Imperial College London

PhD Thesis

Model-based solvent selection for pharmaceutical process development

Imperial College London

Department of Chemical Engineering

Doctoral Candidate

Mohamad Muhieddine

Supervisors

Claire S. Adjiman, Amparo Galindo, and Alan Armstrong

23 March 2023

Statement of Originality

I, Mohamad Muhieddine, declare that this work is my own and that work from others is appropriately referenced.

Copyright Declaration

Data Statement

Data underlying this work can be accessed on Zenodo at 10.5281/zenodo.7252184, and used under the Creative Commons Attribution licence.

Acknowledgements

I would like to thank my supervisors Professors Claire Adjiman, Amparo Galindo and Alan Armstrong for their impeccable guidance and constant support during my PhD journey: I am lucky to have worked with pioneering figures in the fields of Process Systems Engineering, Molecular Thermodynamics and Chemistry, and I will forever be grateful for the knowledge I gained from your expertise and insights. I would also like to thank Dr Shekhar Viswanath, our industrial collaborator at Eli Lilly, for the interesting discussions and valuable feedback on my research work which made my PhD findings more industrially relevant.

It was a pleasure being part of the lively and friendly Molecular Systems Engineering (MSE) group at Imperial College London. Many thanks to all MSE group members for making the PhD experience quite enjoyable. Special thanks to Eliana, Redzuan, Griffin, Suela, Lauren and Tanuj for the fruitful discussions from which I learned a lot and developed the skill of addressing problems from multiple perspectives. I wish my MSE colleagues all the best for the future and cannot wait to hear about your success stories!

Finally, no words are sufficient to describe my gratefulness to my precious parents, Hassan and Nouhad, and adorable siblings, Tarek and Dima, for always being there for me: Your love, care and support are the main ingredients of my success. I hope I always make you proud.

Abstract

Solvents play a key role in the manufacturing of pharmaceutical products as they are extensively used to accelerate synthetic reactions, enable separation and purification, and facilitate drug product formulation. The production of active pharmaceutical ingredients (APIs) is a multi-step process involving several reaction and workup steps in which large amounts of solvents are consumed. This makes the pharmaceutical industry a very wasteful chemical sector and highlights the need for systematic tools to enhance the resource efficiency of its processes. Recently, there has been growing interest in incorporating green chemistry principles in product design and development to enhance the sustainability of chemical manufacturing. In particular, solvent selection is a promising research area within the chemistry and engineering communities, given the many solvent-related contributions to process performance, including mass utilisation, energy consumption and process economics.

Solvent selection is a difficult and complex design problem that entails molecular-level decisions, such as determining the solvent identities and the compositions of mixtures if mixed solvents are considered, together with process performance objectives, which are often competing. In current practice, most pharmaceutical companies develop in-house solvent selection guides to choose solvents based on physico-chemical properties and safety, health and environment characteristics with the aim to reduce process costs and environmental impact. However, these methods are mostly based on heuristic approaches or time-consuming experimental investigations that often lead to sub-optimal designs and fail to account for the integrated nature of the solvent selection problem.

A novel solvent selection approach based on computer-aided mixture/blend design (CAM^{b}D) is proposed to design integrated pharmaceutical processes and evaluate the process performance of pharmaceutical synthesis routes. Predictive thermodynamic models are used to integrate property prediction within process modelling, and advanced optimisation techniques are employed to search the vast design space of potential solvents and process conditions in order to identify the most promising design options. The CAM^{b}D approach is used to optimise the solvent identities, mixture composition and process conditions in: 1) integrated synthesis and crystallisation processes, and 2) end-to-end drug substance manufacturing processes, based on key performance indicators (KPIs) that quantify resource efficiency and product quality. The one-step synthesis of mefenamic acid from 2,3-dimethylaniline and 2-chlorobenzoic acid is used as a case study to illustrate the use of CAM^{b}D in pharmaceutical process design. The CAM^{b}D approach

generates different designs by considering a variety of solvent design spaces and performance objectives. Furthermore, multi-objective optimisation CAM^{b}D problems are formulated to explore the trade-offs between competing KPIs, such as solvent utilisation and process safety, or energy consumption and process yield, in order to identify best-compromise solutions. An important feature of the proposed approach is that comprehensive design specifications, such as the miscibility of the chosen reaction and crystallisation solvents with the wash solvent in the end-to-end process, can be embedded in the mathematical formulation, ensuring that only practical designs are obtained.

In addition to its use in integrated molecular and process design, the proposed CAM^{b}D approach can be deployed to identify the optimal synthesis route of a pharmaceutical compound based on process performance metrics quantifying resource efficiency, product quality and solvent cost. The two-step synthesis of 4-nitrophenol (NP) via two reaction pathways is used as a case study to illustrate the potential of CAM^{b}D in pharmaceutical process route selection.

The work presented in this thesis constitutes a unique scientific contribution to the area of model-based solvent selection for drug substance manufacturing. For the first time, a CAM^{b}D-based approach is developed and deployed to identify promising solvent choices and operating conditions for integrated, end-to-end drug substance manufacturing processes, while focusing on mixture thermodynamics, i.e., species solubility, and considering a range of KPIs that quantify product and process performance within single and multi-objective design formulations. Furthermore, for the first time, CAM^{b}D is deployed to evaluate synthesis routes based on process performance, i.e., process route evaluation, while using simplified thermodynamic models and considering process-related metrics such as process efficiency and product quality. The model-based tool presented in this PhD thesis is relevant to streamline experiments and guide solvent selection and process design during early-stage pharmaceutical process development.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and Motivation

The pharmaceutical industry has been expanding in recent decades to meet the increasing demand for effective medicines accompanying the unprecedented growth in human population [1, 2]. Traditionally, pharmaceutical process development has primarily focused on manufacturing drug products with guaranteed quality and consistency, with far less emphasis on process-related metrics such as product yield and solvent consumption [3]. For this reason, pharmaceutical processes are often accompanied by significant amounts of waste generated from the multi-step syntheses of complex active pharmaceutical ingredients (APIs), rendering many of the developed process routes unsustainable [1]. Several environmental regulations have been issued to enforce the replacement of conventional wasteful approaches to medicines manufacturing with environmentally-benign methods [4, 5].

In order to promote sustainability in pharmaceutical manufacturing, it is useful to draw upon the principles of green chemistry to develop novel processes [6]. Green chemistry is defined as *"the invention, design, and application of chemical products and processes to reduce or to eliminate the use and generation of hazardous substances"* [7]. There are numerous benefits of incorporating green chemistry in process design and development, such as minimising waste generation, improving reaction performance, developing safe reaction schemes, deploying sustainable feed-stocks and catalytic technologies and reducing energy consumption [3, 6].

Over the last two decades, the integration of green chemistry criteria in process route

selection has become an established practice in pharmaceutical companies for mitigating the environmental burden associated with drug manufacturing [1, 8]. In particular, the American Chemical Society Green Chemistry Institute Pharmaceutical Roundtable (ACS GCIPR) was established in 2005 to facilitate the integration of green chemistry and engineering approaches to sustainable medicines manufacturing [9].

Solvents are extensively used in the pharmaceutical industry for a variety of processing tasks, such as in chemical reactions, separations and formulations [10, 11]. Solvents can significantly influence chemical reaction kinetics and mixture thermodynamics. From a kinetics perspective, solvents that stabilise the activated complex, which is the molecular configuration of the reaction species at the transition state, relative to the reactants decrease the activation energy barrier and accelerate the reaction kinetics. On the other hand, solvents that stabilise the reactants relative to the activated complex increase the activation energy barrier and decelerate the reaction kinetics. When multiple reactions are involved in a reaction scheme, the multiple activated complexes can be stabilized to different extents based on solvent choice, hence varying product selectivity [12]. From a thermodynamics perspective, the solubility of a solute in a specific solvent depends on the activity coefficient of that solute, which is a function of solute-solvent interactions or the intermolecular forces between the two species [13]. This indicates that thermodynamics-based performance indicators that are relevant to pharmaceutical processes, such as crystal yield and impurity rejection, are strong functions of solvent choice.

Given the key impact of the solvent on process performance [6], solvents represent a major green chemistry theme. It is estimated that solvents used in API production contribute to more than 80% of the overall mass utilisation [14], approximately 60% of the total energy consumption and 50% of greenhouse gas emissions [15]. From a mass perspective, solvent use constitutes an important component of the Environmental Factor (E-factor), a common green chemistry metric defined as the mass of waste generated per mass of desired product formed. In pharmaceutical processing, the E-factor typically ranges between 25 and 100 [16] due to the use of solvents and stoichiometric reagents in a large number of synthetic steps that are typically quite complex and include several inter-stage isolation and purification units [17]. From an energy perspective, solvent exchange and solvent recovery are usually achieved by

energy-intensive distillation operations; this increases process costs and further compounds the environmental impact [11, 18]. This makes the pharmaceutical industry one of the most wasteful chemical sectors [19] and highlights the need for systematic solvent selection tools to develop resource-efficient processes [20]. In fact, large pharmaceutical companies such as Pfizer [21], Sanofi [22] and GSK [23, 24] have established their own solvent selection guides to promote the incorporation of green chemistry principles in their process development workflows. Additionally, a collaborative effort between the pharmaceutical industry and academia has led to the establishment of the CHEM21 solvent selection guide [25] which offers a preliminary sustainability assessment of solvents based on Safety, Health and Environment (SHE) criteria using physical properties and toxicity data.

In addition to pure solvents, solvent mixtures are widely used in the process industry to accelerate chemical reactions [26, 27] and facilitate separation processes, including extraction [28, 29] and crystallisation [30, 31]. In product design, optimal performance is commonly achieved by deploying mixtures of compounds, such as those routinely carried out in pesticide formulations or crude oil blends, since mixtures can exhibit properties equivalent or superior to those of pure compounds [32, 33, 34]. For example, Granberg and Rasmuson [35] studied paracetamol solubility in water-acetone binary mixtures and demonstrated that at 303.15 K, a 30:70 (by mass) mixture of water and acetone exhibited maximal paracetamol solubility, specifically a five-fold solubility increase relative to that in pure acetone, despite the fact that paracetamol is poorly soluble in water. Furthermore, mixtures comprised of up to 75% by mass of water were shown to exhibit at least as high a solubility as pure acetone. Such nonlinear behaviour can be attributed to the non-ideal thermodynamics of ternary mixtures [34].

The traditional practice in the pharmaceutical industry is to select solvents based on heuristics or time-consuming and costly experiments [36]. These trial-and-error approaches do not allow for a comprehensive exploration of the design space and may lead to sub-optimal solutions that fail to account for the interdependence between solvent choice and process performance [37]. Furthermore, while the aforementioned solvent selection guides constitute an accessible solvent selection approach to practitioners, the vast amount of information contained within these guides makes it challenging to make well-informed decisions on product development. To

address this limitation, Diorazio et al. [38] developed a computer-aided solvent selection tool that considers process requirements, solvent properties and environmental criteria, hence offering the opportunity to identify solvents that are more novel and sustainable than traditional solvents and removing the human decision-making aspect from the design process. The proposed tool couples experimental data with property prediction to construct a Principal Component Analysis (PCA) model that allows the user to visually explore the correlation between experimental and computed variables. However, a main limitation of this approach is that it does not take into consideration the interrelationship between physical properties and process performance, which is a key requirement for the development of a practical solvent selection approach. It is especially important to adopt a holistic approach to solvent selection for pharmaceuticals, given that drug development is a complex activity involving multiple interrelated key performance indicators (KPIs) that need to be optimised [39, 40].

The application of modelling concepts and techniques, such as those from Process Systems Engineering (PSE), offers a possible route to reduce the environmental impact of pharmaceutical processes. PSE is an interdisciplinary field that uses systematic, model-based tools for the simulation, design, control and optimisation of chemical process systems at multiple production scales [41, 42]. PSE methods offer a data-rich environment that can enhance process understanding, improve system performance and identify optimal processing materials and operating conditions for integrated process design [20]. Although PSE methods have been used successfully to advance the bulk-chemicals industry, they have not yet been fully deployed within the pharmaceutical and fine-chemicals industries [43]. Papadakis et al. [20] developed an integrated framework for pharmaceutical process development that uses systematic model-based approaches to improve process understanding and explore opportunities for continuous processing, hence facilitating decision-making during the early stage development of small-molecule APIs. However, despite the data-rich environment this framework creates, it does not address the solvent selection and design problem from a holistic perspective, as solvents are first designed based on target molecular properties and the best candidates are then assessed based on process performance. This sequential method cannot fully capture the dependence of process improvement on solvent properties, and accordingly, incorporating an integrated solvent and process design formulation

within a process development framework would be invaluable. In the context of continuous pharmaceutical manufacturing, the literature includes several examples on the use of model-based techniques to design continuous processes (reaction and workup steps) to produce a range of pharmaceutical compounds, such as aliskiren [44], ibuprofen [45] , artemisinin [46], diphenhydramine [47], nevirapine [48] and atropine [49]. Diab et al. [19] performed a plant-wide techno-economic evaluation and systematic selection of separation solvents for the continuous production of ibuprofen, artemisinin and diphenhydramine. In addition to regulatory and cost considerations, thermodynamic modelling using UNIFAC [50] and NRTL [51] was used to study multi-component mixture phase equilibria and assess separation solvent candidates based on API recovery and mass efficiency. However, the aforementioned studies focused on a limited set of solvent choices for the design of separation processes, which limits the possibility of exploring different solvent classes and identifying high-performance solvents that could be used across multiple process steps; for example, identifying potential solvents for integrated synthesis and crystallisation would promote the design of telescoped processes with significantly lower costs and environmental impact.

An increasingly interesting and industrially-relevant area in PSE is the use of optimisation-based solvent selection techniques that consider multiple performance metrics such as product yield, process economics and environmental impact [20], hence capturing the synergies between chemistry and engineering and promoting green production processes [52]. The general class of computer-aided molecular design (CAMD) approaches, including computer-aided mixture/blend design (CAM^{b}D), is concerned with the identification of compounds or blends that best achieve a specified performance objective, given a set of atom groups or molecules and a predefined set of target properties [53]. In tools such as CAM^{b}D, predictive thermodynamic models are embedded to integrate property prediction and process performance modelling. Efficient optimisation algorithms are used to search a vast space of design options and identify promising solvent choices and process operating conditions [54].

CAM^{b}D has been successful in identifying optimal solvent mixtures for a wide range of applications, including pharmaceutical manufacturing. Jonuzaj et al. [34, 55] developed a CAM^{b}D formulation to select optimal solvent mixtures for separation processes, including

crystallisation, in which the number, identity and composition of mixture ingredients were simultaneously identified. Watson et al. [56, 57] formulated a solvent-mixture design problem to identify the optimal solvents, mixture composition and crystallisation temperatures for the hybrid cooling and anti-solvent crystallisation of APIs. Building on the latter work, Jonuzaj et al. [58] proposed a comprehensive mixture design formulation for the integrated crystallisation and isolation of pharmaceutical compounds that includes mixture property constraints, such as the miscibility of the crystallisation and wash solvents, for both purification stages. A main advantage of considering the simultaneous design of processes and solvents is that it allows linking molecular-level decisions to process-level decisions, hence offering a more comprehensive assessment of process performance [59]. Studies on model-based solvent selection for pharmaceuticals have also considered other process structures, such as integrated crystallisation and distillation [54], reactive crystallisation [60], and integrated synthesis, liquid-liquid extraction and crystallisation [61]. However, despite the ongoing efforts in developing holistic solvent selection approaches for pharmaceutical manufacturing tasks, a tool that can identify high-performance solvents for end-to-end processes (i.e., including synthesis/reaction as well as crystallisation and isolation) is still lacking. The implementation of such an integrated approach would be of utmost importance as it would deliver designs for safer and greener processes by reducing solvent use and eliminating the need for resource-intensive solvent swaps. Furthermore, a process-wide solvent selection approach can help screen synthetic routes based on process performance metrics (mass utilisation, energy consumption, process economics, etc.) during early-stage process development.

1.2 Scope

The aim of this thesis is to develop and assess a novel CAM^{b}D formulation for the integrated synthesis, crystallisation and isolation of pharmaceutical compounds to support the development of high-performance processes. The solvent mixture, its composition and the process conditions that optimise one or more KPIs are simultaneously identified. The formulation developed is used to evaluate the process performance of pharmaceutical synthesis routes. It is also important to mention that although several phenomena need to be considered when quantifying solvent effects on KPIs [62], the focus of this PhD thesis is on thermodynamic driving forces, such as

species solubility, in order to shortlist a set of solvent candidates that can then be evaluated based on more detailed process analysis. To illustrate this point further, consider the traditional drug process development workflow as shown in Figure 1.1, which is adapted from Cervera et al. [63]. Solvent selection is a critical activity which typically relies heavily on experimental approaches and is addressed in a sequential manner early on in the workflow, initially focusing on finding a suitable solvent/solvent mixture for one process unit at a time. This can lead to issues of solvent compatibility between different unit operations and to sub-optimal designs when considered from a process-wide perspective. There is growing evidence that integration can bring significant benefits, as studies of isolation have demonstrated [58, 64]. Accordingly, in this thesis, an alternative approach to solvent selection in which CAM^{b}D is first used to identify promising candidate solvents based on the simultaneous consideration of several unit operations, thereby assessing solvent performance within an integrated process perspective, is proposed. Initially, KPIs that quantify solvent effects on process performance from a thermodynamics perspective are selected. Then, process modelling, in the form of mass and energy balances, and property prediction tools are used to relate thermodynamic molecular properties, such as species solubility, to process performance metrics. This model is used to select solvents from a list of commonly used and available solvents in the pharmaceutical industry and to identify process conditions that optimise the selected KPIs, resulting in a list of solvents that can be further assessed using criteria not included in the design problem. The solvents selected in this way can be taken forward to subsequent stages of process development, as has been considered extensively in the literature, including the design of crystallisation systems [65, 66, 67, 68], the design of end-to-end pharmaceutical processes, [20, 43, 44, 69, 70, 71], the development of optimal control strategies for pharmaceutical manufacturing [72, 73, 74] and the consideration of economic criteria [47, 66, 67, 75, 76, 77]. It is worthy to mention that the Drug Discovery and Development step shown in the workflow in Figure 1.1 entails identifying a list of potential reaction routes for the synthesis of the target drug product; these routes can then be evaluated based on process-related metrics, such as solvent consumption, as will be demonstrated in Chapter 5.

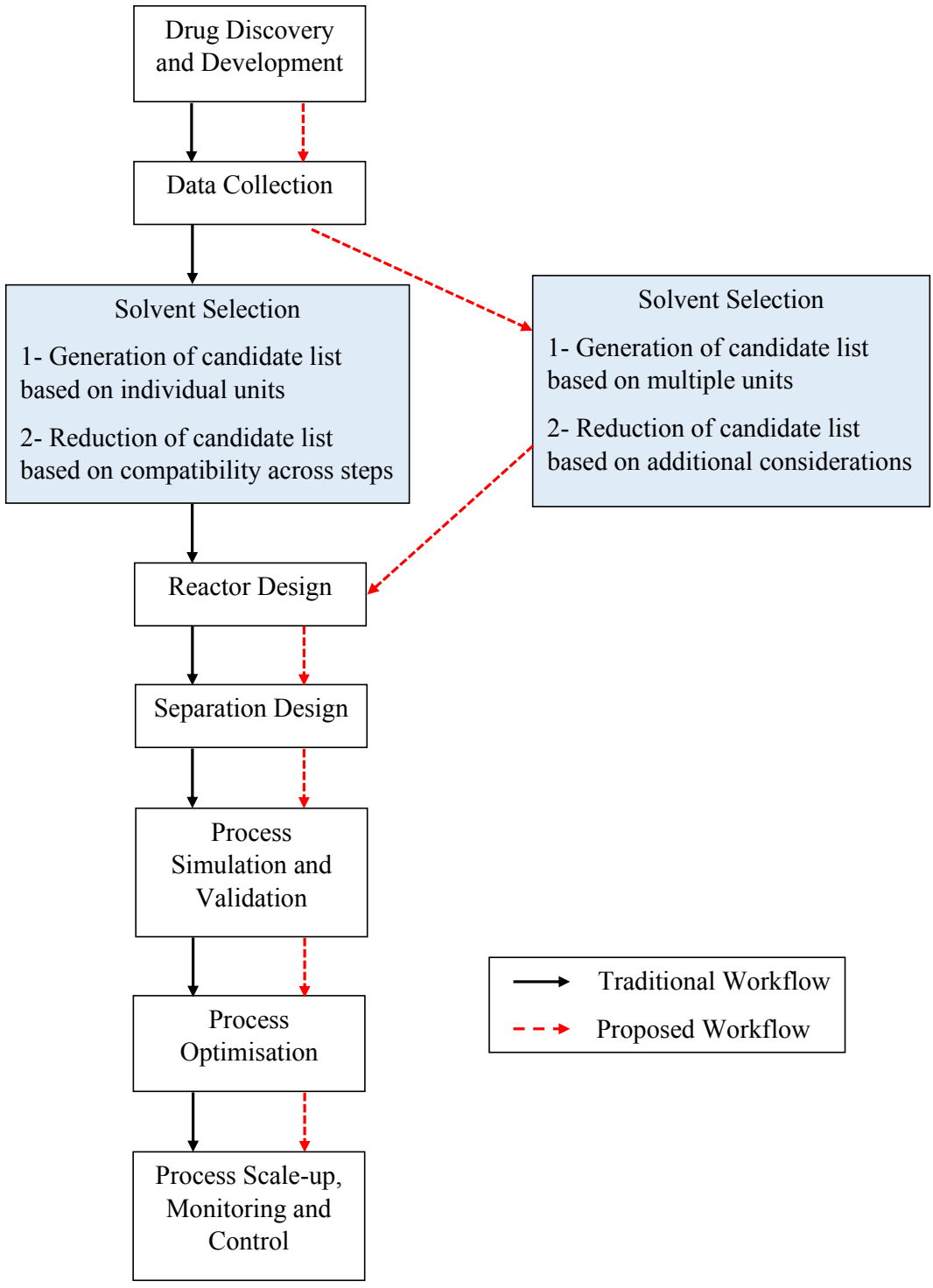

Figure 1.1: The traditional and proposed workflows for drug product development, adapted from Cervera et al. [61].

1.3 Thesis Outline

The thesis is structured as follows:

- Chapter 2 provides an overview of the different classes of CAMD approaches, including a review of key research works on computer-aided solvent and process design, with a focus on pharmaceutical manufacturing applications, and chemical route selection, and presents key concepts in mathematical optimisation.

- Chapter 3 introduces a CAM^{b}D formulation for identifying the optimal solvent/anti-solvent mixtures, mixture composition and process conditions for integrated synthesis and crystallisation processes.

- Chapter 4 presents an extended CAM^{b}D formulation that includes modelling the isolation steps that follow crystallisation, namely filtration, washing and drying, hence identifying the optimal solvent/anti-solvent mixtures, mixture composition and process conditions for end-to-end drug substance manufacturing processes.

- Chapter 5 demonstrates the use of the proposed CAM^{b}D tool to evaluate two-step process routes based on sustainability and cost metrics.

- Chapter 6 draws conclusions on the potential of the proposed CAM^{b}D approach to solvent selection, process design and route screening, and discusses future research directions.

Chapter 2

Background Theory

2.1 Introduction

This chapter introduces computer-aided molecular design (CAMD) as a model-based tool for solvent selection. The main classes of CAMD are presented with example applications, and some of the important property prediction tools used to estimate molecular and mixture properties in CAMD are covered. Furthermore, a literature survey on model-based solvent selection/process design and route selection is presented, highlighting the gap in developing a systematic solvent selection approach for pharmaceutical process design that incorporates process-wide KPIs and that could be used for synthetic route screening. Finally, key concepts in mixed-integer and multi-objective optimisation are discussed.

2.2 Computer-Aided Molecular, Mixture and Process Design

2.2.1 Computer-aided molecular design (CAMD)

As mentioned in Chapter 1, methods from PSE can be deployed for the systematic design of sustainable materials and processes. In the pharmaceutical industry, implementing model-based tools from PSE can promote an easier integration of sustainability metrics in process development [41, 78]. This can lead to the discovery of environmentally-benign process routes

that replace traditional, wasteful manufacturing strategies. In particular, tools that can, in reasonable time and cost, identify solvents that offer enhanced process performance in terms of process economics, yield and environmental impact would significantly accelerate drug discovery and development. The different computer-aided tools that have been developed to design or select molecules that exhibit enhanced performance are classified as computer-aided molecular design (CAMD) methods [79]. The main objectives of a CAMD methodology are: 1) optimising the physical properties of molecules [80], and/or 2) optimising process performance, such as maximising productivity [62] or minimising process costs [81]. This section gives an overview of property-based CAMD.

CAMD is a reverse engineering method that generates molecules from a set of structural building blocks, evaluates them based on a specified set of target performance measures and identifies promising candidates that best satisfy these criteria [53]. The generated molecules can be evaluated through property prediction methods, such as quantitative structure-property relationships (QSPRs), and their performance can be further validated by experiments [82].

CAMD methods can be broadly categorised in two groups, namely generate-and-test and mathematical optimisation approaches. In the generate-and-test approach, all chemically-feasible compounds are enumerated *in silico* and their properties evaluated using predictive models, resulting in a ranked list of candidates. In the mathematical optimisation approach, the CAMD problem is formulated as an optimisation problem, and an optimisation algorithm is used to guide the search for optimal molecular structures that maximise or minimise some performance metric, subject to a set of constraints on molecular and application-related aspects. In this sense, optimisation-based CAM^{b}D resembles automated molecular synthesis, where an initial molecular structure is iteratively modified until an optimal design is achieved [83].

Generate-and-test approach The generate-and-test approach for CAMD was first developed for solvent design and selection [28, 29] and consists of two steps, namely "generation" and "testing". The main idea of this approach is to use a set of structural groups or building blocks to design molecules that match a set of pre-defined target properties. In this sense, CAMD is classified as the reverse problem of property prediction, where molecular properties are estimated based on a given molecular structure. In the first step (generation), a set of functional groups is

used to generate a large number of molecular structures based on a set of molecular feasibility rules. In the second step (testing), the properties of the generated structures are computed and assessed to form a ranked list of candidate molecules [12]. The reader is referred to section 2.2.4 for more details on property prediction in CAMD.

Mathematical optimisation approach Although generate-and-test methods can be effective in designing solvents out of a small number of structural groups, they are not suitable for problems with a large solvent design space as this can lead to a combinatorial explosion/large computational time due to the enumeration of many solvent combinations. Enumeration methods are also not suitable when there are many continuous variables, such as temperature and composition, or if process performance is expensive to evaluate; for example, process performance evaluation may require solving a mathematical optimisation problem. In such cases, optimisation-based methods are more advantageous than enumeration approaches. A general mathematical representation of a CAMD problem, which is usually a Mixed-Integer Nonlinear Program (MINLP), is given as follows [79]:

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & f(\mathbf{x},\mathbf{y}) \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{x},\mathbf{y}) = \mathbf{0} \\ & \mathbf{g}(\mathbf{x},\mathbf{y}) \leq \mathbf{0} \\ & \mathbf{x} \in [\mathbf{x}^L, \mathbf{x}^U] \subset \mathbb{R}^n \\ & \mathbf{y} \in \{0,\,1\}^u \end{aligned} \tag{2.1}$$

where $f(\mathbf{x},\mathbf{y})$ is a performance objective function that needs to be optimised, subject to a set of equality constraints, $\mathbf{h}(\mathbf{x},\mathbf{y}) = \mathbf{0}$, that represent physical property relationships, and inequality constraints, $\mathbf{g}(\mathbf{x},\mathbf{y}) \leq \mathbf{0}$, that represent feasibility rules and design specifications. $\mathbf{x}$ is an n-dimensional vector of continuous variables representing physical properties or process variables (temperature, mixture composition, etc.) and $\mathbf{y}$ is a u-dimensional vector of binary variables representing the identities of the functional groups in the molecules, when the CAMD problem involves designing molecules from a set of functional groups, or the identities of the molecules, when the CAMD problem involves selecting molecules from a predefined list of

candidate species.

CAMD has been applied to the design and selection of solvents/solvent mixtures for chemical reactions [26, 84, 85, 86, 87] and separation processes [30, 34, 55, 88, 89], as well as the design of refrigerants [90], heat transfer fluids [91], and polymer blends [92]. The reader is referred to Austin et al. [93] and Ng et al. [94] for an overview of the tools, challenges and solution approaches of CAMD problems, and to Chemmangattuvalappil et al. [95] for an overview of recent developments in CAMD for solvent design.

2.2.2 Computer-aided mixture/blend design (CAM^{b}D)

In addition to designing pure compounds, property-based CAMD can be extended to computer-aided mixture/blend design (CAM^{b}D) to design or select mixtures of compounds which can exhibit properties superior to those of pure compounds [30, 96].

The development of solvent mixture design tools, such as CAM^{b}D, is crucial for promoting sustainable chemical manufacturing, given the increasingly stringent environmental regulations, such as the REACH regulations [97], on the list of allowable chemical compounds. In pharmaceutical manufacturing, for instance, the list of allowable organic solvents is being constantly reduced to conform to toxicological guidelines, especially since solvents cannot be completely separated from the final drug product [98]. Additionally, the use of unsuitable solvents may result in undesired drug properties and deteriorating drug performance [99]; for example, some crystallisation solvents may lead to undesired crystal shapes, such as needles, which can negatively impact product characteristics and downstream processing [31].

A general mathematical representation of CAM^{b}D is given as follows [34]:

$$
\begin{aligned}
\min_{\mathbf{x},\mathbf{y}} \quad & f(\mathbf{x},\mathbf{y}) \\
\text{s.t.} \quad & \mathbf{g_1}(\mathbf{y}) \leq \mathbf{0} \\
& \mathbf{g_2}(\mathbf{y}) \leq \mathbf{0} \\
& \mathbf{g_3}(\mathbf{x},\mathbf{y}) \leq \mathbf{0} \\
& \mathbf{g_4}(\mathbf{x},\mathbf{y}) = \mathbf{0} \\
& \mathbf{x} \in [\mathbf{x}^L, \mathbf{x}^U] \subset \mathbb{R}^n \\
& \mathbf{y} \in \{0,\,1\}^u
\end{aligned}
\tag{2.2}
$$

where $f(\mathbf{x},\mathbf{y})$ is a performance objective function that needs to be optimised, subject to a set of structural constraints, $\mathbf{g_1}(\mathbf{y}) \leq \mathbf{0}$, pure component property constraints, $\mathbf{g_2}(\mathbf{y}) \leq \mathbf{0}$, mixture property constraints, $\mathbf{g_3}(\mathbf{x},\mathbf{y}) \leq \mathbf{0}$, and process model constraints, $\mathbf{g_4}(\mathbf{x},\mathbf{y}) = \mathbf{0}$. $\mathbf{x}$ is an n-dimensional vector of continuous variables representing physical properties or process variables (temperature, mixture composition, etc.) and $\mathbf{y}$ is a u-dimensional vector of binary variables representing the identities of the functional groups in the mixture components or the identities of those components.

Formulation (2.2) usually takes the form of a highly nonconvex and challenging MINLP problem due to its combinatorial nature, the nonlinearity of the embedded physical property models, the difficulty of solving the phase equilibrium equations that are usually included in the problem, and the inclusion of additional optimisation variables in the mathematical formulation relative to pure component CAMD, such as the number of mixture components and mixture composition.

Most optimisation-based CAM^{b}D methods can be used in the context of the hybrid CAMD framework developed by Harper et al. [100] which entails three main design steps, namely: (1) pre-design, (2) design, and (3) post-design. In the pre-design step, the design objectives are fully defined and all qualitative information is expressed quantitatively. More specifically, the properties of interest and the tools for estimating those properties are chosen. Properties can be classified as either primary properties, which can be predicted based on molecular structure, e.g., group contribution methods, or secondary properties, which can be derived from primary properties. In the design step, molecules and mixtures that optimise a given

performance objective are designed, subject to a set of molecular structure constraints, pure compound/mixture property constraints and design constraints. Examples of molecular structure constraints include chemical feasibility constraints that ensure only chemically-meaningful combinations of functional groups are assembled (e.g., no violation of the octet rule [87]), and chemical complexity constraints that limit the number of occurrences of a specific functional group in the designed molecule(s). In the post-design step, the generated molecular candidates are evaluated with respect to additional aspects not included in the design formulation, such as legislative thresholds, cost, environmental impact and experimental findings.

CAM^{b}D approaches have been widely used in pharmaceutically-relevant applications such as the design of optimal solvent blends for enhanced crystallisation processes [31, 34, 55, 101, 102]. In these studies, optimal solvents are identified using solvent physical properties (e.g., boiling point, toxicity) to formulate performance metrics that may be related to process objectives (crystal yield). Karunanithi et al. [31] proposed a decomposition-based approach to design binary solvent and/or anti-solvent mixtures that maximise the potential recovery of APIs. This approach involves decomposing the original problem into a series of sub-problems for both pure solvent and solvent mixture designs. This CAM^{b}D problem was also addressed by Austin et al. [101] who presented an alternative decomposition-based approach by which the problem is projected onto low-dimensional pure component property space: the first step involves finding the optimal structure of each mixture component that corresponds to a given candidate property vector (discrete molecular design optimisation problem), while the second step involves optimising the mole fractions of the designed components (continuous mole fraction optimisation problem). Chen et al. [102] developed a computational tool for the rational selection of solvent mixtures that improve the aspect ratio of needle-like crystals by using a modified attachment energy model that incorporates solvent effects on crystal morphology. Jonuzaj et al. [34] proposed a CAM^{b}D formulation based on Generalised Disjunctive Programming (GDP) [103] to select multi-component solvent mixtures for separations, in which the number, identity and composition of mixture ingredients are simultaneously determined based on a predefined list of candidate compounds. This approach was extended to design novel solvents and solvent mixtures from a set of functional groups [55], an attractive feature given the increasingly

stringent environmental and safety legislation on the list of allowed compounds in the chemical industry [104].

2.2.3 Integrated molecular and process design

In order to understand better the effect of solvent choice on process performance and explore a wider design space, recent work has focused on developing tools for simultaneously identifying the optimal solvent mixtures and operating conditions based on a process-wide performance objective [59, 93].

Zhou et al. [105] proposed an integrated reaction solvent and process design formulation and applied it to a Diels-Alder reaction between acrolein and 1,3-cyclopentadiene, with the objective of maximising annual profit. A reaction kinetic model relating a set of experimentally-derived rate constants in several solvents with the associated theoretical solvent descriptors, obtained from a COSMO calculation [106], was built and embedded in the integrated formulation. Siougkrou et al. [84] developed an optimisation-based approach to design a CO_2-expanded liquid for the Diels-Alder reaction between anthracene and 4-phenyl-1,2,4-triazoline-3,5-dione. Their methodology identified the optimal organic co-solvent out of three candidate molecules, namely acetonitrile, methanol and acetone, as well as the CO_2 mole fraction that would minimise process costs. For modelling the effect of co-solvent choice on reaction kinetics, a solvatochromic equation was used in conjunction with a preferential solvation model. The authors' approach highlights the importance of considering the trade-off between the reaction rate constant and solubility when designing economical chemical processes. McBride et al. [107] developed a data-driven screening approach to identify an optimal two-solvent thermomorphic multicomponent system (TMS) for promoting the hydroformylation of 1-dodecene in the presence of a Rh-BiPhePhos catalyst complex. Thermodynamic calculations to estimate catalyst solubility, liquid-liquid equilibrium behaviour and product and catalyst partition coefficients were performed using COSMO-RS [108] and identified N,N–dimethylformamide (DMF) as a promising solvent candidate for catalyst recovery. In order to identify solvent pairs that are both functional and environmentally benign for the hydroformylation application, the latter work was extended to embed Safety, Health and Environment (SHE) criteria in the selection process using quantitative structure-activity

relationships (QSARs) and identified diethyl sulfoxide (DESO) as a green replacement solvent to DMF that exhibits higher catalyst separation and similar product separation efficiencies [109].

Recent research efforts have also focused on developing systematic frameworks for integrated solvent selection and process design in the context of pharmaceutical manufacturing. Wang et al. [110] presented a systematic approach to identify the optimal solvents and process conditions for continuous anti-solvent crystallisation with solvent recycling via single-stage flash separation. The authors used the perturbed-chain statistical associating fluid theory (PC-SAFT) model [111, 112] to predict relevant thermodynamic properties (solubility and vapour-liquid equilibrium) and the continuous-molecular targeting (CoMT) approach [113] to efficiently solve the optimisation problem. In CoMT, solvents are represented by a set of pure component parameters derived from the PC-SAFT model. In this approach, the integrated molecular and process design problem is solved in two steps: in the first step (continuous targeting), the set of thermodynamic model parameters that describe real solvent molecules are treated as continuous variables and allowed to vary, reducing the MINLP problem into a nonlinear program (NLP), and an NLP solver is used to optimise the continuous solvent parameters (within a predefined convex hull) and process variables, hence leading to an ideal hypothetical solvent molecule with parameters $\mathbf{y}^*$ and an optimised process with variables $\mathbf{x}^*$; in the second step (structural mapping), a real solvent molecule with thermodynamic parameters $\mathbf{y}$ closest to those of the hypothetical solvent from the first step is identified, for example by means of database search to select solvents from a predefined list [113] or GC methods to design novel molecules [114]. Since the optimal process variables $\mathbf{x}^*$ for the hypothetical solvent may not be optimal for the real solvent, the process is re-optimised while fixing solvent identity to that of the real solvent. Similarity between the hypothetical and real solvents is quantified using a Taylor approximation of the objective function around the CoMT-optimum parameter set $\mathbf{y}^*$. Mapping performance is then evaluated by comparing the objective function value corresponding to the real solvent with that predicted from the Taylor expansion. Wang et al. [54] developed a hybrid stochastic-deterministic optimisation framework to determine simultaneously economically favourable solvents and process operating conditions for a continuous process involving crystallisation and anti-solvent separation and recycling by multi-stage distillation.

In this work, the PC-SAFT model is used to predict thermodynamic and caloric properties needed to calculate phase equilibria and energy balances, and the solution algorithm for the optimisation problem consists of two nested layers: in the inner layer, the CoMT method is used to optimise the solvent parameters for a given process (fixed process topology and conditions) hence converting the MINLP problem into an NLP and solving it deterministically, while in the outer layer, the variables representing process topology and conditions are optimised using a genetic algorithm. The same research group recently extended their methodology to consider reactive crystallisation and raw material recycling [60], as well as continuous synthesis, extraction and crystallisation processes [61]. Watson et al. [56, 57] proposed a solvent mixture design formulation for the hybrid cooling and anti-solvent crystallisation of APIs where the initial and final crystallisation temperatures are treated as decision variables and relevant thermodynamic properties are calculated using the SAFT-γ Mie equation of state [115]. The design of a cooling crystallisation process and associated pure solvent was also considered by Chai et al. [116] using the decomposition approach of Karunanithi et al. [30]. Building on the work of Watson et al. [56], Jonuzaj et al. [58] formulated a comprehensive mixture design problem for the integrated crystallisation and isolation of pharmaceutical compounds where mixture property constraints, such as the miscibility of the crystallisation and wash solvents, are imposed for both purification stages and species solubility is predicted using the UNIFAC model. Furthermore, a bi-objective optimisation problem that involves minimising solvent use and maximising process safety, as quantified by solvent SHE scores obtained from the GSK solvent selection guide [23], is formulated and solved to highlight the trade-offs between the two KPIs. One of the benefits of considering the simultaneous design of the process and solvent(s) is that it is possible to conduct a more holistic assessment of environmental impact: Watson et al. [56, 57] and Jonuzaj et al. [58] embedded solvent consumption in the design formulation as a green chemistry metric. In fact, many approaches to CAMD/CAM^{b}D include some consideration of environmental, health and safety metrics via solvent properties. This is the case for instance in the recent work of Chai et al. [116], who have applied the Grand Product Design framework [117] to select and design solvents for crystallisation and isolation (filtration, washing and drying), and who include several sub-models that can accommodate multiple performance criteria, including

environmental impact.

2.2.4 Property prediction in CAMD

Property prediction is a central task in CAMD problems. Physical properties, such as boiling/melting points, heats of melting and heats of vaporisation, need to be computed for a wide range of mixtures, from knowledge of molecular structure only, to select or design optimal molecules for a particular application. This is especially challenging in pharmaceutical applications due to the diversity of molecular structures and the lack of experimental data in the early stages of process development. In this section, group contribution methods are introduced as tools for estimating molecular and mixture properties in CAMD. This is followed by a discussion of the UNIFAC model and the prediction of liquid-phase activity coefficients for solubility calculations.

Group contribution (GC) methods

Group contribution (GC) methods have been developed to estimate the properties of a wide range of molecules using a small number of structural building groups, where the contribution of each building group to molecular properties is independent of the target molecule and of the molecular environment. In this sense, GC methods are classified as QSPRs that relate the structure of a chemical compound to its properties, hence enabling the quantification of its performance for a given application [93]. Examples of structural groups include CH_3, CH_2, OH and CH_3NO_2. In GC methods, a target property of a molecule can be computed as the sum of the structural group contributions to that property multiplied by the number of group occurrences in the molecule. This can be represented mathematically by the following equation:

$$P_m = f\left(\sum_{g \in G} n_g P_g\right), \tag{2.3}$$

where f is a function, P_m is an estimated property of molecule m, n_g is the number of occurrences of functional group g in m, and P_g quantifies the contribution of group g to P_m. The vector of coefficients P_g is obtained by regression over a large data set of the target molecular property

P_m for different compounds.

GC methods have been used to estimate properties such as boiling point, melting point and standard enthalpy of vaporization [118], as well as the acentric factor and liquid molar volume at 298.15 K [119]. They have also been developed to predict hydrogen-bond basicity, hydrogen-bond acidity, the Hildebrand solubility parameter, surface tension, dipole moment, index of refraction and dielectric constant [120], in addition to solvent dipolarity/polarisability [26]. However, the additive nature of GC approaches means that structural groups are treated as independent of each other. This limits the ability of GC methods to distinguish between isomers, so that multiple molecules can be designed using the same set of building blocks; for example, both 2-Pentanol and 3-Pentanol can be constructed from the groups CH_3, CH_2 and OH. Another consequence of the additive-contribution assumption is that proximity effects, i.e., the polarisation effects that take place when some structural groups are close to one another in a molecule, such as CH_2 and OH, are neglected [50]. For these reasons, there have been considerable efforts to develop GC methods that incorporate higher-order groups, i.e., groups that combine simple structural groups (first-order groups) as structural building blocks that capture proximity effects and advance molecular design to differentiate between isomer candidates [121, 122]. For example, if COOH, CH and NH_2 are classified as first-order structural groups, CHCOOH may be classified as a second-order group with first-order groups as building blocks, and $CHNH_2COOH$ may be classified as a third-order group with a second-order group and a first-order group as building blocks.

The Marrero-Gani method [121] is commonly used to predict molecular properties using higher-order groups. This method can be represented by the following equation:

$$P_m = f\left(\sum_{g\in F} a_g X_g + \sum_{g\in S} b_g Y_g + \sum_{g\in T} c_g Z_g\right), \tag{2.4}$$

where a_g, b_g and c_g represent, respectively, the number of occurrences of a first-order, second-order and third-order functional group g in molecule m, X_g, Y_g and Z_g represent, respectively, the contributions of a first-order, second-order and third-order group g to property P_m, and F, S and T represent, respectively, first-order, second-order and third-order functional groups.

The Marrero-Gani method has been used to predict properties of large heterocyclic and

polyfunctional acyclic molecules. It was found that using higher-order groups decreased the standard deviation and absolute average error associated with the prediction of boiling points, melting points and standard enthalpies of vaporization [121].

GC methods are useful in CAMD as they can be easily embedded in mathematical formulations and can describe a large design space from a small set of structural groups. However, as mentioned previously, there are some weaknesses associated with these methods. For example, GC methods cannot distinguish between isomers unless they are based on higher-order groups, which may increase the computational cost of performing property calculations as more binary variables would be needed for representing molecular structure [123]. Second, structural groups used to estimate different properties in GC methods are not consistent; this can lead to difficulties in formulating a CAMD problem mathematically.

While the estimation of pure compound properties is relatively simple as it is mainly dependent on molecular structure, the evaluation of mixture properties, which can either be single-phase thermodynamic properties such as heat capacities or phase equilibrium properties such as activity coefficients, is more challenging as these properties are highly influenced by several variables such as temperature, pressure and composition. Furthermore, although simple mixing rules and linear models can be used to study ideal systems, more sophisticated, nonlinear models should be used when considering non-ideal mixtures [94]. In the chemical industry, a wide variety of solvent mixtures are deployed in manufacturing processes that operate under a wide range of conditions. This makes it difficult to derive a general property prediction method that applies to all sorts of mixtures. Accordingly, developing mathematical models that can predict the thermodynamic properties of as many mixtures as possible is invaluable [124]. Thermodynamic properties of mixtures are often estimated using GC methods, which are used in this work, or computational chemistry/quantum mechanical approaches [108, 125, 126].

Estimating mixture properties

In the context of CAM^{b}D, mixtures are modelled as mixtures of structural groups that constitute their individual components. Accordingly, GC methods can be used to estimate mixture properties based on the contributions of the different functional groups present in the mixture

of interest. According to Papaioannou et al. [127], GC methods for estimating phase equilibria can be classified into two types: (1) models to estimate liquid-phase activity coefficients, and (2) approaches combining GC methods with equations of state to model both liquid and vapour phases.

Activity coefficient GC methods are considered the state-of-the-art tools to predict phase behaviour of mixtures at a specific temperature and pressure through the calculation of activity coefficients [127]. In phase equilibrium calculations, activity coefficients are used to express the partial properties of each component in a mixture [124]. Consider a multi-component system of Z different phases and N components. At phase equilibrium, the fugacity (or "escaping tendency") of each component in the mixture is the same in all phases at a given temperature T and pressure P [128]:

$$f_i^1 = f_i^2 = ... = f_i^Z, \quad i = 1, 2, ..., N. \tag{2.5}$$

Considering the case of vapour-liquid equilibrium (VLE), the vapour-phase fugacity of a component i, f_i^V, can be expressed as follows:

$$f_i^V = y_i P \phi_i^V(P, T, \underline{y}), \quad i = 1, 2, ..., N, \tag{2.6}$$

where y_i is the mole fraction of component i in the vapour phase, $\underline{y}$ is the concentration of the vapour phase and ϕ_i^V is the vapour-phase fugacity coefficient.

