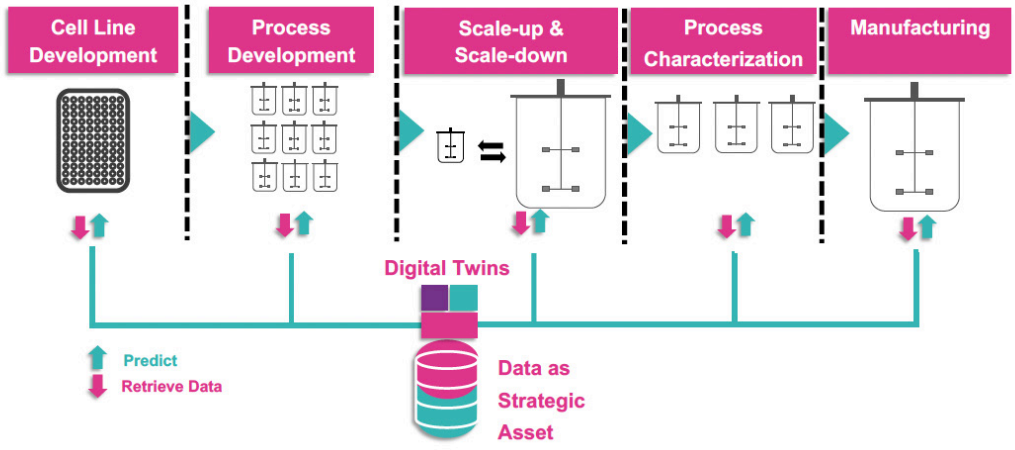
In the biopharmaceutical industry, manufacturing success relies on the ongoing development, understanding, and improvement of involved processes. Today, digitalization and data-driven methodologies are opening the way for greater virtual exploration in the development and manufacturing design space. In parallel, real-time analytical technologies deliver the information required for performance understanding and data-driven bioprocess improvements. This article first examines how using a digital twin platform can increase speed to market and improve process performance, and then explores recent progress in applying process analytical technologies [PAT] for the automation of bioreactor systems.

PART 1: THE DIGITAL TWIN PLATFORM

The estimated average cost of developing a new drug is currently 1.3 billion USD and the process can take around ten years to complete.1 Within just one and a half months, Moderna brought the SARS-Cov-2 gene sequence from publication to shipping for the first vaccine for clinical trials. Their previous means of building digital capabilities and de-siloing data enabled seamless data flow throughout the company, supporting the rapid development and commercialization of the vaccine.

When developing any biopharmaceutical product, every month saved in development time potentially contributes millions of dollars to the top line. Digital twins have been used for years in a variety of industrial settings for process development, optimization, and control. While they are relatively new to the biopharmaceutical industry, it's clear that using a digital twin model of bioreactor to de-silo data can save time while providing greater insight into the process.

FIGURE 1: The role of data for de-siloing biprocess lifecycle.



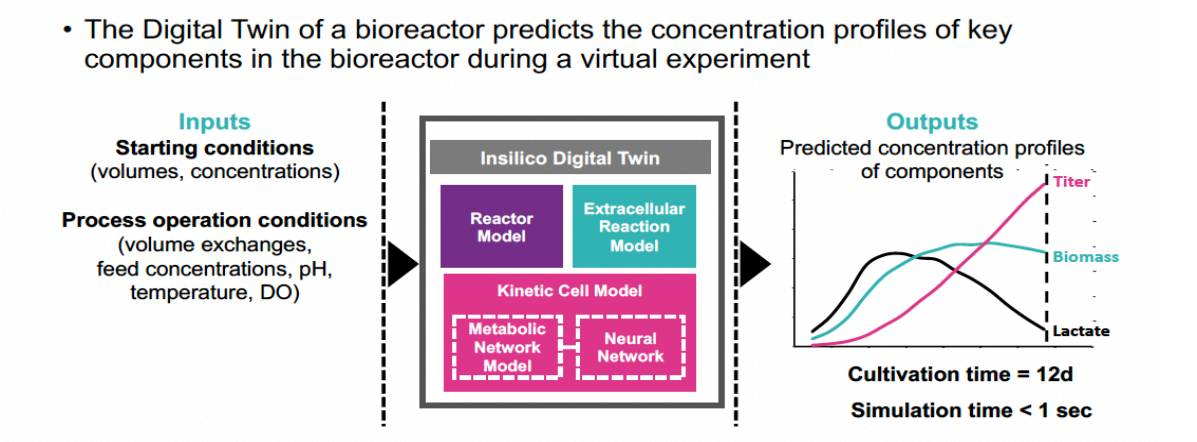
De-Siloing Data Across the Bioprocess Lifecycle

