

Optimization-Based Synthesis and Assessment Frameworks for Bio-Based Chemicals Production

By

Wenzhao Wu

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Chemical Engineering)

at the

UNIVERSITY OF WISCONSIN–MADISON

2018

Date of final oral examination: 9/13/2018

The dissertation is approved by the following members of the Final Oral Committee:

Christos T. Maravelias, Professor, Chemical and Biological Engineering

Brian F. Pfleger, Professor, Chemical and Biological Engineering

Victor M. Zavala, Associate Professor, Chemical and Biological Engineering

Jennifer L. Reed, Professor, Chemical and Biological Engineering

Timothy J. Donohue, Professor, Bacteriology

Abstract

Recent advances in metabolic engineering enable the production of chemicals via bio-conversion using microbes. However, comprehensive frameworks that analyze such bio-production processes while comparing alternative technologies and target chemical products have been limited. To this end, we develop optimization-based synthesis and assessment frameworks for the production of bio-based chemicals (or “products”).

Specifically, we first develop a method for the efficient representation, generation, and modeling of superstructures, which imbed all relevant alternative technologies and interconnections, for process synthesis. This method is then used to develop a framework for the synthesis of cost-effective downstream bio-separation processes, which account for 60-80% of the total production cost in many cases. Thus, a general superstructure is generated to account for all classes of products, and a superstructure reduction method is developed to solve specific cases, based on product attributes, technology availability, case-specific considerations, and final product stream specifications. The bio-separation process synthesis framework is further used to study two major categories of products: extracellular and intracellular. We analyze the influence of a combination of key parameters, such as titer and technology performance, on optimal technology selection and cost.

Next, we develop a framework for the identification of techno-economically promising products. Specifically, we first develop a genome-scale metabolic modeling approach, which is used to identify a candidate pool of 209 bio-derivable products with high production volume. Then, we design three screening criteria based on a product’s profit margin, market volume and market size. In calculating the profit margin, a set of cost-titer curves generated with the aforementioned bio-separation process synthesis method is used to estimate cost. Thus, we identify 32 products

currently sold on the market as promising targets for bio-production if maximum yields can be achieved, and 22 products if maximum productivities can be achieved. Additionally, some bio-derivable products may currently have little market demand but could potentially replace existing products due to environmental or economic advantages in the future. Therefore, we further identify the molecular characteristics of such promising replacement products in terms of the number of carbon atoms, oxygen atoms and functional groups, as well as the specific function groups, in a molecule.

Acknowledgments

First, I would like to express my sincere gratitude to my advisor, Professor Christos Maravelias, for being a wonderful supervisor, mentor and role model. He has all the great qualities of a critical thinker, dedicated researcher and charismatic leader, which have influenced me profoundly in value and skill development. He has provided me with significant guidance on effectively forming ideas, conducting research, writing scientific papers and presenting results. It has been my privilege working with him in the past five years, and the memories will last forever.

Next, I would like to thank my dissertation committee members, Professors Brian Pfleger, Victor Zavala, Jennifer Reed and Timothy Donohue, for reviewing this thesis, attending my defense and sharing valuable feedbacks.

Many thanks to former and current members of the Maravelias Group for providing suggestions for my research, especially Jeff Herron, Kirti Yenkie, Kefeng Huang, Rex Ng, Yachao Dong, Jeehoon Han, Murat Sen, Lingxun Kong, Andres Merchan, Peyman Fasahati and Bruno Calfa.

My research has been funded by the U.S. Department of Energy Great Lakes Bioenergy Research Center and National Science Foundation. Thanks for the financial support.

Special thanks to all my friends. Those memories related to lunches in the Grad Lounge, sunset by Lake Mendota, trips to NYC, flipped canoes at Wildcat Mountain... will stay with me forever.

Finally, I would like to thank my parents for their endless support and encouragement during our frequent video chats, through which we have developed the strongest friendship. Also, I would like to thank them for giving me a meaningful name that often reminds me of how to get things done: start with a strong will (Wu 武); identify the best method (Wen 文); and persevere (Zhao 钊) until success happens. This has guided me through the most confusing and challenging times in life. My deepest love and gratitude go to them.

Table of Contents

1 Introduction.....	1
1.1 Superstructure optimization	1
1.2 Bio-based chemicals production	4
1.3 Thesis scope.....	6
2 Superstructure representation, generation and modeling framework.....	7
2.1 Introduction.....	7
2.2 Superstructure representation	8
2.2.1 Units.....	8
2.2.2 Ports	9
2.2.3 Streams.....	10
2.3 Superstructure generation	12
2.3.1 Fully connected superstructure.....	12
2.3.2 Superstructure connectivity rules	13
2.3.3 Superstructure generation example: HDA process.....	17
2.4 Superstructure modeling.....	19
2.4.1 Units.....	20
2.4.2 Ports	21
2.4.3 Streams.....	23
2.4.4 General form of the superstructure element models.....	26
2.4.5 Logic constraints	28
2.4.6 Objective function	28
2.5 Implementation.....	29
2.6 Remarks.....	31
2.6.1 Incorporation of non-conditioning streams	31
2.6.2 Single-stream ports.....	33
2.6.3 Conditioning reduction.....	35
2.7 Example: bio-separation process	36
2.8 Conclusions	38
3 Bio-separation process synthesis framework.....	40
3.1 Introduction.....	40
3.2 Framework overview	41
3.3 Stage-superstructure generation	43
3.3.1 Stage-wise analysis.....	43
3.3.2 Module-superstructure generation	47
3.3.3 Combination of module-superstructures.....	50
3.4 Stage-superstructures	52

3.4.1 Stage 1 superstructure.....	52
3.4.2 Stage 2 superstructure.....	52
3.4.3 Stage 3 superstructure.....	54
3.4.4 Stage 4 superstructure.....	56
3.5 General superstructure	57
3.6 Reduction method	60
3.7 Modeling approach	63
3.8 Conclusions	64
4 Separation process analysis for extracellular and intracellular chemicals	66
4.1 Introduction.....	66
4.2 Methods	67
4.2.1 Stage-wise separation scheme.....	68
4.2.2 Superstructure generation and modeling.....	69
4.2.3 Analysis framework	70
4.3 Results and discussions	70
4.3.1 EX NSL LT products.....	70
4.3.1.1 Superstructure and optimal solution.....	70
4.3.1.2 Analysis for EX NSL LT LQD CMD product.....	72
4.3.1.3 Extension to other classes of EX NSL LT products	74
4.3.2 EX NSL HV products.....	74
4.3.2.1 Superstructure and optimal solution.....	75
4.3.2.2 Analysis for EX NSL HV SLD CMD product.....	76
4.3.2.3 Extension to other classes of EX NSL HV products.....	78
4.3.3 EX SOL products	79
4.3.3.1 Superstructure and optimal solution.....	79
4.3.3.2 Alternative concentration options in Stage II.....	80
4.3.3.3 Analysis for EX SOL LQD NVL CMD product.....	82
4.3.3.4 Extension to other classes of EX SOL products	84
4.4 Conclusions	85
5 Framework for the identification of promising products	87
5.1 Introduction.....	87
5.2 Methods	89
5.2.1 Generation of a candidate pool and production pathways.....	89
5.2.2 Estimation of yield, productivity and residence time.....	92
5.2.3 Development of screening criteria.....	93
5.3 Results & discussions.....	95
5.3.1 Identification of the most cost-favorable systems	95
5.3.2 Screening results.....	97
5.3.3 Remarks	103

5.4 Conclusions	104
6 Framework for the identification of promising replacement chemicals.....	105
6.1 Introduction.....	105
6.2 Methods	106
6.3 Results	107
6.3.1 Number of carbon atoms (#C).....	107
6.3.2 Number of oxygen atoms (#O)	109
6.3.3 Number of (distinct) functional groups (#FG and #DFG)	111
6.3.4 Specific functional groups (FG)	112
6.4 Discussions.....	113
6.4.1 Suitability of bio-production	113
6.4.2 Profitability of bio-production	116
6.4.3 Additional remarks	117
6.5 Conclusions	118
7 Conclusions and future work.....	120
Appendix.....	122
A1 Explanations to Chapter 2.....	122
A1.1 Specific formulations of thermodynamic relations	122
A1.2 Stream model involving expansion valves	124
A1.3 Non-conditioning stream pressure relation.....	124
A2 Bio-separation model	126
A3 Explanations to Chapter 4.....	152
A3.1 Key parameters for EX NSL LT LQD CMD product.....	152
A3.2 Extension to EX NSL HV LQD CMD product.....	152
A3.3. Variation in Ext partition coefficient	153
A3.4. Model parameters and input data for base cases.....	153
A4 Process economics in Chapter 5.....	158
Bibliography	160

List of Figures

Figure 1.1. Superstructure representation approaches.....	2
Figure 1.2. A generic bio-production process	4
Figure 2.1. Representation elements: units, ports, and conditioning streams.....	8
Figure 2.2. Graphic representation of unit, port and stream elements, and related subsets	10
Figure 2.3. Impact of the sequence of rule implementation on the number of iterations	15
Figure 2.4. Introduction of mixer-splitter.....	15
Figure 2.5. HDA process superstructure generated	19
Figure 2.6. Graphic representation of the unit model.....	20
Figure 2.7. Graphic representation of the inlet port model.....	21
Figure 2.8. Graphic representation of the outlet port model.....	22
Figure 2.9. Graphic representation of the stream model.....	24
Figure 2.10. GAMS implementation of a general superstructure model	30
Figure 2.11. Three possible connection types between two major unit ports.....	32
Figure 2.12. Single-stream port example.....	34
Figure 2.13. Conditioning reduction at single-stream ports.....	35
Figure 2.14. Example: bio-separation process superstructure.....	36
Figure 2.15. Optimal process for the bio-separation example.....	38
Figure 3.1. Overview of the framework for bio-separation network synthesis.....	42
Figure 3.2. Stage-wise analysis of general bio-separation processes.....	43
Figure 3.3. Simplification of the superstructure for the IN module in Stage 1	47
Figure 3.4. Module- and stage- superstructures for Stage 1	52
Figure 3.5. Module- and stage- superstructures for Stage 2	53
Figure 3.6. Module- and stage- superstructures for Stage 3	55
Figure 3.7. Module- and stage- superstructures for Stage 4	56
Figure 3.8. General bio-separation superstructure	59
Figure 3.9. Example superstructure reduction for EX NSL LT NVL LQD CMD product.....	62
Figure 3.10. Final reduced superstructure for the example in Figure 3.9	62
Figure 4.1. Stage-wise separation scheme and its simplification	68
Figure 4.2. Superstructure and optimal solution for EX NSL LT LQD CMD product.....	71
Figure 4.3. Analysis on technology selection and cost for EX NSL LT LQD CMD product.....	72

Figure 4.4. Superstructure and optimal solution for EX NSL HV SLD CMD product.....	75
Figure 4.5. Analysis on technology selection and cost for EX NSL HV SLD CMD product.....	77
Figure 4.6. Superstructure and optimal solution for EX SOL LQD NVL CMD product.....	79
Figure 4.7. Technologies selected for Ext and direct Dst options in stage II for EX SOL LQD NVL CMD product.....	81
Figure 4.8. Cost analysis for EX SOL LQD NVL CMD product.....	82
Figure 4.9. Analysis with varying product titer for EX SOL LQD NVL CMD product.....	83
Figure 5.1. Economically promising products identification framework.....	89
Figure 5.2. Solution space for the production of ethanol relative to biomass growth rate.....	93
Figure 5.3. Separation cost-product titer curves for different classes of products.....	94
Figure 5.4. Comparisons of the upstream costs of the 209 products	95
Figure 5.5. Graphic representation of the screening results with the optimal microbe.....	99
Figure 5.6. Graphic representation of the screening results at the optimal localization	100
Figure 6.1. Results for #C.....	107
Figure 6.2. Results for #O	110
Figure 6.3. Results for #FG and # DFG	111
Figure 6.4. Results for specific FGs	112
Figure 6.5. Microbial conversion yield of chemicals	114
Figure 6.6. Analysis of bio-production profitability	116
Figure A1.1. Explanation to the non-conditioning stream pressure relation	125
Figure A2.1. New labels for unit ports.....	128
Figure A3.1. Superstructure and optimal solution for EX NSL HV LQD CMD product.....	152
Figure A3.2. Variation in partition coefficient versus solubility in water, and partition coefficient versus cost of solvent.....	153
Figure A4.1. Upstream process and costs in the base case.....	158

List of Tables

Table 2.1. Processing functions and port component sets for units in the HDA process.....	17
Table 3.1. Number of variables and equations of each unit model	64
Table 4.1. Technology options available for performing the tasks listed in the three stages	68
Table 5.1. Screening results for NSL LT products	101
Table 5.2. Screening results for NSL HV products.....	101
Table 5.3. Screening results for SOL products	102
Table A2.1. Component set	127
Table A2.2. Unit port numbering.....	127
Table A3.1. Key parameters for EX NSL LT LQD CMD product.....	152
Table A3.2. Important input parameters and product specifications.....	153
Table A3.3. Particle size and density information.....	153
Table A3.4. Utility and labor costs	154
Table A3.5. Standard capacities, costs, scaling factors, and labor requirements of technologies...	154

Chapter 1

1 Introduction¹

Recent advances in metabolic engineering enable the production of chemicals via bio-conversion using microbes. However, studies on developing comprehensive frameworks to synthesize and analyze such bio-production processes while comparing alternative technologies and target chemical products have been limited. To this end, we develop optimization-based synthesis and assessment frameworks for bio-based chemicals production.

This chapter describes superstructure optimization (**Section 1.1**), bio-based chemicals production (**Section 1.2**), and the thesis scope (**Section 1.3**).

1.1 Superstructure optimization

One of the fundamental problems in chemical engineering is the synthesis of a process, that is, the selection of unit operations (“units”), their interconnections and operational conditions to generate a flowsheet that meets given goals and constraints [1], [2]. Methods used for process synthesis generally include enumeration of alternatives, evolutionary modification, and superstructure optimization [3]. In enumeration of alternatives, each alternative design is generated and evaluated. In evolutionary modification, designers make changes to known flowsheets for similar processes to meet new objectives and constraints. An optimization model can also be formulated to facilitate the comparison between different flowsheets and determination of process variables such as flowrates, operating temperatures, and pressures [1], [4]–[13]. However, these two methods are not feasible

¹ This chapter includes content from: W. Wu, C. A. Henao, and C. T. Maravelias, “A superstructure representation, generation, and modeling framework for chemical process synthesis,” *AIChE J.*, vol. 62, pp. 3199–3214, 2016; and W. Wu, K. Yenkie, and C.T. Maravelias, “A superstructure-based framework for bio-separation network synthesis,” *Comput. Chem. Eng.*, vol. 96, pp. 1–17, 2017.

for cases where a large number of different alternative technologies is available for comparison. On the other hand, superstructure optimization is a model-based approach that compares alternative process networks simultaneously [14]–[16].

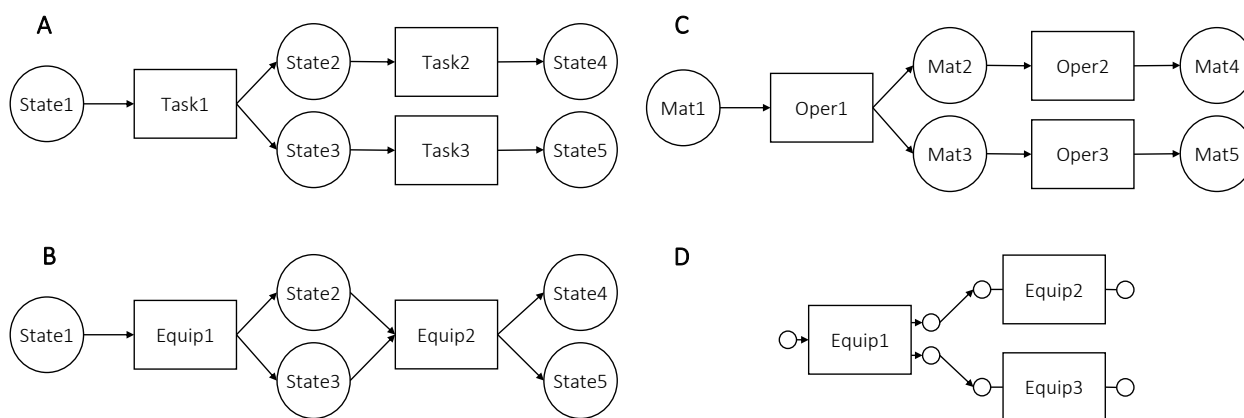


Figure 1.1. Superstructure representation approaches. (A) STN; (B) SEN; (C) P-graph; (D) R-graph.

The first step toward the generation of a superstructure, which embeds all relevant units and interconnections, is to decide how to represent a chemical process in an abstract yet clear fashion. Two most commonly used representations are the State Task Network (STN) and the State Equipment Network (SEN) [17], [18], as shown in **Figure 1.1A** and **Figure 1.1B**, respectively. The building blocks of these representations are “states”, “tasks” and “equipment units”. “States” are described in terms of physicochemical properties that uniquely define process streams. “Tasks” indicate processing steps required to transform one or more states into others. “Equipment units” are the units where a task is carried out. In STN, states and tasks are treated as nodes, whereas the connections between them are arcs. In SEN, nodes are states and equipment units, and the arcs are connections between them.

An alternative representation is P-graph [19] (**Figure 1.1C**), which is composed of two types of nodes –material (similar to “states”) and operation nodes (similar to “tasks”); the arcs represent connections between them. Another approach is R-graph [20], [21] (**Figure 1.1D**) where nodes

represent inlet and outlet ports of units; every outlet port is regarded as a general stream splitter, and every inlet port is regarded as a general mixer. Connections represent process streams going from outlet ports to inlet ports.

After the representation approach is selected, the next step is to generate a *rich* superstructure that covers the structural space (thus increasing the chances of finding the optimum design) but excludes infeasible structures (thus increasing mathematical tractability). Early works focused on (1) the combination of simple promising structures identified using engineering heuristics [22], and (2) the combination of subsystem superstructures (e.g. reaction networks, separation networks, heat recovery networks, etc.) each pre-designed via system-specific methodologies [23]–[30]. Nevertheless, these approaches are limited by the scope of the heuristics and the independence of the subsystems, and thus the resulting superstructure may include effective processes but does not guarantee the inclusion of the optimal one. Later works by Friedler et al. [19], [20] proposed a framework, using the P-graph representation, to generate the simplest superstructure containing all relevant structural alternatives. Also, in some cases, generation of schemes can assist in the generation of superstructures for the synthesis of separation networks [9], [31]–[40]. A scheme incorporates a list of technologies available for a set of tasks, while a superstructure incorporates a number of alternative specific units and relevant interconnections.

Once a superstructure is generated, the next step is the formulation of the corresponding mathematical model. In general, this model includes sets of equations describing units, their interconnections, equations for thermodynamic property calculations, etc. The model is often formulated as a mixed integer non-linear programming (MINLP) problem. The modeling of units can be performed in various ways, generally using simplifications, such as shortcut models (e.g. Fenske-Underwood equations for distillation columns) [41]–[43] and surrogate models [44]–[53]. Also, in pharmaceutical product and process development, ontologies have been used for efficient

model development and management [54]–[56]. The connections between units can be modeled using binary variables to identify structural alternatives via activation/deactivation (i.e. selection/exclusion) of units and streams. In general, the model equations include mass and energy balances, equilibrium relationships, sizing equations, design specifications, and logic constraints. The solution of the optimization model corresponds to the best process along with the optimal operating conditions for all units. For a review of MINLP and global optimization solution algorithms, the readers are pointed to past works [46], [57]–[62].

1.2 Bio-based chemicals production

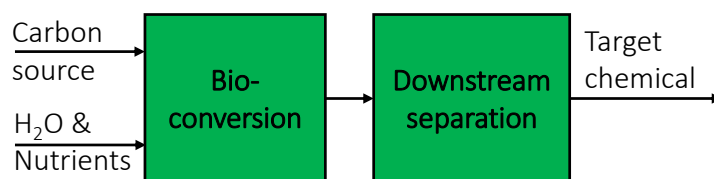


Figure 1.2. A generic bio-production process.

The last decade has seen significant progress in metabolic engineering and synthetic biology [63]–[67]. These advances enable the use of microbes such as *E.coli* and *S. cerevisiae* for the production of chemicals (“bio-based chemicals”) from sugars, made from traditional sources like sugar beet and sugarcane or from alternative sources like cellulosic biomass [66], [68]–[76]. Many of these chemicals are currently derived mainly from fossil fuel feedstocks. In comparison, bio-conversion (or “microbial conversion”) processes can be advantageous for their mild production conditions, good selectivity toward a specific product, and direct conversion instead of step-wise chemical conversions (some steps can have low yield and high cost) [77]. Also, metabolic engineering and bioreactor engineering tools can be used to maximize the yield and selectivity of the desired product and thus minimize the concentrations of coproducts [78]–[81].

A bio-production process typically consists of two major process blocks: bio-conversion and downstream separation, as shown in **Figure 1.2**. A carbon source (typically glucose), water and nutrients are fed into a bioreactor, where a microbe such as *E.coli* and *S. cerevisiae* grow and produce a target chemical (“product”). The effluent of the bioreactor, which contains the product, microbial cells, water and other impurities, is fed to a downstream separation network to obtain the final product in high purity. The costs associated with the feedstocks, bio-conversion and separation, as well as the target product selected (in terms of market demand and price), all have significant impact on the profitability of the overall process.

Note that a bioreactor effluent is often dilute (less than 20 wt % product) [82] and the purity requirement for chemical products is relatively high. Therefore, downstream separation tends to be expensive, accounting for 60–80 % of the total production cost in many cases [77], [83]–[85]. Thus, the synthesis of an effective downstream separation process is one of the most crucial tasks [1], [2], [15], [16], [86]–[89].

The synthesis of bio-separation processes is challenging for the following reasons: (1) multiple technologies are usually available for a given separation task, and thus a large number of alternative process configurations exists; (2) many bio-based chemicals require processing under mild conditions, and thus specific common separation technologies (e.g., distillation) are sometimes excluded; and (3) the product physical properties and the bioreactor effluent composition are not uniform across chemicals, but rather specific. Superstructure optimization has been proposed for the synthesis of separation networks [51], [90]–[92] and the development of bio-refineries and bio-processes [93]–[102]. However, these studies were mostly performed for specific products, on a case-by-case basis.

1.3 Thesis scope

We aim to develop optimization-based frameworks to guide the efficient synthesis and assessment of bio-production processes and the identification of techno-economically promising products.

In **Chapter 2**, we first develop a method for the efficient representation, generation, and modeling of superstructures for process synthesis. Then, in **Chapter 3**, we use this method to develop a framework for the synthesis of cost-effective bio-separation processes. In **Chapter 4**, the bio-separation framework is further used to study extracellular and intracellular products to generate significant insights. In **Chapter 5**, we develop a framework for the identification of products that currently exist in the market and are promising for bio-production, where the results in **Chapter 4** are incorporated. In **Chapter 6**, we further consider bio-derivable chemicals that may currently have little market demand but could potentially replace existing chemicals due to environmental or economic advantages in the future. We identify the characteristics of such promising replacement chemicals.

Chapter 2

2 Superstructure representation, generation and modeling framework²

2.1 Introduction

For design problems, which are not solved online or repeatedly, most of the effort goes into the generation of design alternatives and formulation of the models rather than their solutions. Thus, methods that facilitate this process, as opposed to solution methods, are likely to have greater impact. However, work on systematically generating simple superstructures embedding all relevant connections using an efficient representation approach has been limited (see **Section 1.1**). Thus, we endeavor to develop a superstructure framework, including (1) a novel representation approach that features modularity, (2) a procedure to generate simple superstructures for given sets of units, and (3) a modular modeling approach designed to work coherently with the proposed representation and generation methods. The new representation approach allows straightforward superstructure generation, while modularity allows easy model generation.

The remaining of this chapter is structured as follows. In **Section 2.2**, we discuss the superstructure representation elements. In **Section 2.3**, we develop the generation method and present an example. In **Section 2.4**, we present our modeling methods. In **Section 2.5**, we discuss the software implementation of our framework. In **Section 2.6**, we present several remarks regarding the

² This chapter includes content from: W. Wu, C. A. Henao, and C. T. Maravelias, "A superstructure representation, generation, and modeling framework for chemical process synthesis," *AIChE J.*, vol. 62, pp. 3199-3214, 2016. Carlos Henao proposed the Unit-Port-Conditioning Stream representation and connectivity rules; Wenzhao Wu completed the formulation of the superstructure optimization model based on the representation, development of the case studies, and writing of the manuscript. Major notations for the sets, parameters and variables can be found in Appendix 2.

techniques that can be applied to improve computational tractability. Finally, in **Section 2.7**, we demonstrate the applicability of the framework using a bio-separation example.

2.2 Superstructure representation

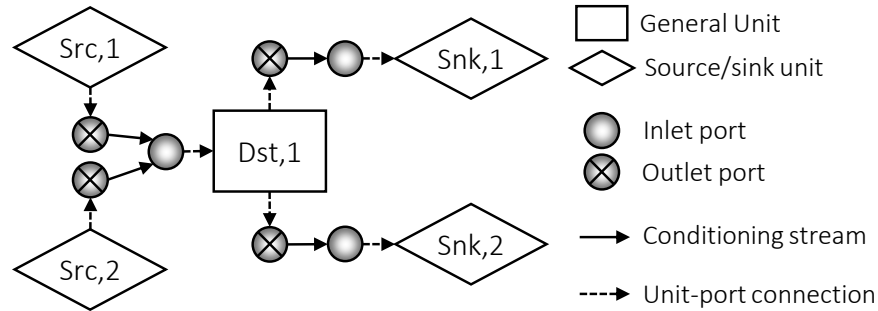


Figure 2.1. Representation elements: units, ports, and conditioning streams. General units (Dst,1), sources (Src,1-2) and sinks (Snk,1-2) are grouped as “units”; inlet and outlet ports are “ports”; “unit-port connections” merely function to denote port locations, but not representation elements.

As shown in **Figure 2.1**, we employ three types of representation elements: units, unit ports (referred to as “ports” hereafter) and general conditioning streams (also referred to as “streams” hereafter unless noted).

2.2.1 Units

In addition to general units, source units and sink units (henceforth referred to as “sources” and “sinks”, respectively) are included as auxiliary elements whose function is to provide raw materials and collect final products and wastes. Two basic sets are used to define a unit:

- (1) $ut \in \mathbf{UT}$: unit types, such as reactors (“Rct”) and separators (“Spr”) generally, and distillation tanks (“Dst”), sources (“Src”), sinks (“Snk”) and mixer-splitters (“Mxs”) specifically;
- (2) $un \in \mathbf{UN}$: unit numbers, i.e. a list of consecutive integers used to distinguish units of the same type.

The unit set is then defined as $u \in \mathbf{U} \subset \mathbf{UT} \times \mathbf{UN}$, which is indexed by unit type and unit number.

Two additional sets are defined:

(1) $c \in \mathbf{C}$: components, used to denote different chemicals;

(2) $r \in \mathbf{R}$: reactions, i.e., a list of consecutive integers.

2.2.2 Ports

Ports correspond to stream inlet and outlet points in every unit (e.g. reactor inlet, distillation column distillate outlet, etc.). Every type of unit has a predefined number of inlet and outlet ports (e.g. flash units have a single inlet port and two outlet ports). In particular, sources only have one outlet port, while sinks only have one inlet port. Every outlet port is regarded as a general splitter able to divide the stream leaving a unit into streams that can be sent to different inlet ports. Every inlet port is regarded as a general mixer able to combine streams coming from different outlet ports. Two additional basic sets are used to define a port:

(1) $pt \in \mathbf{PT}$: port types, i.e. inlet (“in”)/outlet (“out”);

(2) $pn \in \mathbf{PN}$: port numbers.

The port set is then defined as $p \in \mathbf{P} \subset \mathbf{U} \times \mathbf{PT} \times \mathbf{PN}$, which is indexed by unit, port type and port number. We also define $\mathbf{P}_{pt}$ (including $\mathbf{P}_{in}$ and $\mathbf{P}_{out}$) to denote inlet port and outlet port subsets. For clarity, when inlet ports and outlet ports are to be distinguished, we use p to denote inlet ports, and p' to denote outlet ports.

2.2.3 Streams

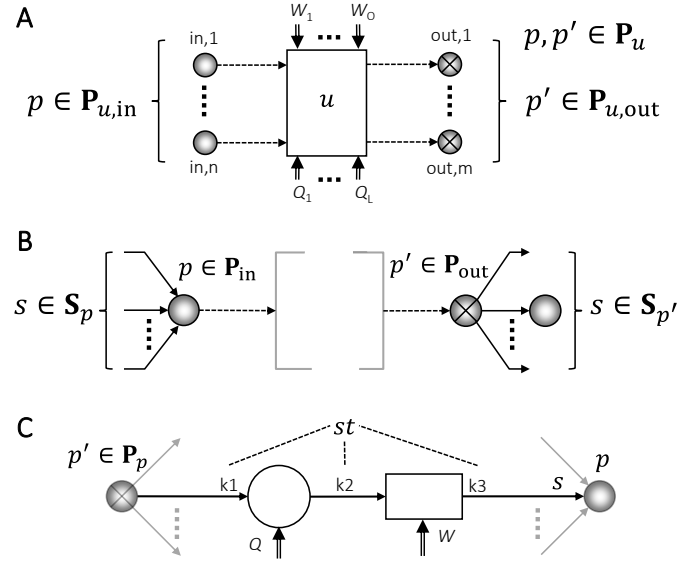


Figure 2.2. Graphic representation of unit, port and stream elements, and related subsets. **(A)** A unit u and its inlet and outlet ports; **(B)** an inlet port p and its incoming streams, and an outlet port p' and its outgoing streams; **(C)** a conditioning stream s , and its head and tail ports, where the large circle and rectangle represent temperature conditioning and pressure conditioning tasks, respectively.

Streams act as connections between outlet ports and inlet ports, while performing conditioning tasks (note that multiple conditioning units may be required for a conditioning task). Since each connection corresponds to a stream, we do not distinguish between the terms “connection” and “stream” (unless noted). A conditioning stream performs a temperature conditioning task (heating/cooling) followed by a pressure conditioning task (compression/expansion/pumping), which allows temperature and pressure changes required to generate inlet port mixtures at conditions necessary for the receiving unit to operate properly. **Figure 2.2C** shows a graphic representation of conditioning streams, which includes three distinct states $k \in \mathbf{K} = \{k1, k2, k3\}$: initial state ($k1$) defined by the conditions of the outlet port p' ; intermediate state ($k2$) after temperature conditioning; and final state ($k3$) after pressure conditioning, at which the next inlet port p operates. Note that the symbols representing conditioning tasks in **Figure 2.2C** do not

appear in the superstructure. Instead, we use a single conditioning stream as shown in **Figure 2.1** to represent both the flow and the conditioning tasks in a compact way. Correction factors can be used to better estimate the net heat and work duties of multi-stage operations (e.g., compressions with intercooling).

The stream set is defined as $s \in \mathbf{S} \subset \mathbf{P}_{\text{out}} \times \mathbf{P}_{\text{in}}$, which is indexed by the “head” and “tail” ports. Thus we can also denote a stream s as $p' \rightarrow p$, or $s \in \mathbf{S}_{p',p}$.

To facilitate the presentation, several additional subsets are defined below. The graphic representation of these subsets is given in **Figure 2.2A** and **Figure 2.2B**. Note that since we use two-letter indices such as unit types (ut) and port types (pt), we add commas in between when multiple indices are used (e.g. $\mathbf{P}_{u,pt}$).

$\mathbf{P}_u/\mathbf{P}_{u,pt}$: ports of unit u / unit u and type pt ;

$\mathbf{P}_{ut,pt}$: ports connected to unit type ut , and of port type pt ;

$\mathbf{P}_p$: outlet ports connected to inlet port p ;

$\mathbf{S}_p$: streams connected to port p .

While ports function as multi-stream mixers and splitters, their direct use facilitates superstructure generation. A connection is constructed between each pair of inlet/outlet ports, which is straightforward and the generated superstructure is unique. In contrast, if mixers and splitters are used, then the number and connections of such mixers/splitters should be decided, and different superstructures may be generated.

In general, the use of our superstructure elements (units, ports and streams) offers the possibility of highly connected superstructures while emphasizing a distinction between the reaction/separation tasks and conditioning tasks. This way, we can focus on the main units and

treat the conditioning tasks in a more unified fashion during modeling. Specifically, if we build a highly connected network using traditional “non-conditioning streams” together with conditioning units (i.e. compressors/expanders/pumps and heaters/coolers), this will unnecessarily increase the total number of streams and ports in the model. In addition, the use of conditioning streams can facilitate the use of aggregated models to deal with the integration of industrial utilities, e.g. Pinch Analysis based on transshipment models [103]. Finally, in principle, surrogate models can also be developed for conditioning stream models.

2.3 Superstructure generation

In this section, we present our approach to superstructure generation, which applies four connectivity rules on a fully connected superstructure. An illustrative example is used for demonstration.

2.3.1 Fully connected superstructure

To generate the fully connected superstructure, we start with the selection of a set of units (reaction and separation only) necessary to generate a series of tasks meeting the specified objectives [3], [104], [105]. In this respect, our representation is similar to STN, but the tasks here are more general since they are not strictly defined. For example, a unit required to perform the separation of components A and B can admit other components (e.g. inert components) in its feed as well, for a wide range of compositions and flow rates. After the units are selected, the inlet and outlet ports of each unit are identified and then each outlet port is connected to each inlet port, thus generating a fully connected superstructure.

2.3.2 Superstructure connectivity rules

Connections in a process structure allow the transfer of components, making them available to each unit as required, in accordance with their intended processing functions (i.e., their normal operations). To obtain a rich superstructure that contains relevant and feasible connections only, the proposed fully connected superstructure has to be simplified. The connection feasibility can be expressed in terms of the components which must be present or absent at each port for normal operation of a unit. Specifically, after removing unnecessary connections, every unit should be reachable by the components required for its normal operation and not reachable by components whose presence can disrupt such operation. In order to present a formal statement of our connectivity rules, four basic component subsets are defined:

- (1) Feasible component of inlet port $p \in \mathbf{P}_{\text{in}}$ $\{\mathbf{C}_p^F\}$: Set of components whose presence in port p will not negatively affect the normal operation of the unit (e.g., for a reaction $A+B \rightarrow C$ in a reactor, component A and an inert component are feasible for the reactor inlet port).
- (2) Minimal component of inlet port $p \in \mathbf{P}_{\text{in}}$ $\{\mathbf{C}_p^M\}$: Set of components whose presence in port p is required during the normal operation of the unit (e.g., components A and B are minimal for the inlet port of a reactor producing component C via the reaction $A+B \rightarrow C$).
- (3) Feasible component of outlet port $p' \in \mathbf{P}_{\text{out}}$ $\{\mathbf{C}_{p'}^F\}$: Components present in the outlet port p' of a unit performing its intended processing function while being fed with the feasible components (e.g., a non-volatile component is feasible for the liquid port of a flash tank).
- (4) Minimal component of outlet port $p' \in \mathbf{P}_{\text{out}}$ $\{\mathbf{C}_{p'}^M\}$: Components present in the outlet port p' of a unit performing its intended processing function while being fed with minimal components (e.g., component C is minimal for the outlet port of a reactor with reaction $A+B \rightarrow C$).

The normal operation of a unit u is feasible if (1) infeasible components can be provided to its inlet ports, and (2) supplying all its inlet ports with all their minimal components is possible. A process superstructure is feasible if the normal operation of all its units is feasible. Below we present our connectivity rules.

Rule 1: To prevent infeasible components from being fed to inlet port $p \in \mathbf{P}_{\text{in}}$, all minimal components for outlet ports $p' \in \mathbf{P}_p$ (outlet ports p' connected to inlet port p) must be feasible for p . Mathematically, this rule is expressed as:

$$\mathbf{C}_{p'}^{\text{M}} \subseteq \mathbf{C}_p^{\text{F}}, \quad p \in \mathbf{P}_{\text{in}}, p' \in \mathbf{P}_p \quad (2.1)$$

Rule 2: The supply of all minimal components to a unit inlet port $p \in \mathbf{P}_{\text{in}}$ is satisfied if the incoming connections from all outlet ports $p' \in \mathbf{P}_p$ collectively contain all such components. Mathematically, this rule is expressed as:

$$\mathbf{C}_p^{\text{M}} \subseteq \bigcup_{p' \in \mathbf{P}_p} \mathbf{C}_{p'}^{\text{F}}, \quad p \in \mathbf{P}_{\text{in}} \quad (2.2)$$

Rule 3: A useful connection to the inlet port of a reactor ($p \in \mathbf{P}_{\text{ut,in}}, ut = \text{Rct}$) is one which supplies at least one of its minimal components (i.e. reactants). Formally, this rule is expressed as:

$$\mathbf{C}_p^{\text{M}} \cap \mathbf{C}_{p'}^{\text{F}} \neq \emptyset, \quad p \in \mathbf{P}_{\text{Rct,in}}, p' \in \mathbf{P}_p \quad (2.3)$$

Rule 4: A useful connection to a separator inlet port $p \in \mathbf{P}_{\text{Spr,in}}$ is one which supplies all its minimal components, in order to avoid unnecessary mixing and separation. For example, the connection between the outlet port of a source supplying A and the inlet port of a unit intended to separate A from C (both are minimal components) does not violate Rules 2.1 and 2.2, but it is obviously counterproductive. Formally, this rule is expressed as:

$$\mathbf{C}_p^{\text{M}} \cap \mathbf{C}_{p'}^{\text{F}} = \mathbf{C}_p^{\text{M}}, \quad p \in \mathbf{P}_{\text{Spr,in}}, p' \in \mathbf{P}_p \quad (2.4)$$

The above four rules should all be applied to the fully connected superstructure. While the four rules must be satisfied simultaneously, the sequence of implementation affects the number of times Rule 2 has to be applied. Notice that Rules 1, 3 and 4 are applied to a single stream, while Rule 2 is applied to the union of incoming streams connected to an inlet port. Thus, applying Rules 1, 3 and 4 first and then Rule 2 reduces the number of iterations. An example is shown in **Figure 2.3**, where the connections between two outlet ports and one inlet port are to be evaluated by Rules 1 and 2.

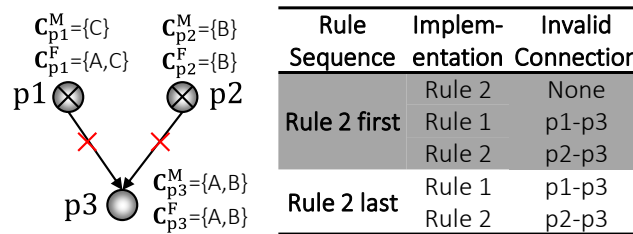


Figure 2.3. Impact of the sequence of rule implementation on the number of iterations of Rule 2. In this simple example, we can directly infer that neither of the connections satisfy both rules. However, applying Rule 2 last requires one less iteration.

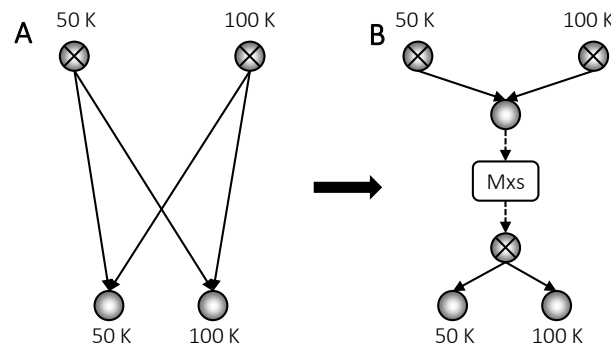


Figure 2.4. Introduction of mixer-splitter. (A) Original complete bipartite graph; (B) simplified graph with mixer-splitter. In general, such simplification decreases exergy. For example, if the optimal process requires no temperature conditioning, i.e. the 50 K (and 100 K) outlet port is connected to the 50 K (and 100 K) inlet port in the original graph, then the simplified graph in B cannot yield the same process because the mixing of the initial 50 K and 100 K stream will render a temperature in between without any conditioning (e.g. 80 K) and thus additional utilities are needed to achieve final temperature targets. However, the exergy decrease is avoided if at most one of the outlet (or inlet) ports can be activated. For example, if only one of the two inlet ports is activated in A, then graph B yields the same conditioning duty because the mixed stream now has a single temperature target.

The next question then becomes: what does the removal of all connections to an inlet port imply? If no connections are generated toward an inlet port of a unit, then the unit cannot function normally,

indicating that the unit should be removed from the superstructure. At this point, we must check whether we have assigned improper tasks to the unit in the first place, or whether we have defined the minimal and feasible components of the corresponding ports incorrectly.

If ports are viewed as vertices of a bipartite graph, with inlet and outlet ports defining the partition, and streams are represented as edges in this graph, then the four rules remove unnecessary edges. However, after all the rules are implemented, complete subgraph (where all the outlet ports are connected to all the inlet ports, e.g. **Figure 2.4A**) may exist. This particular graph can be simplified (but with caution) with a final “polishing step” that introduces a “superstructure mixer-splitter” as shown in **Figure 2.4B**. The “full-connectivity” of the original bipartite graph guarantees satisfaction of all the simplification rules in the simplified structure. Nonetheless, note that the use of mixer-splitters in general decreases exergy in the mix-split process (see **Figure 2.4** for explanation). However, if the mix and the split do not co-exist, then the exergy decrease is avoided, which happens when at most one of the outlet (or inlet) ports can be activated. For example, if the outlet (or inlet) ports in **Figure 2.4A** belong to two competing units respectively, and at most one unit can be activated, then at most one of the outlet (or inlet) ports can be activated, and thus the bipartite graph can be simplified accordingly as shown in **Figure 2.4B**.

Compared to STN representations, our approach tends to generate simpler superstructures not only because the unit models are more flexible (in terms of the description of “tasks”), but also because the conditioning tasks have been assigned to the streams. On the other hand, the unit models in our framework are potentially more complex because one unit operation can be equivalent to multiple tasks (e.g., a reactor running under different conditions resulting in different outlet streams). Compared to SEN representations, our approach leads to simpler connectivity between units. However, for a given process synthesis problem, the number of units used in our representation tends to be higher than in the case of SENs. Nonetheless, having an idea of the

function each unit is supposed to perform allows the formulation of unit models that are less complex. The proposed approach is different from the P-graph approach in that (1) it is based on a new representation using units, ports and conditioning streams, and (2) it employs a fully connected superstructure as a basis before the application of the four connectivity rules.

2.3.3 Superstructure generation example: HDA process

Table 2.1. Processing functions and port component sets for units in the HDA process.

Unit	Processing function ⁽¹⁾	Port ⁽²⁾	C_p^M ⁽¹⁾	C_p^F ⁽¹⁾
Src,1	Supply raw material 1	out,1	1	1
Src,2	Supply raw material 4	out,1	4	4
Snk,1	Retrieve waste material 2	in,1	2	1,2
Snk,2	Retrieve main product 3	in,1	3	3
Snk,3	Retrieve secondary product 5	in,1	5	5
Pfr,1 & Str,1	Consume 1 and 4, to produce 3 while minimizing the production of 5	in,1	1,4,5	1,2,4,5
		out,1	1,2,3,4,5	1,2,3,4,5
Fls,1	Separate lights 1,2 from heavies 3,4,5 (from a mixture containing significant amounts of all of them).	in,1	2,3	1,2,3,4,5
		out,1	2 ⁽³⁾	1,2 ⁽³⁾
		out,2	2,3	2,3,4,5
Fls,2	Strip 2 from a mixture containing traces of 2 and significant amounts 3,4,5.	in,1	2,3	2,3,4,5
		out,1	2,3	2,3
		out,2	3 ⁽⁴⁾	3,4,5 ⁽⁴⁾
Fls,3	Strip 4 from a mixture containing 4 and 5.	in,1	4,5	4,5
		out,1	4,5	4,5
		out,2	5 ⁽⁴⁾	5 ⁽⁴⁾
Abs,1	Separate incondensable 1 & highly volatile 2 using selective absorption of 2 in heavy component 4	in,1	1,2	1,2,3,4,5
		in,2	4	3,4,5
		out,1	1 ⁽⁵⁾	1 ⁽⁵⁾
		out,2	2,4	2,3,4,5
Dst,1	Separate component 3 from heavier components 4 and 5	in,1	3,4	3,4,5
		out,1	3 ⁽⁶⁾	3
		out,2	4 ⁽⁶⁾	4,5 ⁽⁶⁾

Notes:

(1) Component identification number: H₂=1, CH₄=2, C₆H₆=3, C₇H₈=4, C₁₂H₁₀=5

(2) Port naming: light or only inlet = in,1; heavy inlet = in,2; light or only outlet = out,1; heavy outlet = out,2

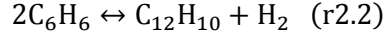
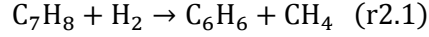
(3) Assuming no heavies in the light product

(4) Assuming no lights in the heavy product

(5) Assuming no lights in the heavy product

(6) Assuming no heavies in the light product, no lights in the heavy product

Consider the synthesis problem for the Hydrodealkylation (HDA) process [106], where the production of benzene from toluene and hydrogen is based on the following stoichiometry:



The main reaction (r2.1) generates the product C_6H_6 (benzene) and a byproduct CH_4 (methane). The side reversible reaction (r2.2) consumes the main product and generates a secondary product $\text{C}_{12}\text{H}_{10}$ (diphenyl). We are mainly interested in the production of C_6H_6 .

The available process information suggests that the production of diphenyl can be inhibited by the presence of both hydrogen and diphenyl in the reactors. Hence, proper reactor operation should involve hydrogen excess and diphenyl recycling. To support this, other operations are necessary for the separation and recycling of H_2 , C_7H_8 and $\text{C}_{12}\text{H}_{10}$, the separation of CH_4 , and the separation of C_6H_6 . The component relative volatilities allow all separations to be performed using operations based on liquid-vapor equilibrium. Also, with the volatility order given by $\alpha_{\text{H}_2} > \alpha_{\text{CH}_4} > \alpha_{\text{C}_6\text{H}_6} > \alpha_{\text{C}_7\text{H}_8} > \alpha_{\text{C}_{12}\text{H}_{10}}$, the specific separation cuts to perform are $\text{CH}_4/\text{C}_6\text{H}_6$, $\text{C}_6\text{H}_6/\text{C}_7\text{H}_8$, $\text{C}_7\text{H}_8/\text{C}_{12}\text{H}_{10}$. To cover these processing functions the superstructure includes one plug flow reactor (Pfr,1), one stirred tank reactor (Str,1), three flash vessels (Fls,1-3) one absorption column (Abs,1) and one distillation column (Dst,1). The superstructure also includes two sources (Src,1-2) and three sinks (Snk,1-3). The specific processing functions and port component sets are presented in **Table 2.1**.