The liquid-phase fugacity of a component i, f_i^L, can be expressed as follows:

$$f_i^L = x_i \gamma_i(P, T, \underline{x}) f_i^0, \quad i = 1, 2, ..., N, \tag{2.7}$$

where γ_i is the liquid-phase activity coefficient of component i in the mixture, $\underline{x}$ is the concentration of the liquid phase and f_i^0 is the fugacity of that component at standard state.

At moderate pressures, f_i^0 is approximated by the saturated vapour pressure, P_i^{sat}, and equation (2.7) is reformulated as follows:

$$f_i^L = x_i \gamma_i(P, T, \underline{x}) P_i^{sat}, \quad i = 1, 2, ..., N. \tag{2.8}$$

Since at VLE the equality of fugacity applies:

$$f_i^V = f_i^L, \quad i = 1, 2, ..., N, \tag{2.9}$$

it follows that:

$$y_i P \phi_i^V(P, T, \underline{y}) = x_i \gamma_i(P, T, \underline{x}) P_i^{sat}, \quad i = 1, 2, ..., N. \tag{2.10}$$

The calculation of ϕ_i^V requires an equation of state. However, at moderate pressures, the vapour phase can be assumed ideal and ϕ_i^V can be set to 1 [128]. Therefore, equation (2.10) reduces to:

$$y_i P = x_i \gamma_i(P, T, \underline{x}) P_i^{sat}, \quad i = 1, 2, ..., N, \tag{2.11}$$

which is an equation that describes the VLE behaviour of non-ideal systems by using γ_i to account for the non-idealities.

While activity coefficient GC methods are widely used for estimating mixture thermodynamics, they are reliable over only a small range of temperatures and pressures, and they are limited to phase behaviour calculations; they cannot be used to compute other properties such as heat capacities, densities and enthalpies. On the other hand, methods that couple equations of state with GC methods (GC-based equations of state) offer several advantages over activity coefficient models, such as a wider range of application, an equal treatment of the liquid and vapour phases, and the ability to predict thermodynamic properties beyond phase behaviour [127]. An interesting example of a molecular GC-based equation of state is SAFT-γ Mie [115], which has been developed to estimate fluid-phase behaviour and thermodynamic derivative properties. More specifically, in SAFT-γ Mie, the molecular model is related to macroscopic thermodynamic properties by developing a functional form of the total Helmholtz free energy, which is formulated in terms of functional group contributions. From the Helmholtz free energy expression, standard thermodynamic relations [129] are used to predict first-order derivative properties such as pressure, chemical potential, enthalpy and Gibbs free energy, as well as second-order derivative properties such as isochoric and isobaric heat capacity and isothermal compressibility [115].

In this PhD thesis, the original Universal Quasi-chemical Functional Group Activity Coefficient (UNIFAC) model [50] is used to predict phase behaviour since: 1) the operating conditions of temperature and pressure in the presented case studies are within the range of applicability of the original UNIFAC model, i.e., below 10–15 atm and within 275–425 K, and 2) this work does not involve the prediction of single-phase thermodynamic properties, which would have required the use of an equation of state. The reader is referred to the next section for an overview of the original UNIFAC model.

The original UNIFAC model for predicting API solubility

In pharmaceutical manufacturing, the selection of suitable process solvents is highly dependent on the solubility of solid compounds in those solvents for a particular processing task; for example, reaction solvents must dissolve reactants to initiate the chemical transformation and speed up the reaction rate, whereas crystallisation solvents must dissolve impurities and limit API solubility in the final crystallisation state. Accordingly, estimating the solubility of pharmaceutical intermediates or APIs is critical for the development a solvent selection procedure for pharmaceutical processes.

In order to develop a thermodynamic equation that can be used to estimate solubility, it should first be noted that solubility is defined as the equality of the chemical potential between the solid state, μ^S, and the liquid state (saturated solution), μ^L, of a given solute i [130, 131]:

$$\mu_i^S = \mu_i^L. \tag{2.12}$$

An expression of μ_i^L is given as follows:

$$\mu_i^L = \mu_i^0 + R_g T \ln(\gamma_i^L x_i^L), \tag{2.13}$$

where μ_i^0 is the chemical potential of a hypothetical pure solute i at a reference state (system temperature), R_g is the universal gas constant, T is temperature, γ_i^L is the activity coefficient of the solute in the saturated solution and x_i^L is the solubility of the solute in the saturated solution.

Combining equations (2.12) and (2.13) and rearranging gives:

$$\ln(\gamma_i^L x_i^L) = \frac{\mu_i^S - \mu_i^0}{R_g T}, \tag{2.14}$$

which can be alternatively expressed as follows:

$$\ln(\gamma_i^L x_i^L) = \frac{\Delta G^m}{R_g T}, \tag{2.15}$$

where ΔG^m describes the difference in the partial molar Gibbs free energy of the solute between the solid state and the reference state at constant temperature and pressure.

Equation (2.15) can be expanded to give the following relation between the solubility of a solute and its activity coefficient:

$$\ln(\gamma_i^L x_i^L) = \frac{\Delta H_m(T_m)}{R_g}\left(\frac{1}{T_m} - \frac{1}{T}\right) - \frac{\Delta C_p(T_m)}{R_g}\left[\ln\left(\frac{T_m}{T}\right) - \frac{T_m}{T} + 1\right], \tag{2.16}$$

where ΔH_m is the enthalpy of melting, T_m is the melting temperature and ΔC_p is the heat capacity difference between the supercooled melt and the solid form of the solute. It is often assumed that the heat capacity term in equation (2.16) is small relative to the other term [131]; accordingly, the following simplified expression can be used to predict solubility:

$$\ln(\gamma_i^L x_i^L) = \frac{\Delta H_m(T_m)}{R_g}\left(\frac{1}{T_m} - \frac{1}{T}\right). \tag{2.17}$$

A study by Bouillot et al. [130] evaluated the performance of a number of predictive and semi-predictive thermodynamic models to predict the solubility of drug or drug-like molecules in different solvents. It was found that the UNIFAC models (original [50] and modified [132]) provided the best results in comparison with COSMO-SAC [106, 125] and NRTL-SAC [133]; hence, UNIFAC can be used as a tool to make rapid solubility predictions and guide experimental solvent screening during pharmaceutical process development.

The UNIFAC model is commonly used to predict liquid-phase activity coefficients at moderate to low pressures [128]. In the UNIFAC model, the activity coefficient of a pure compound is divided into two contributions:

$$\ln \gamma = \ln \gamma^C + \ln \gamma^R, \tag{2.18}$$

where γ^C, the combinatorial term, quantifies the entropic contribution (molecular size and shape)

to the activity coefficient, whereas γ^R, the residual term, quantifies the enthalpic contribution (intermolecular and intramolecular interactions) to the activity coefficient. γ^C is calculated by using the van der Waals group volume, R_k, and the group surface area, Q_k, parameters, where k is a functional group, while γ^R is calculated by using energetic interaction parameters $a_{m,n}$ and $a_{n,m}$ ($a_{m,n} \neq a_{n,m}$ and $m \neq n$) between two different functional groups m and n. The UNIFAC interaction parameters are computed by regression to experimental data, and are compiled in the UNIFAC parameter table [134].

Despite its usefulness in providing estimates of activity coefficients, the original UNIFAC model suffers from a number of limitations. For example, the method can only be applied within a limited range of conditions, namely moderate pressures of less than 10–15 atm and temperatures within the range of 275–425 K. Additionally, the method assumes ideal behaviour for the vapour phase, and its parameters have originally been regressed to VLE data, both of which can lead to inaccuracies in the predictions. Some of these limitations have been addressed, for example, by regressing parameters of 32 structural groups to experimental liquid-liquid equilibria (LLE) data in order to accurately model LLE behaviour [135], and by developing a modified UNIFAC model [132] that better expresses temperature dependence in the residual contribution to the activity coefficient.

In this work, the solubility of chemical species that can exist in solid form during API manufacturing is calculated using equation (2.17) where the liquid-phase activity coefficients are estimated using the original UNIFAC model as the process conditions of the chosen case studies fall within the model range of applicability.

2.3 Chemical Route Selection

The development and commercialisation of novel products is crucial for driving profitability in the fine chemicals and pharmaceutical industries. Route selection is the first step of any chemical process development workflow [136], and decisions taken during this task can greatly impact process economics, safety and operability once the process is scaled-up and commercialised [137]. Accordingly, the selection of efficient synthetic routes is key for the development of profitable and environmentally-benign drug manufacturing processes. Nevertheless, route selection is not

a trivial task but is a complex, multifaceted problem in which several factors must be addressed, such as safety, environmental impact, legal considerations (intellectual property), economics, productivity and quality control [137, 138]. A tool that can quickly screen and evaluate synthetic routes based on the aforementioned criteria with minimal time and cost would significantly accelerate drug development and ensure the design of sustainable and robust pharmaceutical processes for the production of high-quality APIs.

The literature includes several studies that devise methodologies for synthetic route screening based on a variety of performance metrics. Serna et al. [139] proposed a systematic approach to evaluate chemical process routes based on sustainability indicators related to economics, environmental impact, safety and occupational health. Their methodology uses multi-criteria decision analysis (MCDA) to calculate weights and mutual relations or influences between the indicator groups, followed by combining the indicators, weights and influences in a single sustainability metric, called the Sustainable Cumulative Index (SCI), for assessing routes and facilitating decision-making during early design stages. More specifically, this methodology consists of three main steps: 1) Identifying chemical process route alternatives for the production of a specific molecule and gathering information for calculating sustainability indicators, e.g., molecular properties, material costs, process specifications, etc., 2) Selecting and calculating sustainability indicators that are relevant to the process of interest, and 3) Using MCDA to calculate weights, by Analytic Hierarchy Process (AHP) [140], and influences between the indicator groups, by Decision Making Trial and Evaluation Laboratory (DEMATEL) [141], followed by evaluating an integrated sustainability assessment, the SCI, for each studied route. Jacob et al. [142] developed an automated workflow for synthetic route generation and evaluation. In this approach, data mining using the Reaxys database (`https://www.reaxys.com`) is used to construct reaction networks based on input process data, whereas the PROMETHEE algorithm [143] is used for multi-criteria decision making by which routes are evaluated and ranked based on mass and energy-related process metrics and heuristics. This methodology was extended to include chemical structure requirements for restricting the chemical space during route screening, and process simulation of the most optimal route using process models built from literature and proprietary data [144]. Voll and Marquardt [145] developed Reaction Network Flux Analysis

(RNFA) as an optimisation-based approach to route selection that is inspired from metabolic pathway analysis and only requires information on reaction yield and stoichiometry. In this approach, reaction pathways are systematically identified and optimisation techniques are used to enumerate and rank those pathways based on a given performance objective, such as product yield, energy efficiency and raw material cost. RNFA has been applied to identify promising pathways for the production of biofuels [146, 147, 148] and biopolymers [149, 150]. Although RNFA constitutes a simple yet powerful approach to screening reaction routes, it neglects the impact of solvent use on process performance and assumes perfect separation of solvents and byproducts after each reaction step; accordingly, RNFA was extended to Process Network Flux Analysis (PNFA) in order to evaluate the feasibility and energy demand of different separation technologies by incorporating thermodynamically-sound, shortcut separation models in the mathematical formulation of the extended approach [151]. The extension of RNFA, which focuses on mass-based performance criteria, to PNFA, which considers energy-related contributions to process performance, offers a valuable tool to quantify process costs and environmental impact while maintaining the rapid screening feature of RNFA [152]. Similar to RNFA, PNFA has mainly been applied to screen reaction pathways for renewable fuels production [148, 151, 153]. More recently, with the recent advancement in data science and machine learning, there have been efforts to develop computational approaches to retrosynthetic route planning that can be applied on large reaction databases. Based on an input molecular structure of a target compound, these approaches can generate a ranked list of reaction pathways that connect that molecule to commercially available starting materials through a number of chemically feasible synthetic steps by deploying state-of-the-art machine learning algorithms. Accordingly, these approaches address the challenges of early work on retrosynthesis in which heuristics and chemical intuition were used to suggest molecular disconnections but have often led to infeasible, inefficient or incomplete reaction routes [154].

It has been shown in Section 2.2.3 that integrated molecular and process design approaches, such as CAM^{b}D, can be used to identify the optimal solvents/solvent mixtures and operating conditions of integrated pharmaceutical processes. It is therefore possible to deploy such model-based tools to evaluate the process performance of synthetic routes in a given reaction

network and guide both molecular and process-level decision making. A major limitation of synthetic route screening during early-stage process development is that it heavily relies on lab-scale data to compute performance metrics and compare reaction routes. Comparisons based on this scale may fail to provide any useful conclusions or may become invalid at process scale; accordingly, it is necessary to evaluate the process-level implementation of a synthetic route, i.e., the process route, during route selection by considering a variety of process aspects such as solvent use and recycling, energy consumption, byproduct valorisation, etc. [155]. The development of a CAM^{b}D-based approach to route selection, in which multiple KPIs that quantify resource efficiency, product quality and process economics are embedded, can potentially identify promising synthetic routes based on process performance, before taking these routes forward to the next steps of process development.

2.4 Mathematical Optimisation

2.4.1 Mixed-integer optimisation

Optimisation-based CAM^{b}D formulations take the form of mixed discrete and continuous optimisation problems that include continuous variables representing process variables, such as species mole fractions and operating temperatures, and discrete variables representing molecular identity, such as solvent choice in a solvent selection problem. The combinatorial nature of discrete optimisation problems makes them challenging to solve. Furthermore, the existence of multiple local minima in the search space makes it difficult to identify a globally optimal solution [156]. Logic-based modelling, such as Generalised Disjunctive Programming (GDP), and mixed-integer programming (MILP or MINLP) are common approaches used to formulate discrete and continuous optimisation problems [157, 158].

Generalised disjunctive programming (GDP)

Generalised disjunctive programming (GDP) is a logic-based approach to model discrete and continuous optimisation problems that was proposed by Raman and Grossman [103]. In GDP, the discrete decisions of an optimisation problem are expressed as Boolean variables and those

variables are related to continuous variables through disjunctions, algebraic equations and logic propositions. Formulating an MINLP problem via the GDP formalism offers a more structured approach for modelling discrete and continuous decisions and provides a clearer representation of the quantitative and qualitative parts of an optimisation problem.

The GDP formulation can be represented mathematically as follows:

$$\begin{aligned}
\min_{\mathbf{x},\mathbf{Y}} \quad & f(\mathbf{x}) \\
\text{s.t.} \quad & \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \\
& \bigvee_{j \in J_k} \begin{bmatrix} Y_{j,k} \\ h_{j,k}(\mathbf{x}) \leq 0 \end{bmatrix}, k \in K \\
& \veebar_{j \in J_k} Y_{j,k}, k \in K \\
& \Omega(\mathbf{Y}) = \text{True} \\
& \mathbf{x} \in [\mathbf{x}^L, \mathbf{x}^U] \subset \mathbb{R}^n \\
& Y_{j,k} \in \{\text{True}, \text{False}\}, j \in J_k, k \in K
\end{aligned} \tag{2.19}$$

where $f(\mathbf{x})$ is an objective function which depends on the continuous variables represented by an n-dimensional vector $\mathbf{x}$, $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$ is a set of general constraints that are independent of the discrete variables, and $h_{j,k}(\mathbf{x}) \leq 0$ is a set of conditional constraints that depend on the discrete choices. These conditional constraints are expressed within disjunctions and are linked via the OR ($\vee$) operator. K is the index set for the disjunctions and J_k is the index set for the disjunctive terms. $Y_{j,k}$ is an element of a matrix $\mathbf{Y}$ of Boolean variables that specifies whether an element j in a disjunction k is activated ($Y_{j,k}$=True) or not ($Y_{j,k}$=False). $\veebar_{j \in J_k} Y_{j,k}$ is a constraint enforcing that only one of the Boolean variables is true in each disjunction, where $\veebar$ denotes the EXCLUSIVE OR operator. In the context of mixture design, the disjunctive constraints are related to assigning the identities of mixture components. $\Omega(\mathbf{Y})$ represents a set of logic relationships between the Boolean variables, expressed as propositional logic [103, 158].

While in an MINLP problem the logic is expressed in the objective function, $f(\mathbf{x}, \mathbf{y})$, and algebraic constraints, $\mathbf{h}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$ or $\mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0}$, in GDP, the logic is expressed in the disjunctions by relating the Boolean variables, $Y_{j,k}$, to the continuous-variable constraints,

$h_{j,k}(\mathbf{x})$, and connecting the disjunctive sets via the propositional logic $\Omega(\mathbf{Y})$.

GDP can be reformulated as an MINLP by different approaches, such as the big-M (BM) approach [159] which is the simplest representation of GDP as a mixed-integer problem [103], and the focus of this work. The mathematical formulation of GDP as an MINLP problem by BM can be expressed as follows:

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \\ & h_{j,k}(\mathbf{x}) \leq M_{j,k}(1 - y_{j,k}), j \in J_k, k \in K \\ & \mathbf{A}\mathbf{y} \leq \mathbf{a} \\ & \sum_{j=1}^{J_k} y_{j,k} = 1, k \in K \\ & \mathbf{x} \in [\mathbf{x}^L, \mathbf{x}^U] \subset \mathbb{R}^n \\ & \mathbf{y} \in \{0, 1\}^u \end{aligned} \tag{2.20}$$

where $\mathbf{y}$ is a matrix of binary variables that correspond to the elements $Y_{j,k}$ in the Boolean variable matrix, $\mathbf{Y}$, and $M_{j,k}$ are elements of a matrix $\mathbf{M}$ of sufficiently large values as upper bounds in the inequality constraints, i.e., these constraints are deactivated when $y_{j,k} = 0$. The logic propositions $\Omega(\mathbf{Y}) = \text{True}$ in the GDP formulation are transformed to the linear inequalities $\mathbf{A}\mathbf{y} \leq \mathbf{a}$ in the BM formulation via Boolean algebra rules [160].

Mixed-integer nonlinear programming (MINLP)

The general mathematical formulation of an MINLP model is given as follows [161]:

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & f(\mathbf{x}, \mathbf{y}) \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{x}, \mathbf{y}) = \mathbf{0} \\ & \mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0} \\ & \mathbf{x} \in [\mathbf{x}^L, \mathbf{x}^U] \subset \mathbb{R}^n \\ & \mathbf{y} \in \{0, 1\}^u \end{aligned} \tag{2.21}$$

where $f(\mathbf{x})$ is an objective function that needs to be optimised, $\mathbf{h}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$ is a set of equality constraints, $\mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0}$ is a set of inequality constraints, $\mathbf{x}$ is an n-dimensional vector of continuous variables and $\mathbf{y}$ is a u-dimensional vector of binary variables.

Some common methods that have been devised to solve MINLP problems are summarised below.

The Nonlinear Branch and Bound (NLP-BB) method [162, 163]: In NLP-BB, the binary variables in the binary tree are relaxed between 0 and 1 in the first node. Then, the NLP problem is solved to give a lower bound on the objective function value in a minimisation problem. The node is then separated into two branches, each representing a discrete value for one of the binary variables. The NLP problem for each node is solved and the solution added to a list of open nodes. The node with the best objective value in the list is considered for further branching. As the procedure continues and new layers are added, the node with the first feasible solution provides an upper bound on the objective value. Accordingly, nodes with lower bounds greater than the current upper bound or with infeasible solutions can be eliminated.

The Generalised Benders Decomposition (GBD) method [164]: This approach is applied to the class of problems in which the binary variables appear in linear and mixed-bilinear terms in the optimisation problem. The general principles of the GBD method include:

- Partitioning the variable set into complicating (typically binary) variables $\mathbf{y}$ and continuous variables $\mathbf{x}$;
- Decomposing the problem by solving two types of subproblems, namely a primal problem, formulated using fixed values of the binary variables $\mathbf{y}$, to provide an upper bound on the original MINLP problem, and a master problem, formulated using information obtained from the solution of the primal problem, to provide a lower bound on the original problem;
- Iteratively refining the problem by using information from the primal and master problems to construct new subproblems such that the bounds are tightened and

convergence is achieved within a finite number of iterations.

The Outer-Approximation (OA) method [165, 166, 167]: This approach is applied to the class of problems in which the binary variables appear in linear terms of the optimisation problem. The OA method is similar to the GBD method, but differs in formulating the master problem which is constructed by linearisations of nonlinear objective functions and/or constraints at the solution of the primal problem from all previous iterations.

It is important to mention that if the NLP part of the MINLP problem is non-convex in any of the methods, the optimisation algorithm may converge to a local solution, unless a global optimisation solver is used. In this work, the Simple Branch and Bound (SBB) solver [163], which is an implementation of NLP-BB in GAMS, is used to solve the formulated CAM^{b}D problems as SBB has been successfully used in previous work to design integrated crystallisation and isolation processes [58]. The CONOPT solver [168] is used to solve the NLP part of the problem as it is well suited for large models with nonlinear constraints [169]. Unfortunately, the use of a global solver, such as BARON [170], currently fails to identify a solution due to the large model size and embedded non-linearities.

2.4.2 Multi-objective optimisation

Solvent selection is a multi-faceted problem involving several competing performance indicators that need to be optimised. For example, increasing the crystal yield of a pharmaceutical compound (desired) may require the addition of an anti-solvent to the crystallisation mixture, hence increasing solvent consumption (undesired), or may require a larger drop in the crystalliser temperature to further induce product crystallisation, hence increasing the crystalliser duty (undesired). In order to explore these trade-offs and guide rational design decisions during process development, multi-objective optimisation (MOO) can be used to simultaneously optimise several KPIs and identify compromise solutions based on the design objectives. Accordingly, this work includes several instances of multi-objective CAM^{b}D to generate different designs (solvent mixtures, mixture composition and process conditions) in which KPIs are simultaneously optimised. Each of these designs constitutes a Pareto-optimal solution of the MOO problem,

i.e., a solution that cannot be improved with respect to one objective without deteriorating the other objective(s). The general mathematical formulation of an MINLP MOO problem is given by:

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & f_1(\mathbf{x},\mathbf{y}), f_2(\mathbf{x},\mathbf{y}), ..., f_P(\mathbf{x},\mathbf{y}) \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{x},\mathbf{y}) = \mathbf{0} \\ & \mathbf{g}(\mathbf{x},\mathbf{y}) \leq \mathbf{0} \\ & \mathbf{x} \in [\mathbf{x}^L, \mathbf{x}^U] \subset \mathbb{R}^n \\ & \mathbf{y} \in \{0, 1\}^u \end{aligned} \tag{2.22}$$

where each objective function $f_i(\mathbf{x},\mathbf{y})$, $i = 1, \ldots, P$, is a performance objective that needs to be optimised, $\mathbf{h}(\mathbf{x},\mathbf{y}) = \mathbf{0}$ is a set of equality constraints, $\mathbf{g}(\mathbf{x},\mathbf{y}) \leq \mathbf{0}$ is a set of inequality constraints, $\mathbf{x}$ is an n-dimensional vector of continuous variables and $\mathbf{y}$ is a u-dimensional vector of binary variables.

Several approaches exist for solving MOO problems. In this section, two common MOO solution techniques are discussed, namely the weighted-sum method and the ϵ-constraint method [171].

The weighted-sum method

In the weighted-sum method, the original MOO problem is converted to a single objective optimisation (SOO) problem by applying a weighting factor to each objective (scalarisation) and summing the weighted sum of the objectives to form a single function. Optimising the single objective function generates one Pareto-optimal point, while varying the weights associated with the multiple objectives leads to a set of Pareto-optimal solutions that constitute the Pareto front. The general mathematical formulation of the weighted-sum method is given by:

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & \sum_{i=1}^{P} w_i f_i(\mathbf{x},\mathbf{y}) \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{x},\mathbf{y}) = \mathbf{0} \\ & \mathbf{g}(\mathbf{x},\mathbf{y}) \leq \mathbf{0} \\ & \mathbf{x} \in [\mathbf{x}^L, \mathbf{x}^U] \subset \mathbb{R}^n \\ & \mathbf{y} \in \{0,1\}^u \end{aligned} \tag{2.23}$$

where w_i are the weights associated with each objective function, $f_i(\mathbf{x},\mathbf{y})$, such that $\sum_{i=1}^{P} w_i = 1$. Although the weighted-sum method is relatively simple to implement, some of its main limitations include: the challenge of systematically choosing appropriate weights to generate a diverse and well-distributed set of Pareto-optimal points [172], and the inability of the method to identify any non-convex regions of the Pareto front [173, 174, 175].

The ϵ-constraint method

In the ϵ-constraint method, a series of SOO problems are solved sequentially [176] where one objective is optimised while the remaining objectives are transformed into inequality constraints bounded within some user-defined bounds $\boldsymbol{\epsilon}$. The general mathematical formulation of the ϵ-constraint method is given by:

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & f_1(\mathbf{x},\mathbf{y}) \\ \text{s.t.} \quad & f_2(\mathbf{x},\mathbf{y}) \leq \epsilon_2 \\ & f_3(\mathbf{x},\mathbf{y}) \leq \epsilon_3 \\ & \ldots \\ & f_P(\mathbf{x},\mathbf{y}) \leq \epsilon_P \\ & \mathbf{h}(\mathbf{x},\mathbf{y}) = \mathbf{0} \\ & \mathbf{g}(\mathbf{x},\mathbf{y}) \leq \mathbf{0} \\ & \mathbf{x} \in [\mathbf{x}^L, \mathbf{x}^U] \subset \mathbb{R}^n \\ & \mathbf{y} \in \{0,1\}^u \end{aligned} \tag{2.24}$$

where $f_1(\mathbf{x}, \mathbf{y})$, $f_2(\mathbf{x}, \mathbf{y})$, ..., $f_P(\mathbf{x}, \mathbf{y})$ are P objective functions to be optimised and ϵ_2, ϵ_3, ..., ϵ_P are $P-1$ user-defined bounds on those objective functions. The Pareto-optimal solutions are obtained by varying the values of $\boldsymbol{\epsilon}$ in each run of the optimisation problem. Although the ϵ-constraint method can detect the non-convex regions of the Pareto front [177, 178], for p >2, the problem can become computationally demanding and constructing the Pareto front becomes highly dependent on the user-defined values of $\boldsymbol{\epsilon}$. In this work, the ϵ-constraint method is used to solve MOO problems where bi-objective optimisation CAM^{b}D formulations are considered.

The augmented ϵ-constraint method

In order to address the weaknesses of the original ϵ-constraint method, the augmented ϵ-constraint method (AUGMENCON) was developed by Mavrotas [179]. AUGMENCON involves the use of slack or surplus variables to transform the objective function ϵ-constraints to equality constraints and using these variables as a second term in the main objective function with lower priority, hence guaranteeing the the Pareto-optimality of the generated solutions. The reader is referred to the paper by Mavrotas [179] for a proof that AUGMENCON produces only efficient or non-dominated solutions. The general mathematical formulation of AUGMENCON, assuming without loss of generality that the multiple objective functions are to be maximized, is given by:

$$
\begin{aligned}
\max_{\mathbf{x},\mathbf{y}} \quad & f_1(\mathbf{x}, \mathbf{y}) + \kappa\left(\frac{s_2}{r_2} + \frac{s_3}{r_3} + ... + \frac{s_P}{r_P}\right) \\
\text{s.t.} \quad & f_2(\mathbf{x}, \mathbf{y}) - s_2 = \epsilon_2 \\
& f_3(\mathbf{x}, \mathbf{y}) - s_3 = \epsilon_3 \\
& ... \\
& f_P(\mathbf{x}, \mathbf{y}) - s_P = \epsilon_P \\
& \mathbf{h}(\mathbf{x}, \mathbf{y}) = \mathbf{0} \\
& \mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0} \\
& \mathbf{x} \in [\mathbf{x}^L, \mathbf{x}^U] \subset \mathbb{R}^n \\
& \mathbf{s} \subset \mathbb{R}^+ \\
& \mathbf{y} \in \{0, 1\}^u
\end{aligned}
\tag{2.25}
$$

where $f_1(\mathbf{x}, \mathbf{y})$, $f_2(\mathbf{x}, \mathbf{y})$, ..., $f_P(\mathbf{x}, \mathbf{y})$ are P objective functions to be optimised, ϵ_2, ϵ_3, ..., ϵ_P are the $P-1$ right-hand side parameters corresponding to a specific combination of the enforced lower bounds on the objective functions f_2, f_3, ..., f_P, $\mathbf{s}$ is a vector of surplus variables s_2, s_3, ..., s_P used to transform the objective function ϵ-constraints to equality constraints, r_2, r_3, ..., r_P are the ranges of the $P-1$ objective functions f_2, f_3, ..., f_P, and κ is a small number ranging between 10^{-3} and 10^{-6}. In this work, AUGMENCON is used to solve MOO problems where more than 2 objective functions are considered.

2.5 Conclusion

In this chapter, the main classes of CAMD were introduced and discussed and an overview of chemical route selection was presented. A summary of major research works that deploy integrated molecular and process design tools to identify optimal solvents and process conditions for integrated processes, with a focus on pharmaceutical applications, was presented. This highlighted a gap in developing a systematic approach to solvent selection and process design for end-to-end pharmaceutical processes, which would potentially constitute a useful method to screen synthetic routes based on process performance. Finally, since this work involves developing and using optimisation-based CAM^{b}D for solvent and process design, some important concepts in discrete/continuous and multi-objective optimisation were discussed.

The next chapter introduces a CAM^{b}D-based approach to the design of integrated synthesis and crystallisation processes while taking several important KPIs that quantify resource efficiency, product quality and SHE performance into account.

Chapter 3

Model-based Solvent Selection for Integrated Synthesis and Crystallisation

Most of the material of this chapter has been published (open access) in Chemical Engineering Science [180].

3.1 Introduction

Given the many solvent-related contributions to process and environmental performance, pharmaceutical companies and academic research groups are actively developing tools to identify greener approaches for the design of resource-efficient process routes. While model-based solvent selection for pharmaceuticals has received significant interest over the past few years, current methodologies often focus on identifying solvents that are optimal for a single process unit with fixed operating conditions, and hence fail to account for the integrated nature of pharmaceutical processes [57]. In particular, the two key stages of synthesis and separation are often treated independently. This often results in the use of different solvents for each processing task, which is typically achieved by energy-intensive solvent swap operations.

In this chapter, a novel computer-aided approach that couples property prediction with process modelling and optimisation to simultaneously identify the optimal solvent/anti-solvent

mixtures, mixture composition and process conditions in integrated synthesis and crystallisation processes is proposed. Solvents are chosen based on a set of key performance indicators (KPIs) that quantify mass/energy efficiency and product quality. These indicators include the solvent E-factor, the process E-factor, crystal yield and process energy consumption (Section 3.2.4). The proposed methodology is illustrated by identifying reaction and crystallisation solvents for the synthesis of mefenamic acid from 2,3-dimethylaniline and 2-chlorobenzoic acid and its subsequent crystallisation. Furthermore, bi-objective optimisation is deployed to highlight the trade-offs between competing KPIs, such as the solvent E-factor and process safety, the solvent E-factor and crystal yield, and process energy consumption and crystal yield. The inclusion of KPIs and safety specifications in the proposed formulation ensures that only environmentally benign, high-performance solvents are chosen for the integrated reaction-crystallisation process.

3.2 Problem Definition and Formulation

3.2.1 General design problem formulation

The CAM^{b}D problem tackled in this chapter can be described as follows: Given a synthetic step in the manufacturing of a pharmaceutical compound and a specified production rate, reaction conversion and selectivity, product purity, and a list of possible solvents, identify the solvent/anti-solvent mixture, mixture composition and process conditions for the corresponding integrated synthesis and crystallisation process that optimise one or more KPIs.

A generic reaction-crystallisation process is depicted in Figure 3.1. This can be used to represent a batch or continuous process, where in the case of a batch process each stream represents the starting or end point of the batch operation. The process can be described as follows. Stream 1, the feed stream, consists of reactants and impurities dissolved in a pure solvent $s1$ or binary solvent mixture ($s1$+$s2$) that enter a heat exchanger at a temperature T_1 (K). The heated reaction mixture enters the reactor as Stream 2 at a temperature T_2 (K). The reaction proceeds in a single step and may be accompanied by side reactions. The stream exiting the reactor, Stream 3, is a homogeneous (liquid) mixture of the desired product **D**, unconverted reactants, impurities, and undesired side and byproducts at a temperature T_3 (K).

This stream is sent to a crystallisation unit, which can be designed as a cooling crystallisation, an anti-solvent crystallisation, or a hybrid of both techniques. When anti-solvent crystallisation is involved, Stream 4, consisting of pure solvent $s2$ at temperature T_3, is fed to the crystalliser. The crystalliser outlet is a slurry that consists of a liquid phase (Stream 5), that normally contains some amount of all process components, and a solid phase (Stream 6), typically pure **D**. Both streams are at a temperature T_5 (K).

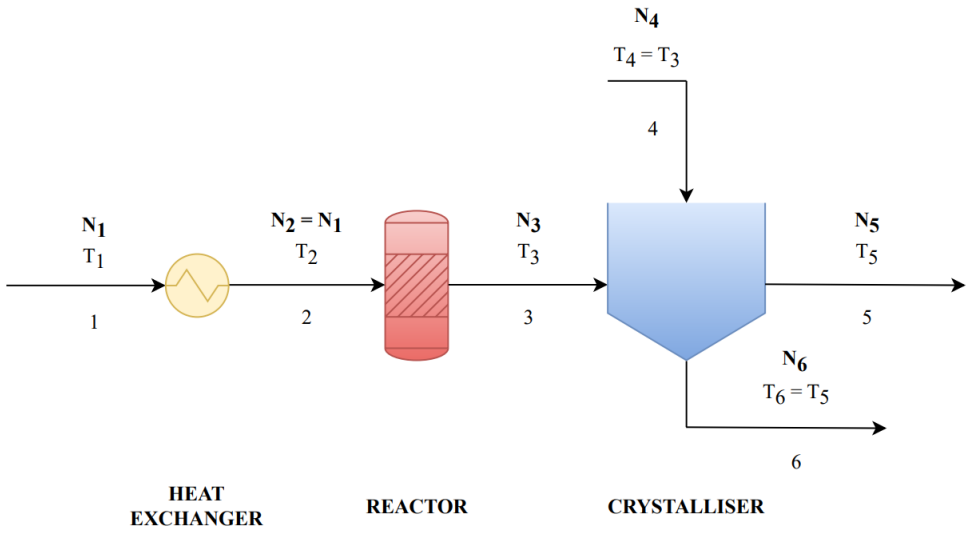

Figure 3.1: The conceptual flowsheet used for the solvent selection problem of the integrated synthesis and crystallisation process. N denotes the vector of component mole numbers (batch process) or component molar flowrates (continuous process) and T the vector of stream temperatures.

While multiple phenomena need to be taken into account to quantify the effect of solvents on KPIs [62], the proposed CAM^{b}D model focuses on thermodynamic driving forces, i.e., species solubility, in order to obtain a shortlist of solvent candidates that can then be tested through other considerations, such as kinetics, transport properties and crystal shape.

To facilitate the development of the model, specific definitions are assigned to the terms "component" and "species". "Component" refers to a constituent of a process stream, i.e., a reactant, an impurity, the desired product, an undesired byproduct or side product, solvent $s1$ or solvent $s2$. "Species" refers to the molecules that are on the list of solvents from which solvent $s1$ and solvent $s2$ can be chosen.

The following sets are defined to derive a generic formulation of the design problem:

- the set C comprises all the components that appear in any of the liquid streams;

- the set C^K comprises all the components that may appear in solid form in the process, i.e., whose melting point is within the range of process operating temperatures. In principle, this may introduce new components if co-crystals or solvates can form. However, here it is assumed that only pure crystal forms can be formed so that $C^K \subset C$;
- the set Q comprises all the solvents in the process – here it is assumed that there are up to two solvents, so that $Q = \{s1, s2\}$ – and therefore $Q \subset C$;
- the set S comprises the solvent candidates from which solvents $s1$ and $s2$ are chosen;
- the set of reaction indices R;
- the set $T = \{1, 2, 3, 4, 5, 6\}$ defines the process streams introduced earlier and shown in Figure 3.1.

The CAM^{b}D problem can be formulated using Generalized Disjunctive Programming (GDP) [103] to identify disjunctions in which one or two solvents are used. This can in turn be transformed into an MINLP using the big-M approach [159]. A multi-objective optimisation (MOO) version is considered so that the overall mathematical structure of the problem can be represented as follows:

$$
\begin{aligned}
\min_{\mathbf{x},\mathbf{y}} \quad & f_1(\mathbf{x}), ..., f_P(\mathbf{x}) \\
\text{s.t.} \quad & \mathbf{h}_0(\mathbf{x}) = \mathbf{0} \\
& \mathbf{g}_0(\mathbf{x}) \leq \mathbf{0} \\
& \mathbf{g}_{i,s}(\mathbf{x}) \leq \mathbf{M}_{i,s}(1 - y_{i,s}\mathbf{e}), i \in \{s1, s2\}; s \in S \\
& \mathbf{A}\mathbf{y} \leq \mathbf{a} \\
& \mathbf{x} \in [\mathbf{x}^L, \mathbf{x}^U] \subset \mathbb{R}^n \\
& \mathbf{y} \in \{0, 1\}^u
\end{aligned}
\tag{3.1}
$$

where each objective function $f_p(\mathbf{x})$, $p = 1, \ldots, P$ is a KPI that needs to be minimised, such as the solvent E-factor. $\mathbf{x}$ is an n-dimensional vector of continuous variables such as the number of moles, temperature and composition, and $\mathbf{y}$ is a u-dimensional vector of binary variables

representing the identities of the process solvents. This includes the variables $y_{i,s}$, $i \in \{s1, s2\}$, $s \in S$, which are defined as follows:

$$y_{i,s} = \begin{cases} 1, \text{ if solvent species } s \text{ is assigned to solvent component } i, \\ 0, \text{ otherwise.} \end{cases} \tag{3.2}$$

The vector of constraints $\mathbf{h}_0(\mathbf{x})$ describes the process models of the process units, the vector of constraints $\mathbf{g}_0(\mathbf{x})$ describes design specifications, the vector of constraints $\mathbf{g}_{i,s}(\mathbf{x})$ describes design and operating specifications that depend on solvent identity (whether solvent species s is assigned to solvent $i = s1$ or $i = s2$), $\mathbf{M}_{i,s}$ is a matrix of big-M values (that may be positive or negative), $\mathbf{e}$ is the unit vector, and $\mathbf{A}$ and $\mathbf{a}$ are a constant matrix and vector, respectively, used to express the logical relationships between the binary variables.

For design problems with multiple objectives ($P > 1$), MOO is deployed to explore the trade-offs between conflicting objectives. In this chapter, bi-objective optimisation (BOO) CAM^{b}D problems are solved using the ϵ-constraint method [176].

3.2.2 Process model and design constraints

Solvent assignment and properties

Solvent assignment constraints The process constraints are expressed in terms of solvents $s1$ and $s2$, which must be related to specific species in set S. This is achieved via assignment constraints.

The following constraint ensures that exactly one solvent from set S is assigned to solvent component $s1$:

$$\sum_{s \in S} y_{s1,s} = 1. \tag{3.3}$$

The following constraint ensures that at most one solvent from set S is assigned to solvent component $s2$:

$$\sum_{s \in S} y_{s2,s} \leq 1. \tag{3.4}$$

The following constraint ensures that each solvent from set S is selected at most once:

$$\sum_{q \in Q} y_{q,s} \leq 1, \quad s \in S. \tag{3.5}$$

Solvent properties The properties of $s1$ and $s2$ are obtained from the properties of the species in set S. The molecular weights of $s1$ and $s2$ are assigned based on the molecular weight MW_s of each solvent candidate s in set S:

$$MW_q = \sum_{s \in S} MW_s y_{q,s}, \quad q \in Q. \tag{3.6}$$

Similarly, the boiling and melting points of $s1$ and $s2$ are assigned based on the experimental values of the boiling point, $T_{b,s}$, and melting point, $T_{m,s}$, of each solvent candidate s in set S:

$$T_{b,q} = \sum_{s \in S} T_{b,s} y_{q,s}, \quad q \in Q, \tag{3.7}$$

$$T_{m,q} = \sum_{s \in S} T_{m,s} y_{q,s}, \quad q \in Q. \tag{3.8}$$

Masses and mole fractions

The mass and mole fraction of component $i \in C$ in stream $t \in T$ are given as follows:

$$M_{i,t} = N_{i,t} MW_i, \quad i \in C; t \in T, \tag{3.9}$$

$$x_{i,t} = \frac{N_{i,t}}{\sum_{i' \in C} N_{i',t}}, \quad i \in C; t \in T, \tag{3.10}$$

where $M_{i,t}$, $N_{i,t}$ and $x_{i,t}$ denote the mass, number of moles and mole fraction of component $i \in C$ in stream $t \in T$, respectively, and MW_i represents the molecular weight of component $i \in C$.

Furthermore, if solvent $s2$ does not exist, then the mole fraction $x_{s2,t}$ must be set to zero in the liquid-phase streams:

$$x_{s2,t} \leq \sum_{s \in S} y_{s2,s}, \quad t \in \{1, 2, 3, 4, 5\}. \tag{3.11}$$

Heat exchanger balances

Material balances must hold across the heat exchanger:

$$N_{i,1} = N_{i,2}, \quad i \in C. \tag{3.12}$$

Reactor balances and equilibrium constraints

Material balances must hold across the reactor:

$$N_{i,3} = N_{i,2} + \sum_{r \in R} \nu_{i,r} \xi_r, \quad i \in C, \tag{3.13}$$

where $\nu_{i,r}$ is the stoichiometric coefficient of component $i \in C$ in reaction $r \in R$ and ξ_r is the extent of reaction r. The extents of reaction are obtained from a user-specified conversion for the limiting reactant, x^c, and from $|R|$-1 user-specified selectivities, where given a set S, $|S|$ denotes the number of elements in the set, i.e., here $|R|$ is the number of reactions. Thus, in the simplest case of a single reaction, $R = \{1\}$, with limiting reactant A, the extent of reaction ξ_1 is given by:

$$\xi_1 = x^c N_{\mathrm{A},2}. \tag{3.14}$$

Next, a set of operating constraints is imposed on the reactor to ensure that all components are in the liquid phase. The following equations ensure that all components that may appear in the solid phase are under-saturated or dissolved in the feed to the reactor and in the reaction mixture leaving the reactor:

$$x_{i,t} \leq \hat{x}_{i,t}^{(i)} - \epsilon_s, \quad i \in C^K; t = \{2, 3\}, \tag{3.15}$$

where $\hat{x}_{i,t}^{(i)}$ is the solubility of component i in a hypothetical stream that is similar to stream t but saturated in i, and ϵ_s is a small, positive number.

To explain this more precisely: in a multi-component system in which no co-crystals are formed, the solubility of a given component is affected by the presence of all other components. If operating conditions are such that one or more components can crystallise ($|C^K| \geq 1$), a liquid

phase consisting of all components can be in equilibrium with 1 to $|C^K|$ pure crystalline phases. Here, to limit the complexity of the mathematical formulation, only cases where a liquid phase is in equilibrium with only one crystalline phase are considered. In such a case, the solubility of a component $i \in C^K$ in stream t can be calculated using the solid-liquid equilibrium expression [181, 182]:

$$\ln \hat{x}_{i,t}^{(i)} + \ln \hat{\gamma}_{i,t}(\hat{\mathbf{x}}_t^{(i)}, T_t) = \frac{\Delta H_{m,i}}{R_g}\left[\frac{1}{T_{m,i}} - \frac{1}{T_t}\right], \quad i \in C^K, \quad t \in \{2,3\}, \tag{3.16}$$

where $\hat{\gamma}_{i,t}$ is the liquid-phase activity coefficient of component i in a stream that is similar to stream t but saturated in i, $\Delta H_{m,i}$ is the molar heat of melting of component i, T_m is its melting point, T_t is the temperature of stream t, R_g is the gas constant, and for each $i \in C^K$, $\hat{\mathbf{x}}_t^{(i)}$ is a mole fraction vector such that:

$$\frac{\hat{x}_{i',t}^{(i)}}{\hat{x}_{s1,t}^{(i)}} = \frac{x_{i',t}}{x_{s1,t}}, \quad i' \in C \setminus \{s1, i\}, \quad i \in C^K, \quad t \in \{2,3\}, \tag{3.17}$$

$$\sum_{i' \in C} \hat{x}_{i',t}^{(i)} = 1, \quad i \in C^K, \quad t \in \{2,3\}. \tag{3.18}$$

Importantly, in the above expressions, the relative proportions of all components but i are the same as those in stream t. Component i is at solid-liquid equilibrium. The relative amount of component i is calculated to correspond to its mole fraction at solid-liquid equilibrium.

The liquid-phase activity coefficients of the components that may crystallise are calculated using the UNIFAC activity coefficient model [50, 181]. The equations proposed by Smith et al. [183] are convenient for implementation in this design problem and have therefore been used to evaluate activity coefficients.

The relevant UNIFAC equations are provided in Appendix A.

Crystalliser balances and equilibrium constraints

The model proposed by Jonuzaj et al. [58] is used to describe crystallisation in the integrated process, with some modifications. Kinetic considerations are neglected and the crystalliser is assumed to operate at thermodynamic equilibrium.

An impurity rejection of 100% is imposed in the crystalliser, so that only the desired product, **D**, is at solid-liquid equilibrium, while other solute impurities are below their saturation compositions at the end of crystallisation (i.e., in Stream 5). This results in the following material balances:

$$N_{i,3} + N_{i,4} - N_{i,5} - N_{i,6} = 0, \quad i \in C, \tag{3.19}$$

$$N_{i,4} = 0, \quad i \in C \setminus \{s2\}, \tag{3.20}$$

$$N_{i,6} = 0, \quad i \in C \setminus \{\mathbf{D}\}. \tag{3.21}$$

Liquid Stream 5 is at equilibrium with solid Stream 6, consisting of pure **D**:

$$\ln x_{D,5} + \ln \gamma_{D,5}(\mathbf{x}_5, T_5) = \frac{\Delta H_{m,D}}{R_g}\left[\frac{1}{T_{m,D}} - \frac{1}{T_5}\right], \tag{3.22}$$

where $\gamma_{D,5}(\mathbf{x}_5, T_5)$ is obtained using UNIFAC [50, 181].

