# Intensification of a Workflow for Particle Engineering and Crystallisation Process Development

Thomas Alexander Pickles

Strathclyde Institute of Pharmacy and Biomedical Sciences

University of Strathclyde

Glasgow, UK

A thesis presented in fulfilment of the requirements for the degree of Doctor of Philosophy in the Faculty of Sciences of the University of Strathclyde

March 2024

## Declaration of authenticity and author's rights

'This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.'

Signed:

Date:        14-March-2024

# Abstract

Crystallisation is a fundamental aspect of chemical and pharmaceutical manufacturing to ensure high purity, bio-availability and desirable physical attributes of quality products. In this thesis, a comprehensive workflow was designed to enhance the efficiency of crystallisation process development through the utilisation of digital tools. This research aims to integrate existing knowledge in crystallisation thermodynamics and kinetics and quality by design (QbD) principles into a unified framework.

To develop and test the workflow's approach, two case studies were conducted involving the crystallisation of lamivudine and aspirin. These studies served as foundational experiments to first develop the workflow and second validate the individual components and the logical flow of the workflow. Data collected was inclusive of solubility, morphological characteristics, particle size, kinetic properties, and solid forms in a time and material-reductive way.

The two case studies highlighted a need for a more intelligent experimental planning and optimisation tool, particularly when contrasted with conventional methodologies. To address this requirement, an Adaptive Bayesian Optimisation (AdBO) tool was developed and applied to the crystallisation processes of lamivudine and aspirin. This acceleration of the workflow demonstrated significant advantages when compared to traditional grid search and design of experiment (DoE) optimisation approaches.

Subsequently, the generalisability of the workflow was validated by applying it to five additional active pharmaceutical ingredient (API) case studies, ibuprofen, ascorbic acid, salicylic acid, benzoic acid and D-mannitol. This evaluation, across a broader chemical scope, demonstrated the versatility and robustness of the workflows approach.

In contrast to conventional industrial research and development methods, which typically operate on a scale exceeding 100 mL, consume substantial time resources, and necessitate a large workforce, our workflow was integrated into an industrial pharmaceutical facility. This integration enabled the systematic validation of the workflow across various stirring methods, crystallisation modes, and vessel sizes,

ultimately leading to the design of a robust hybrid antisolvent-cooling crystallisation process.

This thesis provides a comprehensive framework for the optimisation of crystallisation processes, leveraging digital tools to streamline experimentation, enhance efficiency and promote consistency. This is impactful to the wider community as the workflow and digital tools developed can seamlessly be integrated into existing chemical and pharmaceutical research to yield efficiencies. The generalisability shown by this work also allows for the expansion of similar work packages into areas outside of chemical and pharmaceutical manufacturing.

# Acknowledgements

I would like to start by thanking my supervisors, Prof Alastair Florence, Dr Cameron Brown and Dr Chantal Mustoe, for all their guidance, insight and support throughout the last 3 years. To all the technical staff within the National Facility for their ever present training and maintenance of the lab space. To all my co-authors in getting our research out to the wider community in the most efficient and impactful way possible. Also, to all the new friends I have made within CMAC, the climbing, the hiking and the numerous sub-crawls have kept me grounded.

I would like to extend thanks to everybody at Pfizer that made my 3-month placement very enjoyable and an all-round successful experience. A special thanks is reserved for Václav Svoboda.

Finally, I would like to express my gratitude to my friends and to my family, Ian, Sharon and Rebecca (plus the kids), for all the help, support and welcome distractions.

Thank you.

*Table 1. CRediT (Contributor Roles Taxonomy) for all co-authors listed across the work found in this thesis.*

| Person | Contribution | Chapter(s) |
|---|---|---|
| Alastair Florence | Supervision, Writing – review and editing | All |
| Cameron Brown | | |
| Chantal Mustoe | | |
| Javier Cardona | Writing – review and editing | 4 |
| Chris Boyle | Software, Writing – review and editing | 4 |
| Ivan Marziano | Supervision, Writing – review and editing | 8 |
| Václav Svoboda | | |