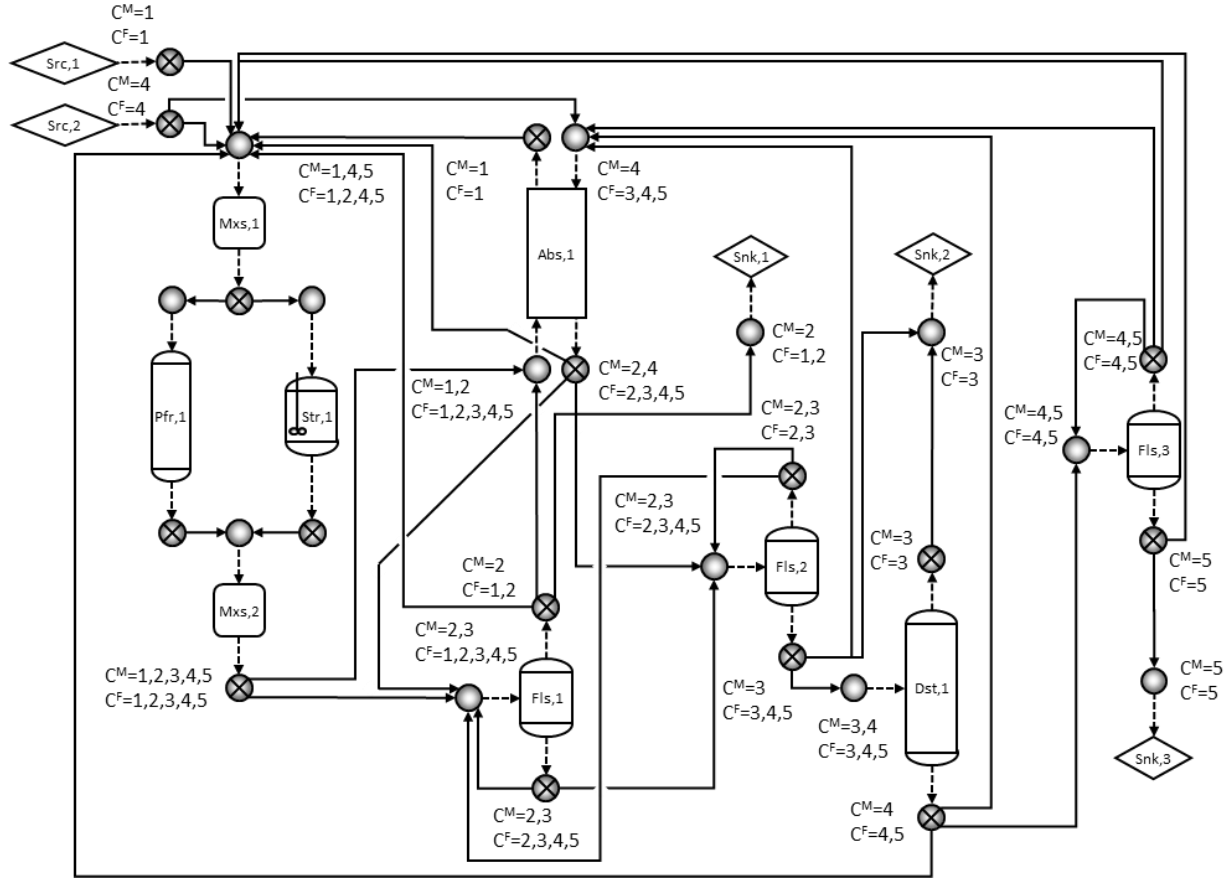


Figure 2.5. HDA process superstructure generated after the application of connectivity rules and the final polishing step. For simplicity, instead of $C_p^M/C_p^F = \{\dots\}$ we use $C^M/C^F = \dots$

The fully connected superstructure includes 14 outlet ports and 11 inlet ports, which define 154 connections. However, the number of connections is reduced to 32 when the connectivity rules and the final polishing step (the use of mixer-splitters before and after Pfr,1 and Str,1) are enforced. The final superstructure is presented in **Figure 2.5**.

2.4 Superstructure modeling

In this section, we present the formulation of a MINLP model based on the proposed representation. First, we discuss specific element models as well as a general form of such models.

Second, we discuss system-level modeling, including the formulation of logic constraints and objective function.

2.4.1 Units

The natural way to view a unit model is that unit “output” variables including outlet port variables (e.g. component flow rates in the outlet ports of the unit) and unit dependent variables (e.g. heat and work requirement, and unit size) can be calculated from unit “input” variables including inlet port variables (e.g. component flow rates in the inlet ports of the unit) and unit “internal” variables (e.g. membrane concentrating factor, and unit internal pressure drop). In a general case, the input variables (denoted as $\mathbf{x}_{in}$) and the output variables ($\mathbf{x}_{out}$) are correlated as $f(\mathbf{x}_{in}, \mathbf{x}_{out}) = 0$; and the output variables are calculated by solving this function in its implicit form (e.g. using underwood equations to calculate component fractions). However, here we present a general explicit form instead: $\mathbf{x}_{out} = f(\mathbf{x}_{in})$, for the sole purpose of demonstrating the logic behind our models (the discussion of specific models is beyond the scope of this paper). The same approach is used in the other superstructure element models.

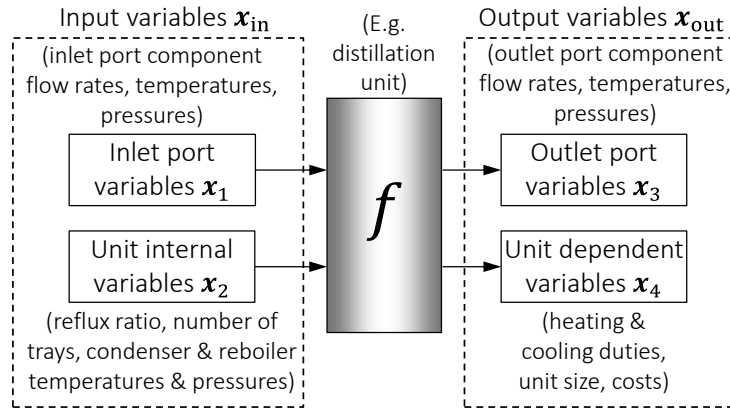


Figure 2.6. Graphic representation of the unit model. Specific variables for an example distillation unit model is shown in brackets.

Accordingly, for a general unit $u \in \mathbf{U}$, the model is as follows (see **Figure 2.6**):

$$(\mathbf{x}_3, \mathbf{x}_4) = f(\mathbf{x}_1, \mathbf{x}_2) \quad (2.5)$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are the input variables, and $\mathbf{x}_3$ and $\mathbf{x}_4$ are the output variables of the unit. Specifically, $\mathbf{x}_1$ denotes the inlet port variables (component flow rates $F_{p,c}$, temperatures T_p and pressures P_p); $\mathbf{x}_2$ denotes unit internal variables; $\mathbf{x}_3$ denotes outlet port variables (component flow rates $F_{p',c'}$, temperatures $T_{p'}$ and pressures $P_{p'}$); $\mathbf{x}_4$ denotes unit dependent variables such as heating duty Q_u , work W_u , unit size A_u , capital cost C_u^{acc} , and operating cost C_u^{oc} ; Note that in **Equation 2.5**, we express the unit dependent variables ($\mathbf{x}_4$) as a function of inlet port variables ($\mathbf{x}_1$) and internal variables ($\mathbf{x}_2$) in a general form. However, in some cases, once calculated from the inlet port ($\mathbf{x}_1$) and unit internal ($\mathbf{x}_2$) variables, outlet port variables ($\mathbf{x}_3$) can be directly used to calculate the unit dependent variables ($\mathbf{x}_4$). For example, the top outlet port component flow rates of a distillation unit together with the unit internal variables (reflux ratio and number of stages) can be directly used to calculate volume of the unit.

2.4.2 Ports

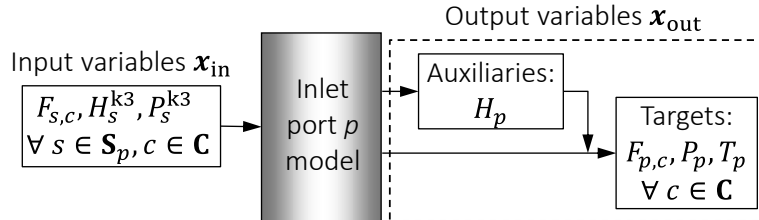


Figure 2.7. Graphic representation of the inlet port model. “Targets” are the key output variables. “Auxiliaries” are intermediate variables used to calculate the target variables.

The inlet port model calculates three key output variables (component flow rates, temperature and pressure in an inlet port) based on the incoming stream variables (as shown in **Figure 2.7**). The inlet ports ($p \in \mathbf{P}_{\text{in}}$) are modeled as adiabatic and isobaric stream mixers as follows:

$$F_{p,c} = \sum_{s \in \mathbf{S}_p} F_{s,c}, \quad c \in \mathbf{C} \quad (2.6)$$

$$H_p = \sum_{s \in \mathbf{S}_p} H_s^{\text{k3}} \quad (2.7)$$

$$P_p = P_s^{\text{k3}}, \quad s \in \mathbf{S}_p \quad (2.8)$$

$$T_p = \hat{f}^{\text{H}}\left([F_{p,c}]_{c \in \mathbf{C}}, H_p, P_p\right) \quad (2.9)$$

Equations 2.6 and **2.7** are mass and energy balances; $F_{p,c}/F_{s,c}$ denotes the flow rate (of unit e.g. kg/hour) of component c in port p /stream s ; H_p/H_s^{k3} denotes enthalpy flow rates (of unit e.g. kJ/hour) in port p /of the k3 state (after conditioning) in stream s , where enthalpy is introduced to cope with the temperature change in mixing. **Equation 2.8** is the isobaric condition. Finally, **Equation 2.9** calculates the port temperature T_p ; $\hat{f}^{\text{H}}$ denotes the inverse function of the thermodynamic equation $H_p = f^{\text{H}}([F_{p,c}]_{c \in \mathbf{C}}, T_p, P_p)$ in terms of T_p (i.e., regarding $[F_{p,c}]_{c \in \mathbf{C}}$ and P_p as constants), where H_p and T_p are correlated. The specific f^{H} and $\hat{f}^{\text{H}}$ expressions are discussed in **Appendix 1.1**. To summarize, the incoming stream variables $F_{s,c}$, H_s^{k3} and P_s^{k3} uniquely determine the inlet port model.

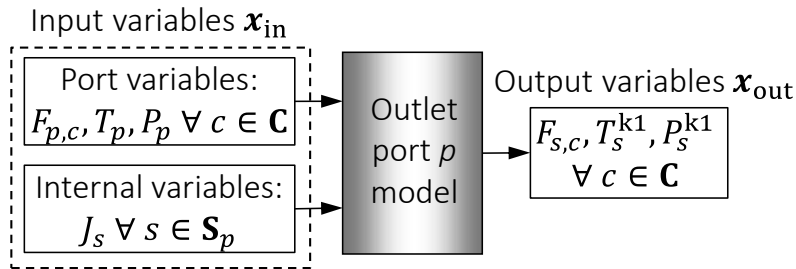


Figure 2.8. Graphic representation of the outlet port model.

The outlet port model is used to calculate all the outgoing stream variables based on the port variables and the internal variable (splitting fraction), as shown in **Figure 2.8**. Outlet ports ($p \in \mathbf{P}_{\text{out}}$) are modeled as general splitters:

$$\sum_{s \in \mathbf{S}_p} J_s = 1 \quad (2.10)$$

$$0 \leq J_s \leq 1, \quad s \in \mathbf{S}_p \quad (2.11)$$

$$F_{s,c} = J_s F_{p,c}, \quad c \in \mathbf{C}, s \in \mathbf{S}_p \quad (2.12)$$

$$T_s^{k1} = T_p, \quad s \in \mathbf{S}_p \quad (2.13)$$

$$P_s^{k1} = P_p, \quad s \in \mathbf{S}_p \quad (2.14)$$

Equations 2.10 and **2.11** describe the basic relationship between stream split fractions J_s for streams $s \in \mathbf{S}_p$ leaving outlet port p . **Equations 2.12-2.14** describe mass and energy balances of a splitter. The outlet port variables and $|s| - 1$ split fractions $(F_{p,c}, T_p, P_p, J_s)$ uniquely determine the outlet port model.

The outlet port model (**Equations 2.10-2.14**) is mostly linear, except for the bilinear term in **Equation 2.12**. These terms can be eliminated by using approximations employing binary variables (e.g., convex hull reformulation [107] and parametric disaggregation [108]).

2.4.3 Streams

The stream model (**Figure 2.9A**) calculates conditioning duties and costs, given the stream flow rate, and the k1 and k3 states (refer to **Figure 2.2C**). Streams ($s \in \mathbf{S}$) are modeled by combining equations describing a general heating-cooling task and a general compression-expansion (or pumping) task.

$$P_s^{k2} = P_s^{k1} \quad (2.15)$$

$$S_s^{k2} = S_s^{k3} \quad (2.16)$$

$$T_s^{k2} = \hat{f}^H \left([F_{s,c}]_{c \in \mathbf{C}}, S_s^{k2}, P_s^{k2} \right) \quad (2.17)$$

$$S_s^{k3} = f^S \left([F_{s,c}]_{c \in \mathbf{C}}, T_s^{k3}, P_s^{k3} \right) \quad (2.18)$$

$$H_s^k = f^H \left([F_{s,c}]_{c \in \mathbf{C}}, T_s^k, P_s^k \right), \quad k \in \{k1, k2, k3\} \quad (2.19)$$

$$Q_s = H_s^{k2} - H_s^{k1} \quad (2.20)$$

$$W_s = \eta(H_s^{k3} - H_s^{k2}) \quad (2.21)$$

$$C_s^{\text{acc}} = f^{\text{acc,HC}}(Q_s, T_s^{k1}, T_s^{k2}, P_s^{k1}) + f^{\text{acc,CE}}(W_s, T_s^{k2}, T_s^{k3}, P_s^{k2}, P_s^{k3}) \quad (2.22)$$

$$C_s^{\text{oc}} = f^{\text{oc,HC}}(Q_s, T_s^{k1}, T_s^{k2}) + f^{\text{oc,CE}}(W_s) \quad (2.23)$$

$$f_{s,c}^{\text{lo}} Y_s \leq F_{s,c} \leq f_{s,c}^{\text{up}} Y_s, \quad c \in \mathbf{C} \quad (2.24)$$

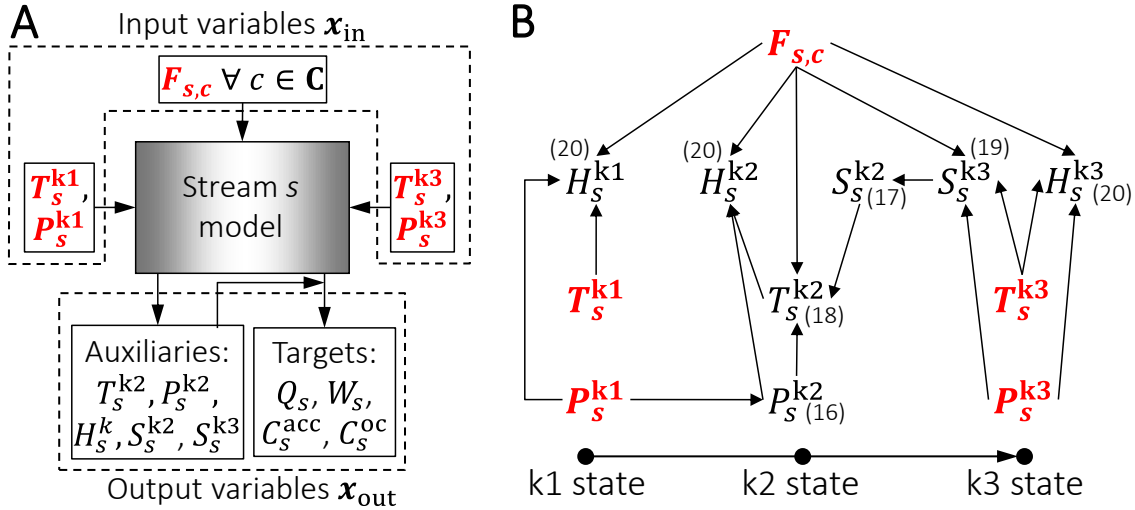


Figure 2.9. Graphic representation of the stream model. **(A)** General model; **(B)** specific calculation procedure for **Equations 2.15-2.19**. The input variables are marked bold and red. The bracketed number next to each output variable in B corresponds to the number of the equation used for the calculation. Once all the variables in B are calculated, the conditioning duties are known according to **Equations 2.20** and **2.21**.

Equation 2.15 describes the pressure equality relation across the heater/cooler. **Equation 2.16** is the isentropic condition for compressors, pumps or expansion turbines in the pressure conditioning task. The model for expansion valves with isenthalpic assumption is shown in

Appendix 1.2. If the specific conditioning task (either isentropic or isenthalpic pressure conditioning here) cannot be determined prior to optimization, then additional binary variables are required to model the activation/deactivation of candidate conditioning tasks (see the conditioning stream model in **Appendix 2** for reference). **Equations 2.17-2.19** relate stream state variables with thermodynamic properties, where $\hat{f}^H$, f^S and f^H denote general thermodynamic correlations. In **Appendix 1.1**, we present specific formulations for these expressions, with different levels of assumptions and simplifications. **Equations 2.20** and **2.21** are the energy balances of the temperature conditioning task (connecting states k1 and k2) and the pressure conditioning task (connecting states k2 and k3), where Q_s and W_s are the stream heating/cooling utility and compression/ pumping work; η is the pressure conditioning efficiency. **Equations 2.22** and **2.23** calculate capital cost and operating cost of the conditioning tasks. **Equation 2.24** ensures that the flow rates of deactivated streams are zero, where $f_{s,c}^{lo}$ and $f_{s,c}^{up}$ denote flow rate lower and upper bounds of component c in stream s, respectively. Note that this constraint is not enforced for enthalpy flow rates because the zero-enthalpy is automatically satisfied due to the zero-flow modeling for deactivated streams. The logic behind model **Equations 2.15-2.19** is illustrated in **Figure 2.9B**. To summarize, T_s^{k1} , T_s^{k3} , P_s^{k1} , P_s^{k3} and $F_{s,c}$ uniquely determine the stream model.

The modeling of streams is a nontrivial task for the following reasons.

First, the thermodynamic **Equations 2.17-2.19** are usually nonlinear. To avoid excessively complex expressions, as shown in **Appendix 1.1**, simplifying assumptions are typically made, e.g. ideal gas, adiabatic, isobaric, isothermal, isenthalpic and isentropic assumptions, etc. [109]–[111]. To further simplify the model, the heating/cooling duty and the compression /pumping work can be directly correlated with the variables at k1 and k3 states, i.e., $F_{s,c}$, $(P_s^{k1} - P_s^{k3})$ and $(T_s^{k1} - T_s^{k3})$, without

introducing variables at $k+2$ states, thus eliminating both entropy and enthalpy in the formulation (see the conditioning stream model in **Appendix 2**).

Second, phase change may occur during conditioning, further adding complexity to the thermodynamic expressions. Moreover, phase change can occur inside the pressure conditioning units. For example, if phase change occurs when a gas stream is compressed in a compressor (where gas becomes liquid), a pump must then take over to increase the liquid pressure to a target level. In this work, we disregard such a complex switch of units and assume no phase change during pressure conditioning.

Third, uncertain conditioning tasks may exist; e.g. it is not known, a priori, whether a cooling task or a heating task is required. To model the condition that cooling is needed if temperature decreases and heating if temperature increases (and similar for pressure conditioning), additional constraints (see the conditioning stream model in **Appendix 2**) are required, thus further increasing model complexity.

2.4.4 General form of the superstructure element models

In this section, we present a general form of the unit models. We categorize variables in terms of input/output, and constrained/free.

(1) Input variables $\mathbf{x}_{in}$: variables uniquely describing the element once they are determined;

Output variables $\mathbf{x}_{out}$: variables that can be calculated from input variables.

The input and output variables of each superstructure element are shown in **Figures 2.6-2.9**.

(2) Constrained variables $\mathbf{x}^c$: variables that must equal zero when the element is deactivated.

Free variables $\mathbf{x}^f$: variables that can take arbitrary values when the element is deactivated.

If $\mathbf{x}^c \neq 0$ when the element is deactivated, then the solution may violate our modeling purpose or make the model infeasible. For example, if we do not enforce zero-flow for deactivated streams, then the resulting solution may include multiple streams with $Y_s = 0$ but with non-zero flows, violating the purpose of, say, a single stream activation constraint (with $\sum_{s \in \mathbf{s}_p} Y_s = 1$). Variables $\mathbf{x}^c$ include extensive properties (e.g. flow), and other variables that scale with extensive properties (e.g. utility cost). Variables $\mathbf{x}^f$ can take arbitrary values when the corresponding element is deactivated; they describe internal unit (e.g. distillation reflux ratio, membrane concentrating factor), outlet port variables, and intensive properties (temperature and pressure).

The combinations of the two variable categories yield all types of variables: $\mathbf{x}_{\text{in}}^c/\mathbf{x}_{\text{out}}^c/\mathbf{x}_{\text{in}}^f/\mathbf{x}_{\text{out}}^f$. Thus, for any given superstructure element, the model has the following general form:

$$\left[\begin{array}{l} \mathbf{x}_{\text{out}}^c = f(\mathbf{x}_{\text{in}}^f, \mathbf{x}_{\text{in}}^c, Y) \\ \mathbf{x}_{\text{out}}^f = g(\mathbf{x}_{\text{in}}^f, \mathbf{x}_{\text{in}}^c) \\ \mathbf{x}_{\text{in}}^{\text{clo}} \cdot Y \leq \mathbf{x}_{\text{in}}^c \leq \mathbf{x}_{\text{in}}^{\text{cup}} \cdot Y \\ \mathbf{x}_{\text{in}}^{\text{flo}} \leq \mathbf{x}_{\text{in}}^f \leq \mathbf{x}_{\text{in}}^{\text{fup}} \end{array} \right] \quad (2.25)$$

where the binary Y is used to activate/deactivate the element. The first two equations represent all relationships between input and output variables. The last two equations specify variable domains. The binary variable Y is introduced into the 1st and the 3rd block to enforce $\mathbf{x}_{\text{in}}^c = 0$ and $\mathbf{x}_{\text{out}}^c = 0$ if $Y = 0$.

Note that the superstructure element models are all formulated based on component flow rates (F_c formulation), but they can also be modeled based on total flow rate and component fractions (F/X_c formulation). In fact, in specific cases the F/X_c formulation may be more natural (e.g. the use of Fenske equations). These two types of formulations will lead to the same solutions due to the basic relation $F_c = FX_c$.

2.4.5 Logic constraints

To formulate problems with lower complexity, the need for logic constraints has been previously recognized [112], [113], especially when dealing with superstructures featuring a large number of units and/or rich connectivity.

All the ports in a unit are required for its normal operation, and every port in the superstructure belongs to a particular unit. Hence, if a unit port is activated then the unit should be activated and, if a unit is activated then all its ports should be activated. Thus, port activation can be handled using appropriate unit variables, eliminating the need for port variables. Further, when a unit u is activated, at least one stream per unit port should be activated as well, so the unit has all the required connections to operate properly. Formally, this is expressed as:

$$Y_u \leq \sum_{s \in \mathbf{S}_p} Y_s, \quad u \in \mathbf{U}, p \in \mathbf{P}_u \quad (2.26)$$

Similarly, if a stream s is activated, the units connected to it (denoted by $u \in \mathbf{U}_s$) should be activated.

$$Y_s \leq Y_u, \quad s \in \mathbf{S}, u \in \mathbf{U}_s \quad (2.27)$$

Logic constraints other than these simple connectivity implications can be used to incorporate particular technology evaluation purposes [112], [113]. For example if our goal is to evaluate the potential of a series of competing technologies, we can enforce the selection of at most one of them.

2.4.6 Objective function

A general objective function is given as follows:

$$\max \sum_{p \in \mathbf{P}_{\text{Sink}}, c \in \mathbf{C}} \pi_p F_{p,c} - \sum_{u \in \mathbf{U}} (C_u^{\text{acc}} + C_u^{\text{oc}}) - \sum_{s \in \mathbf{S}} (C_s^{\text{acc}} + C_s^{\text{oc}}) \quad (2.28)$$

where $\mathbf{P}_{\text{Snk}}$ denotes ports connected to sinks; π_p denotes the price of the stream to/from sink/source port p ; and C_u^{acc} and C_u^{oc} are the annualized capital and operating costs of unit u . Changes can be made to account for other objectives, e.g. $\sum_{u \in \mathbf{U}} (C_u^{\text{acc}} + C_u^{\text{oc}}) + \sum_{s \in \mathbf{S}} (C_s^{\text{acc}} + C_s^{\text{oc}})$ can be used for the minimization of total annual cost.

2.5 Implementation

In this section, we present the software implementation architecture, including a description of the system of files and subroutines used to implement our framework. Notably, the implementation techniques presented here are designed to cope with large process synthesis problems, promoting ease of construction and modification of the model.

The formulation of the optimization model requires the use of a modeling environment. The description presented here assumes the use of GAMS (General Algebraic Modeling System), but the same architecture and models can be implemented in any other modeling language having comparable capabilities. **Figure 2.10** presents the general structure of the main file (ProcessNetwork.gms) as well as the required supporting files containing model data (XX-Input.gms, XX-Specifications.gms) and element models (ZZ-Model-YY.gms, WW_Model.gms) which are combined with the main model at compilation time.

The main model file is divided into eight parts. Part 1 includes the declaration and assignment of key control variables specifying the data and model files to use. These control variables include the "problem name" and the "model type". The problem name is a string "XX" used to identify files containing problem specific information. The string is incorporated in the files named "XX-Input.gms" and "XX-Specifications.gms". The model type is a string "YY" used to distinguish

between different math programming model types (e.g. “LP”, “MIP”, “MINLP”) for the same element “ZZ”. All element models are named “ZZ-Model-YY.gms”.

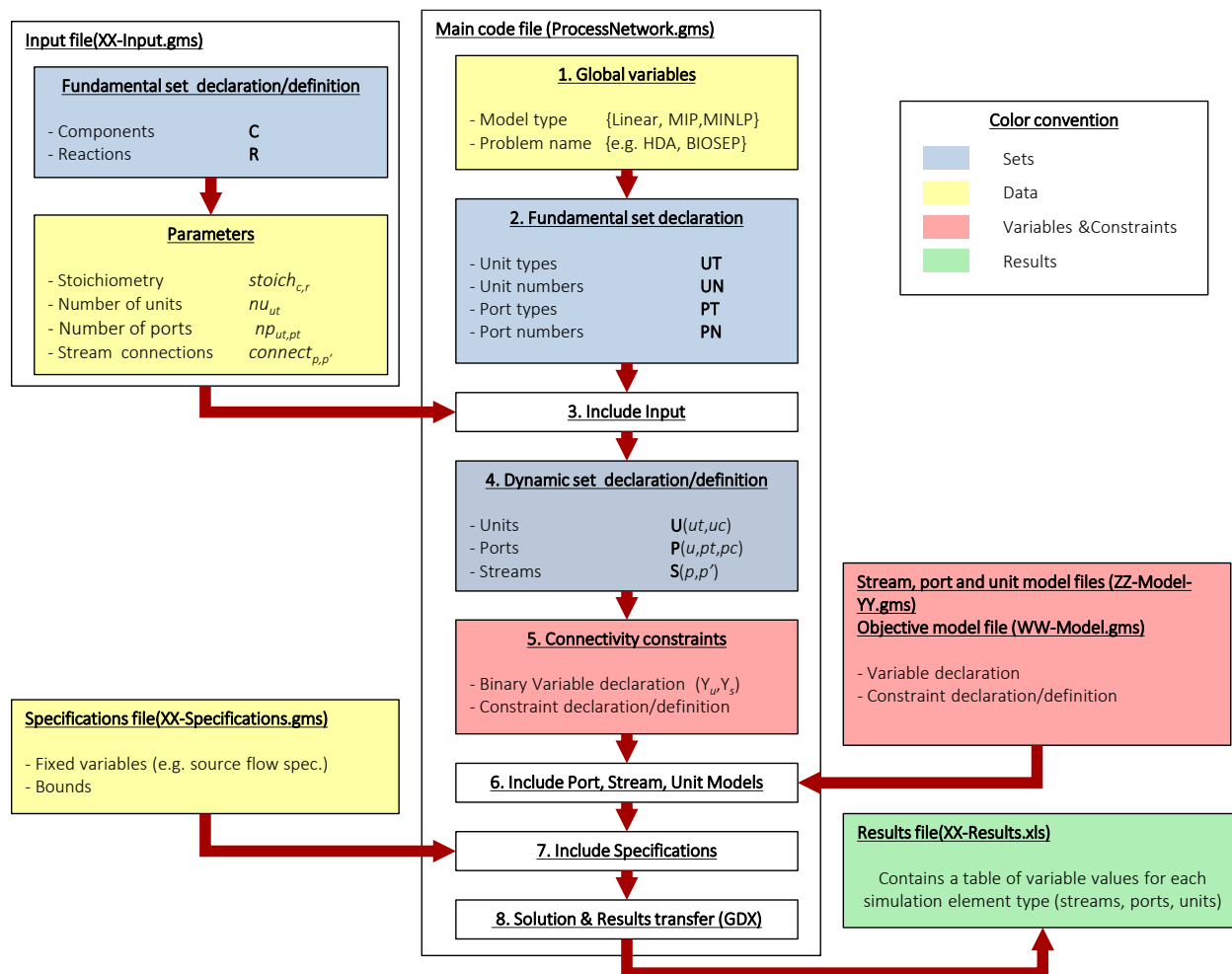


Figure 2.10. GAMS implementation of a general superstructure model.

Part 2 contains the declaration and definition of fundamental sets such as unit types **UT**, unit numbers **UN**, port types **PT**, and port numbers **PN**; as well as the declaration of all the sets and subsets. It also includes the declaration and assignment of parameter $np_{ut,pt}$ indicating the number of inlet and outlet ports in each unit type.

Part 3 incorporates a problem-specific input file “XX-Input.gms”. This file declares problem specific sets for components **C** and reactions **R**. It also declares and assigns problem specific parameters

such as stoichiometric coefficient $stoich_{r,c}$, vector nu_{ut} indicating the number of units per unit type included in the superstructure, and the connectivity binary matrix $connect_{p,p'}$ indicating the pairs of ports connected by streams. Thus, when a change in the superstructure connectivity is required we only need to update the connectivity matrix, without compromising the integrity of the model.

In Part 4, sets **U**, **P**, **S** are generated using the basic sets and the parameters previously defined. Specifically, **U** is generated based on parameter nu_{ut} , **P** is generated based on **U** and parameter $np_{ut,pt}$, and **S** is generated based on **P** and the connectivity matrix $connect_{p,p'}$. Using the same approach, the remaining subsets and multidimensional sets are generated.

Following the set generation, Part 5 contains unit and stream binaries declarations, as well as logic constraints. In Part 6, appropriate stream, port and unit model files "ZZ-Model-YY.gms" are included. Part 7 incorporates specification files, where the values of certain variables are fixed or bounded (e.g. raw material characteristics, final product requirements). Part 8 includes the solution statement, and other commands to gather and present the numerical results (i.e. spreadsheet files "XX-Results.xls").

2.6 Remarks

In this section, we discuss extensions that would allow us to apply special techniques to improve computational tractability.

2.6.1 Incorporation of non-conditioning streams

General conditioning streams provide a flexible way to model both temperature and pressure changes required for the normal operation of units. However, in some cases, additional information may be available that allows to remove unnecessary conditioning tasks, thus leading to the formulation of simpler models. Toward this end, traditional non-conditioning streams

(representing simple connections) along with conditioning units can be employed to improve computational tractability. For clarity, we distinguish the terms “connection” and “stream” here. Specifically, as shown in **Figure 2.11**, the connection between two ports of main reaction/separation units can be made using a conditioning stream, a non-conditioning stream, or a combination of non-conditioning streams and conditioning units. For example, if the conditions of a unit outlet port are adequate for the normal operation of a downstream unit inlet port, and the objective function is independent of the conditions of this inlet port, then a non-conditioning stream can be adopted for the connection. If it is known prior to optimization that only temperature conditioning is required, then a non-conditioning stream together with a heat exchanger can be adopted.

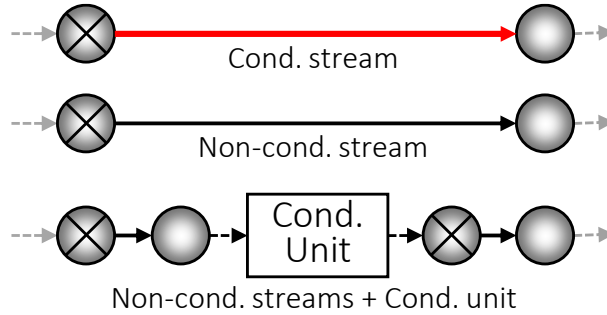


Figure 2.11. Three possible connection types between two major unit ports: a conditioning stream; a non-conditioning stream; non-conditioning streams and one or more conditioning units. Black arrows represent non-conditioning streams; the thick red arrow represents a conditioning stream.

The non-conditioning stream ($s \in \mathbf{S}^{\text{NC}}$) model is as follows.

$$-\Delta p_s^{\text{up}}(1 - Y_s) \leq p_s^{k1} - p_s^{k3} \leq \Delta p_s^{\text{up}}(1 - Y_s) \quad (2.29)$$

$$T_s^{k3} = T_s^{k1} \quad (2.30)$$

$$Q_s, W_s, C_s^{\text{acc}}, C_s^{\text{oc}} = 0 \quad (2.31)$$

$$F_{s,c}^{\text{lo}} Y_s \leq F_{s,c} \leq F_{s,c}^{\text{up}} Y_s, \quad c \in \mathbf{C} \quad (2.32)$$

where Δp_s^{up} is the upper bound of $|P_s^{k1} - P_s^{k3}|$. **Equation 2.29** enforces that if the stream is activated ($Y_s = 1$) then $P_s^{k3} = P_s^{k1}$; otherwise ($Y_s = 0$), the equality is void. The necessity of such a formulation (instead of a simple $P_s^{k1} = P_s^{k3}$ equation) is illustrated in **Appendix 1.3**.

2.6.2 Single-stream ports

In some cases, we can identify “single-stream” outlet ports, which allow the activation of at most one outgoing stream [22]. For example, in separation processes, we typically do not split a stream and feed the splits to parallel units for separation (unless for recycling). Therefore, if the outgoing streams from an outlet port p' are fed to parallel separation units, then p' can be regarded as a single-stream outlet port; in the optimal solution only one stream will be selected. The single-stream outlet port model is as follows.

$$F_{p,c} = \sum_{s \in \mathbf{S}_p} F_{s,c}, \quad c \in \mathbf{C} \quad (2.33)$$

$$T_p = T_s^{k1}, P_p = P_s^{k1}, \quad s \in \mathbf{S}_p \quad (2.34)$$

$$\sum_{s \in \mathbf{S}_p} Y_s = Y_u, \quad u \in \mathbf{U}, p \in \mathbf{P}_u \quad (2.35)$$

A key benefit of using single-stream outlet ports is the removal of the stream split fraction variables J_s in **Equation 2.12**, thus avoiding the bilinear terms.

Single-stream inlet ports also exist in some cases. For example, as shown in **Figure 2.12**, if the outlet port of unit Mxs,1 has been assumed to be a single-stream outlet port, then port p must be a single-stream inlet port (see explanation in the figure caption). The single-stream inlet port model is as follows.

$$F_{p,c} = \sum_{s \in \mathbf{S}_p} F_{s,c}, \quad c \in \mathbf{C} \quad (2.36)$$

$$T_p = T_s^{k3}, P_p = P_s^{k3}, \quad s \in \mathbf{S}_p \quad (2.37)$$

$$\sum_{s \in \mathbf{S}_p} Y_s = Y_u, \quad u \in \mathbf{U}, p \in \mathbf{P}_u \quad (2.38)$$

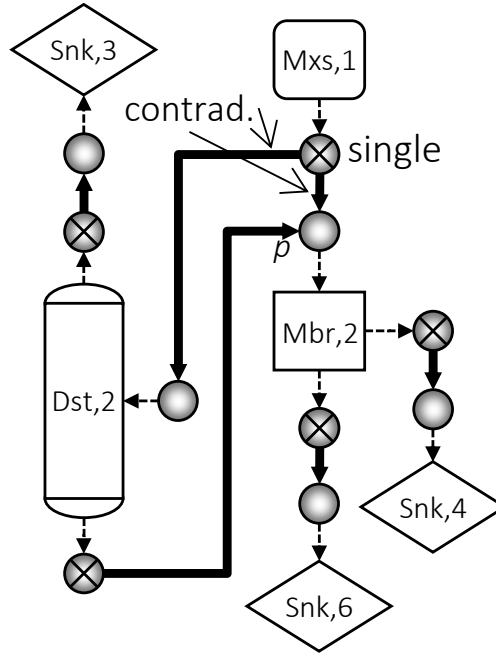


Figure 2.12. Single-stream port example. The outlet port of Mxs,1 is already known to be a single-stream port (marked as “single”). If both streams connected to port p are activated, then all the streams represented with thick black arrows must also be activated, leading to contradiction with the single-stream outlet port assumption for Mxs1. Thus, port p is a single-stream inlet port.

An advantage of using single-stream inlet ports is the avoidance of the non-linear enthalpy **Equation 2.9**. Also note that if a non-conditioning stream is connected to a single-stream inlet port, then the non-conditioning stream temperature relation (**Equation 2.30**) has to be replaced by the following equation to remain consistent with the single-stream inlet port model.

$$-\Delta t_s^{\text{up}}(1 - Y_s) \leq T_s^{k1} - T_s^{k3} \leq \Delta t_s^{\text{up}}(1 - Y_s) \quad (2.39)$$

where Δt_s^{up} is the upper bound on $|T_s^{k1} - T_s^{k3}|$.

2.6.3 Conditioning reduction

Conditioning tasks are usually modeled using non-linear thermodynamic equations. To improve computational tractability, we can reduce the number of conditioning streams and conditioning units without reducing the feasible space.

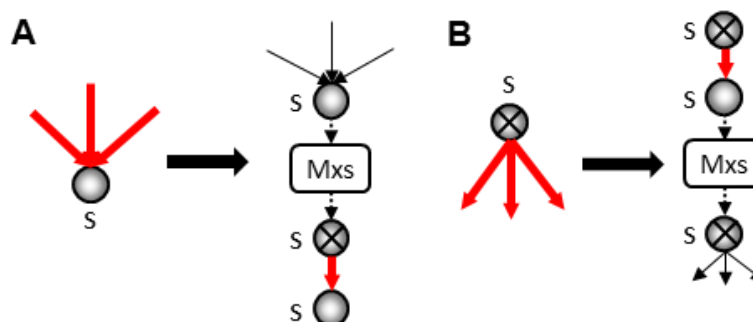


Figure 2.13. Conditioning reduction at single-stream ports. **(A)** Reduction at single-stream inlet port; **(B)** reduction at single-stream outlet port. “S” represents single-stream port. The red arrows represent conditioning streams; the black arrows represent non-conditioning streams. The ports in the simplified structure are all single-stream ports.

The use of single-stream ports has the additional benefit of reducing conditioning. For example, as shown in **Figure 2.13**, where multiple conditioning streams enter/leave a single-stream inlet/outlet port, the structure can be simplified through the use of a mixer-splitter whose ports are single-stream ports, resulting in fewer conditioning streams but maintaining the same conditioning ability. Note that such modification cannot be applied if the original port is not a single-stream port, because the single conditioning stream in the modified structure cannot meet the conditioning requirement of all the activated streams.

step that further purifies the product if necessary. In this example, we consider the bio-separation process superstructure presented in **Figure 2.14**, with a dilute mixture of extracellular water-soluble product (500 kg/hour), cells (500 kg/hour), and water (9000 kg/hour) as the feed. We aim to obtain a product stream with a concentration of at least 90 wt% satisfying a minimum recovery of 90%. We assume that all impurities are inside the cells; thus the extracellular product does not mix with the intracellular impurities. The feed components therefore include the product, cells and water. The superstructure consists of the following major units: two sources (Src,1-2), one product sink (Snk,6), five waste sinks (Snk,1-5), one sedimentation unit (Sdm,1), two distillation units (Dst,1, Dst,2), one extraction unit (Ext,1), and two membrane units (Mbr,1, Mbr,2), where Mbr,1 functions to separate the cell while Mbr,2 separates the product from water. The total number of inlet ports, outlet ports and streams is 14, 14 and 18, respectively. The goal is to find the best process configuration, system utility requirements, and sizing of the units, minimizing total annualized cost.

The MINLP model is presented in **Appendix 2** (for unit models, refer to equations only applicable to the units existing in this case). The objective function is $U = \sum_u C_u^{\text{atc}} + \sum_s C_s^{\text{atc}}$.

The formulation involves a total of 490 variables, 30 binary variables, and 486 equations. The optimization is performed using GAMS 24.4.6 –BARON [57]. There are $2 \times 4 = 8$ potential configurations. The optimal path is shown in **Figure 2.15**. The solution involves the following units, with characteristic sizes in parentheses: a sedimentation tank (total area = 227 m³); an extraction unit (volume = 6.6 m³); and a distillation unit (volume = 6 m³). The objective value (total annualized cost) is 2.1 million \$/year, leading to a unit cost of 0.6 \$/kg product (pure product basis); the total utility consumption is 1355 KW (distillation heating/cooling: 565/565 KW; conditioning heating/cooling/pumping electricity: 205/20/0.1 KW). The specific utility consumptions, component flow rates of the final product stream, and costs of units and conditioning streams are

presented in **Figure 2.15**. The optimization problem was solved on a PC with an Intel (R) Core (TM) i5-3570 CPU @ 3.40 GHz. The model is solved to optimality in 13 minutes. The time required to solve to 5% gap is 3 minutes.

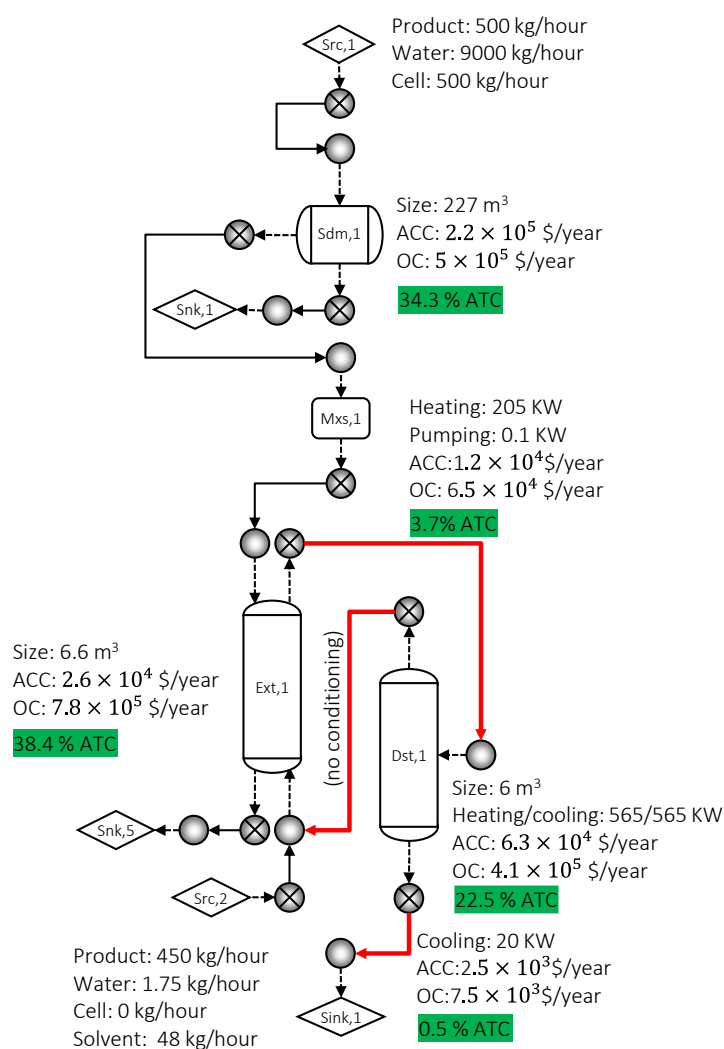


Figure 2.15. Optimal process for the bio-separation example.

2.8 Conclusions

In this work, we developed a comprehensive framework for superstructure representation, generation, and modeling, facilitating efficient initial formulation and later modification of superstructure-based optimization models. In terms of representation, we introduced unit ports as

the key connecting elements. Also, the use of conditioning streams allows us to treat the main reaction/separation tasks and the conditioning tasks separately in a more flexible manner. In terms of generation, we proposed four connectivity rules based on minimal and feasible component sets for each unit port, which lead to the simplest superstructure containing all feasible configurations. In terms of modeling, we developed general models for the superstructure elements (units, ports, and streams) and an implementation architecture. Of particular importance is the modular nature of our modeling framework due to its tight combination with the proposed generation approach. Our representation approach, which features modularity, allows straightforward superstructure generation and easy model generation. This work lays the foundation for the studies in the subsequent chapters.

Chapter 3

3 Bio-separation process synthesis framework³

3.1 Introduction

As discussed in **Chapter 1**, separation accounts for 60-80% of the total bio-production cost in many cases, and the synthesis of efficient separation networks considering competing alternative technologies is essential. Superstructure optimization has been proposed for the synthesis of separation networks [36], [51], [90]–[92], [114] as well as the development of bio-processes [93]–[101]. However, these studies were mostly performed for specific products on a case-by-case basis. In this chapter, we develop a general framework for bio-separation network synthesis, which allows the generation of the optimal separation processes for liquid and solid products produced through microbial conversion. The framework provides guidance on the preliminary synthesis of separation networks, thereby aiding the analysis of bio-based chemical production technologies.

The remaining of the paper is structured as follows. In **Section 3.2**, we provide a general overview of our framework. In **Section 3.3**, we discuss the generation of stage-superstructures. In **Section 3.4**, we present the stage-superstructures. In **Section 3.5**, we discuss the general superstructure. In **Section 3.6**, we present methods to generate a reduced superstructure for specific instances. In **Section 3.7**, we discuss our modeling approach.

³ This chapter includes content from: W. Wu, K. Yenkie, and C.T. Maravelias, “A superstructure-based framework for bio-separation network synthesis,” *Comput. Chem. Eng.*, vol. 96, pp. 1-17, 2017. Wenzhao Wu developed the framework (including superstructure generation, reduction and modeling) and wrote the manuscript; Kirti Yenkie participated in the discussion of superstructure modeling.

3.2 Framework overview

Chemical products can be divided into three classes: commodity chemicals, fine chemicals, and specialty chemicals [115], [116]. Commodity and fine chemicals are pure chemical substances with specified purity grades. Specialty chemicals are formulations of chemicals containing one or more pure chemical substances as active ingredients. Our framework does not incorporate the blending of chemicals for the formulation of specialty chemicals containing more than one pure ingredients. Instead, we assume a single chemical substance as being the *product*. Moreover, some technologies are usually suitable only for fine and specialty chemicals in industrial production (e.g. chromatography). Hence, as a distinction, we refer to fine and specialty chemicals with single ingredients as “SPC”, and commodity chemicals as “CMD”.

We consider a general effluent (i.e. the “initial product stream”) from an upstream bioreactor, which typically contains cells, the product (if extracellular), large amount of water, and other impurities. It must be concentrated and purified to a specific industrial grade before entering the market (thus becoming the “final product stream”). We first identify the key stages in bio-separation processes, by investigating general process synthesis heuristics, e.g., removing the most plentiful and easiest-to-remove impurities first [31]; general bio-separation principles [5], [31], [117]; and insights obtained from industrial separation processes of specific products, such as lactic acid, citric acid, polyhydroxyalkanoate (PHA), 3-Hydroxypropionic acid (3HP), ethanol, and β -galactosidase [118]–[125]. Consequently, four key separation stages are identified:

Stage 1 - cell treatment, where cells are harvested and then disrupted to release intracellular products (present if the product is intracellular; bypassed if the product is extracellular);

Stage 2 – product phase isolation, where the phase that contains the product is isolated;

Stage 3 - concentration and purification, where water and impurities are removed; and

Stage 4 – refinement, where the product is further refined.

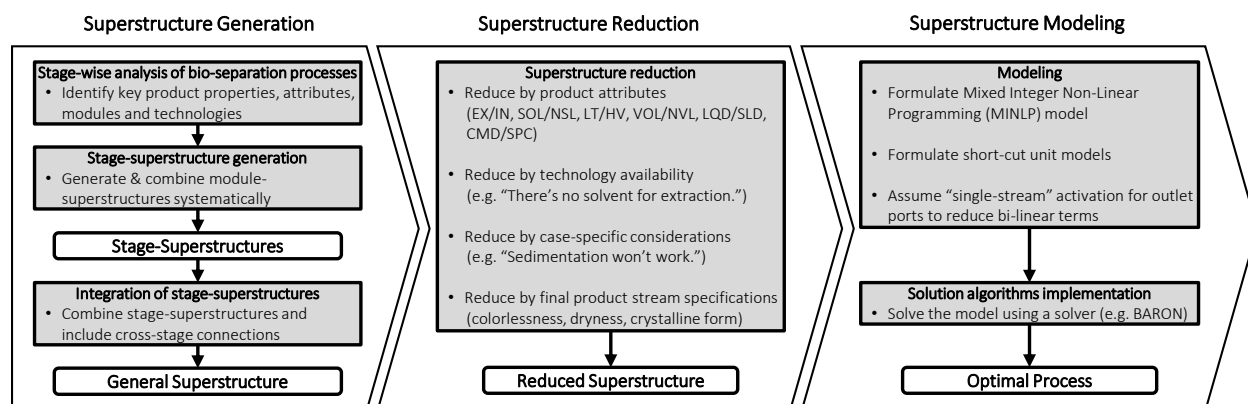


Figure 3.1. Overview of the superstructure-based framework for bio-separation network synthesis.

