# Combining Machine Learning and Traditional Experimental Approaches to Streamline Bioprocess Development for Recombinant Protein Production in *Komagataella phaffii*

by Brittney C. Sunday

Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY IN CHEMICAL ENGINEERING

The biosynthesis of value-added products via cells in cultivation is a burgeoning and far-reaching industry with applications ranging from biopharmaceuticals to diagnostics to food additives to industrial processing agents and beyond. To meet the growing demand for these products, rapid and efficient development of productive, cost-effective, scalable biomanufacturing processes is necessary. Yet, the development and optimization of biomanufacturing processes suitable for commercial production is an arduous task, in part due to the complexity of the associated organisms and in part due to the diversity of physical systems it traverses. This thesis presents a streamlined approach to bioprocess development using both emerging (machine learning) and traditional experimental approaches. Using production of recombinant proteins for biopharmaceutical applications in *Komagataella phaffii*, a yeast species with great promise for laboratory- and industrial-scale production of high-quality recombinant proteins, this work demonstrates how the individual advantages of machine learning and traditional approaches make them uniquely poised to work in tandem for comprehensive bioprocess development lifecycles, encompassing optimization, transition across cultivation systems, and scale-up.

First, we explored the application of machine learning algorithms to guide the efficient optimization of cell cultivation process conditions at the small (multiwell plate) scale. Through the novel coupling of process-constrained batch Bayesian optimization with transfer learning via a multioutput Gaussian process model, we improved the cultivation outputs 3-fold and demonstrated meaningful reductions in the experimental burden compared to alternative experimental approaches. Next, we evaluated approaches for transitioning cultivation protocols across different bioprocessing systems, specifically from shake flasks to laboratory-scale bioreactors. Here, to minimize the experimental burden at the bioreactor scale, we focused on effective, traditional approaches grounded in monothetic analysis and physically based scaling laws. Finally, we demonstrated successful application of the combined, comprehensive workflow, with machine learning used to optimize process conditions at the small scale, followed by scaling up and across, from the multiwell plate to the bioreactor, using on physically based scaling strategies.

In totality, this thesis demonstrated that the combination of machine learning-based optimization with traditional scaling principles provided an efficient, practical framework for bioprocess development. While this demonstration was conducted using the yeast *Komagataella phaffii* as an illustrative example, the concepts are broadly extendable to other host organisms and lay the foundation for the continued integration of machine learning tools into existing workflows to further improve bioprocess development lifecycles.

Thesis supervisor: J. Christopher Love
Title: Raymond A. (1921) and Helen E. St. Laurent Professor of Chemical Engineering