# Contents

# 1. Introduction.

## 1.1. Crystallisation

### 1.1.1. Crystallisation in the Pharmaceutical Industry

Crystallisation is a significant step in the development and manufacture of commercial pharmaceuticals and holds significance to the quality and properties of drug substances. The pharmaceutical industry faces unique challenges with crystallisation so that parameters such as crystal form and particle size distribution are met whilst ensuring good manufacturing practice (GMP), approval from regulatory boards and strict batch-to-batch consistency.[1]
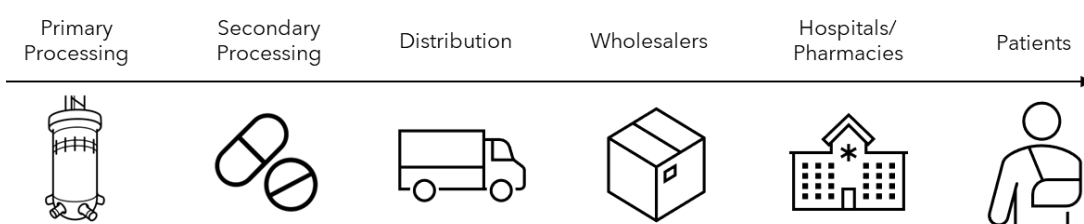


*Figure 1. A schematic illustration of the many steps involved in drug manufacture.*

As shown in Figure 1 there are many steps from initial active pharmaceutical ingredient (API) synthesis before prescription to the patient. Crystallisation occurs as one of the key stages throughout the primary processing steps including synthesis and separation of intermediates, purification of the drug substance and ensuring processability in downstream manufacture.

### 1.1.2. Thermodynamics

Equilibrium solubility is the point at which a solvent is saturated with a solute at a specific temperature, it is denoted by the clear point line in Figure 2. Dissolution is the process where a solute, generally in the solid phase but can be liquid or gaseous, dissolves in a solvent to form a solution, this is a reversible process. Solubility typically varies with temperature with a positive exponential relationship. This is approximately described by Black's rule that solubility doubles with every increase of 20 °C[2] however some systems have higher or lower temperature dependence. It is important to consider solvent choices when designing a crystallisation process as a strong temperature dependence would give the best

yield. Differences in solubility is essential for crystallisation as the composition and/or temperature defines the supersaturation (i.e. thermodynamic driving force) which governs the crystallisation rate processes that dictate the outcome.
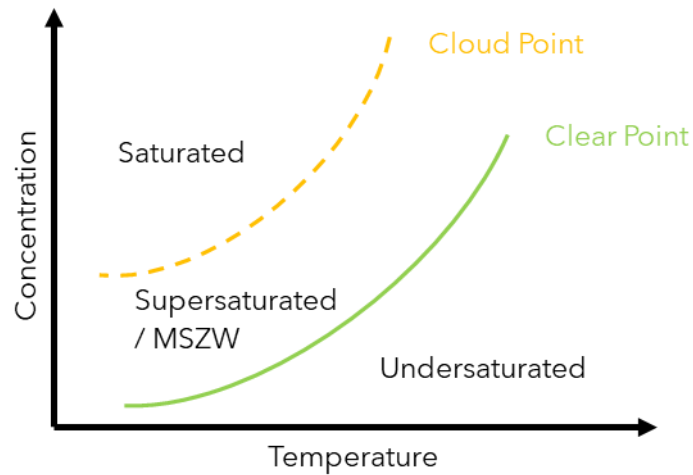


*Figure 2. Simple crystallisation phase diagram representative of a solute in a solvent.*

Figure 2 shows the three phases of a solute in a solvent system and is representative only. Above the cloud point line is the saturated area, between the cloud and clear point is the metastable zone width (MSZW) and below the clear point is the undersaturated area.

A system can become supersaturated when the amount of dissolved solute is greater than the equilibrium solubility which can be achieved by cooling, changing the solvent composition using antisolvent and evaporation of some of the solvent. The highest level of supersaturation (refer to Equation 1) possible in a system is shown by the cloud point line in Figure 2 where if passed, the formation of solid particles will spontaneously occur. Supersaturation can be expressed and related to chemical potential by the following equations.

*Equation 1.*   $SS = \left(\dfrac{C_{ss}}{C_{eq}}\right)$

*Equation 2.*   $\Delta\mu = ln\,(SS)$

where SS is the supersaturation ratio, $C_{ss}$ is the supersaturated concentration, $C_{eq}$ is the equilibrium concentration and $\Delta\mu$ is the change in chemical potential.[3]

2

A positive change in chemical potential will drive a transition towards being spontaneous. It is important to control supersaturation for quality assurance as nucleation and growth are both affected by this property.