Based on the four stages, we first perform a stage-wise analysis of general bio-separation processes. Then, for each stage, we systematically develop a stage-superstructure. Subsequently, all of the stage-superstructures are integrated to generate a general superstructure that accounts for all types of products. To address specific instances (e.g. when the product is extracellular, soluble, volatile, etc.), a reduction method is proposed, leading to a reduced superstructure that excludes non-applicable configurations. Finally, we formulate and solve an optimization model to identify the optimal process. These steps are summarized in **Figure 3.1**. Note that we employ a systematic method to generate the general superstructure. Therefore, when dealing with a specific case, the generation step can be bypassed (unless the general superstructure needs to be modified), and start from the reduction step.

Note that we do not explicitly account for the separation of co-products. However, the current framework is sufficient in some cases where co-products exist: e.g., if the product is volatile and the co-product is non-volatile, then a distillation unit used for separating the product (top stream) from water (bottom stream) will also remove the co-product.

3.3 Stage-superstructure generation

3.3.1 Stage-wise analysis

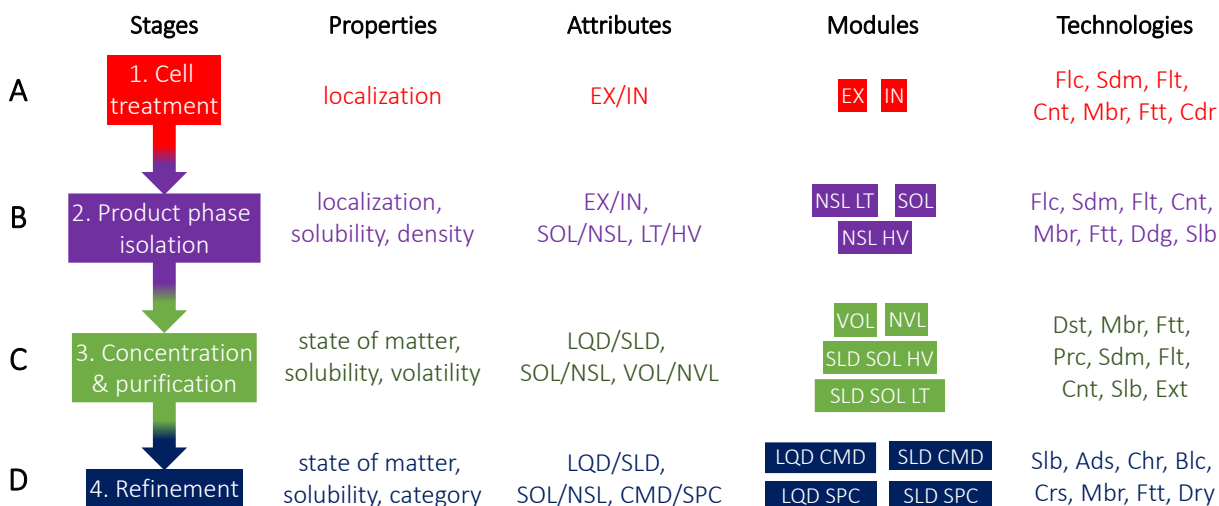


Figure 3.2. Stage-wise analysis of general bio-separation processes. (A)-(D): key product properties, attributes, modules and technologies for the corresponding stages. Key product properties (and the corresponding attributes) include: localization (extracellular-EX/intracellular-IN), solubility in water (soluble-SOL/insoluble-NSL), density compared to water (light-LT/heavy-HV), volatility with respect to water (volatile-VOL/non-volatile-NVL), state of matter (liquid-LQD/solid-SLD), and category (commodity chemicals-CMD/single-component specialty chemicals and fine chemicals-SPC). Technologies include flocculation (Flc), sedimentation (Sdm), flotation (Flt), centrifugation (Cnt), membrane (Mbr), filtration (Ftt), cell disruption (Cdr), cell debris differential digestion (Ddg), distillation (Dst), extraction (Ext), precipitation (Prc), solubilization (Slb), adsorption (Ads), chromatography (Chr), bleaching (Blc), crystallization (Crs), and Drying (Dry).

For each stage, we identify the following (see **Figure 3.2**):

- (1) Key properties of the product, such as solubility and density, and the corresponding attributes (specifications of the properties), e.g. soluble/insoluble in water as a specification of solubility;
- (2) Key modules, which are superstructures generated for different combinations of product attributes; the combination of these modules in each stage leads to a complete stage-superstructure;
- (3) Technologies that can accomplish the separation task in the stage.

Relevant product properties (and the corresponding attributes) include localization (extracellular-EX/intracellular-IN), solubility in water (soluble-SOL/insoluble-NSL), density compared to water (light-LT/heavy-HV), volatility compared to water (volatile-VOL/non-volatile-NVL), state of matter at normal conditions (liquid-LQD/solid-SLD), and category (commodity chemicals-CMD/specialty chemicals-SPC). For simplicity, hereafter, we refer to product attributes using their abbreviations. The attributes can be used in combination to describe a product, e.g., a product that is insoluble and light is denoted by NSL LT. They can also describe modules, e.g., a module generated for an NSL LT product is also denoted by NSL LT. Note that the properties, attributes and modules relate to the product (a pure component), but not the streams containing the product (a mixture of components). For example, a SLD product can be dissolved in water and thus the product-containing stream is a liquid.

Stage 1 – cell treatment

In Stage 1, an EX product (secreted to the outside of the cells) requires no processing, while an IN product (located inside the cells) undergoes cell harvesting and cell disruption (see **Figure 3.2**). Cell harvesting increases cell concentration by removing extracellular liquid (mainly water). Next, the harvested cells (assumed to be insoluble in water and have density higher than water) are disrupted to release the intracellular product; i.e., the cells are converted to the product, cell debris (assumed to be insoluble in water and have density higher than water), and other soluble components. Accordingly, the key property determining the separation process in this stage is the localization of the product (IN/EX). Hence, only two key modules exist in Stage 1: the IN module leading to cell harvesting followed by cell disruption, and the EX module leading to no processing (i.e. bypass). Typical technologies include sedimentation, flotation, centrifugation, membrane, filtration, and cell disruption. Flocculation can also be added in the beginning to pre-treat the stream for more efficient harvesting.

Stage 2 – product phase isolation

In Stage 2, the phase containing the product is isolated (see **Figure 3.2**). Distribution of the product is in general determined by three properties: localization (EX/IN), solubility in water (SOL/ NSL), and relative density with respect to water (LT/ HV). The treatment in Stage 1 guarantees that the product has been released out of the cells. Thus, the product distribution is determined by the combination of SOL/NSL and HV/LT attributes. Specifically, the product is dissolved in the water phase if it is SOL (no matter HV or LT); it floats to the top (given enough time) if the product is NSL LT; it settles to the bottom together with the cells (or debris) and other insoluble impurities if the product is NSL HV. Hence, the key modules in this stage are SOL, NSL LT, and NSL HV. For the NSL HV case, if the product is LQD, then it can be directly separated from the cells or debris using membrane or filtration. If the product is SLD, then it can be dissolved with a solubilizing agent, and thus the product is transferred into the water phase, leaving mainly the cells or debris at the bottom. Alternatively, if the product is IN SLD, differential digestion can be used to dissolve cell debris [123], [124], thus leaving mainly the product at the bottom for easier separation. Technologies applicable for phase separation in this stage include flocculation, sedimentation, flotation, centrifugation, membrane, filtration, differential digestion, and solubilization.

Stage 3 – concentration and purification

In Stage 3, the product concentration is increased (large amount of water is removed) and impurities are removed, by utilizing the differences between the product and the other components in terms of volatility, molecular size, diffusivity, solubility in solvents, ability to precipitate, etc (see **Figure 3.2**). The key properties in this stage are product state at normal condition (LQD/ SLD), solubility (SOL/NSL), and relative volatility with respect to water (VOL/NVL). The product state and solubility properties determine whether precipitation is a viable option: precipitation is only

viable if the product is SLD SOL. Volatility determines the outlet port from which the product leaves a distillation column (either used directly, or following extraction to recover the solvents). Accordingly, we examine the SLD SOL LT and SLD SOL HV modules to account for the use of precipitation, and the VOL and NVL modules to account for the use of distillation and extraction. Relevant technologies include distillation, membrane, extraction, precipitation, and phase separation technologies (sedimentation, flotation, centrifugation, membrane, and filtration) required after precipitation. Also, solubilization may be needed after precipitation to convert the product back to its original dissolved form. The use of sedimentation, flotation, filtration and centrifugation in this stage is excluded because these technologies have been employed already in the first two stages. On the other hand, due to the variety of membrane types (e.g. micro-, ultra- and nano-filtrations), membrane technologies are still included in this stage.

Stage 4 – refinement

In Stage 4, trace impurities are further removed and refinement operations are performed to satisfy special final product stream specifications, such as colorlessness, complete dryness, and crystalline form (see **Figure 3.2**). The key properties in this stage are state of matter (LQD/SLD) and category (CMD/ SPC). Hence, the key modules in this stage are LQD CMD, SLD CMD, LQD SPC and SLD SPC. Typical technologies in this stage include adsorption, chromatography, bleaching, crystallization, drying, solubilization (to dissolve solid particles, thus allowing a solid-free stream required by the adsorption, chromatography, bleaching, and crystallization feed conditions), and sedimentation and filtration required after crystallization (since the crystal size is usually large, membrane, centrifugation and flotation are excluded). Chromatography and crystallization are usually applicable to SPC chemicals but not to CMD chemicals. In addition, crystallization and drying require the product to be SLD.

3.3.2 Module-superstructure generation

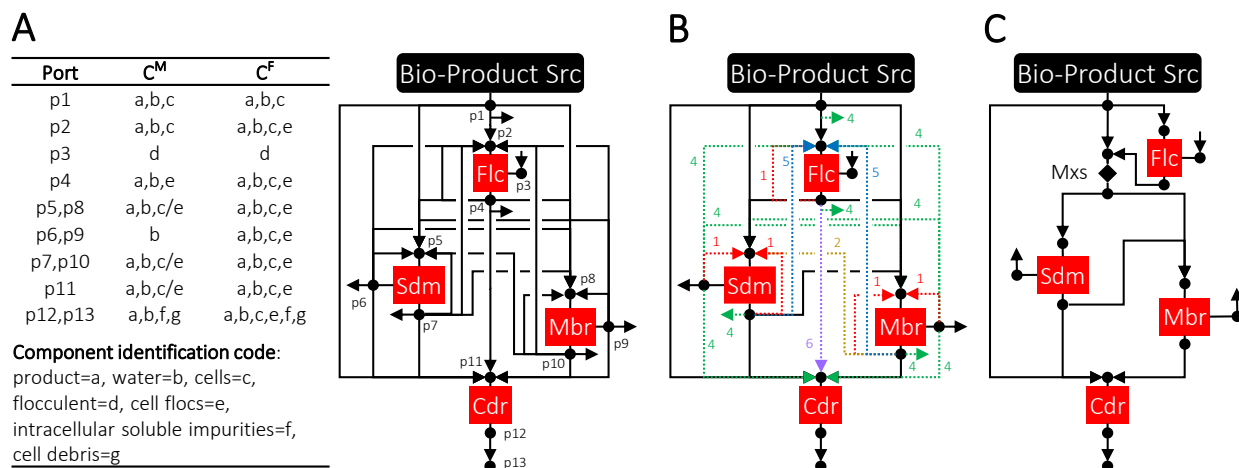


Figure 3.3. Simplification of the superstructure for the IN module in Stage 1, with reduced number of technologies as an example. **(A)** The superstructure obtained after applying four connectivity rules on the fully connected superstructure; **(B)** further simplification based on the six additional connectivity rules; **(C)** final simplified superstructure. Colored boxes represent units. Circles that are connected to units represent their ports; the port type (inlet/outlet) can be identified based on the direction of the connected streams. In **(A)**, C^M and C^F represent minimal and feasible component sets, respectively, for the corresponding ports p1, p2, etc. In **(B)**, the removed connections are marked with dashed lines, and the numbers of the corresponding rules applied are labeled accordingly.

Based on the identified key modules in each stage, we first generate the module-superstructures using the systematic approach developed in **Chapter 2**. Specifically, we first generate a fully connected superstructure for each module based on the technologies in the corresponding stage. We assume that each unit in the fully connected superstructure can complete its assigned task, and thus we do not consider the same type of units used in series or in parallel for a single task. Then we apply the four connectivity rules to the fully connected superstructure to obtain a simplified superstructure. An example of the simplified superstructure is shown in **Figure 3.3A**, for the IN module in Stage 1, but only flocculation, sedimentation, membrane and cell disruption are included for the sake of demonstration. Also, to simplify the presentation, unit numbers are omitted in **Figure 3.3** and hereafter, except in the reduced superstructures to be discussed in **Sections 3.6**.

Furthermore, since we are interested in a preliminary assessment, we apply the following six additional connectivity rules which are based on simplifications.

Rule 1: The outlet ports of a unit should not be connected to the unit's own inlet ports. The outlet ports of a separation unit are either product lean or product rich, and thus feeding the separated streams back to the inlet port is counter-productive. For further separation, the outlet ports should be connected to other separation units. For instance, as shown in **Figure 3.3A**, the concentration of the cells/flocculated cells in the sedimentation unit (Sdm) outlet port p7 is higher than that in the inlet port p5. Therefore, p7 should not be connected to p5. Furthermore, units that function to pre-treat streams for more efficient downstream separation (which we refer to as “pre-treatment units”), e.g. flocculation units used before phase separation, are assumed to be highly efficient, and thus recycling outgoing streams back to their own inlet ports is also considered unnecessary.

Note that in principle, each of the purged streams (see **Figure 3.3A**) can be connected to a sink unit. Similarly, source units can be used to provide additional input streams (e.g. the flocculent stream entering the flocculation unit). However, for simplicity, we use a single dummy unit that functions as a “reservoir” (“Rsv”), not shown in the superstructures, with one inlet port and one outlet port. The reservoir provides additional input streams and receive all waste streams.

Rule 2: If multiple technologies are used in series for the same task (e.g. sedimentation and centrifugation for cell harvesting), then more expensive ones (usually with higher performance such as separation resolution and throughput) should be used after cheaper ones (e.g. centrifugation is used after sedimentation; chromatography is used after adsorption). If the cost and performance of the technologies are similar or difficult to assess (e.g., because different driving forces are used), then we sequence them based on industrial best practices, if available.

Rule 3: Filtration and membrane should not be used in series. This is because the use of filtration or membrane, as well as the type of membrane used, is pre-determined based on the characteristics (e.g. molecular/particle size) of the components to be separated.

Rule 4: Typically, only one of the outlet ports of a unit is rich in the product, therefore, only the product-rich outlet port should be connected to the inlet ports of other units, while the other outlet ports are connected to the reservoir inlet port (i.e. the streams from these ports are purged), except for the following cases:

If the product is dissolved in water in the inlet port of a sedimentation, flotation or centrifugation unit (i.e. the product-rich stream will exit the unit through the outlet port for the middle water phase), then the product-lean stream (in the outlet port for the bottom phase) can also be fed to other units for further separation. This is because the bottom phase may still have a substantial amount of dissolved product.

If a distillation unit is used to recycle solvent from an upstream extraction unit, then one outlet port of the distillation unit is connected to other units for further purification of the product-containing stream, while the other is connected to the inlet port of the extraction unit for solvent recycling.

Rule 5: Pre-treatment technologies should be placed at the beginning of each stage, to facilitate separation in the current stage, or the end, to facilitate separation in the next stage.

Rule 6: Connections that do not facilitate any processing tasks in the stage should be removed. For example, the direct connection from the flocculation unit to the cell disruption unit in Stage 1 facilitates neither the harvesting of cells, nor the disruption of the cells. Thus, it should be removed.

The implementation of these rules is demonstrated in **Figure 3.3B** and **Figure 3.3C**. Note that mixer-splitters can be used to simplify the superstructure when multiple units are fully-interconnected as shown in **Section 2.3**. A mixer-splitter (represented by a diamond as shown in

Figure 3.3C) has one inlet port and one outlet port, which is used to mix and split streams at the same time. For example, it is used after the flocculation unit in **Figure 3.3C** to represent the by-pass of the flocculation unit in a simpler fashion, instead of connecting the outlet ports of the source and flocculation units directly to the inlet ports of the sedimentation and membrane units. In addition, we name the ports connecting the upstream and downstream stages in the superstructures as “upstream/ downstream stage-connecting ports”. For example, as shown in **Figure 3.3A**, the downstream stage-connecting port p13 at the bottom of the superstructure represents the connection to Stage 2.

3.3.3 Combination of module-superstructures

Next, we combine module-superstructures to generate stage-superstructures. The specific steps are as follows.

Step 1: Using the stage-connecting ports as anchoring points, align in parallel all the module-superstructures (see examples in **Section 3.4**).

Step 2: Further simplify the combined superstructures for Stages 3 and 4, by reducing the number of units of the same type, while ensuring that all the module-superstructures are still uniquely embedded in the simplified stage-superstructures. The specifics are discussed in **Sections 3.4.3** and **3.4.4**.

The combined superstructures include both units and modules. Modules are represented by dashed rounded rectangles. Also, we use dashed rectangles to indicate the product attributes to which the units enclosed by the dashed rectangles are applicable. Specific examples can be found in **Section 3.4**.

To improve graphic representation of the superstructures, placement and numbering of unit ports should be arranged intuitively as follows:

(1) On each unit, we place outlet ports with lower density (or higher volatility) streams above those with higher density (or lower volatility) streams, so the function of a unit can be easily identified. For example, if the stream leaving the bottom outlet port of a membrane unit is purged (while the one leaving the top outlet port is fed to other units), then this membrane unit functions to remove the concentrate stream containing e.g., cells.

(2) An extraction unit is assumed to be a counter-current exchange system, where the solvent stream enters at the bottom port and leaves at the top port, and the product-containing stream enters at the other top port and leaves at the other bottom port.

(3) For each unit, we assign smaller numbers to ports with lower density (or higher volatility) streams, and assign larger ones to ports with higher density (or lower volatility) streams.

(4) For each unit, we number the inlet port receiving the product-containing stream as 1, and the other inlet ports as 2,3...

Specific placement and numbering of each unit port can be found in **Appendix 2**.

3.4 Stage-superstructures

3.4.1 Stage 1 superstructure

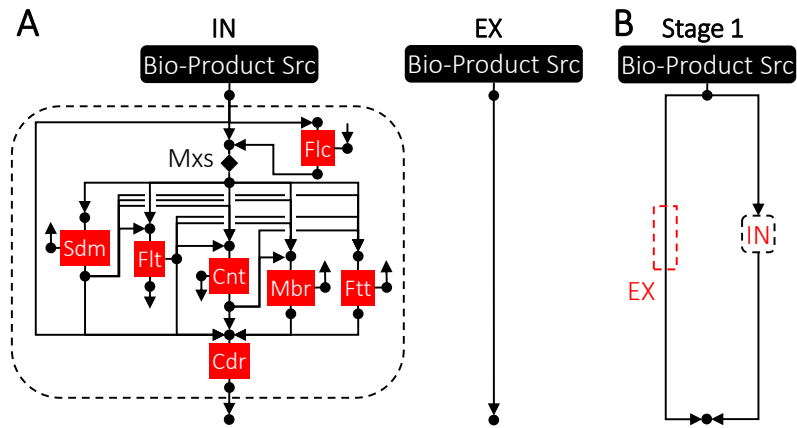


Figure 3.4. Module- and stage- superstructures for Stage 1. (A) Module-superstructures; (B) stage-superstructure, obtained by combining the module-superstructures in (A). Dashed rectangles indicate the attributes (in this case, EX) to which the enclosed units (in this case, none) are applicable. Dashed rounded rectangles indicate modules generated for specific combinations of attributes (in this case, IN). The same presentation is used in the subsequent figures.

After applying the methods discussed in **Sections 3.3.2** and **3.3.3**, the module-superstructures and the stage-superstructure for Stage 1 are shown in **Figure 3.4A** and **Figure 3.4B**, respectively. Flocculent in flocculation units only binds to cells and cell debris (no influence on the product), and is assumed to be completely consumed in the units. Different types of cell disruption are available (such as bead milling and homogenization), however, for the purpose of preliminary assessment, it is assumed that the specific type is pre-determined.

3.4.2 Stage 2 superstructure

The module-superstructures and the stage-superstructure for Stage 2 are shown in **Figure 3.5A** and **Figure 3.5B**, respectively.

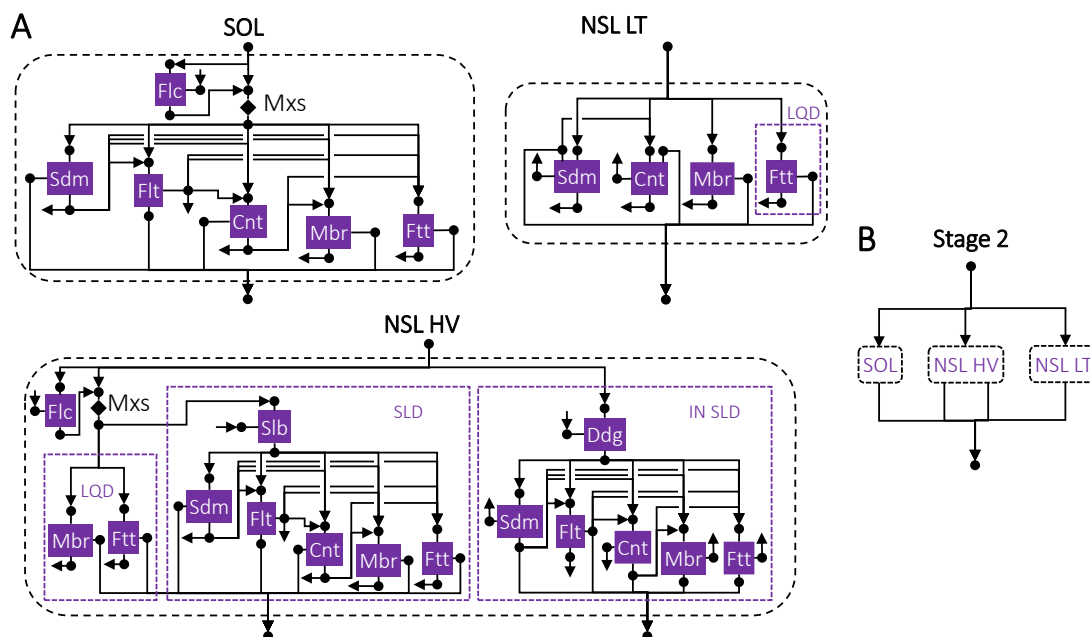


Figure 3.5. Module- and stage- superstructures for Stage 2. (A) Module-superstructures; (B) stage-superstructure, obtained by combining the module-superstructures in (A).

For the sedimentation, flotation and centrifugation in the SOL module in **Figure 3.5A**, the outlet ports containing undesired particles (cells or cell debris) are also connected to other units for further separation (see Rule 4a in **Section 3.3.2**), because the corresponding stream may still contain a substantial amount of dissolved product (due to the limitation of concentrating factors).

For the NSL LT module, sedimentation and centrifugation with three outlet ports are used [125]–[128]. One outlet port (at the top of the units) is for the top phase containing the LT product, one (in the middle) is for the water phase, and the other (at the bottom) is for the bottom phase containing cells/debris. Note that flocculation is not used prior to sedimentation and centrifugation, because we assume that no cells/debris end up in the top phase, and no product ends up in the bottom phase. Thus, addition of a flocculent does not facilitate the separation of the product. Also, flotation is not considered because it will cause the cells/debris to rise to the top phase, which is counterproductive. Membrane and filtration remove cells/debris, and thus the product-rich

streams (which are cell/debris-free) from the sedimentation and centrifugation units are not fed to the membrane and filtration units. Also note that filtration can only be used when the product is LQD, because the cells/debris cannot be separated from an otherwise NSL SLD product.

For the NSL HV module, flocculation is not used prior to differential digestion, because cells/debris are to be dissolved in the digestion unit, not separated. Also, it is assumed that only the product is dissolved in solubilization units.

3.4.3 Stage 3 superstructure

The module-superstructures for Stage 3 are shown in **Figure 3.6A**. It is assumed that the addition of precipitant leads to the precipitation of the product only. The combined stage-superstructure obtained after applying Step 1 in **Section 3.3.3** is shown in **Figure 3.6B**. The final simplified Stage 3 superstructure is obtained after applying Step 2 in **Section 3.3.3** (see **Figure 3.6C**). For example, the four membrane units shown in **Figure 3.6B** serve the same separation task, namely, isolating the product in the concentrate stream and removing the other components (mainly water) through the permeate stream. Therefore, these membrane units are combined into one (which is applicable to all product attributes), and all the connections to/from the original units are combined accordingly. Thus, as shown in **Figure 3.6C**, the inlet port of the combined membrane unit is connected to five outlet ports: the upstream stage-connecting port, the outlet ports of the distillation units directly used for VOL and NVL products, and the outlet ports of the distillation units (for VOL and NVL products) that are used for solvent recovery after extraction. Note that the four distillation units cannot be combined, because their separation tasks are different. For instance, the distillation unit directly used for VOL products has the product-rich stream at the top while the one for the NVL products has the product-rich stream at the bottom. Accordingly, both distillation units are kept in the final Stage 3 superstructure, and they are applicable to VOL and

NVL products, respectively. Note that all module-superstructures (**Figure 3.6A**) are still uniquely embedded in the simplified stage-superstructure (**Figure 3.6C**). To verify, in **Figure 3.6C**, if we eliminate all the inapplicable units for a given module (VOL, NVL, SLD SOL LT, or SLD SOL HV) and the connections to/from them, then the resulting superstructure is the same as the corresponding module-superstructure in **Figure 3.6A**.

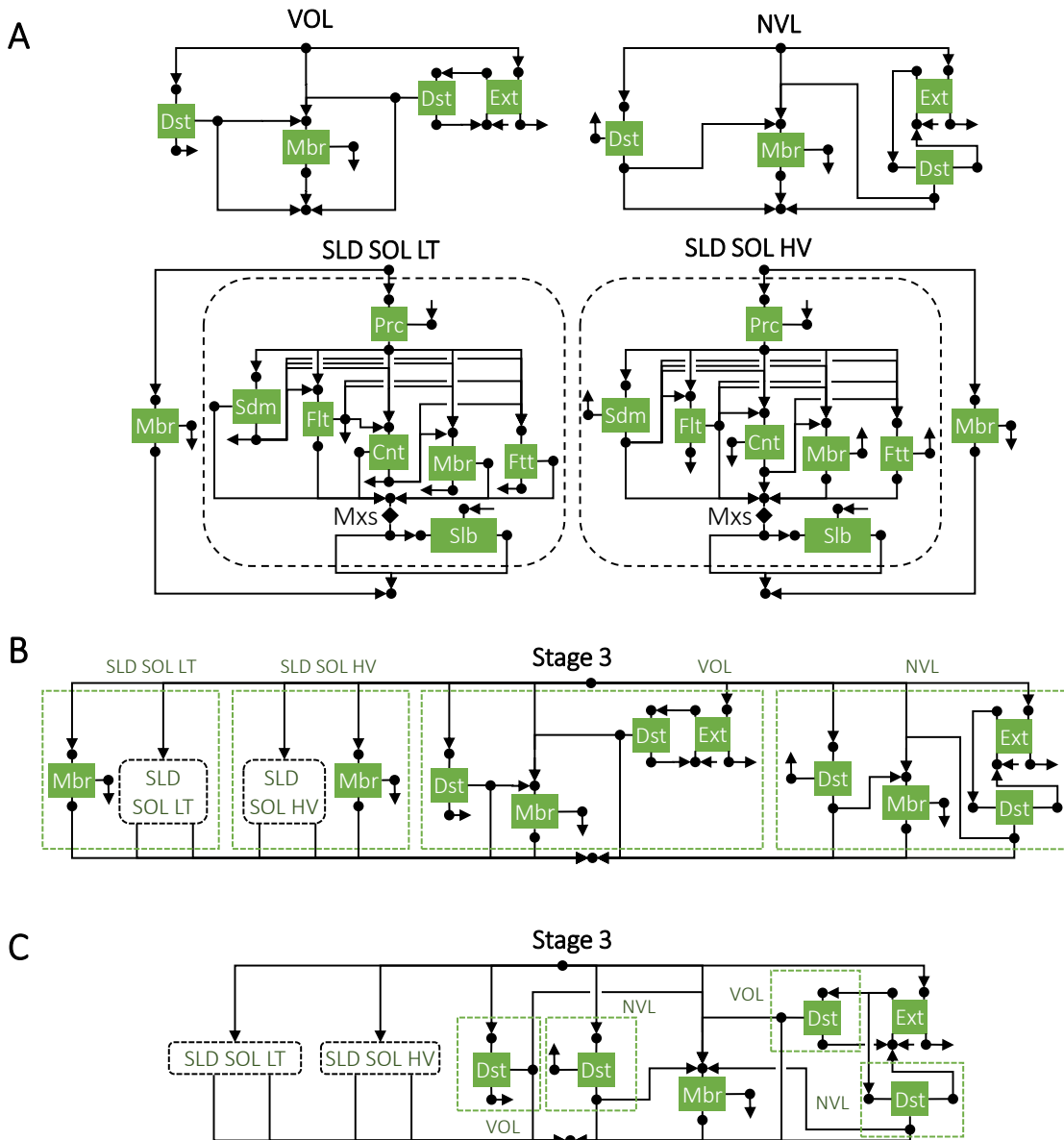


Figure 3.6. Module and stage superstructures for Stage 3. **(A)** Module-superstructures; **(B)** combined stage-superstructure with all the module-superstructures aligned in parallel; **(C)** simplified stage-superstructure.

3.4.4 Stage 4 superstructure

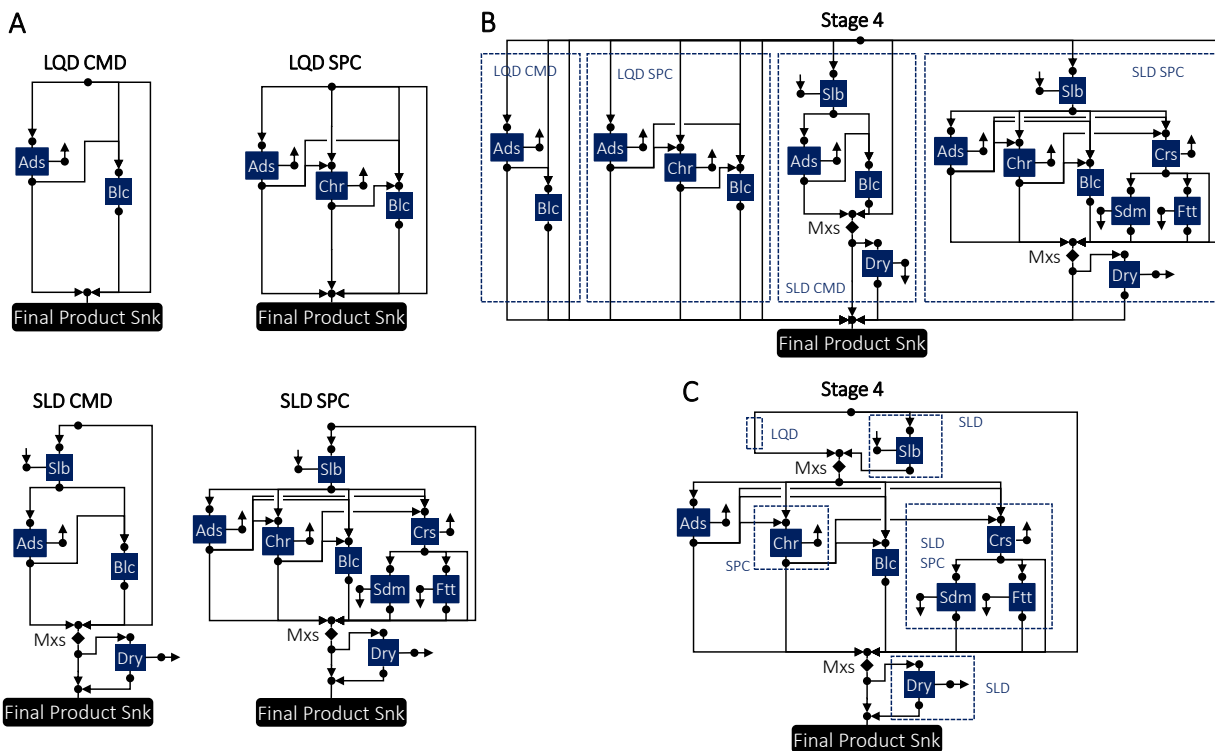


Figure 3.7. Module- and stage-superstructures for Stage 4. **(A)** Module-superstructures; **(B)** combined stage-superstructure with all the module-superstructures aligned in parallel; **(C)** simplified stage-superstructure. The Blc unit has only one inlet port and one outlet port, because we assume that the stream will become color-free once it passes through a Blc unit.

The module-superstructures are shown in **Figure 3.7A**; the combined stage-superstructure obtained after applying Step 1 in **Section 3.3.3** is shown in **Figure 3.7B**; the final simplified superstructure for Stage 4 is shown in **Figure 3.7C**. Solubilization is required for SLD products because adsorption, chromatography, bleaching and crystallization require solid-free feed streams. It is assumed that the type of crystallization (evaporation/cooling) is pre-selected. For simplicity, we always assume that evaporative crystallization is selected in the discussions hereafter, and thus crystallization units have two outlet ports (see **Figure 3.7**). However, if cooling is selected, then the crystallization units should have only one outlet port. A mixer-splitter is used in **Figure 3.7C** to

concisely represent the different processes for SLD products (using solubilization) and LQD products (by-passing solubilization).

3.5 General superstructure

To generate the general superstructure, we combine all the stage-superstructures (**Figure 3.4B**, **Figure 3.5B**, **Figure 3.6C** and **Figure 3.7C**), by connecting them through the stage-connecting ports, and then include additional cross-stage connections. The final superstructure is shown in **Figure 3.8**. The specific cross-stage connections are as follows.

First, since extraction and distillation can handle solid particles (up to certain concentrations), the stream exiting Stage 1 can be directly fed to the extraction or distillation unit in Stage 3 if the product is VOL (thus bypassing Stage 2). Also, a by-pass stream is added to Stage 3.

Second, after processed in the solubilization and phase separation units in the NSL HV module in Stage 2 (see **Figure 3.5**), the NSL SLD product is dissolved and thus precipitation can be directly used afterwards in Stage 3. Accordingly, the SLD SOL HV module in **Figure 3.6** is also applicable to SLD NSL HV products that are solubilized in Stage 2, and thus we rename the module as “SLD HV” in **Figure 3.8**. This module only admits upstream connections from the SOL module and the left part of the NSL HV module (without differential digestion treatment) in Stage 2 (see **Figure 3.8**). Similarly, the SLD SOL LT module in Stage 3 is only connected to the SOL module in Stage 2. The output stream from the left part of the NSL HV module (without differential digestion treatment) can also bypass Stage 3 and the solubilization unit in Stage 4 entirely, because it has been pre-treated for the refinement in Stage 4.

Third, the stream without the solubilization treatment at the end of the SLD SOL LT and SLD HV modules in Stage 3 can be directly fed to the mixer-splitter before drying in Stage 4. The streams

after solubilization at the exit of these modules can bypass the solubilization unit at the beginning of Stage 4.

Note that the general superstructure can also be modified to account for customized modeling. Examples of such customization include: (1) inclusion of more technologies in each stage, such as multiple types of cell disruption in Stage 1; (2) inclusion of different connectivity rules to allow more or fewer connections in the module-superstructures; and (3) inclusion of more or fewer cross-stage connections.

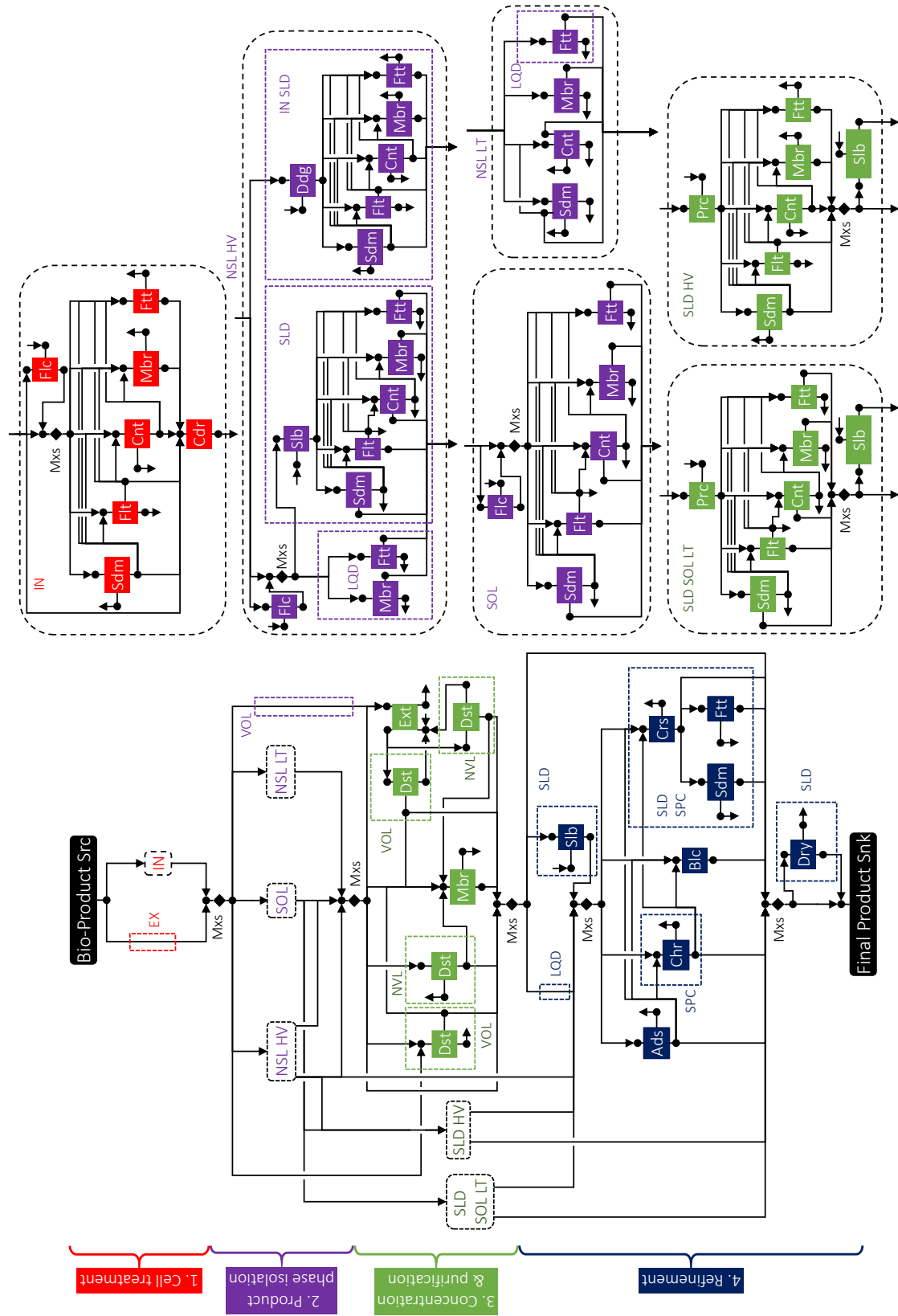


Figure 3.8. General bio-separation superstructure.

3.6 Reduction method

The general superstructure discussed in **Section 3.5** embeds millions of bio-separation configurations. Modeling and solving for such a large superstructure is both impractical and unnecessary when product-specific information is available. For example, if the product is known to be EX SOL, then all the units and modules only applicable to IN or NSL products in the general superstructure can be removed. We propose the following steps to generate a simpler superstructure (“reduced superstructure”) which yet includes all relevant configurations.

Step 1: Remove all the units and modules that are not applicable to the product attributes. For example, if the product is EX, then the IN module in Stage 1 should be removed.

Step 2: Remove all the units and modules that are not available or suitable. For example, if no extraction solvent is known for the product in question, then extraction units should be removed.

Step 3: Remove additional units, modules, or streams based on case-specific considerations. For example, if the product is EX SOL with an extremely low titer, and based on experiments we know that sedimentation is too slow and expensive even after flocculation, then sedimentation can be removed. Similarly, if we know that the concentration of solid particles (e.g. cells) after Stage 1 is greater than the handling limit of distillation, then the bypass stream from Stage 1 to the distillation unit in Stage 3 should be removed.

Step 4: Remove irrelevant units and streams based on final product stream specifications, including colorlessness, dryness and crystalline form. Specifically, if the final product stream is required to be colorless, then all the streams bypassing the bleaching unit should be removed. On the other hand, if colorlessness is not required, then the bleaching unit in Stage 4 should be removed. If the final product stream is required to be completely dry, or be in its crystalline form, then all the streams bypassing the drying unit or the crystallization unit should be removed. However, if no such

requirements are specified, they should not be removed because, unlike the bleaching units, drying and crystallization can be used to concentrate the product-containing stream and remove impurities.

Step 5: Choose between filtration and membrane (if filtration is not removed from the previous steps), based on the characteristics (e.g. molecular weight or particle size) of the components to be separated. The specific type of membrane (e.g. micro-, ultra- and nano-filtration) can also be pre-determined accordingly.

Step 6: After the above five steps, we apply the following two rules iteratively to remove all the remaining irrelevant connections: (i) if a unit is removed, then all of its ports and the streams connected to its ports should be removed; (ii) if a port is not connected by any streams, then the unit to which the port belongs to should be removed. For example, if crystallization in Stage 4 is removed in the previous steps, then based on Rule (i), all of its ports and the connected streams should be removed. Further, the downstream membrane and sedimentation should be removed based on Rule (ii).

To illustrate the proposed reduction method, we consider a product, which is EX NSL LT NVL LQD CMD. The final product stream is required to be colorless. Membrane is selected instead of filtration in Stage 2. After applying steps 1, 4, 5 and 6, the reduced superstructure is shown in **Figure 3.9**.

The superstructure thus obtained still contains redundant mixer-splitters. A final “polishing” step is performed to eliminate them, leading to the final reduced superstructure shown in **Figure 3.10**, where all units are numbered accordingly.

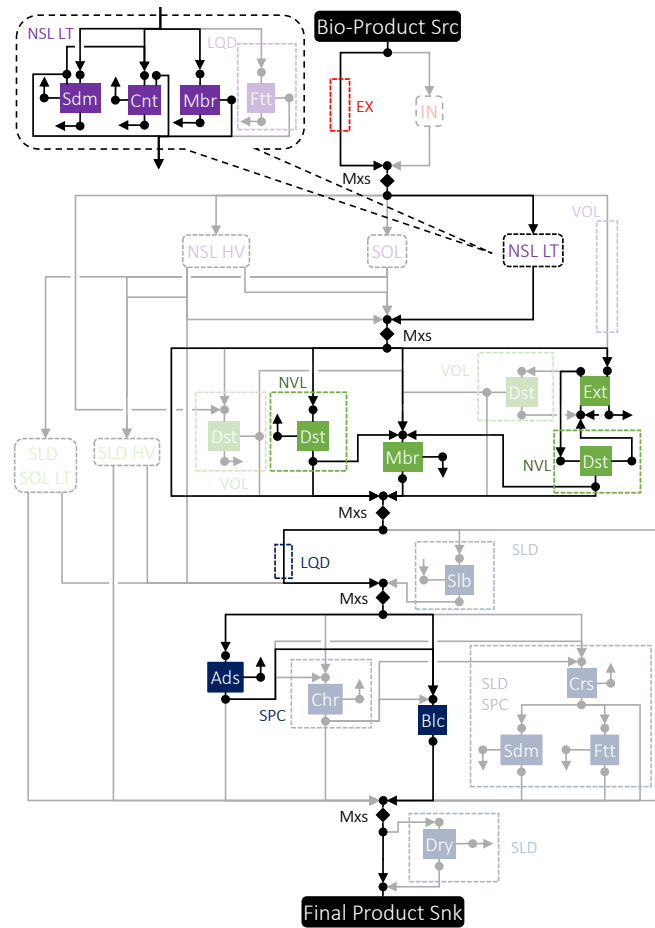


Figure 3.9. Example superstructure reduction for EX NSL LT NVL LQD CMD product, and the final product stream is required to be colorless. The dimmed parts are removed after applying steps 1, 4, 5 and 6.

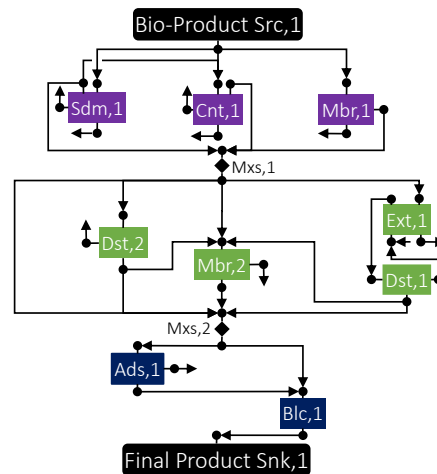


Figure 3.10. Final reduced superstructure for the example in Figure 10.

3.7 Modeling approach

We adopt the modeling approach proposed in **Chapter 2**. Specifically, we first define the basic sets (unit types, unit numbers, port types, port numbers, and components) and three key multidimensional sets (units, ports, and streams), as well as multiple subsets to facilitate modeling. Then we employ standard short-cut unit models using port variables and unit internal variables. The port and stream models are simplified in the current work because stream conditioning is incorporated into the unit models. Moreover, we adopt the single-stream assumption [22] for outlet ports, which allows the activation of at most one outgoing stream; in the optimal solution at most one stream will be selected. Also note that we use binary attributes for superstructure generation but adopt continuous property values when necessary; e.g., although a product is categorized as VOL or NVL, its relative volatility is considered. Furthermore, for components that cannot be strictly classified, we can include more modules in the superstructure (e.g., for both VOL and NVL products).

Implementing the above general modeling approach, we formulate the general bio-separation optimization model as a MINLP problem, as shown in **Appendix 2**. The model includes unit models for all the technologies shown in the general superstructure in **Figure 3.8**. Due to the modular nature of our modeling approach, modeling for the reduced superstructure is performed simply by (1) removing irrelevant unit models, and (2) adding case-specific constraints, including logic constraints inferred by the final product stream specifications, and other case-specific considerations (e.g., no more than two units can be activated in series). Also, since this framework is used for preliminary assessment, we neither consider multiple units of the same type used in parallel due to unit size restrictions, nor do we set a minimum size for an activated unit. Instead, we calculate the equivalent total size required to complete its separation task.

Table 3.1. Number of variables and equations of each unit model. “c” represents the number of components in the corresponding unit.

Unit	Number of var. ^{a,c}	Number of eqn. ^{b,c}	Unit	Number of var. ^{a,c}	Number of eqn. ^{b,c}
Adsorption/chromatography	$3c+1$	$2c+1$	Extraction	$7c-2$	$6c-4$
Bleaching	$2c+1$	$c+1$	Membrane/filtration	$3c+7$	$2c+6$
Cell disruption	$2c+1$	$c+1$	Flocculation	$3c+1$	$2c+1$
Crystallization (cooling)	$2c+2$	$c+2$	Flotation ^f	$3c+c'+1$	$2c+c'-1$
Crystallization (evaporation) ^d	$3c+c'+2$	$c+c'$	Precipitation	$3c+1$	$2c+1$
Differential digestion	$3c+1$	$2c+1$	Sedimentation/centrifugation ^f	$4c+c'+1$	$3c+c'-2$
Distillation ^e	$4c+c'+9$	$3c+c'+9$	Solubilization	$3c+1$	$2c+1$
Drying ^d	$3c+c'+2$	$2c+2$			

(a) Var.=variables;

(b) Eqn.=equations;

(c) The counts exclude variables and equations associated with costs;

(d) c': number of components more volatile than the product;

(e) c': number of components less volatile than the light key, and more volatile than the heavy key;

(f) c': number of insoluble components in the unit.

The unit models are formulated based on multiple sources: membrane and filtration [127], [129]; sedimentation and centrifugation [126]–[128]; extraction [130], [131]; distillation [3], [132]–[134]; other units [125], [127], [130]. A summary of these models is given in **Table 3.1**, while the specific models are presented in **Appendix 2**.

3.8 Conclusions

In this work, we developed a superstructure-based framework for bio-separation network synthesis. We first identified four stages: cell treatment, product phase isolation, concentration and purification, and refinement. Then, for all four stages, we systematically developed stage-superstructures, which, together with cross-stage connections, leads to the generation of the general bio-separation superstructure. We further developed a superstructure reduction method to solve case-specific problems, based on product attributes, technology availability, case-specific considerations, and final product stream specifications. An MINLP optimization model, including short-cut models for all technologies considered in the framework, was formulated. The proposed

framework can be used to quickly evaluate separation processes for different bioreactor effluent streams and different products, thereby aiding the identification of products that can be produced effectively using bio-conversion.

