Next-Generation Bioprocess Optimization

An ArrayXpress White Paper
August 2014

Pharmaceutical companies are making significant investments in biologics and dedicating up to 40% of their R&D efforts into their biopharmaceutical pipelines in lieu of classical, small-molecule drugs (Rader, 2013; Aggarwal, 2014; Evaluate Pharma, 2013). Experts forecast continuous strong market growth with increasing revenue reliance and contribution to gross margin. Within the top 100 pharmaceutical products, biologics are expected to account for more than 50% of prescription sales by 2018 (Evaluate Pharma 2013; Figure 1). With the increased sales in biologics, dramatic improvements have been required throughout the manufacturing process. Over the past several decades, titers have jumped more than a 100-fold, from sub-single digit yields (in g/liter) to today's double-digit production levels. Early gains in production capacity were achieved simply by using larger bioreactors. Smaller incremental gains resulted from process optimizations in which higher cell density, viability, increased product expression levels, and higher specific productivities were gradually achieved.

At the same time, initiatives implemented by the regulatory agencies such as Quality by Design (QbD) and Process Analytical Technology (PAT) led to improved manufacturing processes and product quality. The primary objective of these initiatives was to direct the industry away from the empirical in-process development

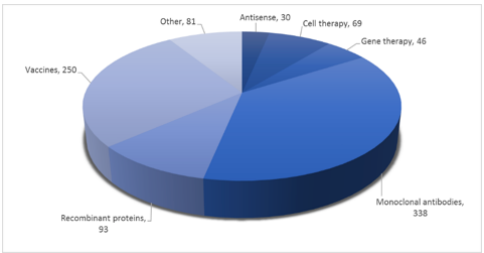

Figure 1. Biologics in development organized by product category. Data obtained from PhRMA, 2013.

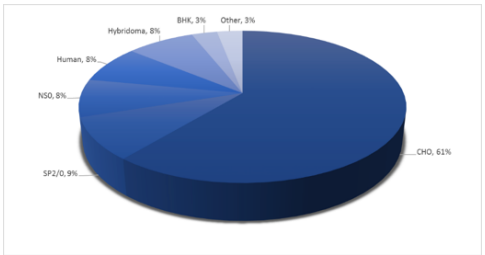

Figure 2. Mammalian cell lines used in biomanufacturing by number of biologics (until 2012). Modified from Kantardjieff and Zhou, 2014.

approach. The goal now is to build in quality starting at the design stage (Glassey, 2011). This approach relies on integrative systems and data-driven methods that contribute to the understanding of biomanufacturing processes, and where critical process parameters are identified, monitored, and controlled. Ultimately, the goal is to develop processes that are predictable, consistent and ensure high product quality and titers (FDA, 2004).

The need to innovate

Biomanufacturing performance is determined by the interaction of the BIO and MANUFACTURING components. While significant progress has been accomplished for the latter, including physical production systems (bioreactors), media formulations, and process optimization strategies, much less effort has been dedicated towards the BIO component.

Because of their molecular complexity and unique quality attributes, most biologics require complex production systems. Mammalian cell lines are ideally suited for this purpose due to their ability to generate complex human-like glycan profiles and other post-translational modifications that are critical for product efficacy and safety. Currently, 51% of all approved biologics are manufactured in mammalian cells, including 83% of all recombinant blood factors, 95% of all monoclonal antibodies, and 74% of all other recombinant products (Kantardjieff and Zhou, 2014). Chinese hamster ovary (CHO) cells have become the cell line of choice and the industry's workhorse, primarily due to their proven safety record and adaptation to high-density suspension growth. CHO cells are currently responsible for producing over 60% of all mammalian cell-based biologics (Kantardjieff and Zhou 2014; Figure 2). Other mammalian cell lines are used to a less extent, including baby hamster kidney, mouse myeloma cell, and human cell lines.

The inherent complexity of biological systems is the primary contributor to biomanufacturing process variability and inconsistency. Process inconsistencies are commonplace and cannot support the expected growth in market demand nor the economic and regulatory challenges faced by the industry. The biopharmaceuti-

cal industry can greatly benefit from technological innovations that drive rapid and adaptive change, provide competitive advantage and allow it to focus on efficiency, flexibility, convenience, and quality (Rader, 2013; Langer, 2012; Gottschalk, 2012; Gottschalk, 2013; Davidson and Farid, 2014; Carinhas, 2012).