The MSZW is a supersaturated region where the nucleation of crystals can occur but will not spontaneously unless triggered by another driving force such as shear, agitation or seeding. Performing a crystallisation within the MSZW allows for a higher degree of control of the crystallisation and subsequent quality of the final product.

Crystallisation in a system depends on the availability of sufficient free energy to facilitate crystal formation, indicated by a negative Gibbs free energy change. This thermodynamic driving force is met when there is a disparity between the chemical potential of the pure solute and the solute in solution.[4]

$$\text{Equation 3.} \quad \mu_L = \mu_L^* + kT \ln a$$

$$\text{Equation 4.} \quad \mu_S = \mu_S^{eq} = \mu_L^{eq} = \mu_L^* + kT \ln a_{eq}$$

$$\text{Equation 5.} \quad \Delta_\mu = kT \ln \frac{a}{a_{eq}}$$

where $\mu_L$ is the chemical potential of solution, $\mu_S$ is the chemical potential of crystal compound, k is the Boltzmann constant, T is the temperature, a is the activity and eq refers to equilibrium.

Hence, thermodynamics play a pivotal role as the primary driving force behind crystallisation. By comprehending the thermodynamic properties of the crystallisation system, it is possible to select the optimal crystallisation mode to effectively control yield and recovery. This understanding enables informed decisions regarding the choice of crystallisation techniques to achieve desired outcomes in terms of product quality and process efficiency.

### 1.1.3. Kinetics

The formation of new crystal particles from a solution occurs initially via nucleation and can be done by two pathways, named primary and secondary nucleation. Once nucleation occurs the entropy of the solute is lowered due to the formation of more

organized crystal structures.[5] Once nucleation has occurred, the crystal nucleus can then grow and thus change size, shape and form[6].

Primary nucleation is the formation of new crystal nuclei from a supersaturated solution and can be either homogenous or heterogeneous. Homogenous primary nucleation is entirely dependent on the transition from a system at equilibrium to one that is supersaturated. Heterogenous primary nucleation occurs due to the presence of fine foreign material e.g., dust, impurities or blemishes on the wall of the reactor.[7]
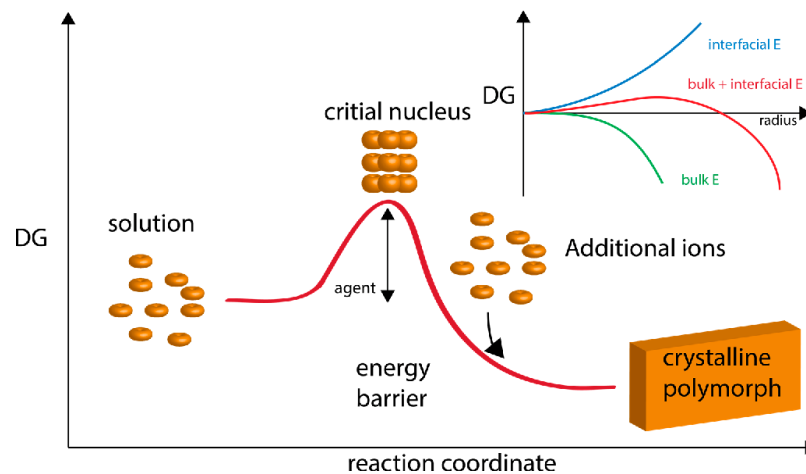


Figure 3. A schematic showing classical nucleation theory, taken from reference. [8]

The phenomenon of primary nucleation has been extensively investigated and has aimed to be described by a range of theories. Classical nucleation theory (CNT), as illustrated in Figure 3 provides a framework for understanding how molecules aggregate into clusters until they reach a critical size at which they can adopt the crystalline form and exhibit ordered geometry.[5] Assuming that nucleation takes place with the formation of a spherical nucleus, Equation 6 can be derived by differentiating the change in Gibbs free energy (ΔG) with respect to the nucleus radius. The resulting equation represents the nucleation rate, expressed in the Arrhenius form.[7]

*Equation 6.* $\quad J = A \, exp \left[ -\dfrac{16 \, \pi \, \gamma^3 v^2}{3 \, K_B^3 \, T^3 (ln \, S)^2} \right]$

where J is the homogenous nucleation rate, A is a pre-exponential factor, γ is interfacial tension, n is the single molecule volume, $K_B$ is the Boltzmann constant, T is the temperature and S is the supersaturation. In the context of this thesis,

Equation 6 shows a negative exponential relationship between nucleation kinetics and supersaturation and temperature.