Chapter 4

4 Separation process analysis for extracellular and intracellular chemicals⁴

4.1 Introduction

In **Chapter 3**, we have presented a systematic approach to generating and modeling bio-separation superstructures for different *classes* of products defined in terms of a set of properties including product localization, solubility, density, volatility, physical state and intended use. In this chapter, we use the approach to analyze two major categories of products: extracellular and intracellular. Note that only the analysis for extracellular products is shown in this chapter; the same analysis for intracellular products can be found in [135].

When a product of interest is produced by microbial cells, it is then localized either inside the cells or released to the extracellular phase. In fact, most products are initially produced intracellularly, but some products are localized extracellularly to the aqueous medium through passive diffusion or active transport [136]. Previous work on economic assessment for the separation of extracellular chemicals has been mainly restricted to specific examples such as hyaluronic acid [137]–[140], limonene [141]–[147], xanthan gum [148]–[150], butanediol [151]–[156], lactic acid [157]–[159] and penicillin V [160], [161]. Also, assessment studies have been performed on individual separation technologies [8], [162]–[164]. However, technology selection is nontrivial because many competing alternatives are often available as discussed in **Chapter 3**. Furthermore, traditional

⁴This chapter includes content from: W. Wu, K. Yenkie, and C. T. Maravelias, “Synthesis and analysis of separation processes for extracellular chemicals generated from microbial conversions,” *Biotechnol. Bioeng.*, submitted, 2018. Wenzhao Wu completed the development of Case 1 in Section 4.3.1, the sensitivity analyses of Cases 2 and 3 in Sections 4.3.2 and 4.3.3, and writing of the manuscript; Kirti Yenkie generated the base cases and heat maps in Cases 2 and 3 in Sections 4.3.2 and 4.3.3.

analyses have usually focused on sensitivity analyses where the technologies in the separation network are fixed and only one parameter is varied at a time to analyze its influence on the process economics [165]–[167]. To these ends, we synthesize and analyze separation processes for extracellular products using superstructure optimization, aiming to convert a dilute effluent (containing product, microbial cells, water, and a small amount of co-product impurities) from a bioreactor to a high-purity product stream. Note that we only consider liquid or solid products entering the separation networks.

In **Section 4.2**, we briefly discuss our methods, including the stage-wise separation scheme, superstructure generation and modeling, and an analysis framework. In **Section 4.3**, we present the results. Specifically, we first categorize extracellular products into three different categories based on their physical properties: (1) insoluble light (with density lower than that of water), (2) insoluble heavy, and (3) soluble. Such categorization is necessary because, e.g., separation of extracellular insoluble light products tends to be easier (via simple decantation, filtration, etc.) than that of extracellular soluble ones (via distillation, membrane, precipitation, etc.), especially when the product titer is low. Second, in each category, we develop a base case, which is solved to generate the cost-minimal process with the optimal technology selection, and we identify key cost drivers. Third, we analyze the influence of key parameters (such as product titer and technology performances) on optimal technology selection and cost. Finally, we extend the discussion to account for other classes of products in the category.

4.2 Methods

The recovery of an extracellular (EX) product is divided into three stages [168], [169]: (I) product phase isolation (including pretreatment as well as cell removal and phase isolation), (II) product concentration, and (III) product purification and refinement, as shown in **Figure 4.1A**. Note that in

Chapter 3, we adopt a four-stage scheme. However, only three stages are kept here because Stage I in **Figure 3.8** of **Chapter 3** exists only for intracellular products; for EX products it is bypassed. Therefore, we redefine the stage numbers here. Each stage has multiple technologies available for the same task, as shown in **Table 4.1**. We will use the abbreviations when referring to the specific technologies and product properties hereafter.

4.2.1 Stage-wise separation scheme

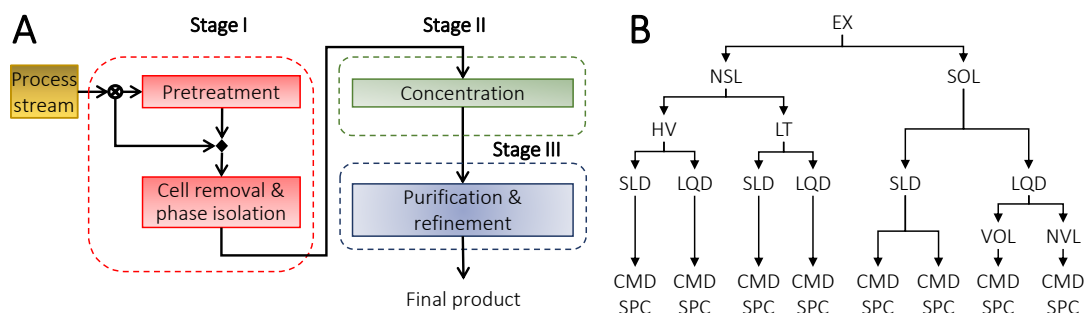


Figure 4.1. Stage-wise separation scheme and its simplification for extracellular products. (A) Representation of the three-stage separation scheme for extracellular products; (B) simplification of the separation scheme for superstructure generation based on product properties (solubility, density, physical state, volatility and intended use). Abbreviations: solubility in water [insoluble (NSL) or soluble (SOL)], density with respect to water [heavy (HV) or light (LT)], physical state [solid (SLD) or liquid (LQD)], relative volatility with respect to water [volatile (VOL) and non-volatile (NVL)], and intended use [commodity (CMD) or specialty (SPC)].

Table 4.1. Technology options available for performing the tasks listed in the three separation stages. Abbreviations of the technologies are shown in parentheses.

Tasks	Technologies
Pretreatment	Flocculation (Flc)
Cell removal & phase isolation	Sedimentation (Sdm), filtration (Ftt), centrifugation (Cnt), flotation (Flt), membranes (Mbr- MF [microfiltration], UF [ultrafiltration], and RO [reverse osmosis]), Differential digestion (Ddg), solubilization (Slb)
Concentration	Extraction (Ext), aqueous two phase extraction (Atpe), evaporation (Evp), precipitation (Prc), sedimentation (Sdm), filtration (Ftt), centrifugation (Cnt), membranes (MF, UF, NF (nanofiltration), RO), distillation (Dst)
Purification & refinement	Adsorption (Ads), chromatography (Chr), crystallization (Crs), pervaporation (Pvp), membranes (Mbr-MF, UF, NF, RO), Drying (Dry), bleaching (Blc)

4.2.2 Superstructure generation and modeling

The potential separation stages and the relevant technology options in the separation scheme (see **Figure 4.1A** and **Table 4.1**) can be narrowed down based on other distinguishing properties of an EX product such as the product's solubility in water [insoluble (NSL) or soluble (SOL)], density with respect to water [heavy (HV) or light (LT)], physical state [solid (SLD) or liquid (LQD)], relative volatility with respect to water [volatile (VOL) and non-volatile (NVL)], and intended use [commodity (CMD) or specialty (SPC)], as shown in **Figure 4.1B**. Each combination of these properties corresponds to a specific *class* of products, e.g., 2,3-butanediol is in the EX SOL NVL LQD CMD class. Thus, for the class of products chosen in the base case in each product category, we generate a superstructure based on the approach in **Chapter 3** and previous work on separation network synthesis using schemes and superstructures [169].

Next, we formulate an optimization model (mixed-integer non-linear programming - MINLP) for the superstructure, with binary variables denoting the activation/deactivation of technologies present in the superstructure. The model involves constraints describing the separation technologies, stream flows, input specifications and product purity requirements. The objective is to minimize the overall process cost, including annualized capital cost and operating cost (input feedstock, consumables, labor, utility, etc.). We assume base case values for process and economic parameters such as product titer, technology efficiencies and material cost. The specific model equations can be found in **Appendix 2**, and the parameter values used can be found in **Appendix 3**. The model is developed in GAMS 25.1.1 environment and solved using BARON [170], a global optimization solver. For discussions about specific separation technologies (including common parameter values), the readers are pointed to [135], [169].

4.2.3 Analysis framework

In each product category, after solving the base case to determine the cost-minimal separation network and the key cost drivers, we further analyze how variations in the aforementioned model parameters affect the cost and technology selection, with the following steps:

Step 1. Vary one or a combination of key cost contributing parameters and solve an optimization problem for each combination, to determine the threshold values where a shift in optimal technology selection happens.

Step 2. Extend the analysis to other product classes in the category based on (1) the results for the base case, if the same technology options are suitable for the other classes, or (2) individual technology considerations, if new technologies should be included.

4.3 Results and discussions

4.3.1 EX NSL LT products

For the EX NSL LT category, we choose an EX NSL LT LQD CMD product as a representative base case. Note that EX NSL LT products float to the top and are thus naturally separated from the microbial cells. The key parameters used are shown in **Appendix 3.1**.

4.3.1.1 Superstructure and optimal solution

By simplifying the separation scheme of an EX NSL LT product, we obtain the superstructure for EX NSL LT LQD CMD product, as shown in **Figure 4.2**. Note that unit ports are omitted in **Figure 4.2** and hereafter for presentation simplicity. In Stage I – phase isolation, Sdm1 and Cnt1 separate the product as a top phase (isolated from the cells at the bottom), while removing water at the same time. Ftt functions to only remove the cells. In Stage II – concentration, Sdm2, Cnt2 and Mbr can

separate the product from water, which can be bypassed if enough concentration has been achieved in Stage I. In Stage III – refinement, Ads further purifies the product, which can also be bypassed if the optimization model decides that the final product specifications have been satisfied.

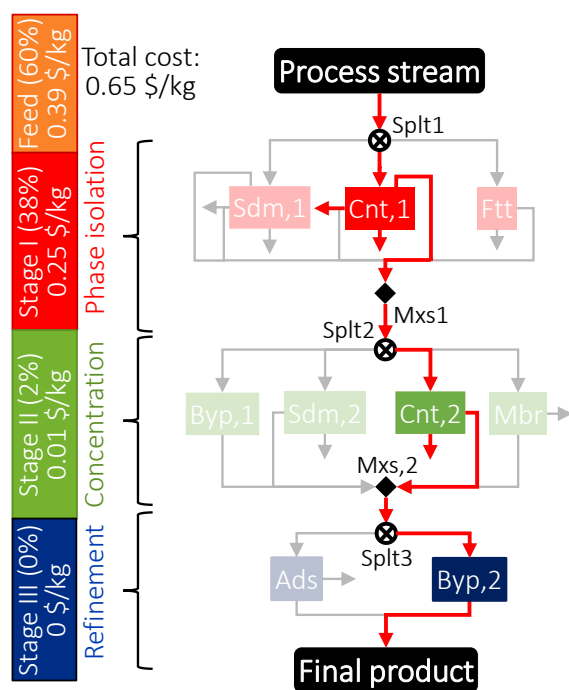


Figure 4.2. Superstructure (including all the units and streams) and optimal solution (the highlighted parts) for EX NSL LT LQD CMD product. The active streams are shown by bold red lines and selected technologies are highlighted in different colors corresponding to each stage: red for stage I, green for stage II, and blue for stage III. Cost distribution is shown by the numbers on the left bar. Byp=bypass.

After solving the superstructure optimization model, we obtain the cost-minimal separation network (Cnt1 in Stage I followed by Cnt2 in Stage II), as represented by the highlighted parts in **Figure 4.2**, as well as the corresponding cost distribution. The total minimum cost is 0.65 \$/kg, with the feed accounting for 60%, Stage I separation 38%, and Stages II separation 2%. Cnt2 is also selected because Cnt1 alone is not able to concentrate the product stream from 5 g/L (~0.5 wt%) to the required final purity of 95 wt% due to concentrating factor limitations (see Cnt parameters in **Appendix 3.4**).

4.3.1.2 Analysis for EX NSL LT LQD CMD product

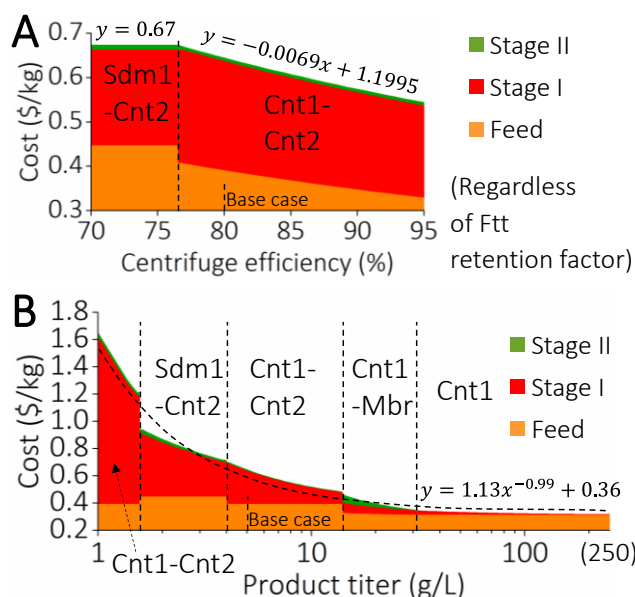


Figure 4.3. Analysis on technology selection and cost for EX NSL LT LQD CMD product. **(A)** Analysis with varying Cnt1 efficiency and Ftt retention factor; **(B)** analysis with varying product titer. The optimal technologies selected are labeled in the corresponding regions. The fitted functions are shown, where y represents the cost, and x represents the Cnt1 efficiency and product titer in (A) and (B), respectively. The based cases are marked with short dashed lines with “Base case” labels next to them.

Performance of phase isolation technologies

Since the major cost component is Stage I in the optimal network, we vary parameters related to the phase isolation technologies (Sdm1, Cnt1 and Ftt). The performance for Sdm1 and Cnt1 is defined in terms of “efficiency” of the separation of product from the aqueous phase. For Ftt, it is defined as the retention factor of cells on the retentate side of the filter. We vary the Cnt1 efficiency and Ftt retention factor between 70% and 95% simultaneously [171]–[173] (see **Appendix 3.1**) and run the optimization model to obtain the optimal separation network and cost for each combination of these two parameters. Note that for the efficiency of Sdm1, there is limited scope for performance enhancement [171], therefore, we fix it to 70%.

As a result, regardless of the Ftt retention factor variation, Ftt is not selected in the optimal solution (see **Figure 4.3A**). This is because Ftt can only separate the cells from the product and water, while Cnt1 or Sdm1 functions to remove water at the same time. The optimal technologies selected are noted in the corresponding regions in **Figure 4.3A**. The readers can identify the optimal separation network in **Figure 4.2** accordingly. The same notation is used in the subsequent figures. It can be seen that when the Cnt1 efficiency is below 76.5%, Sdm1 is selected in Stage I, and Cnt2 is selected in Stage II due to lower cost; otherwise, Cnt1 is selected. Since the efficiency of Cnt2 is still fixed to 80%, Cnt2 is selected as the optimal technology in Stage II.

Product titer

Sdm1 and Cnt1 equipment sizes are the major phase isolation cost drivers, and they depend on the product titer in the feed entering the separation network (affecting the total input stream flow rate). Product titer depends on the microbial strain, substrate utilization, microbial-conversion pathways, and bioreactor design. It has a potential to be altered by metabolic engineering tools [78], [82], [174]–[176]. We vary the product titer from 1 to 250 g/L and obtain the costs and the corresponding optimal technology selection in **Figure 4.3B**.

It can be seen that at high titer (≥ 32 g/L), Cnt1 alone is able to achieve the required product purity. When the titer is 14–32 g/L, another concentration technology is required in Stage II, and Mbr is preferred to Sdm2 and Cnt2 because the product loss is lower. When the titer is 4–14 g/L, Cnt2 becomes a better option than Mbr in Stage II because the low titer requires large Mbr equipment size and more costs associated with Mbr replacements. When the titer is 1.57–4 g/L, Sdm1 is a cheaper option than Cnt1 in Stage I because the major cost of Sdm is equipment cost, which is scaled with the equipment size based on the power scaling rule; however, apart from equipment cost, Cnt also has electricity cost, which is scaled linearly with the equipment size. Therefore, when

the titer is low (leading to large equipment size), Cnt1 electricity cost is high, and thus Sdm1 is a cheaper option. Finally, when the titer is < 1.57 g/L, Sdm1 in Stage I is limited by its maximum concentrating factor and thus is not able to concentrate the product enough for Stage II to reach the required purity. Therefore, Cnt1, with a higher concentrating factor, is selected.

4.3.1.3 Extension to other classes of EX NSL LT products

The other properties determining the superstructure for EX NSL LT products are physical state (LQD/SLD) and intended use (CMD/SPC) (see **Figure 4.1B**).

For a SLD product, Ftt in Stage I cannot separate the product from the cells, and thus it should be removed from the superstructure. However, since Ftt is actually not selected in the optimal network, our base case is still considered representative.

For a SPC product, separation technologies in Stage III need to be more stringent to meet the purity requirements. Thus technologies such as Chr and Blc (to remove pigments) can be included in the superstructure, and a similar analysis can be performed to identify the impact of variation in technology parameters in Stage III. However, the Stage III parameters are not selected for further analysis because for high-value specialty chemicals, quality is the major concern and cost minimization becomes secondary [177].

4.3.2 EX NSL HV products

For this category, we choose EX NSL HV SLD CMD product as a representative base case. The key parameters used are the same with those in the EX NSL LT base case. Other parameters can be found in the **Appendix 3.4**.

4.3.2.1 Superstructure and optimal solution

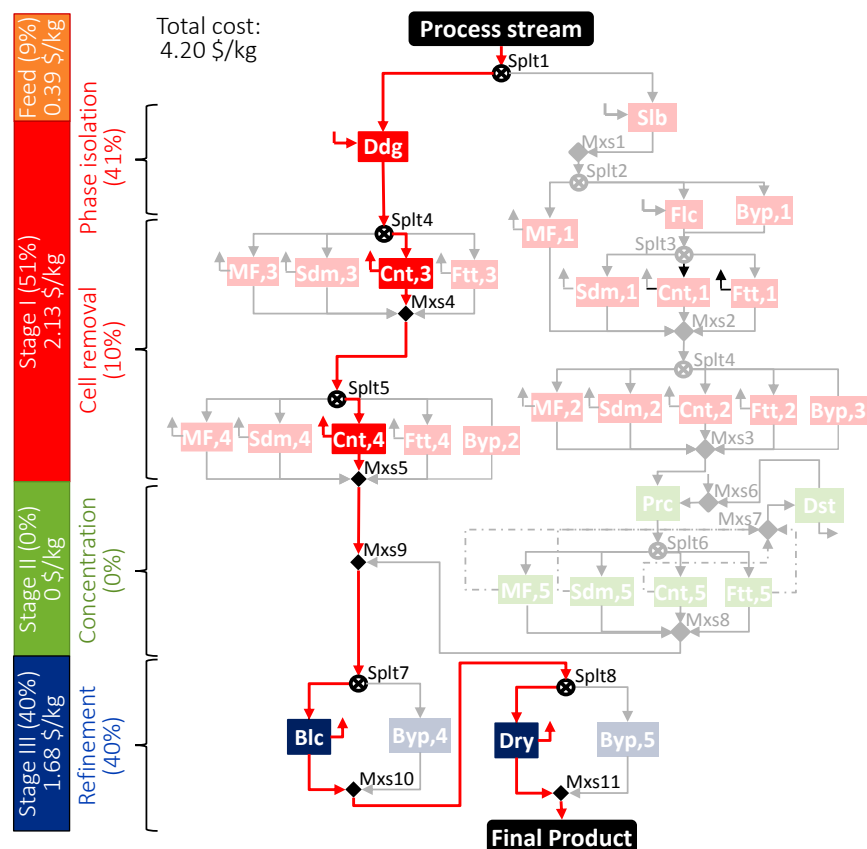


Figure 4.4. Superstructure (including all the units and streams) and optimal solution (the highlighted parts) for EX NSL HV SLD CMD product. The active streams are shown by bold red lines and selected technologies are highlighted in different colors corresponding to each stage: red for stage I, green for stage II, and blue for stage III. Cost distribution is shown by the numbers on the left bar.

By simplifying the separation scheme of an EX NSL HV product, we obtain the superstructure for EX NSL HV SLD CMD product, as shown in **Figure 4.4**. Since the product is EX, Stage I consists of phase isolation and cell removal. Phase isolation i.e., the separation of product-containing phase from other components in the stream, can be achieved using Slb or Ddg. Slb is used to dissolve the product in a suitable solvent to separate it from cells and other solid impurities. Ddg is used to dissolve the non-product containing materials (NPCM). Cell removal technologies include Sdm, Cnt, Ftt, and Mbr. Multiple technologies may be required in series depending on the initial product and

cell concentration in the inlet stream. Flc, as an optional pretreatment technology, can enhance the separation efficiency of subsequent tasks by enabling the formation of flocs of cells, which are then easier to isolate from the aqueous phase.

Stage II is required if the process stream undergoes Slb in Stage I. Stage II includes Prc followed by the phase separation technologies which are similar to the ones used for cell removal. If Ddg is selected in Stage I, then Stage II is not required. Stage III involves Blc and Dry options for final product purification.

After solving the superstructure optimization model, we obtain the cost-minimal separation network, as represented by the highlighted parts in **Figure 4.4**, as well as the corresponding cost distribution. The technologies selected in Stage I are Ddg for phase isolation and Cnt3 and Cnt4 for cell removal. The final product refinement involves Blc to remove undesired color imparting impurities and Dry to retrieve product in the solid form. The overall process cost is 4.20 \$/kg, where the separation cost contribution is \$3.81/kg (91%). Stage I is the highest cost contributor (51%).

4.3.2.2 Analysis for EX NSL HV SLD CMD product

Performance of phase isolation and cell removal technologies

Since Ddg for phase isolation in Stage I is the major cost component in the optimal configuration (41% of the overall cost), and its competing technology is Slb, we vary the required amount of Ddg agent and Slb solvent, as shown in **Figure 4.5A**. Ddg is the preferred option even if the digestion agent is required in higher amounts as compared to Slb solvent. This is because Slb selection adds additional cost in Stage II, where Prc followed by phase separation is required.

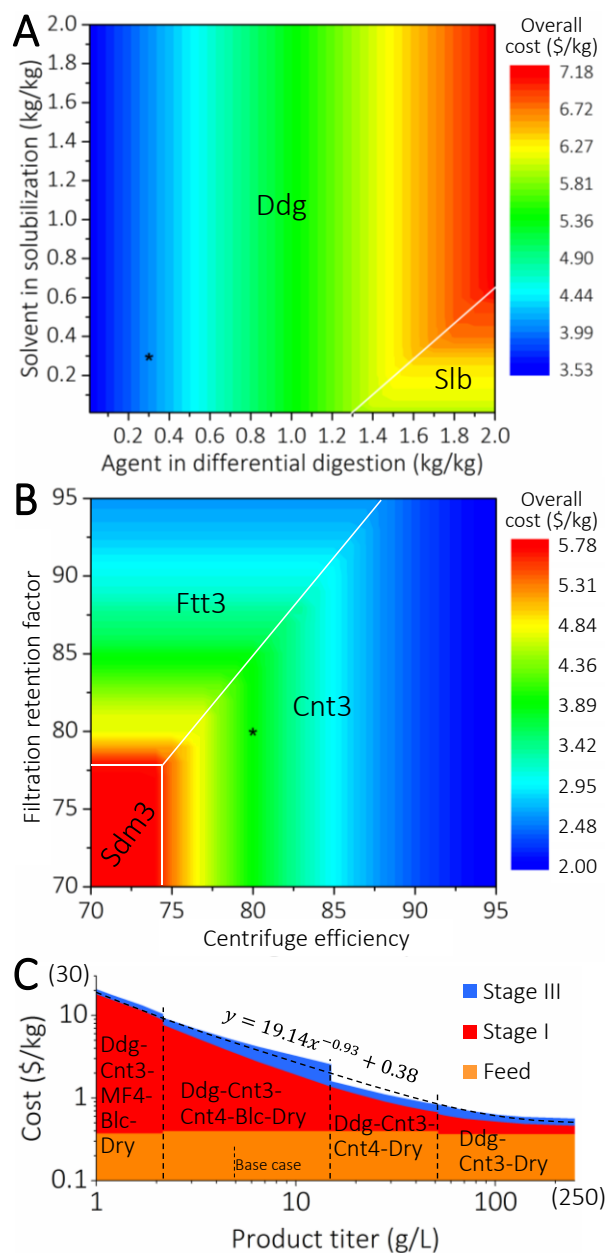


Figure 4.5. Analysis on technology selection and cost for EX NSL HV SLD CMD product. **(A)** Analysis with varying required amount of Ddg agent and Slb solvent; **(B)** analysis with varying Cnt3 efficiency and Ftt3 retention factor; **(C)** analysis with varying product titer. The fitted cost-titer functions are shown in (C), where y represents the cost, and x represents the product titer. The based cases are marked with asterisks in (A) and (B) and a short dashed line with “Base case” label next to it in (C).

The second major cost component is Blc (28% of the overall cost). However, there is limited room for performance improvements since the Blc efficiency is already assumed to be 99%.

The next major cost component is Cnt3 (8.4% of the overall cost) for cell removal, and its performance affects product loss and Stage III cost. Therefore, we also vary the Cnt3 efficiency and Ftt3 retention factor between 70% and 95% simultaneously, while fixing Sdm3 efficiency to 70%, for the same reason discussed in **Section 4.3.1.2**. The analysis is shown in **Figure 4.5B**. Cnt3 is the preferred option in most cases because its capital cost is lower than that of Sdm3, and Ftt3 membrane replacement cost is high.

Product titer

We vary the product titer from 1 to 250 g/L. The costs and the corresponding optimal technology selection are obtained and presented in **Figure 4.5C**. We observe that when the titer is greater than 52 g/L, Ddg-Cnt3-Dry is the optimal selection; from 15 to 52 g/L, further concentration by Cnt4 is needed to achieve the final purity requirement; from 2.2 to 15 g/L, the amount of color imparting impurities is substantial compared with the amount of product present, and thus Blc is required; when the titer is less than 2.2 g/L, MF4 replaces Cnt4 due to limitations of Cnt4 concentrating factor.

4.3.2.3 Extension to other classes of EX NSL HV products

The other properties determining the superstructure for EX NSL HV products are physical state (LQD/SLD) and intended use (CMD/SPC) (see **Figure 4.1B**).

For a LQD product, instead of using Ddg or Slb (see **Figure 4.4**), Ftt or Mbr (depending on the size of solid impurities) can be used directly to separate the product from cells, followed by concentration using Sdm, Cnt or Mbr. Also, Dry in Stage III is only applicable to SLD products and thus should be removed. Therefore, for LQD product, the separation cost will be lower. We modify the base case superstructure to account for LQD product, and the optimal technology selections are

Flc-Ftt-MF-Blc, and the cost is decreased from 4.20 \$/kg (for SLD) to 3.52 \$/kg (for LQD), as shown in the **Appendix 3.2**. For a SPC product, the same argument in **Section 4.3.1.3** holds.

4.3.3 EX SOL products

For this category, we choose EX SOL LQD NVL CMD product as the base case. The key parameters used are the same with those in the EX NSL LT base case. Other parameters can be found in the **Appendix 3.4**.

4.3.3.1 Superstructure and optimal solution

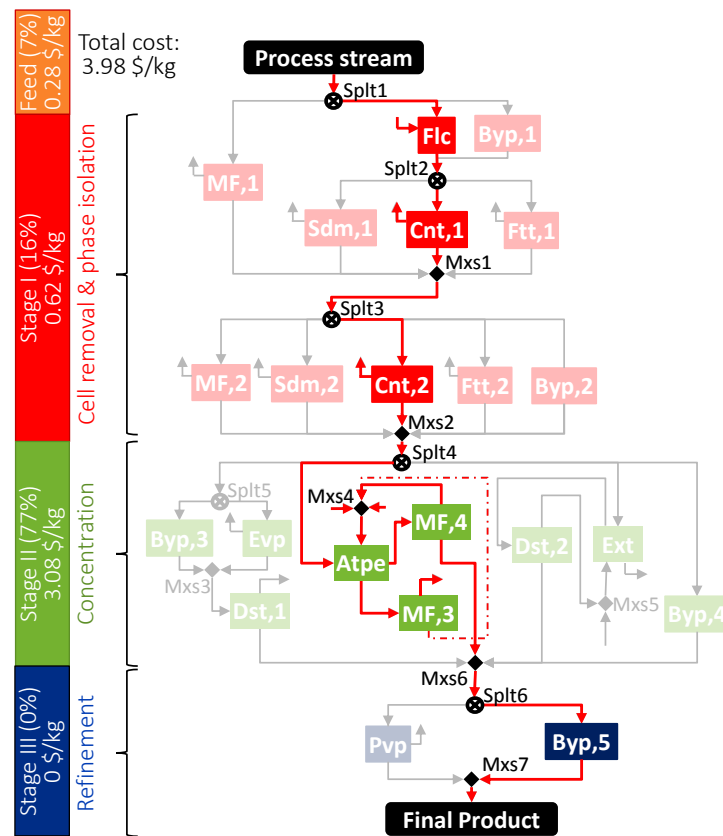


Figure 4.6. Superstructure (including all the units and streams) and optimal solution (the highlighted parts) for EX SOL LQD NVL CMD product. The active streams are shown by bold red lines and selected technologies are highlighted in different colors corresponding to each stage: red for stage I, green for stage II, and blue for stage III. Cost distribution is shown by the numbers on the left bar.

By simplifying the separation scheme of an EX SOL product, we obtain the superstructure for EX SOL LQD NVL CMD product, as shown in **Figure 4.6**. In Stage I, cells are first removed, and thus the aqueous phase containing the product is isolated. Then, in Stage II, Dst, Atpe and Ext are considered as concentrating technologies. Finally, in Stage III, Pvp can be used to remove small amount of remaining impurities, if necessary.

After solving the superstructure optimization model, we obtain the cost-minimal separation network, as represented by the highlighted parts in **Figure 4.6**, as well as the corresponding cost distribution. The technologies selected in stage I are Flc, Cnt1, and Cnt2. Atpe, followed by MF3 and MF4, is selected for concentration in Stage II. Stage III is bypassed. The total minimum cost is 3.98 \$/kg, with the feed accounting for 7%, Stage I 16% and Stage II 77%. Stage II is the major cost component because the separation for SOL products requires concentration of the product present in water-rich phase.

4.3.3.2 Alternative concentration options in Stage II

Atpe is selected in the optimal network for the base case. However, we also analyze the separation networks when Ext or Dst1 is selected in Stage II, thus accounting for cases where effective Atpe polymer and/or salt for the product may not exist.

When Ext (as well as its auxiliary Dst 2) is selected (by setting the binary variable for Ext to 1 to ensure selection), the technologies selected (see **Figure 4.7A**) in Stage I include Flc, Cnt1 and Cnt2. Stage III is bypassed. The overall process cost is 4.06 \$/kg, and the separation cost contribution is 3.77 \$/kg (94%). Stage II is still the major cost contributor (80%).

When Dst1 is selected, the technologies selected (see **Figure 4.7B**) in Stage I include Flc, Cnt1 and Cnt2. Stage III is bypassed. The overall process cost is 8.19 \$/kg, and the separation cost

contribution is 7.93 \$/kg (97%). Stage II is still the major cost contributor (90%). Direct Dst is costly because a large amount of water needs to be vaporized in Dst1.

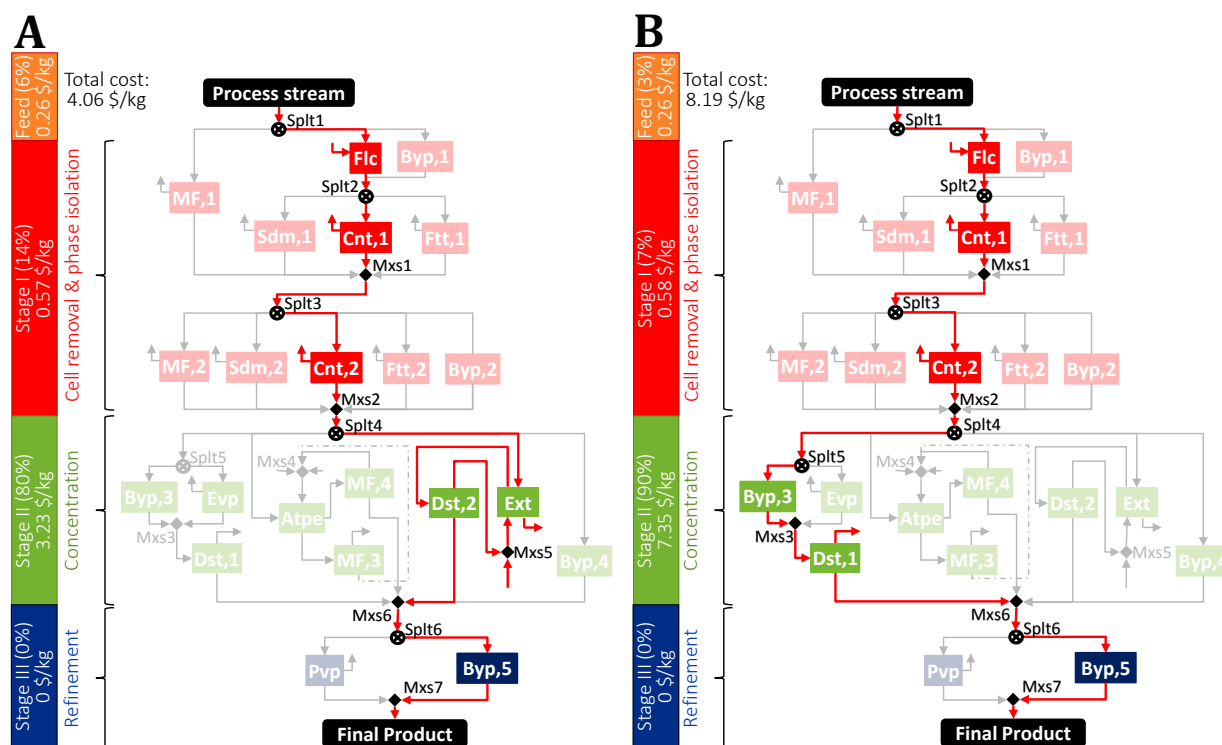


Figure 4.7. Technologies selected for Ext and direct Dst options in stage II for EX SOL LQD NVL CMD product. (A) Technologies selected when Ext is selected; (B) technologies selected when Dst1 is selected. The active streams are shown by bold red lines and selected technologies are highlighted in different colors corresponding to each stage: red for stage I, green for stage II, and blue for stage III. Cost distribution is shown by the numbers on the left bar.

3.3.3 Analysis for EX SOL LQD NVL CMD product

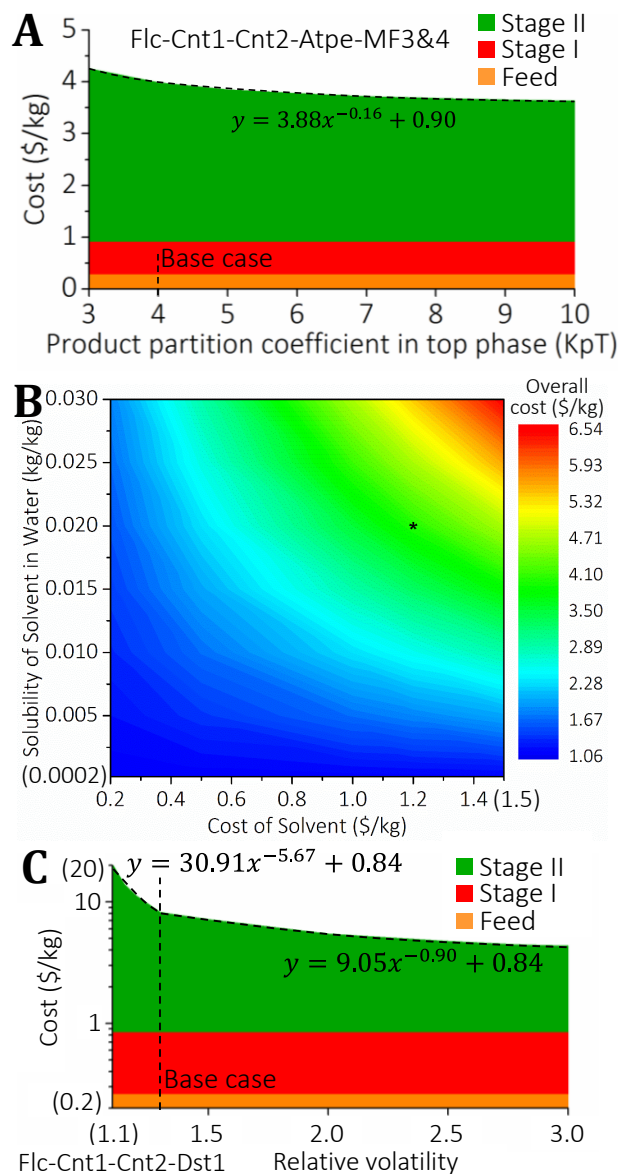


Figure 4.8. Cost analysis for EX SOL LQD NVL CMD product. (A) Analysis with varying Atpe partition coefficient when Atpe selection is fixed; (B) analysis with varying solvent solubility and cost when Ext selection is fixed; (C) analysis with varying relative volatility when Dst1 selection is fixed. The fitted cost-titer functions are shown in (A) and (C), where y represents the cost, and x represents the Atpe product partition coefficient and Dst1 relative volatility, respectively. The based cases are marked with asterisks in (B) and short dashed lines with “Base case” labels next to them in (A) and (C).

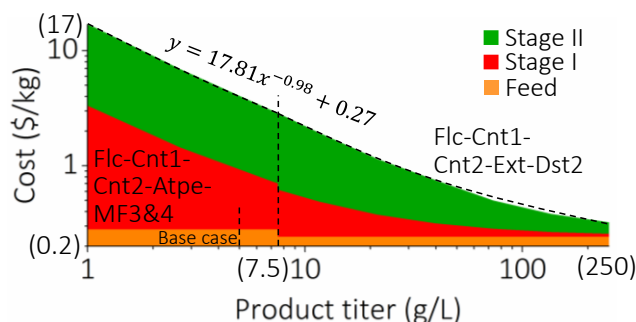


Figure 4.9. Analysis with varying product titer for EX SOL LQD NVL CMD product. The fitted cost-titer functions are shown, where y represents the cost, and x represents the product titer.

Since Stage II is the major cost component, we perform the analysis with varying parameters related to Atpe, Ext and Dst1.

Partition coefficient for Atpe

The major cost driver of Atpe is the equipment size, which is a function of the flowrates of feed and added separating agents (polymer for top phase and salt for bottom phase). The amount of agents required is affected by the partition coefficient for the top phase (K_{pT}), which we vary from 3 to 10 while fixing Atpe selection. In **Figure 4.8A**, we observe that the overall cost decreases from 4.25 \$/kg to 3.63 \$/kg with the increase of partition coefficient from 3 to 10, which enables more product to be extracted into the top phase and thus reduces the amount of separating agents required. When the Atpe partition coefficient is below 3.7 (thus cost > 4.06 \$/kg), Atpe becomes a more expensive option than Ext, assuming the base case parameters for Ext (see **Figure 4.7A**).

Ext parameters

For Ext, solvent is the major cost contributor, and thus we vary solubility of solvent in water from 0.0002 to 0.03 kg/kg and vary cost of solvent from 0.2 to 1.5 \$/kg while fixing Ext selection, and the analysis result is shown in **Figure 4.8B**. Compared with the base case, if the solubility decreases from 0.02 to 0.0002 kg/kg and the solvent cost decreases from 1.2 to 0.2 \$/kg, then the overall cost will be reduced from 4.06 to 1.06 \$/kg, a 74% reduction. Thus, if both parameters can be improved

such that the overall cost is lower than 3.98 \$/kg, then Ext becomes a cheaper option than Atpe, assuming the base case parameters for Atpe (see **Figure 4.6**). Also note that the change in partition coefficient usually does not have a significant impact on the Ext cost when compared with solvent cost and solubility (see **Appendix 3.3**). Therefore, targeting solvents with low water solubility and cost, even if the partition coefficients are low, can help reduce cost because less solvent will be lost.

Dst1 relative volatility

For Dst1, we vary the volatility of water relative to the product from 1.1 to 3, and the analysis result is shown in **Figure 4.8C**, where the cost ranges from 4.34 to 20.35 \$/kg. Thus, even if a relative volatility of 3 can be achieved, direct Dst is still more expensive than Atpe or Ext, assuming their base case parameters.

Product titer

We vary the product titer from 1 to 250 g/L. As a result, the costs and the corresponding optimal technology selection are obtained and presented in **Figure 4.9**. We observe that when the titer is greater than 7.5 g/L, Ext selection is optimal; otherwise, Atpe selection is optimal.

4.3.3.4 Extension to other classes of EX SOL products

The other properties determining the superstructure for EX SOL products are physical state (LQD/SLD), volatility (VOL/NVL), and intended use (CMD/SPC) (see **Figure 4.1B**).

For a SLD product (such as a soluble salt), Mbr and Prc can be used for product concentration as an alternative to Dst, Atpe and Ext in Stage II. Also, in Stage III, Ads, Cry and Dry can be considered. However, the cost will not likely be influenced because the desired product purity is already achieved without Stage III in the base case.

For a VOL product, the product will be obtained at the top instead of at the bottom in Dst. Also, direct Dst is typically cheaper than Ext or Atpe when the relative volatility is greater than 1.05 [135], [178]. For a SPC product, the same argument in **Section 4.3.1.3** holds.

4.4 Conclusions

This work focuses on the synthesis and analysis of separation processes for EX chemicals generated from microbial conversions. We first categorized EX products into (1) NSL LT, (2) NSL HV, and (3) SOL, based on their physical properties. For each category, we presented a representative base case, for which a superstructure was generated, modeled and solved to identify the cost-minimal process and key cost drivers. Next, we analyzed the influence of key parameters on technology selection and cost, which is depicted in the form of sensitivity curves and heat maps. Finally, we extended the discussion to account for other classes of products in the category.

For NSL LT products, the overall cost (including feedstock cost and separation cost) of the base case (5 g/L product titer) is 0.65 \$/kg. Out of the separation cost of 0.26 \$/kg, Stage I (phase isolation) accounts for 96%, and Stage II (concentration) accounts for 4%. Cnt efficiency and product titer are identified to be the major influencers for technology selection and cost. Cnt is the preferred option in most cases.

For NSL HV products, the base case cost is 4.20 \$/kg. Out of the separation cost of 3.81 \$/kg, Stage I accounts for 56%, and Stage III (refinement) accounts for 44%. The required amount of Ddg agent and Slb solvent, Cnt efficiency, Ftt retention factor, and product titer are identified to be the major influencers for technology selection and cost. Ddg and Cnt are the preferred options in most cases.

For SOL products, the base case cost is 3.98 \$/kg. Out of the separation cost of 3.7 \$/kg, Stage I accounts for 17%, and Stage II accounts for 83%. Atpe partition coefficient, Ext solvent solubility

and cost, Dst relative volatility, and product titer are identified to be the major influencers for technology selection and cost. Atpe or Ext is the preferred option in most cases.

In comparison, a NSL LT product has the lowest separation cost because it floats to the top and is thus naturally separated from the microbial cells settling to the bottom. Also, concentrating an NSL product is easier than concentrating a SOL product.

In this work, we have included most of the common technologies to generate reliable insights. However, new technologies can be incorporated by changing model parameters and/or adding new constraints for the corresponding technologies. The insights from the base case results, as well as the predictions associated with the varying model parameters, provide important guidance on the selection of economically promising chemicals generated from microbial conversions and on the design of cost-efficient separation processes.

Chapter 5

5 Framework for the identification of promising products⁵

5.1 Introduction

Recent advances in metabolic engineering and synthetic biology enable the use of microbes for the production of chemicals. *E. coli* and *S. cerevisiae* are the major microbes used for such bio-conversion [179]–[183]. As discussed in **Section 1.2**, compared to traditional fossil fuel-based processes, biological processes can be advantageous for their mild production conditions and good selectivity toward a specific product [77]. The deployment of bio-based chemicals production can be a more attractive near-term goal than that of biofuels, because chemicals have a higher selling price, and their production has potential cost advantages due to two main factors. First, they are produced directly using microbes instead of being converted via multiple conversion steps (some of which can have low yield and high cost) from fossil fuel feedstocks. Second, the “effective hydrogen to carbon ratio” [74], [184] $(H/C_{\text{eff}} = \frac{n(H)-2n(O)}{n(C)})$, where $n(X)$ represents the number of element X in the chemical’s formula; used to evaluate the similarity between the starting and the final material for a conversion) of sugar (~ 0), which is the substrate for microbial cultivation or the carbon storage compound in microbial cells, is closer to that of chemicals (0–1.5) than fuels (1–2.3, which is similar to that of crude oil), indicating potentially simpler production of bio-based chemicals than bio-fuels[185].

⁵ This chapter includes content from: W. Wu, M. R. Long, X. Zhang, J. L. Reed, and C. T. Maravelias, “A framework for the identification of promising bio-based chemicals,” *Biotechnol. Bioeng.*, accepted, 2018. Wenzhao Wu completed the generation of the candidate pool (partially), market analysis, process synthesis and techno-economic analysis, screening criteria development and implementation, and writing of the manuscript (except Sections 5.2.1 and 5.2.2); Matt Long and Xiaolin Zhang performed the yield and productivity estimation based on Flux Balance Analysis and participated in the generation of candidate pool. Matt Long wrote most of Sections 5.2.1 and 5.2.2.

Past studies[186]–[194] mostly focused on common platform chemicals (that could potentially be converted to higher-value ones) and/or products that currently attract investment in development and demonstration. However, some high-value end products are directly producible using microbes; more importantly, the economic prospect of the identified chemicals was not studied thoroughly, and the analysis on the process cost, especially downstream separation cost, was limited. Note that the effluent of a bioreactor is often dilute and the purity requirement for chemicals is higher than that of fuels, and thus the downstream separation tends to be expensive (see **Chapters 1** and **3**). Therefore, bio-separation needs to be carefully considered. For a rough cost estimate, the cost-titer curves generated in **Chapter 4** and [135] can be used directly. For a more accurate estimation, the bio-separation network synthesis framework developed in **Chapter 3** can be used to identify an optimal process configuration. Other methods to synthesize and assess the separation processes are also applicable[1], [3], [5], [10], [31], [36], [38], [45], [51], [90]–[92], [117], [195].

In this chapter, we develop a framework for the identification of economically promising chemicals (only liquid and solid are considered) that can be produced using *E. coli* and *S. cerevisiae* (see **Figure 5.1**). Specifically, we develop a genome scale metabolic modeling-based approach to identify the pool of producible products (“candidate pool”) as well as an estimate of yield (g product/g glucose), productivity (the amount of product produced per unit time and volume, e.g., $\text{g}\cdot\text{L}^{-1}\cdot\text{day}^{-1}$) and residence time (hour) for each. Then, we design three screening criteria based on a product’s (1) profit margin (\$/kg), (2) market volume (million MT/year; MT = metric ton) and (3) market size (million \$/year). The total process cost, including the downstream separation cost, is incorporated into the evaluation of economic prospect. Finally, we apply the screening criteria on the candidate pool to identify promising products.

5.2 Methods

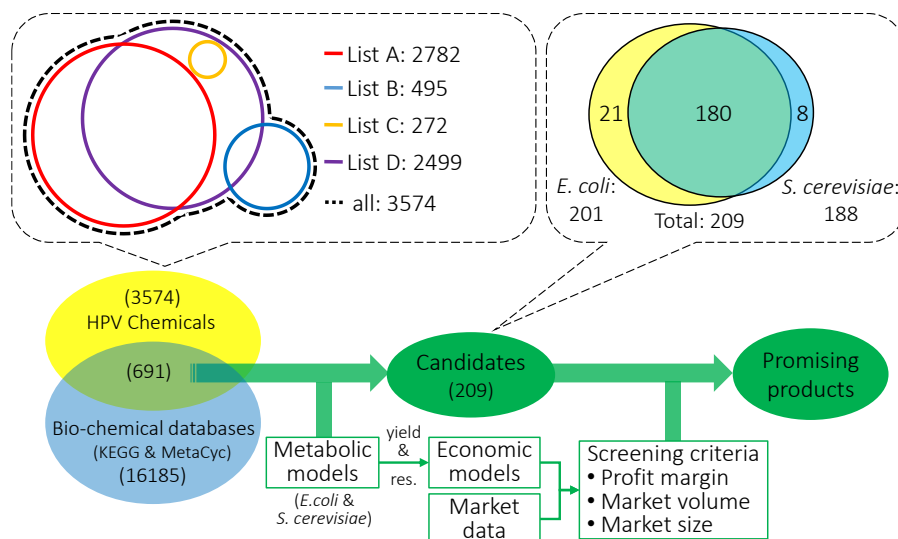


Figure 5.1. Economically promising products identification framework, including the compilation of HPV chemicals, identification of the candidate pool and development of the screening criteria. “Res.” = residence time.