Thermodynamic constraints must be included to ensure that only **D** is at solid-liquid equilibrium. A hypothetical mixture is considered for each solute i other than **D**, with mole fraction vector $\hat{\mathbf{x}}_5^{(i)}$, $i \in C^K \setminus \{\mathbf{D}\}$, in which the proportions of all components are the same as in Stream 5, but component i is at solid-liquid equilibrium. As for the reactor, this is given by:

$$\ln \hat{x}_{i,5}^{(i)} + \ln \hat{\gamma}_{i,5}(\hat{\mathbf{x}}_5^{(i)}, T_5) = \frac{\Delta H_{m,i}}{R_g}\left[\frac{1}{T_{m,i}} - \frac{1}{T_5}\right], \quad i \in C^K \setminus \{\mathbf{D}\}, \tag{3.23}$$

where $\hat{\gamma}_{i,5}(\hat{\mathbf{x}}_t^{(i)}, T_5)$ is calculated via UNIFAC and $\hat{\mathbf{x}}_5^{(i)}$ is such that:

$$\frac{\hat{x}_{i',5}^{(i)}}{\hat{x}_{s1,5}^{(i)}} = \frac{x_{i',5}}{x_{s1,5}}, \quad i' \in C \setminus \{s1, i\}, i \in C^K \setminus \{\mathbf{D}\}, \tag{3.24}$$

$$\sum_{i' \in C} \hat{x}_{i',5}^{(i)} = 1, \quad i \in C^K \setminus \{\mathbf{D}\}. \tag{3.25}$$

Under-saturation of the mixture is then achieved by imposing:

$$x_{i,5} \leq \hat{x}_{i,5}^{(i)} - \epsilon_s, \quad i \in C^K \setminus \{\mathbf{D}\}, \tag{3.26}$$

where ϵ_s is a small, positive number. Strictly speaking, one should also check the solubility limits obtained when the liquid phase is in equilibrium with two or more solid phases. The solubility limits for individual components are often higher in these situations than those observed for one solid phase only, so that neglecting this possibility leads to a conservative choice of solvent(s); accordingly, a small number of options that may be feasible cannot be identified.

3.2.3 Process design constraints

Constraints on operating temperatures

Since the process solvents must be in the liquid phase, constraints are imposed to ensure that the stream and unit temperatures are between the boiling and melting points of the selected solvents. Although it is possible to apply such constraints to solvent mixtures [57], here the simpler approach of applying these to each solvent on a pure-component basis is chosen, thus neglecting the dependence of the phase envelope on mixture composition. This can result in an overly conservative design as eutectic behaviour or higher mixture boiling points are not taken into account. In the case of low-boiling azeotropes, this can on the contrary lead to an overestimate of the liquid range. The mixture properties should therefore be checked after the optimisation is complete. The constraints on the operating temperatures are formulated via the big-M (BM) approach [159], using a temperature offset T_o, to ensure operation does not take place too close to solvent phase-change temperatures, and a large constant M_T:

$$T_t - (T_{b,q} - T_o) \leq M_T(1 - \sum_{s \in S} y_{q,s}), \quad q \in Q, t \in T, \tag{3.27}$$

$$(T_{m,q} + T_o) - T_t \leq M_T(1 - \sum_{s \in S} y_{q,s}), \quad q \in Q, t \in T, \tag{3.28}$$

where the solvent boiling and melting points are calculated using equations (3.7) and (3.8).

Solvent miscibility

When two solvents are selected, it is important to ensure that they form a mixture with a single liquid phase at the composition and conditions of interest. Accordingly, solvent miscibility constraints are included in the formulation. It is not possible to express such a constraint

in closed form for a mixture consisting of more than 2 components. An algebraic miscibility constraint can however be formulated for a binary mixture and this is used as an approximation here, as has been done in previous work [30, 31, 34, 184].

The hypothetical binary mixture is represented by mole fractions $\tilde{x}_{s1,t}$ and $\tilde{x}_{s2,t}$. These mole fractions are calculated by:

$$-\tilde{M}_x(1-\sum_{s\in S} y_{s2,s}) \leq \tilde{x}_{s1,t} - \frac{x_{s1,t}}{x_{s1,t}+x_{s2,t}} \leq \tilde{M}_x(1-\sum_{s\in S} y_{s2,s}), \quad t \in \{1,2,5\}, \tag{3.29}$$

where a BM formulation has been used with constant $\tilde{M}_x$. The miscibility constraint for the binary pair of solvent components $s1$ and $s2$ is similarly given by [183]:

$$\frac{\partial \ln \tilde{\gamma}_{s1,t}(\tilde{\mathbf{x}}_t, T_t)}{\partial \tilde{x}_{s1,t}} + \frac{1}{\tilde{x}_{s1,t}} \geq -M_\gamma(1-\sum_{s\in S} y_{s2,s}), \quad t \in \{1,2,5\}, \tag{3.30}$$

where $\frac{\partial \ln \tilde{\gamma}_{s1,t}(\tilde{\mathbf{x}}_t, T_t)}{\partial \tilde{x}_{s1,t}}$ is the analytical derivative of the UNIFAC model with respect to $\tilde{x}_{s1,t}$, $\tilde{\mathbf{x}}_t$ is a vector of mole fractions $\tilde{x}_{s1,t}$ and $\tilde{x}_{s2,t}$, and M_γ is a sufficiently large constant.

3.2.4 KPIs

The KPIs of interest in the CAM^{b}D formulation include the process E-factor [185], the solvent E-factor, crystal yield, process safety and process energy consumption. These KPIs are defined as follows:

- Process E-factor or PEF (kg Waste/kg Crystals) is defined as the mass of waste generated per mass of crystallised target product **D**:

$$\text{PEF} = \frac{\sum_{i\in C} M_{i,5}}{M_{D,6}}; \tag{3.31}$$

- Solvent E-factor or SEF (kg Solvents/kg Crystals) is defined as the solvent contribution to the PEF:

$$\text{SEF} = \frac{\sum_{q\in Q} M_{q,5}}{M_{D,6}}. \tag{3.32}$$

The SEF is bounded from below to ensure that a sufficient quantity of solvent is available

to facilitate slurry mixing in the crystalliser:

$$\text{SEF} \geq \text{SEF}_{\min}. \tag{3.33}$$

A value of $\text{SEF}_{\min} = 3.5$ kg/kg is typically chosen based on process experience [58].

- Crystal Yield or Y_C (%) is defined as the percentage in moles of product **D** crystallised out of solution:

$$Y_C = \frac{N_{D,6}}{N_{D,3}} \times 100\%. \tag{3.34}$$

The crystal yield is usually treated as a constraint, with a minimum target value $Y_{C,\min}$, typically set to 90%:

$$Y_C \geq Y_{C,\min}. \tag{3.35}$$

- Process safety is quantified by an overall safety indicator, I_T, for the selected solvent or solvent mixture. This is taken as the smallest indicator amongst a set F of indicators for each solvent component q in the mixture, denoted by $I_{q,f}$, $f = 1, \ldots, F$. Hence, I_T is given by

$$I_T = \min_{q \in \{s_1, s_2\}, f = 1, \ldots, F} \{I_{q,f}\}, \tag{3.36}$$

where the safety indicators of the selected solvents are expressed in terms of the safety indicators of the solvent species considered, $I_{s,f}$, as follows:

$$I_{q,f} = \sum_{s \in S} I_{s,f}\, y_{q,s}, \quad q \in Q; f \in F. \tag{3.37}$$

In order to remove the discontinuity imposed by equation (3.36), the problem is reformulated as a set of inequality constraints that can be used provided that I_T is being maximised. Each safety indicator of solvent component q in the mixture is thus bounded

from below as follows:

$$I_{s1,f} \geq I_T, \quad f \in F, \tag{3.38}$$

$$I_{s2,f} \geq I_T - M_S(1 - \sum_{s \in S} y_{s2,s}), \quad f \in F, \tag{3.39}$$

where M_S is a sufficiently large constant.

The process safety KPI is embedded in a BOO CAM^{b}D formulation in which the SEF or PEF is minimised and I_T is maximised or bounded from below. For example, the objective of the problem involving SEF minimisation and I_T maximisation can be represented as follows:

$$\min \text{SEF}, \quad \max I_T. \tag{3.40}$$

This is then reformulated into the equivalent series of single level optimisation problems:

$$\begin{aligned} \min \quad & \text{SEF} \\ \text{s.t.} \quad & I_{s1,f} \geq \epsilon, \quad f \in F \\ & I_{s2,f} \geq \epsilon - M_S(1 - \sum_{s \in S} y_{s2,s}), \quad f \in F, \end{aligned} \tag{3.41}$$

where the equivalence is achieved by using a single value of ϵ.

- Process energy consumption Q_T (kJ/kg API) is expressed as the sum of the heat duties of the individual process units. This includes the heat exchanger duty to pre-heat the reactor feed, the reactor duty to maintain a specific reaction temperature, and the crystalliser duty to cool the API mixture and induce product crystallisation. In this section, the process heat duties are expressed by performing energy balances around the process units, under the assumption of ideal mixing.

  1. Heat exchanger duty Q_H. The energy balance around the heat exchanger is given by:

$$Q_H = \sum_{i \in C} N_{i,1} \int_{T_1}^{T_2} C_{p_i}^{(L)} dT, \tag{3.42}$$

where $C_{p_i}^{(L)}$ is the liquid-phase molar heat capacity of component $i \in C$.

2. Reactor duty Q_R. The energy balance around the reactor is given by:

$$Q_R = \sum_{i \in C} N_{i,2} \int_{T_2}^{T_R} C_{p_i}^{(L)} dT + \xi_1 \Delta H_R(T_R) + \sum_{i \in C} N_{i,3} \int_{T_R}^{T_3} C_{p_i}^{(L)} dT, \tag{3.43}$$

where T_R is a reference temperature (usually 298.15 K) and ΔH_R is the standard heat of reaction.

3. Crystalliser duty Q_C^-. The energy balance around the crystalliser is given by:

$$Q_C^- = -Q_C = H_F - (H_L + H_S), \tag{3.44}$$

where H_F is the crystalliser feed enthalpy, H_L is the mother liquor enthalpy and H_S is the crystallised product enthalpy. The minus sign in the superscript of Q_C^- is used to indicate that the negative crystalliser duty is multiplied by -1 so that all energy contributions are reported as positive quantities for consistency. Assuming that only the API crystallises, the enthalpies in equation (3.44) are reformulated as follows:

$$H_F = \sum_{i \in C} N_{i,3} \int_{T_0}^{T_3} C_{p_i}^{(L)} dT + \sum_{i \in C} N_{i,4} \int_{T_0}^{T_3} C_{p_i}^{(L)} dT, \tag{3.45}$$

$$H_L = \sum_{i \in C} N_{i,5} \int_{T_0}^{T_5} C_{p_i}^{(L)} dT, \tag{3.46}$$

$$H_S = N_{D,6} \left(\int_{T_0}^{T_{m,D}} C_{p_{API}}^{(L)} dT + \Delta H_{solid}(T_{m,D}) + \int_{T_{m,D}}^{T_5} C_{p_{API}}^{(S)} dT \right), \tag{3.47}$$

where T_0 is a reference temperature, ΔH_{solid} is the enthalpy of solidification of the API and $C_{p_{API}}^{(S)}$ is the solid-phase molar heat capacity of the API.

Setting $T_0 = T_3$ and combining equations (3.44)–(3.47) gives:

$$Q_C^- = -\left[\sum_{i \in C} N_{i,5} \int_{T_3}^{T_5} C_{p_i}^{(L)} dT + N_{D,6} \left(\int_{T_3}^{T_{m,D}} C_{p_{API}}^{(L)} dT + \Delta H_{solid}(T_{m,D}) + \int_{T_{m,D}}^{T_5} C_{p_{API}}^{(S)} dT \right) \right], \tag{3.48}$$

or equivalently:

$$Q_C^- = \sum_{i \in C} N_{i,5} \int_{T_5}^{T_3} C_{p_i}^{(L)} dT + N_{D,6} \left(\int_{T_{m,D}}^{T_3} C_{p_{API}}^{(L)} dT + \Delta H_{m,D}(T_{m,D}) + \int_{T_5}^{T_{m,D}} C_{p_{API}}^{(S)} dT \right). \tag{3.49}$$

The energy KPI can be expressed as:

$$Q_T = Q_H + |Q_R| + Q_C^-, \tag{3.50}$$

where the absolute value of Q_R and the minus sign in Q_C^- are used to ensure consistent signs of the energy duties, i.e., Q_H, Q_R and $Q_C^- \geq 0$.

This KPI can be minimised in a single or multi-objective optimisation problem. The absolute value term in equation (3.50) can be removed by reformulating the energy KPI objective function as follows:

$$\min \qquad Q_H + Q'_R + Q_C^-, \tag{3.51}$$

and adding the following two constraints to the full formulation:

$$Q_R \leq Q'_R, \tag{3.52}$$

$$-Q_R \leq Q'_R, \tag{3.53}$$

where Q'_R is a variable used to incorporate the absolute value of Q_R in the expression of Q_T.

3.3 Case Study: The Synthesis and Crystallisation of Mefenamic Acid

The single-step synthesis and crystallisation of mefenamic acid (MA), a non-steroidal anti-inflammatory drug with analgesic properties, from 2,3-dimethylaniline (DMA) and 2-chlorobenzoic acid (CBA) is used as a case study to illustrate the computer-aided solvent selection approach.

This Ullmann reaction is conducted in the presence of a copper-based catalyst and a carbonate salt, and the MA salt produced by the reaction is acidified to MA. HCl and KCl are typically obtained as byproducts [186]. The overall reaction scheme is shown in Figure 3.2, omitting the inorganic species. The reaction has been reported to proceed effectively at temperatures between 358 and 403 K [187], and the reaction has been carried out with a DMA-to-CBA ratio of 2:1 in several experimental studies [186, 187, 188, 189]. Accordingly, the reactor temperature range and feed ratio considered in this work will match those reported in the aforementioned studies.

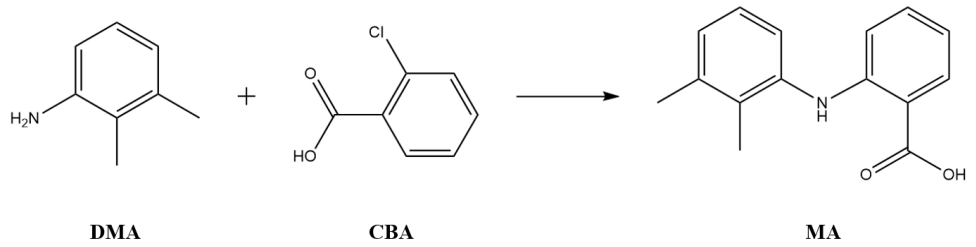

Figure 3.2: Reaction of 2,3-dimethylaniline (DMA) and 2-chlorobenzoic acid (CBA) to form mefenamic acid (MA).

The proposed methodology is applied to the synthesis and crystallisation of mefenamic acid as this relatively simple manufacturing process, with its one-step synthesis, provides a useful case study to explore a range of problem formulations and gain some insights into trade-offs in solvent design. To adapt the general model of Section 3.2, the following sets are specified: $C = \{DMA, CBA, MA, s1, s2\}$, $C^K = \{DMA, CBA, MA\}$ and $Q = \{s1, s2\}$. The set S of candidate solvents considered in Jonuzaj et al. [58] is used in this chapter as it is representative of pharmaceutical industry solvents, with potentially reactive solvent pairs excluded. The solvent list thus consists of 58 solvent molecules and is provided in Appendix B.2. The CAM^{b}D model is developed using the specifications and parameters shown in Tables 3.1 and 3.2, respectively. The CAM^{b}D formulation is used as a screening tool to identify the most promising solvents on the basis of KPIs considered. In developing the model, the following considerations are neglected:

- the effect of the solvent on reaction and crystallisation kinetics;
- the effect of the inorganic salts and catalyst on mixture thermodynamics.

Table 3.1: Specifications of the CAM^{b}D formulation of the MA case study.

Constraint	Equation	Units (where required)
Feed temperature	$T_1 = 298.15$	K
Isothermal reactor operation	$T_2 = T_3$	K
2:1 DMA:CBA Feed Ratio	$N_{DMA,2} = 2N_{CBA,2}$	mol
No MA in process feed	$N_{MA,1} = 0$	mol
Reaction conversion	$x^c = 0.9$	–
Throughput	$M_{MA,6}$=80	kg
Reaction temperature range	$358 \leq T_2 \leq 403$	K
Final crystallisation temperature range	$290 \leq T_5 \leq 380$	K

Table 3.2: Parameters of the CAM^{b}D formulation of the MA case study.

Parameter	Value	Units (where required)
SEF_{min}	3.5	kg/kg
$Y_{C,min}$	90	%
T_o	10	K
M_T	600	–
$\tilde{M}_x$	10	–
M_γ	100	–
M_S	100	–

Accordingly, the prioritised list of solvents generated by CAM^{b}D will need to be explored further experimentally to assess these effects.

The impurities considered in this case study are the unreacted amounts of DMA and CBA as these compounds belong to set C^K and can hence crystallise. However, since DMA exists in the liquid phase within the process temperature range considered, the mathematical formulation includes solubility equations for only CBA and MA.

Eight CAM^{b}D formulations are considered for this process. In the first 3 formulations, the SEF is used as the objective function, while the PEF is used in the fourth formulation. In the fifth and sixth formulations, a BOO optimisation problem is formulated in which the SEF or the PEF is minimised while the safety performance of the design is maximised. In the seventh formulation, a BOO optimisation problem is formulated in which the SEF is minimised while Y_C is maximised. In the eighth and final formulation, a BOO optimisation problem is formulated in which Q_T is minimised while Y_C is maximised. Through the first 3 formulations, the outcomes of considering sequential or integrated design and of limiting the set of possible solvents based on the International Council for Harmonisation (ICH) guidelines for residual solvents [190] are compared. In the first design problem, the optimal solvents are selected from the list of 43

solvent candidates (List I) given in Appendix B.1. In design problems 2–7, the optimal solvents are selected from the list of 58 solvent candidates (List II) given in Appendix B.2. In the final design problem, the optimal solvents are selected from the list of 49 solvent candidates (List III) given in Appendix B.3 – this final list is a shortened version of List II that includes solvents with available thermal properties needed for energy balance calculations.

Solvent melting and boiling points are obtained from the Chemical Book (`https://www.chemicalbook.com/ProductIndex_EN.aspx`) [191]. The melting points of CBA, DMA and MA are obtained from the Chemical Book (`https://www.chemicalbook.com/ProductIndex_EN.aspx`), the heats of melting of CBA and MA are obtained from the National Institute of Standards and Technology (NIST) Chemistry WebBook (`https://webbook.nist.gov/chemistry`) [192], and the heat of melting of DMA is estimated using the Marrero-Gani GC method [121]. Values of $T_{m,i}$ and $\Delta H_{m,i}$ for $i \in C^K$ are reported in Appendix C.1.

In all energy balances, heat capacity is assumed to be phase-independent, i.e., $C_{p_i}^{(L)} \approx C_{p_i}^{(S)}$ for $i \in C$, and to be a weak function of temperature, i.e., standard constant-pressure heat capacities of the chemical species are used. Standard constant-pressure heat capacities of solvent candidates in set S are obtained from the NIST Chemistry WebBook (`https://webbook.nist.gov/chemistry`), whereas standard constant-pressure heat capacities of reactants and products are calculated using the GC method of Kolská et al. [193]. Values of $C_{p_i}^{(L)}$ for $i \in C^K$ are reported in Appendix C.1. The standard heat of reaction of MA synthesis is calculated from the heats of formation of MA, CBA and DMA, which are estimated using the Marrero-Gani GC method [121], and is reported in Appendix C.2.

The optimisation problems are solved in GAMS version 32.2.0 [169] using Simple Branch and Bound (SBB) [163], a local branch and bound MINLP solver. Multiple high-performance solutions are generated for each design problem by including integer cuts in the MINLP formulation, and are ranked with respect to the objective function value. For each solution, the following quantities are reported: the identities of solvent species $s1$ and $s2$, the value of the objective function(s), the reaction temperature T_2, the final crystallisation temperature T_5, the composition of the solvent in the reactor, as the ratio of the mole fraction of solvent 1 to that of solvent 2 in the reactor outlet (Stream 2), the composition of the solvent in the crystalliser, as

the ratio of the mole fraction of solvent 1 to that of solvent 2 in the crystalliser outlet (Stream 5) and the total solvent mass in Stream 5. Furthermore, for all integrated, single-objective design formulations, both the SEF and the PEF are reported.

3.3.1 Minimising the SEF

Integrated reaction-crystallisation design: Reduced solvent space

The first formulation of the design problem involves minimising the SEF of the integrated reaction-crystallisation process, while considering a design space that excludes solvents with significant toxic potential as defined by the ICH Guideline for Residual Solvents [190], which classifies solvents based on their risk to human health. Solvents belonging to Classes 1 and 2, which represent solvents that must be avoided or limited in pharmaceutical manufacturing, respectively, are discarded from set S (Appendix B.2), and the optimisation problem is solved to identify optimal solvents out of a list of 43 candidates (Appendix B.1). The list does not contain N,N-dimethylformamide (DMF) nor N,N-dimethylacetamide (DMAC), two ICH Class 2 solvents which have previously been used for mefenamic acid synthesis [188, 187] and/or crystallisation [194, 195]. It contains other reported crystallisation solvents such as Ethyl acetate, Acetone, Isopropanol and Ethanol [194, 195]. The results are shown in Table 3.3 and it is found that none of the solvents previously reported in the literature is identified as high-performance. It can be seen that the SEF of all top ten solutions is at the lower bound of 3.5 kg/kg, which corresponds to a total solvent mass of 280 kg. Furthermore, all reported solutions correspond to a cooling crystallisation process with no anti-solvent addition involved, as evidenced by the temperature decrease from T_2 to T_5 and the constant ratio of solvent mole fractions across the crystalliser. An interesting finding is that Nitrobenzene (Solution 1) is identified as a pure solvent that can be deployed in both synthesis and crystallisation, highlighting the flexibility of the proposed formulation to identify both pure and binary solvent mixtures for integrated process design. However, since Nitrobenzene has reproductive toxicity risk and is a suspected carcinogen, its use would not be acceptable in processes and hence the other solutions that do not include Nitrobenzene as a solvent component would be more suitable solvent mixture choices.

Table 3.3: Top ten solutions of the SEF minimisation problem while considering a list of solvents that excludes ICH Classes 1 and 2 solvents.

Solution	Solvents	SEF (kg/kg)	PEF (kg/kg)	T_2 (K)	T_5 (K)	$\frac{x_{s1,2}}{x_{s2,2}}$	$\frac{x_{s1,5}}{x_{s2,5}}$	$\sum_{q\in Q} M_{q,5}$ (kg)
1	s1: Nitrobenzene s2: –	3.5	4.37	403	320	-	-	280
2	s1: Isobutyl acetate s2: Anisole	3.5	4.25	389	297	0.80	0.80	280
3	s1: n-Butanol s2: Anisole	3.5	4.28	376	293	0.30	0.30	280
4	s1: Cyclohexanone s2: Anisole	3.5	4.24	383	290	0.30	0.30	280
5	s1: Cyclohexanone s2: Acetic acid	3.5	4.26	381	301	0.28	0.28	280
6	s1: Cyclohexanone s2: Nitrobenzene	3.5	4.23	388	290	0.60	0.60	280
7	s1: Nitrobenzene s2: Isobutyl acetate	3.5	4.28	389	305	9.00	9.00	280
8	s1: Acetic acid s2: Isobutyl acetate	3.5	4.26	378	299	6.06	6.06	280
9	s1: Nitrobenzene s2: 1,2-Propanediol	3.5	4.37	382	314	0.57	0.57	280
10	s1: Acetic acid s2: Anisole	3.5	4.35	373	306	2.31	2.31	280

Integrated reaction-crystallisation design: Full solvent space

The second formulation of the design problem involves minimising the SEF of the integrated process while considering a larger solvent design space of 58 solvents, including some solvents belonging to ICH Classes 1 and 2 (Appendix B.2). This allows exploring the applicability of the proposed approach to larger solvent design spaces, as the introduction of an additional 15 solvents leads to an 80% increase in the number of pure solvents and binary mixtures considered. ICH Class 1 and 2 solvents can be used in some circumstances and the expansion of the solvent

set makes it possible to increase the diversity of the solvent candidates that can be investigated in subsequent stages of solvent selection. The top ten solutions are shown in Table 3.4. As expected with the increased design space, the optimal value of the SEF obtained in all cases remains at the lower bound of 3.5 kg/kg. Furthermore, all reported designs involve a binary solvent mixture and correspond to a cooling crystallisation process. The constant ratio of solvent mole fractions in the reactor and the crystalliser indicates that no anti-solvent addition is involved in the reported solutions. The slight difference in PEF values across the different solutions is attributed to the difference in crystal yields, where lower PEFs correspond to higher yields (Section 3.3.2). There is a variation of only 3% in the values of the PEF even though it is not used as an explicit objective.

Table 3.4: Top ten solutions of the SEF minimisation problem while considering the full list S of solvent candidates.

Solution	Solvents	SEF (kg/kg)	PEF (kg/kg)	T_2 (K)	T_5 (K)	$\frac{x_{s1,2}}{x_{s2,2}}$	$\frac{x_{s1,5}}{x_{s2,5}}$	$\sum_{q\in Q} M_{q,5}$ (kg)
1	s1: Chlorobenzene s2: Anisole	3.5	4.25	394	305	2.07	2.07	280
2	s1: n-Butyl acetate s2: Anisole	3.5	4.24	386	290	0.40	0.40	280
3	s1: Acetic acid s2: Anisole	3.5	4.31	373	299	0.98	0.98	280
4	s1: 1,2-Propanediol s2: Cyclohexanone	3.5	4.37	403	332	3.78	3.78	280
5	s1: o-Xylene s2: Cyclohexanone	3.5	4.37	401	332	0.12	0.11	280
6	s1: Water s2: DMF	3.5	4.33	363	290	0.55	0.55	280
7	s1: 2-Pentanone s2: DMF	3.5	4.37	364	290	1.08	1.08	280
8	s1: Acetic acid s2: o-Xylene	3.5	4.37	381	318	2.40	2.40	280
9	s1: DMAC s2: o-Xylene	3.5	4.37	403	338	0.15	0.15	280
10	s1: DMAC s2: t-AmOH	3.5	4.37	365	290	0.80	0.80	280

The CPU times (in seconds) needed to generate the solutions in Tables 3.3 and 3.4 are reported in Table 3.5. It can be seen that expanding the design space does not lead to a significant increase in computational time (approximately 300 CPU seconds only) which highlights the benefits of using a computer-aided tool to explore a vast design space of solvent candidates and process conditions.

Table 3.5: CPU times of the reduced space and full space design formulations.

Solution	CPU Time – Reduced Space (s)	CPU Time – Full Space (s)
1	1352.45	1374.19
2	96.04	108.60
3	422.17	67.39
4	136.56	560.68
5	480.54	121.11
6	342.67	78.69
7	618.79	376.68
8	241.26	963.71
9	480.26	180.84
10	976.95	1617.24
Total CPU Time (s)	5147.69	5449.13

Sequential reaction-crystallisation design

In order to quantify the benefits of an integrated approach to reaction and crystallisation, the CAM^{b}D problem is solved sequentially to compare the outcomes of a sequential and an integrated approach to process design, using the larger design space (Appendix B.2). This is achieved by formulating two subproblems. In the first subproblem, the optimal solvents and process conditions that minimise the SEF in the reactor only are identified. Here the reactor SEF, SEF_R, is defined as the ratio of the mass of process solvents consumed to MA produced in the reaction mixture based on the previously-defined conversion:

$$\text{SEF}_\text{R} = \frac{\sum_{q \in Q} M_{q,3}}{M_{MA,3}}. \tag{3.54}$$

This problem includes equations (3.3) – (3.18), (3.27) – (3.30), (3.54) and the reactor-specific equations in Table 3.1. This reactor design problem yields optimal solvent identities, $\mathbf{y}^*_{s1}$ and $\mathbf{y}^*_{s2}$, as well as Stream 3 composition $\mathbf{x}^*_3$ and temperature $T^*_3 = T^*_2$. Then, a second optimisation

subproblem, which is a restricted version of the integrated CAM^{b}D problem with the objective of minimising the overall SEF, is formulated by fixing the solvent identities throughout the process to $\mathbf{y}_{s1}^*$ and $\mathbf{y}_{s2}^*$ and the temperature in Stream 3 to T_3^*. This problem includes equations (3.3) – (3.35), as well as the specifications listed in Table 3.1. Note that the composition of the solvent mixture is allowed to vary to ensure that a feasible solution can be found.

The second optimisation subproblem is solved for the ten best solutions generated by solving the reactor design problem. The resulting designs, ranked with respect to the overall SEF, are shown in Table 3.6. In all cases, the crystallisation process is selected to be a cooling crystallisation, as is evident from the decrease in temperature from T_2 to T_5 and the fact that the ratios of solvent mole fractions in streams 2 and 5 are identical, indicating that no anti-solvent is added. Furthermore, all solutions correspond to a binary solvent mixture that is used across both synthesis and crystallisation. The solvent blends identified by the algorithm are diverse, giving the designer several viable options to investigate further.

Focusing specifically on the implications of taking a sequential approach, it can be seen that in all cases, the SEF is greater than the SEF_R, which means that a greater solvent amount is required to meet the design objectives and specifications when extending the reaction system to a reaction-crystallisation process. This is because solvents which may be efficient for synthesis in terms of mass consumption may not be so for crystallisation. This can be seen particularly for solvent blends such as Isobutyl Acetate + 1,4-Dioxane (Solution 9) or 2-Pentanone + DMAC (Solution 10), which exhibit comparatively low SEF_R values but an overall SEF 20-30% larger than the best solutions (1 and 2). This highlights some of the limitations of a sequential approach to solvent and process design. In fact, there might be other solvents/solvent blends with a higher SEF_R that perform better in terms of overall SEF than any of the blends listed in Table 3.6. Identifying these solvents would require solving more instances of the sequential problem or addressing the integrated problem directly.

The results of the sequential and integrated approaches highlight one of the main benefits of allowing process conditions such as the reaction and crystallisation temperatures to vary, namely the generation of different designs/solvent mixtures that may not be optimal for fixed conditions. Accordingly, the set of possible choices is broadened and the interrelation between

process conditions and solvent choice is thoroughly captured. This can be even more clearly realised when comparing the SEF values in Table 3.4 and the results of the sequential design problem in Section 3.3.1, with SEF values reported in Table 3.6. None of the solutions identified using the sequential approach reach the minimum value of 3.5 kg/kg for the overall SEF. Yet, it is clear from Tables 3.3 and 3.4 that all top ten solutions make it possible to reach this value, but this requires investigating different chemical families. For instance, the integrated design solutions indicate that the greater use of ketones and aromatic compounds is beneficial.

Table 3.6: Top ten solutions of the sequential reaction-crystallisation SEF minimisation problem. Columns 2-4 show the optimal solution of the reactor-only problem, while columns 5-9 show the results of the restricted problem, with solvent identities and reactor temperature fixed at the reactor-only solution.

Solution number	Reactor design only			Restricted problem				
	Solvents	SEF_R (kg/kg)	T_2 (K)	SEF (kg/kg)	T_5 (K)	$\frac{x_{s1,2}}{x_{s2,2}}$	$\frac{x_{s1,5}}{x_{s2,5}}$	$\sum_{q \in Q} M_{q,5}$ (kg)
1	s1: 1,2-Propanediol s2: n-Butyl acetate	3.04	389	3.51	290	1.64	1.64	281
2	s1: 1,2-Propanediol s2: Isobutyl acetate	3.03	389	3.51	290	1.64	1.64	281
3	s1: Cyclopentyl methyl ether s2: Formic acid	3.52	364	3.75	292	0.69	0.69	300
4	s1: Cyclopentyl methyl ether s2: Acetic acid	3.53	369	3.80	299	0.39	0.39	304
5	s1: 1,2-Propanediol s2: 1,4-Dioxane	2.61	364	3.83	295	0.63	0.63	307
6	s1: Cyclopentyl methyl ether s2: 1,4-Dioxane	2.82	364	4.03	295	0.78	0.78	323
7	s1: Isobutyl acetate s2: Cyclohexanone	3.93	389	4.04	290	0.11	0.11	324
8	s1: Cyclopentyl methyl ether s2: Nitromethane	3.89	364	4.11	290	0.66	0.66	329
9	s1: Isobutyl acetate s2: 1,4-Dioxane	2.98	364	4.21	295	0.40	0.40	338
10	s1: 2-Pentanone s2: DMAC	1.29	365	4.67	290	7.08	7.08	374

3.3.2 Minimising the PEF

Given the large number of solutions that achieve the lowest SEF value, the PEF is considered as a broader metric. As explained in Section 3.2.4, the PEF is an extension of the SEF that additionally includes the uncrystallised amounts of DMA, CBA and MA in the waste term. In the fourth formulation of the design problem, the PEF is set as the objective function and the larger design space of 58 solvent candidates is used. The results are shown in Table 3.7. Similar to Section 3.3.1, Nitrobenzene appears as an optimal solvent in solution 1. Nevertheless, the generation of multiple optimal solutions using integer cuts in the mathematical formulation allows the practitioner to choose alternative safer solvents for the integrated process. In all cases, the solutions obtained combine the benefit of a lower PEF with a SEF of 3.5 kg/kg. The use of the PEF metric, as a more holistic objective function, leads to significant changes in the designs generated. All solutions involve a binary solvent mixture. Solutions 1, 5 and 6 correspond to a hybrid cooling and anti-solvent crystallisation process as shown in the change in both temperature and ratio of solvent mole fractions between streams 2 and 5 due to the addition of $s2$ as an anti-solvent to the crystallisation mixture. The more frequent appearance of solutions with anti-solvent crystallisation can be attributed to the fact that the PEF can be decreased by reducing the amount of MA in the final waste stream, or equivalently, increasing the crystal yield of MA. Hence, it is likely that a combination of both cooling and anti-solvent addition would be required to achieve this more demanding process requirement. It is also worthy of mention that the difference between temperatures T_2 and T_5 is larger than that in most of the previously reported solutions: T_2 is at its upper bound in most solutions and T_5 is at its lower bound in all solutions. This can again be attributed to the drive towards process conditions that favour the minimisation of the PEF.

Table 3.7: Top ten solutions of the PEF minimisation problem.

Solution	Solvents	PEF (kg/kg)	SEF (kg/kg)	T_2 (K)	T_5 (K)	$\frac{x_{s1,2}}{x_{s2,2}}$	$\frac{x_{s1,5}}{x_{s2,5}}$	$\sum_{q\in Q} M_{q,5}$ (kg)
1	s1: Nitrobenzene s2: o-Xylene	4.20	3.5	403	290	0.80	0.75	280
2	s1: Cyclohexanone s2: o-Xylene	4.20	3.5	403	290	2.57	2.57	280
3	s1: Anisole s2: o-Xylene	4.20	3.5	403	290	0.40	0.40	280
4	s1: DMAC s2: o-Xylene	4.20	3.5	403	290	0.12	0.12	280
5	s1: DMAC s2: Water	4.20	3.5	363	290	9.00	0.11	280
6	s1: DMF s2: 1,2-Propanediol	4.21	3.5	403	290	0.25	0.24	280
7	s1: 1,2-Propanediol s2: Anisole	4.21	3.5	403	290	8.64	8.64	280
8	s1: DMF s2: Chlorobenzene	4.21	3.5	394	290	0.15	0.15	280
9	s1: Chlorobenzene s2: 1,2-Propanediol	4.22	3.5	394	290	0.32	0.32	280
10	s1: DMAC s2: Chlorobenzene	4.22	3.5	392	290	0.11	0.11	280

3.3.3 Multi-objective optimisation: Exploring trade-offs between competing KPIs

Minimising solvent use and maximising safety

To ensure that only safe solvents are selected for the process, the design problem is extended to a BOO CAM^{b}D formulation in which the SEF or the PEF is minimised while the safety, health and environmental (SHE) performance of the design is maximised. This formulation includes the equations and specifications of the integrated design problem, as well as equations (3.37)

– (3.39). As mentioned in Section 3.2.4, the overall safety indicator I_T for a given solvent mixture is taken as the smallest (i.e., worst) amongst a set of SHE indicators such as flammability, health and environmental impacts, and reactivity, for each solvent component in the mixture. The specific indicator values for the solvent species considered, $I_{s,f}$, are obtained from the GSK solvent selection guide [23]. The indicator scales are such that a greater value indicates better performance, with values ranging from 1 to 10. The BOO problem is solved using the ϵ-constraint method [176] in which the SEF or the PEF is minimised while I_T is constrained by a given lower bound ϵ, or equivalently, the safety indicators of each solvent q in the mixture, $I_{q,f}$, are constrained by the same value of ϵ, which represents the minimum acceptable value of I_T. For example, if a value $\epsilon = 6$ is used, then all solvent SHE indicators, $I_{q,f}$, must be greater than or equal to 6. Each BOO problem is solved for $1 \leq \epsilon \leq 7$ to generate its corresponding set of optimal solutions (the Pareto front) [171], since for $8 \leq \epsilon \leq 10$, no feasible solutions were identified despite trying different initial guesses. This means that for the given design formulation and considered case study, no feasible designs with GSK SHE indicators beyond 7 can be obtained.

Minimising the SEF and maximising SHE performance The fifth formulation of the design problem is a BOO problem in which the SEF is minimised while the value of I_T is maximised. Solutions of this problem are shown in Figure 3.3 and details of each solution are provided in Table 3.8. The solutions consist of different solvent mixtures, mixture compositions and process conditions. It can be seen that the point with $I_T = 6$ corresponds to three distinct Pareto-optimal designs (solvent choices and process conditions). Furthermore, the points (SEF,I_T) = (3.5,6) and (5.8,7) represent the Pareto-optimal solutions of the problem as they highlight the trade-off between the SEF and I_T. More specifically, solutions for which I_T is greater than 6 can only be found with a significant increase in the SEF from 3.5 kg/kg to at least 5.8 kg/kg. Additional solutions that achieve the minimum SEF value of 3.5 kg/kg, but with a deterioration of I_T, are also reported and plotted as open blue squares in Figure 3.3. The small number of Pareto-optimal solutions, despite the existence of many solvents and solvent mixtures that achieve the minimum SEF, illustrates how highly constrained the solvent selection problem is and how unlikely one is to find solvents that achieve a good balance

between competing objectives if using a trial-and-error approach. The solutions of the sequential problem in Section 3.3.1 are plotted as green triangles in Figure 3.3. All sequential solutions are dominated, i.e., they are not on the Pareto frontier, highlighting the benefits of both taking an integrated approach to solvent selection and considering multiple objectives simultaneously.

Table 3.8: Pareto-optimal solutions of the BOO problem for minimising the SEF and maximising SHE performance. Additional solutions which achieve the lowest value of SEF with a deterioration of SHE performance are also reported and indicated with a *.

Solution	ϵ	Solvents	SEF (kg/kg)	I_T	T_2 (K)	T_5 (K)	$\frac{x_{s1,2}}{x_{s2,2}}$	$\frac{x_{s1,5}}{x_{s2,5}}$	$\sum_{q\in Q} M_{q,5}$ (kg)
1	2	s1: Anisole s2: n-Butyl acetate	3.5	6	390	290	1.18	1.18	280
2	5	s1: Acetic acid s2: Anisole	3.5	6	374	299	9.00	9.00	280
3	6	s1: Anisole s2: 1,2-Propanediol	3.5	6	376	290	9.00	9.00	280
4	7	s1: Isobutyl acetate s2: Isobutanol	5.8	7	371	290	0.40	0.40	464
5*	1	s1: Nitrobenzene s2: Acetic acid	3.5	1	381	316	0.11	0.11	280
6*	3 or 4	s1: Anisole s2: Chlorobenzene	3.5	4	385	318	9.00	9.00	280

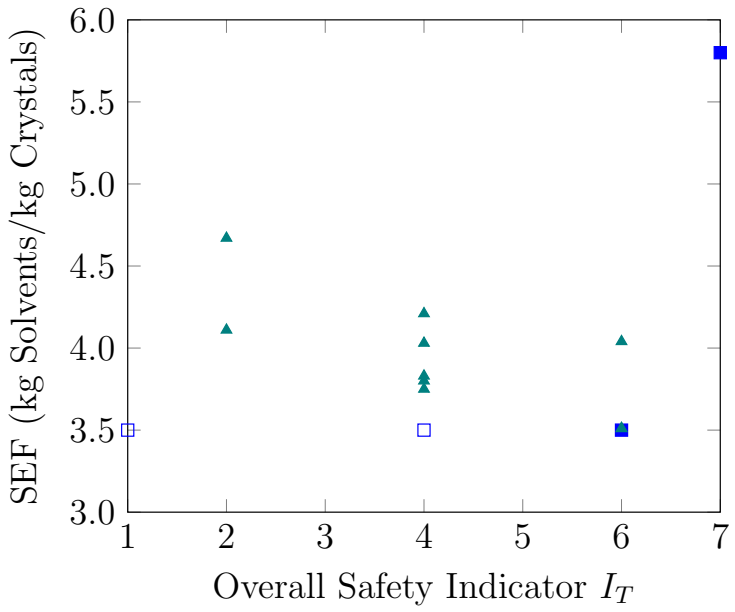

Figure 3.3: Pareto front of the BOO problem for minimising the SEF and maximising SHE performance shown with solid blue squares. The point with $I_T = 6$ represents 3 distinct solutions. The open blue squares denote points that achieve the global optimum value in SEF, but that are not strictly Pareto-optimal. The green triangles denote the solutions of the sequential design problem.

Minimising the PEF and maximising SHE performance The sixth formulation of the design problem is a BOO problem in which the PEF is minimised while the value of I_T is maximised. Solutions of this problem are shown in Figure 3.4 and details of each solution are provided in Table 3.9. It can be seen that the points with $I_T = 2$ and $I_T = 6$ each corresponds to two distinct Pareto-optimal designs. Furthermore, several optimal solutions achieve approximately the same low PEF, yet a small impact of the SHE constraint is seen on the solutions achieved, with a deterioration in the crystal yield. A solution for which I_T is greater than 6 can only be found with a significant increase in the PEF from 4.21 kg/kg to at least 6.57 kg/kg. As in Section 3.3.3, the solutions of the sequential problem are plotted as green triangles in Figure 3.4 and are all found to be dominated solutions.

Table 3.9: Pareto-optimal solutions of the BOO problem for minimising the PEF and maximising SHE performance. An additional solution, obtained with ϵ values of 5 or 6 and that achieves a relatively low PEF value with a deterioration of SHE performance is also reported and indicated with a *.

Solution	ϵ	Solvents	PEF (kg/kg)	I_T	T_2 (K)	T_5 (K)	$\frac{x_{s1,2}}{x_{s2,2}}$	$\frac{x_{s1,5}}{x_{s2,5}}$	$\sum_{q\in Q} M_{q,5}$ (kg)
1	1	s1: o-Xylene s2: Anisole	4.20	2	403	290	2.47	2.47	280
2	2	s1: Cyclohexanone s2: o-Xylene	4.20	2	403	290	2.57	2.57	280
3	3	s1: Anisole s2: 1,2-Propanediol	4.21	6	403	290	0.12	0.12	280
4	4	s1: 1,2-Propanediol s2: Cyclohexanone	4.21	6	403	290	7.00	7.00	280
5	7	s1: Isobutyl acetate s2: Isobutanol	6.57	7	371	290	0.40	0.40	464
6*	5 or 6	s1: Anisole s2: Cyclohexanone	4.23	6	403	290	0.11	0.11	280

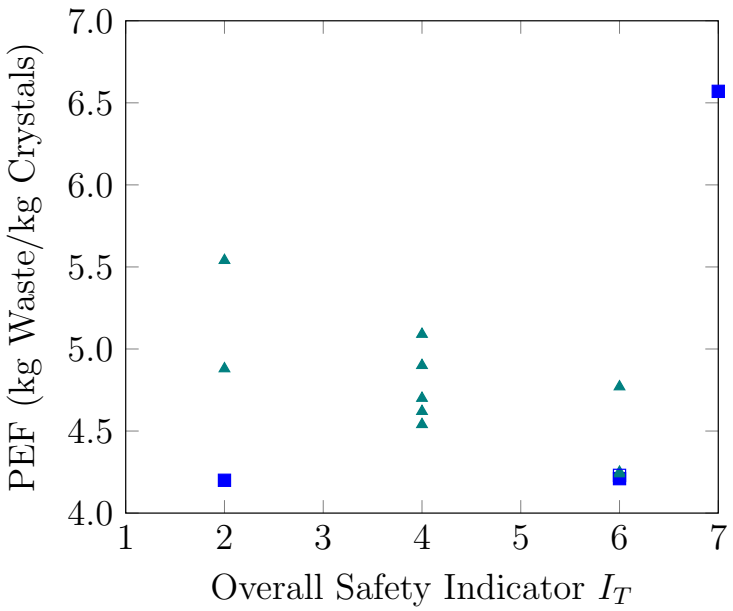

Figure 3.4: Pareto front of the BOO problem for minimising the PEF and maximising SHE performance shown with solid squares. There are 2 distinct solutions at $I_T = 2$ and 2 distinct solutions at $I_T = 6$. Additionally, the open square at $I_T = 6$ (Solution 6) denotes a point that achieves a low PEF value but is not strictly Pareto-optimal. The green triangles denote the solutions of the sequential design problem.

From the results of the previous two design formulations, it can be seen how the proposed solvent selection approach identifies SHE friendly solvent mixtures, such as Anisole/n-Butyl acetate, Acetic acid/Anisole and Isobutyl acetate/Isobutanol (Table 3.8), and Anisole/1,2-Propanediol and Anisole/Cyclohexanone (Table 3.9), all of which involve solvent components with good SHE scores (at least 6 on a scale from 1 to 10) according to the GSK solvent selection guideline [23].

Minimising the SEF and maximising Y_C

Increasing the crystal yield of pharmaceutical compounds is usually achieved at the expense of resource consumption, such as solvent use. To explore the competing relationship between solvent use and crystal yield, the seventh formulation of the design problem is a BOO problem that involves minimising the SEF while maximising Y_C. The problem is solved using the ϵ-constraint method, in which the SEF is optimised while Y_C is constrained by a given lower bound ϵ ranging between 98.6% and 99.4%. Solutions of this problem are shown in Figure 3.5, and the details of each solution are given in Table 3.10. It can be seen that a marginal increase in Y_C beyond 99.1% requires a significant increase in the SEF from 3.76 kg/kg to at least 4.54 kg/kg. In other words, more solvent consumption is required to satisfy the MA solubility conditions which favour crystal yield maximisation. Accordingly, the solution corresponding to (SEF,Y_C) = (3.76,99.1) would be a good compromise solution. All solutions correspond to cooling crystallisation; accordingly, T_5 is at its lower bound in all solutions and T_2 is at its upper bound in most solutions (except in Solution 1 where it is very close to the upper bound) to create a large temperature drop across crystallisation which would also contribute to crystal yield maximisation in the absence of anti-solvent effects. Furthermore, both Solutions 2 and 4 correspond to the same solvent identities and process conditions, i.e., Cyclohexanone/o-Xylene at T_2= 403 K and T_5 = 290 K, yet to different solvent mixture proportions, which impacts process performance. This highlights an important feature of the proposed CAM^{b}D approach: generating different solvent and process designs to highlight the trade-offs between competing KPIs, hence guiding rational design decisions during process development.

