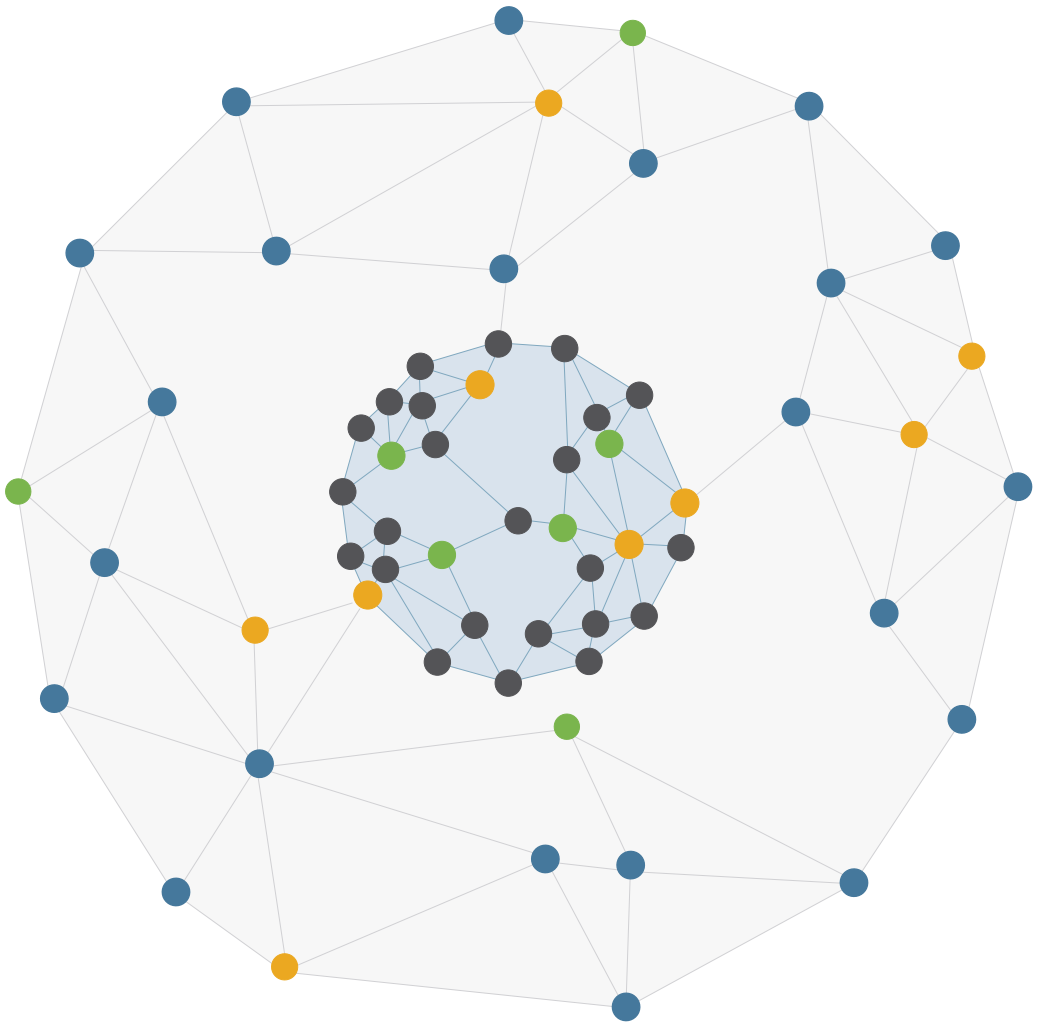
SARTORIUS + Genedata

White Paper

0100011101100100
0100011101100101011
0111001100101

Accelerate Cell Line Development Using Data Management, Modeling, and Analytics



Cell Line Development

Cell Line Development (CLD) is a fast-growing critical market projected to reach USD 2.44B by 2030 [1]. However, generating the best cell lines at scale and in the desired quality to manufacture biological products is a long and costly process, largely relying on wet lab experimentation. Huge amounts of heterogeneous data are generated at every step, but this data is often not captured in a structured way, nor analyzed using comprehensive analytical and modelling approaches. By combining the power of the structured data management backbone in Genedata Bioprocess® with the *in silico* modelling capabilities of Cell Insights by Umetrics® Studio, users can get automatic data-driven insights and significantly accelerate CLD processes (Fig.1).