5.2.1 Generation of a candidate pool and production pathways

A US High-Production-Volume (HPV) chemical is manufactured in or imported into the US in amounts equal to or greater than 500 MT/year [196]. We develop an HPV database (including 3574 chemicals) by compiling four HPV lists [197]–[200] published by the Environmental Protection Agency (EPA) (see **Figure 5.1**): List A (published in 1990), List B (published in 1994), List C (an additional supplementary list) and List D (a list currently maintained as a focus for regulation purposes). We target HPV chemicals, which include both commodity chemicals and a portion of fine chemicals, for greater impact.

We then identify all of the HPV chemicals that are contained in the KEGG [201]–[203] and MetaCyc [204] databases using their CAS number annotations. These metabolite and reaction databases include chemicals produced by characterized reactions in a wide variety of biological organisms. The direct overlap with the HPV database is 613 chemicals. In addition, we find 78 HPV chemicals

which are known to be bio-producible [186]–[188] but for which no obvious reaction pathways are found in the KEGG and MetaCyc databases. We manually curate the alignment of these 78 chemicals to correct for potential CAS number mismatches between databases due to the representation of chemicals in different forms (e.g., 50-21-5 is a mixture of D- and L- lactic acid in the HPV database, which does not match the pure-form product CAS number in the KEGG and MetaCyc databases). These types of mixtures, even if present in a database (e.g., KEGG compound C01432 is a non-stereospecific form of lactic acid), do not generally participate in any reactions since enzyme catalyzed reactions tend to be stereo and isomer specific and thus have no production pathway. We then align both the iJO1366 [205] and iMM904 [206] genome scale metabolic models for *E. coli* and *S. cerevisiae* to both the KEGG and MetaCyc using the BIGG database [207] alignment for metabolites and reactions between the metabolic models and databases. Both models are run under glucose-limited conditions with no ATP maintenance (ATPM) energy. The ATPM requirement is removed to allow for linear scaling of the solutions based upon changes in the glucose uptake rate. This means the calculated values of the biomass growth rate and production fluxes can be scaled by the change between the original simulated glucose uptake rate and any other glucose uptake rate, and would have the same results if the model were re-run utilizing the new glucose uptake rate. With a glucose uptake rate of 10 mmol/gDW/hr, removing the ATPM maintenance constraint results in only a 1.6% and 0.6% increase in predicted growth rate for the iJO1366 and iMM904 models, respectively. This indicates that the ATPM constraints have a small impact on the model predictions. The glucose uptake rate is set so that the aerobic wild type growth rate is consistent with a doubling time of 0.7 hour for *E. coli* [208] and 2.3 hours for *S. cerevisiae* [209].

To allow the metabolic models to interact with the databases, we allow each common metabolite to be transferred from the metabolic models to the respective databases. For KEGG, we utilize only reactions which are atom-balanced; however, due to differences in protonation state between the

database and physiologically relevant conditions, we allow reactions which are not proton balanced. Furthermore, since KEGG does not specify reaction directionality, all reactions are treated as reversible. For MetaCyc, we utilize only reactions which are marked as balanced and we utilize the suggested reaction directionality annotated in the database. HPV chemical production pathways are initially analyzed with the KEGG database and only chemicals which could not be produced with KEGG are analyzed with MetaCyc.

For each product, we first identify which additional reactions, if any, need to be added from a universal database to enable production. This process is similar in principle to OptStrain [210]. Specifically, we first maximize the yield of the product using Flux Balance Analysis [211] allowing for the model to utilize the entirety of the reaction database. If the product could be made, we next minimize the number of non-native reactions utilized while requiring at least 1% of the previously calculated maximum. Note that while requiring at least 1% production, we are not limiting the production to only 1%, and pathways often have a maximum yield in excess of 50% of the theoretical carbon mole yield (i.e., all carbons in the substrate are contained in the product). Utilizing a small minimum threshold is necessary since the maximum production rate calculated in the first step may be larger than the theoretical carbon mole yield due to CO₂ fixation in *E. coli* or other thermodynamically infeasible loops which may generate free mass, energy, and reducing units. Note that *E. coli* is able to uptake CO₂ and is thus permitted to do so by the iJ01366 model while *S. cerevisiae* is unable to do so and therefore blocked by the iMM904 model. Such infeasible steps arise because database reactions are generally reversible and are not necessarily proton balanced; therefore, the solver may include a set of extra reactions which improve the predicted yield by generating free resources even if such processes are not thermodynamically possible. By minimizing the number of database reactions utilized while still requiring some level of production, we ensure that only reactions necessary for making the product are utilized, and the shortest path

necessary to produce the product is identified. Furthermore, minimizing the number of additional reactions minimizes the number of heterologous enzymes required to engineer a production strain, which is a desirable trait for a production pathway. After identifying the minimum number of additional reactions, we again maximize the yield while allowing only the minimum number of additional reactions needed for production to be used as identified in the previous step. This step, while the most computationally expensive, ensures that we identify the exact minimum length pathway with the highest yield and not an alternate minimum length pathway with lower yield. For example, a pathway utilizing NADH as a cofactor may have a different yield than one utilizing NADPH even though the same number of additional reactions are utilized. Some of these minimum length pathways include additional reactions with CO₂ fixation, resulting in extremely high theoretical yields in *E. coli* where CO₂ is permitted to be directly fixed. All products created by this type of pathway in *E. coli* are manually removed from the results. The specific calculation methods can be found in the Supporting Material of [212].

The above process results in 209 products (candidate pool) which are identified to be producible by *E. coli* and *S. cerevisiae*, often with the addition of heterologous reactions. Of the 209 products, 180 can be produced with both *E. coli* and *S. cerevisiae*; 21 can only be produced with *E. coli*, and 8 only with *S. cerevisiae* (see **Figure 5.1**).

5.2.2 Estimation of yield, productivity and residence time

After finding the production pathway for each product, we then identify the production space constraints utilizing the Flux Envelope Analysis approach (submitted paper attached together with our current submission) in both aerobic and anaerobic environments. This allows us to quickly calculate the yield and growth rate for a variety of different objectives: the maximum theoretical yield, maximum productivity in a batch reactor, and the maximum productivity in a chemostat

(operated continuously). Utilizing the growth rate and yield, we then calculate the corresponding residence time and productivity for each product in different reactors. Specifically, we assume a 240 g/L input glucose concentration into the fermenter and an output of 4.52 g/L; the input cell concentration is 9.8 gDW/L (generated from a seed fermenter and mixed with the glucose feed; DW = dry weight). These numbers are adopted from a simulation model[125] of the NREL process [213] that evaluates the economics of converting 2000 MTDW/day cellulosic biomass into ethanol. **Figure 5.2** shows an example of the differences between maximum yield, batch productivity, and chemostat productivity for the aerobic production of ethanol in *E. coli*.

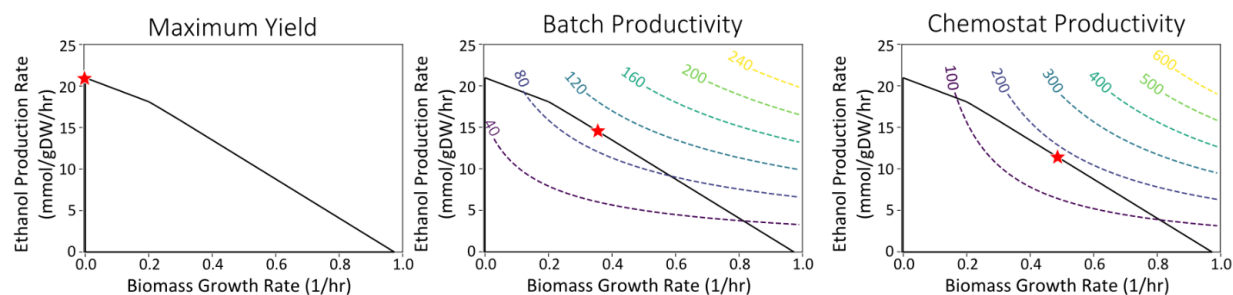


Figure 5.2. Solution space for the production of ethanol relative to biomass growth rate for glucose-limited aerobic *E. coli*. The maximum yield, batch reactor productivity, and chemostat productivity are shown with red stars on their respective plots. The contours represent productivities.

5.2.3 Development of screening criteria

We first collect price and volume data estimates for the 209 products from the ICIS [214], CDAT [215], Alibaba.com, and IUR [216] databases. Next, we develop three screening criteria for each product to quantify economic prospect.

Criterion 1: The *profit margin* (\$/kg) should be positive.

The economic prospect of a chemical is largely influenced by its profit margin, i.e., the difference between a product's selling price and its total unit production cost. However, the downstream separation process configuration is highly product-dependent (and unknown), and thus an accurate

separation cost is difficult to estimate for each product. On the other hand, the upstream cost (related to bio-conversion, including costs of feedstocks, capital, utilities, etc.) is easier to estimate because there is limited configuration variations (see calculation in **Appendix 4.**) Therefore, we define the “separation cost margin”, which is the difference between a product’s price and its upstream cost, to represent the maximum allowable downstream separation cost rendering the product breakeven (zero profit). Hence, for a specific product, if the separation cost margin is negative, then we conclude that the product fails Criterion 1 (because separation costs are always positive); otherwise, we further evaluate whether it is greater than the separation cost. We use the cost-titer curves (see **Figure 5.3**) generated for different classes of products based on superstructure optimization (see **Chapter 4** and [135]) as an example of the separation cost approximate.

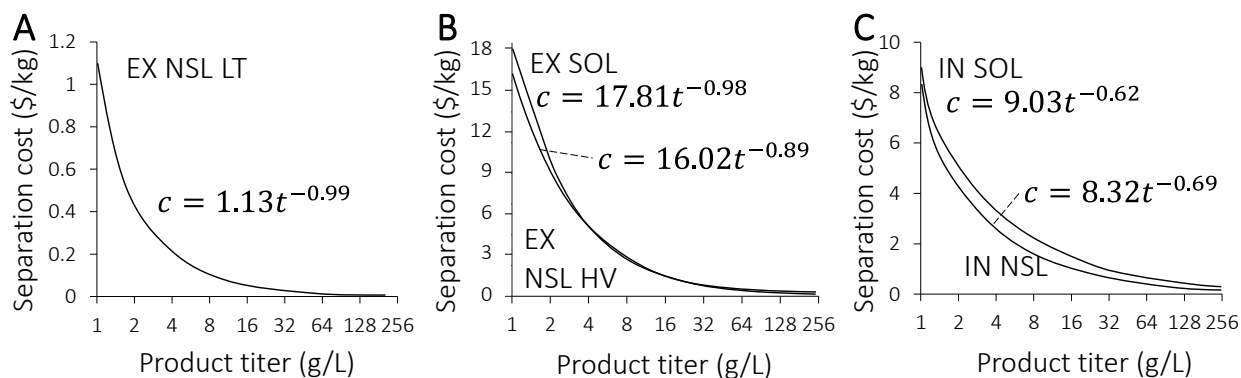


Figure 5.3. Separation cost-product titer curves (logarithmic scale) for different classes of products. (A) EX NSL LT products; (B) EX SOL and EX NSL HV products; (C) IN SOL and IN NSL products. The equations are fitted functions generated based on the curves, where c represents the separation cost and t represents the product titer. IN/EX = intracellular/extracellular; SOL/NSL = soluble/ insoluble in water; HV/LT = heavy/light (in terms of density) in water.

From the functions representing each curve in **Figure 5.3**, we see that EX NSL LT products have the lowest costs because a simple decantation usually suffices.

Criterion 2: The *market volume* (million MT/year) should be greater than a minimum production capacity.

The minimum production capacity is assumed to be 30000 MT/year product here, which is ~20% of the NREL bio-refinery size (in terms of the amount of product). Note that this value can be raised for higher impact or lowered if low-volume chemicals (such as fine chemicals) need to be considered.

Criterion 3: The *market size* (price $\times$ volume; million \$/year) should meet revenue expectations so the capital investment can be recovered.

We choose a minimum revenue of 5 million \$/year and assume that no more than 20% of the current volume can be replaced. Therefore, the market size should be at least 25 million \$/year (i.e., 5 million \$/year $\div$ 20%).

5.3 Results & discussions

5.3.1 Identification of the most cost-favorable systems

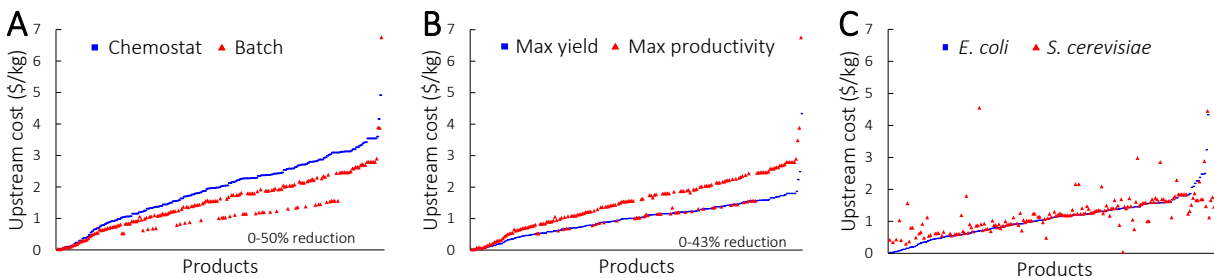


Figure 5.4. Comparisons of the upstream costs of the 209 products. (A) Chemostat vs. batch at maximum productivity (B) maximum yield vs. maximum productivity; (C) *E. coli* vs. *S. cerevisiae*. The products are ordered such that the costs of the blue markers (representing chemostat, maximum yield and *E. coli*, respectively) increase monotonically. For each option in a specific feature to be compared (e.g., production goal in B), the options in the other features (reactor type and microbe) are chosen such that the cost reaches the lowest possible level.

A product can be produced in a system described by the following features (each with 2 options): production goal (maximum yield or maximum productivity) and microbe (*E. coli* or *S. cerevisiae*).

Also, at maximum productivity, the reactor type (chemostat or batch) also influences yield and thus cost. We first identify the most cost-favorable systems for the products.

The calculated upstream costs of the 209 products for both options in a given feature are presented in **Figure 5.4**. For each option in a specific feature to be compared (e.g., production goal in **Figure 5.4B**), the options in the other features (reactor type and microbe) are chosen such that the cost reaches the lowest possible level. From **Figure 5.4**, we have the following observations:

- (1) *Batch vs. chemostat*: at maximum productivity, operating in batch reactors renders 0-50% lower costs than operating in chemostats for the 209 products, mainly because the estimated yield in a batch reactor is higher than that in a chemostat. This is due to the fact that the productivity equation for a batch reactor is equivalent to that of a chemostat with a penalty for larger growth rates. Thus, a smaller growth rate in a batch reactor results in a higher yield and thus a lower cost since yield is the main contributor to cost;
- (2) *Maximum yield vs. maximum productivity*: operating at maximum yield renders 0-43% lower costs than at maximum productivity (in a batch reactor) since yield is the main cost contributor;
- (3) *E. coli vs. S. cerevisiae*: 21 products can only be produced with *E. coli*, and 153 are cheaper to produce with *E. coli*; 8 can only be produced with *S. cerevisiae*, and 27 are cheaper.

We also compare aerobic and anaerobic options: the anaerobic yield and productivity are estimated to be lower than (for 15 products) or equal to (for the other 194) those of aerobic, since our metabolic modeling approach for the anaerobic option is the same with that for the aerobic one but with an additional constraint excluding oxygen uptake. The cost of using compressors to maintain the aerobic production environment is fairly low. For example, with a VVM (aeration rate - volume of gas per volume of liquid per minute) of 0.013 min^{-1} (the NREL value) increases the unit capital cost (\$/kg) by roughly 5.5% and unit utility cost by 11% in the base case (see **Appendix 4**), but the

total unit cost is increased by just 0.4% compared with the anaerobic process because the major cost component is feedstock (even with a VVM of 0.1 min^{-1} , which is 8 times the NREL value, the total unit cost is increased by 4%). Therefore, due to such a small additional cost and the potential improvement in yield, the aerobic option will in general lead to better economics.

5.3.2 Screening results

In the most cost-favorable systems (operating at maximum yield, and deciding on *E. coli* vs. *S. cerevisiae* based on the specific products) discussed in **Section 5.3.1** for all the 209 products, 45 products satisfy Criteria 2 and 3 in **Section 5.2.3** and have positive separation cost margins. Therefore, we further estimate the separation costs of the 45 products.

The 45 products are first classified into NSL LT (5), NSL HV (12) and SOL (28). We assume that each product can be produced extracellularly (EX) or intracellularly (IN). Then, the titers of each product at the maximum yield and maximum productivity are used to estimate the separation costs based on curves in **Figure 5.3**. We consider operating at maximum productivity as well because it is a relatively more achievable goal in real productions today. Finally, the profit margin is calculated by subtracting the separation cost from the separation cost margin. Also, the breakeven titer that renders zero profit margin for each product is calculated. The screening results are shown in **Tables 5.1-5.3** and **Figure 5.5**. For a given product, if the profit margin is positive, then the product is economically promising (marked “yes” in bold and red in **Tables 5.1-5.3**). To summarize, all classes of products combined, 32 products are economically promising at maximum yield, and 22 are promising at maximum productivity. The minimum breakeven titer, among all products, is 15 g/L. Note that “promising” does not mean that industrial-scale production today would be profitable. Instead, we use the term to identify products whose production can be profitable, if reasonable further advances in metabolic engineering and separation technologies, are achieved in

the near future. To that end, we have deliberately made optimistic assumptions, such as maximum yield/productivity and minimum separation cost, so that no products are “cut-off”. For a more accurate techno-economic evaluation of a process for a specific product, experimental data (e.g., to determine yield and residence time) and detailed process synthesis, simulation, and analysis are necessary.

From the comparison between EX and IN in **Tables 5.1-5.3** and **Figure 5.5**, we have the following observations, which can be explained by the curves in **Figure 5.3**:

(1) For any NSL LT product, EX production leads to a higher profit margin than IN production due to a much lower separation cost; (2) for a product belonging to the other classes: (i) at a relatively high product titer (e.g., at maximum yield for all products), EX production leads to a higher profit margin than IN production due to the additional cost of cell disruption to release the IN product; (ii) at a relatively low product titer (e.g., at maximum productivity for some products), IN production leads to a higher profit margin than EX production because separating the product from large amount of EX water is difficult.

In addition, to facilitate the users in choosing between microbes, a comparison between *E. coli* and *S. cerevisiae* (assuming an optimal localization - EX/IN) is shown in **Figure 5.6**.

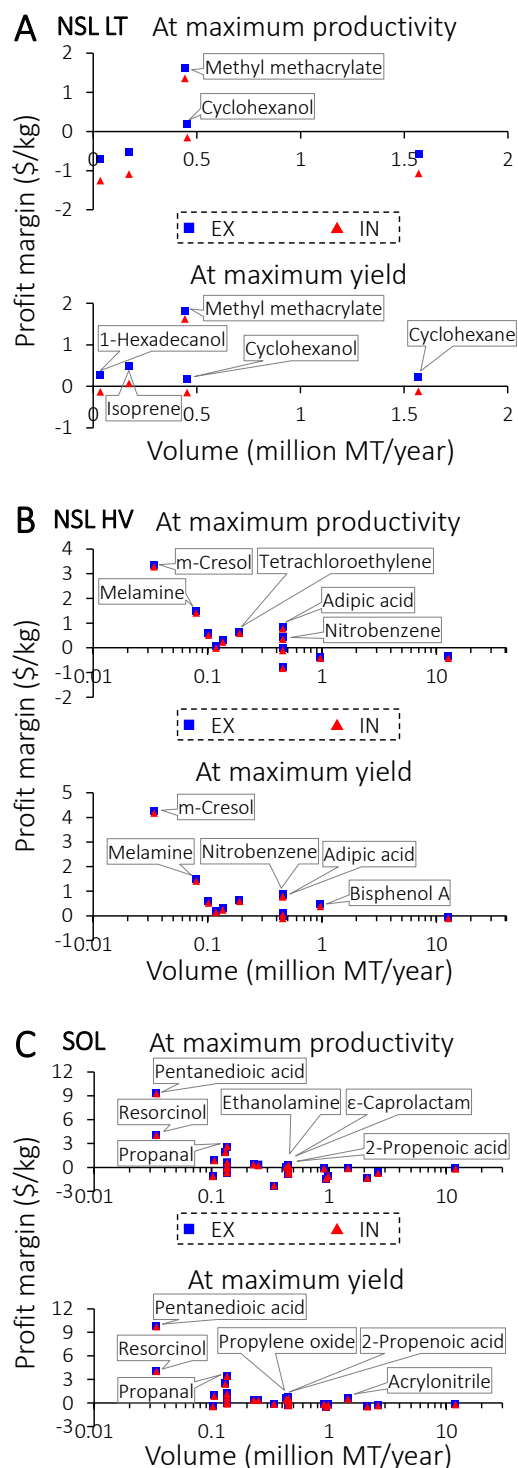


Figure 5.5. Graphic representation of the screening results at maximum yield and maximum productivity, where EX and IN productions are compared when the optimal microbe is selected (corresponding to **Tables 5.1-5.3**). **(A)** NSL LT products; **(B)** NSL HV products; **(C)** SOL products. **(B)** and **(C)** are plotted on logarithmic scales. Promising products with top 3 profit margins and volumes are labeled on each plot.

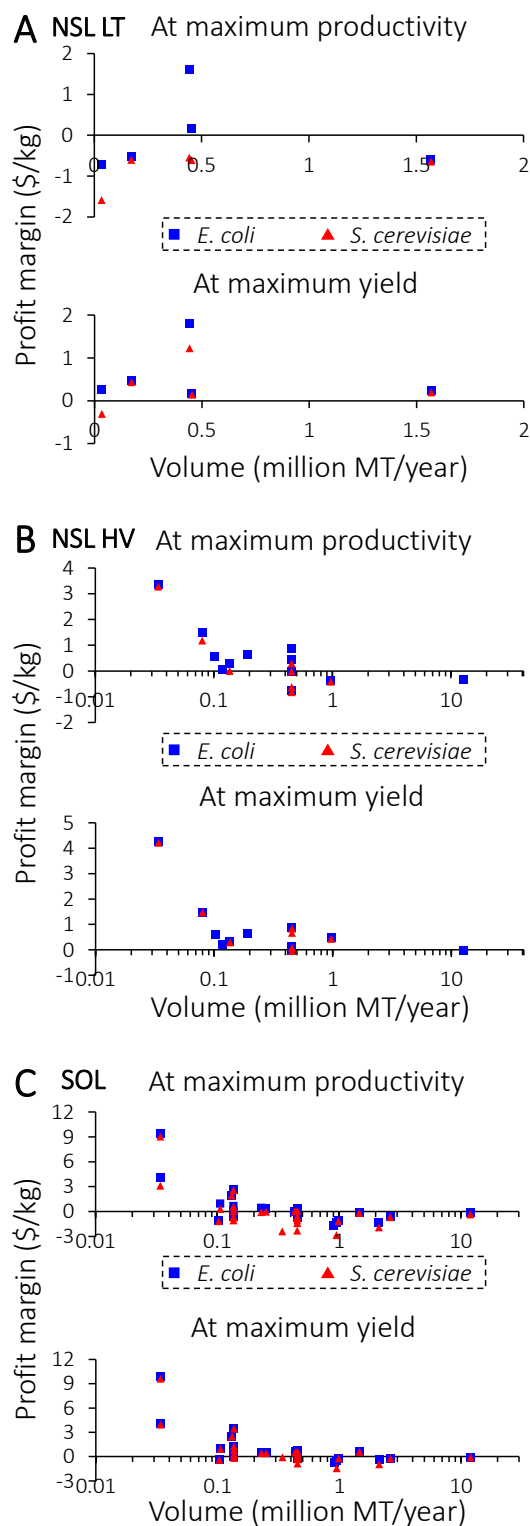


Figure 5.6. Graphic representation of the screening results at maximum yield and maximum productivity, where microbes are compared at the optimal localization. (A) NSL LT products; (B) NSL HV products; (C) SOL products. (B) and (C) are plotted on logarithmic scales.

Table 5.1. Screening results for NSL LT products. Promising products are marked “yes” in bold and red. Products are ordered by market volume. Size = market size; prom. = promising; prod. = productivity; yr = year; max. = maximum; opt. = optimal; mic. = “microbe; MM = million; *E.* = *E. coli*. Feed glucose concentration is assumed to be 24 wt% in calculating titer.

CAS #	Product name	Volume (MM MT/yr)	Size (MM \$/yr)	Opt. mic.	EX/IN	Prom. at max. yield?	Prom. at max. prod.?	At max. yield Titer (g/L)	Prof. margin (\$/kg)	At max. prod. Titer (g/L)	Prof. margin (\$/kg)	Break-even titer (g/L)
36653-82-4	1-Hexadecanol	0.03	68	<i>E.</i>	EX IN	Yes No	No No	79	0.26 -0.13	50	-0.71 -1.26	69 85
78-79-5	Isoprene	0.17	397	<i>E.</i>	EX IN	Yes Yes	No No	76	0.48 0.07	48	-0.53 -1.09	60 73
80-62-6	Methyl methacrylate	0.44	1062	<i>E.</i>	EX IN	Yes Yes	Yes Yes	235	1.81 1.62	154	1.62 1.36	57 70
108-93-0	Cyclohexanol	0.45	680	<i>E.</i>	EX IN	Yes No	Yes No	105	0.18 -0.15	105	0.18 -0.15	92 115
110-82-7	Cyclohexane	1.57	2665	<i>E.</i>	EX IN	Yes No	No No	94	0.23 -0.12	60	-0.59 -1.07	81 101

Table 5.2. Screening results for NSL HV products. Promising products are marked “yes” in bold and red. Products are ordered by market volume. Size = market size; prom. = promising; prod. = productivity; yr = year; max. = maximum; opt. = optimal; mic. = “microbe; MM = million; *E.* = *E. coli*. Feed glucose concentration is assumed to be 24 wt% in calculating titer.

CAS #	Product name	Volume (MM MT/yr)	Size (MM \$/yr)	Opt. mic.	EX/IN	Prom. at max. yield?	Prom. at max. prod.?	At max. yield Titer (g/L)	Prof. margin (\$/kg)	At max. prod. Titer (g/L)	Prof. margin (\$/kg)	Break-even titer (g/L)
108-39-4	m-Cresol	0.03	201	<i>E.</i>	EX IN	Yes Yes	Yes Yes	100	4.26 4.18	63	3.33 3.28	28 27
108-78-1	Melamine	0.08	158	<i>E.</i>	EX IN	Yes Yes	Yes Yes	329	1.49 1.42	329	1.49 1.42	83 86
79-01-6	Trichloroethylene	0.10	92	<i>E.</i>	EX IN	Yes Yes	Yes Yes	515	0.58 0.52	515	0.58 0.52	183 200
75-09-2	Dichloromethane	0.12	53	<i>E.</i>	EX IN	Yes Yes	Yes No	665	0.20 0.15	423	0.06 0.00	367 425
108-80-5	Cyanuric acid	0.14	108	<i>E.</i>	EX IN	Yes Yes	Yes Yes	337	0.31 0.24	337	0.31 0.24	208 229
127-18-4	Tetrachloroethylene	0.19	172	<i>E.</i>	EX IN	Yes Yes	Yes Yes	650	0.65 0.59	650	0.65 0.59	183 200
62-53-3	Aniline	0.45	770	<i>E.</i>	EX IN	Yes Yes	No No	104	0.12 0.04	66	-0.77 -0.82	97 101
98-95-3	Nitrobenzene	0.45	860	<i>E.</i>	EX IN	Yes Yes	Yes Yes	161	0.87 0.79	111	0.43 0.35	87 90
124-04-9	Adipic acid	0.45	820	<i>E.</i>	EX IN	Yes Yes	Yes Yes	174	0.86 0.78	174	0.86 0.78	91 94
100-21-0	Terephthalic acid	0.45	450	<i>E.</i>	EX IN	No No	No No	163	-0.02 -0.10	163	-0.02 -0.10	166 180
80-05-7	Bisphenol A	0.97	1881	<i>E.</i>	EX IN	Yes Yes	No No	112	0.47 0.39	71	-0.36 -0.42	85 88
107-06-2	1,2-Dichloroethane	12.75	6375	<i>E.</i>	EX IN	No No	No No	310	-0.03 -0.10	197	-0.33 -0.41	330 378

Table 5.3. Screening results for SOL products. Promising products are marked “yes” in bold and red. Products are ordered by market volume. Size = market size; prom. = promising; prod. = productivity; yr =year; max. = maximum; opt. = optimal; mic. = “microbe; MM = million; *E.* = *E. coli*; *S.* = *S. cerevisiae*. Feed glucose concentration is assumed to be 24 wt% in calculating titer.

CAS #	Product name	Volume (MM MT/yr)	Size (MM \$/yr)	Opt. mic.	EX/ IN	Prom. at max. yield?	Prom. at max. prod.?	At max. yield		At max. prod.		Break- even titer (g/L)
								Titer (g/L)	Prof. margin (\$/kg)	Titer (g/L)	Prof. margin (\$/kg)	
108-46-3	Resorcinol	0.03	188	<i>E.</i>	EX IN	Yes Yes	Yes Yes	133	4.10 4.03	133	4.10 4.03	33 31
110-94-1	Pentanedioic acid	0.03	366	<i>E.</i>	EX IN	Yes Yes	Yes Yes	207	9.84 9.76	131	9.34 9.28	17 15
141-78-6	Ethyl acetate	0.10	103	<i>E.</i>	EX IN	No No	No No	138	-0.36 -0.43	88	-1.11 -1.12	190 207
79-09-4	Propanoic acid	0.11	221	<i>E.</i>	EX IN	Yes Yes	Yes Yes	166	0.95 0.88	166	0.95 0.88	89 90
79-06-1	2-Propenamide	0.13	464	<i>E.</i>	EX IN	Yes Yes	Yes Yes	186	2.51 2.43	118	1.95 1.90	52 50
102-71-6	Triethanolamine	0.14	162	<i>E.</i>	EX IN	No No	No No	156	-0.02 -0.10	156	-0.02 -0.09	159 170
107-92-6	Butanoic acid	0.14	204	<i>E.</i>	EX IN	Yes Yes	No No	136	0.12 0.06	87	-0.63 -0.64	125 130
111-42-2	Diethanolamine	0.14	327	<i>E.</i>	EX IN	Yes Yes	Yes Yes	165	1.26 1.18	105	0.63 0.59	77 77
123-38-6	Propanal	0.14	699	<i>E.</i>	EX IN	Yes Yes	Yes Yes	114	3.49 3.44	72	2.59 2.61	35 33
71-23-8	1-Propanol	0.14	285	<i>E.</i>	EX IN	Yes Yes	No No	105	0.31 0.27	66	-0.67 -0.63	89 90
75-07-0	Acetaldehyde	0.14	300	<i>E.</i>	EX IN	Yes Yes	Yes Yes	138	0.84 0.78	88	0.10 0.08	84 85
56-81-5	Glycerol	0.23	313	<i>E.</i>	EX IN	Yes Yes	Yes Yes	207	0.44 0.36	207	0.44 0.36	138 145
77-92-9	Citric acid	0.25	304	<i>E.</i>	EX IN	Yes Yes	Yes Yes	247	0.44 0.35	209	0.31 0.23	156 166
106-89-8	Epichlorohydrin	0.34	512	<i>S.</i>	EX IN	No No	No No	120	-0.09 -0.14	47	-2.41 -2.27	125 130
141-43-5	Ethanolamine	0.43	669	<i>E.</i>	EX IN	Yes Yes	Yes No	195	0.57 0.49	124	0.04 -0.02	121 126
105-60-2	ϵ -Caprolactam	0.45	860	<i>E.</i>	EX IN	Yes Yes	Yes Yes	118	0.31 0.26	118	0.31 0.26	98 100
123-72-8	Butanal	0.45	910	<i>E.</i>	EX IN	Yes Yes	No No	103	0.19 0.15	65	-0.80 -0.76	93 94
57-55-6	Propylene Glycol	0.45	550	<i>E.</i>	EX IN	No No	No No	148	-0.06 -0.13	95	-0.74 -0.77	156 166
75-56-9	Propylene oxide	0.45	910	<i>E.</i>	EX IN	Yes Yes	No No	114	0.36 0.31	72	-0.54 -0.52	93 94
75-86-5	Acetone cyanohydrin	0.45	820	<i>E.</i>	EX IN	No No	No No	89	-0.28 -0.30	89	-0.29 -0.30	103 106
79-10-7	2-Propenoic acid	0.45	770	<i>E.</i>	EX IN	Yes Yes	Yes Yes	188	0.69 0.61	119	0.14 0.09	110 113
71-36-3	1-Butanol	0.91	1089	<i>S.</i>	EX IN	No No	No No	146	-0.08 -0.15	146	-0.08 -0.15	157 168

67-63-0	2-Propanol	0.95	1257	E.	EX IN	No No	No No	106	-0.46 -0.50	66	-1.44 -1.40	142 150
108-94-1	Cyclohexanone	0.99	1485	E.	EX IN	No No	No No	110	-0.20 -0.25	70	-1.13 -1.11	125 130
107-13-1	Acrylonitrile	1.46	2900	E.	EX IN	Yes Yes	No No	139	0.63 0.56	88	-0.12 -0.13	94 95
108-95-2	Phenol	2.13	2908	E.	EX IN	No No	No No	105	-0.41 -0.45	67	-1.38 -1.34	138 145
64-19-7	Acetic acid	2.63	1566	E.	EX IN	No No	No No	235	-0.22 -0.30	149	-0.66 -0.73	323 379
57-13-6	Urea	12	140	E.	EX IN	No No	No No	470	-0.10 -0.18	470	-0.10 -0.18	635 832

5.3.3 Remarks

The proposed screening framework is flexible in that it can account for updated price and volume data, additional products, and modified screening criteria (e.g., applying market volume and market size criteria that suit the users' unique goals). Also, the current framework can be extended to analyze other microbes (such as algae and cyanobacteria) and bio-conversion systems (such as closed photo-bioreactors and open ponds) by modifying the metabolic and cost models. The specific threshold values in the screening criteria can also be modified accordingly. For example, the upstream cost in Criterion 1 depends on the yield and residence time for the specific microbe adopted, as well as the cost of the bio-conversion system; the production capacity in Criterion 2 for an open pond system may be estimated based on the expected land area available for the ponds.

Note that we do not consider replacement chemicals in this work, because they will compete with the existing ones as potential substitutes. With the current framework, we can identify promising existing products whose substitution, if successful, would generate significant economic impact on the market.

5.4 Conclusions

In this work, we developed an identification framework for promising bio-based chemicals, aiming at market economic impact. Specifically, a metabolic modeling-based approach was developed to identify 209 products producible using *E. coli* and *S. cerevisiae* (together with the estimated yield, productivity and residence time for each) from the intersection of HPV and bio-chemical databases (KEGG and MetaCyc). Cost-titer curves were also generated based on superstructure optimization for the estimation of separation costs. Then, three screening criteria were developed to identify promising products based on their physical properties (SOL/NSL and LT/HV). Given the three assumed criteria in this work, we identify 32 products as economically promising if the maximum yields can be achieved, and 22 products if the maximum productivities can be achieved. It was also found that the following three measures usually have cost benefits: (1) increasing product yield even if residence time will be increased as a result; (2) operating in batch if the goal is to maximize productivity; (3) deciding on *E. coli* vs. *S. cerevisiae* based on specific products (*E. coli* usually leads to lower costs). A major challenge the proposed framework overcomes is that the total process cost, especially downstream separation cost, is now systematically incorporated into the evaluation of economic prospect. The framework is also flexible in that the databases, yield estimations, and criteria can be modified to customize the screening.

Chapter 6

6 Framework for the identification of promising replacement chemicals⁶

6.1 Introduction

Most organic chemicals are produced from fossil fuel feedstocks today, which poses environmental concerns. Also, such production processes usually involve multiple conversion steps, some of which can have low yield and high cost. Therefore, as discussed in **Chapter 1**, alternative production methods, especially bio-conversion, have been the focus of many studies. In **Chapter 5**, we have identified specific existing chemicals as promising targets for bio-production. However, some bio-derivable chemicals may currently have little or no demand but could potentially replace existing chemicals due to environmental or economic advantages, upon successful deployment of bio-production in the future. The goal of this study is to help identify such promising replacement chemicals. Directly evaluating the potential of such chemicals, which either have almost no market and production data or have not even been discovered yet, is extremely difficult. Therefore, instead, we aim to identify the molecular *characteristics* of potentially promising ones as a surrogate to the original problem. Specifically, these characteristics should be: (1) demanded by the market, and (2) difficult to obtain from fossil fuel feedstocks (thus favoring replacement). Further, in the context of bio-production, the replacement chemicals should be relatively easy to obtain through bio-conversion.

⁶ This chapter includes content from: W. Wu and C. T. Maravelias, "Identifying the characteristics of potentially promising bio-based replacement chemicals," in preparation, 2018.

6.2 Methods

We first gather the molecular structures, market prices and market volumes of 44 organic chemicals (totaling 161 million MT/year; MT=metric ton; each has volume between 0.2 and 41 million MT/year), including the top 20 commodity organic chemicals by US volume [217]–[219] and 24 other High-Production-Volume (HPV) chemicals that can be produced through microbial conversion using *E. coli* and/or *S. cerevisiae*. These 44 chemicals roughly represent 85% of the total US organic chemical market volume [220], [221].

Next, we analyze 5 molecular characteristics for each chemical, including (1) number of carbon atoms (abbreviated as #C hereafter), (2) number of oxygen atoms (#O), (3) number of function groups (#FG), (4) number of distinct functional groups (#DFG; e.g., 2 hydroxyl FGs are counted as 1 DFG), and (5) existence of specific functional groups (FG) such as alkenyl, hydroxyl and phenyl groups as well as their combinations. Each characteristic corresponds to multiple attributes, e.g., #C=1 is an attribute of the #C characteristic. For each of the 5 characteristics, we summarize the market volume and price data for all the 44 chemicals. Market volume can be regarded as a surrogate for the demand of a set of attributes and price as a surrogate for the current cost of obtaining the attributes (mainly from fossil fuel feedstocks such as natural gas and petroleum). Thus, we identify the molecular characteristics that are demanded by the current market (indicated by market volume) and difficult to obtain from fossil fuel feedstocks (indicated by price), as shown in **Section 6.3**. Finally, we discuss the difficulty of obtaining these characteristics through bio-production, as shown in **Section 6.4**.

Accuracy of the proposed analysis is influenced by (1) accuracy of the market price and volume data for each chemical, and (2) percentage of the total organic chemical market volume represented by the 44 chemicals.

6.3 Results

In this section, we present the analyses for #C, #O, #FG, #DFG and specific FGs. Major insights are presented in bullet forms.

6.3.1 Number of carbon atoms (#C)

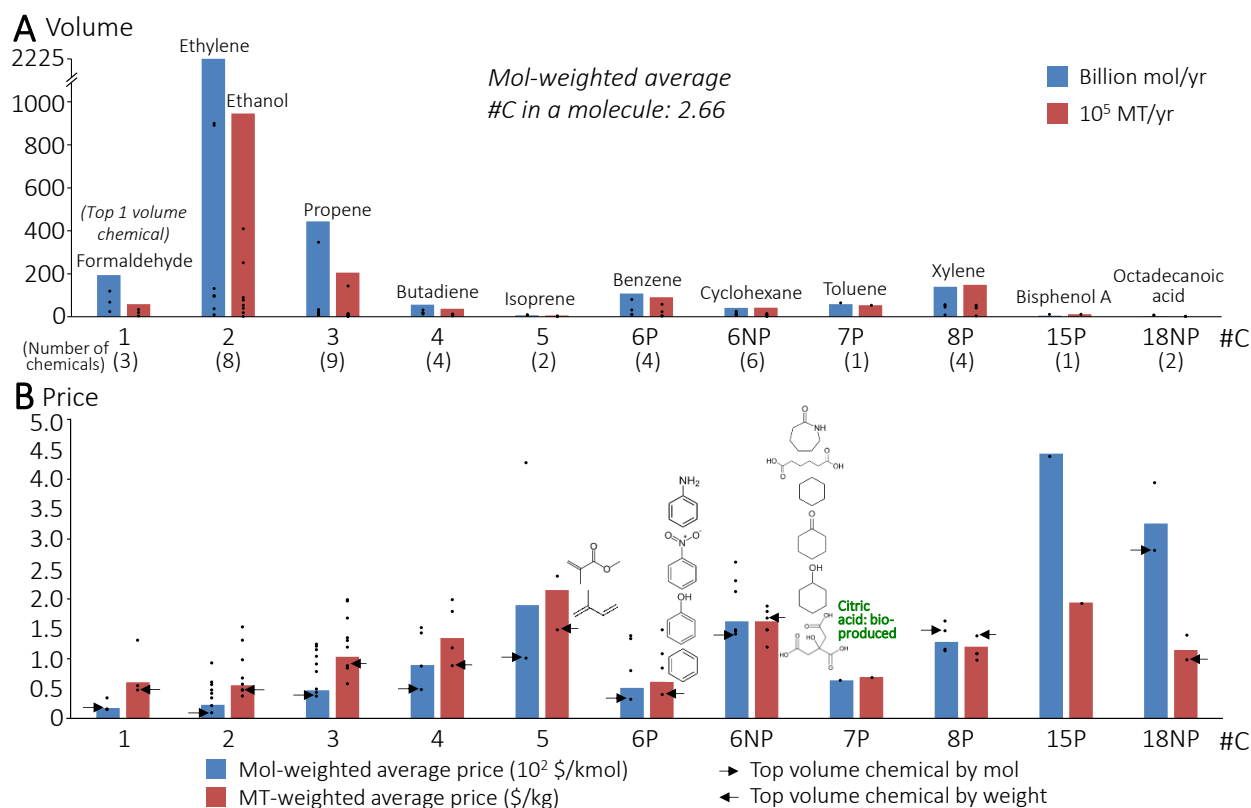


Figure 6.1. Results for #C. (A) Volume vs. #C; each blue or red bar represents the total volume (billion mol/year or 10^5 MT/year) of chemicals with the corresponding #C; number of chemicals with the corresponding #C are labeled at the bottom; top 1 volume chemical for each #C is labeled above the two bars, and only one is labeled if the top 1 volume chemical is the same for both bars; each black dot represents a chemical; “P” in #C represents phenyl groups, e.g., 6P denotes a chemical with phenyl groups and 6 carbon atoms; “NP” denotes the non-existence of phenyl groups; (B) price vs. #C; each blue or red bar represents the mol-weighted average price (10^2 \$/kmol) or MT-weighted average price (\$/kg) of chemicals with the corresponding #C; arrows represent top volume chemicals; molecular structures of chemicals with #C = 5, 6P and 6NP are shown, ordered in consistency with the MT-weighted average price (\$/kg) of each chemical; chemicals that are currently bio-produced are labeled bold in green.

The volume and price data for chemicals with different #C are shown in **Figure 6.1**, where “P” in #C represents phenyl groups, e.g., 6P denotes a chemical with phenyl groups and 6 carbon atoms; “NP” denotes the non-existence of phenyl groups.

Key insights from **Figure 6.1A** are summarized as follows:

- C2 and C3 chemicals are most demanded, followed by C1, C8P, C6P, etc. Thus, replacement chemicals with large #C (e.g., #C>18) will have little demand.
- On a mol-weighted basis, there are 2.66 carbon atoms in a molecule on average.
- For #C≥6, chemicals with phenyl groups have much larger demand than those without.

To facilitate the understanding of insights from **Figure 6.1B**, the industrial production methods of chemicals in different #C categories are summarized below:

- C1 chemicals are produced from natural gas, which is relatively cheap.
- C2 and C3 chemicals are produced from either natural gas or petroleum.
- C4 chemicals are mainly produced from petroleum.
- C5 chemicals are byproducts from the production of other chemicals from petroleum and usually have higher costs.
- C6P chemicals are produced from benzene (or its derivatives), which is produced mainly through catalytic reforming.
- C6NP chemicals are produced from cyclohexane (or its derivatives), which are initially produced from benzene.
- C7P, C8P and C15P chemicals are produced mainly through the addition of carbons to benzene (or its derivatives).

- C18NP chemicals (octadecanoic acid and oleic acid) are bio-produced.

Key insights from **Figure 6.1B** are summarized as follows:

- Price of #C, i.e., difficulty of obtaining from fossil fuel feedstocks and thus potential advantage of replacement, on the mol-weighted average basis, is sequenced as $1 < 2 < 3 < 6P < 7P < 4 < 8P < 6NP < 5 < 15P$. The sequence on the MT-weighted average basis is $1 < 6P < 7P < 2 < 3 < 8P < 4 < 6NP < 15P < 5$, where chemicals with larger #C (e.g., 6P) have lower price than on the mol-weighted average basis because they have larger molecular weight. We use the mol-weighted average as the basis for further discussion.
- Although a larger #C renders a higher price in general, C4 and C5 chemicals are more expensive than C6P chemicals (produced from benzene) because benzene cost is relatively low.
- C6NP chemicals are more expensive than C6P chemicals because C6NP are produced from cyclohexane or its derivatives, which are initially produced from benzene, thus more expensive than C6P chemicals produced directly from benzene.

6.3.2 Number of oxygen atoms (#O)

The volume and price data for different #O are shown in **Figure 6.2**. Key insights from **Figure 6.2A** are summarized as follows:

- Chemicals with lower #O have higher demand, and there is little demand for chemicals with more than 4 oxygen atoms except citric acid (O7), which is bio-produced today.
- On a mol-weighted basis, there are 0.51 oxygen atoms in a molecule on average.

Key insights from **Figure 6.2B** are summarized as follows:

- Chemicals with higher #O generally have higher price (thus posing potential for replacement), because fossil fuels are just slightly oxidized, which facilitates the production of less oxidized chemicals.

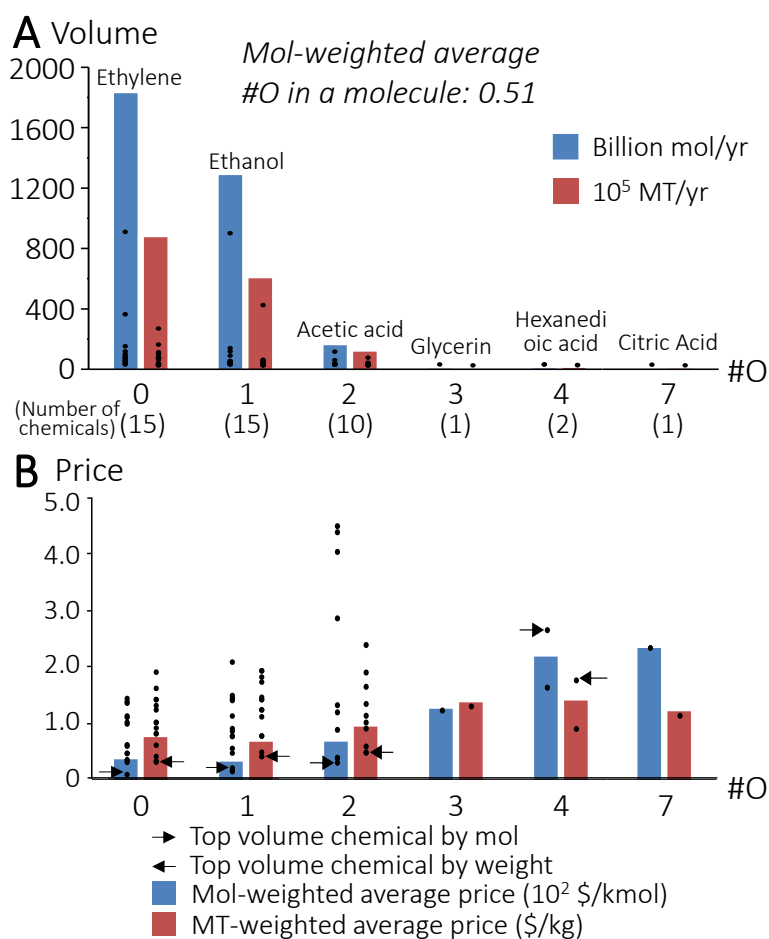


Figure 6.2. Results for #O. (A) Volume vs. #O; each blue or red bar represents the total volume (billion mol/year or 10^5 MT/year) of chemicals with the corresponding #O; number of chemicals with the corresponding #O are labeled at the bottom; top 1 volume chemical for each #O is labeled above the two bars; (B) price vs. #O; each blue or red bar represents the mol-weighted average price (10^2 \$/kmol) or MT-weighted average price (\$/kg) of chemicals with the corresponding #O; arrows represent top volume chemicals. Each black dot represents a chemical.

