Rule-based Decision Framework for the Digital Synthesis of Optimal Pharmaceutical Processes

Yash Barhate,a Daniel J. Laky,a Daniel Casas-Orozco,a Gintaras Reklaitis,a Zoltan Nagya\*

a Davidson Schoool of Chemical Engineering, Purdue University, West Lafayette, IN 47907, USA.

\* znagy@purdue.edu

Abstract

This study introduces a rule-based decision framework employing PharmaPy for the optimization and rapid in-silico design comparisons of different end-to-end optimal (E2EO) pharmaceutical manufacturing flowsheets. The framework is designed to incorporate hybrid pharmaceutical flowsheets, incorporating unit operations in both batch and continuous modes. The methodology involves the conceptualization and application of heuristic-based synthesis rules on the master superstructure to generate smaller-scale superstructure realizations that can be readily optimized with moderate computational efforts. Lastly, the effectiveness of the framework is demonstrated using a case study, comparing various manufacturing pathways for the synthesis and purification of the anti-cancer drug Lomustine.

**Keywords**: Process synthesis, Optimization, Pharmaceutical manufacturing.

* 1. Introduction

The landscape of pharmaceutical manufacturing is undergoing a significant transformation, driven by emerging paradigms such as quality-by-design (QbD) and quality-by-control (QbC). In this rapidly evolving context, model-based digital design tools are essential for informed decision-making in process design and operation due to their in-silico optimization and design comparison capabilities for various unit operations and process flowsheets (Su et al., 2019; Yu et al., 2014). Furthermore, with the ongoing batch-to-continuous transition of the pharmaceutical industry, there is a rising demand for modeling tools capable of flexibly simulating flowsheets with different operating modes such as batch, continuous, or hybrid (incorporating both batch and continuous unit operations). Addressing this need, PharmaPy emerges as a user-friendly, open-source Python-based tool with the capability to configure and simulate various manufacturing setups (Casas-Orozco et al., 2021).

Flowsheet optimization and synthesis are commonly used for comparing and evaluating different manufacturing routes in terms of process efficiency, techno-economic metrics, environmental impact, regulatory compliance, and other relevant factors. Traditional approaches often involve the formulation of a superstructure network, typically solved through mixed-integer nonlinear optimization (MINLP) or generalized disjunctive programming (Chen & Grossmann, 2017). While these equation-oriented methods are frequently applied to simplified models, their effectiveness in optimizing flowsheets modeled with rigorous unit operation models is limited (Navarro-Amorós et al., 2014). Simulation-optimization-based frameworks are used to optimize flowsheets that contain prohibitive complexity for equation-oriented models. However, addressing complex process synthesis problems through a simulation-optimization approach presents challenges, either due to the computational cost associated with exhaustively optimizing each flowsheet or the necessity for specialized interfaces between MINLP optimization algorithms and process simulators (Corbetta et al., 2016).

In this response, this study proposes a rule-based decision process framework for the optimization of end-to-end optimal (E2EO) pharmaceutical manufacturing flowsheets. By applying a set of heuristic process synthesis rules influenced by user preferences, the original process superstructure is systematically condensed into smaller, more manageable sets for subsequent optimization. Thus, this approach reduces the computational cost associated with the optimization of different flowsheets due to the reduction in the search space.

The remainder of the paper is organized as follows: firstly, the rule-based decision framework for the optimal synthesis of pharmaceutical flowsheets is presented. Subsequently, the devised framework is applied in a case study for optimizing various manufacturing pathways to synthesize and purify an active pharmaceutical ingredient (API), Lomustine.

* 1. Proposed methodology

A generic manufacturing route for producing small-molecule APIs comprises of a few general processing steps: (i) synthesis, (ii) purification, and (iii) isolation. Each of these steps has operating decisions that impact unit performance for a given flowsheet. However, each processing step may be represented with a multitude of different choices that impact what set of operating decisions to make (i.e., desired product, raw materials, complexity of each processing step, unit operations for each step, operating mode of each unit operation, etc.). Each of these decisions impacts the process superstructure. Subsequently, the search space for possible flowsheet alternatives expands exponentially with the increasing number of options (i.e., decision variables) for each decision at each step. Also, in many pharmaceutical processing applications it is desirable to maintain high model fidelity, rendering the aforementioned mixed integer programming frameworks, (Chen & Grossmann, 2017), intractable or insufficient. This presents a substantial computational bottleneck for the optimization of every possible flowsheet within the superstructure using simulation-optimization. Here, a heuristic rule-based synthesis algorithm becomes crucial, enabling the reduction of the superstructure into smaller, manageable sub-structures for optimization with moderate computational efforts.