As shown in **FIGURE 1**, there are five distinct stages in the cell culture process: cell line development, process development, scale-up and scale-down, process characterization, and manufacturing. Each stage is significant, requiring and generating high-value data. Historically, this data has remained within each individual stage, and has not been used to build further knowledge of the overall process. A digital twin can bridge the gaps where data is under-utilized at the different lifecycle stages.

A digital twin is a virtual representation of a physical object, system, or process (in this case a bioreactor in the cell culture process) that is used to predict performance. The digital twin described in **FIGURE 2** (Insilico Digital Twin, Yokogawa) predicts the concentration profiles of key components within the bioreactor during a virtual experiment. This hybrid digital twin is comprised of the following models:

- The reactor model tracks concentration changes due to feeding or sample removal, and represents concentration and volume changes due to feeding and sampling.

FIGURE 2: Digital twin of a bioreactor.



- The extra-cellular reaction model includes reactions that take place in the cultivation media, but do not involve cells (e.g., oxidation of metabolites).
- The kinetic cell model tracks changes in concentrations primarily within the living cell that results from cell metabolism, growth, and culture status.

The kinetic cell model is arguably the most significant and is comprised of two sub-models. The first sub-model is the metabolic network model, a continuously refined, genome-based network model of a particular cell line, in this case the Chinese Hamster Ovary [CHO-K1]. The second sub-model is the neural network artificial intelligence [AI] machine learning [ML]; this is the part of the model that drives the learning of the cellular metabolism dynamics.

Inputs for the digital twin are the same as for a wet experiment: starting concentrations, volumes, and concentrations at inoculation. Process operating conditions must also be input. The outputs are the predicted concentration profiles of critical components, including biomass, titer, and lactate. In the wet lab, equivalent experimentation typically takes 10 to 12 days for mammalian cell culture, whereas virtualization in the digital twin is completed in a matter of seconds. This makes it possible to carry out thousands of virtual experiments to scan the design space of the cell culture process at various stages. Simply changing the feed concentrations, pH, temperature, or other parameters enables the optimal design space which can achieve higher titers or other favorable effects on critical quality attributes of the biologic being produced.

Connecting R&D and Manufacturing Workflows

The digital twin connects research and development [R&D] and manufacturing workflows. This is in part due to the design-enabling modification of each individual model within the overall hybrid digital twin model, independently of others. For example, if the data used to train the kinetic model originated from deep-well plates, the process model can be modified to reflect a higher-scale bioreactor to understand how those cells would respond in a different process. The reactor model reflects gas and liquid inflows and outflows, and tracks volume over time. Therefore, it can track the volumetric mass transfer coefficient [kLa] and shear stress, which are critical parameters for accurate prediction of behavior at larger scales.

Application Across the Lifecycle

This digital twin can also de-silo data across different scales. In cell line development, the digital twin can predict the performance of multiple clones in a variety of media and feeding schemes, and upon scale-up to bioreactors. It simultaneously accomplishes both smart clone and process development. Inserting a gene into a pool of cells produces a variety of different clones; knowing which will perform best saves substantial amounts of time and resources. Use of the digital twin reduces cell line development time by an estimated one to two months. Moving into process development, virtual optimization of process parameters means achieving process lockdown within one to two rounds of experimentation, again saving between one and two months of time.

The next stage is scale-up or scale-down. Having already experienced dynamic data during cell line and process development, the digital twin can predict process performance at scales up to 2000 L. It can also predict the set points of process parameters to establish a scale-down model. Minimizing engineering batches in this manner is estimated to save two to three months of development time.