The derivation showed that the radius of the nucleus cluster is critical in passing the barrier of nucleation and some studies have shown it can be between 10 – 1000 molecules.[9] [10] As CNT has been shown to not be widely applicable in practice, there have been other theories proposed such as the two-step mechanism which proposes that crystallisation proceeds through an intermediate stage.[11] This is graphically represented in Figure 4.
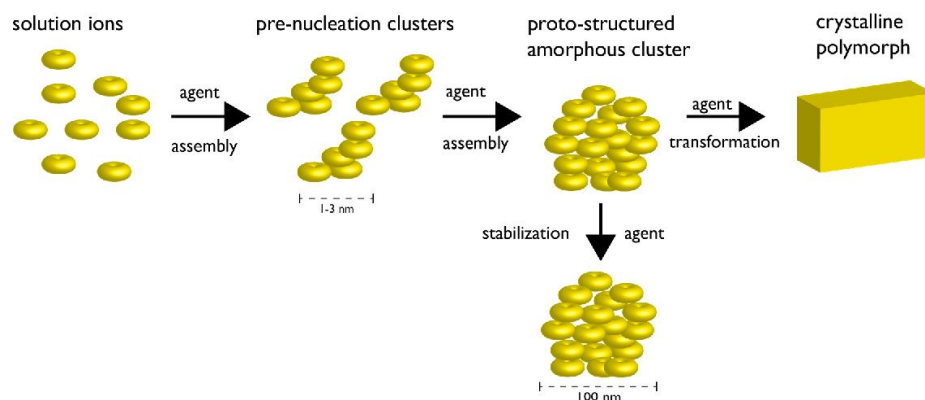


*Figure 4. A schematic showing non-classical nucleation theory, taken from reference [8]*

Secondary nucleation is the formation of new crystal nuclei caused by crystals already present in the system. Secondary nucleation can occur from many sources such as attrition, breakages and abrasion.[3] Collisions can be crystal-crystal, crystal-stirrer and crystal-vessel therefore choice over the stir rate, impellor type and vessel type are important considerations. Secondary nucleation occurs at much lower supersaturation than what is needed to cause primary nucleation. Often a system is seeded, which triggers secondary nucleation, and allows a crystallisation process to be carried out at a manageable supersaturation.

Crystal growth occurs by the addition of solute molecules onto the organized structure called the unit cell which then forms a lattice structure in a layer-by-layer method. Crystal growth can be controlled by diffusion of solute i.e., the movement of solute along concentration gradient from the bulk to the surface of the crystal, therefore an agitated solution would have a higher growth rate. Growth can also, and more commonly, be controlled by limiting surface integration i.e., the inclusion of molecules into the crystal lattice.[12] However high stir rates can cause more crystal

collisions therefore higher secondary nucleation which can lead to production of fines. Growth rate can be expressed as follows:

$$\text{Equation 7.} \quad R_{growth} = \frac{\delta x}{\delta t}$$

where $R_{growth}$ is the growth rate typically in µm/min, $\partial_x$ is the change in the size of the particle and $\partial_t$ is the change in the time taken.

The surface structure, typology and degree of solvation and the shape of the crystal can affect the growth due to steric hindrance and orientation for the surface integration of solute. Another phenomenon of note is Ostwald ripening where small crystals dissolve so as to be deposited onto larger particles.[13]

Fouling and oiling out can often be driven by kinetic factors and are deemed as unfavourable outcomes of a crystallisation process. Fouling can include encrustation where solid collects on the vessel walls whereas oiling out is where the solution splits into two layers rather than forming crystalline material.[14] These problems are often caused by a system being excessively supersaturated.

### 1.1.4. Solid state

Solid state chemistry, in relation to API crystallography, is the study of the structure of drugs in the solid phase i.e., the packing of molecules to form crystals. Understanding the solid-state properties of APIs is crucial as different solid forms can exhibit distinct physical and mechanical characteristics despite being chemically identical.