6.3.3 Number of (distinct) functional groups (#FG and #DFG)

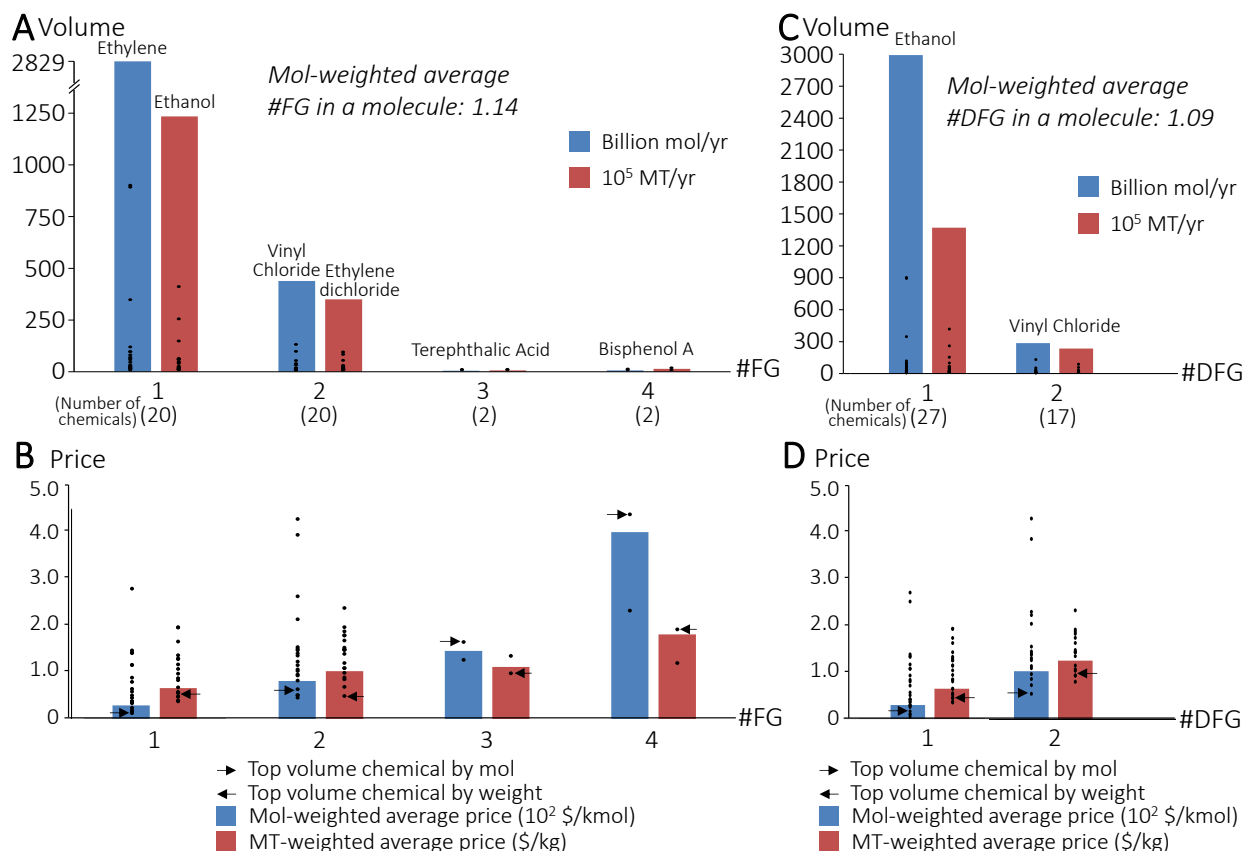


Figure 6.3. Results for #FG and #DFG. **(A)** Volume vs. #FG; each blue or red bar represents the total volume (billion mol/year or 10^5 MT/year) of chemicals with the corresponding #FG; number of chemicals with the corresponding #FG are labeled at the bottom; top 1 volume chemical for each #FG is labeled above the two bars, and only one is labeled if the top 1 volume chemical is the same for both bars; **(B)** price vs. #FG; each blue or red bar represents the mol-weighted average price (10^2 \$/kmol) or MT-weighted average price (\$/kg) of chemicals with the corresponding #FG; arrows represent top volume chemicals; **(C)** and **(D)** show results for #DFG, and the legends are the same with those for **(A)** and **(B)**, respectively. Each black dot represents a chemical.

The volume and price data for different #FG and #DFG are shown in **Figure 6.3**, and key insights are summarized as follows:

- Chemicals with lower #FG or #DFG have higher demand, and there is little demand for chemicals with more than 4 FGs or 2 DFGs.
- On a mol-weighted basis, there are 1.14 FGs or 1.09 DFGs in a molecule on average.

- Chemicals with lower #FG or #DFG have lower price because fossil fuel feedstocks have small number of FGs and DFGs.

6.3.4 Specific functional groups (FG)

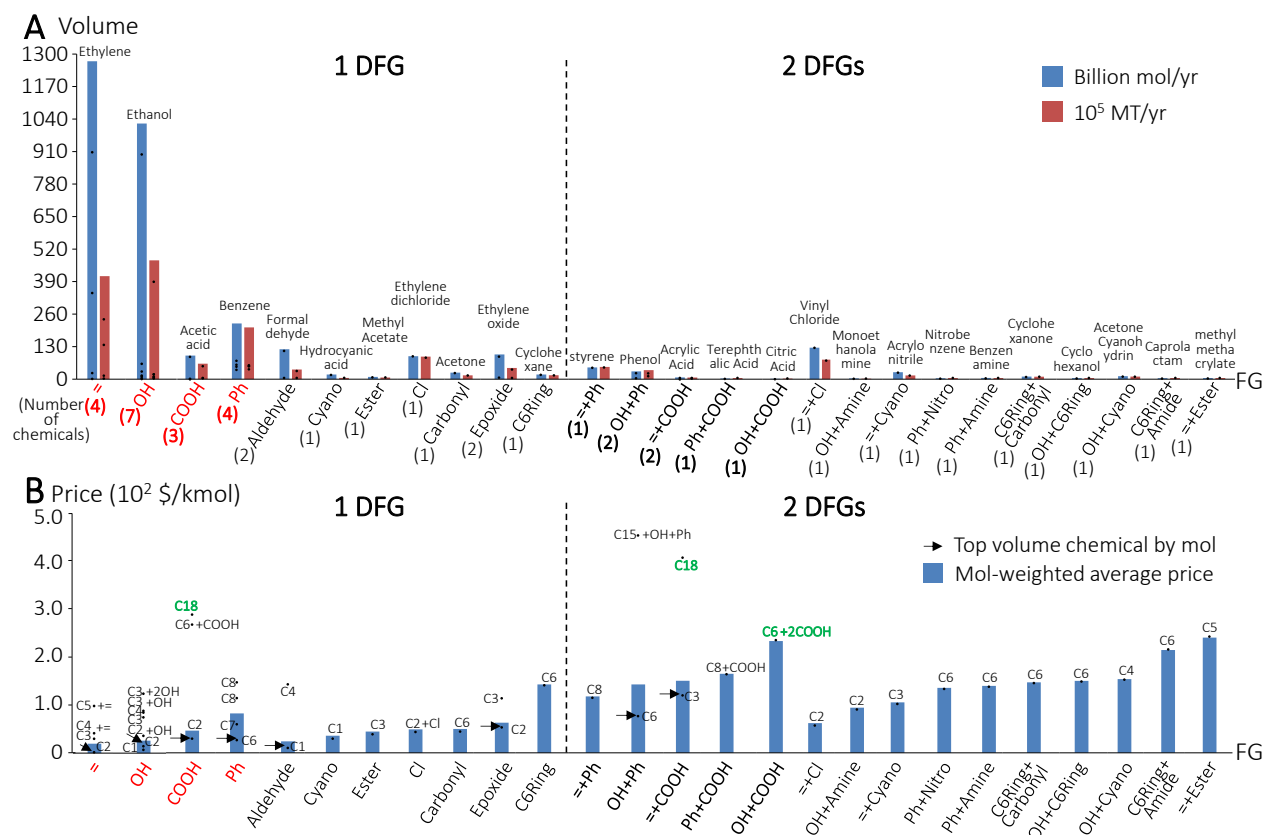


Figure 6.4. Results for specific FGs. **(A)** Volume vs. FGs; each blue or red bar represents the total volume (billion mol/year or 10⁵ MT/year) of chemicals with the corresponding FG (or a combination of FGs); number of chemicals with the corresponding FG are labeled at the bottom; top 1 volume chemical for each FG is labeled above the two bars; **(B)** price vs. specific FGs; each blue bar represents the mol-weighted average price (10² \$/kmol) of chemicals with the corresponding FG; arrows represent top volume chemicals; #C is marked next to each chemical, and if a chemical has more FGs than labeled on the x axis, then the additional FGs are also marked, e.g., the chemical marked “C3” and “+2OH” in the “OH” column denotes a C3 chemical that has 3 hydroxyl groups; chemicals that are currently bio-produced are marked bold in green. Each black dot represents a chemical; = denotes alkenyl; Cl denotes chloro; Ph denotes phenyl; COOH denotes carboxyl; OH denotes hydroxyl; C6Ring denotes non-aromatic ring with 6 carbon atoms; in the mono-DFG region (left), each attribute with more than or equal to three data points are marked bold in red and ranked based on the mol-weighted average price, from low to high (=, OH, COOH, Ph); in the bi-DFG region, FG1+FG2 means that a chemical has both FGs, and each attribute that corresponds to a combination of the aforementioned four mono-DFG attributes is marked bold.

The volume and price data for different specific FGs are shown in **Figure 6.4**. We draw insights based on attributes with at least three data points in the mono-DFG region (marked bold in red) and the combination of these attributes in the bi-DFG region (marked bold). Key insights from **Figure 6.4A** are summarized as follows:

- The most demanded FGs are alkenyl (=), hydroxyl (OH), phenyl (Ph) and carboxyl (COOH) groups.
- Demand for all combinations of these 4 FGs exists except alkenyl (=) combined with hydroxyl (OH).

Key insights from **Figure 6.4B** are summarized as follows:

- Price of FGs, on the mol-weighted average basis, is sequenced as alkenyl (=)<hydroxyl (OH)<carboxyl (COOH)<phenyl (Ph), which is consistent with the price of the combinations of these FGs in the bi-DFG region. The alkenyl (=)<hydroxyl (OH)<carboxyl (COOH) sequence can be explained from an organic chemistry point (increased level of oxidation). Note that phenyl groups have high prices partly because the corresponding chemicals have larger #C.

6.4 Discussions

6.4.1 Suitability of bio-production

In this section, we discuss characteristics of potentially profitable chemicals produced using microbial conversion of glucose.

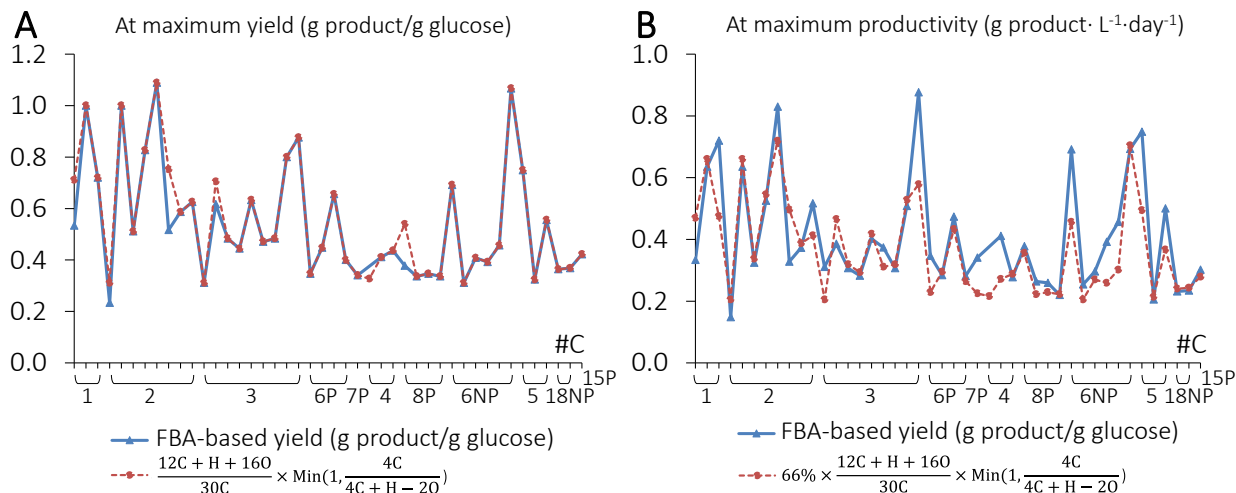


Figure 6.5. Microbial conversion yield (g product/g glucose) of chemicals. **(A)** Maximum yield vs. #C; blue curve represents the maximum yield estimated using flux balance analysis (FBA); red dashed curve represents the stoichiometric maximum yield using glucose as the carbon source, where C, H and O represent the number of carbon, hydrogen and oxygen atoms in a molecule, and MW (g/mol) is the molecular weight of the product; **(B)** yield at maximum productivity (g product · L medium⁻¹ · day⁻¹) vs. #C; blue curve represents the yield at maximum productivity estimated using flux balance analysis (FBA); red dashed curve represents the estimates based on stoichiometric yield using glucose as the carbon source.

Through microbial conversion, a carbon source like glucose is converted to a chemical product by a microbe. In **Chapter 5**, the yields of chemicals produced through microbial conversion of glucose from *E.coli* and *S. cerevisiae* are estimated using Flux Balance Analysis (FBA), at both maximum yield (g product/g glucose fed) and maximum productivity (g product · L medium⁻¹ · day⁻¹) conditions. It was also found that maximum yield represents the production condition with minimum production cost (including microbial conversion and downstream separation), and maximum productivity represents a fairly good condition with low cost. Therefore, we use the FBA-based method to calculate yields of all the chemicals studied here, as shown in **Figure 6.5A** for maximum yield and **Figure 6.5B** for maximum productivity, respectively. It can be seen that there is no clear correlation between #C and yield (thus production cost). We further compare the FBA-based yield with the stoichiometric maximum yield (g product/g glucose), which can be expressed as $\frac{12C+H+16O}{30C} \times \frac{4C}{4C+H-20}$ in most cases (calculated from r6.1), where C, H and O represent the

number of carbon, hydrogen and oxygen atoms in a molecule of the chemical product; $\frac{4C}{4C+H-2O}$ represent the molar carbon efficiency (mol of carbon in product/mol of carbon in glucose); $12C+H+16O$ represents the molecular weight of the product.

$$\frac{x(4C + H - 2O)}{24} C_6H_{12}O_6 = xC_C H_H O_O + \frac{x(4C - H - 2O)}{4} H_2O + \frac{x(H - 2O)}{4} CO_2 \quad (r6.1)$$

$$CC_6H_{12}O_6 + \frac{6O - 3H}{2} O_2 = 6C_C H_H O_O + (6C - 3H)H_2O \quad (r6.2)$$

Note that for citric acid and terephthalic acid, their molar carbon yields, if calculated using $\frac{4C}{4C+H-2O}$ based on r6.1, will be greater than 100% because CO_2 will be considered as feedstock in order to balance r6.1. Instead, we consider r6.2 for such products, where the yield can be expressed as $\frac{12C+H+16O}{30C}$. Therefore, the stoichiometric maximum yield can be expressed as $\frac{12C+H+16O}{30C} \times \text{Min}(1, \frac{4C}{4C+H-2O})$, which is close to the maximum yield calculated based on FBA, as shown in **Figure 6.5A**. Also, at maximum productivity, $66\% \times \frac{12C+H+16O}{30C} \times \text{Min}(1, \frac{4C}{4C+H-2O})$ can be regarded as the approximate yield, as shown in **Figure 6.5B**.

- Therefore, in most cases, the stoichiometric yield (g/g) $\frac{12C+H+16O}{30C} \times \frac{4C}{4C+H-2O} = \frac{24+2(\frac{H}{C})+32(\frac{O}{C})}{60+15(\frac{H}{C})-30(\frac{O}{C})}$ can be regarded as an indicator for the suitability of microbial conversion:

the larger the better. Thus, chemicals that are highly oxidized (with high O/C ratio such as those with OH and COOH groups) and/or less hydrogenated (low H/C ratio such as those with alkenyl, ring and other “hydrogen replacing” groups) are likely suitable targets.

As an example, note that citric acid ($C_6H_8O_7$, with a high O/C ratio and low H/C ratio, and indicator value of 1.42; marked green in **Figure 6.1B**) is currently produced through microbial conversion of glucose, and it is relatively cheap compared with other 6-NP chemicals with even less complicated

structures such as hexanedioic acid ($C_6H_{10}O_4$, with an indicator value of 0.75), which could be a sign

that $\frac{24+2(\frac{H}{C})+32(\frac{O}{C})}{60+15(\frac{H}{C})-30(\frac{O}{C})}$ is a reasonable indicator for the suitability of microbial conversion.

6.4.2 Profitability of bio-production

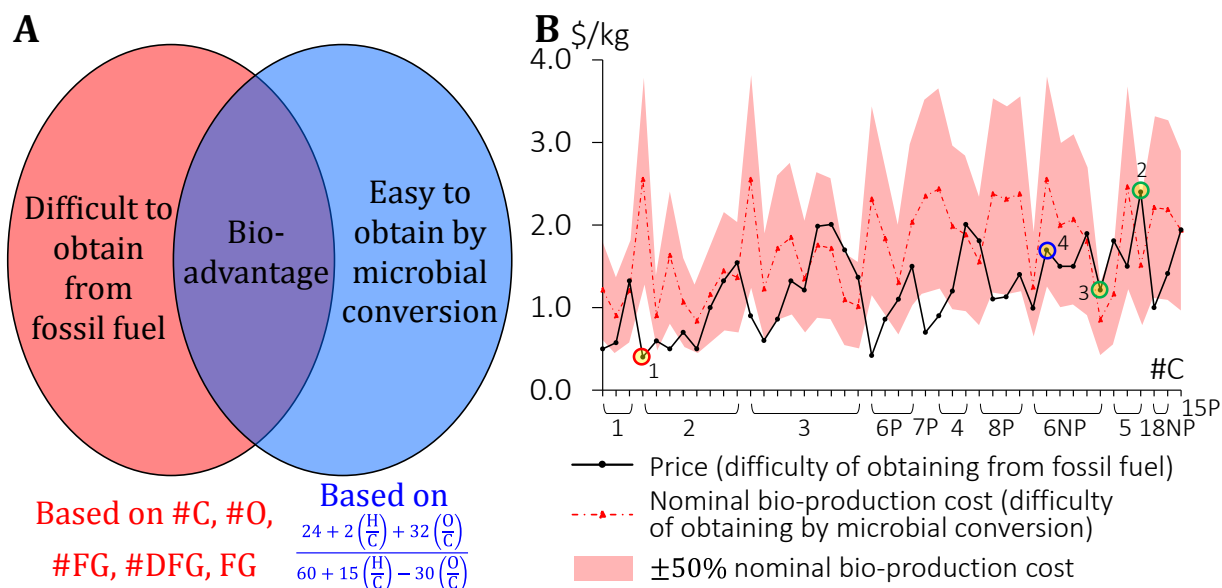


Figure 6.6. Analysis of bio-production profitability. **(A)** General conceptual analysis framework; **(B)** Current market prices and bio-production costs vs. #C, where #C is ordered such that the mol-weighted average prices of chemicals increase from left to right.

We present a conceptual analysis method for the identification of potentially profitable bio-based replacement chemicals in **Figure 6.6A**. Specifically, potentially profitable replacement chemicals should be difficult to obtain from fossil fuel feedstocks (with the corresponding characteristics discussed in **Section 6.3**) and relatively easy to obtain by microbial conversion (highly oxidized and less hydrogenated as discussed in **Section 6.4.1**). Profitable targets can be identified at the intersection.

As further demonstration, we present the current market prices and bio-production cost estimates (including both bio-conversion and downstream separation) of all the 44 chemicals against #C as an example in **Figure 6.6B**. Specifically, we calculate the *nominal* cost based on the estimated

maximum yield using the approach discussed in **Chapter 5**, assuming the product is intracellular and soluble in water, which corresponds to a medium separation cost (see **Figure 5.3**). Also, the red shaded band represents $\pm 50\%$ of the nominal cost to account for uncertainty. A few key observations are as follows:

- Some chemicals have low market prices (with #C characteristic relatively easy to obtain from fossil fuel in the example) and high bio-production costs (with low $\frac{24+2(\frac{H}{C})+32(\frac{O}{C})}{60+15(\frac{H}{C})-30(\frac{O}{C})}$ values), e.g., Entry 1 marked by a red circle. Such chemicals, especially if they are below the $\pm 50\%$ band, are unlikely to be profitable.
- Chemicals above or close to the upper boundary of the band, which have either (1) high prices and low bio-production costs (such as Entry 2), or (2) relatively low prices and very low bio-production costs (such as Entry 3), are likely to be profitable.
- Profitability of chemicals in the middle range of the band (such as Entry 4) is uncertain.

It can be seen that fossil fuel-based production and bio-production of chemicals could occupy different sectors of the market. Their natural suitability regions for producing chemicals are different (even opposite).

6.4.3 Additional remarks

The analysis in this work is based on microscopic characteristics of chemicals such as FGs. Based on group contribution theory, the macroscopic physical properties of a chemical is a function of FGs. Therefore, insights from the current analysis is applicable to most cases. However, in rare cases, if a specific chemical has microscopic attributes identified as undesirable in this work but has properties similar to chemicals of top volumes, then it can still be a promising target. For analysis based on macroscopic properties, the readers can consider characteristics such as density, boiling

pints, melting points, heat of vaporization, partition coefficients, viscosity, surface tension, thermal conductivity and solubility [222].

Also note that although most chemicals that could be profitable for bio-production appear to have very low volume in this work, they are actually all HPV chemicals. A chemical with even the smallest volume in this study has a volume of 200,000 MT/year, which may still be worth investigating.

In addition, we do not claim absolute completeness with respect to the identification of promising replacement chemicals in this work. Instead, we seek to identify general patterns and insights. Other considerations for successful replacement include efficient supply chain, cost reduction of bio-conversion and separation, and microbial strain engineering to improve yield and productivity. Also, if polymers are target replacement chemicals, then linkages between monomers and the specific three-dimensional structure of the polymers are also important.

6.5 Conclusions

We developed a framework to identify the characteristics of promising replacement chemicals. By studying the market volume of major organic chemicals, we identified the demanded characteristics. Further, by studying the market price, we identified the characteristics that are difficult to obtain from fossil fuel feedstocks. Finally, we identified the characteristics that could lead to profitable bio-production.

Specifically, chemicals with one or more of the following characteristics are likely to have higher demand: (1) with two or three carbon atoms; (2) with smaller number of oxygen atoms; (3) with fewer functional groups; and (4) with alkenyl, hydroxyl, phenyl or carboxyl groups. However,

chemicals with >18 carbon atoms, >4 oxygen atoms, >4 functional groups, or >2 distinct functional groups, will likely have little demand.

Chemicals with one or more of the following characteristics are expensive to obtain from fossil fuel feedstocks: (1) with large number of carbon atoms, oxygen atoms or functional groups; (2) with six carbon atoms but without a phenyl group, or with five carbon atoms; and (3) with carboxyl groups.

Chemicals that are suitable for bio-conversion will likely have a large value in $\frac{24+2(\frac{H}{C})+32(\frac{O}{C})}{60+15(\frac{H}{C})-30(\frac{O}{C})}$, where

C, H and O represent the number of carbon, hydrogen and oxygen atoms in a molecule of the chemical product. These chemicals are highly oxidized and lightly hydrogenated.

Chapter 7

7 Conclusions and future work

We developed optimization-based synthesis and assessment frameworks for bio-based chemicals production to provide guidance on the efficient synthesis of bio-production processes and the identification of promising products.

Specifically, we first developed a method for the efficient representation, generation, and modeling of superstructures for process synthesis, which features modularity. This method was then used to develop a framework for the synthesis of downstream bio-separation processes. A general superstructure was generated to account for all classes of products, and a superstructure reduction method was developed to solve specific cases. The bio-separation framework was further used to study two major categories of products: extracellular and intracellular. The influence of a combination of key parameters, such as titer and technology performance, on technology selection and cost, was presented.

Next, we developed a framework for the identification of techno-economically promising products, including the generation of a candidate pool of 209 bio-producible chemicals with high production volume, and the development of three screening criteria based on a product's profit margin, market volume and market size. Thus, we identified 32 products currently sold on the market as promising targets for bio-production if maximum yields can be achieved, and 22 products if maximum productivities can be achieved. We further identified the characteristics of promising replacement products in terms of the number of carbon atoms, oxygen atoms, and functional groups, as well as the specific function groups, in a molecule.

Detailed conclusions can be found in the Conclusions section of each chapter.

Potential future directions of this work are as follows:

1. Superstructure-based bio-separation process synthesis with rigorous unit models. Rigorous unit models involving detailed kinetics and unconventional units can provide significant insights for the development of production processes. However, most current research focuses on either rigorous unit models in simple networks or simple unit models in complex networks (e.g., in **Chapter 3**). The development of unit models involving more detailed kinetics for both conventional and unconventional units in a bio-separation superstructure can be valuable. To gain such details, property estimation modules, e.g., used to predict solvent properties, and lab experiments would be essential. Since a complicated non-convex MINLP model is likely to be formulated for the superstructure, an improvement on the solution tractability would be critical. Therefore, efficient solution techniques such as tightening, reformulation and solution algorithms would need to be applied or developed for this purpose.

2. A superstructure approach to the holistic assessment of promising products. As mentioned in **Chapter 1**, microbial production of chemicals has unique advantages, and the product identification framework discussed in **Chapters 5** and **6** can help identify techno-economically promising products. Therefore, a holistic assessment of the bio-production of such products could generate significant impact. Specifically, we could develop superstructure optimization approaches to assess the different bio-conversion and separation technologies, microbial strains, and feedstocks available to produce the promising products, thus identifying the combination (production strategy) that optimizes a specific objective (such as cost minimization). Key design parameters that can potentially improve the objective would also be identified. In addition, the rigorous unit models, integration with experimental results and solution techniques discussed in Point 1 could also be utilized here.

Appendix

A1 Explanations to Chapter 2

A1.1 Specific formulations of thermodynamic relations

We specify the general f^H , f^S , $\hat{f}^H$ and $\hat{f}^S$ functions (see **Equation 2.17** for example) used in the element models.

First, we discuss the specific formulation of enthalpy and entropy expressions. The enthalpy and entropy of a system at a given state (with temperature T and pressure P) can be calculated by constructing a route where the system changes from the standard state (T^0, P^0) to the given state, as shown in **Equations A1.1** and **A1.2**. Here we consider an isobaric cooling/heating followed by isothermal compression/expansion. If there is a phase change from the standard state to the given state, we assume that the phase change occurs during heating/cooling, not during compression/expansion.

$$h(T, P) = h^0 + \int_{T^0}^{T^m} c_1 dT + \Delta h_m + \int_{T^m}^T c_2 dT + X \quad (\text{A1.1})$$

$$s(T, P) = s^0 + \int_{T^0}^{T^m} c_1 \frac{dT}{T} + \frac{\Delta h_m}{T^m} + \int_{T^m}^T c_2 \frac{dT}{T} + Y \quad (\text{A1.2})$$

where $h(T, P)$ and $s(T, P)$ are specific enthalpy and entropy at temperature T and pressure P ; h^0 and s^0 are standard specific enthalpy and entropy; T^0 and P^0 are standard temperature and pressure; c_1 is specific heat capacity of the initial phase; T^m is phase change temperature; $\int_{T^0}^{T^m} c_1 dT$ and $\int_{T^0}^{T^m} c_1 \frac{dT}{T}$ denote specific enthalpy and entropy changes due to temperature change in the initial phase; Δh_m and $\frac{\Delta h_m}{T^m}$ are specific enthalpy and entropy changes due to phase change; c_2 is specific

heat capacity of the final phase; $\int_{T_m}^T c_2 dT$ and $\int_{T_m}^T c_2 \frac{dT}{T}$ denote specific enthalpy and entropy changes due to temperature change in the final phase; X and Y denote specific enthalpy and entropy changes due to pressure change. In general,

$$X = \int_{P^0}^P \left[-T \left(\frac{\partial V}{\partial T} \right)_P + V \right] dP, Y = - \int_{P^0}^P \frac{V dP}{T} \quad (\text{A1.3})$$

For non-ideal gas $P(V - a) = RT$,

$$X = a(P - P^0), Y = -R \ln \frac{P}{P^0} - \frac{a(P - P^0)}{T} \quad (\text{A1.4})$$

For ideal gas $PV = RT$,

$$X \approx 0, Y = -R \ln \frac{P}{P^0} \quad (\text{A1.5})$$

For liquid and solid,

$$X \approx 0, Y \approx 0 \quad (\text{A1.6})$$

With different simplified assumptions, the specific enthalpy expression then becomes the following.

$$h(T, P) = h^0 + c_1(T - T^0) + \Delta h_m + X \text{ (constant } c, \text{ constant } \Delta h_m) \quad (\text{A1.7})$$

$$h(T, P) = h^0 + c_1(T - T^0) + X \text{ (constant } c, \text{ no phase change)} \quad (\text{A1.8})$$

$$h(T, P) = h^0 + c_1(T - T^0) + a(P - P^0) \text{ (constant } c, \text{ no phase change, non-ideal gas)} \quad (\text{A1.9})$$

$$h(T, P) = h^0 + c_1(T - T^0) \text{ (constant } c, \text{ no phase change, ideal gas, liquid or solid)} \quad (\text{A1.10})$$

Similarly, simplified expressions for specific entropy can also be obtained. With the above equations, the general $f^H, f^S, \hat{f}^H$ and $\hat{f}^S$ expressions can be specified based on any assumptions we make. For example, based on the assumption made in **Equation A1.10**, the f^H and $\hat{f}^H$ expressions are specified as follows.

$$H_s^{k1} = f^H([F_{s,c}]_{c \in \mathcal{C}}, T_s^{k1}, P_s^{k1}) = \sum_{c \in \mathcal{C}} h^0_{F_{s,c}} + \sum_{c \in \mathcal{C}} c_c F_{s,c} (T_s^{k1} - T^0) \quad (\text{A1.11})$$

$$T_s^{k1} = \hat{f}^H([F_{s,c}]_{c \in \mathcal{C}}, H_s^{k1}, P_s^{k1}) = \frac{H_s^{k1} - \sum_{c \in \mathcal{C}} h^0_{F_{s,c}}}{\sum_{c \in \mathcal{C}} c_c F_{s,c}} + T^0 \quad (\text{A1.12})$$

A1.2 Stream model involving expansion valves

For conditioning streams using expansion valves, **Equations 2.15-2.23** become the following.

$$H_s^{k2} = H_s^{k3} \quad (\text{A1.13})$$

$$P_s^{k2} = P_s^{k1} \quad (\text{A1.14})$$

$$H_s^k = f^H([F_{s,c}]_{c \in \mathcal{C}}, T_s^k, P_s^k), \forall k \in \{k1, k3\} \quad (\text{A1.15})$$

$$T_s^{k2} = \hat{f}^H([F_{s,c}]_{c \in \mathcal{C}}, H_s^{k2}, P_s^{k2}) \quad (\text{A1.16})$$

$$Q_s = H_s^{k2} - H_s^{k1}, \quad W_s = 0 \quad (\text{A1.17})$$

$$C_s^{\text{acc}} = f^{\text{acc,HC}}(Q_s, T_s^{k1}, T_s^{k2}, P_s^{k1}) \quad (\text{A1.18})$$

$$C_s^{\text{oc}} = f^{\text{oc,HC}}(Q_s, T_s^{k1}, T_s^{k2}) \quad (\text{A1.19})$$

A1.3 Non-conditioning stream pressure relation

We use an example, as shown in **Figure A1.1**, to illustrate the necessity of **Equation 2.29** in the non-conditioning stream model, i.e., $-\Delta p_s^{\text{up}}(1 - Y_s) \leq P_s^{k1} - P_s^{k3} \leq \Delta p_s^{\text{up}}(1 - Y_s)$: if the stream is activated ($Y_s = 1$), then $P_s^{k3} = P_s^{k1}$; otherwise ($Y_s = 0$), the equality is void. Supposing only stream p1-p4 is activated, then $P_{p1} = P_{p1-p4}^{k1} = P_{p1-p4}^{k3} = P_{p4} = 50$ bar according to the activated port and stream models. If the pressure relation is formulated simply as $P_s^{k1} = P_s^{k3}$ (**Figure A1.1-A**), then $P_{p2-p4}^{k3} = 1$ bar, and $P_{p3-p4}^{k3} = 10$ bar. However, these values contradict with **Equation 2.8** the

isobaric condition, leading to model infeasibility; On the contrary, if the bigM formulation $-\Delta p_s^{\text{up}}(1 - Y_s) \leq P_s^{k1} - P_s^{k3} \leq \Delta p_s^{\text{up}}(1 - Y_s)$ is adopted (**Figure A1.1-B**), then no contradiction is incurred because P_{p2-p4}^{k3} and P_{p3-p4}^{k3} are free to take any values, and thus the model assigns 50 bar to both variables to satisfy the isobaric condition. Also note that, the bigM formulation is unnecessary for the temperature relation (see **Equation 2.30**), because isothermal condition $T_p = T_s^{k3}$ is not required at the inlet ports. Instead, T_p is calculated using enthalpy **Equation 2.9**.

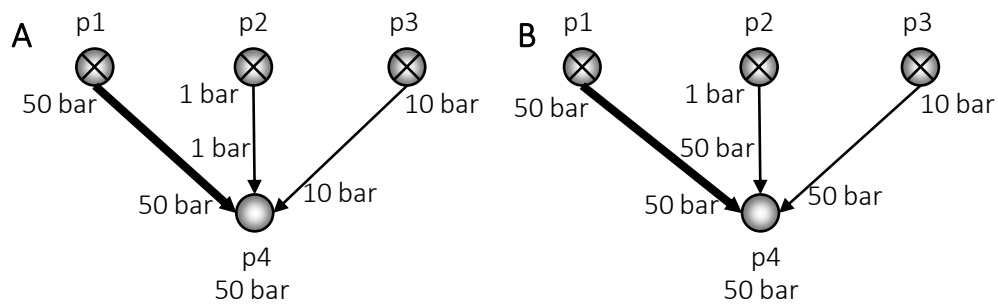


Figure A1.1. Explanation to the non-conditioning stream pressure relation **Equation 2.29**. Pressures at ports p1, p2 and p3 are known to be 50 bar, 1 bar and 10 bar, respectively. All the streams are non-conditioning streams. (A) $P_s^{k1} = P_s^{k3}$ formulation; (B) $-\Delta p_s^{\text{up}}(1 - Y_s) \leq P_s^{k1} - P_s^{k3} \leq \Delta p_s^{\text{up}}(1 - Y_s)$ formulation.

A2 Bio-separation model

This section contains models related to **Chapters 2-4**.

Component set specification

We first present a method to specify the component set (**C**) for the reduced superstructure in **Chapter 3**. In general, if the product is EX, then the initial product stream contains the product, water, cells, impurities that are soluble in water, and impurities that are insoluble in water. If the product is IN, then the initial product stream contains water and cells (see columns $\mathbf{C}^{\text{Basic}}$ in **Table A2.1**). We refer to the set composed of these elements as the “basic component set”. The existence of flocculation (Flc), cell disruption (Cdr), precipitation (Prc), extraction (Ext), solubilization (Slb), crystallization (Cry), and differential digestion (Ddg) technologies in the reduced superstructure will lead to the introduction of new elements (see **Table A2.1**). For example, in cell disruption, the cells are converted to product (“prodt”), insoluble debris (“debris”) and other soluble impurities (“solimp”). Thus, if cell disruption is included in the reduced superstructure, then “prodt”, “solimp” and “debris” should be added to the component set. Mathematically, it is expressed as $\mathbf{C} = \mathbf{C}^{\text{Basic}} \cup \mathbf{C}^{\text{Flc}} \cup \mathbf{C}^{\text{Cdr}} \cup \dots$, where $\mathbf{C}^{\text{Flc}}$, $\mathbf{C}^{\text{Cdr}}$... represent sets of components specific for technologies Flt, Cdr, etc. Note that the components in **Table A2.1** can be further modified, e.g. “solimp” can be replaced by “proteins”, “pigments”, etc.

Table A2.1. Component set. Prodt=product, solimp=impurities soluble in water, insolimp=impurities insoluble in water, floct=flocculent, cellfc=flocculated cells, debrisfc= flocculated debris, prodtprc=precipitated product, prcpt=precipitant, extsolv=solvent used for extraction, prodtsol=solubilized product, solbr=solubilizing agent, prodctcrs=crystalized product, digr=digesting agent for the cell debris, digdebris=digested debris.

	C^{Basic}	C^{Flc}	C^{Cdr}	C^{Prc}	C^{Ext}	C^{Slb}	C^{Crs}	C^{Ddg}
EX	prodt, water, cell, solimp, insolimp	floct, cellfc	NA	prodtprc, prcpt	extsolv	prodtsol, solbr	prodctcrs	NA
IN	water, cell	floct, cellfc, debrisfc	prodt, solimp, debris	prodtprc, prcpt	extsolv	prodtsol, solbr	prodctcrs	digr, digdebris

Port numbering

Table A2.2. Unit port numbering.

unit	port	port	port	unit	port	port	port
type <i>ut</i>	type <i>pt</i>	No. <i>pn</i>	usage	type <i>ut</i>	type <i>pt</i>	No. <i>pn</i>	usage
Cdr	in	1	feed stream	Sdm/Cnt	in	1	feed stream
	out	1	disrupted cells			1	light phase
Blc	in	1	feed stream		out	2	water
	out	1	bleached stream			3	heavy phase
Flt	in	1	feed stream	Flc	in	1	feed stream
		1	insoluble			2	flocculent
	out	2	compoents water		out	1	flocs
Mbr/Ftt	in	1	feed stream	Slb	in	1	feed stream
		1	permeate			2	solvent
	out	2	concentrate		out	1	solubilized product
Dst	in	1	feed stream	Ddg	in	1	feed stream
		1	top			2	digesting agent
	out	2	bottom		out	1	digested debris
Ads/Chr	in	1	feed stream	Prc	in	1	feed stream
		1	non-product			2	precipitant
	out	2	product		out	1	precipitated product
Crs	in	1	feed stream	Ext	in	1	feed stream
		1	vapor/liquid			2	solvent
	out	2	crystal			1	extract
Dry	in	1	feed stream		out	2	raffinate
		1	vapor	Src	out	1	initial product stream
	out	2	solid	Snk	in	1	final product stream

The specific numbering of ports is shown in **Table A2.2**.

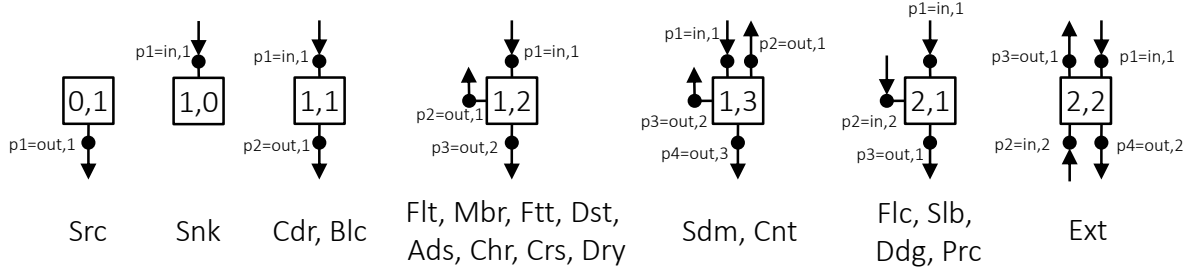


Figure A2.1. New labels for unit ports. The boxes represent technologies. The “a,b” label in each box indicates the number of inlet ports (“a”) and outlet ports (“b”) of the corresponding technology. The corresponding technologies with the specified number of ports are labeled below the boxes. “pn=in/out,m” labels around the ports mean that the inlet/outlet port with identification number m is renamed as “pn” for the corresponding unit type. For example, the third box on the left has one inlet port (“in,1”) and one outlet port (“out,1”), which corresponds to Cdr and Blc. The ports are renamed “p1” and “p2”, respectively. If crystallization is performed by cooling instead of evaporation, then Crs has one outlet port.

Further, to facilitate reading of the unit models below, new port labels (p1, p2...) are assigned, as shown in **Figure A2.1**.

Indices and sets

$ut \in \mathbf{UT}$ = unit types

$un \in \mathbf{UN}$ = unit numbers

$u \in \mathbf{U} \subset \mathbf{UT} \times \mathbf{UN}$ = units

$pt \in \mathbf{PT}$ = port types, i.e. in/out

$pn \in \mathbf{PN}$ = port numbers

$p \in \mathbf{P} \subset \mathbf{U} \times \mathbf{PT} \times \mathbf{PN}$ = ports

$s \in \mathbf{S} \subset \mathbf{P} \times \mathbf{P}$ = streams defined by two ports

$c \in \mathbf{C}$ = components

$l \in \mathbf{L}$ = utility types, i.e. cooling/heating/electricity/labor

$k \in \mathbf{K} = \{k1, k2, k3\}$ = states of a conditioning stream

$\mathbf{C}_p^{\text{FI}}/\mathbf{C}_p^{\text{MI}}/\mathbf{C}_{p'}^{\text{FO}}/\mathbf{C}_{p'}^{\text{MO}}$ = feasible/minimal components for inlet port p /outlet port p'

$\mathbf{C}_p^{\text{M}}/\mathbf{C}_p^{\text{F}}$ = minimal/feasible components for port p

$r \in \mathbf{R}$ = reactions

Subsets

$\mathbf{U}_{ut}/\mathbf{U}_s$ = units of type ut / connected to stream s

$\mathbf{P}_u/\mathbf{P}_{pt}/\mathbf{P}_{u,pt}$ = ports of unit u / type pt / unit u and type pt

$\mathbf{P}_{ut,pt}$ = ports of unit type ut and port type pt

$\mathbf{P}_p$ = outlet ports that are connected with inlet port p

$\mathbf{S}_{p',p}$ = streams defined by port p' and port p

$\mathbf{S}_{pt}$ = streams connected to type pt ports

$\mathbf{S}_p$ = streams connected to port p

$\mathbf{S}^{\text{C}}/\mathbf{S}^{\text{NC}}$ = conditioning/non-conditioning streams

$\mathbf{C}_u^{\text{LK}}/\mathbf{C}_u^{\text{HK}}$ = light key/heavy key in unit $u \in \mathbf{U}_{\text{Dst}}$

$\mathbf{C}_u^{\text{NSLHV}}$ = insoluble heavy components in unit $u \in \mathbf{U}_{\text{Sdm}} \cup \mathbf{U}_{\text{Cnt}}$

$\mathbf{C}_u^{\text{NSLLT}}$ = insoluble light components in unit $u \in \mathbf{U}_{\text{Sdm}} \cup \mathbf{U}_{\text{Cnt}}$

$\mathbf{C}_u^{\text{NSL}}$ = insoluble components in unit $u \in \mathbf{U}_{\text{Flt}}$

$\mathbf{C}_u^{\text{SOLID}}$ = insoluble solid components in unit $u \in \mathbf{U}_{\text{Dst}}$

$\mathbf{C}_u^{\text{DIS}}$ = components dissolved in water in unit $u \in \mathbf{U}_{\text{Sdm}} \cup \mathbf{U}_{\text{Cnt}} \cup \mathbf{U}_{\text{Flt}}$

$$\mathbf{C}_u^{\text{SOL}}/\mathbf{C}_u^{\text{SOLV}} = \text{solute/solvent components in unit } u \in \mathbf{U}_{\text{Ext}}$$

$$\mathbf{C}^{\text{prodt}} = \text{components that are considered as the final product}$$

Note that $\mathbf{C}_u^{\text{LK}}$, $\mathbf{C}_u^{\text{HK}}$ and $\mathbf{C}_u^{\text{SOLV}}$ each have only one element per unit u .

General parameters

$$\rho_c = \text{density [kg/m}^3\text{] of component } c$$

$$h_c^{\text{vap}} = \text{latent heat at boiling point [KWh/kg] of component } c$$

$$\tau_u = \text{residence time [hr] of unit } u$$

$$\beta_{u,c} = \text{fraction of component } c \text{ converted in pre-treatment unit } u$$

General variables

$$F_s/F_p = \text{mass flow rate [kg/hr] of stream } s/\text{port } p$$

$$F_{s,c}/F_{p,c} = \text{mass flow rate [kg/hr] of component } c \text{ in stream } s/\text{port } p$$

$$X_{p,c} = \text{mass fraction [wt\%] of component } c \text{ in port } p$$

$$A_u = \text{size of unit } u$$

Ports & Streams

We use the following variables:

$$F_s/F_p = \text{mass flow rate [kg/hr] of stream } s/\text{in port } p$$

$$F_{s,c}/F_{p,c} = \text{mass flow rate [kg/hr] of component } c \text{ in stream } s/\text{port } p$$

$$X_{p,c} = \text{mass fraction [wt\%] of component } c \text{ in port } p$$

Mass balance in ports:

$$\sum_{s \in \mathcal{S}_p} F_{s,c} = F_{p,c}, \quad p, c \quad (\text{A2.1})$$

Component & total flow relation in ports:

$$\sum_c F_{p,c} = F_p, \quad p \quad (\text{A2.2})$$

Mass fraction definition in ports:

$$F_{p,c} = F_p X_{p,c}, \quad p, c = c_1, \dots, c_{|C|-1} \quad (\text{A2.3})$$

Mass fractions summation to one:

$$\sum_c X_{p,c} = 1, \quad p \quad (\text{A2.4})$$

Note that **Equations A2.3** and **A2.4** are only formulated for ports that have $X_{p,c}$ variables in the corresponding unit models. For example, $X_{p2,c}$ (mass fraction of component c in the permeate port $p2$) is used in the membrane (Mbr) & filtration (Ftt) models to define rejection coefficient, while $X_{p1,c}$ (mass fraction of component c in the feed port $p1$) does not appear in any equations. Thus, **Equations A2.3** and **A2.4** are formulated for port $p2$, but not formulated for port $p1$.