When it comes to process characterization, the digital twin can predict the design space of a process and minimize the number of experiments at the edge-of-failure that need validation in the wet lab. This reduces the establishment of robust normal operating ranges by an estimated three to six months. Since the digital twin has been trained with various amounts of data at all stages of the process, it can be applied for predictive model control in manufacturing. Such control delivers real-time

prediction of future trajectories of manufacturing performance; this provides proper warning of potential deviations and enables preventative action to reduce batch failure. Not only does the digital twin save time, but it can significantly impact the cost of goods and experimentation.

Re-using Data Across Projects and Scales

In the hybrid-model digital twin, the neural network sub-model (part of the kinetic cell model) is split into two parts. One part learns the behavior of the host cell line while the other retains knowledge from past projects. So, for every new project, the kinetic cell model has a different layer that will learn the specifics of particular clones. The more the digital twin is involved in projects, the higher its predictive accuracy is, and the less new data it will require for subsequent projects.

The Digital Twin – Summary

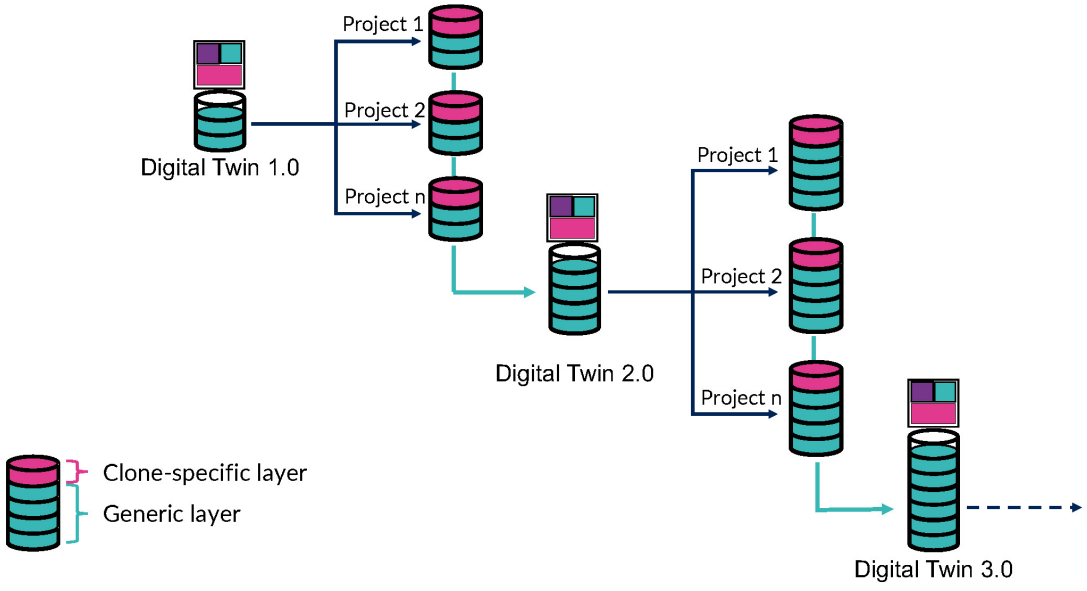
FIGURE 3 illustrates how knowledge builds as each project adds to the clone-specific data and generic layers of the twin. The digital twin's knowhow grows, reducing the need for excessive data across scales and lessening the time required for experimentation. Overall, the hybrid-model digital twin can save between eight and 15 months of time during chemistry, manufacturing, and controls [CMC] development.

PART 2: ADVANCES IN PAT-DRIVEN BIOREACTOR CONTROL AND AUTOMATION

Cell culture-based manufacturing processes are complex and have multiple critical process parameters [CPPs]. Deviation of a CPP from its optimal range can result in variability of critical quality attributes [CQAs] with potentially adverse consequences for a drug's potency, performance, and safety. Maintaining proper cell health during manufacturing requires tight process control and techniques that allow real-time monitoring of process performance.