Table 3.10: Pareto-optimal solutions of the BOO problem for minimising the SEF and maximising Y_C. Additional solutions which achieve the lowest value of SEF with a deterioration of Y_C are also reported and indicated with a *.

Solution	ϵ (%)	Solvents	SEF (kg/kg)	Y_C (%)	T_2 (K)	T_5 (K)	$\frac{x_{s1,2}}{x_{s2,2}}$	$\frac{x_{s1,5}}{x_{s2,5}}$	$\sum_{q\in Q} M_{q,5}$ (kg)
1	99.00	s1: DMF s2: o-Xylene	3.50	99.00	401	290	0.14	0.14	280
2	99.10	s1: Cyclohexanone s2: o-Xylene	3.76	99.10	403	290	1.97	1.97	301
3	99.20	s1: Anisole s2: o-Xylene	4.54	99.20	403	290	0.22	0.22	363
4	99.30	s1: Cyclohexanone s2: o-Xylene	4.89	99.30	403	290	0.90	0.90	391
5	99.40	s1: Nitrobenzene s2: o-Xylene	5.65	99.40	403	290	0.15	0.15	452
6*	98.60	s1: 1,2-Propanediol s2: DMAC	3.50	98.60	403	290	4.82	4.82	280
7*	98.70	s1: Nitrobenzene s2: 1,2-Propanediol	3.50	98.70	403	290	0.13	0.13	280
8*	98.80	s1: o-Xylene s2: DMF	3.50	98.80	403	290	5.77	5.77	280
9*	98.90	s1: o-Xylene s2: DMAC	3.50	98.97	403	290	8.10	8.10	280

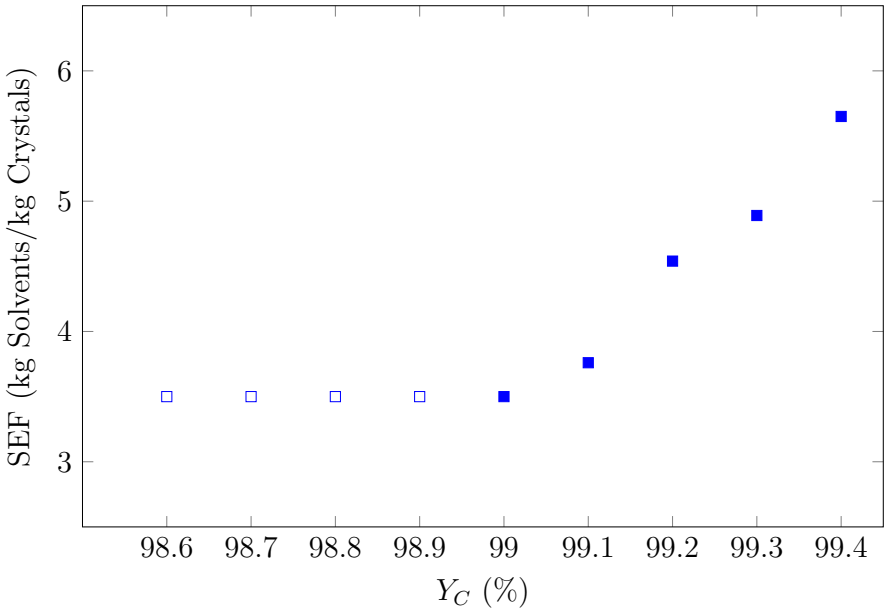

Figure 3.5: Pareto front of the BOO problem for minimising the SEF and maximising Y_C shown with solid squares. The open squares denote points that achieve the global optimum value in SEF, but that are not strictly Pareto-optimal.

Minimising Q_T and maximising Y_C

Another interesting relationship between competing KPIs is that of process energy consumption and crystal yield. To explore this relationship, the eighth and final formulation of the design problem is a BOO problem that involves minimising Q_T while maximising Y_C. In addition to the integrated design equations, this problem includes equations (3.42), (3.43), (3.49) and (3.50). In this design problem, the optimal solvents are selected from the list of 49 solvent candidates given in Appendix B.3. The problem is solved using the ϵ-constraint method, in which Q_T is optimised while Y_C is constrained by a given lower bound ϵ ranging between 92% and 99%. The Pareto front representing eight Pareto-optimal solutions of this problem is shown in Figure 3.6, and the details of each solution are given in Table 3.11. Q_T is reported in kJ per kg API. It can be seen that all solutions correspond to hybrid cooling and anti-solvent crystallisation; this is expected given the demanding process requirements of high Y_C. It can also be seen that a marginal increase in Y_C beyond 95% requires a significant increase in Q_T from 802 kJ/kg API to at least 998 kJ/kg API. Accordingly, the solution corresponding to (Q_T,Y_C) = (802,95) would be a good compromise solution.

Table 3.11: Pareto-optimal solutions of the BOO problem for minimising Q_T and maximising Y_C.

Solution	ϵ (%)	Solvents	Q_T (kJ/kg API)	Y_C (%)	T_2 (K)	T_5 (K)	$\frac{x_{s1,2}}{x_{s2,2}}$	$\frac{x_{s1,5}}{x_{s2,5}}$	$\sum_{q\in Q} M_{q,5}$ (kg)
1	92.00	s1: DMAC s2: Formic acid	722	92.57	358	306	4.53	0.44	280
2	93.00	s1: DMAC s2: Formic Acid	735	93.00	358	305	4.53	0.44	280
3	94.00	s1: DMAC s2: Formic Acid	766	94.00	358	302	4.53	0.44	280
4	95.00	s1: DMAC s2: Formic acid	802	95.00	358	298	4.53	0.44	280
5	96.00	s1: DMF s2: Water	998	96.00	362	306	9.00	0.58	280
6	97.00	s1: DMAC s2: o-Xylene	1188	97.00	390	290	9.00	0.35	280
7	98.00	s1: Nitrobenzene s2: Chlorobenzene	1350	98.00	394	292	1.75	1.38	280
8	99.00	s1: DMF s2: o-Xylene	1711	99.00	403	290	0.17	0.14	280

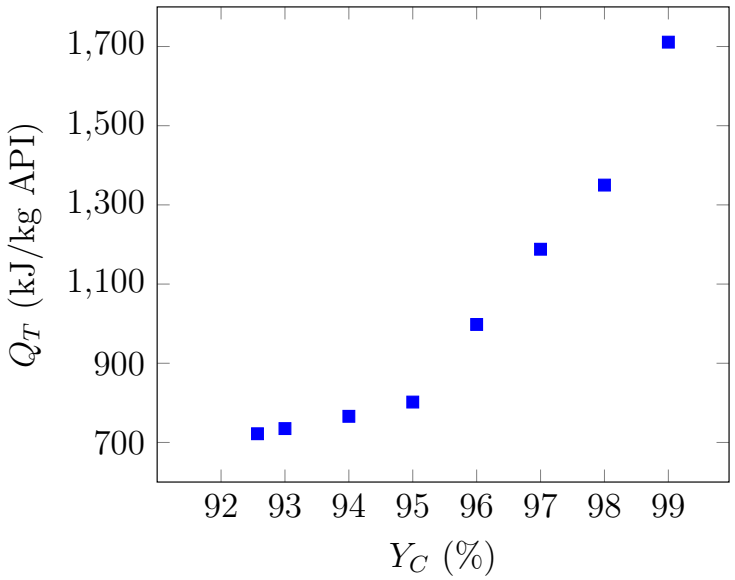

Figure 3.6: Pareto front of the BOO problem for minimising Q_T and maximising Y_C.

It is important to mention that the CPU time (in seconds) needed to generate the solutions

of each of the integrated design problems presented in this chapter is similar to that reported for the full space design formulation in Table 3.5. This shows that the proposed solvent selection approach can be used for a relatively quick thermodynamic assessment of a wide range of solvent candidates for integrated synthesis and crystallisation processes while considering different design objectives.

3.4 Conclusion

In this chapter, a systematic methodology for the identification of optimal solvent mixtures and process conditions for the integrated synthesis and crystallisation of pharmaceutical compounds was presented. KPIs that quantify resource efficiency and product quality were used to guide the design, and the model equations that describe the integrated process were formulated. The methodology was applied to identify optimal solvent mixtures and process conditions for the synthesis and crystallisation of mefenamic acid. The results demonstrate the importance of taking a holistic view and integrating different process steps to achieve more sustainable processes; for example, substantially better E-factors were obtained than when considering each step in turn.

The flexibility of the approach was demonstrated by using different solvent design spaces and different objective functions. Furthermore, BOO was used to highlight the trade-offs between the solvent or process E-factors and the health and environmental impacts of the chosen solvent molecules, those between the solvent E-factor (SEF) and crystal yield (Y_C) and those between process energy consumption (Q_T) and crystal yield (Y_C). The findings of this chapter demonstrate the value of using model-based solvent selection to explore a vast range of design options while capturing the interdependence between molecular-scale and process-level decisions. Additionally, the integrated nature of the proposed approach allows identifying solvents that could potentially (i.e., with experimental validation) be used across both synthesis and crystallisation, without the need for resource-intensive intermediate workups such as solvents swaps, hence reducing process energy consumption and environmental impact.

The next chapter focuses on extending the design formulation to account for subsequent purification steps deployed in pharmaceutical processing, namely filtration, washing and drying.

Chapter 4

Model-based Solvent Selection for Integrated Synthesis, Crystallisation and Isolation

4.1 Introduction

In this chapter, the solvent selection approach proposed in Chapter 3 is extended and applied to the integrated synthesis, crystallisation and isolation of pharmaceutical compounds. The proposed method simultaneously identifies the solvent/anti-solvent mixture, mixture composition and process conditions that optimise process-wide KPIs in the end-to-end process. Furthermore, the method entails comprehensive design specifications such as the miscibility of the selected solvents with the wash solvent, and the API solubility across synthesis, crystallisation and washing. Accordingly, the work presented in this chapter builds on the solvent selection and process design approach developed by Jonuzaj et al. [58] by considering synthesis as a preceding step to crystallisation and by incorporating a wider range of pharmaceutically relevant KPIs, such as product purity and process yield, in the CAM^{b}D formulation. Additionally, MOO CAM^{b}D formulations are developed to explore the trade-offs between competing KPIs. The design approach is illustrated by identifying optimal solvent mixtures and process conditions for end-to-end mefenamic acid production.

4.2 Problem Definition and Formulation

4.2.1 General design problem formulation

The extended CAM^{b}D problem can be described as follows: Given a synthetic step in the manufacturing of a pharmaceutical compound and a specified production rate, reaction conversion and selectivity, product purity, and a list of possible solvents, identify the solvent mixture, mixture composition and process conditions for the corresponding integrated synthesis, crystallisation and isolation (filtration, washing and drying) process that optimise one or more KPIs. In this formulation, KPIs that describe process-wide performance, rather than KPIs that are limited to a single process unit, are considered.

A generic end-to-end drug substance manufacturing process is depicted in Figure 4.1. Similar to the reaction-crystallisation process considered in Chapter 3, this process configuration can be used to represent a batch or continuous process. After crystallisation, the slurry that exits the crystalliser is sent to a filtration unit where solvents and impurities are removed in the filtrate stream (Stream 7), leaving behind a wet filer cake consisting of a liquid phase of solvents/dissolved impurities trapped inside the pore volume of the cake (Stream 8) and a solid phase of crystallised API (Stream 9). The wet filter cake is then sent to a washing unit where it is washed with a stream of pure wash solvent (Stream 10), denoted by sw, to remove any excess impurities and increase product purity. Stream 11 represents the filtrate stream that exits the washing unit, while Streams 12 and 13 represent the liquid and solid phases of the filter cake obtained after washing, respectively. Finally, the wet filter cake from the washing unit is sent to the dryer where the solvents are evaporated and a dry product is produced.

Due to the integrated nature of the explored pharmaceutical process, comprehensive or interlinked design decisions are embedded in the formulation to design feasible processes with enhanced process efficiency. These include specifications related to API solubility across the synthesis, crystallisation and isolation steps, as well as the miscibility of the reaction/crystallisation solvent(s) $s1$ and $s2$ present in the pore volume of the cake after washing with solvent sw.

A number of sets are defined to derive a generic formulation of the design problem, most of which are the same as those defined in Section 3.2.1:

- the set C comprises all the components that appear in any of the liquid streams;
- the set C^K comprises all the components that may appear in solid form in the process, i.e., whose melting point is within the range of process operating temperatures. In principle, this may introduce new components if co-crystals or solvates can form. However, here it is assumed that only pure crystal forms can be formed so that $C^K \subset C$;
- the set J comprises all solid components that may appear as impurities in the final dried product, i.e., all components in set C^K except the desired product;
- the set Q comprises solvents $s1$ and $s2$ that are selected for the integrated process;
- the set V comprises all the solvents in the process, including the set $Q = \{s1, s2\}$ and the wash solvent sw; therefore, $Q \subset V$;
- the set S comprises the solvent candidates from which solvents $s1$ and $s2$ are chosen;
- the set of reaction indices R;
- the set T defines the process streams introduced earlier and shown in Figure 4.1.

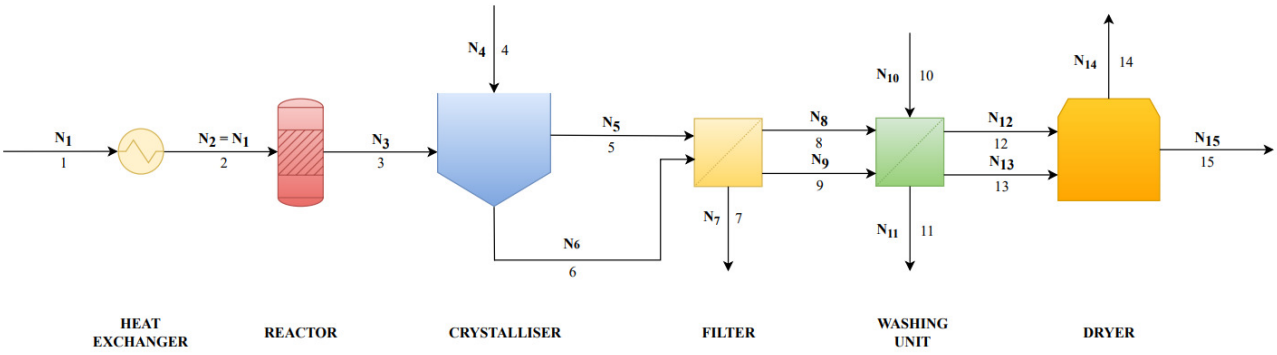

Figure 4.1: The conceptual flowsheet used for the solvent selection problem of the integrated synthesis, crystallisation and isolation process. N denotes the vector of component mole numbers (batch process) or component molar flowrates (continuous process).

The overall mathematical structure of the multi-objective optimisation version of the extended CAM^{b}D formulation is given by:

$$\begin{aligned}
\min_{\mathbf{x},\mathbf{y}} \quad & f_1(\mathbf{x}), ..., f_P(\mathbf{x}) \\
\text{s.t.} \quad & \mathbf{h}_0(\mathbf{x}) = \mathbf{0} \\
& \mathbf{g}_0(\mathbf{x}) \leq \mathbf{0} \\
& \mathbf{g}_{i,s}(\mathbf{x}) \leq \mathbf{M}_{i,s}(1 - y_{i,s}\mathbf{e}), i \in \{s1, s2\}; s \in S \\
& \mathbf{A}\mathbf{y} \leq \mathbf{a} \\
& \mathbf{x} \in [\mathbf{x}^L, \mathbf{x}^U] \subset \mathbb{R}^n \\
& \mathbf{y} \in \{0, 1\}^u
\end{aligned} \tag{4.1}$$

The main differences between the above problem and problem (3.1) is that now the objective functions $f_p(\mathbf{x})$, $p = 1, \ldots, P$ represent KPIs that describe process-wide performance, such as the process-wide SEF or process yield, the vector of constraints $\mathbf{h}_0(\mathbf{x})$ describes the process models of synthesis, crystallisation and isolation, and the vector of constraints $\mathbf{g}_0(\mathbf{x})$ and $\mathbf{g}_{i,s}(\mathbf{x})$ describes comprehensive design and operating specifications such as the miscibility of $s1$ and $s2$ with sw as well as API solubility across synthesis, crystallisation and isolation.

4.2.2 Process model and design constraints

Solvent assignment and properties

Solvent assignment constraints The assignment constraints presented in equations (3.3), (3.4) and (3.5) are also included in the end-to-end CAM^{b}D formulation. It is again possible to choose up to two solvents for synthesis/crystallisation, with the possibility of adding an anti-solvent ($s2$) in the crystalliser.

Solvent properties The properties of $s1$ and $s2$ are obtained from the properties of the species in set S. The densities of $s1$ and $s2$ are assigned based on the density ρ_s of each solvent candidate s in set S:

$$\rho_q = \sum_{s \in S} \rho_s y_{q,s}, \quad q \in Q. \tag{4.2}$$

The molecular weights, boiling points and melting points of $s1$ and $s2$ are obtained using equations (3.6), (3.7) and (3.8), respectively.

Masses, volumes, mass fractions and mole fractions

The mass, volume, mass fraction and mole fraction of component $i \in C$ in stream $t \in T$ are given as follows:

$$M_{i,t} = N_{i,t} MW_i, \quad i \in C; t \in T, \tag{4.3}$$

$$M_{i,t} = \rho_t V_{i,t}, \quad i \in C; t \in T, \tag{4.4}$$

$$w_{i,t} = \frac{M_{i,t}}{\sum_{i' \in C} M_{i',t}}, \quad i \in C; t \in T, \tag{4.5}$$

$$x_{i,t} = \frac{N_{i,t}}{\sum_{i' \in C} N_{i',t}}, \quad i \in C; t \in T, \tag{4.6}$$

where ρ_t denotes the mixture density of stream $t \in T$ which is calculated as a function of the pure component densities ρ_i of the mixture components using the following expression, assuming ideal mixture behaviour:

$$\frac{1}{\rho_t} = \sum_{i \in C} \frac{w_{i,t}}{\rho_i}. \tag{4.7}$$

Equation (3.11) is included in the end-to-end design formulation to ensure that $s2$ is not present in any of the liquid-phase process streams if it is not selected, i.e., if pure solvent $s1$ is selected for synthesis and crystallisation.

The process model and design constraints of all process units preceding the filter, i.e., the heat exchanger, reactor and crystalliser, are identical to those presented in Chapter 3. Therefore, the following subsections will only include models of the filter, washing unit and dryer.

Filter balances

The following material balances apply across the filter:

$$N_{i,5} + N_{i,6} - N_{i,7} - N_{i,8} - N_{i,9} = 0, \quad i \in C, \tag{4.8}$$

$$N_{i,9} = 0, \quad i \in C \setminus \{\mathbf{D}\}, \tag{4.9}$$

$$N_{D,6} = N_{D,9}. \tag{4.10}$$

The volume of the filter cake can be estimated as the sum of the volume occupied by the solid and that occupied by the liquid:

$$V_f^c = \frac{M_{D,9}}{\rho_D^S} + \sum_{i \in C} V_{i,8}, \tag{4.11}$$

where ρ_D^S is the density of solid product **D**.

The volume of solution in the cake constitutes the void volume of the cake. This can be expressed as follows:

$$\sum_{i \in C} V_{i,8} = p_f V_f^c, \tag{4.12}$$

where p_f is a user-defined value of the cake porosity after filtration.

Since the filter acts as a stream splitter with no phase change, the mole fractions of each component in Streams 5, 7 and 8 are equal:

$$x_{i,5} = x_{i,7} = x_{i,8}, \quad i \in C. \tag{4.13}$$

Washing unit balances and equilibrium constraints

Washing is usually achieved using the crystallisation solvent, an anti-solvent or a mixture of both, in several washing stages [196]. In this work, a single-stage washing unit in which the optimal volume of wash solvent, $V_{sw,10}$, is identified by the CAM^{b}D formulation to meet the product purity requirements is considered.

The wash solvent sw is an element of set C of components and is non-existent in all streams

that precede the washing unit:

$$N_{sw,t} = 0, \quad t \in \{1, 2, 3, 4, 5, 6, 7, 8, 9\}. \tag{4.14}$$

The following material balances apply across the washing unit:

$$N_{i,8} + N_{i,9} + N_{i,10} - N_{i,11} - N_{i,12} - N_{i,13} = 0, \quad i \in C, \tag{4.15}$$

$$N_{i,10} = 0, \quad i \in C \setminus \{sw\}, \tag{4.16}$$

$$N_{i,13} = 0, \quad i \in C \setminus \{\mathbf{D}\}. \tag{4.17}$$

The volume of the cake after washing can be estimated as the sum of the volume occupied by the crystalline material and of the volume occupied by the solvent:

$$V_w^c = \frac{M_{D,13}}{\rho_D^S} + \sum_{i \in C} V_{i,12}. \tag{4.18}$$

Similar to filtration, the volume of solution in the cake after washing constitutes the void volume of the cake. This can be expressed as follows:

$$\sum_{i \in C} V_{i,12} = p_w V_w^c, \tag{4.19}$$

where p_w is a user-defined value of the cake porosity after washing.

The volume of wash solvent used is expressed as a factor of the cake void volume:

$$V_{sw,10} = F_w p_f V_f^c, \tag{4.20}$$

where F_w typically ranges between 1 and 5. A factor F_w lower than 1 is not considered in the formulation as previous studies have shown that more than one cake void volume of wash solvent is practically used to purify the API cake [196].

Since the washing filter acts as a stream splitter with no phase change, the mole fractions of

each component in Streams 11 and 12 are equal:

$$x_{i,11} = x_{i,12}, \quad i \in C. \tag{4.21}$$

Depending on the wash solvent used, the washing step may lead to the partial dissolution of the crystallised target product, **D**, in case the crystallisation solvent is used for washing, or the undesired crystallisation of dissolved impurities, in case an anti-solvent is used. Accordingly, it is important to include equilibrium constraints that ensure that no dissolution of **D** can occur and that all impurities remain under-saturated.

The following constraint enforces that **D** remains at solid-liquid equilibrium in Stream 12, i.e., after washing:

$$\ln x_{D,12} + \ln \gamma_{D12}(\mathbf{x}_{12}, T_{12}) = \frac{\Delta H_{m,D}}{R_g}\left[\frac{1}{T_{m,D}} - \frac{1}{T_{12}}\right], \tag{4.22}$$

where $\gamma_{D,12}(\mathbf{x}_{12}, T_{12})$ is obtained using UNIFAC.

A hypothetical mixture is considered for each solute i other than **D**, with mole fraction vector $\hat{\mathbf{x}}_{12}^{(i)}$, $i \in C^K \setminus \{\mathbf{D}\}$, in which the proportions of all components are the same as in Stream 12, but component i is at solid-liquid equilibrium. As illustrated previously in Chapter 3, this is given by:

$$\ln \hat{x}_{i,12}^{(i)} + \ln \hat{\gamma}_{i,12}(\hat{\mathbf{x}}_{12}^{(i)}, T_{12}) = \frac{\Delta H_{m,i}}{R_g}\left[\frac{1}{T_{m,i}} - \frac{1}{T_{12}}\right], \quad i \in C^K \setminus \{\mathbf{D}\}. \tag{4.23}$$

where $\hat{\gamma}_{i,12}(\hat{\mathbf{x}}_t^{(i)}, T_{12})$ is calculated via UNIFAC and $\hat{\mathbf{x}}_{12}^{(i)}$ is such that:

$$\frac{\hat{x}_{i',12}^{(i)}}{\hat{x}_{s1,12}^{(i)}} = \frac{x_{i',12}}{x_{s1,12}}, \quad i' \in C \setminus \{s1, i\}, i \in C^K \setminus \{\mathbf{D}\}, \tag{4.24}$$

$$\sum_{i' \in C} \hat{x}_{i',12}^{(i)} = 1, \quad i \in C^K \setminus \{\mathbf{D}\}. \tag{4.25}$$

Impurity under-saturation after washing is then achieved by enforcing:

$$x_{i,12} \leq \hat{x}_{i,12}^{(i)} - \epsilon_s, \quad i \in C^K \setminus \{\mathbf{D}\}, \tag{4.26}$$

where ϵ_s is a small, positive number.

Although the current model describes a simple washing step, alternative washing models can be explored. For example, Shahid et al. [196] developed a methodology to evaluate the propensity of crystallisation-wash solvent combinations in inducing API and/or impurity precipitation during isolation. An important qualitative finding of their work is that the appropriate washing strategy depends on the design objective. If the objective is maximising impurity removal, then using the crystallisation solvent as the wash solvent would be an option; however, there is a risk of API yield loss and consequent agglomeration in drying. On the other hand, if the objective is maximising impurity removal and minimising effects on crystal product, then a two-(or more) part washing strategy should be devised. In the paper, paracetamol is crystallised from Ethanol, first washed by an equivolume mixture of Ethanol and n-Heptane as this increases the solubility limit and hence reduces the risk of API and/or impurity precipitation, and second washed by pure n-Heptane (anti-solvent) to remove the equivolume wash solution from the API crystal cake. The use of n-Heptane in the second wash facilitates drying as it is relatively easily evaporated and the solubility of paracetamol in n-Heptane is less than that in the equivolume mixture, which minimises agglomeration once the solvents are evaporated in the dryer.

Dryer balances

The set V of process solvents is an extension of set Q that additionally includes the wash solvent sw. In the drying model, it is assumed that all the solvents evaporate and all dissolved solutes are left behind as solids. Accordingly, the following material balances apply across the dryer:

$$N_{i,12} + N_{i,13} - N_{i,14} - N_{i,15} = 0, \quad i \in C, \tag{4.27}$$

$$N_{i,14} = 0, \quad i \in C^S, \tag{4.28}$$

$$N_{i,15} = 0, \quad i \in V. \tag{4.29}$$

4.2.3 Process design constraints

Constraints on operating temperatures

The extended CAM^{b}D formulation includes equations (3.27) and (3.28) as constraints on operating temperatures.

Solvent miscibility

Equations (3.29) and (3.30) remain in the extended CAM^{b}D formulation to ensure that the selected solvents form a homogeneous liquid phase in the process streams. Miscibility constraints in Stream 12 are additionally included in the formulation to ensure that each of the identified solvents $s1$ and $s2$ (if selected) are miscible with sw, hence enabling an effective washing procedure.

The mole fraction of solvent $s1$ in a binary mixture of $s1$ and sw, $\tilde{x}_{s1,12}$, and the mole fraction of solvent $s2$ in a binary mixture of $s2$ and sw, $\tilde{x}_{s2,12}$, are calculated using the following expressions:

$$\tilde{x}_{s1,12} = \frac{x_{s1,12}}{x_{s1,12} + x_{sw,12}}, \tag{4.30}$$

$$-\tilde{M}_x(1 - \sum_{s \in S} y_{s2,s}) \leq \tilde{x}_{s2,12} - \frac{x_{s2,12}}{x_{s2,12} + x_{sw,12}} \leq \tilde{M}_x(1 - \sum_{s \in S} y_{s2,s}), \tag{4.31}$$

where a big-M formulation has been used with constant $\tilde{M}_x$. The miscibility constraint for the binary solvent pairs ($s1$, sw) and ($s2$, sw) is given by:

$$\frac{\partial \ln \tilde{\gamma}_{q,12}(\tilde{\mathbf{x}}^q_{12}, T_{12})}{\partial \tilde{x}_{q,12}} + \frac{1}{\tilde{x}_{q,12}} \geq -M_\gamma(1 - \sum_{s \in S} y_{s2,s}), \quad q \in Q, \tag{4.32}$$

where $\frac{\partial \ln \tilde{\gamma}_{q,12}(\tilde{\mathbf{x}}^q_{12}, T_{12})}{\partial \tilde{x}_{q,12}}$ is calculated as described above for $\frac{\partial \ln \tilde{\gamma}_{s1,t}(\tilde{\mathbf{x}}_t, T_t)}{\partial \tilde{x}_{s1,t}}$, $\tilde{\mathbf{x}}^q_{12}$ is a vector of mole fractions $\tilde{x}_{q,12}$ and $\tilde{x}_{sw,12}$, and M_γ is a sufficiently large constant.

4.2.4 KPIs

The following KPIs are considered in the extended design formulation:

- Crystal Yield or Y_C (%) as defined in Section 3.2.4;
- Process Yield or Y_P (%), defined as the number of moles of product **D** per moles of limiting reagent A fed to the process, multiplied by a stoichiometric factor SF:

$$Y_P = \frac{N_{D,15}}{N_{A,1}} \times SF \times 100\%; \tag{4.33}$$

- Solvent E-factor or SEF (kg Solvents/kg API) is defined as the mass of solvent consumed throughout the end-to-end process per mass of dried product **D**:

$$\text{SEF} = \frac{\sum_{i\in Q} M_{i,7} + \sum_{i\in Q} M_{i,11} + \sum_{i\in Q} M_{i,14} + M_{sw,11} + M_{sw,14}}{M_{D,15}}; \tag{4.34}$$

- Process E-factor or PEF (kg Waste/kg API) is defined as the mass of waste produced in the end-to-end process per mass of dried product **D**:

$$\text{PEF} = \frac{\sum_{i\in C} M_{i,7} + \sum_{i\in C} M_{i,11} + \sum_{i\in C} M_{i,14} + \sum_{i\in J} M_{i,15}}{M_{D,15}}; \tag{4.35}$$

- Product Purity or PP (%) is defined as the mass of product **D** per total mass of dried product:

$$\text{PP} = \frac{M_{D,15}}{\sum_{i\in C^K} M_{i,15}} \times 100\%; \tag{4.36}$$

- Process safety as defined in Section 3.2.4;

- Process energy consumption Q_T (kJ/kg API) as defined in Section 3.2.4 with the inclusion of the drying energy requirement Q_D, defined as the amount of energy supplied to the dryer that is required to evaporate the process solvents. An energy balance around the dryer is performed in order to obtain an expression for Q_D:

$$Q_D + H_{12} + H_{13} = H_{14} + H_{15}, \tag{4.37}$$

where H_{12} is the enthalpy of liquid Stream 12, H_{13} is the enthalpy of solid Stream 13, H_{14} is the enthalpy of vapour Stream 14, and H_{15} is the enthalpy of the dried product/Stream

15. Assuming ideal mixing, the enthalpies are computed as follows:

$$H_{12} = \sum_{i \in C} N_{i,12} \int_{T_0}^{T_{12}} C_{p_i}^{(L)} dT, \tag{4.38}$$

$$H_{13} = \sum_{i \in C^K} N_{i,13} \int_{T_0}^{T_{12}} C_{p_i}^{(S)} dT, \tag{4.39}$$

$$H_{14} = \sum_{i \in V} N_{i,14} \Bigg(\int_{T_0}^{T_{dew}} C_{p_i}^{(L)} dT + \Delta H_{V_i} \Bigg), \tag{4.40}$$

$$H_{15} = \sum_{i \in C^K} N_{i,15} \int_{T_0}^{T_{dew}} C_{p_i}^{(S)} dT, \tag{4.41}$$

where T_{12} is the temperature of Stream 12 and that of the washing unit, T_0 is a reference temperature, T_{dew} is the dew point of the solvent mixture being evaporated at the operating pressure of the dryer, P_{dry}, and ΔH_{V_i} is the heat of vaporisation of solvent species $i \in V$ at its boiling point, $T_{b,i}$.

Setting $T_0 = T_{12}$ and combining equations (4.37)–(4.41) gives the following expression for Q_D:

$$Q_D = \sum_{i \in V} N_{i,14} \Bigg(\int_{T_{12}}^{T_{dew}} C_{p_i}^{(L)} dT + \Delta H_{V_i} \Bigg) + \sum_{i \in C^K} N_{i,15} \int_{T_{12}}^{T_{dew}} C_{p_i}^{(S)} dT. \tag{4.42}$$

In order to calculate the dew point, T_{dew}, of the solvent mixture, the following assumptions are taken:

- The liquid and vapour phases are ideal, i.e., vapour-phase fugacity coefficients and liquid-phase activity coefficients are assumed to be equal to 1;
- ΔH_{V_i} is constant;
- The solid solutes (API and impurities) have very low volatility and fully crystallise upon solvent evaporation, i.e., $N_{i,14} = 0, \;\; i \in C^S$.

Accordingly, the following equations are used to estimate T_{dew}:

$$K_i(T) = \frac{P_{sat,i}(T)}{P_{dry}}, \quad i \in V, \tag{4.43}$$

$$\sum_{i \in V} \frac{x_{i,12}}{K_i(T_{dew})} = 1, \tag{4.44}$$

where K_i is the vapour-liquid equilibrium value of solvent $i \in V$ and $P_{sat,i}$ is the saturation pressure of solvent $i \in V$.

The Clausius-Clapeyron equation is used to estimate $P_{sat,i}$ in Pascal [197]:

$$\ln P_{sat,i}(T) = \frac{\Delta H_{V_i}}{R_g}\left(\frac{1}{T_{b,i}} - \frac{1}{T}\right) + 11.5261, \quad i \in V. \tag{4.45}$$

In order to lift the assumption of constant ΔH_{V_i} in equation (4.45), the semi-empirical Antoine equation [198] can be used to estimate $P_{sat,i}$. However, the Antoine coefficients of some solvents may not cover a sufficiently large temperature range that is suitable for the application of interest. Furthermore, some solvents may have different sets of Antoine coefficients based on the experimental temperature ranges on which they were regressed. Accordingly, the Clausius-Clapeyron equation can be more easily embedded in the mathematical formulation of CAMD problems, as has been done in previous works [18, 105].

Finally, the energy KPI can be expressed as:

$$Q_T = Q_H + |Q_R| + Q_C^- + Q_D, \tag{4.46}$$

where, similar to Section 3.2.4, the absolute value of Q_R and the minus sign in Q_C^- ensure consistent signs of the energy duties, i.e., Q_H, Q_R, Q_C^- and $Q_D \geq 0$.

This KPI can be minimised in a single or multi-objective optimisation problem. The absolute value term in equation (4.46) can be removed by reformulating the energy KPI

objective function as follows:

$$\min \qquad Q_H + Q'_R + Q_C^- + Q_D, \tag{4.47}$$

and adding the following two constraints to the full formulation:

$$Q_R \leq Q'_R, \tag{4.48}$$

$$-Q_R \leq Q'_R, \tag{4.49}$$

where Q'_R is a variable used to incorporate the absolute value of Q_R in the expression of Q_T.

4.3 Case Study: The Synthesis, Crystallisation and Isolation of Mefenamic Acid

After exploring the integrated synthesis and crystallisation of mefenamic acid (MA) in Chapter 3, in this chapter, the integrated synthesis, crystallisation and isolation of MA from 2,3-dimethylaniline (DMA) and 2-chlorobenzoic acid (CBA) as starting material is used as a case study. The CAM^{b}D formulation identifies solvent mixtures of at most 2 components that optimise selected KPIs, while focusing on the thermodynamic aspects of the process and neglecting solvent effects on reaction and crystallisation kinetics, as well as the effect of inorganic species on mixture thermodynamics.

Eight different CAM^{b}D problems are formulated: in the first formulation the SEF is minimised, in the second, third and fourth formulations Y_P is maximised while considering different design formulations (restricted and flexible), in the fifth formulation the SEF is minimised while PP is maximised, in the sixth formulation the PEF is minimised while SHE performance is maximised, in the seventh formulation Q_T is minimised while Y_P is maximised, and in the eighth formulation the SEF, Y_P and PP are simultaneously optimised.

All formulations include a set of process model equations describing the material balances across the process units, as well as equilibrium constraints describing species solubility, i.e.,

under-saturation of CBA during synthesis, crystallisation and washing (DMA exists in the liquid phase under the process operating temperatures), under-saturation of MA during synthesis and that MA is at solid-liquid equilibrium during crystallisation and washing. The formulations also include a set of design specifications such as process temperature constraints to ensure solvents are in the liquid phase during operation, miscibility constraints to reduce the likelihood of a second liquid phase appearing and constraints on the target values/ranges of the selected KPIs. In this case study, n-Heptane is used as the wash solvent due to the limited solubility of MA in this solvent [58]. The main model equations and design specifications are given in Table 4.1, whereas the model parameters are provided in Table 4.2. It can be seen that the final crystallisation temperature range has been narrowed down from the range used for the integrated synthesis and crystallisation case study (Section 3.3) to 290 K $\leq T_5 \leq$ 306 K so that T_5 is not too far from the washing temperature $T_{12} = 298.15$K, hence reducing the possibility of undesired solute precipitation. The solvent lists from which $s1$ and $s2$ are selected are given in Appendix B – depending on the design formulation and the availability of data on solvent thermodynamic properties, one of two lists, List I or List III, is considered. In the first six design problems, the optimal solvents are selected from the list of 48 solvent candidates (List I) given in Appendix B.1. In the seventh and eighth design problems, the optimal solvents are selected from the list of 49 solvent candidates (List III) given in Appendix B.3.

Table 4.2: Parameters of the CAM^{b}D formulation of the MA case study.

Parameter	Value	Units (where required)
T_o	10	K
M_T	600	–
$\tilde{M}_x$	10	–
M_γ	100	–
M_S	100	–
SF	1	mol/mol

In all energy balances, heat capacity and heat of vaporisation are assumed to be weak functions

Table 4.1: Main CAM^{b}D model equations and design specifications of the MA case study.

Constraint	Equation	KPI Range (where applicable)	Units (where required)
Solvent E-factor	$\text{SEF} = \frac{\sum_{i \in Q} M_{i,5} + M_{Heptane,10}}{M_{MA,15}}$	3.5-20	kg Solvents/kg API
Process E-factor	$\text{PEF} = \frac{\sum_{i \in C} M_{i,5} + M_{Heptane,10}}{M_{MA,15}}$	3.5-20	kg Waste/kg API
Crystal Yield	$Y_C = \frac{N_{MA,6}}{N_{MA,3}} \times 100\%$	90-100	%
Process Yield	$Y_P = \frac{N_{MA,15}}{N_{CBA,1}} \times 100\%$	75-95	%
Product Purity	$\text{PP} = \frac{M_{MA,15}}{\sum_{i \in C^K} M_{i,15}} \times 100\%$	95-100	%
Feed temperature	$T_1 = 298.15$	–	K
Isothermal reactor operation	$T_2 = T_3$	–	K
2:1 DMA:CBA Feed Ratio	$N_{DMA,2} = 2N_{CBA,2}$	–	mol
No MA in process feed	$N_{MA,1} = 0$	–	mol
Reaction conversion	$x^c = 90$	–	%
Throughput	$M_{MA,15}$=80	–	kg
Reaction temperature range	$358 \leq T_2 \leq 403$	–	K
Final crystallisation temperature range	$290 \leq T_5 \leq 306$	–	K
Dryer pressure	$P_{dry} = 1$	–	bar
Cake porosity (Filtration)	$p_f = 0.4$	–	–
Cake porosity (Washing)	$p_w = 0.4$	–	–

of temperature, i.e., standard constant-pressure heat capacities (of the chemical compounds) and constant heats of vaporisation (of the solvent candidates) are used. Standard constant-pressure heat capacities and heats of vaporisation of solvent candidates in set S are obtained from the NIST Chemistry WebBook (`https://webbook.nist.gov/chemistry`) [192], whereas standard constant-pressure heat capacities of reactants and products are calculated using the GC method of Kolská et al. [193]. Furthermore, heat capacity is assumed to be phase-independent, i.e., $C_{p_i}^{(L)} \approx C_{p_i}^{(S)}$ for $i \in C$. In addition, pure component densities, ρ_i for $i \in C^K$, are obtained from the Chemical Book (`https://www.chemicalbook.com/ProductIndex_EN.aspx`)[191]. All properties are reported in Appendix C.

The optimisation problems are solved in GAMS version 32.2.0 [169] using the SBB solver [163]. Integer cuts are included in the MINLP formulation to generate multiple optimal solutions for each design problem that are ranked with respect to the objective function value. For each solution of a single-objective design problem, the values of SEF, Y_P and PP are reported. For each solution of a bi-objective design problem, the values of the objective functions are reported. For each solution of all single and bi-objective design problems, the following are reported: the identities of solvent components $s1$ and $s2$, the reaction temperature T_2, the final crystallisation temperature T_5, the composition of the solvent in the reactor, defined as the ratio of the mole fraction of solvent 1 to that of solvent 2 in the reactor outlet (Stream 2), the composition of

the solvent in the crystalliser, defined as the ratio of the mole fraction of solvent 1 to that of solvent 2 in the crystalliser outlet (Stream 5), and the total solvent mass used in the process, i.e., $\sum_{i \in Q} M_{i,5} + M_{sw,10}$. For each solution of the three-objective design problem, the solvent identities and the values of the objective functions are reported.

4.3.1 Minimising the SEF

The first CAM^{b}D formulation involves minimising the SEF. Ten optimal solutions, ranked with respect to the SEF, are listed in Table 4.3. It can be seen that solutions 1, 6, 7, 8, 9 and 10 correspond to hybrid cooling and anti-solvent crystallisation as is evident in the temperature decrease from T_2 to T_5 and the change in the ratio of solvent mole fractions across crystallisation, whereas all other solutions correspond to cooling crystallisation. Furthermore, most solutions achieve approximately the same SEF, with values ranging between 3.81 and 3.83 kg/kg. This narrow range of SEF values does not offer a clear distinction between the reported solutions. Accordingly, it would be useful to consider alternative objective functions, such as the process yield Y_P, to generate a ranked list of distinct solutions from which better insights on optimal solvent selection and process design can be derived.

Comparing the results shown in Table 4.3 with those reported for the integrated synthesis and crystallisation formulation in Section 3.3.1 (same objective function and design space), it can be seen that several solvent mixtures are identified as optimal in both CAM^{b}D formulations, such as Anisole/n-Butanol, Cyclohexanone/Anisole and Anisole/Acetic acid, although solvent proportions are different between the two formulations. Furthermore, due to the extension of the process presented in Chapter 3 to include isolation in this chapter, additional solvent consumption, e.g., the consumption of an additional process solvent, *sw*, in the washing stage, leads to an increase in the SEF from 3.5 kg/kg (Section 3.3.1) to the 3.81–3.83 kg/kg range in the end-to-end process.

The inclusion of interlinked design decisions in the CAM^{b}D formulation allows identifying solvents that not only form a miscible mixture, but that are also miscible with the wash solvent, n-Heptane. Had the solvent selection problem been solved sequentially, for example, by identifying optimal solvent identities for an integrated synthesis and crystallisation process

and then fixing those identities before optimising the end-to-end process, solvents that form an immiscible mixture with n-Heptane, such as pure Water, could have been identified, leading to ineffective washing with limited impurity rejection.

Table 4.3: Ten optimal solutions of the CAM^{b}D problem with the objective of minimising the SEF.

Solution	Solvents	SEF (kg/kg)	Y_P (%)	PP (%)	T_2 (K)	T_5 (K)	$\frac{x_{s1,2}}{x_{s2,2}}$	$\frac{x_{s1,5}}{x_{s2,5}}$	Total solvent use (kg)
1	s1: n-Butanol s2: Nitrobenzene	3.81	81.95	95.13	381	306	0.25	0.24	305
2	s1: 2-Pentanol s2: Nitrobenzene	3.82	81.93	95.22	380	306	0.32	0.32	306
3	s1: Anisole s2: n-Butanol	3.82	81.83	95.76	376	306	2.36	2.36	306
4	s1: Nitrobenzene s2: Anisole	3.82	83.53	95.31	382	306	1.01	1.01	306
5	s1: Anisole s2: Acetic acid	3.82	82.52	95.56	374	306	1.08	1.08	306
6	s1: 2-Pentanol s2: Anisole	3.82	82.58	95.81	380	306	0.45	0.44	306
7	s1: Cyclohexanone s2: Anisole	3.83	84.73	95.70	403	306	0.11	0.06	307
8	s1: n-Butyl acetate s2: Anisole	3.83	84.90	95.74	389	306	0.11	0.09	307
9	s1: Isobutyl acetate s2: Anisole	3.83	84.90	95.74	389	306	0.11	0.09	307
10	s1: Cyclohexanone s2: Nitrobenzene	3.83	85.58	95.10	403	306	0.11	0.07	307

4.3.2 Maximising Y_P

Crystallisation can have a significant influence on process performance, given the interrelationship between crystallisation performance, i.e., the crystal yield, and process-wide KPIs such as throughput, process yield and energy consumption. In order to explore the effect of crystallisation on process-wide performance in terms of process yield, the second, third and fourth design formulations involve maximising Y_P while considering different crystallisation conditions (fixed

or variable operating temperatures, with or without anti-solvent addition) and comparing the results obtained for each of them.

Restricted Design Problem 1: Fixed crystallisation temperatures with anti-solvent addition allowed

The second CAM^{b}D formulation involves maximising Y_P while fixing the initial crystallisation (or reaction) temperature T_2 to its upper bound, i.e., 403 K, and the final crystallisation temperature T_5 to its lower bound, i.e., 290 K, and allowing anti-solvent addition to occur. Therefore, this problem includes restrictions on the process operating temperatures and flexibility in the crystalliser mode of operation (cooling or hybrid cooling and anti-solvent crystallisation are both possible). Temperatures are fixed to their upper or lower bounds as a practitioner might tend to maximise the temperature drop throughout crystallisation to maximise crystal yield. Five optimal solutions, ranked with respect to Y_P, are listed in Table 4.4. Solution 1 corresponds to a purely cooling crystallisation process, whereas Solutions 2–5 involve cooling and anti-solvent effects. The downside of fixing the operating temperatures is that it restricts the design space of process solvents to those whose liquid range falls within 403 K and 290 K. Accordingly, potentially promising solvents that could result in equal or better objective function values than those obtained by solving this restricted formulation, as well as solvents that may exhibit high SHE performance, such as n-Butyl acetate [23], are no longer feasible.

Table 4.4: Five optimal solutions of the CAM^{b}D problem with the objective of maximising Y_P with fixed crystallisation temperatures and anti-solvent addition allowed.