Understanding the system

Current methods for cell line development and process optimization are very time consuming, expensive, and labor intensive and only lead to incremental improvements. Commonly used approaches in cell line development include gene amplification strategies and selection of stable, high expressing clones. Traditionally, process optimization for media and cell culture conditions is achieved either through Design of Experiment, experience, or trial-and-error approaches (Zhang, 2011). Most importantly, this practice must be repeated for every new production cell line and associated protein product. At best, it results in a highly variable and unpredictable process, both in terms of productivity, as well as product quality. These heuristic approaches lack the mechanistic understanding of how and why process conditions, or any implemented changes, bring about the desired outcome.

It is not possible to fully understand the processes without considering the cell lines used and their relationship to the products they synthesize. An understanding of the BIO component, that is, the intracellular processes relevant to biomanufacturing including protein translation, post-translational modifications, folding, aggregation, trafficking, and secretion is key to overcoming the inconsistency and variability. Without this knowledge, any optimizations that lead to production gains observed for one cell line are not likely transferrable to another and could not be fully implemented across the entire production portfolio. Developing an in-depth understanding of the biology of these production cell lines is critical for sustained biomanufacturing.

'Omics and Systems Biology

'Omics-based technologies rely on the generation and interpretation of high-throughput data from an organism's DNA, RNA, proteins and metabolites (Figure 3). These technologies are commonly used to discover novel targets for therapeutics, identify biomarkers, pharmacogenomics,

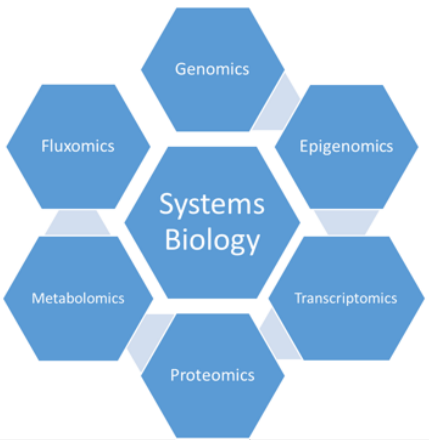

Figure 3. Systems Biology employs a suite of 'Omics technologies to decipher the functional dynamics and interactions across the various celluar organization levels: DNA, RNA, proteins, and metabolites.

personalized medicine, and for disease diagnosis and classification. Although each has its own application, individual 'Omics-technologies do not provide a holistic view or capture the complex interactions occurring within the cells. With advances in processing and computational capacity, the concept of Systems Biology has evolved by combining the individual 'Omics technologies. As a result, we can now view cellular systems as a complex network with intricate interactions across their distinct organizational components that define the cells (Oltvai and Barabasi 2002; Kitano, 2002). More than merely visualizing these interactions, Systems Biology aims to model and understand the structure and dynamics of the cells' functional networks. The increasingly rapid release of draft genomes for various mammalian systems, including the recent CHO genomes, is expected to further drive the incorporation of systems biology in bioprocessing (Xu et al., 2011; Brinkrolf et al., 2013; Lewis et al., 2013).

Integrated Cellular 'Omics Platform - iCOP™

From the need to understand the production system and technological advances in Systems Biology, ArrayXpress developed iCOP – an integrated Cellular 'Omics Platform for Bioprocess Optimization (Figure 4). iCOP was designed using the principles of Systems Biology to enable the systematic and directed engineering of production cell lines. Consistent with the QbD philosophy, iCOP is fundamentally a data-driven approach to bioprocessing. The ultimate goal is to increase product titer, quality, and achieve predictable and stable biomanufacturing, resulting in Next-Generation Bioprocessing.

iCOP is comprised of two interdependent components: an Integrative Systems Biology component and an Engineering component that drive the Next Generation Bioprocess. The first component relies on system-wide molecular characterization of the cell's functional dynamics that is then used to direct cell line development and process optimization efforts.

Integrative Systems Biology Component

The task of generating systems-level data was dramatically simplified and became cost effective by improvements in the latest generations of 'Omics technologies. The major obstacle now is the analysis and subsequent biological "data mining" of

these very large and highly dimensional datasets. Perhaps even more critical when dealing with Systems Biology is the fact that these various 'Omics datasets are generated from inherently different technologies and possess intrinsic characteristics that must be properly accounted for using sophisticated statistical and bioinformatics methods. Integration of the various 'Omics datasets and their subsequent interpretation are computationally intensive processes that require highly skilled expertise across various fields in biology and bioinformatics. Rather than merely looking for simplistic linear relationships and drawing static interconnecting diagrams, data integration in Systems Biology aims to reveal and define the basic governing principles that describe the functional dynamics of the cells (Kitano, 2002). With iCOP, ArrayXpress has developed proprietary data analyses pipelines for the successful integration of these various datasets to derive process relevant insights.