These synthesis rules are formulated and generalized, considering common commercial and process engineering decision points encountered in pharmaceutical manufacturing. They are systematically organized into five distinct groups: (1) process feasibility rules, (2) regulatory considerations, (3) equipment availability constraints, (4) experience- or knowledge-based rules, and (5) scenario analysis. Regulatory considerations are particularly crucial in pharmaceutical manufacturing to ensure compliance with regulations from agencies such as the FDA. Equipment availability constraints account for practical limitations in manufacturing, ensuring that the synthesis rules are aligned with real-world scenarios. Experience- or knowledge-based rules integrate prior known process knowledge to simplify the synthesis workflow. Finally, scenario analysis rules are designed to facilitate the comparison and optimization of various industrially relevant and interesting manufacturing scenarios, such as end-to-end continuous manufacturing and/or telescoped reaction synthesis, etc. While presenting an exhaustive list of all rules exceed the scope, Table 1 provides illustrative examples from each category, showcasing the flexibility offered by synthesis rules during pharmaceutical flowsheet synthesis.

|  |
| --- |
| Table 1: Process synthesis rules with their categories |
| No. | Category | Example |
| 1 | Regulatory aspects | R-1: Isolation of ISO intermediate after Rxn1, increases drug safety |
| 2 | Feasibility rules | F-1: For semi-batch reactors, all reactions are assumed to happen in a single reactor |
| 3 | Equipment availability constraints | A-1: Batch distillation is not available |
| 4 | Scenario Analysis | S-1: End-end continuous manufacturingS-2: Telescoped reaction synthesis |
| 5 | Experience or Knowledge-based rules | E-1: Solvent switch is preferred in batch mode of operation.E-2: If reaction 1 is fast and performed in a continuous reactor, avoid batch reactor for performing the next reaction. |

From the rule-based knowledge database, a specific set of rules is first selected based on user preferences. These chosen rules are then applied to the comprehensive superstructure, generating small-scale superstructure realizations, as depicted in Figure 1. Configurations violating one or more synthesis rules were systematically excluded from the search space, and only feasible flowsheets were subjected to deterministic optimization. As a result of the explicit formulation of the synthesis rules, arbitrary configurations can be analyzed transparently and systematically, enabling a thorough examination of the superstructure design space.

* 1. Case Study

Figure 1: Heuristics reduce the search space of valid superstructure layouts

In this study, the described methodology is applied to analyze the end-to-end manufacturing of the anti-cancer API, Lomustine. Recent literature highlights ongoing efforts in the development of lab-scale continuous synthesis workflows for Lomustine production (Ewan et al., 2017). This case study aims to supplement these studies by exploring and comparing various manufacturing pathways designed for the synthesis and purification of Lomustine.

Following the proposed methodology, a process-specific superstructure template was drafted considering the process chemistry as depicted in Figure 3. The manufacturing process encompasses a two-step reaction synthesis carried out in reactors R01 and R02, succeeded by a solvent-switch operation (VAP01), followed by crystallization (CR01) and filtration (F01) processing steps. There are multiple alternative unit operations and operating modes for each step in Figure 3. The comprehensive superstructure, encompassing all potential flowsheet configurations, is depicted in Figure 2. A thorough enumeration of all combinatorically generated flowsheets within this superstructure results in a total of 256 distinct configurations. Process feasibility rule F-1, equipment availability rule A-1, and experience-based rule E-1 were first applied to the above superstructure to screen out disallowed flowsheet configurations. With these rules, the search space narrows down from 256 flowsheets to 40 flowsheets to be optimized.

Figure 2: Exhaustive enumeration of possible flowsheet alternatives.

Figure 3: Process-specific superstructure template for Lomustine manufacturing.

* + 1. Design Problem:

The design problem is formulated as a nonlinear constrained optimization problem (Equation 1) with the dual objectives of maximizing both mass production rate and number-based mean crystal size of the filtered API product. To prevent premature crystallization in the vaporizer, an inequality constraint was introduced as a path constraint by maintaining the concentration of API ( less than or equal to a specified threshold given the fixed operating temperature of the vaporizer (.

|  |  |
| --- | --- |
|  | (1) |
|  |  |
|  |  |
|  |  |
|  |  |

|  |
| --- |
| Table 2: Decision variables used  |
| Unit operations | Variable |
| Reactor 1 | *CA/B,in (mol/L), tR01*/𝜏*R01, VR01* |
| Reactor 2 | *CD,in (mol/L), tR02*/𝜏*R02* |
| Evaporator | *tVAP01 (s), P (bar), RC7|R02* |
| Batch Crystallizer | *T1*, *T2, T3, tCR01* |
| MSMPR | *TCR0n*, 𝜏*CR0n* |

The optimization problem was solved using a simulation-optimization approach. The state variables, inequality constraints, and objective functions were evaluated by using callbacks to the pharmaceutical manufacturing process written in PharmaPy. A mesh adaptive direct search (MADS) algorithm using the Python interface pyNOMAD (Audet et al., 2021) was used to solve the optimization problem. The decision variables used for the flowsheets are shown in Table 2, where reactor variables one of *tR01* or 𝜏*R01* are used for batch or continuous operation respectively.