*Figure 5. Schematic of various solid-state forms (from left to right): polymorph form I, polymorph form II, amorphous, solvate, salt and co-crystal.*

Polymorphs refer to crystalline forms of drugs that have identical chemical compositions, but are capable of organizing their molecules into distinct crystalline structures, resulting in more than one unique crystalline state. As depicted in the first two diagrams of Figure 5, the same molecule can adopt different arrangements

to form distinct crystalline states, which may be influenced by factors such as solvent choice, seeding, and growth.[3] Consequently, polymorphism yields drugs with the same chemical composition but varying physical properties such as solubility, crystal habit, and melting point. [15] For instance, Ritonavir, an industrially synthesized drug, exhibits polymorphism wherein storage temperature changes can result in its conversion to form II, characterized by reduced solubility.[16] Thus, polymorphism in drugs presents significant challenges in ensuring their efficacy, safety, and quality, highlighting the need for a comprehensive understanding of this phenomenon. An example of a molecule with lots of polymorphs is flufenamic acid.[17]

Amorphous materials lack the organised lattice structure and are sometimes caused by fast precipitation or overprocessing, hindering the ability of molecules to orient into regular repeating long range ordered systems.[18] This disordered arrangement causes them to have increased solubility over crystalline materials as there are fewer interactions for the solvents to break down during dissolution.[19] Amorphous materials are generally avoided due to the degree of amorphousness being hard to control which can lead to drugs with varied efficacies.

Solvates are crystalline forms of drugs that, if formed in solution, can incorporate solvent molecules within the lattice structure. For example, drugs crystallised in water could form crystalline structures containing water molecules, these are named hydrates. Solvates are easily characterised and identified by differential scanning calorimetry (DSC) as the solvent is removed, often at a lower temperature than the API.[16] An example of a molecule with a lot of solvates is Axitinib.[20]

Salts are formed from the ionisation of the API during crystallisation and the generated charge is then balanced by a counter ion that binds within the lattice. Counter ions can come from the solvent and are formed due to an acid/ base reaction. Salts are often used in drug manufacturing as they enhance the solubility, dissolution and bioavailability of the API.[21] An example of an API marketed as different salt forms for different applications is morphine.[22]

Co-crystals are crystalline materials that contain two or more different molecules that exhibit the same structural shape, have a stoichiometric relationship and are uncharged. It is often the case that an API interacts with a non-API molecule, a coformer, by intermolecular forces such as hydrogen bonds, van-der-Waals or π- π

stacking.[16] Co-crystals are useful in drug manufacture as they improve the physiochemical properties of APIs, can be used as combination therapy, and also expand patent portfolios.[23-25] Some examples of co-crystal drugs currently on the market are Depakote, Entresto and Suglat.[26]

### 1.1.5. Crystallisation Methods

**Cooling Crystallisation:**
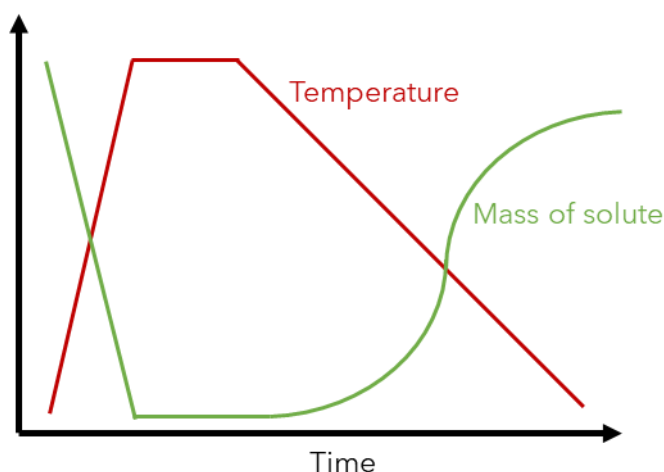


*Figure 6. Simple temperature profile representative of a cooling crystallisation.*

Cooling crystallisation as depicted in Figure 6 has the underlying theory that a solute will dissolve in solution upon heating and then when cooled will crystallise. The solute will remain in solution below the equilibrium solubility temperature, as it becomes supersaturated, up until a point where the driving force of $\Delta G$ allows for spontaneous nucleation to occur. The supersaturation increases as the temperature decreases, thus nucleation can dominate until the bulk of the API has crystallised. Well-designed cooling crystallisations give good control of the output quality but sometimes return a low yield.[27] However, scaled-down studies have shown 'well-behaved' cooling crystallisations with good yield.[28]

A study by Black *et al.*[27] highlighted four rules for a successful cooling crystallisation, with good recovery and minimal risk of impurity precipitation and fouling:

1. The elevated temperature should exceed 60 °C.
2. The solubility at the elevated temperature should be between 50 – 150 g/L.
3. The working temperature range should span over 60 °C.
4. The solubility at the low temperature should exceed 5 g/L.