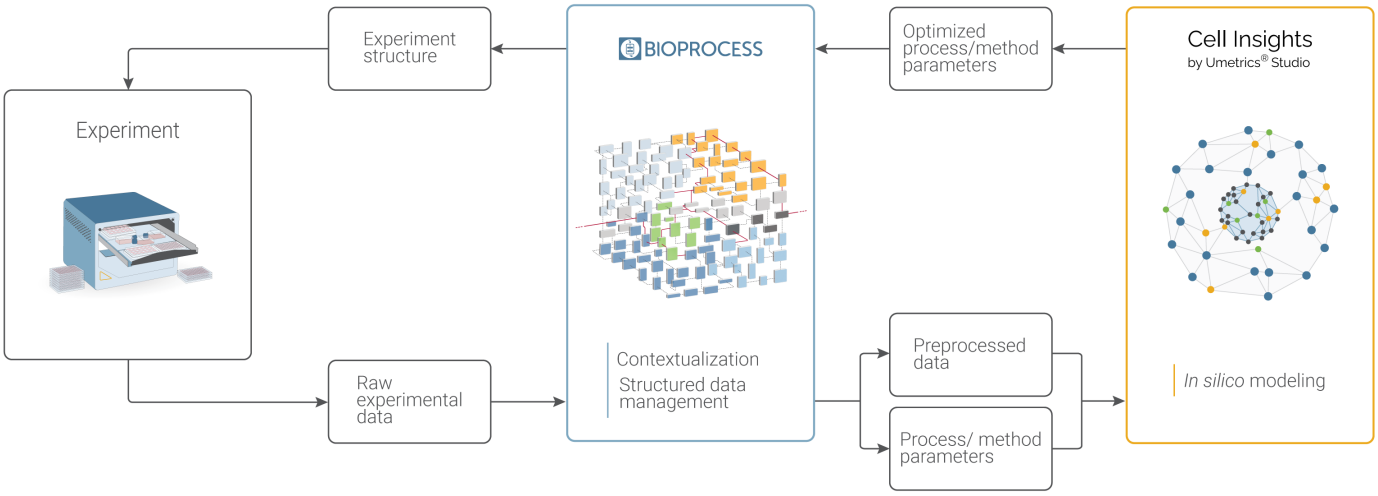


Figure 1 – Data flow between hybrid modeling and wet lab experimentation through the integration of Cell Insights by Umetrics Studio with Genedata Bioprocess. Genedata Bioprocess captures and contextualizes process parameters and experimental data and passes it on for use in modeling. The improved process parameter set calculated with Cell Insights by Umetrics Studio is then directly used to generate representations of planned experiments in Genedata Bioprocess. This closes the loop between lab and model and drastically reduces the effort in data transformation and communication.

Process Modeling:
A Systematic Approach to Improvement

Process intensification is a strategy that can improve efficiency and lower the cost of goods in CLD [2]. It offers options such as seed-train intensification (usually at the N – 1 stage), concentrated fed-batch production, and dynamic perfusion (at the production bioreactor stage) [3]. These options can be assessed before selecting the final production strategy. Understanding the initial potential of each clone using different intensification strategies can be key to selecting the optimal clone and process. However, the large number of different clones available early in development makes it impossible to test all combinations with different process strategies during CLD.

Modelling processes can simulate the performance of different clones in an intensified process without the need to exhaustively test each clone in the laboratory. It is challenging to develop a holistic model explaining all processing strategies without a detailed understanding of intra-cellular kinetics of the cell line. However, as shown by Richelle et al. [4], a straight-forward model with some pre-configured generic assumptions can sufficiently simulate intensified processes even with limited data. To enable automation of model-building without expert knowledge, the model structure requires limited inputs and offers configurable options in a standardized workflow. Better insights based on simulated clone behavior can reduce the time needed for decision making, and significantly speed up the development of intensified processes.