The culmination of several iCOP iterations is the creation of an in silico cellular model. The model is constantly tested and refined during the process to ensure that it is robust, accurate, and can be used for prediction of cellular behavior in culture. It is a data driven process-optimization based on mechanistic insights of the production host physiology and metabolism. Cellular and process targets are identified for immediate optimization and capitalized on in the Engineering component.

The Engineering Component

Based on the empirical understanding of the cell, production processes, and corresponding aberrations, the genes and proteins related to and regulating the key metabolic or cellular phenotypes can be targeted. These targets are manipulated through engineering strategies that fall into four main areas: genetic, cellular, metabolic, and process engineering.

In genetic engineering transcriptomic data may be used to identify transcriptional hotspots to support high levels of transgene expression. Site-specific integration has been heralded by many and is highly pursued by the biomanufacturing industry. Genetic information may be used to design site-specific transgene integration strategies using engineered nucleases such as zinc-finger nucleases (ZFNs), CRISPR/Cas9-based systems, and transcription activator-like effector nucleases (TALENs). The limited availability of a well-characterized genome, such as in CHO, imposes some obvious limitations to this approach.
Intracellular processes that are key to recombinant protein production and quality, such as post-translational modification and protein folding and secretion, are optimized using cellular engineering approaches. For example, by manipulating specific genes specific glycan profiles can be achieved leading to improved product safety and efficacy. Likewise, increased protein folding and secretory capacities can be improved by the manipulation of genes involved in these processes and leading to higher titers. Similarly, other production-relevant cellular processes including apoptosis and cell growth can be manipulated to increase cell viability and product titers. Metabolic by-products such as lactate and ammonia are a common problem in mammalian cell culture that adversely impact production and are a prime engineering target. In some instances, cultures are able to efficiently shift from lactate production to lactate consumption without adverse effects. The ability to accomplish this transition is associated with increased titers. Similarly, redox homeostasis and carbon utilization are critically important for productivity and quality. Adverse metabolic reactions related to any of these processes can be characterized and identified in the Systems Biology component and targeted for modification in order to optimize central metabolism efficiency and maximize cellular productivity.
Furthermore, improved process understanding, as outlined by QbD, requires the comprehensive and integrated analysis of process data and phenotypic cellular-level data. ArrayXpress' data analysis pipelines were designed to allow for the integration of these intrinsically distinct but relevant datasets. This integration of PAT process data with cellular-level molecular data allows for the identification of early biomarkers and process parameters indicative of final production phenotypes. These biomarkers and process parameters can be implemented in biomanufacturing as early process indicators that can be monitored in real-time for process monitoring, quality control, and troubleshooting. iCOP therefore offers the distinct advantage to enable cell development scientists and process engineers to work synergistically, accelerating and streamlining cell line development and process optimization.

iCOP: eliminating the black box

Currently, process optimization relies on data derived mainly from the cells culture environment leaving the biological component as a "black box". In contrast to this traditional "black box" approach for process optimization, iCOP process engineering is primarily driven by the biological knowledge accumulated from the Systems Biology component. Implementation of iCOP is an iterative process in which experiments are designed to answer a specific question. The workflow is characterized by two main phases: a

data phase and an integration phase, each comprised of three steps (Figure 4). The generation of an empirically based novel hypothesis reinitiates the cycle leading to continuous process improvements.

The first step in the data phase is defining the project and hypothesis. The ArrayXpress team will work with you to define a list of prioritized goals. The experimental design reflects the hypothesis in question and takes into consideration the type of comparison (e.g., time, media, and production rate), sample replication, and the 'Omics technology to be used. The cumulative years of experience in 'Omics experimental design and the partnerships with academic and private leaders in this field are distinct advantages that ArrayXpress brings to our clients.