Different spectroscopy techniques are used widely as PAT tools in upstream bioprocesses since they provide continuous real-time monitoring and can be tuned and tailored for analysis of multiple targets. Not only do spectroscopic sensors allow continuous measurement of CPPs, but they are scalable, thus ensuring consistency from

FIGURE 3: The digital twin knowledge builds-upon projects.



the lab scale right through to manufacturing. However, the use of bioreactor in-line sensors in process development and manufacturing comes with challenges and can be costly to implement because:

- Conversion and interpretation of spectroscopic data requires expertise in both spectroscopy and programming.
- Commercial solutions usually employ proprietary logic for data interpretation, raising questions of suitability for specific media and cell types.
- Sensor data must be calibrated.
- Use of PAT for process control requires cell model development. This means massive datasets with accuracy potentially limited to a particular cell line, medium, and process.
- Process control for biologics must be highly predictive (rather than reactive) because there is constant dynamic change to the existing state, and response times to control initiatives are slow.

The drive to simplify this multifaceted challenge has led to the development of a new lab scale mammalian bioreactor system [Advanced Control Bioreactor System BR1000, Yokogawa]. This reactor includes fully integrated hardware and software for implementation of upstream PAT without requiring technical or engineering expertise. It brings the benefits of PAT into a configuration that is accessible and allows users to achieve consistent results.

PAT-Driven Bioreactor Design – Capabilities and Data

The Advanced Control Bioreactor System is designed to monitor growth and energy needs, and to stabilize cell culture (**FIGURE 4**). Near infrared [NIR] spectroscopy and dielectric impedance measurement [biocapacitance] are the key technologies. It is crucial to have engineering approaches employed to work around the obstacles end users encounter when trying to use, integrate, or calibrate PAT systems. Users can still customize their own cell models for high precision and dynamic control. In part, this is due to the model predictive control software that automates the

FIGURE 4: BR1000 sensing & control.

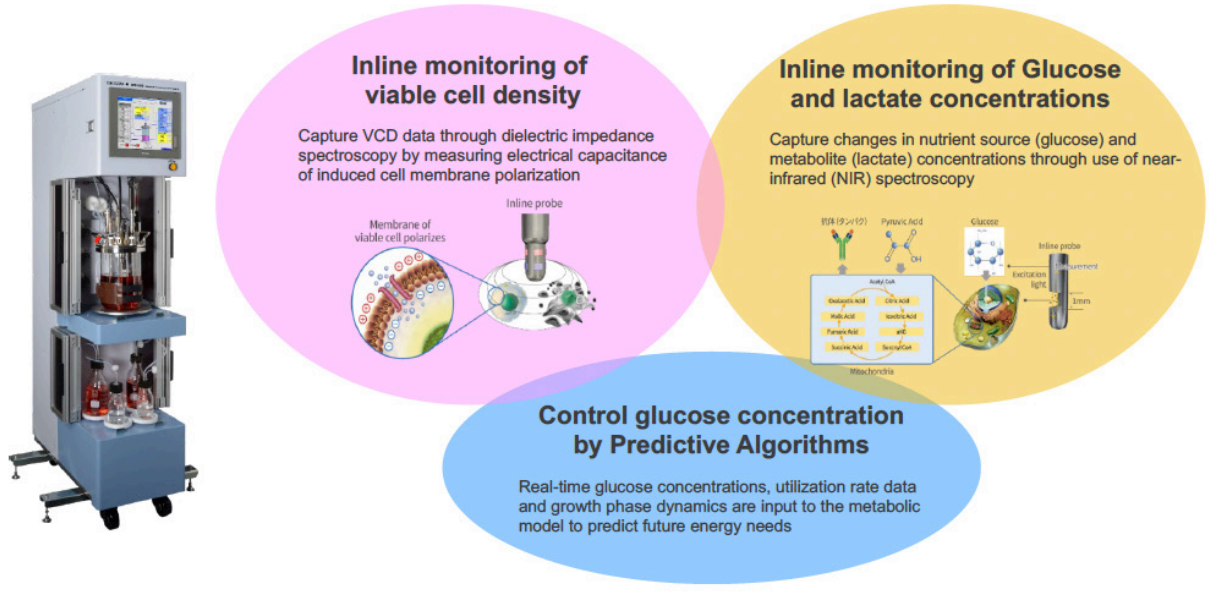
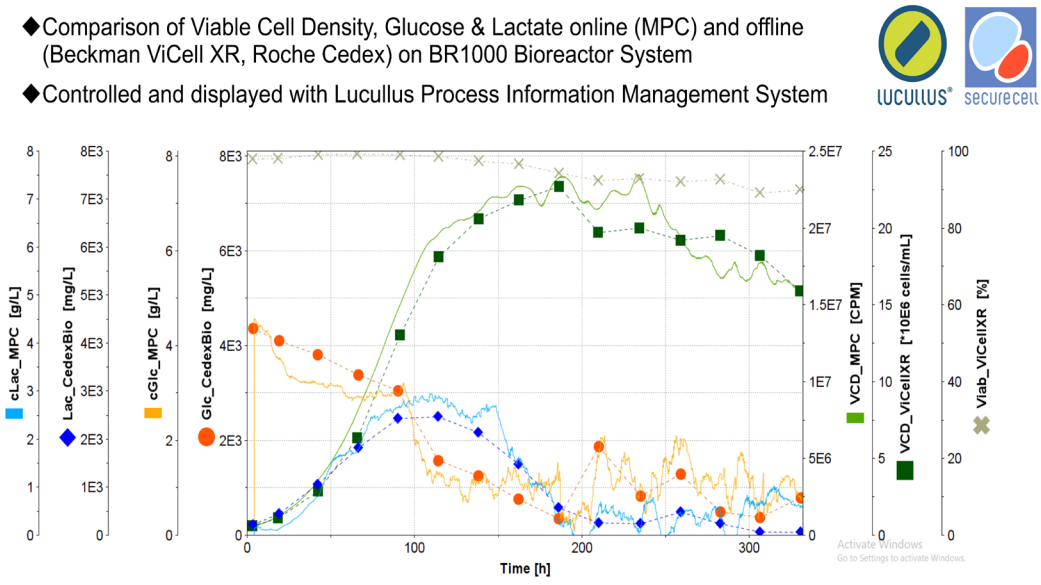