Three Key Challenges on the Way to a Model

When it comes to automated, rapid, and effective modeling, we see three obstacles:

1.

The accuracy of predictive models is limited by the completeness and accuracy of the data being used. However, data is often not easily findable across organizations and data scientists can spend 50 to 80% of their time gathering, structuring, and integrating relevant data for each project [6]. The reason often lies in the way that data is stored from the get-go. ELNs provide only limited structure and rely on creation of individualized data models by each group. This introduces variability and disconnect in data capture across organizations, making comparable data harder to find and access. All in all, in the long run, quality, completeness, and interoperability of data are problematic when not stored in a structured, standardized, and enforced format.

2.

Sharing the correct data with the right person in the right way is very challenging since the same person usually does not perform both the experiments and modeling of the data. Data modelers in turn must convey changed parameters and protocols for follow-up experiments based on their modeling efforts. Manual data and information handovers between lab scientists and modelers can be error-prone and result in poor data quality and development delays.

3.

Modeling is often considered a project-specific or one-time effort. This prevents modeling from having the maximum impact it could have when regularly applied with the most current data available. A major obstacle is often the perception that one needs a holistic model that can fit every possible scenario and lots of data for each clone. Practical application of less demanding and more targeted models to solve specific use cases (e.g., the process mode switch) significantly simplifies modeling. Such models can be automatically generated (e.g., at the end of each clone screening experiment) and the resulting simulations can be made easily available to a larger user group for the biggest effect.

A Direct Path from Experiment to Model

A simple way of overcoming obstacles via an automated method uses both data and process knowledge to generate a mathematical model of a process.

The integration of Genedata Bioprocess with Cell Insights by Umetrics Studio enables modelling that is seamlessly integrated into the workflow. It removes the extra effort of transferring data to the modeling tool and creating protocols and work instructions for lab scientists. This method can be used to simulate behaviors under intensified process conditions and enables easy ranking and selection of optimum clones based on both the actual and the predicted behavior. By providing a single comprehensive end-to-end data model, data cannot get lost or mislabeled and data entry is harmonized across the entire process. Data is automatically structured based on underlying biological and experimental conditions, making it possible to mine the data and make data-based decisions (Figure 2).