Solution	Solvents	Y_P (%)	SEF (kg/kg)	PP (%)	T_2 (K)	T_5 (K)	$\frac{x_{s1,2}}{x_{s2,2}}$	$\frac{x_{s1,5}}{x_{s2,5}}$	Total solvent use (kg)
1	s1: Nitrobenzene s2: Cyclohexanone	88.12	3.85	95.82	403	290	0.24	0.24	308
2	s1: Anisole s2: Cyclohexanone	87.93	3.85	95.97	403	290	0.11	0.07	308
3	s1: Anisole s2: Nitrobenzene	87.64	5.28	98.33	403	290	0.11	0.07	422
4	s1: 1,2-Propanediol s2: Isobutyl acetate	87.40	5.28	98.61	403	290	0.11	0.06	422
5	s1: 1,2-Propanediol s2: Anisole	86.73	5.28	98.55	403	290	0.11	0.05	422

Restricted Design Problem 2: Variable crystallisation temperatures with no anti-solvent addition allowed

The third CAM^{b}D formulation involves maximising Y_P while allowing T_2 and T_5 to vary in a purely cooling crystallisation process, i.e., no anti-solvent addition is allowed. This condition can be enforced in the mathematical formulation by the inclusion of $N_{s2,4} = 0$ as a design constraint. Therefore, this problem includes restrictions on the crystalliser mode of operation (only cooling crystallisation allowed) and flexibility in setting the process operating temperatures. Ten optimal solutions, ranked with respect to Y_P, are listed in Table 4.5. It can be seen that solvent components with superior SHE performance, such as n-Butyl acetate, can now be identified due to the removal of the operating temperature constraint. Four out of the ten solutions result in $Y_P \geq 88\%$, none of which has an initial crystallisation temperature at the upper bound of 403 K. In comparison, only one of the solutions identified by solving Restricted Design Problem 1 results in $Y_P \geq 88\%$; additionally, the solvent combinations of Restricted Design Problem 1 are also identified by solving Restricted Design Problem 2, but with a smaller crystallisation temperature drop to achieve the same Y_P. This can be explained by the fact that the crystallisation yield limit could have already been reached due to the properties of the solvents and the solutes. More specifically, looking at the Nitrobenzene/Cyclohexanone design

(Solution 1) in Tables 4.4 and 4.5, it can be seen that both a T_2 of 403 K and 391 K result in the same Y_P, SEF and PP; accordingly, there is no point of having the maximum temperature drop for this particular design, which increases process costs without improving any of the KPIs. This emphasizes the importance of implementing a flexible approach to solvent selection and process design. Solving Restricted Design Problem 2 hence offers solutions that are superior to those obtained by solving Restricted Design Problem 1.

Table 4.5: Ten optimal solutions of the CAM^{b}D problem with the objective of maximising Y_P with variable crystallisation temperatures and no anti-solvent addition allowed.

Solution	Solvents	Y_P (%)	SEF (kg/kg)	PP (%)	T_2 (K)	T_5 (K)	$\frac{x_{s1,2}}{x_{s2,2}}$	$\frac{x_{s1,5}}{x_{s2,5}}$	Total solvent use (kg)
1	s1: Nitrobenzene s2: Cyclohexanone	88.12	3.85	95.82	391	290	0.24	0.24	308
2	s1: Nitrobenzene s2: n-Butyl acetate	88.10	5.58	98.56	389	290	2.10	2.10	446
3	s1: Nitrobenzene s2: Isobutyl acetate	88.10	5.56	98.56	389	290	2.15	2.15	445
4	s1: n-Butyl acetate s2: Isobutyl acetate	88.00	7.15	97.86	389	290	9.00	9.00	572
5	s1: Anisole s2: Cyclohexanone	87.91	5.28	98.62	392	290	0.11	0.11	422
6	s1: Anisole s2: n-Butyl acetate	87.85	7.07	98.95	389	290	0.11	0.11	566
7	s1: Anisole s2: Isobutyl acetate	87.83	5.28	98.63	389	290	1.18	1.18	422
8	s1: Cyclohexanone s2: n-Butyl acetate	87.80	7.38	99.03	389	290	0.11	0.11	590
9	s1: Isobutyl acetate s2: Cyclohexanone	87.80	7.34	99.03	362	290	9.00	9.00	587
10	s1: t-AmOH s2: n-Butanol	86.28	11.36	98.71	365	290	0.90	0.90	909

Flexible Design Problem: Variable crystallisation temperatures with anti-solvent addition allowed

The fourth CAM^{b}D formulation involves maximising Y_P while allowing T_2 and T_5 to vary and considering the possibility of anti-solvent addition. Ten optimal solutions, ranked with respect to Y_P, are listed in Table 4.6. Solutions 5 and 10 correspond to a hybrid cooling and anti-solvent crystallisation process as shown by both the decrease in temperature from T_2 to T_5 and the decrease in the ratio of solvent mole fractions between synthesis and crystallisation. All other solutions on the other hand correspond to cooling crystallisation as shown by the drop in temperature and the constant ratio of solvent mole fractions across the crystalliser. Since most solutions, including the top four, are cooling crystallisation designs, it can be said that anti-solvent use for this case study is not as crucial/impactful as cooling to induce product crystallisation and meet the design objectives. It can be seen that in all solutions, the SEF is greater than the lower bound of 3.5 kg/kg and the difference between T_2 and T_5 is wide: T_2 is close to its upper bound of 403 K in most solutions and T_5 is at its lower bound of 290 K in all ten solutions. This can be explained by the demanding process requirements that favour process yield maximisation. Several of the reported solutions are identical to those reported in Table 4.5, indicating that both the Restricted Design Problem 2 and Flexible Design Problem formulations yield similar results, and that the impact of anti-solvent use on the selected KPI target values for this case study is small. Specifically, designs including Nitrobenzene/n-Butyl acetate, Nitrobenzene/Isobutyl acetate, Anisole/n-Butyl acetate and Anisole/Isobutyl acetate are identical in terms of solvent choices and process conditions in the two formulations, whereas designs including Nitrobenzene/Cyclohexanone, n-Butyl acetate/Isobutyl acetate, Anisole/Cyclohexanone and Cyclohexanone/n-Butyl acetate share the same solvent identities but different process conditions (temperatures + mole fractions) which impacts the selected KPIs; for example, in Solution 1 in Table 4.5 (Nitrobenzene/Cyclohexanone), the design corresponds to an SEF of 3.85 kg/kg with minimal wash solvent use ($F_w = 1$) and a PP of 95.82 %, whereas in Solution 1 in Table 4.6, the design corresponds to an SEF of 5.29 kg/kg and a PP of 98.56% due to more wash solvent addition at the washing stage ($F_w = 5$). However, each of these two designs could have been identified by any of the two formulations since they

both correspond to cooling crystallisation. An interesting finding is that the product purity constraint is not active at the reported solutions in Table 4.6, and PP beyond the lower bound of 95% can be achieved when generating different solutions.

Table 4.6: Ten optimal solutions of the CAM^{b}D problem with the objective of maximising Y_P with variable crystallisation temperatures and anti-solvent addition allowed.

Solution	Solvents	Y_P (%)	SEF (kg/kg)	PP (%)	T_2 (K)	T_5 (K)	$\frac{x_{s1,2}}{x_{s2,2}}$	$\frac{x_{s1,5}}{x_{s2,5}}$	Total solvent use (kg)
1	s1: Nitrobenzene s2: Cyclohexanone	88.15	5.29	98.56	403	290	0.26	0.26	423
2	s1: Nitrobenzene s2: n-Butyl acetate	88.10	5.58	98.56	389	290	2.10	2.10	446
3	s1: Nitrobenzene s2: Isobutyl acetate	88.10	5.56	98.56	389	290	2.15	2.15	445
4	s1: n-Butyl acetate s2: Isobutyl acetate	88.00	7.96	98.98	389	290	9.00	9.00	637
5	s1: Anisole s2: Cyclohexanone	87.96	5.28	98.62	403	290	0.11	0.07	422
6	s1: Anisole s2: n-Butyl acetate	87.85	7.07	98.95	389	290	0.11	0.11	566
7	s1: Anisole s2: Isobutyl acetate	87.83	5.28	98.63	389	290	1.18	1.18	422
8	s1: Cyclohexanone s2: n-Butyl acetate	87.78	6.27	97.56	389	290	0.11	0.11	502
9	s1: n-Butanol s2: 2-Pentanol	87.62	7.57	99.22	381	290	0.11	0.11	606
10	s1: Acetic acid s2: 2,2,4-Trimethylpentane	83.94	9.24	99.30	362	290	6.50	4.13	739

It can be seen that constrained KPIs may vary from one solution to another, and it is therefore interesting to explore the competing relationships or trade-offs between those KPIs through multi-objective optimisation. In all subsequent sections, hybrid cooling and anti-solvent crystallisation (the Flexible Design Problem) is considered.

4.3.3 Multi-objective optimisation: Exploring trade-offs between competing KPIs

Minimising the SEF and maximising PP

To better explore the trade-offs between the KPIs, the fifth CAM^{b}D formulation is a BOO problem that involves minimising the SEF and maximising PP. The problem is solved using the ϵ-constraint method [176], in which the SEF is minimised while PP is constrained by a given lower bound ϵ ranging between 98.6% and 99.5%. The Pareto front representing the Pareto-optimal solutions of this problem is shown in Figure 4.2, and the details of each solution are given in Table 4.7. It can be seen that a marginal increase in purity beyond 98.8% requires a significant increase in the SEF. Since the Pareto curve before PP = 98.8% is relatively flat, indicating a small increase in solvent consumption with increasing purity, the solution corresponding to (SEF,PP) = (5.45,98.8) would be a good compromise solution. It can also be seen that some solutions correspond to the same solvent identities and process temperatures, such as Anisole/n-Butanol in Solutions 2 and 3 and Isobutyl acetate/n-Butanol in Solutions 5 and 6, yet to different solvent proportions in the process units; this clearly influences the SEF, PP and Y_P and highlights the capability of CAM^{b}D to generate different designs based on the KPI of interest.

Table 4.7: Pareto-optimal solutions of the BOO problem for minimising the SEF and maximising PP.

Solution	ϵ (%)	Solvents	SEF (kg/kg)	PP (%)	Y_P (%)	T_2 (K)	T_5 (K)	$\frac{x_{s1,2}}{x_{s2,2}}$	$\frac{x_{s1,5}}{x_{s2,5}}$	Total solvent use (kg)
1	98.60	s1: Nitrobenzene s2: 2-Pentanol	5.17	98.60	86.84	382	290	0.28	0.28	414
2	98.70	s1: Anisole s2: n-Butanol	5.21	98.70	87.11	381	290	0.16	0.16	417
3	98.80	s1: Anisole s2: n-Butanol	5.45	98.80	87.21	381	290	0.13	0.13	436
4	98.90	s1: Cyclohexanone s2: n-Butanol	5.80	98.90	87.23	381	290	0.12	0.12	464
5	99.00	s1: Isobutyl acetate s2: n-Butanol	6.22	99.00	87.17	381	290	0.11	0.10	498
6	99.10	s1: Isobutyl acetate s2: n-Butanol	6.79	99.10	86.98	381	290	0.11	0.09	543
7	99.20	s1: Cyclohexanone s2: n-Butanol	7.54	99.20	86.60	381	290	0.11	0.08	603
8	99.30	s1: n-Butanol s2: 2-Pentanol	8.33	99.30	87.42	381	290	0.11	0.10	666
9	99.40	s1: Isobutanol s2: n-Butanol	9.52	99.40	86.57	371	290	4.26	4.26	762
10	99.50	s1: n-Heptane s2: 1-Propanol	10.91	99.50	85.43	360	290	0.28	0.28	873

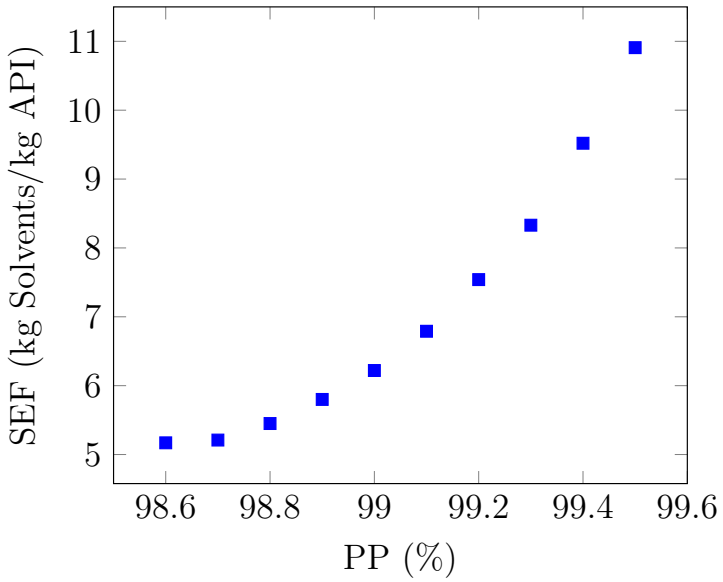

Figure 4.2: Pareto front of the BOO problem for minimising the SEF and maximising PP.

Minimising the PEF and maximising SHE Performance

The sixth CAM^{b}D formulation involves minimising the PEF and maximising SHE performance, which is quantified by the safety indicators introduced in Section 3.2.4. This BOO problem is solved using the ϵ-constraint method in which the PEF is minimised while I_T is constrained by a given lower bound ϵ, or equivalently, the safety indicators of each solvent q in the mixture, $I_{q,f}$, are constrained by the same value of ϵ, which represents the minimum acceptable value of I_T. The problem is solved for $1 \leq \epsilon \leq 7$ to generate the corresponding Pareto front as ϵ values beyond 7 lead to infeasible solutions. Solutions of this problem, as well as the SEF of each solution, are shown in Figure 4.3 and details of each solution are provided in Table 4.8. It can be seen that, similar to the results reported in Section 3.3.3, only two Pareto-optimal solutions are identified and other designs that achieve a relatively low PEF can be identified, although these might not necessarily be Pareto-optimal. Designs which have an I_T of at least 7 can only be achieved with a significant increase in the PEF. It is also clear that there is a correlation between the PEF and SEF of each reported solution, the gap being due to unreacted CBA and DMA, left in the process as impurities, as well as uncrystallised MA, and which are captured by the PEF but not the SEF.

Table 4.8: Pareto-optimal solutions of the BOO problem for minimising the PEF and maximising SHE performance of the end-to-end process. Dominated (non-Pareto optimal) solutions are reported with a *.

Solution	ϵ	Solvents	PEF (kg/kg)	I_T	SEF (kg/kg)	T_2 (K)	T_5 (K)	$\frac{x_{s1,2}}{x_{s2,2}}$	$\frac{x_{s1,5}}{x_{s2,5}}$	Total solvent use (kg)
1	5 or 6	s1: Anisole s2: Cyclohexanone	4.58	6	3.85	403	290	0.11	0.07	308
2	7	s1: Isobutanol s2: Isobutyl acetate	6.92	7	6.15	371	290	0.39	0.39	492
3*	1	s1: Nitrobenzene s2: n-Butyl acetate	4.58	1	3.85	389	290	0.19	0.19	308
4*	2	s1: Trifluoroethanol s2: n-Butylacetate	4.58	2	3.85	389	290	1.18	1.18	308
5*	3	s1: Anisole s2: 2-Pentanol	4.60	6	3.85	382	290	9.00	9.00	308
6*	4	s1: Acetic acid s2: n-Butanol	4.72	5	3.96	381	299	1.26	1.26	317

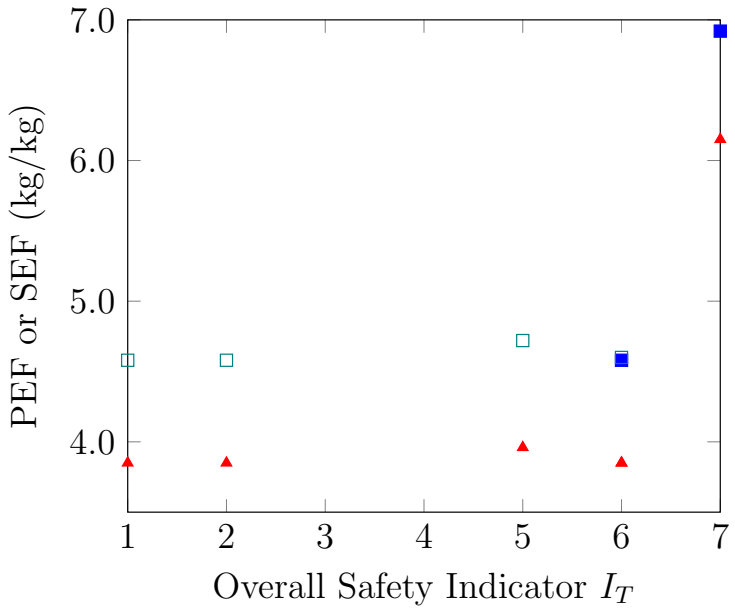

Figure 4.3: Pareto front of the BOO problem for minimising the PEF and maximising SHE performance shown as solid blue squares. The open green squares represent dominated solutions, whereas the red triangles represent the SEF value of each reported solution in Table 4.8.

Minimising Q_T and maximising Y_P

It is expected that higher process yields, which are linked to higher crystal yields, require higher process energy consumption; for example, significant temperature drop requirements during crystallisation (higher Q_C^-) may contribute significantly to Q_T. Accordingly, the seventh formulation of the end-to-end design problem involves minimising Q_T and maximising Y_P to explore the competing relationship between these two KPIs. This BOO problem is solved using the ϵ-constraint method, in which Q_T is minimised while Y_P is constrained by a given lower bound ϵ ranging between 81% and 89%. The Pareto front representing the Pareto-optimal solutions of this problem is shown in Figure 4.4, and the details of each solution are given in Table 4.9. An interesting finding is that all solutions, except Solution 4, involve hybrid cooling and anti-solvent crystallisation. It can be seen that an increase in Y_P from 85% to 87% requires a significant increase in Q_T from 1502 kJ/kg API to 1967 kJ/kg API; accordingly, the design point (Q_T,Y_P) = (1502,85) would be a good compromise solution.

Table 4.9: Pareto-optimal solutions of the BOO problem for minimising Q_T and maximising Y_P.

Solution	ϵ (%)	Solvents	Q_T (kJ/kg API)	Y_P (%)	Y_C (%)	T_2 (K)	T_5 (K)	$\frac{x_{s1,2}}{x_{s2,2}}$	$\frac{x_{s1,5}}{x_{s2,5}}$	Total solvent use (kg)
1	81	s1: 1,4-Dioxane s2: Formic acid	1315	82	90.00	364	301	1.70	1.14	419
2	83	s1: DMAC s2: o-Xylene	1342	83	91.58	364	297	9.00	0.53	466
3	85	s1: DMAC s2: Chlorobenzene	1502	85	93.80	384	292	1.10	0.36	364
4	87	s1: Toluene s2: DMF	1967	87	96.45	374	290	3.23	3.23	433
5	89	s1: Chlorobenzene s2: o-Xylene	2922	89	98.86	394	297	9.00	1.40	874

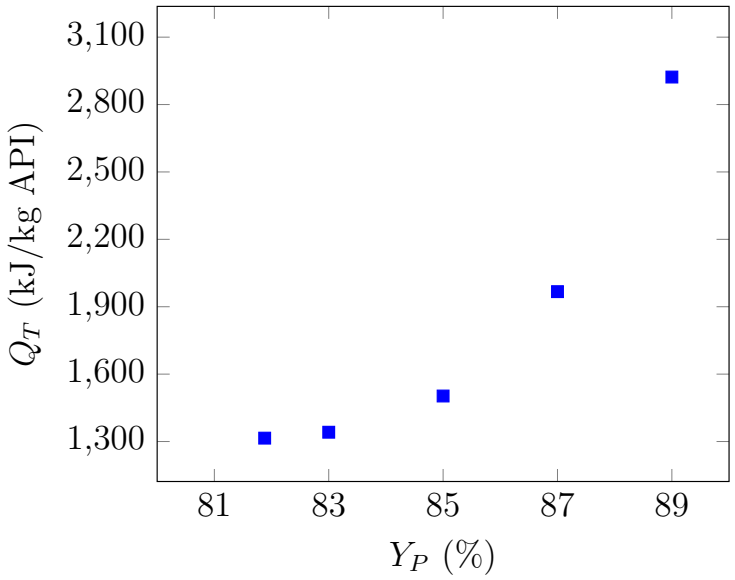

Figure 4.4: Pareto front of the BOO problem for minimising Q_T and maximising Y_P.

Figure 4.5 shows the energy-related contributions of the process units to Q_T, where Q_H and Q_R are combined in a single contribution (heating the reactor feed + maintaining isothermal reactor operation), with values reported in Table 4.10. It can be seen that Q_C^- is the largest of all energy contributions in each of the reported solutions, and increases with increasing Y_P; hence, Q_C^- is the main reason for the competing relationship between Q_T and Y_P. This is expected since a greater Y_P generally requires a greater Y_C (see Table 4.9) as more product needs to be crystallised from solution and sent to the isolation steps. This is driven by larger temperature drops in the crystalliser and/or anti-solvent addition, leading to higher Q_C^- requirements. Finally,

Table 4.10: Process unit energy contributions and the total process energy consumption for the five reported Pareto-optimal solutions in Table 4.9.

Solution	$Q_H + \|Q_R\|$ (kJ/kg API)	Q_C^- (kJ/kg API)	Q_D (kJ/kg API)	Q_T (kJ/kg API)
1	166	785	364	1315
2	160	818	364	1342
3	110	951	441	1502
4	448	1102	417	1967
5	715	1810	397	2922

it can be seen that the Q_D values of the reported solutions are close to each other and are significantly less than their Q_C^- counterparts; this is because there is a significant reduction in the amount of reaction/crystallisation solvent(s) that reaches the dryer due to the two filtration stages involved in the end-to-end process. Accordingly, the Q_D values mainly correspond to the energy required to vaporise the wash solvent, n-Heptane, which is the main component of the solvent mixture entering the dryer (Stream 12) as seen in Table 4.11.

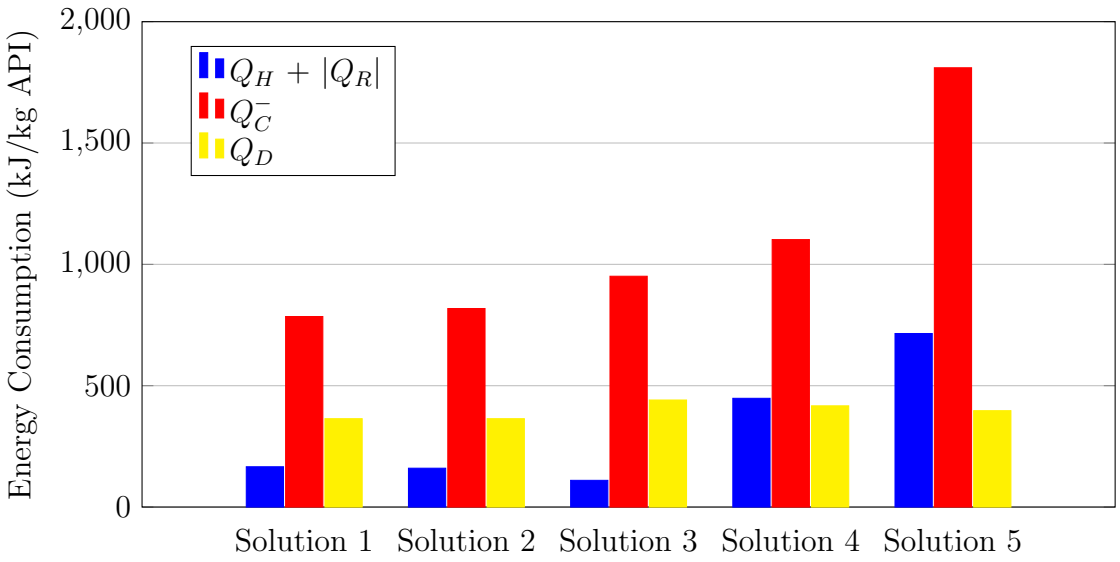

Figure 4.5: Process unit energy contributions to Q_T for the five reported Pareto-optimal solutions in Table 4.9.

In order to better understand solvent effects on process energy consumption, Figure 4.6 shows the molar flows of solvents $s1$, $s2$ and sw, which are the main components of the process streams, in the end-to-end process of Solution 4 and that of Solution 5 (Table 4.9). The molar quantities highlighted in blue, red and yellow correspond to those being subject to heating or cooling in the heat exchanger/reactor, crystalliser and dryer, respectively. The larger solvent

quantities entering the heat exchanger/reactor in Solution 5 relative to Solution 4 contribute to the larger reactor duty in Solution 5. Similarly, the larger solvent quantities being subject to cooling in the crystalliser in Solution 5, which involves anti-solvent addition ($N_{s2,4} \neq 0$), relative to Solution 4 (cooling crystallisation) contribute to the larger crystalliser duty in Solution 5. However, the larger solvent quantities entering the dryer in Solution 4 relative to Solution 5 contribute to the slightly larger dryer duty in Solution 4, even though washing in Solution 5 involves a larger amount of wash solvent ($F_w = 5.00$) relative to Solution 4 ($F_w = 2.48$). It should also be noted that in addition to solvent molar quantities, solvent properties such as heat capacity and heat of vaporisation contribute to the process unit duties as clearly seen in the corresponding, previously-derived expressions of Q_H (Equation (3.42)), Q_R (Equation (3.43)), Q_C^- (Equation (3.49)) and Q_D (Equation (4.42)).

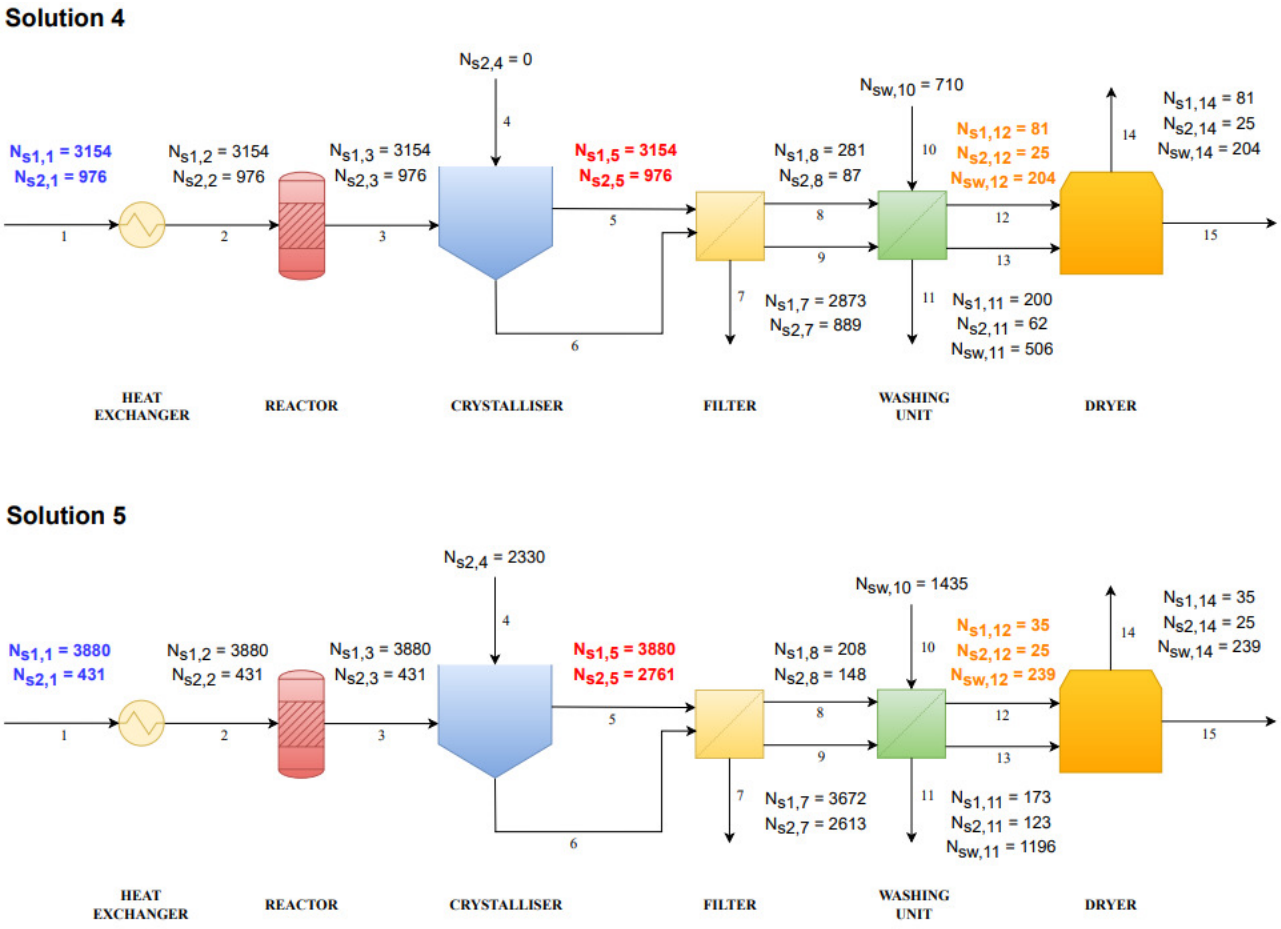

Figure 4.6: The molar quantities, in moles and rounded to the nearest integer, of solvents $s1$, $s2$ and sw in the end-to-end processes corresponding to Solutions 4 and 5 in Table 4.9. Quantities highlighted in blue, red and yellow are those contributing to the energy duty of the heat exchanger/reactor, crystalliser and dryer, respectively.

Table 4.11: Molar amounts of s1, s2 and *sw* in Stream 12 and F_w values of the reported Pareto-optimal solutions of the BOO problem involving Q_T and Y_P.

Solution	Solvent $q \in V$	$N_{q,12}$ (moles)	F_w (dimensionless)
1	s1: 1,4-Dioxane	47.03	5.00
	s2: Formic acid	41.20	
	sw: n-Heptane	239.50	
2	s1: DMAC	21.56	4.22
	s2: o-Xylene	40.57	
	sw: n-Heptane	232.43	
3	s1: DMAC	22.80	3.00
	s2: Chlorobenzene	62.61	
	sw: n-Heptane	215.93	
4	s1: Toluene	81.17	2.48
	s2: DMF	25.13	
	sw: n-Heptane	204.65	
5	s1: Chlorobenzene	34.70	5.00
	s2: o-Xylene	24.69	
	sw: n-Heptane	239.17	

Minimising the SEF, maximising Y_P and maximising PP

All of the presented MOO design problems so far were limited to 2 objective functions. In order to explore the applicability of the proposed methodology to more than 2 objectives, the eighth and final formulation of the end-to-end design problem involves minimising the SEF, maximising Y_P and maximising PP. This MOO problem is solved using AUGMENCON, where the SEF is set as the main objective function while Y_P and PP are transformed into constraints as demonstrated in Section 2.4.2. The corresponding mathematical formulation can hence be represented as follows:

$$\begin{aligned} \max \quad & -SEF + \kappa\left(\frac{s_2}{r_2} + \frac{s_3}{r_3}\right) \\ \text{s.t.} \quad & Y_P - s_2 = \epsilon_2 \\ & \text{PP} - s_3 = \epsilon_3 \end{aligned} \tag{4.50}$$

where κ is set to 10^{-3} and the ranges r_2 and r_3 are determined by recording the values of the objective functions at the optimal solution of each of the three single-objective optimisation problems. These values are provided in Table 4.12 where it can be seen that $81.80\% \leq \epsilon_2 \leq 88.10\%$ and $95.70\% \leq \epsilon_3 \leq 99.60\%$. Five different values for each of ϵ_2 and ϵ_3 are considered by dividing their respective ranges into four equal intervals. Accordingly, this MOO problem

involves solving 25 instances of the problem to cover the different combinations of ϵ_2 and ϵ_3.

Table 4.12: Optimal KPI values used in determining the objective function ranges in the MOO problem involving the SEF, Y_P and PP.

Objective function	KPI		
	SEF (kg/kg)	Y_P (%)	PP (%)
Min SEF	3.82	81.80	95.70
Max Y_P	5.29	88.10	98.60
Max PP	12.82	86.00	99.60

The Pareto-optimal solutions of this MOO problem are given in Table 4.13. It can be seen that there is an inverse relationship between the SEF and PP. This is expected as achieving higher purity requires the consumption of additional amounts of wash solvent. The relationship between Y_P and the other two KPIs is not as straightforward. In general, the relationship between Y_P and PP is direct since a larger amount of solid-phase API throughout the process would contribute to a high Y_P and, if that amount is significantly larger than that of impurities, to a high PP. However, at very high purity, the large use of wash solvent (large SEF) may lead to product loss if more API is rejected with the impurities. This could result in a reduction in Y_P if the ratio of the solid-phase API molar quantity to that of the limiting reagent decreases, even if PP is increased due to a further reduction in the amount of impurities relative to that of the API. This explains why the 2-Pentanol/Anisole design (Solutions 4, 9 and 19) exhibits a slight decrease in Y_P relative to the preceding solution. Ultimately, the relationship between the SEF, Y_P and PP highly depends on the chemical and process system considered; for example, a solvent mixture that induces high crystal yield of the API may lead to enhanced Y_P even at high PP levels. This is the case in Solution 14 (Anisole/Isobutyl acetate with $Y_C = 97.36\%$) which exhibits an increase in all 3 KPIs relative to the previous solutions for which $\epsilon_2 = 84.95\%$. Some ϵ_2-ϵ_3 combinations lead to infeasible solutions due to high PP targets (Solutions 5, 10, 15 and 20) as well as high Y_P and PP targets (Solutions 24 and 25). For $\epsilon_2 = 88.10\%$ and $\epsilon_3 = 95.70\%$, 96.68% and 97.65%, the same solution (Nitrobenzene/Cyclohexanone) is identified. It is in fact the solution reported in Table 4.12 for the Max Y_P problem.

Table 4.13: Pareto-optimal solutions of the MOO problem involving the SEF, Y_P and PP.

Solution	ϵ_2 (%)	ϵ_3 (%)	Solvents	SEF (kg/kg)	Y_P (%)	PP (%)
1	81.80	95.70	s1: Anisole	3.84	84.90	95.70
			s2: Acetic acid			
2		96.68	s1: Anisole	3.93	87.12	96.68
			s2: n-Butanol			
3		97.65	s1: Nitrobenzene	4.29	87.14	97.65
			s2: n-Butanol			
4		98.63	s1: 2-Pentanol	5.16	86.81	98.63
			s2: Anisole			
5		99.60		Infeasible Solution		
6	83.38	95.70	s1: 2-Pentanol	3.82	83.38	95.74
			s2: Anisole			
7		96.68	s1: Anisole	4.00	87.82	96.68
			s2: Isobutyl acetate			
8		97.65	s1: n-Butyl acetate	4.37	87.83	97.65
			s2: Anisole			
9		98.63	s1: 2-Pentanol	5.16	86.81	98.63
			s2: Anisole			
10		99.60		Infeasible Solution		
11	84.95	95.70	s1: Anisole	3.83	84.95	95.76
			s2: Isobutyl acetate			
12		96.68	s1: Anisole	3.95	86.78	96.68
			s2: 2-Pentanol			
13		97.65	s1: Anisole	4.27	87.10	97.65
			s2: n-Butanol			
14		98.63	s1: Anisole	5.26	87.84	98.63
			s2: Isobutyl acetate			
15		99.60		Infeasible Solution		

Solution	ϵ_2 (%)	ϵ_3 (%)	Solvents	SEF (kg/kg)	Y_P (%)	PP (%)
16	86.53	95.70	s1: Nitrobenzene s2: Cyclohexanone	3.84	86.53	95.70
17		96.68	s1: Nitrobenzene s2: n-Butanol	3.94	87.18	96.68
18		97.65	s1: Anisole s2: n-Butanol	4.27	87.10	97.65
19		98.63	s1: 2-Pentanol s2: Anisole	5.16	86.81	98.63
20		99.60		Infeasible Solution		
21, 22, 23	88.10	95.70, 96.68, 97.65	s1: Nitrobenzene s2: Cyclohexanone	5.29	88.10	98.56
24		98.63		Infeasible Solution		
25		99.60		Infeasible Solution		

The feasible solutions reported in Table 4.13 are displayed in Figures 4.7, 4.8 and 4.9, and are labelled by solution number. In each of these figures, the Pareto-optimal solutions are represented by the values of two KPIs and are color-coded by the value of the third KPI.

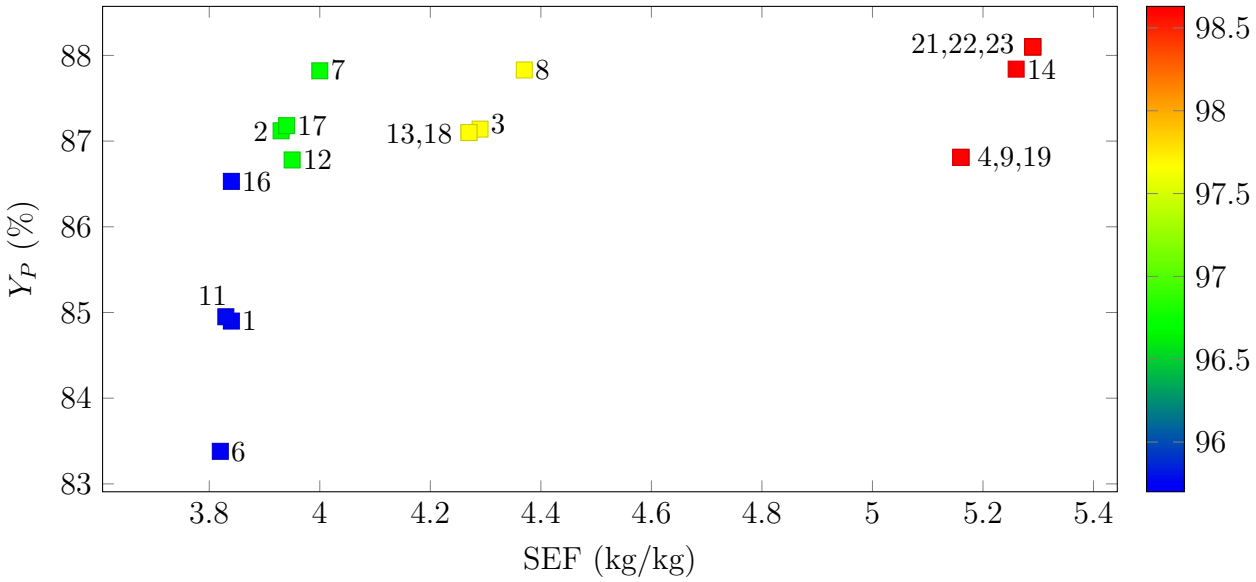

Figure 4.7: Plot of the Y_P (%) and SEF (kg/kg) values of the Pareto-optimal solutions of the MOO problem involving the SEF, Y_P and PP. The solutions are labelled by solution number as given in Table 4.13. The color scale corresponds to the PP value (%).

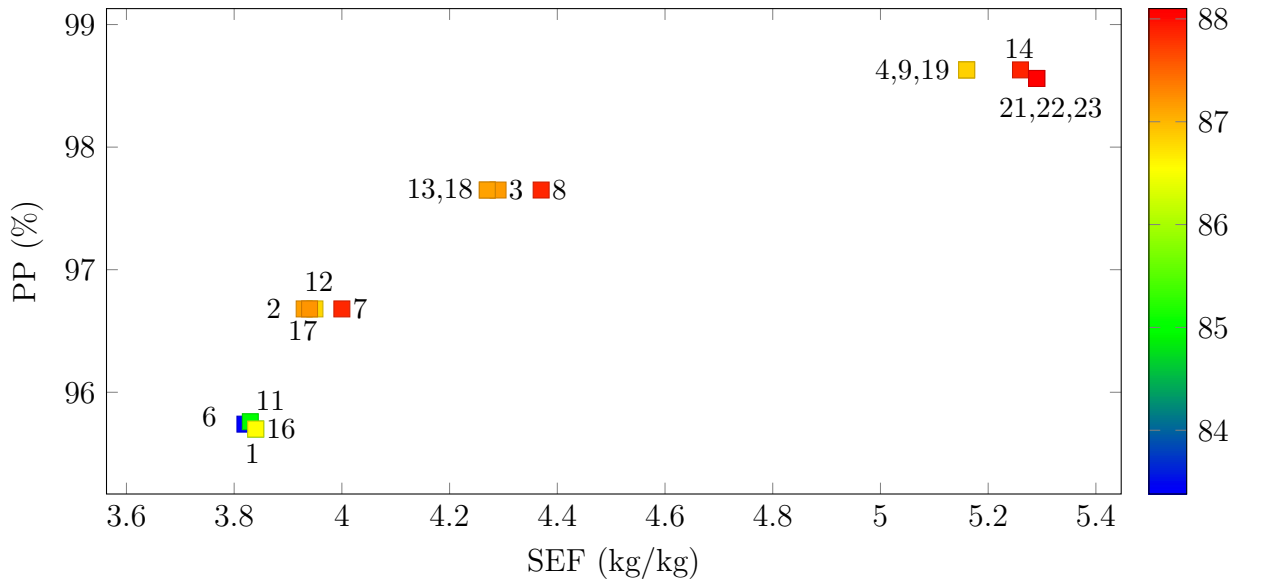

Figure 4.8: Plot of the PP (%) and SEF (kg/kg) values of the Pareto-optimal solutions of the MOO problem involving the SEF, Y_P and PP. The solutions are labelled by solution number as given in Table 4.13. The color scale corresponds to the Y_P value (%).

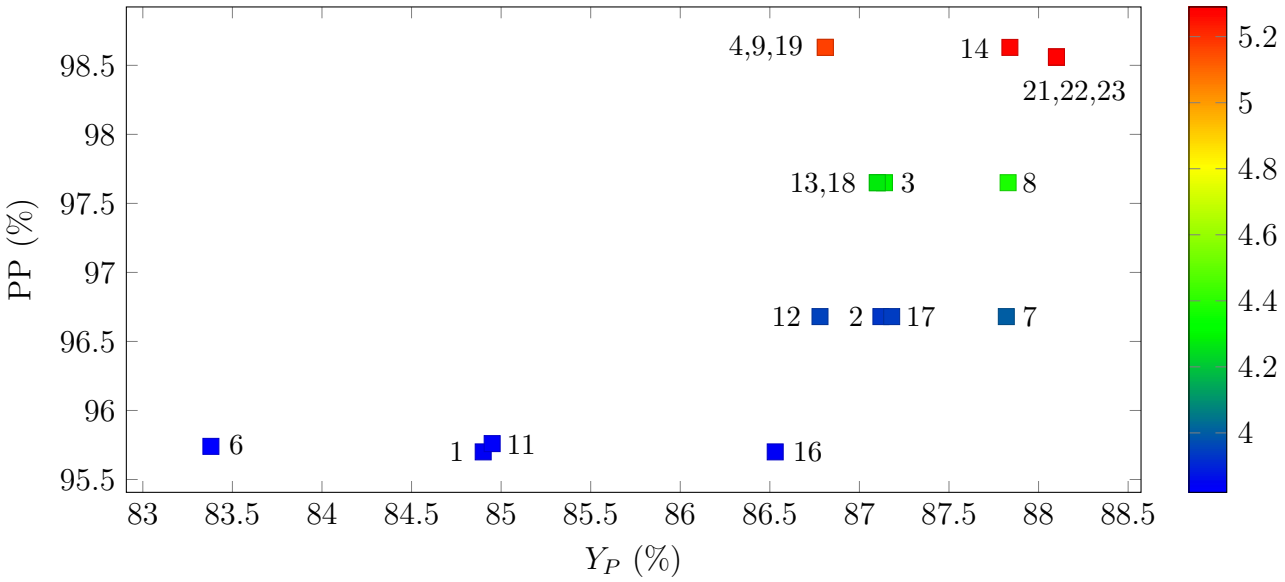

Figure 4.9: Plot of the PP (%) and Y_P (%) values of the Pareto-optimal solutions of the MOO problem involving the SEF, Y_P and PP. The solutions are labelled by solution number as given in Table 4.13. The color scale corresponds to the SEF value (kg/kg).

It is important to mention that the CPU time (in seconds) needed to generate the solutions of each of the integrated end-to-end design problems presented in this chapter is approximately 2-3 times that of the integrated synthesis and crystallisation design problem considered in Chapter 3. This shows that even with the extended design formulation, the proposed CAM^{b}D-based

approach can identify a range of solvent and process design options that optimise process performance in reasonable computational time.

4.4 Conclusion

In this chapter, a systematic solvent selection approach for the integrated synthesis, crystallisation and isolation of APIs was presented. The proposed CAM^{b}D formulation identifies the solvent mixtures, mixture composition and process conditions that optimise selected KPIs, while meeting comprehensive design specifications such as the miscibility of the reaction/crystallisation solvent(s) with the wash solvent, and species solubility across synthesis, crystallisation and washing.

The approach was illustrated by identifying optimal solvent and process designs for the integrated synthesis, crystallisation and isolation of mefenamic acid. Furthermore, three multi-objective CAM^{b}D design problems were formulated to explore the trade-offs between competing KPIs, namely the solvent E-factor (SEF) and product purity (PP), process energy consumption (Q_T) and process yield (Y_P), and the SEF, Y_P and PP. Some interesting findings were observed; for example, solving the SEF/PP bi-objective design problem generated different designs with varying SEF/PP values, showing that a PP of 98.8% can be achieved without significantly deteriorating the SEF. Additionally, solving the Q_T/Y_P bi-objective design problem showed that the crystalliser energy consumption, Q_C^-, was the greatest contributor to the total energy consumption, Q_T, of the end-to-end process due to the link between high crystal yield and process yield maximisation. Finally, solving the SEF/Y_P/PP multi-objective design problem highlighted some interesting trade-offs that exist between the three objective functions, particularly between Y_P and the other two KPIs.

The next chapter focuses on using CAM^{b}D to evaluate synthetic routes in a given reaction network based on process performance metrics, including resource efficiency, product quality and process economics.

Chapter 5

Solvent Selection for Telescoped Multi-step Process Routes

5.1 Introduction

In this chapter, the solvent selection approach proposed in Chapter 3 is extended and applied to screen multi-step synthetic routes based on process performance. For each synthetic route considered, a conceptual process route consisting of integrated reaction and separation steps is developed, and the CAM^{b}D method is used to simultaneously identify, for each route, the solvent/anti-solvent mixture, mixture composition and process conditions that optimise selected KPIs of resource efficiency, product quality and solvent cost. In the case of competing KPIs, route comparison is achieved by solving bi-objective optimisation (BOO) CAM^{b}D problems and visually exploring the Pareto-fronts of the studied routes. The design approach is illustrated by evaluating two reaction pathways for the two-step synthesis of 4-nitrophenol.