It is important to control the rate of nucleation and the rate of growth to dictate end particle size and shape, where supersaturation is one of the most important factors (refer to Section 1.1.3.). In general: a faster cooling rate causes more supersaturation in the system which in turn leads to more nucleation and smaller particles and vice versa.

**Anti-Solvent Crystallisation:**
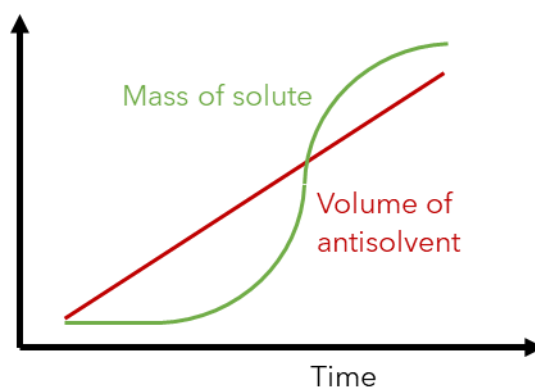


*Figure 7. Simple addition profile representative of an antisolvent crystallisation.*

The addition of an anti-solvent, that is miscible with the other crystallisation solvent and has low solubility for the API, will cause an increase in supersaturation and thus cause crystallisation to occur. Fast antisolvent addition can result in very high nucleation of meta-stable particles which increases the risk of fouling and oiling out.[27] Despite the potential quality and processing risks, anti-solvent crystallisation remains a popular method as it is fast and returns a high yield of crystalline products.[29]

**Evaporative Crystallisation:**
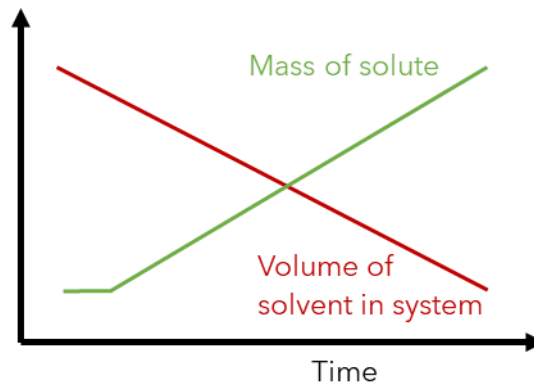


*Figure 8. Simple rate of solvent loss profile representative of an evaporative crystallisation.*

Evaporation of a solvent from a system that is at the equilibrium solubility point will result in a supersaturated solution which then has the potential to crystallise. As shown in Figure 8 there is a delay before crystallisation as the system builds supersaturation until crystallisation is feasible. This method is not often used in industry as the large time required and the high degree of impurity influence on the crystallisation product quality. However, for some systems that are not temperature dependent then this can be a preferred method. In general, evaporative crystallisation yields larger crystals than other methods as supersaturation is well controlled and slowly increases with solvent evaporation resulting in growth dominating instead of nucleation.

**Batch vs Continuous Crystallisation:**

Generally, crystallisation is performed as a batch process where reactants go in and products come out as distinct single steps.[30] This is a simple concept and subsequently easy to develop and implement but comes with high power surge usage and extended holding time required. Continuous crystallisation however has a constant feed of reactant and a constant output of product so it is deemed more efficient and has less dependency on human interaction.[31] [32] Examples of continuous crystallisers are oscillatory baffled reactors (OBR), tubular plug flow and mixed-suspension mixed-product removal (MSMPR).[33]