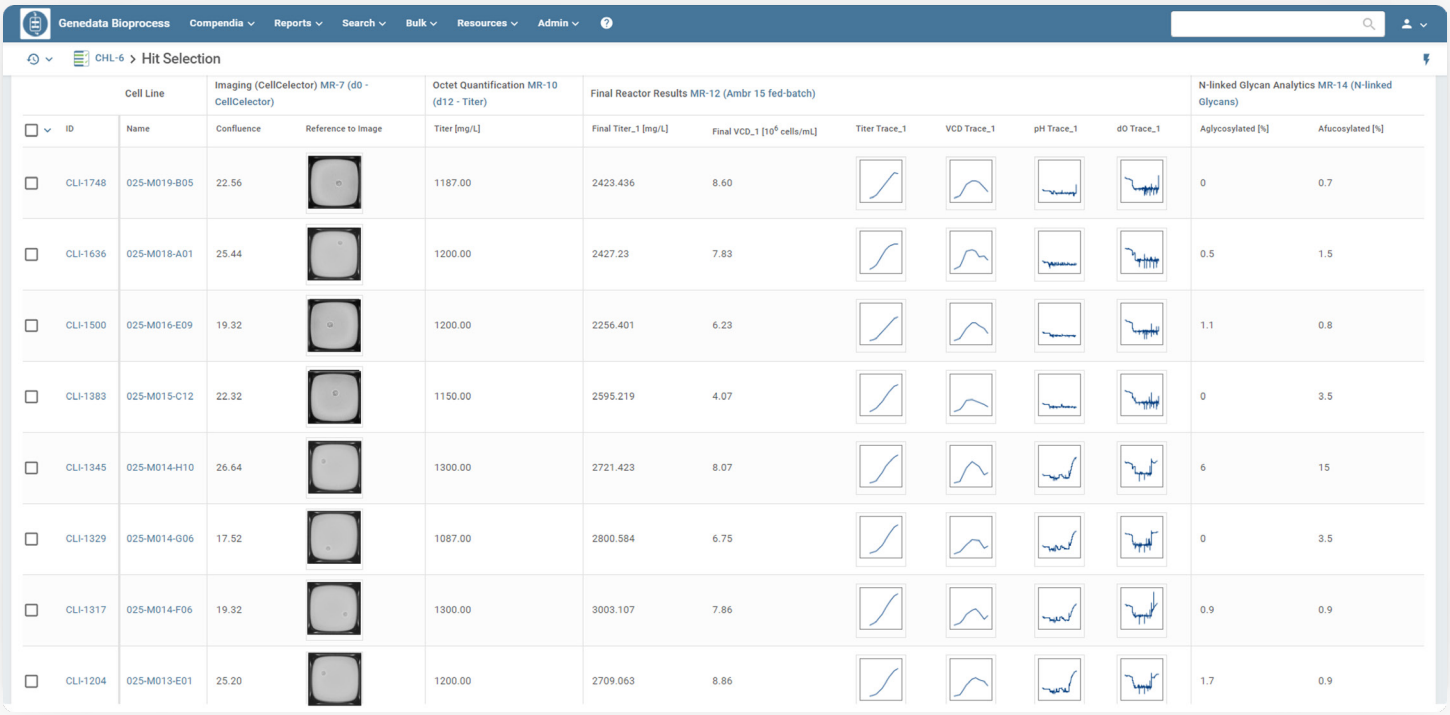


Figure 2 – Genedata Bioprocess CLD hit selection dashboard. Genedata Bioprocess represents a structured workflow backbone for bioprocess development, integrating and automating CLD, upstream process (USP), downstream process (DSP), formulation and analytical development (AD) workflows. This screenshot shows a panel of stably expressing cell lines, with each row representing a distinct transfected cell line. The columns represent characterization data for the various clones, such as product titer, product quality and monoclonality information. Genedata Bioprocess automates the capture from all CLD steps and aggregates the data generated by the various laboratory systems, including CellCelector, Octet, Ambr bioreactors, and MS-based glycan analytics product information, as shown in the table above. By integrating all relevant CLD data, the system enables an integrated data-driven decision-making to identify the optimal manufacturing cell line for a given drug product.

For instance, a process model based on fed-batch data can be used to simulate the performance of different clones in an intensified process mode. The first step to generate such a model is to select the processes to model and the relevant data. In this example, viable cell density and viability for one batch process of different clones were selected, and the available and correctly formatted data was automatically exported from Genedata Bioprocess into Cell Insights by Umetrics Studio.

The model to simulate intensified strategies based on fed-batch processes needs correct contextualization in addition to the data for viable cell density, viability, and optionally product concentration available for each clone, which can be obtained from Genedata Bioprocess on demand and made available to Cell Insights (Figures 3 and 4).