As discussed in Section 2.3, current approaches to synthetic route selection are mostly based on experimental findings, with little to no consideration of process performance metrics. This can lead to the identification of routes that perform optimally at the lab scale but poorly at the process or production level. This chapter hence introduces a CAM^{b}D-based approach to route selection in which multiple KPIs related to process performance are optimised in order to identify potential process routes that can be taken forward for further evaluation during process

development.

5.2 Problem Definition and Formulation

5.2.1 General design problem formulation

The CAM^{b}D problem tackled in this chapter can be described as follows: Given several routes for the multi-step synthesis of a pharmaceutical compound, a specified throughput, reaction conversion and selectivity, and a list of possible solvents, identify a telescoped route and its corresponding process design (solvent/anti-solvent mixture, process stream compositions and process conditions) that optimise one or more KPIs. While pharmaceutical process routes usually entail several reaction steps, each of which has a train of workup steps (e.g., liquid-liquid extraction, crystallisation, isolation), the current CAM^{b}D approach considers a telescoped process, i.e., it identifies solvents that can be used across the different reaction and separation steps of each route, due to the material, energy and cost-related benefits of process telescoping [199]. The shortlisted solvent candidates corresponding to the optimal route can be taken forward to the next steps in process development, e.g., experimental studies, and if their process performance is validated, they can be used in practice, eliminating the need for resource-intensive solvent swaps.

A simple extension of the process shown in Figure 3.1 includes two consecutive reactors, with no inter-stage workup, followed by a crystallisation unit. The corresponding conceptual flowsheet is shown in Figure 5.1. This process configuration can apply to a batch or continuous process; in the case of a batch process, each stream would represent the initial or final point of the batch operation. The process can be described as follows. Stream 1, the feed stream, is made up of reactants and impurities dissolved in either a pure solvent $s1$ or a binary solvent mixture ($s1$+$s2$) that enter a heat exchanger at a temperature T_1 (K). The heated mixture enters the first reactor as Stream 2 at a temperature T_2 (K) where the first reaction proceeds in a single step and can be accompanied by competing unwanted/side reactions. The product stream exiting the first reactor, Stream 3, is a liquid mixture of the desired product **D**, unreacted material, impurities, and unwanted side and byproducts at a temperature T_3 (K). This stream is combined

with a pure stream of reactant(s) involved in the second reaction, Stream 4, at a temperature T_3 (K) and sent to a second heat exchanger as Stream 5 to change its temperature to that of the second reaction, T_6 (K), after which it is fed to the second reactor as Stream 6. Similar to the first reaction, the second reaction proceeds in a single step and may also be accompanied by side reactions. The stream exiting the second reactor, Stream 7, is a homogeneous (liquid) mixture of the desired product **D** as well as unconverted reactants, impurities, and undesired side and byproducts, from the first and second reactions, at a temperature T_7 (K). Stream 7 is sent to a crystalliser, which can be designed as a cooling, anti-solvent or hybrid cooling and anti-solvent crystallisation unit. In the case of anti-solvent crystallisation, Stream 8, composed of pure anti-solvent $s2$ at temperature T_7 (K), is fed to the crystalliser. The crystalliser outlet is a slurry composed of a liquid phase (Stream 9), normally containing dissolved amounts of all process components, and a solid phase (Stream 10), typically a pure solid form of **D**. Both Streams 9 and 10 are at a temperature T_9 (K).

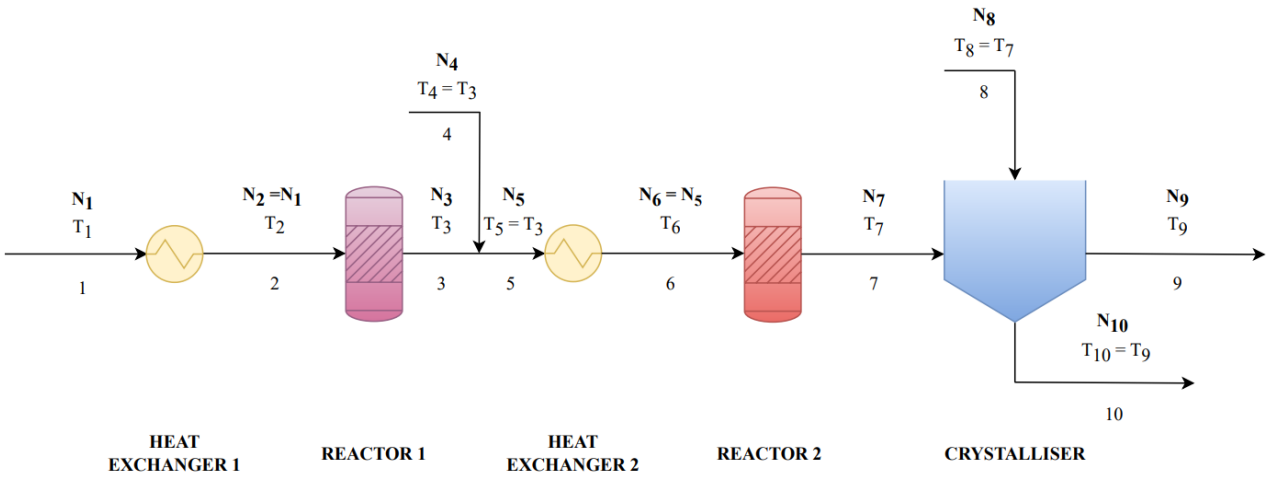

Figure 5.1: The conceptual flowsheet used for the solvent selection problem of the integrated, two-step synthesis and crystallisation process. N denotes the vector of component mole numbers (batch process) or component molar flowrates (continuous process) and T the vector of stream temperatures.

The sets and the big-M formulation (equation (3.1)) discussed in Section 3.2.1 apply in the route selection model, taking into account the additional process streams, i.e., the set $T = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, and the constraints related to the additional heat exchanger and reactor when formulating the MINLP problem.

5.2.2 Process model and process design constraints

Table 5.1 summarises any additional equations or changes in the proposed two-step synthesis and crystallisation CAM^{b}D problem relative to the design problem developed in Chapter 3.

An important difference between the single-step and two-step design formulations is that the presence of impurities (unconverted material) from the first reaction is considered in the material balances and equilibrium constraints of all process units, i.e., the two heat exchangers, the two reactors and the crystalliser. For example, unconverted material from the first reactor must remain under-saturated in the outlet of that reactor (Stream 3), in the second reactor (Streams 6 and 7) and after crystallisation (Stream 9). This makes the extended CAM^{b}D formulation more industrially-relevant by addressing the rejection of impurities generated from the multi-step syntheses of pharmaceutical compounds.

Table 5.1: Summary of any additional equations in the multi-step synthesis CAM^{b}D formulation relative to the single-step formulation developed in Chapter 3.

Model equation	Two-step compared to single-step synthesis formulation
Solvent assignment and properties	Same as single-step formulation
Masses and mole fractions	Same as single-step formulation but applied to the two input streams to the additional heat exchanger and reactor
Heat exchanger balances	Same as single-step formulation but applied to the additional heat exchanger
Reactor balances and equilibrium constraints	Same as single-step formulation but applied to the additional reactor; Reactor 2 equations involve species from the first and second reactions
Crystalliser balances and equilibrium constraints	Same as single-step formulation; Crystalliser equations involve species from the first and second reactions
Operating temperature constraints	Same as single-step formulation but applied to the additional reactor
Solvent miscibility constraints	Same as single-step formulation but applied to the two input streams to the additional heat exchanger and reactor

5.2.3 KPIs

The KPIs of interest in the route selection CAM^{b}D formulation include the previously introduced PEF, Y_C, and Q_T (Section 3.2.4), as well as a solvent cost KPI, C_T. These KPIs are defined as

follows:

- Process E-factor or PEF (kg Waste/kg Crystals) is defined as the mass of waste generated per mass of crystallised target product **D**:

$$\text{PEF} = \frac{\sum_{i \in C} M_{i,9}}{M_{D,10}}; \tag{5.1}$$

- Crystal Yield or Y_C (%) is defined as the percentage in moles of product **D** crystallised out of solution:

$$Y_C = \frac{N_{D,10}}{N_{D,7}} \times 100\%; \tag{5.2}$$

- Process energy consumption Q_T (kJ/kg API) is expressed as the sum of the heat duties of the individual process units. This includes the duties of the two heat exchangers, the two reactors, and the crystalliser. In this section, the process heat duties are mathematically expressed by performing energy balances around the process units, under the assumption of ideal mixing.

  1. First heat exchanger duty Q_{H1}. The energy balance around the first heat exchanger is given by:

$$Q_{H1} = \sum_{i \in C} N_{i,1} \int_{T_1}^{T_2} C_{p_i}^{(L)} dT. \tag{5.3}$$

     Here Q_{H1} is always positive as $T_2 \geq T_1$.

  2. First reactor duty Q_{R1}. The energy balance around the first reactor is given by:

$$Q_{R1} = \sum_{i \in C} N_{i,2} \int_{T_2}^{T_R} C_{p_i}^{(L)} dT + \xi_1 \Delta H_{R1}(T_R) + \sum_{i \in C} N_{i,3} \int_{T_R}^{T_3} C_{p_i}^{(L)} dT, \tag{5.4}$$

     where T_R is a reference temperature (usually 298.15 K), ξ_1 is the extent of the first reaction and ΔH_{R1} is the standard heat of the first reaction.

  3. Second heat exchanger duty Q_{H2}. The energy balance around the second heat exchanger is given by:

$$Q_{H2} = \sum_{i \in C} N_{i,5} \int_{T_3}^{T_6} C_{p_i}^{(L)} dT. \tag{5.5}$$

4. Second reactor duty Q_{R2}. The energy balance around the second reactor is given by:

$$Q_{R2} = \sum_{i \in C} N_{i,6} \int_{T_6}^{T_R} C_{p_i}^{(L)} dT + \xi_2 \Delta H_{R2}(T_R) + \sum_{i \in C} N_{i,7} \int_{T_R}^{T_7} C_{p_i}^{(L)} dT, \tag{5.6}$$

where ξ_2 is the extent of the second reaction and ΔH_{R2} is the standard heat of the second reaction.

5. Crystalliser duty Q_C^-. As derived in Section 3.2.4, the energy balance around the crystalliser is given by:

$$Q_C^- = \sum_{i \in C} N_{i,9} \int_{T_9}^{T_7} C_{p_i}^{(L)} dT + N_{D,10} \Bigg(\int_{T_m}^{T_7} C_{p_{API}}^{(L)} dT + \Delta H_m(T_m) + \int_{T_9}^{T_m} C_{p_{API}}^{(S)} dT \Bigg). \tag{5.7}$$

The energy KPI can be mathematically expressed as:

$$Q_T = Q_{H1} + Q_{H2} + |Q_{R1}| + |Q_{R2}| + Q_C^-, \tag{5.8}$$

where absolute values are used to ensure the cumulative energy requirements, whether heating or cooling, are accounted for.

The absolute value terms in equation (5.8) can be removed by reformulating the energy KPI objective function as follows:

$$\min \qquad Q_{H1} + Q_{H2} + Q'_{R1} + Q'_{R2} + Q_C^-, \tag{5.9}$$

and adding the following four constraints to the full formulation:

$$Q_{R1} \leq Q'_{R1}, \tag{5.10}$$

$$-Q_{R1} \leq Q'_{R1}, \tag{5.11}$$

$$Q_{R2} \leq Q'_{R2}, \tag{5.12}$$

$$-Q_{R2} \leq Q'_{R2}, \tag{5.13}$$

where Q'_{R1} and Q'_{R2} are variables used to incorporate the absolute values of Q_{R1} and Q_{R2}

in the expression of Q_T, respectively.

- Solvent cost C_T (£/kg API) is defined as the cost of solvents consumed in the process per mass of target product **D**:

$$C_T = \frac{\sum_{q \in Q} c_q V_{q,9}}{M_{D,10}}, \tag{5.14}$$

where c_q (£/L) and $V_{q,9}$ (L) are the purchase cost and consumed volume of solvent $q \in Q$, respectively. Values for c_q are obtained from the Chemical Availability Search section of the Physical Sciences Data-Science Service (PSDS): `https://chase-home-offsite.psds.ac.uk/`, and are provided in Appendix D. In selecting the solvent prices, the following considerations are taken into account:

  - The purity of the advertised solvent must be at least 99%; if this purity level is not available, then the price of the solvent with the highest available purity is selected.
  - The advertised solvent volume must be as high possible to benefit from available discounts.

  It is important to mention that solvent prices may differ depending on the source from which they are obtained, which adds a level of uncertainty to the designs generated by the model. Nevertheless, since solvent prices are model parameters in the CAM^{b}D formulation, they can be easily adjusted, for example, based on the solvent prices of a specific supplier.

5.3 Case Study: Two-step Synthesis of 4-Nitrophenol

The investigation of two-reaction pathways for the synthesis of 4-nitrophenol (NP), a key intermediate in the synthesis of acetaminophen (APAP), is used as a case study to illustrate the proposed CAM^{b}D approach to route selection. Since the chemical structure of NP is relatively simple, it can be synthesised in high yields via a small number of steps [155]. The two reaction pathways that will be explored are shown in Figure 5.2. Route 1 involves the nitration of chlorobenzene (CB) to 4-nitrochlorobenzene (NCB), followed by the hydroxylation of NCB to NP, whereas Route 2 involves the oxidation of cumene (CMN) to phenol (PNL), followed by the nitration of PNL to NP. As has been done in the previous chapters, all inorganic species are

omitted from the reaction schemes as they are not considered in the current design approach. The reader is referred to Joncour et al. [155] for a more detailed reaction network for NP synthesis, showing all reagents and species.

To model each reaction shown in the network in Figure 5.2, a literature review on typical temperatures at which these reactions proceed is carried out. This information can be used in CAM^{b}D by either: 1) fixing T_2 and T_6 to those values, or 2) defining ranges for T_2 and T_6 in which the aforementioned typical values fall for each route in the CAM^{b}D formulation. To increase the flexibility of the route selection approach, the option of defining reaction temperature ranges is chosen. Typical reaction temperatures (with references) and the temperature range used in the CAM^{b}D formulation for each reaction are given in Table 5.2.

Table 5.2: Typical temperature data and defined temperature ranges for the route selection CAM^{b}D formulation.

Reaction	Typical Temperature (K)	Typical Temperature Reference	CAM^{b}D Temperature Range (K)
Nitration of CB	293.15	Maleki et al. [200]	$293.15 \leq T_2 \leq 340$
Hydroxylation of NCB	393.15	Thakur et al. [201]	$350 \leq T_6 \leq 400$
Oxidation of CMN	323.15	Silva et al. [202]	$320 \leq T_2 \leq 350$
Nitration of PNL	293.15	Maleki et al. [200]	$293.15 \leq T_6 \leq 315$

For Route 1, the index sets are defined as follows: $C_1 = \{CB, NCB, NP, s1, s2\}$, $C_1^K = \{CB, NCB, NP\}$ and $Q_1 = \{s1, s2\}$, whereas for Route 2: $C_2 = \{CMN, PNL, NP, s1, s2\}$, $C_2^K = \{CMN, PNL, NP\}$ and $Q_2 = \{s1, s2\}$. The CAM^{b}D model for each of Route 1 and Route 2 is developed using the specifications given in Table 5.3. The model parameters (where required) are the same as those given in Table 3.2. The full mathematical CAM^{b}D formulation of the NP case study is given in Appendix E for Route 1 as an example.

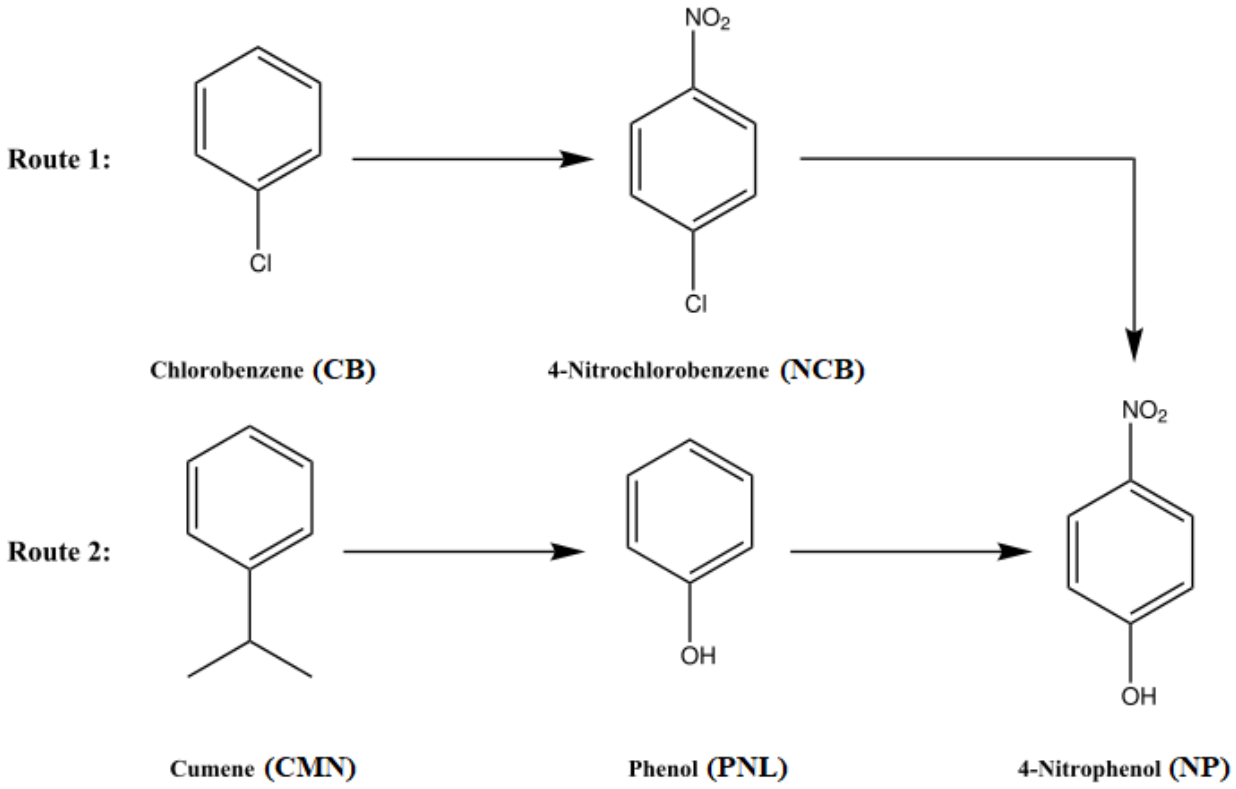

Figure 5.2: Two routes for the synthesis of NP.

Table 5.3: Specifications for the route selection CAM^{b}D formulation of the NP case study.

	Route 1	
Constraint	Equation	Units (where required)
Feed temperature	$T_1 = 298.15$	K
Isothermal reactor operation	$T_2 = T_3$ and $T_6 = T_7$	K
No NCB/NP in process feed	$N_{NCB,1} = N_{NP,1} = 0$	mol
Reaction 1 conversion	$x^{c_1} = 0.9$	–
Reaction 2 conversion	$x^{c_2} = 0.9$	–
Throughput	$M_{NP,10}$=80	kg
Reaction temperature range	See Table 5.2	K
Final crystallisation temperature range	$293.15 \leq T_9 \leq 320$	K
	Route 2	
Constraint	Equation	Units (where required)
Feed temperature	$T_1 = 298.15$	K
Isothermal reactor operation	$T_2 = T_3$ and $T_6 = T_7$	K
No PNL/NP in process feed	$N_{PNL,1} = N_{NP,1} = 0$	mol
Reaction 1 conversion	$x^{c_1} = 0.9$	–
Reaction 2 conversion	$x^{c_2} = 0.9$	–
Throughput	$M_{NP,10}$=80	kg
Reaction temperature range	See Table 5.2	K
Final crystallisation temperature range	$293.15 \leq T_9 \leq 320$	K

Since CB and CMN have very low melting points, they exist in the liquid phase within the

process temperature ranges considered. Accordingly, the mathematical formulation corresponding to Route 1 includes solubility equations for NCB and NP only, and that corresponding to Route 2 includes solubility equations for PNL and NP only.

Five CAM^{b}D formulations, that differ by the objective function(s), are considered to compare Routes 1 and 2 for NP synthesis. In the first formulation, the two routes are compared in terms of material efficiency, which is quantified by the PEF. In the second formulation, the two routes are compared in terms of energy efficiency, which is quantified by Q_T. In the third formulation, the two routes are compared in terms of solvent cost, which is quantified by C_T. In the fourth formulation, the two routes are compared in terms of both material and energy efficiency (resource efficiency) by solving a BOO version of the CAM^{b}D formulation, in which both the PEF and Q_T are minimised, and visualising the resultant Pareto fronts of the two routes. In the fifth and final formulation, the two routes are compared in terms of solvent cost and product quality by solving a second BOO problem, in which C_T is minimised while Y_C is maximised, and analysing the resultant Pareto fronts. In all design formulations, the optimal solvents are selected from the list of 49 solvent candidates (List III) given in Appendix B.3.

Solvent melting and boiling points are obtained from the Chemical Book (`https://www.chemicalbook.com/ProductIndex_EN.aspx`) [191]. The melting points of CB, NCB, CMN and PNL are obtained from the Chemical Book (`https://www.chemicalbook.com/ProductIndex_EN.aspx`), whereas the heats of melting of the aforementioned species are obtained from the NIST Chemistry WebBook (`https://webbook.nist.gov/chemistry`) [192]. Values of $T_{m,i}$ and $\Delta H_{m,i}$ for $i \in C^K$ are reported in Appendix C.1.

In all energy balances, heat capacity is assumed to be phase-independent, i.e., $C_{p_i}^{(L)} \approx C_{p_i}^{(S)}$ for $i \in C$, and to be a weak function of temperature, i.e., standard constant-pressure heat capacities of the chemical species are used. Standard constant-pressure heat capacities of species in set C are obtained from the NIST Chemistry WebBook (`https://webbook.nist.gov/chemistry`). Values of $C_{p_i}^{(L)}$ for $i \in C^K$ are reported in Appendix C.1. The standard heats of reaction for all reaction steps in the NP network are calculated from the heats of formation of the corresponding species, which are estimated using the Marrero-Gani GC method [121]. Standard heats of reaction are reported in Appendix C.2.

The optimisation problems are solved in GAMS version 32.2.0 [169] using the SBB solver [163]. Multiple high-performance solutions are generated for each design problem by including integer cuts in the MINLP formulation. For each solution, the following quantities are reported: the identities of solvent species $s1$ and $s2$, their proportions as molar percentages in Stream 9 (in parentheses), and the value of the objective function(s). The optimal process temperatures, T_2, T_6 and T_9, in all formulations are provided in Appendix F.

5.3.1 Route comparison in terms of minimum PEF

In the first formulation, Routes 1 and 2 are compared in terms of material efficiency by generating designs that minimise the PEF. Table 5.4 shows 10 optimal solutions ranked with respect to the PEF for each of Route 1 and Route 2. For Route 1, 2,2,4-Trimethyl pentane appears as an optimal solvent component in 6 out of the 10 reported solutions, and is a major solvent component in most of these solutions, namely Solutions 1, 2, 3 and 8. This indicates that 2,2,4-Trimethyl pentane can potentially contribute to dissolving the species of the nitration and hydroxylation reactions of Route 1, and induce the subsequent crystallisation of NP. Other recurrent solvent candidates include 2-Pentanone, Isopropyl acetate and Acetic acid as minor components, and Methyl cyclohexane and n-Heptane as major components in the optimal solvent mixtures. For Route 2, although 2,2,4-Trimethyl pentane is the most frequently identified optimal solvent component as in Route 1, several other candidates that were not identified as optimal for Route 1 are deemed so by the model for Route 2, including Methyl acetate, Chloroform, and Carbon tetrachloride as minor solvent components, and 2-Methyl pentane as a major solvent component.

Since environmental impact was not included as a design objective in the route selection formulations, it can happen that some solvents that must be avoided (ICH Class 1) or limited (ICH Class 2) in pharmaceutical manufacturing [190] are identified as optimal for the studied routes. For example, in terms of minimum PEF, Carbon tetrachloride (ICH Class 1), Chloroform (ICH Class 2) and n-Hexane (ICH Class 2) are identified as optimal solvents for Route 2, while 1,4-Dioxane (ICH Class 2) is identified as an optimal solvent for Route 1. This highlights the benefit of generating multiple solutions using CAM^{b}D for: 1) diversifying the solvent design space and hence the shortlisted solvent candidates so that multiple solvent combinations can

be tested when more detailed process-related criteria are considered, and 2) identifying several design options for the practitioner to choose from if one of the identified designs violates some desired criteria not fully considered in the mathematical formulation, such as environmental impact.

As a multi-faceted problem, route selection must be explored based on different design criteria; therefore, the next formulation looks at route comparison from an energy perspective.

Table 5.4: Comparison of Routes 1 and 2 in terms of minimum PEF.

Solution	Route 1		Route 2	
number	Solvents (Mole % in Stream 9)	PEF (kg/kg)	Solvents (Mole % in Stream 9)	PEF (kg/kg)
1	s1: n-Butanol (28.82%) s2: 2,2,4-Trimethyl pentane (71.18%)	3.5	s1: Acetone (34.92%) s2: 2,2,4-Trimethyl pentane (65.08%)	8.27
2	s1: Propyl acetate (49.84%) s2: 2,2,4-Trimethyl pentane (50.16%)	3.5	s1: Methyl acetate (24.10%) s2: n-Hexane (75.90%)	8.27
3	s1: 2-Pentanone (28.20%) s2: 2,2,4-Trimethyl pentane (71.80%)	3.5	s1: Methyl acetate (27.34%) s2: n-Heptane (72.66%)	8.30
4	s1: 2-Pentanone (21.68%) s2: Methyl cyclohexane (78.32%)	3.5	s1: Methyl acetate (30.52%) s2: 2,2,4-Trimethyl pentane (69.48%)	8.34
5	s1: Isopropyl acetate (21.25%) s2: Methyl cyclohexane (78.75%)	3.5	s1: Ethyl acetate (34.45%) s2: 2,2,4-Trimethyl pentane (65.55%)	9.10
6	s1: Isopropyl acetate (23.28%) s2: n-Heptane (76.72%)	3.5	s1: Chloroform (25.84%) s2: 2-Methyl pentane (74.16%)	9.58
7	s1: Acetic acid (28.63%) s2: n-Heptane (71.37%)	3.5	s1: Chloroform (29.44%) s2: n-Heptane (70.56%)	9.67
8	s1: Acetic acid (31.22%) s2: 2,2,4-Trimethyl pentane (68.78%)	3.5	s1: Chloroform (33.00%) s2: 2,2,4-Trimethyl pentane (67.00%)	9.78
9	s1: 1,4-Dioxane (68.97%) s2: 2,2,4-Trimethyl pentane (31.03%)	3.5	s1: Carbon tetrachloride (26.28%) s2: 2-Methyl pentane (73.72%)	10.54
10	s1: n-Butyl acetate (53.14%) s2: 2,2,4-Trimethyl pentane (46.86%)	3.5	s1: Carbon tetrachloride (33.60%) s2: 2,2,4-Trimethyl pentane (66.40%)	10.82

5.3.2 Route comparison in terms of minimum Q_T

In the second formulation, Routes 1 and 2 are compared in terms of energy efficiency by generating designs that minimise Q_T. Table 5.5 shows 10 optimal solutions ranked with respect

to Q_T for each of Route 1 and Route 2. For Route 1, new solvent components that did not appear in any of the solutions of the previous formulation (Section 5.3.1) are now identified as optimal, such as o-Xylene, Toluene and Nitrobenzene. For Route 2, several designs identified as optimal in the previous formulation (Section 5.3.1) appear again in the list of optimal solutions, namely Solutions 1, 2, 4, 5, 6, 7, 9 and 10, yet with different rankings.

It can be seen that Route 1 is less favourable than Route 2 from an energy perspective, with significantly higher Q_T values across the reported solutions. In fact, within the 10 reported solutions for each route, the worst solution for Route 2 (Q_T = 853 kJ/kg API) is better than the best solution for Route 1 (Q_T = 1141 kJ/kg API). Looking closely at the solutions of Route 2, it can be seen that Acetone/2,2,4-Trimethyl pentane, which was the best identified solvent combination from a material use perspective, is now ranked last when the problem is considered from an energy perspective, even though it still constitutes a good design relative to other solvent combinations as it appears in the top 10 solvents. The interesting observation is that the solution ranking is not preserved when the objective KPI is changed. As for Carbon tetrachloride/2,2,4-Trimethyl pentane, which was ranked last from a material use perspective, it now ranks first from an energy perspective. This highlights the importance of considering different design objectives when selecting an optimal process route. It is important to mention that each of the solutions Acetone/2,2,4-Trimethyl pentane and Carbon tetrachloride/2,2,4-Trimethyl pentane corresponds to the same design (solvent choices, mixture compositions and process conditions) in both design formulations, i.e., the minimum PEF and the minimum Q_T formulations, but is ranked differently depending on the objective function.

Figure 5.3 shows the energy breakdown of each of the top solutions of Route 1 and Route 2. It can be seen that the energy requirements of each of the process units in Route 1 exceed those in Route 2, which explains why Route 2 is superior to Route 1 from an energy perspective. Furthermore, it is clear that in both routes, the crystallisation energy consumption Q_C^- is the greatest contributor to the total energy consumption Q_T, which is similar to the findings reported in Section 4.3.3. This highlights the benefit of telescoping from an energy perspective, as the exclusion of inter-stage crystallisation steps can significantly reduce the total energy requirements of drug substance manufacturing processes.

Table 5.5: Comparison of Routes 1 and 2 in terms of minimum Q_T.

Solution	Route 1		Route 2	
number	Solvents (Mole % in Stream 9)	Q_T (kJ/kg API)	Solvents (Mole % in Stream 9)	Q_T (kJ/kg API)
1	s1: o-Xylene (90.00%) s2: Toluene (10.00%)	1141	s1: Carbon tetrachloride (33.60%) s2: 2,2,4-Trimethyl pentane (66.40%)	572
2	s1: Chlorobenzene (10.00%) s2: o-Xylene (90.00%)	1148	s1: Carbon tetrachloride (30.00%) s2: n-Heptane (70.00%)	587
3	s1: Nitrobenzene (37.12%) s2: Methyl cyclohexane (62.88%)	1151	s1: Chloroform (33.00%) s2: 2,2,4-Trimethyl pentane (67.00%)	586
4	s1: Nitrobenzene (44.55%) s2: n-Heptane (55.45%)	1161	s1: Carbon tetrachloride (26.28%) s2: 2-Methyl pentane (73.72%)	592
5	s1: Nitrobenzene (47.00%) s2: 2,2,4-Trimethyl pentane (53.00%)	1197	s1: Chloroform (29.44%) s2: n-Heptane (70.56%)	603
6	s1: n-Heptane (67.24%) 1,4-Dioxane (32.76%)	1218	s1: Chloroform (25.84%) s2: 2-Methyl pentane (74.16%)	609
7	s1: Methyl cyclohexane (72.33%) s2: 1,4-Dioxane (27.67%)	1221	s1: Methyl acetate (30.52%) s2: 2,2,4-Trimethyl pentane (69.48%)	626
8	s1: Acetic acid (37.33%) s2: n-Heptane (62.67%)	1221	s1: Carbon tetrachloride (26.30%) s2: n-Hexane (73.70%)	631
9	s1: Nitrobenzene (10.00%) s2: Toluene (90.00%)	1222	s1: Ethyl acetate (34.45%) s2: 2,2,4-Trimethyl pentane (65.55%)	805
10	s1: 2,2,4-Trimethyl pentane (62.41%) s2: 1,4-Dioxane (37.59%)	1230	s1: Acetone (34.92%) s2: 2,2,4-Trimethyl pentane (65.08%)	853

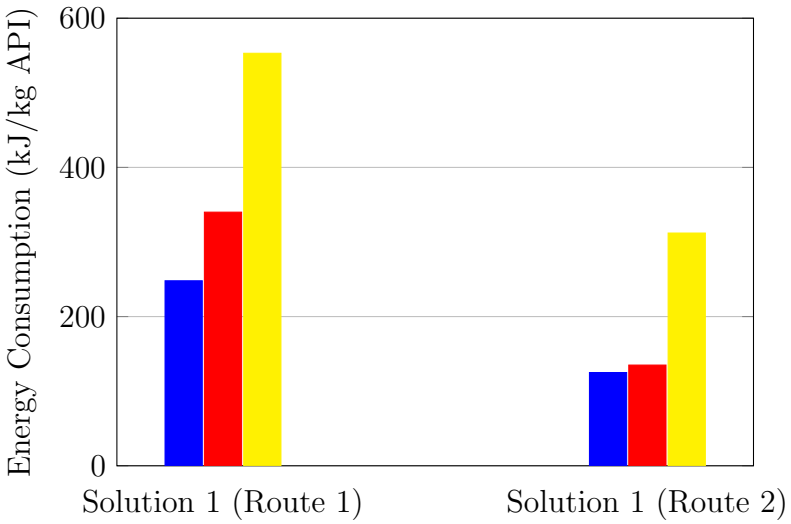

Figure 5.3: Process unit energy contributions to Q_T for the top solutions of Route 1 and Route 2 in the Q_T minimisation problem as given in Table 5.5. The blue and red bars represent the energy requirements for heating and maintaining isothermal operation in reactor 1 and reactor 2, respectively, and the yellow bar represents the crystallisation energy requirement.

5.3.3 Route comparison in terms of minimum C_T

In the third formulation, Routes 1 and 2 are compared in terms of the solvent cost KPI, C_T. Table 5.6 shows 10 optimal solutions ranked with respect to C_T for each of Route 1 and Route 2. For Route 1 designs, the C_T values are approximately the same at around £50/kg. It can be seen that 2,2,4-Trimethyl pentane, which was frequently identified as a major solvent component in the optimal solvent mixtures of the previous formulations, now only appears as a minor solvent component in Solution 4. Interestingly, Toluene appears as a major solvent component in Solutions 2, 7, 8, 9 and 10, and as a minor solvent component in Solution 1, and Nitrobenzene appears as a major solvent component in Solutions 1, 3, 4, 5 and 6. This could be explained by the cheaper prices of Toluene and Nitrobenzene, at £14.93/L and £18.55/L, respectively, relative to that of 2,2,4-Trimethyl pentane at £20.25/L.

For Route 2 designs, the C_T values range between £134.00/kg and £797.77/kg, which are significantly higher than those reported for Route 1. An interesting finding is that although Carbon tetrachloride/2,2,4-Trimethyl pentane is the best solvent combination for Route 2 in terms of Q_T, (Section 5.3.2), it is the worst solvent combination, out of 10 reported solutions, in terms of the PEF (Section 5.3.1) and C_T. This further proves that route selection is a

multi-objective design problem that cannot be studied based on a single KPI. Accordingly, from a solvent cost perspective, Route 1 is superior to Route 2.

In order to explore the contributions of $s1$ and $s2$ to C_T in the Route 2 designs, Table 5.7 reports the solvent volumes and unit prices corresponding to each of the 10 solutions. Solutions 7–10 correspond to distinctly high C_T values:

- in Solution 7, the high volume of 2,2,4-Trimethyl pentane ($V_{s2,9}$ = 703.13 L) contributes significantly to the high C_T value,
- in Solution 8, the high price of Methyl cyclopentane (c_{s2} = £101.72/L) contributes significantly to the high C_T value, and
- in Solutions 9 and 10, the high price of Carbon tetrachloride (c_{s2} = £268.00/L) contributes significantly to the high C_T value.

Table 5.6: Comparison of Routes 1 and 2 in terms of minimum C_T.

Solution	Route 1		Route 2	
number	Solvents (Mole % in Stream 9)	C_T (£/kg API)	Solvents (Mole % in Stream 9)	C_T (£/kg API)
1	s1: Toluene (4.84%) s2: Nitrobenzene (95.16%)	48.51	s1: Chloroform (39.04%) s2: Methyl cyclohexane (60.96%)	134.00
2	s1: Toluene (74.22%) s2: Isopropyl acetate (25.78%)	48.96	s1: Methyl acetate (30.49%) s2: Methyl cyclohexane (69.51%)	150.95
3	s1: n-Heptane (5.26%) s2: Nitrobenzene (94.74%)	50.00	s1: Chloroform (50.39%) s2: 2,2,4-Trimethyl pentane (49.61%)	151.65
4	s1: 2,2,4-Trimethyl pentane (5.45%) s2: Nitrobenzene (94.55%)	50.52	s1: Methyl acetate (37.25%) s2: n-Heptane (62.75%)	163.00
5	s1: Triethyl amine (6.10%) s2: Nitrobenzene (93.90%)	50.62	s1: Methyl acetate (41.87%) s2: 2,2,4-Trimethyl pentane (58.13%)	165.45
6	s1: o-Xylene (4.90%) s2: Nitrobenzene (95.10%)	51.38	s1: Ethyl acetate (40.44%) s2: 2,2,4-Trimethyl pentane (59.56%)	180.96
7	s1: n-Butanol (10.95%) s2: Toluene (89.05%)	51.42	s1: Acetone (35.55%) s2: 2,2,4-Trimethyl pentane (64.45%)	228.10
8	s1: Toluene (79.15%) s2: Acetic acid (20.85%)	51.83	s1: Chloroform (33.02%) s2: Methyl cyclopentane (66.98%)	621.51
9	s1: 2-Butanol (10.95%) s2: Toluene (89.05%)	52.10	s1: Carbon tetrachloride (41.15%) s2: Methyl cyclohexane (58.85%)	731.55
10	s1: Toluene (70.30%) s2: n-Butyl acetate (29.70%)	52.18	s1: Carbon tetrachloride (51.87%) s2: 2,2,4-Trimethyl pentane (48.13%)	797.77

5.3.4 Route comparison in terms of multiple KPIs

Given the competing relationship between route selection KPIs, the next two formulations involve BOO problems in which, for each route, 1) Q_T and the PEF are minimised, and 2) C_T is minimised while Y_C is maximised. The subsequent construction of the Pareto front corresponding to each route allows a visual comparison of route performance with respect to the optimised KPIs, hence guiding process and route design.

Table 5.7: Consumed volumes and units prices of s1 and s2 in the top 10 optimal Route 2 designs.

Solution	Solvent $q \in Q$	$V_{q,9}$ (L)	c_q (£/L)	$c_q V_{q,9}$ (£)	C_T (£/kg API)
1	s1: Chloroform	160.31	29.25	4689.07	134.00
	s2: Methyl cyclohexane	399.03	15.12	6033.33	
2	s1: Methyl acetate	144.60	28.20	4077.72	150.95
	s2: Methyl cyclohexane	529.00	15.12	7998.48	
3	s1: Chloroform	172.41	29.25	5042.99	151.65
	s2: 2,2,4-Trimethyl pentane	350.10	20.25	7089.53	
4	s1: Methyl acetate	149.17	28.20	4206.60	163.00
	s2: n-Heptane	463.12	19.08	8836.33	
5	s1: Methyl acetate	152.87	28.20	4310.93	165.45
	s2: 2,2,4-Trimethyl pentane	440.76	20.25	8925.40	
6	s1: Ethyl acetate	224.75	24.00	5394.00	180.96
	s2: 2,2,4-Trimethyl pentane	559.46	20.25	11329.07	
7	s1: Acetone	172.48	23.25	4010.16	228.10
	s2: 2,2,4-Trimethyl pentane	703.13	20.25	14238.40	
8	s1: Chloroform	155.86	29.25	4558.91	621.51
	s2: Methyl cyclopentane	443.96	101.72	45159.61	
9	s1: Carbon tetrachloride	197.31	268.00	52879.10	731.55
	s2: Methyl cyclohexane	373.40	15.12	5645.81	
10	s1: Carbon tetrachloride	212.64	268.00	56987.52	639.57
	s2: 2,2,4-Trimethyl pentane	337.46	20.25	6833.57	

Route comparison in terms of minimum PEF and minimum Q_T

In the fourth formulation, Routes 1 and 2 are compared in terms of two design objectives, namely the PEF and Q_T. The BOO problems are solved using the ϵ-constraint method in which Q_T is minimised while the PEF is bounded from below by a range of ϵ values ranging between 3.5 and 6.0 for Route 1, and between 7.0 and 9.5 for Route 2 (ϵ values less than 6 for Route 2 lead to infeasible solutions despite trying different initial guesses), with ϵ being incremented by 0.5 per iteration in each of the two problems. Figure 5.4 shows the Pareto-optimal solutions for Route 1 and Route 2, and the details of each solution are reported in Table 5.8.

The two separate clusters shown in Figure 5.4 indicate that when optimising the process route with respect to the PEF and Q_T, Route 2 is superior to Route 1 from an energy consumption perspective but inferior to Route 1 from a material use perspective as the corresponding Pareto front lies in the upper left region of the plot (high PEF, low Q_T), whereas Route 1 is superior to Route 2 from a material use perspective but inferior to Route 2 from an energy consumption perspective as the corresponding Pareto front lies in the lower right region of the plot (low PEF,

high Q_T). These findings are in agreement with the results reported for the single objective optimisation problems in Sections 5.3.1 and 5.3.2.

It is interesting to see that for Route 1, o-Xylene (Solution 5) is identified as an optimal pure solvent that can be used across the two synthesis steps and crystallisation. The flexibility of CAM^{b}D in identifying both pure solvents and solvent mixtures for process design gives the user/practitioner a wider choice of design options to meet their specific objectives.

Table 5.8: Pareto-optimal solutions of the BOO problem for minimising the PEF and Q_T in Routes 1 and 2.

Solution	Route 1			Route 2		
number	Solvents (Mole % in Stream 9)	PEF (kg/kg)	Q_T (kJ/kg API)	Solvents (Mole % in Stream 9)	PEF (kg/kg)	Q_T (kJ/kg API)
1	s1: 3-Pentanone (10.00%) s2: o-Xylene (90.00%)	3.50	1243	s1: Chloroform (47.00%) s2: 2,2,4-Trimethyl pentane (53.00%)	7.00	675
2	s1: o-Xylene (90.00%) s2: Nitrobenzene (10.00%)	4.00	1228	s1: Methyl acetate (33.68%) s2: 2,2,4-Trimethyl pentane (66.32%)	7.50	664
3	s1: Nitrobenzene (10.00%) s2: Toluene (90.00%)	4.34	1222	s1: Chloroform (40.76%) s2: 2,2,4-Trimethyl pentane (59.24%)	8.00	643
4	s1: Nitrobenzene (10.00%) s2: Toluene (90.00%)	4.34	1222	s1: Methyl acetate (34.00%) s2: 2,2,4-Trimethyl pentane (66.00%)	8.34	626
5	s1: o-Xylene (100.00%) s2: –	5.50	1150	s1: Chloroform (36.00%) s2: 2,2,4-Trimethyl pentane (64.00%)	9.00	610
6	s1: Chlorobenzene (10.00%) s2: o-Xylene (90.00%)	6.00	1149	s1: Chloroform (34.00%) s2: 2,2,4-Trimethyl pentane (66.00%)	9.50	594

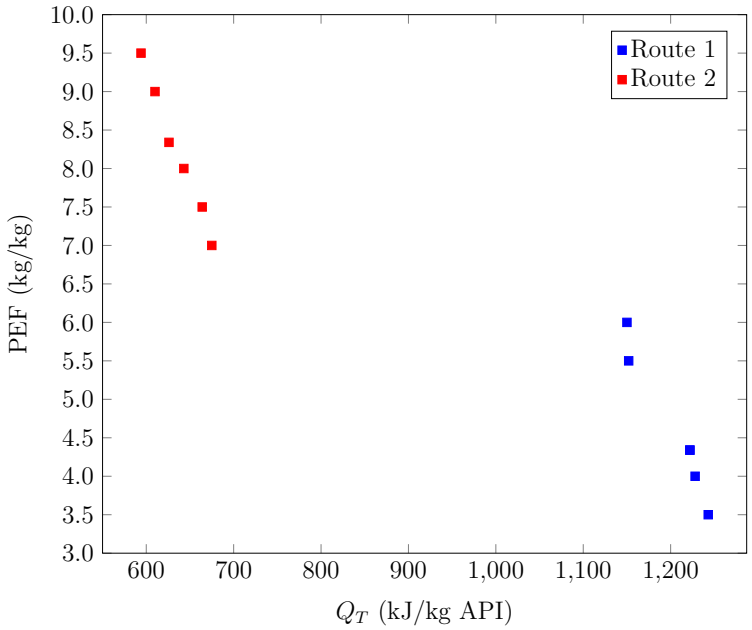

Figure 5.4: Pareto fronts of the BOO problem for minimising the PEF and Q_T in Routes 1 and 2.

Route comparison in terms of minimum C_T and maximum Y_C

In the fifth and final formulation, Routes 1 and 2 are once again compared in terms of two design objectives, namely C_T and the crystal yield Y_C. The BOO problems are solved using the ϵ-constraint method in which C_T is minimised while Y_C is bounded from below by a range of ϵ values ranging between 90% and 95%. Figure 5.5 shows the Pareto-optimal solutions for Route 1 and Route 2, and the details of each solution are reported in Table 5.9.

It can be see that for Route 1, when Y_C is increased from 90% to 99%, C_T increases from £48.52/kg to £89.30/kg, an increase of 84%. As for Route 2, when Y_C is increased from 90% to 95%, C_T increases from £134.03/kg to £843.17/kg, an increase of 529%. Accordingly, Route 1 is again superior to Route 2, now in terms of both solvent cost, quantified by C_T, and product quality, quantified by the crystal yield Y_C. However, an interesting observation is that Y_C is strictly greater than ϵ for $92\% \leq \epsilon \leq 95\%$, which explains the gap observed in the Pareto front of Route 1 in Figure 5.5. This indicates the difficulty of obtaining compromise solutions for $92\% \leq Y_C \leq 95\%$ for this specific route. For Route 2, a clear compromise design would be that corresponding to (C_T,Y_C)=(208.10,94) as beyond that point there is a large increase in solvent cost for a unit increase in crystal yield.

By examining the solvent components that promote a high crystal yield in both routes, it is found that in Route 1, Nitrobenzene and Methyl cyclohexane each appears as a major solvent component in multiple solutions, and the same goes for Methyl cyclohexane and 2,2,4-Trimethyl pentane in Route 2. Accordingly, it is worthwhile to further explore the effectiveness of these solvents in inducing product crystallisation with high yield, e.g., by experimental investigations. In fact, this is an important feature of the proposed CAM^{b}D tool: guiding experiments during process development to reduce time and cost and accelerate process design.

Table 5.9: Pareto-optimal solutions of the BOO problem for minimising C_T and maximising Y_C in Routes 1 and 2.