* + 1. Results

Using the optimization framework outlined earlier, all 40 flowsheets were optimized in approximately 63 hours on a MacBook Pro with macOS Catalina 2.6 GHz Quad-core Intel core i7 and 16 GB RAM. A Pareto front (Figure 4) was obtained by solving the dual-objective optimization problem for each possible flowsheet configuration of the process superstructure. Each flowsheet subsequently has optimal operating conditions which lead to varying process outputs, namely crystal size and hourly production. Figure 4A labels each optimal operating output by the first reaction’s operating mode (i.e., R01 is batch or continuous). Flowsheet configurations with reactors operating in continuous mode for the first reaction step (R01) show higher production rates compared to batch reactors. Additionally, Figure 4B labels the optimal operating output based on different crystallizer types. Flowsheets employing batch-cooling crystallizers yield larger mean-size crystals than those with MSMPR crystallizers for this case. Therefore, the batch operating mode for the crystallization step is more effective than the continuous mode considering the set of operating constraints given and the flowsheet configurations explored. These insights demonstrate the application of the proposed framework for informed decision-making in selecting optimal operating modes for different manufacturing steps.

Furthermore, to explore the computational benefits of integrating heuristic rules, an additional experience-based knowledge rule (E-2) was added to the original set of rules. This led to the further condensation of the superstructure to a smaller sub-structure of 32 flowsheets instead of 40. Optimizing this refined sub-structure took approximately 47 hours, resulting in a notable 25% reduction in computational burden underscoring the advantageous impact of incorporating heuristic rules in enhancing the computational efficiency of the optimization process.

Figure 4: A) Pareto optimal fronts for flowsheets with R01 operating in batch versus continuous modes, B) Pareto front for flowsheets with different types of crystallizers.

* 1. Conclusions

In this study, a rule-based decision framework for synthesizing and optimizing various end-to-end optimal (E2EO) pharmaceutical manufacturing flowsheets was developed. The framework was applied effectively in a case study focusing on Lomustine production, wherein various manufacturing pathways for the synthesis and purification of the API were analyzed. It was shown that PharmaPy can be used to automate superstructure synthesis and optimization across a wide range of flowsheets featuring unit operations in both batch and/or continuous operating modes. The incorporation of heuristic-based synthesis rules selected based on user preferences facilitated the generation of smaller sub-structure realizations from the original process superstructure. Subsequently, these smaller sub-structures were then optimized using a simulation-optimization framework with significantly reduced computational requirements. The analysis of optimized flowsheets within the Lomustine case study provided valuable insights into selecting optimal operating modes for different unit operations across various manufacturing steps, enhancing overall process efficiency and facilitating informed decision-making.

Acknowledgments

Funding for this work was made possible, by the US Food and Drug Administration (FDA) through grant number U01FD006738. Views expressed herein do not necessarily reflect the official policies of the Department of Health and Human Services; nor does any mention of trade names, commercial practices, or organizations imply endorsement by the United States Government.

References

Audet, C., Digabel, S. Le, Montplaisir, V. R., & Tribes, C. (2021). *NOMAD version 4: Nonlinear optimization with the MADS algorithm*.

Casas-Orozco, D., Laky, D., Wang, V., Abdi, M., Feng, X., Wood, E., Laird, C., Reklaitis, G. V., & Nagy, Z. K. (2021). PharmaPy: An object-oriented tool for the development of hybrid pharmaceutical flowsheets. *Computers and Chemical Engineering*, *153*, 107408.

Chen, Q., & Grossmann, I. E. (2017). Recent Developments and Challenges in Optimization-Based Process Synthesis.

Corbetta, M., Grossmann, I. E., & Manenti, F. (2016). Process simulator-based optimization of biorefinery downstream processes under the Generalized Disjunctive Programming framework. *Computers & Chemical Engineering*, *88*, 73–85.

Ewan, H. S., Iyer, K., Hyun, S. H., Wleklinski, M., Cooks, R. G., & Thompson, D. H. (2017). Multistep Flow Synthesis of Diazepam Guided by Droplet-Accelerated Reaction Screening with Mechanistic Insights from Rapid Mass Spectrometry Analysis. *Organic Process Research and Development*, *21*(10), 1566–1570.

Navarro-Amorós, M. A., Ruiz-Femenia, R., & Caballero, J. A. (2014). Integration of modular process simulators under the Generalized Disjunctive Programming framework for the structural flowsheet optimization. *Computers & Chemical Engineering*, *67*, 13–25.

Su, Q., Ganesh, S., Moreno, M., Bommireddy, Y., Gonzalez, M., Reklaitis, G. V., & Nagy, Z. K. (2019). A perspective on Quality-by-Control (QbC) in pharmaceutical continuous manufacturing. *Computers & Chemical Engineering*, *125*, 216–231.

Yu, L. X., Amidon, G., Khan, M. A., Hoag, S. W., Polli, J., Raju, G. K., & Woodcock, J. (2014). Understanding Pharmaceutical Quality by Design. *The AAPS Journal*, *16*(4), 771.