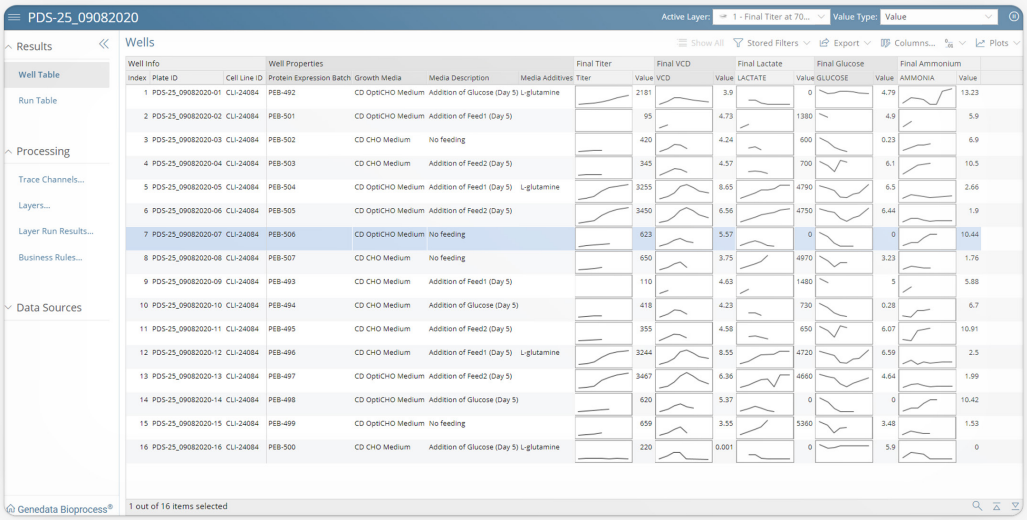


Figure 3 – Genedata Bioprocess is designed to capture and process time resolved data on bioreactor experiments from a variety of offline and online sources. Any necessary preprocessing of data can be fully automated before the data is passed on to Cell Insights by Umetrics Studio.

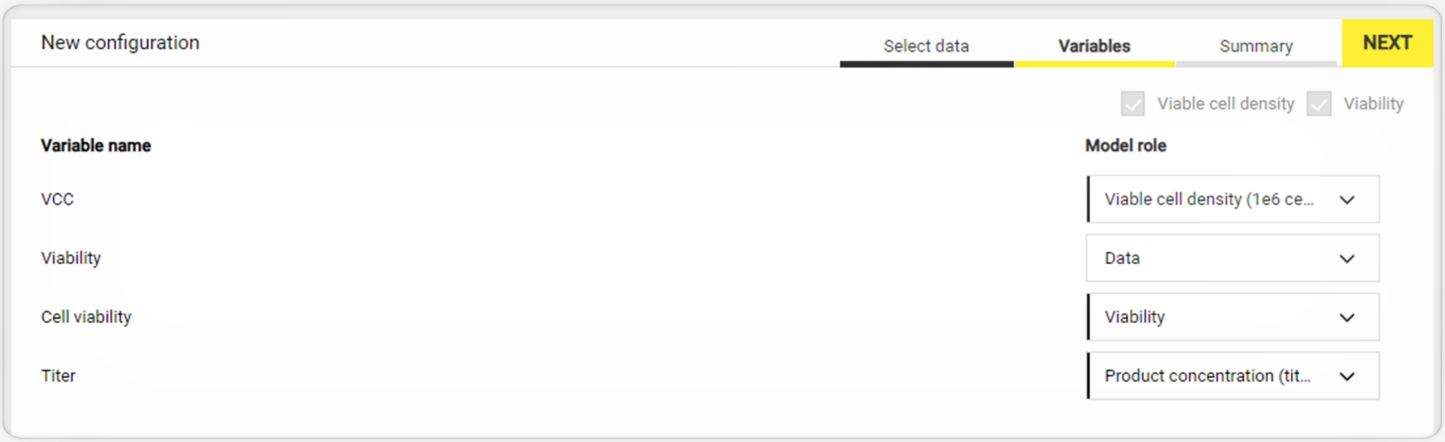


Figure 4 – Cell Insights by Umetrics Studio utilizes out of the box mechanistic cell models to simulate and predict the behavior of cells *in silico*. To get started, contextualize data for viable cell density, viability, and optionally product concentration.

An individual model for each clone is fitted in Cell Insights by Umetrics Studio. The model is represented by the straight line and the measured data by the dots. The closer the match between the line and the dots, the better the model describes the reality (Figure 5).