FIGURE 5: Accuracy.



glucose-feeding requirements of the cell culture. The system enables unique calculations for cell growth dynamics, such as divisions per cell per hour and glucose and lactate metabolism as picograms per cell per hour. The model predictive control uses these calculations to evaluate current and future states of the culture to control feed requirements and overall glucose concentration.

The advanced capabilities of this bioreactor are largely the result of software innovations, technology integration, and the cell culture modelling approach. **FIGURE 5** shows typical data from the PAT sensors employed. The continuous lines are real-time readings from the NIR or biocapacitance sensors, and the dashed lines connect periodic offline analytical values. These results indicate good correlation between the inline and reference values for this CHO cell culture.

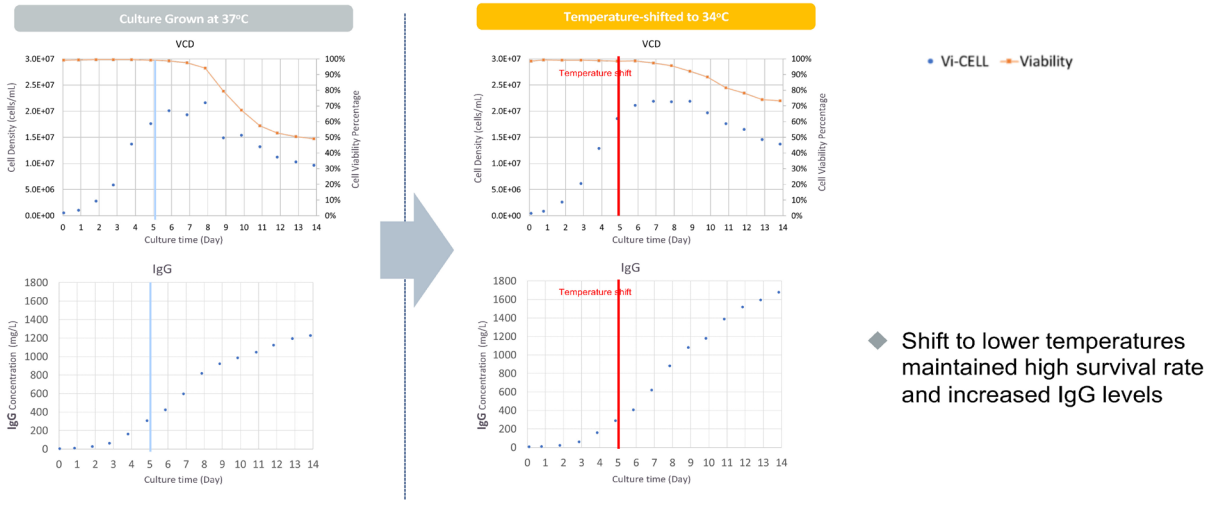
Reproducibility is always a concern for developers, process engineers, and manufacturers. This was evaluated through simultaneous testing of three identical bioreactors using the same cell line model, all monitored and controlled with Lucullus PIMS Bioreactor Software [SecureCell]. The root-mean-square error of prediction [RMSEP] value was selected to evaluate reproducibility. It measured the correlation of inline values for glucose, lactate, and viable cell density [VCD] versus offline reference measurements. Comparison of the offline RMSEP values across all three bioreactors indicated a high degree of reproducibility and fidelity in the detection and control capabilities of the systems.