Solution	Route 1			Route 2		
number (ϵ (%))	Solvents (Mole % in Stream 9)	Y_C (%)	C_T (£/kg)	Solvents (Mole % in Stream 9)	Y_C (%)	C_T (£/kg)
1 (90)	s1: Toluene (4.84%) s2: Nitrobenzene (95.16%)	90	48.52	s1: Chloroform (39.00%) s2: Methyl cyclohexane (61.00%)	90	134.03
2 (91)	s1: Methyl cyclohexane (5.04%) s2: Nitrobenzene (94.96%)	91	49.08	s1: Methyl acetate (28.32%) s2: Methyl cyclohexane (71.68%)	91	160.18
3 (92)	s1: n-Butyl acetate (8.46%) s2: Toluene (91.54%)	96	54.96	s1: Methyl acetate (25.91%) s2: Methyl cyclohexane (74.09%)	92	172.47
4 (93)	s1: Isopropyl acetate (25.10%) s2: Methyl cyclohexane (74.90%)	97	56.41	s1: Chloroform (41.08%) s2: 2,2,4-Trimethyl pentane (58.92%)	93	185.97
5 (94)	s1: Nitrobenzene (47.80%) s2: n-Heptane (52.20%)	98	69.12	s1: Chloroform (36.85%) s2: 2,2,4-Trimethyl pentane (63.15%)	94	208.10
6 (95)	s1: Nitrobenzene (21.52%) s2: Methyl cyclohexane (78.48%)	99	89.30	s1: Carbon tetrachloride (34.70%) s2: 2,2,4-Trimethyl pentane (65.30%)	95	843.17

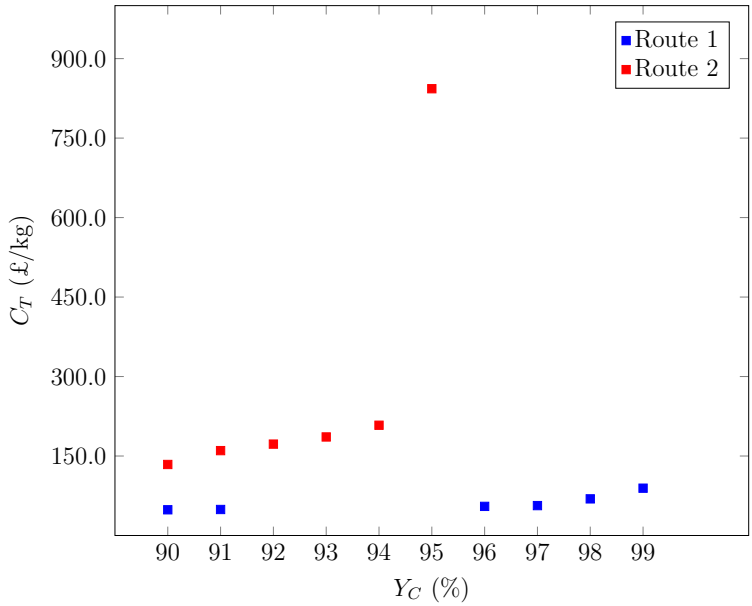

Figure 5.5: Pareto fronts of the BOO problem for minimising C_T and maximising Y_C in Routes 1 and 2.

In terms of computational time, the CPU time (in seconds) needed to generate the solutions of each design problem presented in this chapter is similar to that of the integrated synthesis and crystallisation design problem considered in Chapter 3. This shows that the proposed model-based approach to route selection can be used to quickly assess synthesis routes based on process performance and significantly reduce the number of resource-intensive and time-consuming screening experiments during early-stage product development.

5.4 Conclusion

In this chapter, the solvent selection tool presented in Chapter 3 was used to develop a model-based approach to route design selection which, given a reaction network/multiple routes for the synthesis of a pharmaceutical compound, identifies the route, solvents/solvent mixtures and process conditions that optimise a set of KPIs, including the process E-factor (PEF), process energy consumption (Q_T), solvent cost (C_T) and crystal yield (Y_C).The benefit of this approach is that it allows assessing pharmaceutical synthesis routes from a process perspective, especially since process-related metrics are not usually taken into account when designing routes in early-stage process development. Furthermore, while the proposed modelling approach is thermodynamics-based and does not consider key process aspects such as reaction kinetics,

it allows identifying a shortlist of solvent candidates from a large design space that perform optimally from a species solubility perspective, and which can then be taken forward to the next stages of process development and assessed based on additional process-related criteria.

The approach was illustrated by identifying optimal process routes for the two-step synthesis and crystallisation of 4-nitrophenol (NP). Two routes were considered: Route 1, involving the nitration of chlorobenzene followed by the hydroxylation of 4-nitrochlorobenzene, and Route 2, involving the oxidation of cumene followed by the nitration of phenol. By solving single-objective optimisation formulations of the route selection problem, it was found that Route 1 is superior to Route 2 in terms of the PEF and C_T, whereas Route 2 is superior to Route 1 in terms of Q_T. These findings were in agreement with the results obtained by solving BOO formulations in which both the PEF and Q_T are minimised, and in which C_T is minimised while Y_C is maximised. The CAM^{b}D approach can not only identify the optimal process route based on defined KPI(s), but can also highlight the extent of difference between the process routes of multiple reaction pathways, which offers insights on process-level improvements that can be applied to the sub-optimal routes. In other words, the process metrics that can be achieved based on the idealised representation of the process routes in the proposed design approach are vastly different from route to route, which highlights the importance of assessing these metrics early on in process development. In the NP case study for instance, the best Route 1 design has an approximately 58% lower PEF relative to the best Route 2 design in the PEF minimisation problem, while the best Route 2 design has an approximately 50% lower Q_T relative to the best Route 1 design in the Q_T minimisation problem. The proposed route screening approach can always be extended to include additional shortlisting considerations, such as those related to safety, health and the environment.

Chapter 6

Conclusions and Future Work

6.1 Summary

A process-wide systematic approach based on computer-aided mixture/blend design (CAM^{b}D) for the design of solvents and reaction/separation processes has been proposed, which facilitates the development of fully integrated pharmaceutical processes that incorporate multiple reaction, separation and purification steps, and which can be used to evaluate pharmaceutical synthesis routes based on process performance metrics.

In Chapter 3, a CAM^{b}D-based solvent selection and process design approach was developed for integrated synthesis and crystallisation processes. The solvents or solvent mixtures, mixture compositions and process conditions (e.g., reaction and final crystallisation temperatures) that optimise selected KPIs such as the solvent E-factor (SEF), the process E-factor (PEF), crystal yield (Y_C), process safety (I_T) and process energy consumption (Q_T) are identified. The one-step synthesis of mefenamic acid (MA) from 2,3-dimethylaniline (DMA) and 2-chlorobenzoic acid (CBA) was used as a case study to illustrate the proposed design approach. The flexibility of the method was highlighted by considering different design spaces and optimising different objective functions/KPIs. Furthermore, the integrated nature of the approach allowed the generation of better designs in comparison with the sequential approach to solvent and process design, where reductions of up to 25% in solvent consumption, quantified by the SEF, were observed. Finally, bi-objective optimisation (BOO) formulations of the CAM^{b}D tool were developed to explore the trade-offs between competing KPIs, namely the SEF and I_T, the PEF and I_T, and Q_T and Y_C,

and to identify compromise solutions.

In Chapter 4, the CAM^{b}D approach to solvent and process design was extended to include the modelling of isolation processes, i.e., filtration, washing and drying, by including the corresponding process model and design constraints in the mathematical formulation. Accordingly, the extended approach identifies the solvents or solvent mixtures, mixture compositions and process conditions in end-to-end drug substance manufacturing processes that optimise process-wide KPIs such as the process yield (Y_P), product purity (PP) and process-wide SEF, PEF and Q_T, in addition to Y_C and I_T. The synthesis, crystallisation and isolation of MA, which is an extension of the process described in Chapter 3, was used as a case study. Different design problems were formulated and some important findings were realised. For example, it was found that varying the initial and final crystallisation temperatures provided more high-performance designs to choose from (several solutions with $Y_P \geq 88\%$) than when fixing crystallisation temperatures (only one solution with $Y_P \geq 88\%$), as the latter condition restricts the solvent design space to candidates whose liquid range falls within the bounds of operating temperatures; accordingly, potentially promising solvents may be missed out. The ability to choose from a wider range of design options allows addressing several challenges in process design such as material availability and environmental restrictions. Additionally, by solving a BOO formulation of the design problem in which Q_T is minimised and Y_P is maximised, it was found that the crystallisation energy consumption, Q_C^-, was the greatest contributor to Q_T due to the positive monotonic behaviour between Y_C, which is increased by a larger crystallisation temperature drop or a larger Q_C^-, and Y_P. Furthermore, by solving a multi-objective formulation of the design problem in which the SEF, Y_P and PP are simultaneously optimised, interesting trade-offs were realised; for example, it was found that higher solvent consumption or SEF leads to higher PP, which is expected due to a higher consumption of the wash solvent sw, and generally to higher Y_P due to the relation between the latter KPI and crystal yield (Y_C) which is influenced by solvent and/or anti-solvent use. However, at very high PP values, the excessive consumption of wash solvent may lead to API loss at the washing strange, hence negatively impacting Y_P.

In Chapter 5, the CAM^{b}D approach developed in Chapter 3 was used to evaluate pharmaceutical synthesis routes based on process performance in order to obtain a more comprehensive

assessment beyond lab-scale findings that might become invalid at process scale. The performance metrics included material use (PEF) and product quality (Y_C), energy efficiency (Q_T) and solvent cost (C_T). An important feature of the route selection model is the consideration of impurities generated from previous synthetic steps in the solubility calculations; for example, all unconverted material from the multi-step synthesis of an API must be under-saturated (dissolved) in the final crystallisation step. To illustrate the use of CAM^{b}D in route selection, the two-step synthesis of 4-nitrophenol (NP) via 2 reaction routes was used as a case study. It was found that Route 1, involving the nitration of chlorobenzene followed by the hydroxylation of 4-nitrochlorobenzene, is superior to Route 2, involving the oxidation of cumene followed by the nitration of phenol, in terms of minimising the PEF and C_T, whereas Route 2 is superior to Route 1 in terms of minimising Q_T. These findings suggest that route selection is a multi-objective design problem in which the optimisation of multiple KPIs must be addressed, and that the proposed approach, even with simplified process models, can provide insights on what routes to pursue based on process performance.

6.2 Main Contributions

The main contributions of the thesis are summarised as follows:

- Extending a CAM^{b}D approach for integrated purification design to include synthesis, which is a key step in drug substance manufacturing, in the mathematical formulation by developing the corresponding process model and design constraints. This is the first time that CAM^{b}D has been applied to end-to-end API manufacturing;

- Demonstrating the value of using simplified thermodynamics-based models to describe pharmaceutical process systems within a solvent selection and process design framework;

- Demonstrating the value of integrating process units such as reactors and crystallisers in drug substance manufacturing by comparing the KPIs of CAM^{b}D-generated solutions from both sequential and integrated design formulations;

- Formulating MOO versions of the CAM^{b}D problem to explore the trade-offs between

conflicting performance objectives such as material use (SEF/PEF) and environmental impact (I_T), SEF and product purity (PP), energy consumption (Q_T) and crystal yield (Y_C), Q_T and process yield (Y_P), and crystal yield (Y_C) and solvent cost (C_T);

- Deploying the advanced CAM^{b}D model for process route selection by evaluating different reaction pathways for the synthesis of pharmaceutical compounds based on process performance metrics, including the PEF, Q_T, C_T and Y_C.

6.3 Future Work

6.3.1 Solvent effects on process performance: Kinetics, impurity distribution and crystal morphology

The CAM^{b}D-based solvent selection and process design approach presented in this thesis focused on thermodynamic driving forces, i.e., species solubility, in order to identify potential solvent candidates that can be further assessed with respect to other process-related criteria. The current approach can be extended to consider these criteria. For example, it is important to study solvent effects on reaction kinetics [203] as they can impact reaction conversion and impurity levels, especially since pharmaceutical synthesis typically involves multiple reactions (series + parallel reactions) that generate critical impurities which must be purged by washing or rejected during crystallisation.

After incorporating reaction kinetics in the design methodology, it becomes useful to study solvent effects on impurity distribution during crystallisation [65], as critical genotoxic/mutagenic impurities must be rejected to secure regulatory approval. Models that quantify these effects can be constructed from experimentally-derived or model-based impurity distribution coefficients, and can then be embedded in the solvent selection formulation to identify solvents that promote maximal impurity rejection or minimal impurity partitioning to the solid phase.

An additional important consideration is the effect of solvents on crystal morphology which can greatly impact the feasibility of downstream processing steps such as milling and compaction [102]. Modelling these effects in the CAM^{b}D approach, which currently focuses on upstream

processes, can generate designs that would facilitate the operation of downstream processes; accordingly, an integrated solvent selection tool for drug substance manufacturing, that takes into account design specifications in drug product manufacturing, would be developed.

6.3.2 Effect of ionic species on mixture thermodynamics

The current solvent selection approach uses UNIFAC to estimate activity coefficients and solubilities of organic molecules in solution, but neglects the effect of ionic species on mixture thermodynamics. Since most pharmaceutical syntheses involve inorganic/ionic compounds, advanced thermodynamic models, such as the SAFT-γ Mie equation of state [204], that can account for the presence of such species must be developed and embedded in the CAM^{b}D formulation. Ionic species might constitute critical impurities that can impact important KPIs, such as reaction conversion, crystal yield/morphology and process yield.

6.3.3 Inter-stage workup and solvent recovery

The conceptual flowsheets considered in this thesis can be further advanced in the future to match industrial reality better. For example, in many pharmaceutical processes, liquid-liquid extraction (LLE) or membrane separation units are deployed after the synthesis step in order remove impurities prior to crystallisation. It would be useful to extend the current CAM^{b}D approach to design LLE units, for example, by selecting an extraction solvent that is immiscible with the reactor effluent but in which target impurities have a high solubility, as has been done in the work of Diab et al. [48]. Furthermore, solvent swap is a common unit operation in drug substance manufacturing, which is needed when the solvent used in a certain process unit is not compatible with a subsequent unit, e.g., a reaction solvent may be ineffective as a crystallisation solvent [11]. The modelling of solvent swaps in the CAM^{b}D formulation would be of great interest to pharmaceutical manufacturers, given the resource intensiveness of these operations. In addition, the inclusion of solvent recovery units within the design methodology would be of significant importance to reduce the environmental impact and material/energy costs associated with excessive solvent utilisation in API manufacturing [205].

6.3.4 Multi-step process routes and process synthesis

The CAM^{b}D approach developed for route selection was illustrated on a simple case study involving two-step syntheses of 4-nitrophenol. Pharmaceutical syntheses, however, involve a much larger number of reaction steps; accordingly, it would be useful and straightforward to extend the current methodology to evaluate synthetic routes with multiple steps (>2), and to consider whether each of these steps requires a sequence of workup tasks (LLE, solvent swap, crystallisation, filtration, etc.) by including binary variables in the CAM^{b}D mathematical formulation that denote the existence or non-existence of each of these units based on the design objective. This constitutes a process synthesis approach that allows exploring opportunities for telescoping/process intensification across the process units when it is feasible to do so. Developing optimisation algorithms/solution strategies to address the model complexities arising from such valuable extensions would of course be necessary as well.

6.4 Publications at time of submission

6.4.1 Journal articles

Muhieddine, M.H., Viswanath, S.K., Armstrong, A., Galindo, A. and Adjiman, C.S., Model-based solvent selection for the synthesis and crystallisation of pharmaceutical compounds. *Chemical Engineering Science*, 2022, **264**, 118125.

6.4.2 Refereed conference papers

- Muhieddine, M.H., Jonuzaj, S., Viswanath, S.K., Armstrong, A., Galindo, A. and Adjiman, C.S. Model-based solvent selection for integrated synthesis, crystallisation and isolation processes. *Computer Aided Chemical Engineering*, 2022, **51**, 601-606.
- Muhieddine, M.H., Viswanath, S.K., Armstrong, A., Galindo, A. and Adjiman, C.S., Multi-objective optimisation for early-stage pharmaceutical process development. *Computer Aided Chemical Engineering*, 2022, **49**, 2161-2166.

Bibliography

[1] A. Bhattacharya, S. Akasapu, and R. Bandichhorb, "Green chemistry in drug development," *Scalable Green Chemistry: Case Studies from the Pharmaceutical Industry*, p. 25, 2013.

[2] J. M. Woodley, "Innovative process development and production concepts for small-molecule API manufacturing," in *Computer Aided Chemical Engineering*, vol. 41, pp. 67–84, Elsevier, 2018.

[3] H.-J. Federsel, "En route to full implementation: driving the green chemistry agenda in the pharmaceutical industry," *Green Chemistry*, vol. 15, no. 11, pp. 3105–3115, 2013.

[4] C. S. Slater, M. J. Savelski, W. A. Carole, and D. J. Constable, "Solvent use and waste issues," *Green Chemistry in the Pharmaceutical Industry*, pp. 49–82, 2010.

[5] K. V. Gernaey, A. E. Cervera-Padrell, and J. M. Woodley, "A perspective on PSE in pharmaceutical process development and innovation," *Computers & Chemical Engineering*, vol. 42, pp. 15–29, 2012.

[6] M. M. Pereira and M. J. Calvete, *Sustainable Synthesis of Pharmaceuticals: Using Transition Metal Complexes as Catalysts*, vol. 54. Royal Society of Chemistry, 2018.

[7] P. J. Dunn, "The importance of green chemistry in process research and development," *Chemical Society Reviews*, vol. 41, no. 4, pp. 1452–1461, 2012.

[8] P. J. Dunn, "Pharmaceutical green chemistry process changes–how long does it take to obtain regulatory approval?," *Green Chemistry*, vol. 15, no. 11, pp. 3099–3104, 2013.

[9] C. Jiménez-González, C. S. Ponder, Q. B. Broxterman, and J. B. Manley, "Using the right green yardstick: why process mass intensity is used in the pharmaceutical industry

to drive more sustainable processes," *Organic Process Research & Development*, vol. 15, no. 4, pp. 912–917, 2011.

[10] P. Kolář, J.-W. Shen, A. Tsuboi, and T. Ishikawa, "Solvent selection for pharmaceuticals," *Fluid Phase Equilibria*, vol. 194, pp. 771–782, 2002.

[11] E. Papadakis, A. K. Tula, and R. Gani, "Solvent selection methodology for pharmaceutical processes: Solvent swap," *Chemical Engineering Research and Design*, vol. 115, pp. 443–461, 2016.

[12] E. Siougkrou, A. Galindo, and C. S. Adjiman, "Recent advances in the molecular engineering of solvents for reactions," *Green Chemical Engineering*, 2018.

[13] J. M. Prausnitz, R. N. Lichtenthaler, and E. G. De Azevedo, *Molecular thermodynamics of fluid-phase equilibria*. Pearson Education, 1998.

[14] D. J. Constable, C. Jiménez-González, and R. K. Henderson, "Perspective on solvent use in the pharmaceutical industry," *Organic Process Research & Development*, vol. 11, no. 1, pp. 133–137, 2007.

[15] C. Jiménez-González, A. D. Curzons, D. J. Constable, and V. L. Cunningham, "Expanding GSK's solvent selection guide—application of life cycle assessment to enhance solvent selections," *Clean Technologies and Environmental Policy*, vol. 7, no. 1, pp. 42–50, 2004.

[16] R. A. Sheldon, "The E-factor 25 years on: the rise of green chemistry and sustainability," *Green Chemistry*, vol. 19, no. 1, pp. 18–43, 2017.

[17] D. Ott, D. Kralisch, I. Denčić, V. Hessel, Y. Laribi, P. D. Perrichon, C. Berguerand, L. Kiwi-Minsker, and P. Loeb, "Life cycle analysis within pharmaceutical process optimization and intensification: case study of active pharmaceutical ingredient production," *ChemSusChem*, vol. 7, no. 12, pp. 3521–3533, 2014.

[18] J. Ooi, D. K. Ng, and N. G. Chemmangattuvalappil, "A systematic molecular design framework with the consideration of competing solvent recovery processes," *Industrial & Engineering Chemistry Research*, vol. 58, no. 29, pp. 13210–13226, 2019.

[19] S. A. Diab, H. G. Jolliffe, and D. I. Gerogiorgis, "Plantwide technoeconomic analysis and separation solvent selection for continuous pharmaceutical manufacturing: Ibuprofen, artemisinin, and diphenhydramine," in *Computer Aided Chemical Engineering*, vol. 41, pp. 85–120, Elsevier, 2018.

[20] E. Papadakis, J. M. Woodley, and R. Gani, "Perspective on PSE in pharmaceutical process development and innovation," in *Computer Aided Chemical Engineering*, vol. 41, pp. 597–656, Elsevier, 2018.

[21] K. Alfonsi, J. Colberg, P. J. Dunn, T. Fevig, S. Jennings, T. A. Johnson, H. P. Kleine, C. Knight, M. A. Nagy, D. A. Perry, *et al.*, "Green chemistry tools to influence a medicinal chemistry and research chemistry based organisation," *Green Chemistry*, vol. 10, no. 1, pp. 31–36, 2008.

[22] D. Prat, O. Pardigon, H.-W. Flemming, S. Letestu, V. Ducandas, P. Isnard, E. Guntrum, T. Senac, S. Ruisseau, P. Cruciani, *et al.*, "Sanofi's solvent selection guide: A step toward more sustainable processes," *Organic Process Research & Development*, vol. 17, no. 12, pp. 1517–1525, 2013.

[23] R. K. Henderson, C. Jiménez-González, D. J. Constable, S. R. Alston, G. G. Inglis, G. Fisher, J. Sherwood, S. P. Binks, and A. D. Curzons, "Expanding GSK's solvent selection guide–embedding sustainability into solvent selection starting at medicinal chemistry," *Green Chemistry*, vol. 13, no. 4, pp. 854–862, 2011.

[24] C. M. Alder, J. D. Hayler, R. K. Henderson, A. M. Redman, L. Shukla, L. E. Shuster, and H. F. Sneddon, "Updating and further expanding GSK's solvent sustainability guide," *Green Chemistry*, vol. 18, no. 13, pp. 3879–3890, 2016.

[25] D. Prat, A. Wells, J. Hayler, H. Sneddon, C. R. McElroy, S. Abou-Shehada, and P. J. Dunn, "CHEM21 selection guide of classical-and less classical-solvents," *Green Chemistry*, vol. 18, no. 1, pp. 288–296, 2016.

[26] M. Folić, C. S. Adjiman, and E. N. Pistikopoulos, "Design of solvents for optimal reaction rate constants," *AIChE journal*, vol. 53, no. 5, pp. 1240–1256, 2007.

[27] C. S. Adjiman, A. J. Clarke, G. Cooper, and P. C. Taylor, "Solvents for ring-closing metathesis reactions," *Chemical Communications*, no. 24, pp. 2806–2808, 2008.

[28] R. Gani and E. Brignole, "Molecular design of solvents for liquid extraction based on UNIFAC," *Fluid Phase Equilibria*, vol. 13, pp. 331–340, 1983.

[29] E. A. Brignole, S. Bottini, and R. Gani, "A strategy for the design and selection of solvents for separation processes.," *Fluid Phase Equilibria*, vol. 29, pp. 125–132, 1986.

[30] A. T. Karunanithi, L. E. Achenie, and R. Gani, "A new decomposition-based computer-aided molecular/mixture design methodology for the design of optimal solvents and solvent mixtures," *Industrial & Engineering Chemistry Research*, vol. 44, no. 13, pp. 4785–4797, 2005.

[31] A. T. Karunanithi, L. E. Achenie, and R. Gani, "A computer-aided molecular design framework for crystallization solvent design," *Chemical Engineering Science*, vol. 61, no. 4, pp. 1247–1260, 2006.

[32] R. Gani, "Chemical product design: challenges and opportunities," *Computers & Chemical Engineering*, vol. 28, no. 12, pp. 2441–2457, 2004.

[33] R. Gani, "Computer-aided methods and tools for chemical product design," *Chemical Engineering Research and Design*, vol. 82, no. 11, pp. 1494–1504, 2004.

[34] S. Jonuzaj, P. T. Akula, P.-M. Kleniati, and C. S. Adjiman, "The formulation of optimal mixtures with generalized disjunctive programming: A solvent design case study," *AIChE Journal*, vol. 62, no. 5, pp. 1616–1633, 2016.

[35] R. A. Granberg and Å. C. Rasmuson, "Solubility of paracetamol in binary and ternary mixtures of water+ acetone+ toluene," *Journal of Chemical & Engineering Data*, vol. 45, no. 3, pp. 478–483, 2000.

[36] T. Zhou, K. McBride, S. Linke, Z. Song, and K. Sundmacher, "Computer-aided solvent selection and design for efficient chemical processes," *Current Opinion in Chemical Engineering*, vol. 27, pp. 35–44, 2020.

[37] I. Mitrofanov, S. Sansonetti, J. Abildskov, G. Sin, and R. Gani, "The solvent selection framework: Solvents for organic synthesis, separation processes and ionic liquids solvents," in *Computer Aided Chemical Engineering*, vol. 30, pp. 762–766, Elsevier, 2012.

[38] L. J. Diorazio, D. R. Hose, and N. K. Adlington, "Toward a more holistic framework for solvent selection," *Organic Process Research & Development*, vol. 20, no. 4, pp. 760–773, 2016.

[39] J. L. Tucker, "Green chemistry, a pharmaceutical perspective," *Organic Process Research & Development*, vol. 10, no. 2, pp. 315–319, 2006.

[40] C. A. Nicolaou and N. Brown, "Multi-objective optimization methods in drug design," *Drug Discovery Today: Technologies*, vol. 10, no. 3, pp. e427–e435, 2013.

[41] I. E. Grossmann and I. Harjunkoski, "Process systems engineering: Academic and industrial perspectives," *Computers & Chemical Engineering*, vol. 126, pp. 474–484, 2019.

[42] B. R. Bakshi, "Toward sustainable chemical engineering: the role of process systems engineering," *Annual review of chemical and biomolecular engineering*, vol. 10, pp. 265–288, 2019.

[43] B. Benyahia, R. Lakerveld, and P. I. Barton, "A plant-wide dynamic model of a continuous pharmaceutical process," *Industrial & Engineering Chemistry Research*, vol. 51, no. 47, pp. 15393–15412, 2012.

[44] P. L. Heider, S. C. Born, S. Basak, B. Benyahia, R. Lakerveld, H. Zhang, R. Hogan, L. Buchbinder, A. Wolfe, S. Mascia, *et al.*, "Development of a multi-step synthesis and workup sequence for an integrated, continuous manufacturing process of a pharmaceutical," *Organic Process Research & Development*, vol. 18, no. 3, pp. 402–409, 2014.

[45] H. G. Jolliffe and D. I. Gerogiorgis, "Process modelling and simulation for continuous pharmaceutical manufacturing of ibuprofen," *Chemical engineering research and design*, vol. 97, pp. 175–191, 2015.

[46] H. G. Jolliffe and D. I. Gerogiorgis, "Process modelling and simulation for continuous pharmaceutical manufacturing of artemisinin," *Chemical Engineering Research and Design*, vol. 112, pp. 310–325, 2016.

[47] S. Diab and D. I. Gerogiorgis, "Process modeling, simulation, and technoeconomic evaluation of separation solvents for the continuous pharmaceutical manufacturing (CPM) of diphenhydramine," *Organic Process Research & Development*, vol. 21, no. 7, pp. 924–946, 2017.

[48] S. Diab, D. T. McQuade, B. F. Gupton, and D. I. Gerogiorgis, "Process design and optimization for the continuous manufacturing of nevirapine, an active pharmaceutical ingredient for HIV treatment," *Organic Process Research & Development*, vol. 23, no. 3, pp. 320–333, 2019.

[49] S. Diab and D. I. Gerogiorgis, "Technoeconomic mixed integer nonlinear programming (MINLP) optimization for design of liquid-liquid extraction (LLE) cascades in continuous pharmaceutical manufacturing of atropine," *AIChE Journal*, vol. 65, no. 11, p. e16738, 2019.

[50] A. Fredenslund, R. L. Jones, and J. M. Prausnitz, "Group-contribution estimation of activity coefficients in nonideal liquid mixtures," *AIChE Journal*, vol. 21, no. 6, pp. 1086–1099, 1975.

[51] H. Renon and J. M. Prausnitz, "Local compositions in thermodynamic excess functions for liquid mixtures," *AIChE journal*, vol. 14, no. 1, pp. 135–144, 1968.

[52] C. Jiménez-González, P. Poechlauer, Q. B. Broxterman, B.-S. Yang, D. Am Ende, J. Baird, C. Bertsch, R. E. Hannah, P. Dell'Orco, H. Noorman, *et al.*, "Key green engineering research areas for sustainable manufacturing: a perspective from pharmaceutical and fine chemicals manufacturers," *Organic Process Research & Development*, vol. 15, no. 4, pp. 900–911, 2011.

[53] R. Gani, "Chemical product design: challenges and opportunities," *Computers & Chemical Engineering*, vol. 28, no. 12, p. 2441–2457, 2004.

[54] J. Wang, L. Zhu, and R. Lakerveld, "A hybrid framework for simultaneous process and solvent optimization of continuous anti-solvent crystallization with distillation for solvent recycling," *Processes*, vol. 8, no. 1, p. 63, 2020.

[55] S. Jonuzaj, A. Gupta, and C. S. Adjiman, "The design of optimal mixtures from atom groups using generalized disjunctive programming," *Computers & Chemical Engineering*, vol. 116, pp. 401–421, 2018.

[56] O. L. Watson, A. Galindo, G. Jackson, and C. S. Adjiman, "Computer-aided design of solvent blends for the cooling and anti-solvent crystallisation of ibuprofen," in *Computer Aided Chemical Engineering*, vol. 46, pp. 949–954, Elsevier, 2019.

[57] O. L. Watson, S. Jonuzaj, J. McGinty, J. Sefcik, A. Galindo, G. Jackson, and C. S. Adjiman, "Computer aided design of solvent blends for hybrid cooling and antisolvent crystallization of active pharmaceutical ingredients," *Organic Process Research & Development*, vol. 25, no. 5, pp. 1123–1142, 2021.

[58] S. Jonuzaj, O. L. Watson, S. Ottoboni, C. J. Price, J. Sefcik, A. Galindo, G. Jackson, and C. S. Adjiman, "Computer-aided solvent mixture design for the crystallisation and isolation of mefenamic acid," in *Computer Aided Chemical Engineering*, vol. 48, pp. 649–654, Elsevier, 2020.

[59] C. S. Adjiman, A. Galindo, and G. Jackson, "Molecules matter: the expanding envelope of process design," in *Computer Aided Chemical Engineering*, vol. 34, pp. 55–64, Elsevier, 2014.

[60] N. P. Mendis, J. Wang, and R. Lakerveld, "A thermodynamic approach for simultaneous solvent and process design of continuous reactive crystallization with recycling," in *Computer Aided Chemical Engineering*, vol. 48, pp. 805–810, Elsevier, 2020.

[61] N. P. Mendis, J. Wang, and R. Lakerveld, "Simultaneous solvent selection and process design for continuous reaction–extraction–crystallization systems," *Industrial & Engineering Chemistry Research*, 2022.

[62] M. Folić, R. Gani, C. Jiménez-González, and D. J. Constable, "Systematic selection of green solvents for organic reacting systems," *Chinese Journal of Chemical Engineering*, vol. 16, no. 3, pp. 376–383, 2008.

[63] A. E. Cervera-Padrell, T. Skovby, S. Kiil, R. Gani, and K. V. Gernaey, "Active pharmaceutical ingredient (API) production involving continuous processes–a process system engineering (PSE)-assisted design framework," *European Journal of Pharmaceutics and Biopharmaceutics*, vol. 82, no. 2, pp. 437–456, 2012.

[64] S. Ottoboni, B. Wareham, A. Vassileiou, M. Robertson, C. J. Brown, B. Johnston, and C. J. Price, "A novel integrated workflow for isolation solvent selection using prediction and modeling," *Organic Process Research & Development*, vol. 25, no. 5, pp. 1143–1159, 2021.

[65] J. Li, T.-t. C. Lai, B. L. Trout, and A. S. Myerson, "Continuous crystallization of cyclosporine: Effect of operating conditions on yield and purity," *Crystal Growth & Design*, vol. 17, no. 3, pp. 1000–1007, 2017.

[66] S. Diab and D. I. Gerogiorgis, "Technoeconomic evaluation of multiple mixed suspension-mixed product removal (MSMPR) crystallizer configurations for continuous cyclosporine crystallization," *Organic Process Research & Development*, vol. 21, no. 10, pp. 1571–1587, 2017.

[67] S. Diab and D. I. Gerogiorgis, "Technoeconomic optimization of continuous crystallization for three active pharmaceutical ingredients: cyclosporine, paracetamol, and aliskiren," *Industrial & Engineering Chemistry Research*, vol. 57, no. 29, pp. 9489–9499, 2018.

[68] J. Liu and B. Benyahia, "Systematic model-based dynamic optimization of a combined cooling and antisolvent multistage continuous crystallization process," in *Computer Aided Chemical Engineering*, vol. 50, pp. 1221–1227, Elsevier, 2021.

[69] S. Mascia, P. L. Heider, H. Zhang, R. Lakerveld, B. Benyahia, P. I. Barton, R. D. Braatz, C. L. Cooney, J. M. Evans, T. F. Jamison, *et al.*, "End-to-end continuous manufacturing of pharmaceuticals: integrated synthesis, purification, and final dosage

formation," *Angewandte Chemie International Edition*, vol. 52, no. 47, pp. 12359–12363, 2013.

[70] B. Benyahia, "Applications of a plant-wide dynamic model of an integrated continuous pharmaceutical plant: design of the recycle in the case of multiple impurities," in *Computer Aided Chemical Engineering*, vol. 41, pp. 141–157, Elsevier, 2018.

[71] C. J. Brown, T. McGlone, S. Yerdelen, V. Srirambhatla, F. Mabbott, R. Gurung, M. L. Briuglia, B. Ahmed, H. Polyzois, J. McGinty, *et al.*, "Enabling precision manufacturing of active pharmaceutical ingredients: workflow for seeded cooling continuous crystallisations," *Molecular Systems Design & Engineering*, vol. 3, no. 3, pp. 518–549, 2018.

[72] R. Lakerveld, B. Benyahia, R. D. Braatz, and P. I. Barton, "Model-based design of a plant-wide control strategy for a continuous pharmaceutical plant," *AIChE Journal*, vol. 59, no. 10, pp. 3671–3685, 2013.

[73] R. Lakerveld, B. Benyahia, P. L. Heider, H. Zhang, A. Wolfe, C. J. Testa, S. Ogden, D. R. Hersey, S. Mascia, J. M. Evans, *et al.*, "The application of an automated control strategy for an integrated continuous pharmaceutical pilot plant," *Organic Process Research & Development*, vol. 19, no. 9, pp. 1088–1100, 2015.

[74] A. Mesbah, J. A. Paulson, R. Lakerveld, and R. D. Braatz, "Model predictive control of an integrated continuous pharmaceutical manufacturing pilot plant," *Organic Process Research & Development*, vol. 21, no. 6, pp. 844–854, 2017.

[75] H. G. Jolliffe and D. I. Gerogiorgis, "Plantwide design and economic evaluation of two continuous pharmaceutical manufacturing (CPM) cases: Ibuprofen and artemisinin," *Computers & Chemical Engineering*, vol. 91, pp. 269–288, 2016.

[76] S. Diab and D. I. Gerogiorgis, "Process modelling, simulation and technoeconomic evaluation of crystallisation antisolvents for the continuous pharmaceutical manufacturing of rufinamide," *Computers & Chemical Engineering*, vol. 111, pp. 102–114, 2018.

[77] J. Liu and B. Benyahia, "Techno-economic evaluation and optimization of batch, fed-batch and multistage continuous crystallization processes," *Chemistry Proceedings*, vol. 9, no. 1, p. 6, 2022.

[78] C. S. Adjiman, N. V. Sahinidis, D. G. Vlachos, B. Bakshi, C. T. Maravelias, and C. Georgakis, "Process systems engineering perspective on the design of materials and molecules," *Industrial & Engineering Chemistry Research*, vol. 60, no. 14, pp. 5194–5206, 2021.

[79] R. Gani, "CAMD: Computer aided molecular design–examples of applications," *Technical Report CAPEC, Department of Chemical Engineering, Technical University of Denmark*, 2004.

[80] C. D. Maranas, "Optimal computer-aided molecular design: A polymer design case study," *Industrial & Engineering Chemistry Research*, vol. 35, no. 10, pp. 3403–3414, 1996.

[81] A. Giovanoglou, J. Barlatier, C. S. Adjiman, E. N. Pistikopoulos, and J. L. Cordiner, "Optimal solvent design for batch separation based on economic performance," *AIChE Journal*, vol. 49, no. 12, pp. 3095–3109, 2003.

[82] Q. Liu, L. Zhang, L. Liu, J. Du, Q. Meng, and R. Gani, "Computer-aided reaction solvent design based on transition state theory and COSMO-SAC," *Chemical Engineering Science*, vol. 202, pp. 300–317, 2019.

[83] A. Papadopoulos, P. Linke, and P. Seferlis, "Integrated multiobjective molecular and process design: Operational and computational frontiers," in *Computer Aided Chemical Engineering*, vol. 39, pp. 269–313, Elsevier, 2016.

[84] E. Siougkrou, A. Galindo, and C. S. Adjiman, "On the optimal design of gas-expanded liquids based on process performance," *Chemical Engineering Science*, vol. 115, pp. 19–30, 2014.

[85] N. D. Austin, N. V. Sahinidis, and D. W. Trahan, "A COSMO-based approach to computer-aided mixture design," *Chemical Engineering Science*, vol. 159, pp. 93–105, 2017.

[86] H. Struebing, Z. Ganase, P. G. Karamertzanis, E. Siougkrou, P. Haycock, P. M. Piccione, A. Armstrong, A. Galindo, and C. S. Adjiman, "Computer-aided molecular design of solvents for accelerated reaction kinetics," *Nature Chemistry*, vol. 5, no. 11, pp. 952–957, 2013.

[87] H. Struebing, S. Obermeier, E. Siougkrou, C. S. Adjiman, and A. Galindo, "A QM-CAMD approach to solvent design for optimal reaction rates," *Chemical Engineering Science*, vol. 159, pp. 69–83, 2017.

[88] S. Jonuzaj and C. S. Adjiman, "Designing optimal mixtures using generalized disjunctive programming: Hull relaxations," *Chemical Engineering Science*, vol. 159, pp. 106–130, 2017.

[89] S. Gopinath, G. Jackson, A. Galindo, and C. S. Adjiman, "Outer approximation algorithm with physical domain reduction for computer-aided molecular and separation process design," *AIChE Journal*, vol. 62, no. 9, pp. 3484–3504, 2016.

[90] A. Duvedi and L. E. Achenie, "On the design of environmentally benign refrigerant mixtures: a mathematical programming approach," *Computers & Chemical Engineering*, vol. 21, no. 8, pp. 915–923, 1997.

[91] M. Lampe, M. Stavrou, H. Bucker, J. Gross, and A. Bardow, "Simultaneous optimization of working fluid and process for organic rankine cycles using PC-SAFT," *Industrial & Engineering Chemistry Research*, vol. 53, no. 21, pp. 8821–8830, 2014.

[92] R. Vaidyanathan and M. El-Halwagi, "Computer-aided synthesis of polymers and blends with target properties," *Industrial & Engineering Chemistry Research*, vol. 35, no. 2, pp. 627–634, 1996.

[93] N. D. Austin, N. V. Sahinidis, and D. W. Trahan, "Computer-aided molecular design: An introduction and review of tools, applications, and solution techniques," *Chemical Engineering Research and Design*, vol. 116, pp. 2–26, 2016.

[94] L. Y. Ng, F. K. Chong, and N. G. Chemmangattuvalappil, "Challenges and opportunities in computer-aided molecular design," *Computers & Chemical Engineering*, vol. 81, pp. 115–129, 2015.

[95] N. G. Chemmangattuvalappil, "Development of solvent design methodologies using computer-aided molecular design tools," *Current Opinion in Chemical Engineering*, vol. 27, pp. 51–59, 2020.

[96] L. Achenie, V. Venkatasubramanian, and R. Gani, *Computer aided molecular design: theory and practice*. Elsevier, 2002.

[97] H. Foth and A. W. Hayes, "Background of REACH in EU regulations on evaluation of chemicals," *Human & Experimental Toxicology*, vol. 27, no. 6, pp. 443–461, 2008.

[98] K. Grodowska and A. Parczewski, "Organic solvents in the pharmaceutical industry," *Acta Poloniae Pharmaceutica. Drug Research*, vol. 67, no. 1, 2010.

[99] S. Perez-Vega, E. Ortega-Rivas, I. Salmeron-Ochoa, and P. Sharratt, "A system view of solvent selection in the pharmaceutical industry: towards a sustainable choice," *Environment, Development and Sustainability*, vol. 15, no. 1, pp. 1–21, 2013.

[100] P. Harper, M. Hostrup, and R. Gani, "A hybrid CAMD method," *Computer Aided Molecular Design: Theory and Practice*, vol. 12, pp. 139–165, 2003.

[101] N. D. Austin, A. P. Samudra, N. V. Sahinidis, and D. W. Trahan, "Mixture design using derivative-free optimization in the space of individual component properties," *AIChE Journal*, vol. 62, no. 5, pp. 1514–1530, 2016.

[102] J. Chen and B. L. Trout, "Computer-aided solvent selection for improving the morphology of needle-like crystals: A case study of 2,6-dihydroxybenzoic acid," *Crystal Growth & Design*, vol. 10, no. 10, pp. 4379–4388, 2010.

[103] R. Raman and I. E. Grossmann, "Modelling and computational techniques for logic based integer programming," *Computers & Chemical Engineering*, vol. 18, no. 7, pp. 563–578, 1994.

[104] A. Garrod, P. Evans, and C. Davy, "Risk management measures for chemicals: the "COSHH essentials" approach," *Journal of Exposure Science & Environmental Epidemiology*, vol. 17, no. 1, pp. S48–S54, 2007.

[105] T. Zhou, K. McBride, X. Zhang, Z. Qi, and K. Sundmacher, "Integrated solvent and process design exemplified for a Diels–Alder reaction," *AIChE Journal*, vol. 61, no. 1, pp. 147–158, 2015.

[106] A. Klamt and G. Schüürmann, "COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient," *Journal of the Chemical Society, Perkin Transactions 2*, no. 5, pp. 799–805, 1993.

[107] K. McBride, T. Gaide, A. Vorholt, A. Behr, and K. Sundmacher, "Thermomorphic solvent selection for homogeneous catalyst recovery based on COSMO-RS," *Chemical Engineering and Processing: Process Intensification*, vol. 99, pp. 97–106, 2016.

[108] A. Klamt, "Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena," *The Journal of Physical Chemistry*, vol. 99, no. 7, pp. 2224–2235, 1995.

[109] S. Linke, K. McBride, and K. Sundmacher, "Systematic green solvent selection for the hydroformylation of long-chain alkenes," *ACS Sustainable Chemistry & Engineering*, vol. 8, no. 29, pp. 10795–10811, 2020.

[110] J. Wang and R. Lakerveld, "Integrated solvent and process design for continuous crystallization and solvent recycling using PC-SAFT," *AIChE Journal*, vol. 64, no. 4, pp. 1205–1216, 2018.

[111] J. Gross and G. Sadowski, "Application of perturbation theory to a hard-chain reference fluid: an equation of state for square-well chains," *Fluid Phase Equilibria*, vol. 168, no. 2, pp. 183–199, 2000.

[112] J. Gross and G. Sadowski, "Perturbed-chain SAFT: An equation of state based on a perturbation theory for chain molecules," *Industrial & Engineering Chemistry Research*, vol. 40, no. 4, pp. 1244–1260, 2001.

[113] A. Bardow, K. Steur, and J. Gross, "Continuous-molecular targeting for integrated solvent and process design," *Industrial & Engineering Chemistry Research*, vol. 49, no. 6, pp. 2834–2840, 2010.

[114] M. Lampe, M. Stavrou, J. Schilling, E. Sauer, J. Gross, and A. Bardow, "Computer-aided molecular design in the continuous-molecular targeting framework using group-contribution PC-SAFT," *Computers & Chemical Engineering*, vol. 81, pp. 278–287, 2015.

[115] V. Papaioannou, T. Lafitte, C. Avendaño, C. S. Adjiman, G. Jackson, E. A. Müller, and A. Galindo, "Group contribution methodology based on the statistical associating fluid theory for heteronuclear molecules formed from Mie segments," *The Journal of Chemical Physics*, vol. 140, no. 5, p. 054107, 2014.

[116] S. Chai, Q. Liu, X. Liang, Y. Guo, S. Zhang, C. Xu, J. Du, Z. Yuan, L. Zhang, and R. Gani, "A grand product design model for crystallization solvent design," *Computers & Chemical Engineering*, vol. 135, p. 106764, 2020.

[117] K. Y. Fung, K. M. Ng, L. Zhang, and R. Gani, "A grand model for chemical product design," *Computers & Chemical Engineering*, vol. 91, pp. 15–27, 2016. 12th International Symposium on Process Systems Engineering & 25th European Symposium of Computer Aided Process Engineering (PSE-2015/ESCAPE-25), 31 May - 4 June 2015, Copenhagen, Denmark.

[118] L. Constantinou and R. Gani, "New group contribution method for estimating properties of pure compounds," *AIChE Journal*, vol. 40, no. 10, pp. 1697–1710, 1994.

[119] L. Constantinou, R. Gani, and J. P. O'Connell, "Estimation of the acentric factor and the liquid molar volume at 298 K using a new group contribution method," *Fluid Phase Equilibria*, vol. 103, no. 1, pp. 11–22, 1995.

[120] T. Sheldon, C. Adjiman, and J. Cordiner, "Pure component properties from group contribution: Hydrogen-bond basicity, hydrogen-bond acidity, hildebrand solubility parameter, macroscopic surface tension, dipole moment, refractive index and dielectric constant," *Fluid Phase Equilibria*, vol. 231, no. 1, pp. 27–37, 2005.

[121] J. Marrero and R. Gani, "Group-contribution based estimation of pure component properties," *Fluid Phase Equilibria*, vol. 183, pp. 183–208, 2001.

[122] J. Marrero and R. Gani, "Group-contribution-based estimation of octanol/water partition coefficient and aqueous solubility," *Industrial & Engineering Chemistry Research*, vol. 41, no. 25, pp. 6623–6633, 2002.