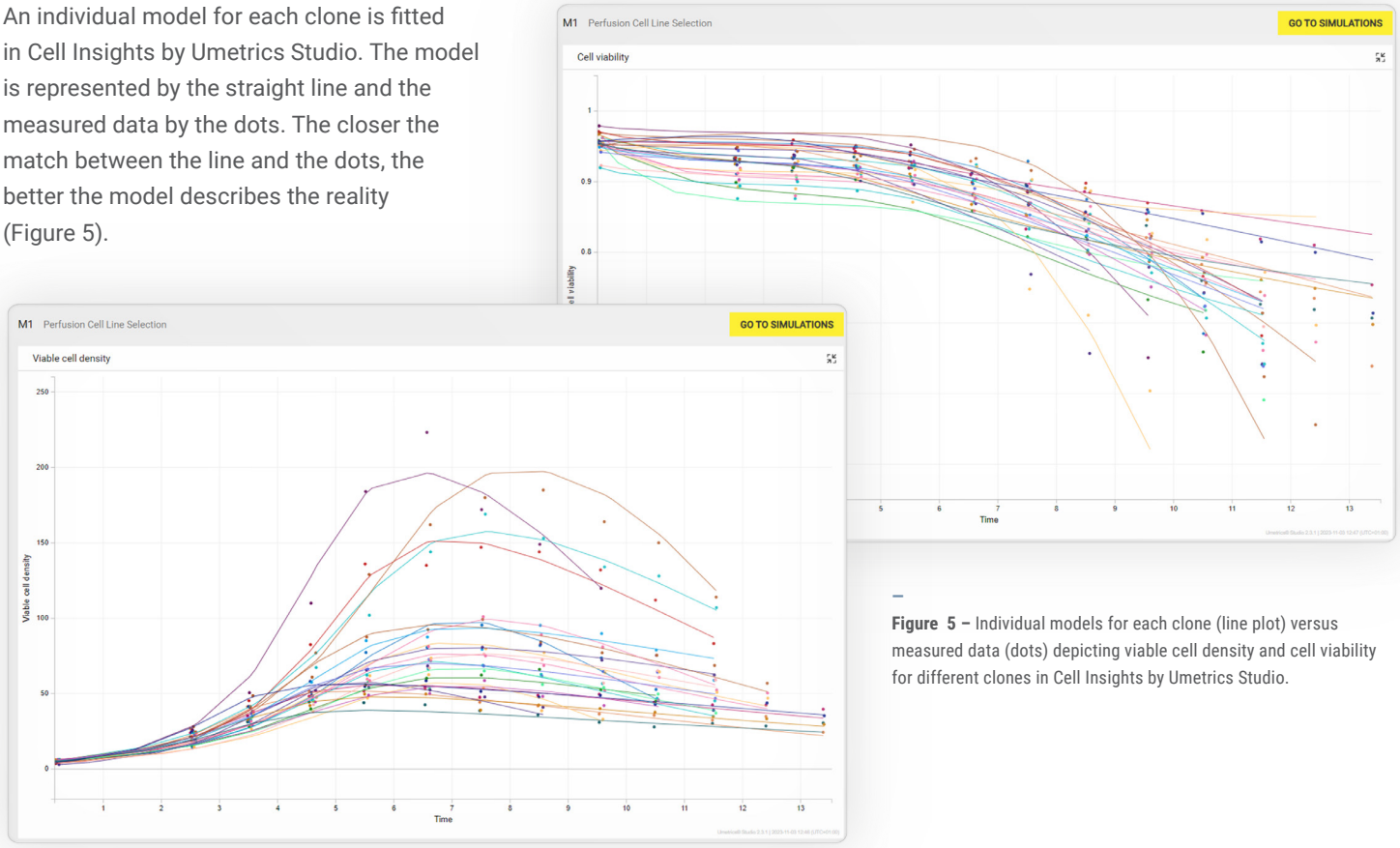


Figure 5 – Individual models for each clone (line plot) versus measured data (dots) depicting viable cell density and cell viability for different clones in Cell Insights by Umetrics Studio.

Based on a good fitting model, new process conditions can be simulated. In this case, we show the different clones' performance in a perfusion-like set-up (Figure 6). The maximum viable cell density that can be achieved, or the media needed to achieve a certain viable cell density, can be used as decision criteria to select different clones when manufacturing a master cell bank.