The Bioreactor in Action – Examples of PAT-Enabled Applications

Accommodating low glucose requirements

Addressing the low glucose concentration requirements of some expression clones can be challenging. For example, trying to maintain a concentration of one g/L with manual operation is resource-intensive and risky, with accidental depletion having the potential to compromise the entire culture. Automated systems reduce the risk of glucose depletion, and having a model predictive control approach achieves the precision required to maintain such a low level in live culture.

FIGURE 6: Temperature shift boosts biologic yield.



Shifting temperature for better yield

FIGURE 6 shows the outcome of a temperature shift experiment with respect to IgG production using CHO cells. Since temperature affects the NIR spectroscopy readings needed for glucose and lactate determination, a new calibration model was created for the application; this is a straightforward task with the system's calibration modeler software. The data on the left in **FIGURE 6** was generated using a constant 37oC temperature throughout the cell culture run, whereas data on the right shows the results after shifting the temperature down to 34oC at day five. This shift correlates roughly to when the culture reached maximum cell density. There was preservation of VCD and cell survival in late-stage culture and elevation of IgG levels from 1.2 g/L to around 1.7 g/L, a 30% increase for this clone.

Working with perfusion cell cultures

Developing an alternating tangential flow [ATF] perfusion cell culture application required some modifications for the NIR to operate accurately at the high cell densities. This resulted in the installation of a custom-designed NIR flow cell downstream of the ATF column to enable reading of glucose and lactate levels in the cell-free media.

Exploring HEK293 and Adeno-associated Virus [AAV] production

To explore the bioreactor's suitability for growing mammalian cells other than CHO, the calibration modeler software was tested for its ability to work with HEK293 cells, whose physical characteristics and growth profile differ from those of CHO cells. Since the calibration modeler software was developed specifically with CHO, this tested its ability to deal with multiple cell parameter deviations. Work to date indicates that, overall, HEK293 growth in AAV production appears to be on a par with, or higher than, optimized shake flask cultures with respect to viral yield, head filling efficiency, transfection efficiency, and overall performance.

The PAT-Enabled Bioreactor – Summary

The Advanced Control Bioreactor System leverages software innovations to enable more robust PAT. Combined with accurate proprietary model predictive control algorithms and customizable cell models, this system adds value to bioprocess development and operation. Such integrated systems are easy to use, reduce the amount of required labor, and eliminate the need to separately assemble and validate upstream bioreactor

PAT systems. From installation, a user needs only six to eight weeks to customize and calibrate a cell model specific to their own media, supplements, cell line and process (vessel, feeding routine, etc.) by using the software tools provided with the BR1000 bioreactor unit.

REFERENCES

1. Wouters OJ, McKee M, Luyten J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. JAMA. 2020 Mar 3;323(9):844-853. Doi: 10.1001/jama.2020.1166. Erratum in: JAMA. 2022 Sep 20;328(11):1111. PMID: 32125404; PMCID: PMC7054832.