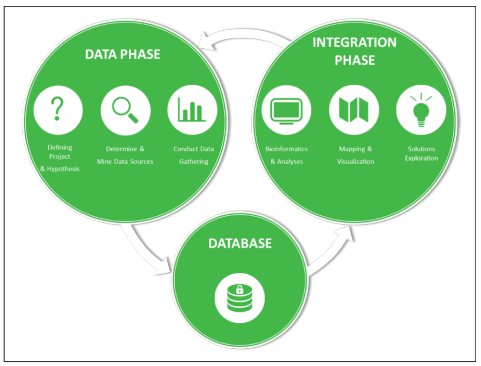

Figure 4. iCOP iterative workflow: during the data phase, empirical data is aggregated and stored in the client's proprietary database. These formerly disparate 'Omics datasets can then be mined during the integration phase, taking into consideration published literature, to derive process relevant insights.

Next, availability of existing public, proprietary, or the client's internal data is determined. The goal is to identify existing client capabilities and avoid duplication while maintaining complementation that can be capitalized on for maximizing resource allocation. As part of this process, a knowledge management and sharing system is outlined and set in place in the first iteration of the iCOP workflow, which is added upon in subsequent projects.

The last step in the data phase is data gathering. Any proprietary pre-existing dataset is obtained for data analysis and integration. Based on the pre-determined experimental design and project specifications, ArrayXpress will utilize state-of-the-art 'Omics technologies to generate any required new data. The data generated or obtained in this phase is deposited in a client specific database. This database becomes the exclusive property of the client, becoming part of your institutional knowledge management system. ArrayXpress follows IT industry best practices for data storage, security, and management.

In the Integration Phase, bioinformatics and data analyses are conducted using ArrayXpress' proprietary statistical and bioinformatics pipelines for initial data assessment, quality control, mapping and annotation. Differential expression and multivariate data analyses follow, along with pathway analyses and systems network reconstruction. These tasks requires careful consideration of appropriate statistical and bioinformatics approaches.

As the second step in the integration phase, the individual 'Omics datasets and the integrated data are visualized using tools that allow the client to display a system-wide snapshot of their data (Figure 5). From the display, data may be explored for a particular interest such as regions identified by differential expression that are correlated, either functionally or empirically. We customize this visualization tool for the different strata of biological data based on the specific study needs. Ultimately, novel functional relationships that govern cellular phenotypes are revealed and their relevance are further investigated

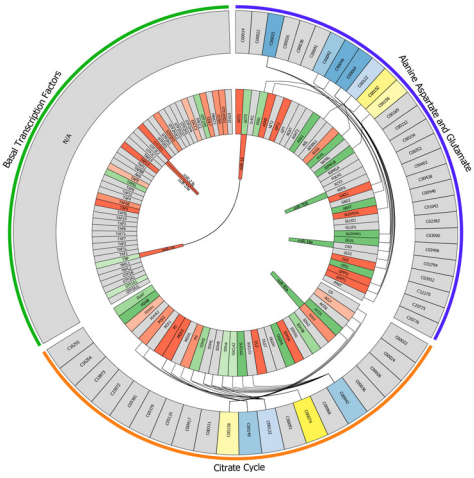

Figure 5. Summary representation of iCOP data analyses. The experiment showed three key enriched pathways represented by the outer most band. All the metabolites associated with the pathway are displayed in the next band and their respective differential concentrations shown as a heatmap with fold changes ranging from -1 (yellow) to 0 (gray) to 1 (blue). Constituent genes of each pathway are presented in the next band with the differential expression shown as a heatmap with fold changes ranging from -4 (green) to 0 (gray) to 4 (red). Metabolites at statistically different levels are connected to genes that are directly involved in the metabolic reaction. The inner-most band represents significant miRNA that are aligned with their most highly probable target gene. miRNA connected by a line denotes the same miRNA with multiple gene targets. One significant pathway, Basal Transcription Factors, consisted of only a network of transcription factors and therefore no metabolites are shown. By only analyzing the metabolic dataset and not examining transcriptomic and the miRomic datasets, the effect of the transcription network would have been overlooked.

during the Solutions Exploration step. In the final step, the integrated data is mined and queried using pathway and network analyses software tools to identify process targets and/or biomarkers that can be used for improvement and real-time process monitoring. Deliverables include the client's proprietary iCOP database and a toolkit for enhanced cell line development and process optimization that can be immediately capitalized on for accelerated market release. Depending on the initial hypothesis and experimental design, this step can provide solutions to very specific challenges and be immediately translated into short-term process improvements. Alternatively, the hypothesis and experimental design may provide a solution to an over-

arching strategy for cell engineering to address a specific cellular or metabolic pathway. All iCOP iterations generate relevant knowledge, leading to an improved understanding of the process and direct the design of hypotheses for the Next Generation Bioprocess.