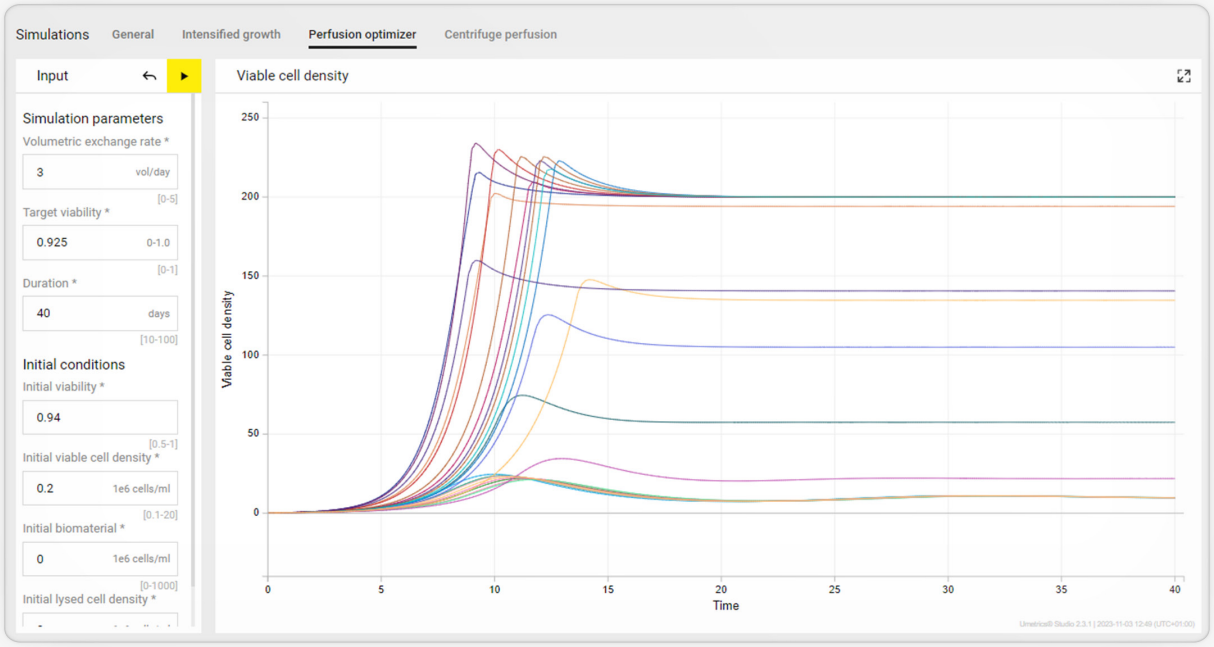


Figure 6 – Efficient modelling of different clones' behavior using simulated parameters in a perfusion cultivation system in Cell Insights by Umetrics Studio.

The entire set of concluded parameters for an optimized process can then be automatically loaded back into Genedata Bioprocess. There, they are immediately available for confirmation runs as well as uploading of all measured data thanks to a fully integrated data flow between Cell Insights by Umetrics Studio and Genedata Bioprocess.

Conclusion

Efficient modelling of CLD processes can provide verifiable insights throughout the entire workflow. It can be sustainably implemented and easily assess every clone across different intensification processes. The presented approach demonstrates the power of mechanistic modeling using complete, structured, and fully traceable data to optimize processes more quickly. Furthermore, it shows that the integration of Genedata Bioprocess and Cell Insights by Umetrics Studio ensures the highest quality and contextualization without the need for manual data collection or time-consuming data processing.

Sources:

[1] https://www.europeanpharmaceuticalreview.com/news/159967/cell-line-development-services-market-to-value-2-4-billion-by-2030/

[2] Process intensification in the biopharma industry: Improving efficiency of protein manufacturing processes from development to production scale using synergistic approaches, Müller et al, 2022, chemical engineering and processing – process intensification

[3] https://bioprocessintl.com/sponsored-content/a-strategic-approach-to-selecting-the-optimal-process-intensification-scenario/

[4] Model-based intensification of CHO cell cultures: One-step strategy from fed-batch to perfusion, A. Richelle et al, 2022, Frontiers in Bioengineering and Biotechnology

[5] https://www.anaconda.com/resources/whitepapers/state-of-data-science-2020

Genedata BIOPROCESS

Genedata Bioprocess® is part of the Genedata portfolio of advanced software solutions, which digitalize and automate data-rich and complex biopharma R&D processes. From early discovery all the way to the clinic, Genedata solutions maximize the ROI in R&D expenditure.