Component & total flow relation in streams:

$$\sum_c F_{s,c} = F_s, \quad s \quad (\text{A2.5})$$

Logic constraints

For logic constraints, we use the following parameters and variables:

$$f^{\text{up}} = \text{global flow rate upper bound [kg/hr]}$$

$$Y_u/Y_s = 1 \text{ if unit } u/\text{stream } s \text{ is activated, otherwise } 0$$

Single-stream outlet port assumption:

$$Y_u = \sum_{s \in \mathbf{S}_p} Y_s, \quad u, p \in \mathbf{P}_{u,\text{out}} \quad (\text{A2.6})$$

General logic constraints:

$$Y_u \leq \sum_{s \in \mathbf{S}_p} Y_s, \quad u, p \in \mathbf{P}_{u,\text{in}} \quad (\text{A2.7})$$

$$Y_s \leq Y_u, \quad s \in \mathbf{S}_{\text{in}}, u \in \mathbf{U}_s$$

Zero-flow enforcement for deactivated streams:

$$F_{sc} \leq f^{\text{up}} Y_s, \quad c, s \in \mathbf{S} \quad (\text{A2.8})$$

Conditioning streams

For $s \in \mathbf{S}^c$, we use the following parameters:

$$\rho_c = \text{density [kg/m}^3\text{] of component } c$$

$$\mu_c = \text{specific heat capacity [KWh/(kg}\cdot\text{K)] of component } c$$

$$\sigma^{\text{kw}} = \text{Pa}\cdot\text{m}^3/\text{hr to KW conversion factor}$$

$$\Delta t_s^{\text{up}} = \text{upper bound of } |T_{p'} - T_p| \text{ [K]}$$

$$\Delta p_s^{\text{up}} = \text{upper bound of } |P_{p'} - P_p| \text{ [Pa]}$$

$$w_s^{\text{up}} = \text{upper bound of } W_s \text{ [KW]}$$

and the following variables:

$$P_p/T_p = \text{pressure [Pa]/temperature [K] of port } p$$

$$Z_s^P = 1 \text{ if pressure increases in stream } s, \text{ otherwise } 0$$

Q_s = temperature conditioning duty [KW] of stream s

Q_s^H/Q_s^C = heating/cooling duty [KW] of stream s

W_s = pressure conditioning duty [KW] of stream s

Temperature conditioning duty:

$$Q_s = Q_s^H + Q_s^C \quad (\text{A2.9})$$

$$Q_s^H - Q_s^C = (T_{p_2} - T_{p_1}) \sum_c \mu_c F_{s,c} \quad (\text{A2.10})$$

Pressure conditioning duty:

$$\sigma^{\text{kw}}(P_{p_2} - P_{p_1}) \sum_c \frac{F_{s,c}}{\rho^c} - w_s^{\text{up}}(1 - Z_s^P) \leq W_s \leq \sigma^{\text{kw}}(P_{p_2} - P_{p_1}) \sum_c \frac{F_{s,c}}{\rho^c} + w_s^{\text{up}}(1 - Z_s^P) \quad (\text{A2.11})$$

$$0 \leq W_s \leq w_s^{\text{up}} Z_s^P \quad (\text{A2.12})$$

Pressure change big M constraint:

$$\Delta p_s^{\text{up}}(Z_s^P - 1) \leq P_{p_2} - P_{p_1} \leq \Delta p_s^{\text{up}} Z_s^P \quad (\text{A2.13})$$

Here we express pumping work/heat exchange duties in terms of the stream flow rate and the pressure/temperature change across the stream for simplification. Moreover, since all the ports in the superstructure (see **Section 2.7**) are single-stream ports, $T_p = T_s^{\text{k3}}$ and $P_p = P_s^{\text{k3}}$ hold for all inlet port p and incoming stream $s \in \mathbf{S}_p$; $T_{p'} = T_s^{\text{k1}}$ and $P_{p'} = P_s^{\text{k1}}$ hold for all outlet port p' and outgoing stream $s \in \mathbf{S}_{p'}$. Therefore, the use of k1-k3 state variables becomes unnecessary, and thus they are replaced with port variables. Since the process streams are liquid, pumps are used to change pressures. We use valves as the pressure decrease conditioning units, and thus the corresponding conditioning duty is zero. **Equations A2.9-A2.12** enforce cooler/heater (and valve/pumping) mode when temperature (and pressure) decreases/increases.

Non-conditioning streams

The model for $s \in \mathbf{S}^{\text{NC}}$ is as follows.

Pressure constraint:

$$-\Delta p_s^{\text{up}}(1 - Y_s) \leq P_{p_2} - P_{p_1} \leq \Delta p_s^{\text{up}}(1 - Y_s) \quad (\text{A2.14})$$

Temperature constraint:

$$-\Delta t_s^{\text{up}}(1 - Y_s) \leq T_{p_2} - T_{p_1} \leq \Delta t_s^{\text{up}}(1 - Y_s) \quad (\text{A2.15})$$

Adsorption, bleaching & chromatography

For $u \in \mathbf{U}_{\text{Ads}} \cup \mathbf{U}_{\text{Blc}} \cup \mathbf{U}_{\text{Chr}}$, we use the following parameters:

$$\pi_{u,c} = \text{removal percentage of component } c \text{ in unit } u$$

Definition of component removal percentage:

$$\pi_{u,c} = \frac{F_{p2,c}}{F_{p1,c}}, \quad c \quad (\text{A2.16})$$

Unit volume:

$$A_u = \tau_u \sum_c \frac{F_{p1,c}}{\rho_c} \quad (\text{A2.17})$$

Cell disruption

For $u \in \mathbf{U}_{\text{Cdr}}$, we use the following parameters:

$$\mu^{\text{prod}}/\mu^{\text{debris}} = \text{mass fraction of product/debris in cells}$$

$$w^{\text{lo}}/w^{\text{up}} = \text{min/max mass fraction of water (wt\%) allowed in the feed}$$

Conversion of cells to product, cell debris & soluble impurities:

$$\begin{aligned}
F_{p2, \text{prodt}} &= \mu^{\text{prodt}} \beta_{u, \text{cell}} F_{p1, \text{cell}} \\
F_{p2, \text{cell}} &= (1 - \beta_{u, \text{cell}}) F_{p1, \text{cell}} \\
F_{p2, \text{debris}} &= \mu^{\text{debris}} \beta_{u, \text{cell}} F_{p1, \text{cell}} \\
F_{p2, \text{solimp}} &= (1 - \mu^{\text{debris}} - \mu^{\text{prodt}}) \beta_{u, \text{cell}} F_{p1, \text{cell}}
\end{aligned} \tag{A2.18}$$

Feed water content range:

$$w^{\text{lo}} \leq \frac{F_{p1, \text{water}}}{F_{p1}} \leq w^{\text{up}} \tag{A2.19}$$

Unit volume:

$$A_u = \tau_u \sum_c \frac{F_{p1, c}}{\rho_c} \tag{A2.20}$$

Crystallization

For $u \in \mathbf{U}_{\text{CrS}}$, we use the following parameters:

$$\theta_u^{\text{crst}} = \text{crystallization temperature in unit } u$$

and the following variables:

$$D_{u, c} = \text{removal percentage of component } c \text{ in unit } u$$

The operating mode is pre-determined (cooling/evaporation). For cooling, the following model is formulated:

Crystallization of product (dissolved):

$$F_{p3, \text{prodt sol}} = (1 - \beta_{u, \text{prodt sol}}) F_{p1, \text{prodt sol}}, F_{p3, \text{prodt crs}} = \beta_{u, \text{prodt sol}} F_{p1, \text{prodt sol}} \tag{A2.21}$$

Energy consumption:

$$Q_u = \sum_c \mu_k (298 - \theta_u^{\text{crst}}) F_{p1,c} \quad (\text{A2.22})$$

Unit volume:

$$A_u = \tau_u \sum_c \frac{F_{p1,c}}{\rho_c} \quad (\text{A2.23})$$

For evaporation, the following model is formulated:

Crystallization of product:

$$F_{p3,\text{prodt sol}} = (1 - \beta_{u,\text{prodt sol}}) F_{p1,\text{prodt sol}}, F_{p3,\text{prodt crs}} = \beta_{u,\text{prodt sol}} F_{p1,\text{prodt sol}} \quad (\text{A2.24})$$

Removal of components:

$$D_{u,c} = \frac{F_{p2,c}}{F_{p1,c}}, \quad c \in \{c | \alpha_c > \alpha_{\text{prodt}}\} \quad (\text{A2.25})$$

$$F_{p2,c} = 0, \quad c \in \{c | \alpha_c \leq \alpha_{\text{prodt}}\}$$

Energy consumption:

$$Q_u = \sum_c h_c^{\text{vap}} F_{p2,c} \quad (\text{A2.26})$$

Unit volume:

$$A_u = \tau_u \sum_c \frac{F_{p1,c}}{\rho_c} \quad (\text{A2.27})$$

Differential digestion

For $u \in \mathbf{U}_{\text{Ddg}}$, it is assumed that the digester only dissolves cell debris. We use the following parameters:

$$v_u^{\text{digr}} = \text{digesting agent requirement [kg/kg debris] in unit } u$$

Differential digestion of debris:

$$F_{p3,debris} = (1 - \beta_{u,debris})F_{p1,debris} \quad (A2.28)$$

$$F_{p3,digdebris} = \beta_{u,debris}F_{p1,debris} + F_{p2,digr}$$

Consumption of digesting agent:

$$F_{p3,digr} = 0 \quad (A2.29)$$

Digesting agent input:

$$F_{p2,digr} = v_u^{digr} F_{p1,debris} \quad (A2.30)$$

$$F_{p2,c} = 0, \quad c \neq digr$$

Unit volume:

$$A_u = \tau_u \sum_c \frac{F_{p1,c}}{\rho_c} \quad (A2.31)$$

Distillation

For $u \in \mathbf{U}_{Dst}$, we use the following parameters:

ξ_u^H / ξ_u^L = splitting fraction of heavy/light key to distillate

ι_u / λ_u = stage height [m]/vapor linear velocity [m/hr] of unit u

θ_u^{feed} = feed temperature (K) of unit u

$\varphi_{u,c}$ = vapor mass-to-volume conversion factor [m³/kg] of component c in unit u (calculated based on ideal gas assumption)

$\alpha_{u,c}$ = relative volatility of component c with respect to the heavy

key in unit u

and the following variables:

$R_u^{\min}$ = minimum reflux ratio of unit u

$M_u^{\text{distmin}}/M_u^{\text{dist}}$ = minimum number/number of stages of unit u

B_u = an auxiliary variable of unit u

Underwood equation:

$$\sum_c \frac{\alpha_{u,c} F_{p1,c}}{\alpha_{u,c} - B_u} = 0 \quad (\text{A2.32})$$

Underwood equation:

$$R_u^{\min} = \sum_c \frac{\alpha_{u,c} X_{p2,c}}{\alpha_{u,c} - B_u} - 1 \quad (\text{A2.33})$$

Fenske equation:

$$X_{p2,c} X_{p3,c'} = X_{p2,c'} X_{p3,c} \alpha_{u,c}^{M_u^{\text{distmin}}}, \quad (\text{A2.34})$$

$$c' \in \mathbf{C}_u^{\text{HK}}, c'' \in \mathbf{C}_u^{\text{LK}}, c = \{c | \alpha_{u,c'} < \alpha_{u,c} < \alpha_{u,c''}\}$$

Specification of heavy/light key splitting fraction to distillate:

$$\frac{F_{p2,c'}}{F_{p1,c'}} = \xi_u^{\text{H}}, \quad c' \in \mathbf{C}_u^{\text{HK}} \quad (\text{A2.35})$$

$$\frac{F_{p2,c}}{F_{p1,c}} = \xi_u^{\text{L}}, \quad c \in \mathbf{C}_u^{\text{LK}}$$

Complete separation assumption:

$$F_{p2,c} = 0, \quad c = \{c | \alpha_{u,c} < \alpha_{u,c'} \in \mathbf{C}_u^{\text{HK}}\} \quad (\text{A2.36})$$

$$F_{p3,c} = 0, \quad c = \{c | \alpha_{u,c} > \alpha_{u,c'} \in \mathbf{C}_u^{\text{LK}}\}$$

Number of stages:

$$\frac{M_u^{\text{dist}} - M_u^{\text{distmin}}}{M_u^{\text{dist}} + 1} = 0.75 - 0.75 \left(\frac{0.3R_u^{\text{min}}}{1.3R_u^{\text{min}} + 1} \right)^{0.57} \quad (\text{A2.37})$$

Feed solid percentage tolerance (10 wt%):

$$\frac{F_{p1,c}}{F_{p1}} \leq 0.1, \quad c \in \mathbf{C}_u^{\text{SOLID}} \quad (\text{A2.38})$$

Heating duty (=cooling duty):

$$Q_u = (1 + 1.3R_u^{\text{min}}) \sum_c h_c^{\text{vap}} F_{p2,c} + \sum_c \mu_k (298 - \theta_u^{\text{feed}}) F_{p1,c} \quad (\text{A2.39})$$

Distillation column volume:

$$A_u = (1 + 1.3R_u^{\text{min}}) \frac{\iota_u M_u^{\text{dist}}}{\lambda_u} \sum_c \varphi_{u,c} F_{p2,c} \quad (\text{A2.40})$$

Note that we regard evaporation as a special distillation. Thus, if a distillation unit with a single stage and zero reflux ratio is activated in the optimal process, then this distillation unit can be regarded as an evaporation unit.

Drying

For $u \in \mathbf{U}_{\text{Dry}}$, the air is considered free and the electricity required to pump the air is neglected.

Therefore, the inlet port for air is omitted. We use the following parameters:

$$\psi_u^{\text{min}} = \text{minimum product mass fraction required in the feed to unit } u$$

Removal of components:

$$D_{u,c} = \frac{F_{p2,c}}{F_{p1,c}}, \quad c \in \{c | \alpha_{u,c} > \alpha_{u,\text{prod}}\} \quad (\text{A2.41})$$

$$F_{p2,c} = 0, \quad c \in \{c | \alpha_{u,c} \leq \alpha_{u,\text{prod}}\}$$

Dryer feed concentration requirement:

$$F_{p1,c} \geq \psi_u^{\min} F_{p1}, \quad c \in \mathbf{C}^{\text{prod}} \quad (\text{A2.42})$$

Energy consumption:

$$Q_u = \sum_c h_c^{\text{vap}} F_{p2,c} \quad (\text{A2.43})$$

Unit volume:

$$A_u = \tau_u \sum_c \frac{F_{p1,c}}{\rho_c} \quad (\text{A2.44})$$

Extraction

For $u \in \mathbf{U}_{\text{Ext}}$, we use the following parameters:

h = minimum fraction of fresh solvent in port p2

$\varepsilon_{cc'}$ = component c solubility [kg/m³] in c'

$b_{u,c}$ = partition coefficient of component c in unit u

$e_{u,c}^{\text{lo}}/e_{u,c}^{\text{up}}$ = extraction coefficient lower/upper bound of component c in unit u

and the following variables:

$E_{u,c}$ = extraction coefficient of component c in unit u

M_u^{ext} = number of stages of unit u

Extraction coefficient definition:

$$E_{u,c} = \frac{b_{u,c} F_{p2,c'}}{F_{p1,\text{water}}}, \quad c \in \mathbf{C}_u^{\text{SOL}}, c' \in \mathbf{C}_u^{\text{SOLV}} \quad (\text{A2.45})$$

Pure solvent feed assumption:

$$F_{\text{rsv},1,\text{out},1,p2,c} = 0, \quad c \notin \mathbf{C}_u^{\text{SOLV}} \quad (\text{A2.46})$$

Minimum fresh solvent fraction:

$$F_{\text{rsv},1,\text{out},1,p2,c} \geq h F_{p2,c}, \quad c \in \mathbf{C}_u^{\text{SOLV}} \quad (\text{A2.47})$$

Extraction coefficient bounds:

$$e_{u,c}^{\text{lo}} \leq E_{u,c} \leq e_{u,c}^{\text{up}} \quad (\text{A2.48})$$

Concentrating effect relation:

$$X_{p1,c} = \left(\frac{E_{u,c}^{M_u^{\text{ext}}+1} - 1}{E_{u,c} - 1} \right) X_{p4,c}, \quad c \in \mathbf{C}_u^{\text{SOL}} \quad (\text{A2.49})$$

Water distribution to solvent phase:

$$F_{p3,\text{water}} = \varepsilon_{\text{water},c} \frac{F_{p3,c}}{\rho_c}, \quad c \in \mathbf{C}_u^{\text{SOLV}} \quad (\text{A2.50})$$

Solvent distribution to water phase:

$$F_{p4,c} = \varepsilon_{c,\text{water}} \frac{F_{p4,\text{water}}}{\rho_{\text{water}}}, \quad c \in \mathbf{C}_u^{\text{SOLV}} \quad (\text{A2.51})$$

Unit volume:

$$A_u = \tau_u M_u^{\text{ext}} \left(\sum_c \frac{F_{p4,c}}{\rho_c} + \sum_{c'} \frac{F_{p3,c'}}{\rho_{c'}} \right) \quad (\text{A2.52})$$

Flocculation

For $u \in \mathbf{U}_{\text{Flc}}$, it is assumed that the flocculent only binds to cells and cell debris, thus increasing particle sizes to facilitate separation. We use the following parameters:

$$v_u^{\text{floct}} = \text{flocculent requirement [kg/kg feed stream] in unit } u$$

Conversion of cells/debris to cell/debris flocs:

$$F_{p3,c} = (1 - \beta_{u,c})F_{p1,c}, \quad c \in \{\text{cell, debris}\}$$

$$F_{p3,\text{cellfc}} = \beta_{u,\text{cell}}F_{p1,\text{cell}} + F_{p2,\text{floct}} \frac{F_{p1,\text{cell}}}{F_{p1,\text{cell}} + F_{p1,\text{debris}}} \quad (\text{A2.53})$$

$$F_{p3,\text{debrisfc}} = \beta_{u,\text{debris}}F_{p1,\text{debris}} + F_{p2,\text{floct}} \frac{F_{p1,\text{debris}}}{F_{p1,\text{cell}} + F_{p1,\text{debris}}}$$

Consumption of flocculent:

$$F_{p3,\text{floct}} = 0 \quad (\text{A2.54})$$

Flocculent input:

$$F_{p2,\text{floct}} = v_u^{\text{floct}}F_{p1} \quad (\text{A2.55})$$

$$F_{p2,c} = 0, \quad c \neq \text{digr}$$

Unit volume:

$$A_u = \tau_u \sum_c \frac{F_{p1,c}}{\rho_c} \quad (\text{A2.56})$$

Note that if the product is EX, then all the terms regarding the debris and debrisfc components are removed.

Flotation

For $u \in \mathbf{U}_{\text{Flt}}$, all the insoluble components float to the top.

Insoluble components removal percentage definition:

$$D_{u,c} = \frac{F_{p2,c}}{F_{p1,c}}, \quad c \in \mathbf{C}_u^{\text{NSL}} \quad (\text{A2.57})$$

Removal percentage bounds:

$$d_{u,c}^{\text{lo}} \leq D_{u,c} \leq d_{u,c}^{\text{up}}, \quad c \in \mathbf{C}_u^{\text{NSL}} \quad (\text{A2.58})$$

Limiting velocity definition:

$$\frac{D_{u,c}}{v_c} = \frac{D_{u,\text{ns11}}}{v_{\text{ns11}}}, \quad c \in \mathbf{C}_u^{\text{NSL}}, c \neq \text{ns11} \quad (\text{A2.59})$$

Distribution of dissolved components:

$$\frac{F_{p1,c}}{F_{p1,\text{water}}} = \frac{F_{p2,c}}{F_{p2,\text{water}}}, \quad c \in \mathbf{C}_u^{\text{DIS}} \quad (\text{A2.60})$$

Concentrating effect relation:

$$\eta_u^{\min} \leq \frac{\sum_{c \in \mathbf{C}_u^{\text{NSL}}} F_{p2,c} / \sum_c F_{p2,c}}{\sum_{c \in \mathbf{C}_u^{\text{NSL}}} F_{p1,c} / \sum_c F_{p1,c}} \leq \eta_u^{\max} \quad (\text{A2.61})$$

Unit volume:

$$A_u = \frac{\chi^{\text{flot}}}{v_{\text{ns11}}} D_{u,\text{ns11}} \sum_c \frac{F_{p1,c}}{\rho_c} \quad (\text{A2.62})$$

Insoluble components follow the $D_{u,c} = \frac{v_c}{V_u^{\text{lim}}}$ rule in the “limiting velocity definition” equation.

However, using V_u^{lim} will cause nonlinearity. Therefore, we use $\frac{D_{u,\text{ns11}}}{v_{\text{ns11}}}$ whenever we need V_u^{lim} in an equation, where “ns11” denotes an arbitrary insoluble component ($c \in \mathbf{C}_u^{\text{NSL}}$) that is present in the unit.

Membrane & filtration

For $u \in \mathbf{U}_{\text{Mbr}} \cup \mathbf{U}_{\text{Ftt}}$, we use the following parameters:

$\gamma_{u,c}$ = rejection coefficient of component c in unit u

ω_u = flux [m/hr] of unit u

$n_u^{\text{lo}}/n_u^{\text{up}}$ = lower/upper bound of concentrating factor of unit u

and the following variables:

N_u = concentrating factor of unit u

Concentrating factor definition:

$$N_u = \frac{F_{p1}}{F_{p3}} \quad (\text{A2.63})$$

Concentrating factor bounds:

$$n_u^{\text{lo}} \leq N_u \leq n_u^{\text{up}} \quad (\text{A2.64})$$

Rejection coefficient definition:

$$\gamma_{u,c} = 1 - \frac{X_{p2,c}}{X_{p3,c}}, \quad c \in \{c | \gamma_{uc} > 0\} \quad (\text{A2.65})$$

Non-rejecting components distribution:

$$\frac{F_{p3,c}}{\sum_{c'' \in \{c | \gamma_{u,c''} = 0\}} F_{p3,c''}} = \frac{F_{p1,c}}{\sum_{c' \in \{c' | \gamma_{u,c'} = 0\}} F_{p1,c'}}, \quad (\text{A2.66})$$

$$c \in \{c | \gamma_{u,c} = 0\}, c = c_1, \dots, c_{|C|-1}$$

Unit area:

$$A_u = \frac{N_u - 1}{\omega_u} \sum_c \frac{F_{p1,c}}{\rho_c} \quad (\text{A2.67})$$

Precipitation

For $u \in \mathbf{U}_{\text{PRC}}$, an added chemical makes the product insoluble. We use the following parameters:

$$v_u^{\text{prcpt}} = \text{precipitant requirement [kg/kg feed stream] in unit } u$$

Precipitation of product:

$$F_{\text{p3,prod}} = (1 - \beta_{u,\text{prod}})F_{\text{p1,prod}}, F_{\text{p3,prod}+\text{prc}} = \beta_{u,\text{prod}}F_{\text{p1,prod}} \quad (\text{A2.68})$$

Precipitant input:

$$F_{\text{p2,prcpt}} = v_u^{\text{prcpt}}F_{\text{p1}} \quad (\text{A2.69})$$

$$F_{\text{p2},c} = 0, \quad c \neq \text{digr}$$

Unit volume:

$$A_u = \tau_u \sum_c \frac{F_{\text{p1},c}}{\rho_c} \quad (\text{A2.70})$$

Note that if the product is SLD NSL, then the “prod” component in the above equations becomes “prodtsol”.

Sedimentation & centrifugation

For $u \in \mathbf{U}_{\text{Sdm}} \cup \mathbf{U}_{\text{Cnt}}$, we use the following parameters:

$$\eta_u^{\min}/\eta_u^{\max} = \text{minimum/maximum concentrating factor of unit } u$$

$$v_c = \text{settling velocity [m/s] of particle } c$$

$$d_{u,c}^{\text{lo}}/d_{u,c}^{\text{up}} = \text{removal percentage lower/upper bound of component } c \text{ in unit } u$$

Insoluble components removal percentage definition:

$$D_{u,c} = \frac{F_{\text{p4},c}}{F_{\text{p1},c}}, \quad c \in \mathbf{C}_u^{\text{NSLHV}} \quad (\text{A2.71})$$

$$D_{u,c'} = \frac{F_{p2,c'}}{F_{p1,c'}} , \quad c' \in \mathbf{C}_u^{\text{NSLLT}}$$

Complete separation between the light and the heavy:

$$F_{p4,c} = 0 , \quad c \in \mathbf{C}_u^{\text{NSLLT}} \quad (\text{A2.72})$$

$$F_{p2,c'} = 0 , \quad c' \in \mathbf{C}_u^{\text{NSLHV}}$$

Removal percentage bounds:

$$d_{u,c}^{\text{lo}} \leq D_{u,c} \leq d_{u,c}^{\text{up}} , \quad c \in \mathbf{C}_u^{\text{NSLHV}} \cup \mathbf{C}_u^{\text{NSLLT}} \quad (\text{A2.73})$$

Limiting velocity definition:

$$\frac{D_{u,c}}{v_c} = \frac{D_{u,\text{ns11}}}{v_{\text{ns11}}} , \quad c \in \mathbf{C}_u^{\text{NSLHV}} \cup \mathbf{C}_u^{\text{NSLLT}} , c \neq \text{ns11} \quad (\text{A2.74})$$

Distribution of dissolved components:

$$\frac{F_{p1,c}}{F_{p1,\text{water}}} = \frac{F_{p4,c}}{F_{p4,\text{water}}} , \frac{F_{p1,c}}{F_{p1,\text{water}}} = \frac{F_{p2,c}}{F_{p2,\text{water}}} , \quad c \in \mathbf{C}_u^{\text{DIS}} \quad (\text{A2.75})$$

Concentrating effect relation:

$$\eta_u^{\min} \leq \frac{\sum_{c \in \mathbf{C}_u^{\text{NSLHV}}} F_{p4,c} / \sum_c F_{p4,c}}{\sum_{c \in \mathbf{C}_u^{\text{NSLHV}}} F_{p1,c} / \sum_c F_{p1,c}} \leq \eta_u^{\max} , \eta_u^{\min} \leq \frac{\sum_{c \in \mathbf{C}_u^{\text{NSLLT}}} F_{p2,c} / \sum_c F_{p2,c}}{\sum_{c \in \mathbf{C}_u^{\text{NSLLT}}} F_{p1,c} / \sum_c F_{p1,c}} \leq \eta_u^{\max} \quad (\text{A2.76})$$

Unit area (sigma factor):

$$A_u = \frac{D_{u,\text{ns11}} \sum_c \frac{F_{p1,c}}{\rho_c}}{v_{\text{ns11}}} \quad (\text{A2.77})$$

Note that if only NSL LT (or NSL HV) components needs to be removed, then the terms containing component $\mathbf{C}_u^{\text{NSLHV}}$ (or $\mathbf{C}_u^{\text{NSLLT}}$) and port p4 (or p2) should be removed from the model.

Solubilization

For $u \in \mathbf{U}_{\text{Sib}}$, the solvent dissolves only the product. We use the following parameters:

$$v_u^{\text{solbr}} = \text{solubilizing agent requirement [kg/kg feed stream] in unit } u$$

Solubilization of product:

$$F_{p3,\text{prodt}} = (1 - \beta_{u,\text{prodt}})F_{p1,\text{prodt}}, F_{p3,\text{prodprc}} = (1 - \beta_{u,\text{prodprc}})F_{p1,\text{prodprc}} \quad (\text{A2.78})$$

$$F_{p3,\text{prodt sol}} = \beta_{u,\text{prodt}}(F_{p1,\text{prodt}} + F_{p1,\text{prodprc}})$$

Solubilizing agent input:

$$F_{p2,\text{solbr}} = v_u^{\text{solbr}} F_{p1} \quad (\text{A2.79})$$

$$F_{p2,c} = 0, \quad c \neq \text{digr}$$

Unit volume:

$$A_u = \tau_u \sum_c \frac{F_{p1,c}}{\rho_c} \quad (\text{A2.80})$$

Costs

For cost calculations, we use the following parameters:

$$o = \text{operating hours [hr] per year}$$

$$\delta_u^C / \delta_u^S = \text{reference cost [\$]/size of unit } u$$

$$\sigma^{\text{pc-cc}} = \text{conversion factor of purchase cost to total capital cost}$$

$$\sigma^{\text{mat_oc}} / \sigma^{\text{util_oc}} = \text{contribution factor of material/utility in operating cost}$$

$$\sigma^{\text{labr_oc}} / \sigma^{\text{pc_oc}} / \sigma^{\text{cons_oc}} = \text{contribution factor of labor/purchase/consumable in operating cost}$$

$$\kappa_c / \kappa_l = \text{purchase price [$/kg] of component } c / \text{price [$/KWh or $/hr]}$$

of utility l

$$v_u^{\text{elec}}/v_u^{\text{labr}} = \text{electricity [KWh/m}^2\text{]}/\text{labor [hr/(unit}\cdot\text{operating hr)]}$$

requirement of unit u

$$sz_u^{\text{max}} = \text{maximum size of unit } u \text{ an operator can handle, only used for}$$

labor cost calculation, not for equipment sizing

$$cs_u^{\text{price}} = \text{consumable price [}/(\text{size}\cdot\text{operating hr)}\text{]} \text{ of unit } u$$

and the following variables:

$$C_u^{\text{pc}} = \text{purchase cost [\$] of unit } u$$

$$C_u^{\text{acc}}/C_u^{\text{labr}}/C_u^{\text{cons}} = \text{capital/labor/consumable annualized cost [$/yr] of unit } u$$

$$C_u^{\text{util}}/C_u^{\text{oc}}/C_u^{\text{atc}} = \text{utility/operating/total annualized cost [$/yr] of unit } u$$

$$U = \text{total annual cost [\$]}$$

Operating costs usually involve components such as overhead, maintenance, lab cost, insurance, administration, etc, which are estimated as a percentage of other costs (fixed capital cost, labor cost, utility cost, etc.). Thus, we express the total operating cost as a linear combination of these other costs, using the σ coefficients.

Unit purchase cost:

$$C_u^{\text{pc}} = \delta_u^{\text{c}} \left(\frac{A_u}{\delta_u^{\text{s}}} \right)^{0.6}, \quad u \quad (\text{A2.81})$$

Unit annualized capital cost (ACC):

$$C_u^{\text{acc}} = \sigma^{\text{pc-cc}} \beta C_u^{\text{pc}}, \quad u \quad (\text{A2.82})$$

Dst utility cost:

$$C_u^{\text{util}} = oQ_u \sum_l \kappa_l \quad (\text{A2.83})$$

Cnt & Cdr utility cost:

$$C_u^{\text{util}} = o\kappa_{\text{elec}} v_u^{\text{elec}} A_u \quad (\text{A2.84})$$

Crs (cooling) utility cost:

$$C_u^{\text{util}} = oQ_u \kappa_{\text{cool}} \quad (\text{A2.85})$$

Crs (evaporation) & Dry utility cost:

$$C_u^{\text{util}} = oQ_u \kappa_{\text{steam}} \quad (\text{A2.86})$$

Mbr, Ftt, Ads, Chr & Blc consumable cost:

$$C_u^{\text{cons}} = o c s_u^{\text{price}} A_u \quad (\text{A2.87})$$

Flc, Slb, Ddg & Ext material cost:

$$C_u^{\text{mat}} = o \sum_c \kappa_c F_{s,c}, \quad s \in \mathbf{S}_{\text{Rsv},1,\text{out},1,u,\text{in},2} \quad (\text{A2.88})$$

Unit Labor cost:

$$C_u^{\text{labr}} = o\kappa_{\text{labr}} v_u^{\text{labr}} \frac{A_u}{SZ_u^{\text{max}}} \quad (\text{A2.89})$$

Cdr, Dst, Crs, Cnt & Dry operating cost:

$$C_u^{\text{oc}} = \sigma^{\text{util_oc}} C_u^{\text{util}} + \sigma^{\text{labr_oc}} C_u^{\text{labr}} + \sigma^{\text{pc_oc}} C_u^{\text{pc}} \quad (\text{A2.90})$$

Mbr, Ftt, Ads, Chr & Blc operating cost:

$$C_u^{\text{oc}} = \sigma^{\text{cons_oc}} C_u^{\text{cons}} + \sigma^{\text{labr_oc}} C_u^{\text{labr}} + \sigma^{\text{pc_oc}} C_u^{\text{pc}} \quad (\text{A2.91})$$

Flc, Slb, Ddg & Ext operating cost:

$$C_u^{\text{oc}} = \sigma^{\text{mat_oc}} C_u^{\text{mat}} + \sigma^{\text{labr_oc}} C_u^{\text{labr}} + \sigma^{\text{pc_oc}} C_u^{\text{pc}} \quad (\text{A2.92})$$

Sdm & Flt operating cost:

$$C_u^{\text{oc}} = \sigma^{\text{labr_oc}} C_u^{\text{labr}} + \sigma^{\text{pc_oc}} C_u^{\text{pc}} \quad (\text{A2.93})$$

Unit annual total cost:

$$C_u^{\text{atc}} = C_u^{\text{acc}} + C_u^{\text{oc}}, \quad u \quad (\text{A2.94})$$

Src,1 specifications

For the source unit $u \in \mathbf{U}_{\text{Src},1}$, we use the following parameters:

$$x_c^{\text{src}} = \text{source mass fraction [wt\%] of component } c$$

Source flow composition specification:

$$X_{p1,c} = x_c^{\text{src}}, \quad c \quad (\text{A2.95})$$

Snk,1 specifications

For the sink unit $u \in \mathbf{U}_{\text{Snk},1}$, we use the following parameters:

$$z = \text{final product purity [wt\%]}$$

$$\zeta = \text{minimum overall product recovery [wt\%]}$$

$$f^{\text{snk}} = \text{sink product flow rate [kg/hr]}$$

$$\mu^{\text{prod}} = \text{mass fraction of product in cells}$$

Note that the minimum recovery value ζ should be small, only to provide a valid (but not necessarily tight) lower bound.

Final product flow specification:

$$\sum_{c \in \mathbb{C}^{\text{prod}}} F_{p1,c} = f^{\text{snk}} \quad (\text{A2.96})$$

Final product stream purity specification:

$$\sum_{c \in \mathbb{C}^{\text{prod}}} F_{p1,c} \geq z F_{p1} \quad (\text{A2.97})$$

Overall product recovery specification:

$$\text{For EX product: } \sum_{c \in \mathbb{C}^{\text{prod}}} F_{p1,c} \geq \zeta F_{\text{src},1,\text{out},1,\text{prod}} \quad (\text{A2.98})$$

$$\text{For IN product: } \sum_{c \in \mathbb{C}^{\text{prod}}} F_{p1,c} \geq \zeta \mu^{\text{prod}} F_{\text{src},1,\text{out},1,\text{cell}}$$

Global bounds

$$F_{s,c} \leq f^{\text{up}} \quad (\text{A2.99})$$

$$X_{p,c} \leq 1 \quad (\text{A2.100})$$

Note that f^{up} is a valid (but not necessarily tight) upper bound.

A3 Explanations to Chapter 4

A3.1 Key parameters for EX NSL LT LQD CMD product

Table A3.1. Key parameters for EX NSL LT LQD CMD product.

Parameter category	Parameter	Base case (nominal)	Range
Operation choices	Production capacity (kg/h)	1000	-
	Annual operation time (days)	330	-
Product streams	Initial product titer (g/L)	5	1-250
	Initial microbial cells (g/L)	2	-
	Final product purity (wt %)	95	-
Separation technologies	Sdm efficiency	70%	-
	Cnt efficiency	80%	70-95%
	Ftt retention factor	80%	70-95%
	Mbr rejection coefficient	97%	-

A3.2 Extension to EX NSL HV LQD CMD product

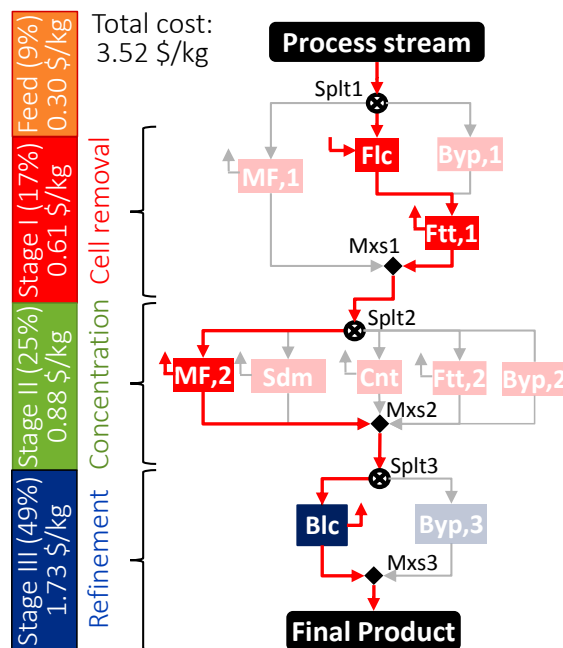


Figure A3.1. Superstructure (including all the units and streams) and optimal solution (the highlighted parts) for EX NSL HV LQD CMD product. The active streams are shown by bold red lines and selected technologies are highlighted in different colors corresponding to each stage: red for stage I, green for stage II, and blue for stage III. Cost distribution is shown by the numbers on the left bar.

A3.3. Variation in Ext partition coefficient

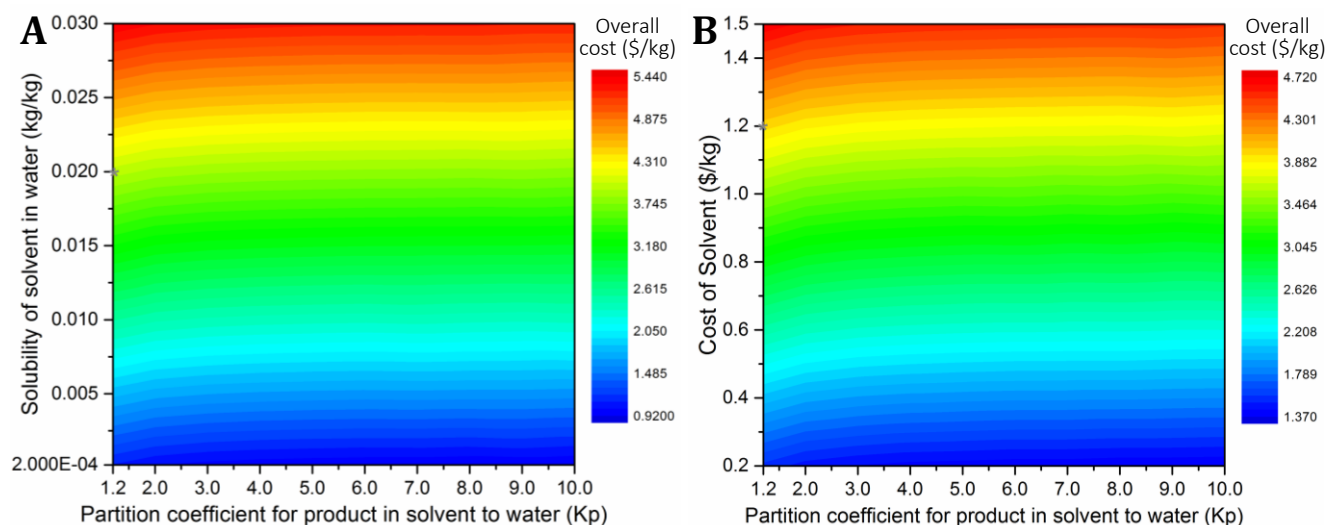


Figure A3.2. Variation in (A) partition coefficient versus solubility in water, and (B) partition coefficient versus cost of solvent. The base case values are represented by the grey asterisks.

A3.4. Model parameters and input data for base cases

Table A3.2. Important input parameters and product specifications.

Parameter	Nominal value	Units
Product titer	5	g/L (kg/m ³)
Cell titer	1.7	g/L (kg/m ³)
Liquid co-product titer	0.85	g/L (kg/m ³)
Solid co-product titer	0.85	g/L (kg/m ³)
Desired production capacity	1000	kg/h
Annual operation time	330	days/year
Final product purity	95	wt% purity

Table A3.3. Particle size and density information.

Component	Particle size (m)	Density (kg/m ³)	Molecular weight (kg/mol)
Product	5E-6	1800	775
Cell	8E-6	1250	24.6
Liquid co-product	3.8E-10	950	85
Solid co-product	5E-7	1100	140
Water	3E-10	1000	18

Table A3.4. Utility and labor costs.

Utility	Cost per unit (\$/unit)
Electricity	0.1 \$/KWH
Cooling water	5E-5 \$/kg
Steam	0.012 \$/kg
Labor	20 \$/laborer-h
Latent heat (steam)	2155.68 KJ/kg

Table A3.5. Standard capacities, costs, scaling factors, and labor requirements of technologies.

Unit operation (costing capacity)	Standard capacity (units)	Base costs (million \$)	Scaling exponent (n)	Labors required (#/h)	Power required (KW/h)
Differential digestion (Flowrate)	40 m ³ /h	0.474	0.5	1	0.1
Solubilization (Flowrate)	40 m ³ /h	0.474	0.5	1	0.1
Flocculation (Volume)	2000 m ³	0.54	0.5	0.1	0.0002
Sedimentation (Area)	2500 m ²	1.13	0.57	0.1	0
Centrifuge (Sigma factor)	60000 m ²	0.275	0.65	1	12.79
Filtration (Area)	80 m ²	0.04	0.55	0.5	0.1
Microfiltration (Area)	80 m ²	0.75	0.55	1	0.1
Precipitation (Flowrate)	40 m ³ /h	0.47	0.5	1	0.1
Freeze drying (Capacity)	600 kg/h	0.11	0.67	0.5	0.3
Distillation (Volume)	22.58 m ³	0.082	2.8	1	0
ATPE (Flowrate)	185 m ³ /h	0.362	0.67	1	0.5
Extraction (Flowrate)	185 m ³ /h	0.362	0.67	1	0.5
Chromatography (Volume)	0.633 m ³	0.775	0.67	1	0.33
Pervaporation (Area)	80 m ²	0.261	0.55	1	0.33
Bleaching (Volume)	0.27 m ³	0.1	0.67	1	0.33

FlocculationFlocculent added – 0.04 kg/m³

Residence time – 0.5 hr

Floc diameter – 5E-4 m

Flocculent cost – 5 \$/kg

Sedimentation tank

Efficiency – 70%

Maximum concentrating factor – 20

Depth – 3m

Residence time – 6 hr

Centrifuge

Efficiency – 80%

Maximum concentrating factor – 30

Rotation speed – 9000 rpm

Filtration

Flux - $0.2 \text{ m}^3\text{m}^{-2}\text{h}^{-1}$

Retention factors (Ftt): 80%

Filter cost – 100 \$/m²

Replacement time – 2000 h

Membrane

Flux – $0.0856 \text{ m}^3\text{m}^{-2}\text{h}^{-1}$

Retention factor: less filtered components (liq co-product, water, solvents) – 0.15, filtered components (product, cell, solid co-product, salt, polymer) – 0.85,

Membrane membrane cost – 500 \$/m²

Replacement time – 2000 h

Differential digestion

Agent required – 0.5 kg/kg NPCM (non-product solid materials – cells, solid co-product)

Cost of agent – 1.5 \$/kg

Density of agent – 1400 kg/m³

Solubilization

Solvent required – 0.5 kg/kg product

Cost of solvent – 1.5 \$/kg

Density of solvent – 1300 kg/m³

Precipitation

Efficiency of product precipitation – 98%

Anti-solvent required – 2 kg/kg product

Cost of anti-solvent – 1.8 \$/kg

Density of anti-solvent – 925 kg/m³

Distillation

Relative volatility: product – 1, water – 1.3, soluble co-product – 1.5

Heat of vaporization (KJ/kg): product – 573, water – 2257, soluble co-product – 1275

Feed quality, $q_f = 1$ (saturated liquid)

Vapor velocity – 3 m/s

Stage efficiency – 80%
 Height of stage – 0.6 m
 Reflux ratio multiplying factor – 1.3

Aqueous two phase extraction

Partition coefficient in top phase: product – 4, water – 1, soluble co-product – 2, heavy solid – 0.001
 Solubility of polymer in bottom phase – 0.005 (kg/kg)
 Solubility of salt in top phase – 0.005 (kg/kg)
 Polymer: Mol. Wt. – 450, Density – 1850 (kg/m³), Cost – 2 \$/kg
 Salt: Mol. Wt. – 136, Density – 1636 (kg/m³), Cost – 0.6 \$/kg

Extraction

Partition coefficient in solvent phase: product – 1.2, soluble co-product – 0.3, heavy solid – 0.0001
 Solubility of solvent in water – 0.02 (kg/kg)
 Solubility of water in solvent – 0.02 (kg/kg)
 Solvent: Mol. Wt. – 78, Density – 810 (kg/m³), Cost – 1.2 \$/kg

Distillation

Relative volatility: solvent – 5.2, product – 1, water – 1.4, soluble co-product – 1.5
 Heat of vaporization (KJ/kg): solvent – 592, product – 2000, water – 573, soluble co-product – 1275, heavy solid – 2257
 Other parameters are the same with those in Dst1

Drying

Sublimation of solvents – 97%
 Heat of sublimation – 5000 KJ/kg
 Ambient temperature: 20 °C
 Freezing temperature: (-1) °C
 Refrigerant inlet temperature – (-10) °C
 Refrigerant outlet temperature – 0 °C
 Specific heat (KJ/kg-°C): Refrigerant – 12, water – 4.2, liquids (co product and added agents) – 1, product – 6.2, solid coproduct – 5.7, cell – 4.2
 Heat transfer coefficient – 180 KJ/m²-K-h

Chromatography

Space time – 0.5 h
 Column capacity – 95%
 Width of chromatogram – 0.05 m
 HETP – 0.0035
 Ratio length to diameter – 0.14

Pervaporation

Flux – 0.055 m³m⁻²h⁻¹

Retention factor: product – 0.0002, water – 0.95, soluble co-product – 0.92, heavy solid – 0.99, salt – 0.99, polymer – 0.99, solvent – 0.001

Membrane cost – 1000 \$/m²

Replacement time – 2000 h

Bleaching

Bleaching efficiency – 99%

Cost of GAC (bleaching agent) – 4\$/kg

Replacement time – 360 h (15 days)

Other parameters

Capital charge factor – 0.11

Base module cost multiplier – 5.4

Annual operation time – 330 days

Cost of feed – 0.25 \$/kg (per kg product basis)

A4 Process economics in Chapter 5

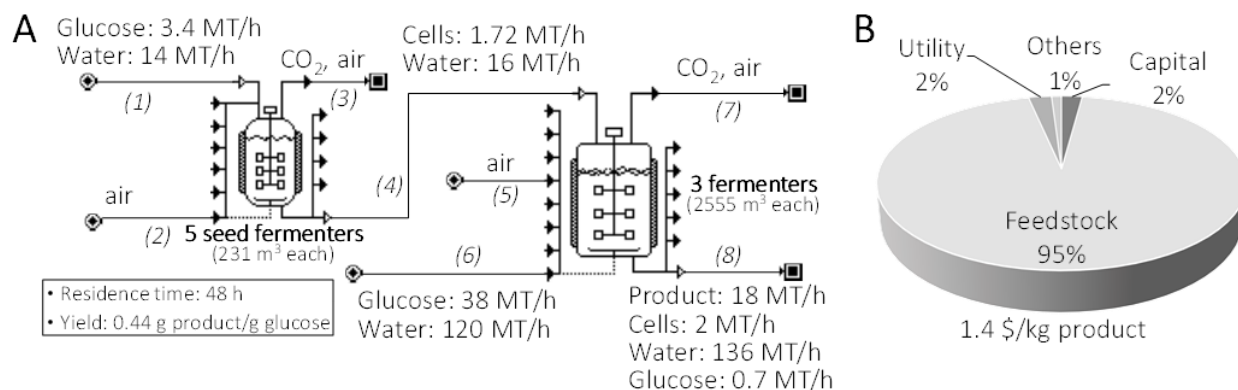


Figure A4.1. Upstream process and costs in the base case. **(A)** Process flow diagram; **(B)** cost distribution. For the yield and residence time estimated for different products in the current study, the flowrates of streams 1-6 are the same; compositions of streams 7 and 8 depend on the yield, and the fermenter size depends on the residence time of different products.

We first synthesize an upstream process (see **Figure A4.1-A**) for a base case (yield=0.44 g product/g glucose; fermentation residence time=48 h) using SuperPro Designer [125], assuming a 994 MT/day glucose supply (240 g/L concentration), which is the amount of glucose (and concentration) generated by hydrolyzing the 2000 MTDW/day biomass in the NREL process [213], where the other process and economic parameters are also adopted. The cost of the compressors used to transport air into the fermentation system is examined to be negligible, so the process flow diagram for both aerobic and anaerobic fermentations are the same here.

Then we perform an economic analysis for the base case to identify the equipment sizes, number of equipment, the product flowrate, capital cost (with a capital charge factor of 0.1 to annualize the cost), feedstock cost, utility cost, etc. (see **Figure A4.1**). The total upstream cost for the base case is calculated to be 1.4 \$/kg (pure product basis; 95% is feedstock cost). Glucose accounts for 99.6 % of the feedstock cost; the main fermenter accounts for 70% of the capital cost. In calculating the total unit cost of 1.4\$/kg product, a 10% product loss in the downstream separation process is assumed.

We also study a case where 2000 MTDW/day cellulosic biomass, instead of pure glucose, is used as the feedstock. In this case, the upstream process includes biomass pre-treatment, enzymatic hydrolysis and fermentation. The stream after hydrolysis contains 240 g/L sugar, and the cost of pre-treatment and hydrolysis is 0.28 \$/kg (pure sugar basis), i.e. a 43% feedstock cost saving compared to the 0.5\$/kg pure glucose price. The upstream cost can thus be reduced from 1.4 to 0.8 \$/kg product. However, we do not consider the use of cellulosic biomass for the screening due to its limited applications in industry. Nonetheless, the users can account for this by reducing the glucose cost by 43%.