The execution of the engineering strategies is accomplished in partnership with the client resulting in the creation of a collection of engineered cell lines, which encompasses the client's integral IP, and is optimized for the client's specific production portfolio.

Conclusion

The recognition of the inherent complexity of biological systems drives the need for a holistic understanding of the production cells' biology. Systems Biology is becoming an integral component of bioprocessing optimization and leads to large performance improvements in the areas of cell line and process development. The biopharmaceutical industry has been observing from the sideline with a few forward thinking companies using 'Omics to address biomanufacturing issues such as production titers and stability, with very few attempts at using an integrated Systems Biology approach (Glassey, 2010; Carrondo, 2012; Estes and Melville, 2014).

iCOP is an integrated Systems Biology solution with an encompassing, comprehensive scope. Its foundation is based on the fact that genetically controlled metabolic and regulatory pathways determine product quantity and quality and can be manipulated to improve process performance and efficiency. By establishing a collection of well characterized cell lines with known, predictable, and stable phenotypes, iCOP provides a stable cell platform for biologics production that accelerates time-to-clinic both in trials and final market release at full scale manufacturing.

For more information please visit www.ArrayXpress.com

References

Aggarwal, S. What's fueling the biotech engine – 2012 to 2013. Nature Biotechnology, 2014, 32(1): 32.

Brinkrolf, Karina, et al. "Chinese hamster genome sequenced from sorted chromosomes." Nature biotechnology 31.8 (2013): 694-695.

Carinhas, N et al. Systems biotechnology of animal cells: the road to prediction. Trends in Biotechnology, 2012, 30(7): 377.

Carrondo, MJT et al. How can measurement, monitoring, modeling and control advance cell culture in industrial biotechnology? Biotechnology Journal, 2012, 7: 1522.

Clarke, C et al. CGCDB: a web-based resource for the investigation of gene coexpression in CHO cell culture. Biotechnology and Bioengineering, 2012, 109(6): 1368.

Davidson, A and Farid, SS. Innovation in biopharmaceutical manufacture. BioProcess International, 2014, 12(1): 12.

Estes, Scott, and Mark Melville. "Mammalian Cell Line Developments in Speed and Efficiency." (2013): 1-23.

EvaluatePharma. World Preview 2013, Outlook to 2018. Available at: www.evaluategroup.com/wp2013.

FDA. "Guidance for industry PAT—a framework for innovative pharmaceutical development, manufacturing, and quality assurance." DHHS, Rockville, MD (2004).

Glassey, J et al. Process analytical technology (PAT) for biopharmaceuticals. Biotechnology Journal, 2011, 6(4): 369

Gottschalk, U et al. The need for innovation in biomanufacturing. Nature Biotechnology, 2012, 30(6): 489.

Gottschalk, U. Biomanufacturing: time for change? Pharmaceutical Bioprocessing, 2013, 1(1): 7.

Hammond, S et al. Chinese Hamster Genome Database: An Online Resource for the CHO Community at www.CHOgenome.org. Biotechnology and Bioengineering, 2012, 109(6): 1353.

Kantardjieff, A and Zhou, W. Mammalian Cell Cultures for Biologics Manufacturing. Advances in Biochemical Engineering Biotechnology, 2014, 139: 1.

Kitano, Hiroaki. Computational systems biology. Nature 420.6912 (2002): 206-210.

Langer, E. A decade of biomanufacturing. BioProcess International, 2012, 10(6): 65.

Lewis, Nathan E., et al. "Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome." Nature biotechnology 31.8 (2013): 759-765.

Mandenius, CF et al. Quality-by-Design for biotechnology-related pharmaceuticals. Biotechnology Journal, 2009, 4: 600.

Oltvai, Zoltán N., and Albert-László Barabási. "Life's complexity pyramid." Science 298.5594 (2002): 763-764.

PhRMA. Medicines in Development, Biologics 2013. Available at: http://www.phrma.org/sites/default/files/pdf/biologicsoverview2013.pdf.

Rader, RA. FDA biopharmaceutical product approvals and trends in 2012. BioProcess International, 2013, 11(3): 18.

Zhang, J. Mammalian cell culture for biopharmaceutical production. (2011) p. 157.

Xu, Xun, et al. "The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line." Nature biotechnology 29.8 (2011): 735-741.