[123] A. Apostolakou and C. Adjiman, "Optimization methods in CAMD–II," in *Computer Aided Chemical Engineering*, vol. 12, pp. 63–93, Elsevier, 2003.

[124] B. E. Poling, J. M. Prausnitz, and J. P. O'connell, *Properties of gases and liquids*. McGraw-Hill Education, 2001.

[125] A. Klamt, V. Jonas, T. Bürger, and J. C. Lohrenz, "Refinement and parametrization of COSMO-RS," *The Journal of Physical Chemistry A*, vol. 102, no. 26, pp. 5074–5085, 1998.

[126] S.-T. Lin and S. I. Sandler, "A priori phase equilibrium prediction from a segment contribution solvation model," *Industrial & Engineering Chemistry Research*, vol. 41, no. 5, pp. 899–913, 2002.

[127] V. Papaioannou, C. Adjiman, G. Jackson, and A. Galindo, "Group contribution methodologies for the prediction of thermodynamic properties and phase behavior in mixtures," *Process Systems Engineering: Volume 6: Molecular Systems Engineering*, vol. 6, pp. 135–172, 2011.

[128] R. Smith, *Chemical process: design and integration*. John Wiley & Sons, 2005.

[129] K. E. Bett, J. S. Rowlinson, and G. Saville, *Thermodynamics for chemical engineers*. Mit Press, 1975.

[130] B. Bouillot, S. Teychené, and B. Biscans, "An evaluation of thermodynamic models for the prediction of drug and drug-like molecule solubility in organic solvents," *Fluid Phase Equilibria*, vol. 309, no. 1, pp. 36–52, 2011.

[131] K. Moodley, J. Rarey, and D. Ramjugernath, "Model evaluation for the prediction of solubility of active pharmaceutical ingredients (APIs) to guide solid–liquid separator design," *Asian Journal of Pharmaceutical Sciences*, vol. 13, no. 3, pp. 265–278, 2018.

[132] U. Weidlich and J. Gmehling, "A modified UNIFAC model. 1. Prediction of VLE, hE, and γ^{∞}," *Industrial & Engineering Chemistry Research*, vol. 26, no. 7, pp. 1372–1381, 1987.

[133] C.-C. Chen and Y. Song, "Solubility modeling with a nonrandom two-liquid segment activity coefficient model," *Industrial & Engineering Chemistry Research*, vol. 43, no. 26, pp. 8354–8362, 2004.

[134] A. Fredenslund and P. Rasmussen, "From UNIFAC to SUPERFAC-and back?," *Fluid Phase Equilibria*, vol. 24, no. 1-2, pp. 115–150, 1985.

[135] T. Magnussen, P. Rasmussen, and A. Fredenslund, "UNIFAC parameter table for prediction of liquid-liquid equilibriums," *Industrial & Engineering Chemistry Process Design and Development*, vol. 20, no. 2, pp. 331–339, 1981.

[136] N. G. Anderson, *Practical process research and development: a guide for organic chemists*. Academic Press, 2012.

[137] R. B. Leng, M. V. Emonds, C. T. Hamilton, and J. W. Ringer, "Holistic route selection," *Organic Process Research & Development*, vol. 16, no. 3, pp. 415–424, 2012.

[138] M. Butters, "Route design and selection," in *Pharmaceutical Process Development*, pp. 90–116, 2011.

[139] J. Serna, E. N. D. Martinez, P. C. N. Rincón, M. Camargo, D. Gálvez, *et al.*, "Multi-criteria decision analysis for the selection of sustainable chemical process routes during early design stages," *Chemical Engineering Research and Design*, vol. 113, pp. 28–49, 2016.

[140] A. Ishizaka and P. Nemery, *Multi-criteria decision analysis: methods and software*. John Wiley & Sons, 2013.

[141] W.-K. Tan, Y.-D. Yeh, S.-J. Chen, Y.-C. Lin, and C.-Y. Kuo, "Using DEMATEL and the smartphone as a case study to investigate how consumers evaluate many features of a product collectively," *International Journal of Applied Mathematics and Informatics*, vol. 6, no. 3, pp. 117–125, 2012.

[142] P.-M. Jacob, P. Yamin, C. Perez-Storey, M. Hopgood, and A. Lapkin, "Towards automation of chemical process route selection based on data mining," *Green Chemistry*, vol. 19, no. 1, pp. 140–152, 2017.

[143] M. Behzadian, R. B. Kazemzadeh, A. Albadvi, and M. Aghdasi, "PROMETHEE: A comprehensive literature review on methodologies and applications," *European Journal of Operational Research*, vol. 200, no. 1, pp. 198–215, 2010.

[144] A. Lapkin, P. Heer, P.-M. Jacob, M. Hutchby, W. Cunningham, S. Bull, and M. Davidson, "Automation of route identification and optimisation based on data-mining and chemical intuition," *Faraday Discussions*, vol. 202, pp. 483–496, 2017.

[145] A. Voll and W. Marquardt, "Reaction network flux analysis: Optimization-based evaluation of reaction pathways for biorenewables processing," *AIChE Journal*, vol. 58, no. 6, pp. 1788–1801, 2012.

[146] M. Hechinger, A. Voll, and W. Marquardt, "Towards an integrated design of biofuels and their production pathways," *Computers & Chemical Engineering*, vol. 34, no. 12, pp. 1909–1918, 2010.

[147] K. Ulonska, A. Voll, and W. Marquardt, "Screening pathways for the production of next generation biofuels," *Energy & Fuels*, vol. 30, no. 1, pp. 445–456, 2016.

[148] A. Konig, K. Ulonska, A. Mitsos, and J. Viell, "Optimal applications and combinations of renewable fuel production from biomass and electricity," *Energy & fuels*, vol. 33, no. 2, pp. 1659–1672, 2019.

[149] D. Zhang, E. A. del Rio-Chanona, and N. Shah, "Screening synthesis pathways for biomass-derived sustainable polymer production," *ACS Sustainable Chemistry & Engineering*, vol. 5, no. 5, pp. 4388–4398, 2017.

[150] T. R. Savage and D. Zhang, "Superstructure reaction network design for the synthesis of biobased sustainable nitrogen-containing polymers," *Industrial & Engineering Chemistry Research*, vol. 59, no. 10, pp. 4688–4697, 2020.

[151] K. Ulonska, M. Skiborowski, A. Mitsos, and J. Viell, "Early-stage evaluation of biorefinery processing pathways using process network flux analysis," *AIChE Journal*, vol. 62, no. 9, pp. 3096–3108, 2016.

[152] A. Konig, W. Marquardt, A. Mitsos, J. Viell, and M. Dahmen, "Integrated design of renewable fuels and their production processes: recent advances and challenges," *Current Opinion in Chemical Engineering*, vol. 27, pp. 45–50, 2020.

[153] K. Ulonska, A. Konig, M. Klatt, A. Mitsos, and J. Viell, "Optimization of multiproduct biorefinery processes under consideration of biomass supply chain management and market developments," *Industrial & Engineering Chemistry Research*, vol. 57, no. 20, pp. 6980–6991, 2018.

[154] C. W. Coley, W. H. Green, and K. F. Jensen, "Machine learning in computer-aided synthesis planning," *Accounts of Chemical Research*, vol. 51, no. 5, pp. 1281–1289, 2018.

[155] R. Joncour, N. Duguet, E. Métay, A. Ferreira, and M. Lemaire, "Amidation of phenol derivatives: a direct synthesis of paracetamol (acetaminophen) from hydroquinone," *Green Chemistry*, vol. 16, no. 6, pp. 2997–3002, 2014.

[156] V. B. Gantovnik, C. M. Anderson-Cook, Z. Gürdal, and L. T. Watson, "A genetic algorithm with memory for mixed discrete–continuous design optimization," *Computers & Structures*, vol. 81, no. 20, 2003.

[157] I. E. Grossmann and F. Trespalacios, "Systematic modeling of discrete-continuous optimization models through generalized disjunctive programming," *AIChE Journal*, vol. 59, no. 9, pp. 3276–3295, 2013.

[158] J. P. Ruiz, J.-H. Jagla, I. E. Grossmann, A. Meeraus, and A. Vecchietti, "Generalized disjunctive programming: Solution strategies," in *Algebraic Modeling Systems*, pp. 57–75, Springer, 2012.

[159] L. A. Wolsey and G. L. Nemhauser, *Integer and combinatorial optimization*, vol. 55. John Wiley & Sons, 1999.

[160] R. Raman and I. E. Grossmann, "Relation between milp modelling and logical inference for chemical process synthesis," *Computers & Chemical Engineering*, vol. 15, no. 2, pp. 73–84, 1991.

[161] I. E. Grossmann, "Review of nonlinear mixed-integer and disjunctive programming techniques," *Optimization and Engineering*, vol. 3, no. 3, pp. 227–252, 2002.

[162] O. K. Gupta and A. Ravindran, "Branch and bound experiments in convex nonlinear integer programming," *Management Science*, vol. 31, no. 12, pp. 1533–1546, 1985.

[163] M. Bussieck and A. Drud, "SBB: Simple branch and bound algorithm for mixed integer nonlinear programming," *URL: https://www. gams. com/latest/docs/solvers/sbb*, 2001.

[164] A. M. Geoffrion, "Generalized benders decomposition," *Journal of Optimization Theory and Applications*, vol. 10, no. 4, pp. 237–260, 1972.

[165] M. A. Duran and I. E. Grossmann, "An outer-approximation algorithm for a class of mixed-integer nonlinear programs," *Mathematical Programming*, vol. 36, no. 3, pp. 307–339, 1986.

[166] R. Fletcher and S. Leyffer, "Solving mixed integer nonlinear programs by outer approximation," *Mathematical Programming*, vol. 66, no. 1, pp. 327–349, 1994.

[167] X. Yuan, S. Zhang, L. Pibouleau, and S. Domenech, "Une méthode d'optimisation non linéaire en variables mixtes pour la conception de procédés," *RAIRO-Operations Research*, vol. 22, no. 4, pp. 331–346, 1988.

[168] A. Drud, "CONOPT: A GRG code for large sparse dynamic nonlinear optimization problems," *Mathematical Programming*, vol. 31, no. 2, pp. 153–191, 1985.

[169] GAMS Development Corporation, "General algebraic modeling system (GAMS)," 2021.

[170] N. V. Sahinidis, "BARON: A general purpose global optimization software package," *Journal of Global Optimization*, vol. 8, no. 2, pp. 201–205, 1996.

[171] U. M. Diwekar, *Introduction to applied optimization*, vol. 22. Springer Nature, 2020.

[172] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Structural and Multidisciplinary Optimization*, vol. 26, no. 6, pp. 369–395, 2004.

[173] Y. S. Lee, E. Graham, G. Jackson, A. Galindo, and C. S. Adjiman, "A comparison of the performance of multi-objective optimization methodologies for solvent design," in *Computer Aided Chemical Engineering*, vol. 46, pp. 37–42, Elsevier, 2019.

[174] J. Burger, N. Asprion, S. Blagov, R. Böttcher, U. Nowak, M. Bortz, R. Welke, K.-H. Küfer, and H. Hasse, "Multi-objective optimization and decision support in process engineering–implementation and application," *Chemie Ingenieur Technik*, vol. 86, no. 7, pp. 1065–1072, 2014.

[175] M. Bortz, J. Burger, N. Asprion, S. Blagov, R. Böttcher, U. Nowak, A. Scheithauer, R. Welke, K.-H. Küfer, and H. Hasse, "Multi-criteria optimization in chemical process design and decision support by navigation on pareto sets," *Computers & Chemical Engineering*, vol. 60, pp. 354–363, 2014.

[176] Y. Haimes, "On a bicriterion formulation of the problems of integrated system identification and system optimization," *IEEE transactions on systems, man, and cybernetics*, vol. 1, no. 3, pp. 296–297, 1971.

[177] A. I. Papadopoulos, I. Tsivintzelis, P. Linke, and P. Seferlis, "Computer aided molecular design: fundamentals, methods and applications," *Chem., Mol. Sci. and Chem. Eng*, 2018.

[178] P. J. Copado-Méndez, C. Pozo, G. Guillén-Gosálbez, and L. Jiménez, "Enhancing the ϵ-constraint method through the use of objective reduction and random sequences: Application to environmental problems," *Computers & Chemical Engineering*, vol. 87, pp. 36–48, 2016.

[179] G. Mavrotas, "Effective implementation of the ϵ-constraint method in multi-objective mathematical programming problems," *Applied Mathematics and Computation*, vol. 213, no. 2, pp. 455–465, 2009.

[180] M. H. Muhieddine, S. K. Viswanath, A. Armstrong, A. Galindo, and C. S. Adjiman, "Model-based solvent selection for the synthesis and crystallisation of pharmaceutical compounds," *Chemical Engineering Science*, vol. 264, p. 118125, 2022.

[181] J. G. Gmehling, T. F. Anderson, and J. M. Prausnitz, "Solid-liquid equilibria using UNIFAC," *Industrial & Engineering Chemistry Fundamentals*, vol. 17, no. 4, pp. 269–273, 1978.

[182] S. I. Sandler, *Chemical, biochemical, and engineering thermodynamics*. John Wiley & Sons, 2017.

[183] J. M. Smith, H. C. Van Ness, M. M. Abbott, and M. T. Swihart, *Introduction to chemical engineering thermodynamics*. McGraw-Hill Singapore, 1949.

[184] A. Buxton, A. G. Livingston, and E. N. Pistikopoulos, "Optimal design of solvent blends for environmental impact minimization," *AIChE Journal*, vol. 45, no. 4, pp. 817–843, 1999.

[185] R. A. Sheldon, "The E-factor: fifteen years on," *Green Chemistry*, vol. 9, no. 12, pp. 1273–1283, 2007.

[186] Z. Li, G. Chen, S. Qiao, R. Liang, C. Lan, and Z. Xia, "Design and synthesis of some new N-Phenylanthranilic Acids from highly sterically hindered anilines," *Synthetic Communications*, vol. 43, no. 9, pp. 1270–1279, 2013.

[187] A. Shah and N. M. Dodia, "An improved process for preparing N-phenylanthranilic acid and its derivatives," 2003.

[188] R. F. Pellón, R. Carrasco, T. Márquez, and T. Mamposo, "Use of N,N-dimethylformamide as solvent in the synthesis of N-phenylanthranilic acids," *Tetrahedron Letters*, vol. 38, no. 29, pp. 5107–5110, 1997.

[189] Y. Sarrafi, M. Mohadeszadeh, and K. Alimohammadi, "Microwave-assisted chemoselective copper-catalyzed amination of o-chloro and o-bromobenzoic acids using aromatic amines under solvent free conditions," *Chinese Chemical Letters*, vol. 20, no. 7, pp. 784–788, 2009.

[190] ICH, "Q3c (R6) on impurities: Guideline for residual solvents," in *International Conference for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH)*, 2016.

[191] "The Chemical Book," *https://www.chemicalbook.com/*, 2023.

[192] "National Institute of Standard and Technology (NIST) Chemistry WebBook," *https://webbook.nist.gov/chemistry/*, 2023.

[193] Z. Kolská, J. Kukal, M. Zábransky, and V. Ruzicka, "Estimation of the heat capacity of organic liquids as a function of temperature by a three-level group contribution method," *Industrial & Engineering Chemistry Research*, vol. 47, no. 6, pp. 2075–2085, 2008.

[194] S. Cesur and S. Gokbel, "Crystallization of mefenamic acid and polymorphs," *Crystal Research and Technology*, vol. 43, no. 7, pp. 720–728, 2008.

[195] S. K. Abdul Mudalip, M. R. Abu Bakar, P. Jamal, F. Adam, R. Che Man, S. Z. Sulaiman, Z. I. Mohd Arshad, and S. Md. Shaarani, "Effects of solvents on polymorphism and shape of mefenamic acid crystals," *MATEC Web of Conferences*, vol. 150, p. 02004, 2018.

[196] M. Shahid, G. Sanxaridou, S. Ottoboni, L. Lue, and C. Price, "Exploring the role of anti-solvent effects during washing on active pharmaceutical ingredient purity," *Organic Process Research & Development*, vol. 25, no. 4, pp. 969–981, 2021.

[197] J. Gmehling and B. Kolbe, "Thermodynamik georg thieme verlag," *Stuttgart—New York*, 1988.

[198] C. Antoine, "Vapor pressure: a new relationship between pressure and temperature," *CR Acad. Sci*, vol. 107, pp. 681–685, 1888.

[199] D. Webb and T. F. Jamison, "Continuous flow multi-step organic synthesis," *Chemical Science*, vol. 1, no. 6, pp. 675–680, 2010.

[200] A. Maleki, M. Aghaei, and R. Paydar, "Highly efficient protocol for the aromatic compounds nitration catalyzed by magnetically recyclable core/shell nanocomposite," *Journal of the Iranian Chemical Society*, vol. 14, no. 2, pp. 485–490, 2017.

[201] K. G. Thakur and G. Sekar, "D-Glucose as green ligand for selective copper-catalyzed phenol synthesis from aryl halides with an easy catalyst removal," *Chemical Communications*, vol. 47, no. 23, pp. 6692–6694, 2011.

[202] G. C. Silva, N. M. Carvalho, A. Horn Jr, E. R. Lachter, and O. A. Antunes, "Oxidation of aromatic compounds by hydrogen peroxide catalyzed by mononuclear iron (III) complexes," *Journal of Molecular Catalysis A: Chemical*, vol. 426, pp. 564–571, 2017.

[203] J. Wang, Z. Song, R. Lakerveld, and T. Zhou, "Solvent selection for chemical reactions toward optimal thermodynamic and kinetic performances: Group contribution and COSMO-based modeling," *Fluid Phase Equilibria*, vol. 564, p. 113623, 2022.

[204] A. J. Haslam, A. Gonzalez-Perez, S. Di Lecce, S. H. Khalit, F. A. Perdomo, S. Kournopoulos, M. Kohns, T. Lindeboom, M. Wehbe, S. Febra, *et al.*, "Expanding the applications of the SAFT-γ Mie group-contribution equation of state: Prediction of thermodynamic properties and phase behavior of mixtures," *Journal of Chemical & Engineering Data*, vol. 65, no. 12, pp. 5862–5890, 2020.

[205] J. D. Chea, A. L. Lehr, J. P. Stengel, M. J. Savelski, C. S. Slater, and K. M. Yenkie, "Evaluation of solvent recovery options for economic feasibility through a superstructure-based optimization framework," *Industrial & Engineering Chemistry Research*, vol. 59, no. 13, pp. 5931–5944, 2020.

Appendix A

UNIFAC Model

The equations proposed by Smith et al. [183] are used as they are formulated in a convenient form for programming purposes. These equations are slightly modified in order to avoid numerical difficulties when evaluating activity coefficients.

Activity coefficient

$$\ln \gamma_{i,t}(x,T) = \ln \gamma^{C}_{i,t}(x,T) + \ln \gamma^{R}_{i,t}(x,T), \quad i \in C^K; t \in T \tag{A.1}$$

Combinatorial part of activity coefficient

$$\ln \gamma^{C}_{i,t}(x,T) = 1 - J_{i,t} + \ln J_{i,t} - 5q_{i,t}\left(1 - \frac{J_{i,t}}{L_{i,t}} + \ln \frac{J_{i,t}}{L_{i,t}}\right), \quad i \in C^K; t \in T \tag{A.2}$$

$$r_i = \sum_{k \in K} n_{i,k} R_k, \quad i \in C^K \tag{A.3}$$

$$q_i = \sum_{k \in K} n_{i,k} Q_k, \quad i \in C^K \tag{A.4}$$

$$r_s = \sum_{k \in K} n_{s,k} R_k, \quad s \in S \tag{A.5}$$

$$q_s = \sum_{k \in K} n_{s,k} Q_k, \quad s \in S \tag{A.6}$$

$$r_q = \sum_{s \in S} r_s y_{q,s}, \quad q \in Q; s \in S \tag{A.7}$$

$$q_q = \sum_{s \in S} q_s y_{q,s}, \quad q \in Q; s \in S \tag{A.8}$$

$$J_{i,t} = \frac{r_i}{\sum_{i \in C} r_i x_{i,t}}, \quad i \in C^K; t \in T \tag{A.9}$$

$$L_{i,t} = \frac{q_i}{\sum_{i \in C} q_i x_{i,t}}, \quad i \in C^K; t \in T \tag{A.10}$$

Residual part of activity coefficient

$$\ln \gamma^R_{i,t}(x,T) = q_i \left[1 - \sum_{k \in K} \left(\theta_{k,t} \frac{\beta_{i,k,t}}{\omega_{k,t}} - e_{i,k} \ln \frac{\beta_{i,k,t}}{\omega_{k,t}} \right) \right], \quad i \in C^K; t \in T \tag{A.11}$$

$$e_{i,k} = \frac{n_{i,k} Q_k}{q_i}, \quad i \in C^K; k \in K \tag{A.12}$$

$$\beta_{i,k,t} = \sum_{m \in K} \frac{n_{i,m} Q_m \psi_{m,k,t}}{q_i}, \quad i \in C^K; k \in K; t \in T \tag{A.13}$$

$$\theta_{k,t} = \frac{\sum_{i \in C} x_{i,t} n_{i,k} Q_k}{\sum_{i \in C} x_{i,t} q_i}, \quad k \in K; t \in T \tag{A.14}$$

$$\omega_{k,t} = \sum_{m \in K} \theta_{m,t} \psi_{m,k,t}, \quad k \in K; t \in T \tag{A.15}$$

$$\psi_{m,k,t} = \exp\left(\frac{-a_{m,k}}{T_t} \right), \quad m \in K; k \in K; t \in T \tag{A.16}$$

Appendix B

Solvent Lists

B.1 List I

- 1,2-Propanediol
- Acetic acid
- Acetone
- Diethoxymethane
- Ethanol
- Ethyl acetate
- Formic acid
- Isobutyl acetate
- Isopropyl acetate
- Isobutanol
- Isopropanol
- Methyl acetate
- Methyl ethyl ketone (MEK)

- Methyl tert-butyl ether (MTBE)
- n-Butanol
- 2-Butanol
- n-Heptane
- tert-Amyl alcohol (t-AmOH)
- t-Butanol
- Triethyl amine
- Water
- Anisole
- n-Butyl acetate
- Trifluoroethanol
- Trifluoroacetic acid
- 1-Propanol
- n-Pentane
- Diethyl ether
- Cyclohexanone
- Ethyl formate
- Nitrobenzene
- Dimethyl sulfide
- 2-Methyl pentane
- 2-Pentanol

- 2-Pentanone
- 3-Pentanone
- Cyclopentyl methyl ether
- 2-2-4-Trimethyl pentane
- Methyl cyclopentane
- Propyl acetate
- t-Amyl methyl ether
- t-Butyl acetate
- t-Butyl ethyl ether

B.2 List II

- 1,2-Dimethoxyethane
- 1,2-Propanediol
- 1,4-Dioxane
- Acetic acid
- Acetone
- Acetonitrile
- Chlorobenzene
- Chloroform
- Cyclohexane
- Diethoxymethane

- DMAC
- Ethanol
- Ethyl acetate
- Formic acid
- Isobutyl acetate
- Isopropyl acetate
- Isobutanol
- Isopropanol
- Methyl acetate
- MIBK
- MEK
- MTBE
- n-Butanol
- 2-Butanol
- n-Heptane
- n-Hexane
- Nitromethane
- o-Xylene
- t-AmOH
- t-Butanol
- Toluene

- Triethyl amine
- Water
- Anisole
- n-Butyl acetate
- Trifluoroethanol
- Trifluoroacetic acid
- 1-Propanol
- n-Pentane
- Carbon tetrachloride
- Diethyl ether
- Cyclohexanone
- Ethyl formate
- Nitrobenzene
- Dimethyl sulfide
- 2-Methyl pentane
- 2-Pentanol
- 2-Pentanone
- 3-Pentanone
- Cyclopentyl methyl ether
- Dimethyl carbonate
- 2,2,4-Trimethyl pentane

- Methyl cyclohexane
- Methyl cyclopentane
- Propyl acetate
- t-Amyl methyl ether
- t-Butyl acetate
- t-Butyl ethyl ether
- DMF

B.3 List III

- 1,2-Dimethoxyethane
- 1,4-Dioxane
- Acetic acid
- Acetone
- Acetonitrile
- Chlorobenzene
- Chloroform
- Cyclohexane
- DMAC
- Ethanol
- Ethyl acetate
- Formic acid

- Isopropyl acetate
- Isobutanol
- Isobutyl acetate
- Isopropanol
- Methyl acetate
- MIBK
- MEK
- MTBE
- n-Butanol
- 2-Butanol
- n-Heptane
- n-Hexane
- Nitromethane
- o-Xylene
- t-AmOH
- t-Butanol
- Toluene
- Triethyl amine
- Water
- Anisole
- n-Butyl acetate

- 1-Propanol
- n-Pentane
- Carbon tetrachloride
- Diethyl ether
- Ethyl formate
- Nitrobenzene
- 2-Methyl pentane
- 2-Pentanone
- 3-Pentanone
- 2,2,4-Trimethyl pentane
- Methyl cyclohexane
- Methyl cyclopentane
- Propyl acetate
- t-Amyl methyl ether
- t-Butyl acetate
- t-Butyl ethyl ether
- DMF

Appendix C

Data

C.1 Melting Points, Heats of Melting and Heat Capacities of Solids

The melting points of CBA, DMA and MA are obtained from the Chemical Book (`https://www.chemicalbook.com/ProductIndex_EN.aspx`), the heats of melting of CBA and MA are obtained from the National Institute of Standards and Technology (NIST) Chemistry WebBook (`https://webbook.nist.gov/chemistry`), and the heat of melting of DMA is estimated using the Marrero-Gani GC method [121].

Standard constant-pressure heat capacities of solvent candidates in set S are obtained from the NIST Chemistry WebBook (`https://webbook.nist.gov/chemistry`), whereas standard constant-pressure heat capacities of reactants and products are calculated using the GC method of Kolská et al. [193].

Compound $i \in C^K$	Melting Point $T_{m,i}$ (K)	Heat of Melting $\Delta H_{m,i}$ (J/mol)	Liquid-Phase Molar Heat Capacity $C_{p_i}^{(L)}$ (J/mol-K)
DMA	276.70	14940	242.20
CBA	414.10	26290	241.22
MA	503.50	38700	447.20
CB	227.95	9550	152.10
NCB	356.15	15030	250.20
CMN	177.15	7326	214.40
PNL	314.05	11514	127.21
NP	386.95	30118	144.00

C.2 Standard Heats of Reaction

The standard heat of each considered reaction is calculated from the heats of formation of the corresponding reactants and products, which are estimated using the Marrero-Gani GC method [121].

Reaction	Standard Heat of Reaction $\Delta H_R(T_R)$ (J/mol)
Synthesis of MA from DMA and CBA	-95067
Synthesis of NCB from CB	-35792
Synthesis of NP from NCB	-147189
Synthesis of PNL from CMN	-97673
Synthesis of NP from PNL	-12861

C.3 Pure Component Densities of Species Involved in the Mefenamic Acid Case Study

Pure component densities, ρ_i for $i \in C^K$, are obtained from the Chemical Book (`https://www.chemicalbook.com/ProductIndex_EN.aspx`).

Compound i $\in C^K$	ρ_i (g/mL)
DMA	0.993
CBA	1.197
MA	1.203

C.4 Properties of Solvents in List III

Solvent $s \in S$	Boiling Point $T_{b,s}$ (K)	Melting Point $T_{m,s}$ (K)	Molar Mass MW_s (g/mol)	Density ρ_s (g/mL)	Heat Capacity $C_{p_s}^{(L)}$ (J/mol-K)	Heat of vaporisation ΔH_{V_s} (J/mol)
1-2-Dimethoxyethane	358.15	204.15	90.12	0.87	191.14	32420
1,4-Dioxane	374.15	285.15	88.11	1.03	149.65	37500
Acetic acid	391.00	289.00	60.05	1.05	123.10	51600
Acetone	329.20	178.50	58.08	0.79	125.45	29100
Acetonitrile	355.00	228.00	41.05	0.78	91.70	29750
Chlorobenzene	404.00	228.00	112.56	1.11	152.10	40100
Chloroform	334.40	209.70	119.38	1.49	114.25	31400
Cyclohexane	353.90	279.62	84.16	0.78	156.00	29970
DMAC	438.20	253.00	87.12	0.94	178.20	45100
Ethanol	351.40	159.00	46.07	0.79	112.40	38560
Ethyl acetate	350.20	189.60	88.11	0.90	168.94	31940
Formic acid	373.90	281.50	46.03	1.22	99.04	22690
Isobutyl acetate	399.20	195.00	116.16	0.87	240.20	39900
Isopropyl acetate	362.00	200.00	102.13	0.87	196.60	32900
Isobutanol	381.04	165.00	74.12	0.80	181.05	43290
Isopropanol	355.80	184.00	60.10	0.79	161.20	39850
Methyl acetate	330.00	175.00	74.08	0.93	141.34	30320
MIBK	390.00	188.50	100.16	0.80	211.90	40650
MEK	353.15	186.15	72.11	0.81	159.00	31300
MTBE	328.30	164.00	88.15	0.74	187.50	27940
n-Butanol	390.80	183.30	74.12	0.81	176.86	43290
2-Butanol	373.00	158.00	74.12	0.81	197.10	40750
n-Heptane	371.53	182.60	100.21	0.68	224.64	31770
n-Hexane	342.00	178.00	86.18	0.66	265.20	28850
Nitromethane	374.30	244.50	61.04	1.13	105.98	33990
o-Xylene	417.50	249.00	106.17	0.88	187.65	36240
t-AmOH	375.00	264.00	88.15	0.81	247.15	40110
t-Butanol	355.00	298.00	74.12	0.81	215.37	39070
Toluene	384.00	178.00	92.14	0.86	157.09	33180
Triethyl amine	362.00	158.45	101.19	0.73	216.43	31010
Water	373.15	273.15	18.02	0.99	75.38	40660
Anisole	427.00	236.00	108.14	0.99	199.00	46910

Solvent $s \in S$	Boiling Point $T_{b,s}$ (K)	Melting Point $T_{m,s}$ (K)	Molar Mass MW_s (g/mol)	Density ρ_s (g/mL)	Heat Capacity $C_{p_s}^{(L)}$ (J/mol-K)	Heat of vaporisation ΔH_{V_s} (J/mol)
n-Butyl acetate	399.20	195.00	116.16	0.88	225.11	36280
1-Propanol	370.00	147.00	60.10	0.80	143.96	41440
n-Pentane	309.00	143.00	72.15	0.63	167.19	25790
Carbon tetrachloride	349.87	250.23	153.82	1.59	131.30	29820
Diethyl ether	307.80	256.80	74.12	0.71	172.50	26520
Ethyl formate	327.10	193.00	74.08	0.92	144.30	32110
Nitrobenzene	484.00	278.80	123.11	1.21	181.13	54500
2-Methyl pentane	334.00	120.00	86.18	0.66	194.19	27790
2-Pentanone	375.00	196.00	86.13	0.81	185.40	38460
3-Pentanone	375.00	234.20	86.13	0.82	196.40	33450
2,2,4-Trimethyl pentane	372.40	165.77	114.23	0.69	242.49	30790
Methyl cyclohexane	374.00	146.60	98.19	0.77	184.38	31270
Methyl cyclopentane	345.00	131.00	84.16	0.75	158.70	29080
Propyl acetate	374.70	178.15	102.13	0.88	194.10	33920
t-Amyl methyl ether	359.40	193.00	102.17	0.77	222.00	27940
t-Butyl acetate	369.80	211.15	116.16	0.86	231.00	36280
t-Butyl ethyl ether	346.00	179.15	102.17	0.74	218.00	29760
DMF	426.15	212.15	73.10	0.95	146.05	43600

C.5 Properties of Additional Solvents in List I

Solvent $s \in S$	Boiling Point $T_{b,s}$ (K)	Melting Point $T_{m,s}$ (K)	Molar Mass MW_s (g/mol)	Density ρ_s (g/mL)
1,2-Propanediol	460.15	213.15	76.10	1.04
Diethoxymethane	361.15	206.65	104.15	0.83
Trifluoroethanol	347.10	229.70	100.04	1.39
Trifluoroacetic acid	345.50	257.80	114.02	1.54
Cyclohexanone	420.80	226.00	98.14	0.95
Dimethyl sulfide	311.00	175.00	62.13	0.85
2-Pentanol	392.00	200.00	88.15	0.81
Cyclopentyl methyl ether	379.00	133.00	100.16	0.86

Appendix D

Solvent Costs

Solvent costs are obtained from the Chemical Availability Search section of the Physical Sciences Data-Science Service (PSDS): `https://chase-home-offsite.psds.ac.uk/`.

Solvent $s \in S$	Purity (%)	Volume (L)	Price (£)	Unit Price (£/L)
1,2-Dimethoxyethane	99.00	1.00	90.00	90.00
1,4-Dioxane	99.00	4.00	266.39	66.60
Acetic acid	$\geq$ 99.00	20.00	259.20	12.96
Acetone	99.50	20.00	465.00	23.25
Acetonitrile	99.50	20.00	1360.00	68.00
Chlorobenzene	99.50	4.00	116.10	29.03
Chloroform	$\geq$ 99.50	4.00	117.00	29.25
Cyclohexane	99.00	20.00	354.00	17.70
DMAC	99.50	20.00	289.60	14.48
Ethanol	99.50	20.00	775.30	38.77
Ethyl acetate	$\geq$ 99.50	20.00	279.89	14.00
Formic acid	$\geq$ 98.00	4.00	100.80	25.20
Isopropyl acetate	99.00	25.00	247.80	9.91

Solvent $s \in S$	Purity (%)	Volume (L)	Price (£)	Unit Price (£/L)
Isobutanol	$\geq$ 99.00	4.00	125.10	31.28
Isopropanol	99.50	4.00	85.50	21.38
Methyl acetate	99.00	4.00	112.80	28.20
MIBK	$\geq$ 99.00	4.00	99.00	24.75
MEK	$\geq$ 99.00	25.00	228.30	9.13
MTBE	99.00	25.00	299.20	11.97
n-Butanol	99.00	25.00	158.50	6.34
2-Butanol	99.00	10.00	82.50	8.25
n-Heptane	$\geq$ 99.00	25.00	476.90	19.08
n-Hexane	$\geq$ 99.00	4.00	81.00	20.25
Nitromethane	99.00	10.00	323.60	32.36
o-Xylene	99.00	1.00	33.70	33.70
t-AmOH	99.00	10.00	459.60	45.96
t-Butanol	$\geq$ 99.00	4.00	104.40	26.10
Toluene	$\geq$ 99.70	4.00	59.70	14.93
Triethyl amine	99.00	4.00	87.20	21.80
Water	Pure, dimineralised	20.00	97.00	4.85
Anisole	99.00	1.00	33.90	33.90
n-Butyl acetate	$\geq$ 99.00	25.00	340.80	13.63
1-Propanol	$\geq$ 99.00	25.00	260.40	10.42
n-Pentane	$\geq$ 99.00	25.00	421.20	16.85
Carbon tetrachloride	$\geq$ 99.50	2.00	536.00	268.00
Diethyl ether	99.00	4.00	160.19	40.04
Ethyl formate	$\geq$ 98.00	25.00	289.70	11.58
Nitrobenzene	99.00	10.00	185.50	18.55
2-Methyl pentane	$\geq$ 99.00	2.50	631.80	252.72
2-Pentanone	99.00	4.00	96.90	24.23

Solvent $s \in S$	Purity (%)	Volume (L)	Price (£)	Unit Price (£/L)
3-Pentanone	99.00	2.50	49.10	19.64
2,2,4-Trimethyl pentane	$\geq$ 99.00	4.00	81.00	20.25
Methyl cyclohexane	99.00	10.00	151.20	15.12
Methyl cyclopentane	95.00	25.00	2543.10	101.72
Propyl acetate	99.00	25.00	363.00	14.52
t-Amyl methyl ether	97.00	0.50	599.00	1198.00
t-Butyl acetate	99.00	1.00	28.00	28.00
t-Butyl ethyl ether	$\geq$ 95.00	0.025	27.00	1080.00
DMF	99.80	20.00	622.00	31.10

Appendix E

CAM^{b}D Formulation of the Route Selection Problem for the NP Case Study (Route 1)

E.0.1 Problem Objective

Minimise KPI

E.0.2 Process Model Constraints

Solvent assignment constraints

$$\sum_{s \in S} y_{s1,s} = 1 \tag{E.1}$$

$$\sum_{s \in S} y_{s2,s} \leq 1 \tag{E.2}$$

$$\sum_{q \in Q} y_{q,s} \leq 1, \quad s \in S \tag{E.3}$$

Solvent properties

$$MW_q = \sum_{s \in S} MW_s y_{q,s}, \quad q \in Q_1 \tag{E.4}$$

$$T_{b,q} = \sum_{s \in S} T_{b,s} y_{q,s}, \quad q \in Q_1 \tag{E.5}$$

$$T_{m,q} = \sum_{s \in S} T_{m,s} y_{q,s}, \quad q \in Q_1 \tag{E.6}$$

Masses and mole fractions

$$M_{i,t} = N_{i,t} MW_i, \quad i \in C_1; t \in T \tag{E.7}$$

$$x_{i,t} = \frac{N_{i,t}}{\sum_{i' \in C} N_{i',t}}, \quad i \in C_1; t \in T \tag{E.8}$$

$$x_{s2,t} \leq \sum_{s \in S} y_{s2,s}, \quad t \in \{1,2,3,4,5,6,7,8,9,10\} \tag{E.9}$$

Heat exchanger 1 balances

$$N_{i,1} = N_{i,2}, \quad i \in C_1 \tag{E.10}$$

Reactor 1 balances and equilibrium constraints

$$N_{i,3} = N_{i,2} + \sum_{r \in R} \nu_{i,r} \xi_r, \quad i \in C_1 \tag{E.11}$$

$$\xi_1 = x^{c_1} N_{\mathrm{CB},2} \tag{E.12}$$

$$x_{i,t} \leq \hat{x}_{i,t}^{(i)} - \epsilon_s, \quad i \in C_1^K; t = \{2,3\} \tag{E.13}$$

$$\ln \hat{x}_{i,t}^{(i)} + \ln \hat{\gamma}_{i,t}(\hat{\mathbf{x}}_t^{(i)}, T_t) = \frac{\Delta H_{m,i}}{R_g} \left[\frac{1}{T_{m,i}} - \frac{1}{T_t} \right], \quad i \in C_1^K, \quad t \in \{2,3\} \tag{E.14}$$

$$\frac{\hat{x}_{i',t}^{(i)}}{\hat{x}_{s1,t}^{(i)}} = \frac{x_{i',t}}{x_{s1,t}}, \quad i' \in C_1 \setminus \{s1, i\}, \quad i \in C_1^K, \quad t \in \{2,3\} \tag{E.15}$$

$$\sum_{i' \in C_1} \hat{x}_{i',t}^{(i)} = 1, \quad i \in C_1^K, \quad t \in \{2,3\} \tag{E.16}$$

Mixer balances

$$N_{i,3} + N_{i,4} = N_{i,5}, \quad i \in C_1 \tag{E.17}$$

Heat exchanger 2 balances

$$N_{i,5} = N_{i,6}, \quad i \in C_1 \tag{E.18}$$

Reactor 2 balances and equilibrium constraints

$$N_{i,7} = N_{i,6} + \sum_{r \in R} \nu_{i,r} \xi_r, \quad i \in C_1 \tag{E.19}$$

$$\xi_1 = x^{c_2} N_{\mathrm{NCB},6} \tag{E.20}$$

$$x_{i,t} \leq \hat{x}_{i,t}^{(i)} - \epsilon_s, \quad i \in C_1^K; t = \{6,7\} \tag{E.21}$$

$$\ln \hat{x}_{i,t}^{(i)} + \ln \hat{\gamma}_{i,t}(\hat{\mathbf{x}}_t^{(i)}, T_t) = \frac{\Delta H_{m,i}}{R_g}\left[\frac{1}{T_{m,i}} - \frac{1}{T_t}\right], \quad i \in C_1^K, \quad t \in \{6,7\} \tag{E.22}$$

$$\frac{\hat{x}_{i',t}^{(i)}}{\hat{x}_{s1,t}^{(i)}} = \frac{x_{i',t}}{x_{s1,t}}, \quad i' \in C_1 \setminus \{s1, i\}, \quad i \in C_1^K, \quad t \in \{6,7\} \tag{E.23}$$

$$\sum_{i' \in C_1} \hat{x}_{i',t}^{(i)} = 1, \quad i \in C_1^K, \quad t \in \{6,7\} \tag{E.24}$$

Crystalliser balances and equilibrium constraints

$$N_{i,7} + N_{i,8} - N_{i,9} - N_{i,10} = 0, \quad i \in C_1 \tag{E.25}$$

$$N_{i,8} = 0, \quad i \in C_1 \setminus \{s2\} \tag{E.26}$$

$$N_{i,10} = 0, \quad i \in C_1 \setminus \{\mathbf{NP}\} \tag{E.27}$$

$$\ln x_{NP,9} + \ln \gamma_{NP,9}(\mathbf{x}_9, T_9) = \frac{\Delta H_{m,NP}}{R_g}\left[\frac{1}{T_{m,NP}} - \frac{1}{T_9}\right] \tag{E.28}$$

$$\ln \hat{x}_{i,9}^{(i)} + \ln \hat{\gamma}_{i,9}(\hat{\mathbf{x}}_9^{(i)}, T_9) = \frac{\Delta H_{m,i}}{R_g}\left[\frac{1}{T_{m,i}} - \frac{1}{T_9}\right], \quad i \in C_1^K \setminus \{\mathbf{NP}\} \tag{E.29}$$

$$\frac{\hat{x}_{i',9}^{(i)}}{\hat{x}_{s1,9}^{(i)}} = \frac{x_{i',9}}{x_{s1,9}}, \quad i' \in C_1 \setminus \{s1, i\}, \quad i \in C_1^K \setminus \{\mathbf{NP}\} \tag{E.30}$$

$$\sum_{i' \in C_1} \hat{x}_{i',9}^{(i)} = 1, \quad i \in C_1^K \setminus \{\mathbf{NP}\} \tag{E.31}$$

$$x_{i,9} \leq \hat{x}_{i,9}^{(i)} - \epsilon_s, \quad i \in C_1^K \setminus \{\mathbf{NP}\} \tag{E.32}$$

E.0.3 Process Design Constraints

Operating temperatures constraints

$$T_t - (T_{b,q} - T_o) \leq M_T(1 - \sum_{s \in S} y_{q,s}), \quad q \in Q_1, t \in T \tag{E.33}$$

$$(T_{m,q} + T_o) - T_t \leq M_T(1 - \sum_{s \in S} y_{q,s}), \quad q \in Q_1, t \in T \tag{E.34}$$

Miscibility constraints

$$-\tilde{M}_x(1-\sum_{s\in S} y_{s2,s}) \leq \tilde{x}_{s1,t} - \frac{x_{s1,t}}{x_{s1,t}+x_{s2,t}} \leq \tilde{M}_x(1-\sum_{s\in S} y_{s2,s}), \quad t \in \{1,2,6,9\} \tag{E.35}$$

$$\frac{\partial \ln \tilde{\gamma}_{s1,t}(\tilde{\mathbf{x}}_t, T_t)}{\partial \tilde{x}_{s1,t}} + \frac{1}{\tilde{x}_{s1,t}} \geq -M_\gamma(1-\sum_{s\in S} y_{s2,s}), \quad t \in \{1,2,6,9\} \tag{E.36}$$

KPIs Equations in Section 5.2.3

Problem Specifications Specifications in Table 5.3

Appendix F

Optimal Process Temperatures in the Route Design Formulations

F.1 Optimal Process Temperatures for the Minimum PEF Formulation

Solution number	Route 1			Route 2		
	T_2 (K)	T_6 (K)	T_9 (K)	T_2 (K)	T_6 (K)	T_9 (K)
1	298	350	293.15	320	314	294
2	298	355	293.15	320	314	303
3	298	355	293.15	320	314	304
4	298	356	293.15	320	314	305
5	298	352	293.15	320	314	299
6	298	352	320.00	320	314	305
7	298	350	299.00	320	314	306
8	298	351	299.00	320	314	307
9	298	356	295.15	320	314	307
10	298	356	293.15	320	314	309

F.2 Optimal Process Temperatures for the Minimum Q_T Formulation

Solution number	Route 1			Route 2		
	T_2 (K)	T_6 (K)	T_9 (K)	T_2 (K)	T_6 (K)	T_9 (K)
1	298	356	318	320	314	309
2	298	355	318	320	314	307
3	298	356	310	320	314	308
4	298	356	310	320	314	307
5	298	356	310	320	314	306
6	298	356	307	320	314	305
7	298	356	308	320	314	305
8	298	356	307	320	314	307
9	293	354	319	320	314	299
10	296	356	306	320	314	294

F.3 Optimal Process Temperatures for the Minimum C_T Formulation

Solution number	Route 1			Route 2		
	T_2 (K)	T_6 (K)	T_9 (K)	T_2 (K)	T_6 (K)	T_9 (K)
1	293.15	355	298.42	320	314	293.15
2	293.15	350	293.15	320	314	293.15
3	293.15	355	299.00	320	314	293.15
4	293.15	355	300.00	320	314	293.15
5	293.15	352	296.00	320	314	293.15
6	293.15	355	299.00	320	314	293.15
7	293.15	352	293.15	320	314	293.15
8	293.15	356	299.00	320	314	293.15
9	293.15	353	293.15	320	314	293.15
10	293.15	351	293.15	320	314	293.15

F.4 Optimal Process Temperatures for the Bi-objective Formulation: Minimum PEF, Minimum Q_T

Solution number	Route 1			Route 2		
	T_2 (K)	T_6 (K)	T_9 (K)	T_2 (K)	T_6 (K)	T_9 (K)
1	298	355	319	320	314	296
2	298	354	319	320	314	302
3	298	354	319	320	314	301
4	298	354	319	320	314	303
5	298	356	320	320	314	305
6	298	355	320	320	314	306

F.5 Optimal Process Temperatures for the Bi-objective Formulation: Minimum C_T, Maximum Y_C

Solution number	Route 1			Route 2		
	T_2 (K)	T_6 (K)	T_9 (K)	T_2 (K)	T_6 (K)	T_9 (K)
1	293.15	355	298.42	320	314	293.15
2	293.15	355	296.64	320	314	293.15
3	293.15	355	293.23	320	314	293.15
4	293.15	350	293.15	320	314	293.15
5	293.15	350	293.15	320	314	293.15
6	293.15	350	293.15	320	314	293.15