Then, assuming the same 994 MT/day glucose supply (240 g/L concentration), we calculate the final cost with the specific residence time and yield calculated from the metabolic modeling approach for each product as shown in Equation B1-B6, where known parameters from the base case are denoted with lowercase letters and unknown variables with uppercase letters; P is flowrate of product produced; Y is yield; T is residence time; V is volume of the fermenter; ACC , UC , FC , OC and TUC represent annualized capital cost of the fermenter, utility cost, feedstock cost, other cost, and total unit cost, respectively; g is flowrate of glucose in the feed; q is total flowrate into the fermenter; v^* , acc^* , uc^* , fc^* and oc^* represent the fermenter volume, fermenter annualized capital cost, utility cost, feedstock cost, and other cost, respectively, in the base case.

$$P = gY \quad (A4.1)$$

$$V = qT \quad (A4.2)$$

$$ACC = acc^* \left(\frac{V}{v^*} \right)^{0.6} \quad (A4.3)$$

$$UC = uc^* \left(\frac{V}{v^*} \right) \quad (A4.4)$$

$$FC = fc^*, OC = oc^* \quad (A4.5)$$

$$TUC = \frac{ACC + UC + FC + OC}{P} \quad (A4.6)$$

Note that yield is the key determinant of upstream cost.

Bibliography

- [1] N. Nishida, G. Stephanopoulos, and A. W. Westerberg, "A review of process synthesis," *AIChE J.*, vol. 27, pp. 321–351, 1981.
- [2] J. J. Siirola, "Strategic process synthesis: Advances in the hierarchical approach," *Comput. Chem. Eng.*, vol. 20, pp. S1637–S1643, 1996.
- [3] L. T. Biegler, I. E. Grossmann, and A. W. Westerberg, *Systematic methods for chemical process design*. Englewood Cliffs, 1997.
- [4] M. E. Lienqueo and J. A. Asenjo, "Use of expert systems for the synthesis of downstream protein processes," *Comput. Chem. Eng.*, vol. 24, pp. 2339–2350, 2000.
- [5] E. N. Lightfoot and J. S. Moscariello, "Bioseparations," *Biotechnol. Bioeng.*, vol. 87, pp. 259–273, 2004.
- [6] M. E. Lienqueo, E. W. Leser, and J. a. Asenjo, "An expert system for the selection and synthesis of multistep protein separation processes," *Comput. Chem. Eng.*, vol. 20, pp. S189–S194, 1996.
- [7] G. Sandoval, D. Espinoza, N. Figueroa, and J. A. Asenjo, "Optimization of a biotechnological multiproduct batch plant design for the manufacture of four different products: A real case scenario," *Biotechnol. Bioeng.*, vol. 114, pp. 1252–1263, 2017.
- [8] E. M. Polykarpou, P. A. Dalby, and L. G. Papageorgiou, "Optimal synthesis of chromatographic trains for downstream protein processing," *Biotechnol. Prog.*, vol. 27, pp. 1653–1660, 2011.
- [9] E. M. Polykarpou, P. A. Dalby, and L. G. Papageorgiou, "An MILP formulation for the synthesis of protein purification processes," *Chem. Eng. Res. Des.*, vol. 90, pp. 1262–1270, 2012.
- [10] E. Simeonidis, J. M. Pinto, and L. G. Papageorgiou, "An MILP model for optimal design of purification tags and synthesis of downstream processing," *Comput. Aided Chem. Eng.*, vol. 20, pp. 1537–1542, 2005.

- [11] M. K. Garg, P. L. Douglas, and J. G. Linderst, "An expert system for identifying separation processes," *Can. J. Chem. Eng.*, vol. 69, pp. 67-75, 1991.
- [12] C. Palaniappan, R. Srinivasan, and R. B. Tan, "Expert system for the design of inherently safer processes. 2. Flowsheet development stage," *Ind. Eng. Chem. Res.*, vol. 41, pp. 6711-6722, 2002.
- [13] E. Vásquez-Alvarez, M. E. Lienqueo, and J. M. Pinto, "Optimal synthesis of protein purification processes," *Biotechnol. Prog.*, vol. 17, pp. 685-696, 2001.
- [14] I. E. Grossmann and M. M. Daichendt, "New trends in optimization-based approaches to process synthesis," *Comput. Chem. Eng.*, vol. 20, pp. 665-683, 1996.
- [15] S. D. Barnicki and J. J. Sirola, "Process synthesis prospective," *Comput. Chem. Eng.*, vol. 28, pp. 441-446, 2004.
- [16] S. Cremaschi, "A perspective on process synthesis: Challenges and prospects," *Comput. Chem. Eng.*, vol. 81, pp. 130-137, 2015.
- [17] H. Yeomans and I. E. Grossmann, "A systematic modeling framework of superstructure optimization in process synthesis," *Comput. Chem. Eng.*, vol. 23, pp. 709-731, 1999.
- [18] E. M. B. Smith and C. C. Pantelides, "Design of reaction/separation networks using detailed models," *Comput. Chem. Eng.*, vol. 19, pp. 83-88, 1995.
- [19] F. Friedler, K. Tarjan, Y. W. Huang, and L. T. Fan, "Graph-theoretic approach to process synthesis: Axioms and theorems," *Chem. Eng. Sci.*, vol. 47, pp. 1973-1988, 1992.
- [20] F. Friedler, K. Tarjan, Y. W. Huang, and L. T. Fan, "Graph-theoretic approach to process synthesis: Polynomial algorithm for maximal structure generation," *Comput. Chem. Eng.*, vol. 17, pp. 929-942, 1993.
- [21] T. Farkas, E. Rev, and Z. Lelkes, "Process flowsheet superstructures: Structural multiplicity and redundancy Part I: Basic GDP and MINLP representations," *Comput. Chem. Eng.*, vol. 29, pp. 2180-2197, 2005.
- [22] G. R. Kocis and I. E. Grossmann, "A modelling and decomposition strategy for the minlp optimization of process flowsheets," *Comput. Chem. Eng.*, vol. 13, pp. 797-819, 1989.
- [23] A. Lakshmanan and L. T. Biegler, "Synthesis of optimal chemical reactor networks," *Ind. Eng. Chem. Res.*, vol. 35, pp. 1344-1353, 1996.
- [24] A. C. Kokossis and C. A. Floudas, "Optimization of complex reactor networks-II.

- Nonisothermal operation," *Chem. Eng. Sci.*, vol. 49, pp. 1037–1051, 1994.
- [25] L. K. E. Achenie and L. T. Biegler, "A superstructure based approach to chemical reactor network synthesis," *Comput. Chem. Eng.*, vol. 14, pp. 23–40, 1990.
 - [26] C. A. Schweiger and C. A. Floudas, "Optimization framework for the synthesis of chemical reactor networks," *Ind. Eng. Chem. Res.*, vol. 38, pp. 744–766, 1999.
 - [27] T. F. Yee and I. E. Grossmann, "Simultaneous optimization models for heat integration: II. Heat exchanger network synthesis," *Comput. Chem. Eng.*, vol. 14, pp. 1165–1184, 1990.
 - [28] C. A. Floudas and G. E. Paules IV, "A mixed-integer nonlinear programming formulation for the synthesis of heat-integrated distillation sequences," *Comput. Chem. Eng.*, vol. 12, pp. 531–546, 1988.
 - [29] C. A. Floudas, A. R. Ciric, and I. E. Grossmann, "Automatic synthesis of optimum heat exchanger network configurations," *AIChE J.*, vol. 32, pp. 276–290, 1986.
 - [30] Z. Novak, Z. Kravanja, and I. E. Grossmann, "Simultaneous synthesis of distillation sequences in overall process schemes using an improved MINLP approach," *Comput. Chem. Eng.*, vol. 20, pp. 1425–1440, 1996.
 - [31] R. G. Harrison, P. Todd, S. R. Rudge, and D. P. Petrides, *Bioseparations science and engineering*. Oxford University Press, 2003.
 - [32] D. L. Pavia, *Introduction to organic laboratory techniques: a small scale approach*. Cengage Learning, 2005.
 - [33] A. A. Linninger, "Industry-wide energy saving by complex separation networks," *Comput. Chem. Eng.*, vol. 33, pp. 2018–2027, 2009.
 - [34] K. Lakhdar, Y. Zhou, J. Savery, N. J. Titchener-Hooker, and L. G. Papageorgiou, "Medium term planning of biopharmaceutical manufacture using mathematical programming," *Biotechnol. Prog.*, vol. 21, pp. 1478–1489, 2005.
 - [35] J. M. Natali, J. M. Pinto, and L. G. Papageorgiou, "Efficient MILP formulations for the simultaneous optimal peptide tag design and downstream processing synthesis," *AIChE J.*, vol. 55, pp. 2303–2317, 2009.
 - [36] G. J. Ruiz, S. B. Kim, J. Moon, L. Zhang, and A. A. Linninger, "Design and optimization of energy efficient complex separation networks," *Comput. Chem. Eng.*, vol. 34, pp. 1556–1563, 2010.
 - [37] L. Zhang and A. A. Linninger, "Towards computer-aided separation synthesis," *AIChE J.*, vol.

52, 2006.

- [38] A. C. Kokossis, M. Tsakalova, and K. Pyrgakis, "Design of integrated biorefineries," *Comput. Chem. Eng.*, vol. 81, pp. 40–56, 2014.
- [39] A. Alva-Argáez, A. C. Kokossis, and R. Smith, "Wastewater minimisation of industrial systems using an integrated approach," *Comput. Chem. Eng.*, vol. 22, pp. 741–744, 1998.
- [40] A. Alva-Argáez, A. Vallianatos, and A. Kokossis, "A multi-contaminant transshipment model for mass exchange networks and wastewater minimisation problems," *Comput. Chem. Eng.*, vol. 23, pp. 1439–1453, 1999.
- [41] J. Bausa and W. Marquardt, "Shortcut design methods for hybrid membrane/distillation processes for the separation of nonideal multicomponent mixtures," *Ind. Eng. Chem. Res.*, vol. 39, pp. 1658–1672, 2000.
- [42] J. A. Caballero and I. E. Grossmann, "Generalized disjunctive programming model for the optimal synthesis of thermally linked distillation columns," *Ind. Eng. Chem. Res.*, vol. 40, pp. 2260–2274, 2001.
- [43] B. Bao, D. K. S. Ng, D. H. S. Tay, A. Jiménez-Gutiérrez, and M. M. El-Halwagi, "A shortcut method for the preliminary synthesis of process-technology pathways: An optimization approach and application for the conceptual design of integrated biorefineries," *Comput. Chem. Eng.*, vol. 35, pp. 1374–1383, 2011.
- [44] J. A. Caballero and I. E. Grossmann, "An algorithm for the use of surrogate models in modular flowsheet optimization," *AIChE J.*, vol. 54, pp. 2633–2650, 2008.
- [45] C. A. Henao and C. T. Maravelias, "Surrogate-based superstructure optimization framework," *AIChE J.*, vol. 57, pp. 1216–1232, 2011.
- [46] S. Burer and A. N. Letchford, "Non-convex mixed-integer nonlinear programming: A survey," *Math. Program.*, vol. 16, pp. 97–106, 2012.
- [47] F. Boukouvala and M. G. Ierapetritou, "Surrogate-based optimization of expensive flowsheet modeling for continuous pharmaceutical manufacturing," *J. Pharm. Innov.*, vol. 8, pp. 131–145, 2013.
- [48] A. Agarwal and L. Biegler, "A trust-region framework for constrained optimization using reduced order modeling," *Optim. Eng.*, vol. 14, pp. 3–35, 2013.
- [49] J. Eason and S. Cremaschi, "Adaptive sequential sampling for surrogate model generation with artificial neural networks," *Comput. Chem. Eng.*, vol. 68, pp. 220–232, 2014.

- [50] A. Cozad, N. V. Sahinidis, and D. C. Miller, "Learning surrogate models for simulation-based optimization," *AIChE J.*, vol. 60, pp. 2211–2227, 2014.
- [51] A. W. Dowling and L. T. Biegler, "A framework for efficient large scale equation-oriented flowsheet optimization," *Comput. Chem. Eng.*, vol. 72, pp. 3–20, 2015.
- [52] M. Yu, D. Miller, and L. T. Biegler, "Dynamic reduced order models for simulating bubbling fluidized bed adsorbers," *Ind. Eng. Chem. Res.*, vol. 54, pp. 6959–6974, 2015.
- [53] A. Rogers and M. Ierapetritou, "Feasibility and flexibility analysis of black-box processes part 2: Surrogate-based flexibility analysis," *Chem. Eng. Sci.*, vol. 137, pp. 1005–1013, 2015.
- [54] V. Venkatasubramanian, C. Zhao, G. Joglekar, A. Jain, L. Hailemariam, P. Suresh, P. Akkisetty, K. Morris, and G. V. Reklaitis, "Ontological informatics infrastructure for pharmaceutical product development and manufacturing," *Comput. Chem. Eng.*, vol. 30, pp. 1482–1496, 2006.
- [55] P. Suresh, S. Hsu, P. Akkisetty, G. V. Reklaitis, and V. Venkatasubramanian, "OntoMODEL: ontological mathematical modeling knowledge management in pharmaceutical product development, 1: Conceptual framework," *Ind. Eng. Chem. Res.*, vol. 49, pp. 175–187, 2010.
- [56] L. Hailemariam and V. Venkatasubramanian, "Purdue ontology for pharmaceutical engineering: Part II. Applications," *J. Pharm. Innov.*, vol. 5, pp. 139–146, 2010.
- [57] M. Tawarmalani and N. V. Sahinidis, "A polyhedral branch-and-cut approach to global optimization," *Math. Progr.*, vol. 103, pp. 225–249, 2005.
- [58] J. G. Optim, C. A. Floudas, and C. E. Gounaris, "A review of recent advances in global optimization," *J. Glob. Optim.*, vol. 45, pp. 3–38, 2009.
- [59] R. Horst and P. Pardalos, *Handbook of global optimization*. Springer Science & Business Media, 2013.
- [60] U. Nallasivam, V. H. Shah, A. A. Shenvi, M. Tawarmalani, and R. Agrawal, "Global optimization of multicomponent distillation configurations: 1. Need for a reliable global optimization algorithm," *AIChE J.*, vol. 59, pp. 971–981, 2013.
- [61] F. Trespalacios and I. E. Grossmann, "Review of mixed-integer nonlinear and generalized disjunctive programming methods," *Chemie-Ingenieur-Technik*, vol. 86, pp. 991–1012, 2014.
- [62] F. Boukouvala, R. Misener, and C. A. Floudas, "Global optimization advances in Mixed-Integer Nonlinear Programming, MINLP, and Constrained Derivative-Free Optimization, CDFO," *Eur. J. Oper. Res.*, vol. 252, pp. 701–727, 2016.

- [63] T. W. Tee, A. Chowdhury, C. D. Maranas, and J. V. Shanks, "Systems metabolic engineering design: Fatty acid production as an emerging case study," *Biotechnol. Bioeng.*, vol. 111, pp. 849–857, 2014.
- [64] S. A. Angermayr, A. Gorchs Rovira, and K. J. Hellingwerf, "Metabolic engineering of cyanobacteria for the synthesis of commodity products," *Trends. Biotechnol.*, vol. 33, pp. 352–361, 2015.
- [65] H. Knoop and R. Steuer, "A computational analysis of stoichiometric constraints and trade-offs in cyanobacterial biofuel production," *Front. Bioeng. Biotechnol.*, vol. 3, pp. 47–55, 2015.
- [66] D. Julleson, F. David, B. Pfleger, and J. Nielsen, "Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals," *Biotechnol. Adv.*, vol. 33, pp. 1395–1402, 2015.
- [67] J. C. Liao, L. Mi, S. Pontrelli, and S. Luo, "Fuelling the future: microbial engineering for the production of sustainable biofuels," *Nat. Rev. Microbiol.*, vol. 14, pp. 288–299, 2016.
- [68] A. Bordbar, J. M. Monk, Z. A. King, and B. O. Palsson, "Constraint-based models predict metabolic and associated cellular functions," *Nat. Rev. Genet.*, vol. 15, pp. 107–120, 2014.
- [69] A. M. Lanza, N. C. Crook, and H. S. Alper, "Innovation at the intersection of synthetic and systems biology," *Curr. Opin. Biotechnol.*, vol. 23, pp. 712–717, 2012.
- [70] L. Miskovic and V. Hatzimanikatis, "Production of biofuels and biochemicals: In need of an ORACLE," *Trends Biotechnol.*, vol. 28, pp. 391–397, 2010.
- [71] S. A. Nicolaou, S. M. Gaida, and E. T. Papoutsakis, "A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: From biofuels and chemicals, to biocatalysis and bioremediation," *Metab. Eng.*, vol. 12, pp. 307–331, 2010.
- [72] U. T. Bornscheuer and A. T. Nielsen, "Editorial overview: Chemical biotechnology: Interdisciplinary concepts for modern biotechnological production of biochemicals and biofuels," *Curr. Opin. Biotechnol.*, vol. 35, pp. 133–134, 2015.
- [73] J. Nielsen, M. Fussenegger, J. Keasling, S. Y. Lee, J. C. Liao, K. Prather, and B. Palsson, "Engineering synergy in biotechnology," *Nat. Chem. Biol.*, vol. 10, pp. 319–322, 2014.
- [74] P. N. R. Vennestrøm, C. M. Osmundsen, C. H. Christensen, and E. Taarning, "Beyond petrochemicals: The renewable chemicals industry," *Angew. Chemie. Int. Ed.*, vol. 50, pp. 10502–10509, 2011.
- [75] S. A. Wilson and S. C. Roberts, "Metabolic engineering approaches for production of

- biochemicals in food and medicinal plants," *Curr. Opin. Biotechnol.*, vol. 26, pp. 174–182, 2014.
- [76] V. G. Yadav, M. De Mey, C. Giaw Lim, P. Kumaran Ajikumar, and G. Stephanopoulos, "The future of metabolic engineering and synthetic biology: Towards a systematic practice," *Metab. Eng.*, vol. 14, pp. 233–241, 2012.
- [77] A. A. Kiss, J. Grievink, and M. Rito-Palomares, "A systems engineering perspective on process integration in industrial biotechnology," *J. Chem. Technol. Biotechnol.*, vol. 90, pp. 349–355, 2015.
- [78] J. W. Lee, D. Na, J. M. Park, J. Lee, S. Choi, and S. Y. Lee, "Systems metabolic engineering of microorganisms for natural and non-natural chemicals," *Nat. Chem. Biol.*, vol. 8, pp. 536, 2012.
- [79] P. Sagmeister, P. Wechselberger, M. Jazini, A. Meitz, T. Langemann, and C. Herwig, "Soft sensor assisted dynamic bioprocess control: efficient tools for bioprocess development," *Chem. Eng. Sci.*, vol. 96, pp. 190–198, 2013.
- [80] C. Herwig, I. Marison, and U. Stockar, "On-line stoichiometry and identification of metabolic state under dynamic process conditions," *Biotechnol. Bioeng.*, vol. 75, pp. 345–354, 2001.
- [81] C. Herwig and U. Stockar, "A small metabolic flux model to identify transient metabolic regulations in *Saccharomyces cerevisiae*," *Bioprocess. Biosyst. Eng.*, vol. 24, pp. 395–403, 2002.
- [82] L. Brennan and P. Owende, "Biofuels from microalgae-a review of technologies for production, processing, and extractions of biofuels and co-products," *Renew. Sustain. Energy Rev.*, vol. 14, pp. 557–577, 2010.
- [83] K. Brandt and G. Schembecker, "Production rate-dependent key performance indicators for a systematic design of biochemical downstream processes," *Chem. Eng. Technol.*, vol. 39, pp. 354–364, 2016.
- [84] J. A. Shaeiwitz, J. D. Henry, and R. Ghosh, "Bioseparation," in *Ullmann's Encyclopedia of Industrial Chemistry*, 2000.
- [85] J. J. Vallino and G. Stephanopoulos, *Frontiers in bioprocessing*. CRC Press, 1990.
- [86] X. Li and A. Kraslawski, "Conceptual process synthesis: Past and current trends," *Chem. Eng. Process. Process Intensif.*, vol. 43, pp. 589–600, 2004.
- [87] R. D. Noble and R. Agrawal, "Separations research needs for the 21st century," *Ind. Eng. Chem. Res.*, vol. 44, pp. 2887–2892, 2005.

- [88] I. E. Grossmann and G. Guillén-Gosálbez, "Scope for the application of mathematical programming techniques in the synthesis and planning of sustainable processes," *Comput. Chem. Eng.*, vol. 34, pp. 1365–1376, 2010.
- [89] Z. Kravanja, "Challenges in sustainable integrated process synthesis and the capabilities of an MINLP process synthesizer MipSyn," *Comput. Chem. Eng.*, vol. 34, pp. 1831–1848, 2010.
- [90] Z. Novak, Z. Kravanja, and I. E. Grossmann, "Simultaneous optimization model for multicomponent separation," *Comput. Chem. Eng.*, vol. 18, pp. S125–S129, 1994.
- [91] M. R. Eden, S. B. Jørgensen, R. Gani, and M. M. El-Halwagi, "A novel framework for simultaneous separation process and product design," *Chem. Eng. Process: Process Intensif.*, vol. 43, pp. 595–608, 2004.
- [92] W. Marquardt, S. Kossack, and K. Kraemer, "A framework for the systematic design of hybrid separation processes," *Chinese J. Chem. Eng.*, vol. 16, pp. 333–342, 2008.
- [93] A. C. Kokossis, A. Yang, M. Tsakalova, and T. C. Lin, "A systems platform for the optimal synthesis of biomass based manufacturing systems," *Comp. Aid. Chem. Eng.*, vol. 28, pp. 1105–1110, 2010.
- [94] A. C. Kokossis and A. Yang, "On the use of systems technologies and a systematic approach for the synthesis and the design of future biorefineries," *Comput. Chem. Eng.*, vol. 34, pp. 1397–1405, 2010.
- [95] S. D. Schaber, D. I. Gerogiorgis, R. Ramachandran, J. M. B. Evans, P. I. Barton, and B. L. Trout, "Economic analysis of integrated continuous and batch pharmaceutical manufacturing: A case study," *Ind. Eng. Chem. Res.*, vol. 50, pp. 10083–10092, 2011.
- [96] M. Rizwan, J. H. Lee, and R. Gani, "Optimal processing pathway for the production of biodiesel from microalgal biomass: A superstructure based approach," *Comput. Chem. Eng.*, vol. 58, pp. 305–314, 2013.
- [97] Z. Yuan, B. Chen, and R. Gani, "Applications of process synthesis: Moving from conventional chemical processes towards biorefinery processes," *Comput. Chem. Eng.*, vol. 49, pp. 217–229, 2013.
- [98] A. Kelloway and P. Daoutidis, "Process synthesis of biorefineries: Optimization of biomass conversion to fuels and chemicals," *Ind. Eng. Chem. Res.*, vol. 53, pp. 5261–5273, 2014.
- [99] J. Kim, S. M. Sen, and C. T. Maravelias, "An optimization-based assessment framework for biomass-to-fuel conversion strategies," *Energy Environ. Sci.*, vol. 6, pp. 1093–1102, 2013.

- [100] S. Mascia, P. L. Heider, H. Zhang, R. Lakerveld, B. Benyahia, P. I. Barton, R. D. Braatz, C. L. Cooney, J. M. B. Evans, T. F. Jamison, K. F. Jensen, A. S. Myerson, and B. L. Trout, "End-to-end continuous manufacturing of pharmaceuticals: Integrated synthesis, purification, and final dosage formation," *Angew. Chemie - Int. Ed.*, vol. 52, pp. 12359–12363, 2013.
- [101] Á. D. González-Delgado, V. Kafarov, and M. El-Halwagi, "Development of a topology of microalgae-based biorefinery: Process synthesis and optimization using a combined forward-backward screening and superstructure approach," *Clean Technol. Environ. Policy*, vol. 17, pp. 2213–2228, 2015.
- [102] L. Kong, S. M. Sen, C. A. Henao, J. A. Dumesic, and C. T. Maravelias, "A superstructure-based framework for simultaneous process synthesis, heat integration, and utility plant design," *Comput. Chem. Eng.*, vol. 91, pp. 68–84, 2016.
- [103] S. A. Papoulias and I. E. Grossmann, "A structural optimization approach in process synthesis—II: Heat recovery networks," *Comput. Chem. Eng.*, vol. 7, pp. 707–721, 1983.
- [104] D. H. S. Tay, D. K. S. Ng, H. Kheireddine, and M. M. El-Halwagi, "Synthesis of an integrated biorefinery via the C-H-O ternary diagram," *Clean Technol. Environ. Policy*, vol. 13, pp. 567–579, 2011.
- [105] R. Karuppiah and I. E. Grossmann, "Global optimization for the synthesis of integrated water systems in chemical processes," *Comput. Chem. Eng.*, vol. 30, pp. 650–673, 2006.
- [106] J. M. Douglas, *Conceptual design of chemical processes*. McGraw-Hill, 1988.
- [107] S. Lee and I. E. Grossmann, "New algorithms for nonlinear generalized disjunctive programming," *Comput. Chem. Eng.*, vol. 24, pp. 2125–2141, 2000.
- [108] S. Kolodziej, P. M. Castro, and I. E. Grossmann, "Global optimization of bilinear programs with a multiparametric disaggregation technique," *J. Glob. Optim.*, vol. 57, pp. 1039–1063, 2013.
- [109] Y. Gengel and B. M., *Thermodynamics: an engineering approach*. McGraw-Hill, 2002.
- [110] R. H. Perry, D. W. Green, and J. O. Maloney, *Perry's chemical engineers' handbook*. McGraw-Hill, 1997.
- [111] E. W and B. Lee, *Applied hydrocarbon thermodynamics*. Gulf Publishing, 1984.
- [112] R. Raman and I. E. Grossmann, "Integration of logic and heuristic knowledge in MINLP optimization for process synthesis," *Comput. Chem. Eng.*, vol. 16, pp. 155–171, 1992.
- [113] R. Raman and I. E. Grossmann, "Symbolic integration of logic in mixed-integer linear

- programming techniques for process synthesis," *Comput. Chem. Eng.*, vol. 17, pp. 909–927, 1993.
- [114] W. Z. Wu, C. A. Henao, and C. T. Maravelias, "A superstructure representation, generation, and modeling framework for chemical process synthesis," *AIChE J.*, vol. 62, pp. 3199–3214, 2016.
- [115] P. Pollak, *Fine Chemicals: The Industry and the business*. John Wiley & Sons, 2011.
- [116] R. Smith, "Chemical Process Design," in *Kirk-Othmer Encyclopedia of Chemical Technology*, John Wiley & Sons, 2015.
- [117] H. J. Huang, S. Ramaswamy, U. W. Tschirner, and B. V. Ramarao, "A review of separation technologies in current and future biorefineries," *Sep. Purif. Technol.*, vol. 62, pp. 1–21, 2008.
- [118] L. Burton, "By-products of milk," *Food Ind.*, vol. 9, pp. 571–575, 1937.
- [119] G. C. Inskeep, "Lactic acid from corn sugar," *Ind. Eng. Chem.*, vol. 44, pp. 1956–1966, 1952.
- [120] Y. Ling, D. R. G. Williams, C. J. Thomas, and A. P. J. Middelberg, "Recovery of poly-3-hydroxybutyrate from recombinant *Escherichia coli* by homogenization and centrifugation," *Biotechnol. Tech.*, vol. 11, pp. 409–412, 1997.
- [121] D. Horowitz and E. Brennan, "Methods for separation and purification of biopolymers," US Patent 20010006802, 2001.
- [122] K. L. Wasewar, "Separation of Lactic Acid : Recent Advances," *Chem. Biochem. Eng. Q.*, vol. 19, pp. 159–172, 2005.
- [123] N. Jacquell, C. W. Lo, Y. H. Wei, H. S. Wu, and S. S. Wang, "Isolation and purification of bacterial poly(3-hydroxyalkanoates)," *Biochem. Eng. J.*, vol. 39, pp. 15–27, 2008.
- [124] M. Koller, H. Niebelschütz, and G. Braunegg, "Strategies for recovery and purification of poly[(R)-3-hydroxyalkanoates] (PHA) biopolyesters from surrounding biomass," *Eng. Life Sci.*, vol. 13, pp. 549–562, 2013.
- [125] Intelligen, "SuperPro Designer." 2013.
- [126] American Petroleum Institute, "Monographs on refinery environmental control- management of water discharges," 1990.
- [127] W. L. McCabe, J. C. Smith, and P. Harriott, *Unit operations of chemical engineering*. McGraw-Hill, 1993.

- [128] C. Ambler, *Handbook of separation techniques for chemical engineers*. MacGraw-Hill, 1988.
- [129] W. McGregor, *Membrane separations in biotechnology*. Marcel Dekker Inc, 1986.
- [130] P. A. Belter, E. L. Cussler, and W.-S. Hu, *Bioseparations: downstream processing for biotechnology*. John Wiley & Sons, 1988.
- [131] R. Smith, *Chemical process design and integration*. John Wiley & Sons, 2005.
- [132] M. F. Doherty and M. F. Malone, *Conceptual design of distillation systems*. McGraw-Hill, 2001.
- [133] M. Assael and M. Magilliotou, *Introduction to unit operations*. Aristotle University Publications, 1988.
- [134] J. D. Seader, E. J. Henley, and D. K. Roper, *Separation process principles with applications using process simulators*. John Wiley & Sons, 2010.
- [135] K. M. Yenkie, W. Wu, and C. T. Maravelias, "Synthesis and analysis of separation networks for the recovery of intracellular chemicals generated from microbial-based conversions," *Biotechnol. Biofuels*, vol. 10, pp. 119-130, 2017.
- [136] C. Jones, K. Armstrong, I. Ingram, E. Bay, J. Dodson, and P. Abrantes, "Wider impacts: general discussion," *Faraday Discuss*, vol. 183, pp. 349-368, 2015.
- [137] K. J. Reddy and K. T. Karunakaran, "Purification and characterization of hyaluronic acid produced by *Streptococcus zooepidemicus*," *J. BioSci. Biotech*, vol. 2, pp. 173-179, 2013.
- [138] H. Zhou, J. Ni, W. Huang, and J. Zhang, "Separation of hyaluronic acid from fermentation broth by tangential flow microfiltration and ultrafiltration," *Sep. Purif. Technol.*, vol. 52, pp. 29-38, 2006.
- [139] L. Liu, Y. Liu, J. Li, G. Du, and J. Chen, "Microbial production of hyaluronic acid: current state, challenges, and perspectives," *Microb. Cell Fact.*, vol. 10, pp. 99-106, 2011.
- [140] V. Rangaswamy and D. Jain, "An efficient process for production and purification of hyaluronic acid from *Streptococcus equi* subsp. *zooepidemicus*," *Biotechnol. Lett.*, vol. 30, pp. 493-496, 2008.
- [141] J. B. Y. H. Behrendorff, C. E. Vickers, P. Chrysanthopoulos, and L. K. Nielsen, "2,2-Diphenyl-1-picrylhydrazyl as a screening tool for recombinant monoterpene biosynthesis," *Microb. Cell Fact.*, vol. 12, pp. 76-90, 2013.

- [142] C. Ignea, M. Pontini, M. E. Maffei, A. M. Makris, and S. C. Kampranis, "Engineering monoterpene production in yeast using a synthetic dominant negative geranyl diphosphate synthase," *ACS Synth. Biol.*, vol. 3, pp. 298–306, 2014.
- [143] H. Kiyota, Y. Okuda, M. Ito, M. Y. Hirai, and M. Ikeuchi, "Engineering of cyanobacteria for the photosynthetic production of limonene from CO₂," *J. Biotechnol.*, vol. 185, pp. 1–7, 2014.
- [144] F. Vararu, J. Moreno-Garcia, J. Moreno, M. Niculaua, B. Nechita, C. Zamfir, C. Colibaba, G.-D. Dumitru, and V. V. Cotea, "Minor volatile compounds profiles of 'Aligoté' wines fermented with different yeast strains," *Not. Sci. Biol.*, vol. 7, pp. 123–128, 2015.
- [145] E. Jongedijk, K. Cankar, J. Ranzijn, S. van der Krol, H. Bouwmeester, and J. Beekwilder, "Capturing of the monoterpene olefin limonene produced in *Saccharomyces cerevisiae*," *Yeast*, vol. 32, pp. 159–171, 2015.
- [146] M. Pourbafrani, G. Forgács, I. S. Horváth, C. Niklasson, and M. J. Taherzadeh, "Production of biofuels, limonene and pectin from citrus wastes," *Bioresour. Technol.*, vol. 101, pp. 4246–4250, 2010.
- [147] W. Wu and C. Maravelias, "Synthesis and techno-economic assessment of microbial-based processes for terpenes production," *Biotechnol. Biofuel.*, submitted, 2018.
- [148] F. García-Ochoa, V. E. Santos, J. A. Casas, and E. Gómez, "Xanthan gum: Production, recovery, and properties," *Biotechnol. Adv.*, vol. 18, pp. 549–579, 2000.
- [149] S. Rosalam and R. England, "Review of xanthan gum production from unmodified starches by *Xanthomonas compestris* sp.," *Enzyme Microb. Technol.*, vol. 39, pp. 197–207, 2006.
- [150] D. Kreyenschulte, R. Krull, and A. Margaritis, "Recent advances in microbial biopolymer production and purification," *Crit. Rev. Biotechnol.*, vol. 34, pp. 1–15, 2014.
- [151] S. D. Birajdar, S. Padmanabhan, and S. Rajagopalan, "Rapid solvent screening using thermodynamic models for recovery of 2,3-butanediol from fermentation by liquid-liquid extraction," *J. Chem. Eng. Data*, vol. 59, pp. 2456–2463, 2014.
- [152] Z. Li, H. Teng, and Z. Xiu, "Aqueous two-phase extraction of 2,3-butanediol from fermentation broths using an ethanol/ammonium sulfate system," *Process Biochem.*, vol. 45, pp. 731–737, 2010.
- [153] A. A. Koutinas, B. Yopez, N. Kopsahelis, D. M. G. Freire, A. M. de Castro, S. Papanikolaou, and I. K. Kookos, "Techno-economic evaluation of a complete bioprocess for 2,3-butanediol production from renewable resources," *Bioresour. Technol.*, vol. 204, pp. 55–64, 2016.

- [154] S. Jeon, D. K. Kim, H. Song, H. J. Lee, S. Park, D. Seung, and Y. K. Chang, "2,3-Butanediol recovery from fermentation broth by alcohol precipitation and vacuum distillation," *J. Biosci. Bioeng.*, vol. 117, pp. 464–470, 2014.
- [155] P. Shao and A. Kumar, "Recovery of 2,3-butanediol from water by a solvent extraction and pervaporation separation scheme," *J. Memb. Sci.*, vol. 329, pp. 160–168, 2009.
- [156] Z. L. Xiu and A. P. Zeng, "Present state and perspective of downstream processing of biologically produced 1,3-propanediol and 2,3-butanediol," *Appl. Microbiol. Biotechnol.*, vol. 78, pp. 917–926, 2008.
- [157] M. A. Abdel-Rahman, Y. Tashiro, and K. Sonomoto, "Recent advances in lactic acid production by microbial fermentation processes," *Biotechnol. Adv.*, vol. 31, pp. 877–902, 2013.
- [158] M. Bishai, S. De, B. Adhikari, and R. Banerjee, "A platform technology of recovery of lactic acid from a fermentation broth of novel substrate *Zizyphus oenophlia*," *Biotech.*, vol. 5, pp. 455–463, 2015.
- [159] Y. Li, A. Shahbazi, K. Williams, and C. Wan, "Separate and concentrate lactic acid using combination of nanofiltration and reverse osmosis membranes," *Appl. Biochem. Biotechnol.*, vol. 147, pp. 1–9, 2008.
- [160] K. Backhaus, M. Lochmüller, M. C. Arndt, O. Riechert, and G. Schembecker, "Knowledge-based conceptual synthesis of industrial-scale downstream processes for biochemical products," *Chem. Eng. Technol.*, vol. 38, pp. 537–546, 2015.
- [161] J. Benavides and M. Rito-Palomares, "Simplified two-stage method to B-phycoerythrin recovery from porphyridium cruentum," *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, vol. 844, pp. 39–44, 2006.
- [162] A. I. Clarkson, I. D. L. Bogle, and N. J. Titchener-Hooker, "Modelling and simulation of fractional precipitation-comparison with pilot plant data," *Comput. Chem. Eng.*, vol. 16, pp. 441–447, 1992.
- [163] M. Schenk, R. Gani, D. Bogle, and E. N. Pistikopoulos, "A hybrid modelling approach for separation systems involving distillation," *Chem. Eng. Res. Des.*, vol. 77, pp. 519–534, 1999.
- [164] T. Winkelkemper and G. Schembecker, "Purification performance index and separation cost indicator for experimentally based systematic downstream process development," *Sep. Purif. Technol.*, vol. 72, pp. 34–39, 2010.
- [165] S. Chhatre, R. Francis, A. R. Newcombe, Y. Zhou, N. Titchener-Hooker, and J. King, "Global sensitivity analysis for the determination of parameter importance in bio-manufacturing

- processes," *Biotechnol. Appl. Biochem.*, vol. 51, pp. 79-90, 2008.
- [166] D. M. Hamby, "A review of techniques for parameter sensitivity analysis of environmental models," *Env. Monit. Assess.*, vol. 32, pp. 135-154, 1994.
- [167] Y. Zhang, M. A. Dubé, D. D. McLean, and M. Kates, "Biodiesel production from waste cooking oil: 2. Economic assessment and sensitivity analysis," *Bioresour. Technol.*, vol. 90, pp. 229-240, 2003.
- [168] W. Wu, K. Yenkie, and C. T. Maravelias, "A superstructure-based framework for bio-separation network synthesis," *Comput. Chem. Eng.*, vol. 96, pp. 1-17, 2017.
- [169] K. M. Yenkie, W. Z. Wu, R. L. Clark, B. F. Pfleger, T. W. Root, and C. T. Maravelias, "A roadmap for the synthesis of separation networks for the recovery of bio-based chemicals: Matching biological and process feasibility," *Biotechnol. Adv.*, vol. 34, pp. 1362-1383, 2016.
- [170] M. R. Kılınç and N. V. Sahinidis, "Exploiting integrality in the global optimization of mixed-integer nonlinear programming problems with BARON," *Optim. Methods Softw.*, vol. 33, pp. 540-562, 2018.
- [171] E. Molina Grima, E. H. Belarbi, F. G. Acien Fernández, A. Robles Medina, and Y. Chisti, "Recovery of microalgal biomass and metabolites: Process options and economics," *Biotechnol. Adv.*, vol. 20, pp. 491-515, 2003.
- [172] J. F. Kennedy, J. M. S. Cabral, and J. S. Cabral, *Recovery processes for biological materials*. John Wiley & Sons, 1993.
- [173] C. M. Ambler, "The fundamentals of separation, including Sharples sigma value for predicting equipment performance," *Ind. Eng. Chem. Res.*, vol. 53, pp. 1404-1410, 1961.
- [174] S. Aikawa, A. Nishida, S.-H. Ho, J.-S. Chang, T. Hasunuma, and A. Kondo, "Glycogen production for biofuels by the euryhaline cyanobacteria *Synechococcus* sp. strain PCC 7002 from an oceanic environment," *Biotechnol. Biofuels*, vol. 7, pp. 88, 2014.
- [175] R. F. Castanha, L. A. S. Morais, A. P. Mariano, and R. T. R. Monteiro, "Comparison of two lipid extraction methods produced by yeast in cheese whey," *Braz. Arch. Biol. Technol.*, vol. 56, pp. 629-636, 2013.
- [176] Q. Li, W. Du, and D. Liu, "Perspectives of microbial oils for biodiesel production," *Appl. Microbiol. Biotechnol.*, vol. 80, pp. 749-756, 2008.
- [177] C. Kontoravdi, N. J. Samsatli, and N. Shah, "Development and design of bio-pharmaceutical processes," *Curr. Opin. Chem. Eng.*, vol. 2, pp. 435-441, 2013.

- [178] G. Towler and R. K. Sinnott, *Chemical engineering design, second edition: principles, practice and economics of plant and process design*. Butterworth-Heinemann, 2012.
- [179] A. M. Feist, D. C. Zielinski, J. D. Orth, J. Schellenberger, M. J. Herrgard, and B. O. Palsson, "Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*," *Metab. Eng.*, vol. 12, pp. 173–186, 2010.
- [180] K. K. Hong and J. Nielsen, "Metabolic engineering of *Saccharomyces cerevisiae*: A key cell factory platform for future biorefineries," *Cell. Mol. Life Sci.*, vol. 69, pp. 2671–2690, 2012.
- [181] A. Krivoruchko, V. Siewers, and J. Nielsen, "Opportunities for yeast metabolic engineering: Lessons from synthetic biology," *Biotechnol. J.*, vol. 6, pp. 262–276, 2011.
- [182] J. Nielsen, C. Larsson, A. van Maris, and J. Pronk, "Metabolic engineering of yeast for production of fuels and chemicals," *Curr. Opin. Biotechnol.*, vol. 24, pp. 398–404, 2013.
- [183] X. Zhang, C. J. Tervo, and J. L. Reed, "Metabolic assessment of *E. coli* as a Biofactory for commercial products," *Metab. Eng.*, vol. 35, pp. 64–74, 2016.
- [184] H. Zhang, Y.-T. Cheng, T. P. Vispute, R. Xiao, and G. W. Huber, "Catalytic conversion of biomass-derived feedstocks into olefins and aromatics with ZSM-5: the hydrogen to carbon effective ratio," *Energy Environ. Sci.*, vol. 4, pp. 2297–2310, 2011.
- [185] E. Scott, F. Peter, and J. Sanders, "Biomass in the manufacture of industrial products-the use of proteins and amino acids," *Appl. Microbiol. Biotechnol.*, vol. 75, pp. 751–762, 2007.
- [186] T. Werpy and G. Petersen, "Top value added chemicals from biomass: Volume I - results of screening for potential candidates from sugars and synthesis Gas," *Program.*, vol. 1, pp. 69–80, 2004.
- [187] J. E. Holladay, J. F. White, J. J. Bozell, and D. Johnson, "Top value-added chemicals from biomass - Volume II - results of screening for potential candidates from biorefinery lignin," *Pacific Northwest Natl. Lab.*, vol. II, pp. 87–95, 2007.
- [188] E. de Jong, A. Higson, P. Walsh, and M. Wellisch, "Bio-based chemicals - value added products from biorefineries," *IEA Biotech.*, 2012.
- [189] N. M. Matsumoto, D. C. González-Toro, R. T. Chacko, H. D. Maynard, and S. Thayumanavan, "Synthesis of nanogel–protein conjugates," *Polym. Chem.*, vol. 4, pp. 2464–2474, 2013.
- [190] Y. S. Jang, B. Kim, J. H. Shin, Y. J. Choi, S. Choi, C. W. Song, J. Lee, H. G. Park, and S. Y. Lee, "Bio-based production of C2–C6 platform chemicals," *Biotechnol. Bioeng.*, vol. 109, pp. 2437–2459, 2012.

- [191] J. Moncada, J. A. Posada, and A. Ramirez, "Early sustainability assessment for potential configurations of integrated biorefineries. Screening of bio-based derivatives from platform chemicals," *Biofuels, Bioprod. Biorefining*, vol. 6, pp. 246–256, 2015.
- [192] J. Becker, A. Lange, J. Fabarius, and C. Wittmann, "Top value platform chemicals: Bio-based production of organic acids," *Curr. Opin. Biotechnol.*, vol. 36, pp. 168–175, 2015.
- [193] J.-V. Bomtempo, F. Chaves Alves, and F. de Almeida Oroski, "Developing new platform chemicals: what is required for a new bio-based molecule to become a platform chemical in the bioeconomy?," *Faraday Discuss.*, vol. 202, pp. 213–225, 2017.
- [194] J. S. Golden, R. B. Handfield, J. Daystar, and T. E. McConnell, "An economic impact analysis of the U.S. biobased products industry: A report to the Congress of the United States of America," *Ind. Biotechnol.*, vol. 11, pp. 201–209, 2015.
- [195] G. Stephanopoulos and A. W. Westerberg, "Synthesis of optimal process flowsheets by an infeasible decomposition technique in the presence of functional non-convexities," *Can. J. Chem. Eng.*, vol. 53, pp. 551–555, 1975.
- [196] USEPA, "HPV chemical hazard characterizations," *High Production Volume Information System*, 2014. [Online]. Available: <https://www.epa.gov/HPV>.
- [197] USEPA, "1994 HPV additions chemical list," 2006.
- [198] USEPA, "Additional chemicals sponsored under the HPV Challenge Program," 2006.
- [199] USEPA, "1990 HPV Challenge Program chemical list," 2006.
- [200] USEPA, "HPV Priority List," 2016. [Online]. Available: <https://www.epa.gov/HPV/pubs/update/hpvchmlt.htm>.
- [201] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa, "KEGG: Kyoto encyclopedia of genes and genomes," *Nucleic Acids Res.*, vol. 27, pp. 29–34, Jan. 1999.
- [202] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, "KEGG as a reference resource for gene and protein annotation," *Nucleic Acids Res.*, vol. 44, pp. D457–D462, Jan. 2016.
- [203] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, "KEGG: New perspectives on genomes, pathways, diseases and drugs," *Nucleic Acids Res.*, vol. 45, pp. D353–D361, Jan. 2017.
- [204] R. Caspi, R. Billington, L. Ferrer, H. Foerster, C. A. Fulcher, I. M. Keseler, A. Kothari, M.

- Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D. S. Weaver, and P. D. Karp, "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases," *Nucleic Acids Res.*, vol. 44, pp. 471–480, 2016.
- [205] J. D. Orth, T. M. Conrad, J. Na, J. A. Lerman, H. Nam, A. M. Feist, and B. Palsson, "A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism-2011," *Mol. Syst. Biol.*, vol. 7, pp. 535-542, 2011.
- [206] M. L. Mo, B. Palsson, and M. J. Herrgård, "Connecting extracellular metabolomic measurements to intracellular flux states in yeast," *BMC Syst. Biol.*, vol. 3, pp. 37-48, 2009.
- [207] J. Schellenberger, J. O. Park, T. M. Conrad, and B. T. Palsson, "BiGG: A Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions," *BMC Bioinformatics*, vol. 11, pp. 213-220, 2010.
- [208] C. S. Thakur, M. E. Brown, J. N. Sama, M. E. Jackson, and T. K. Dayie, "Growth of wildtype and mutant *E. coli* strains in minimal media for optimal production of nucleic acids for preparing labeled nucleotides," *Appl. Microbiol. Biotechnol.*, vol. 88, pp. 771–779, 2010.
- [209] F. Sherman, "Getting started with yeast," *Methods Enzymol.*, vol. 194, pp. 3–21, 1991.
- [210] P. Pharkya, A. P. Burgard, and C. D. Maranas, "OptStrain: A computational framework for redesign of microbial production systems," *Genome Res.*, vol. 14, pp. 2367–2376, 2004.
- [211] E. T. Papoutsakis, "Equations and calculations for fermentations of butyric acid bacteria," *Biotechnol. Bioeng.*, vol. 67, pp. 813–826, 2000.
- [212] W. Wu, M. R. Long, X. Zhang, J. L. Reed, and C. T. Maravelias, "A framework for the identification of promising bio-based chemicals," *Biotechnol. Bioeng.*, accepted, 2018.
- [213] D. Humbird, R. Davis, L. Tao, C. Kinchin, D. Hsu, A. Aden, P. Schoen, J. Lukas, B. Olthof, M. Worley, D. Sexton, and D. Dudgeon, "Process design and economics for biochemical conversion of lignocellulosic biomass to ethanol: Dilute-acid pretreatment and enzymatic hydrolysis of corn stover," *NREL*, 2011.
- [214] ICIS, "ICIS Chemical Business," 2016. [Online]. Available: <https://www.icis.com/>.
- [215] USEPA, "Chemical Data Access Tool (CDAT)," 2013. [Online]. Available: https://java.epa.gov/oppt_chemical_search/.
- [216] USEPA, "2006 IUR," 2006. [Online]. Available: <https://www.epa.gov/chemical-data-reporting/downloadable-2006-iur-public-database>.

- [217] "Essential Chemical Industry." [Online]. Available:
<http://www.essentialchemicalindustry.org/chemicals.html>.
- [218] Innovest, "Overview of the chemicals industry," 2007.
- [219] H. Wittcoff, B. Reuben, and J. Plotkin, *Industrial organic chemicals*. John Wiley & Sons, 2012.
- [220] L. Goldman, "Toxic chemicals and pesticides," in *Stumbling Toward Sustainability*, Environmental Law Institute, 2002.
- [221] FRED, "Industrial Production Organic Chemicals." [Online]. Available:
<https://fred.stlouisfed.org/series/IPG32511A9N#0>.
- [222] R. Gani, "Integrated chemical product design," *CAPEC*, 2007.