## 1.2. Quality by Design

Quality is critical in manufacturing to meet customer needs and expectations. It involves providing a functional product with a dependable, inexpensive, and rapid process, without any flaws.[34] Process development is key in achieving this across various industries, including pharmaceuticals[35], construction[36, 37], and engineering.[38] Joseph Moses Juran's pioneering work advocates forward-thinking in addressing quality issues.[39] Quality by Design (QbD) was introduced as a pre-emptive alternative to correctional methods. With the establishment of the Internet of Things (IoT)[40] and Industry 4.0[41, 42] the growing array of available industrial digital technologies provides a foundation for adopting Industry 5.0[43] principles. Industry 5.0 seeks to go beyond a focus on productivity and optimisation to ensure that the advanced technology and knowledge base enables processes to be more sustainable and resilient. A third key element is to ensure that the technology enables human operators/researchers to be more creative and effective in their roles, placing them at the heart of the manufacturing process. Quality by Digital Design (QbDD) is now possible as process and digital technologies have become more accessible to a wider demographic.
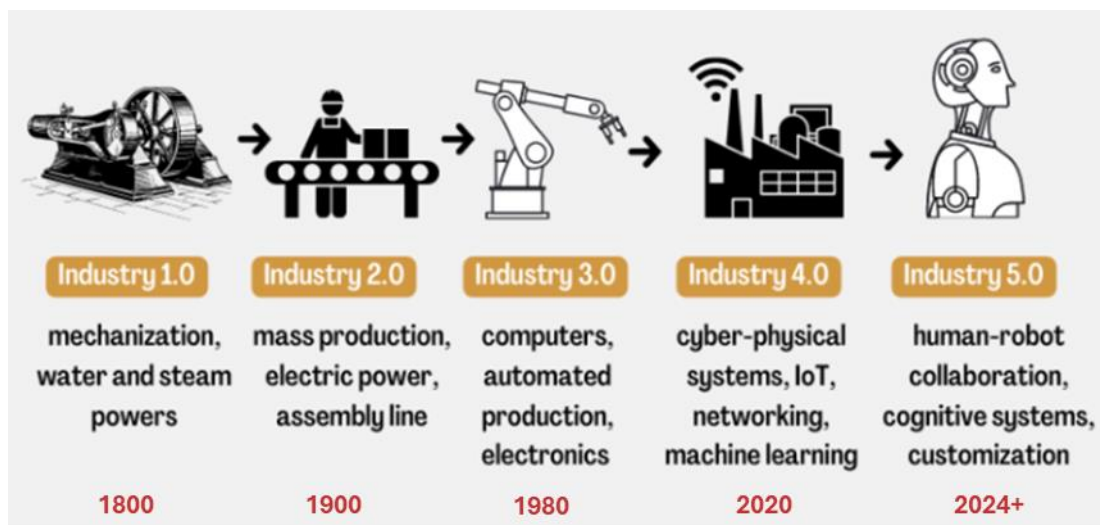


*Figure 9. Timeline of the various milestones within the industrial revolution, adapted from reference.[44]*

### 1.2.1. Quality by Design (QbD)

Quality by Design (QbD) is a systematic framework for ensuring quality in newly developed products.[45] Traditional QbD has been promoted by the US FDA[46] and involves designing the product by defining its target quality profile, identifying

critical quality attributes, assessing risks, establishing an experimental design space, implementing a control strategy, and continuously improving the process.[47] This step-wise sequence is illustrated in Figure 10. QbD has been successfully applied in various areas, such as alcohol synthesis[48], gel manufacturing[49], insulin spray drying[50], and oral drug development.[51]
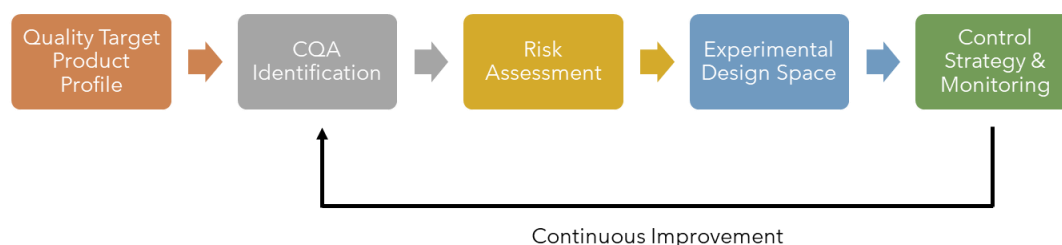


*Figure 10. Quality by Design framework.*

The Food and Drug Administration (FDA) has shifted from quality by testing to QbD in new drug development.[46] Furthermore the international conference of harmonization (ICH) Q8 emphasises the importance of a systematic approach to product development.[52] Despite the advantages of QbD, only a fraction of marketing applications submitted to the European Union from 2014 to 2019 were developed using QbD principles.[53] QbD implementation requires a deep understanding of complex guidance documentation and the submission of quality documents. Nonetheless, the adoption of QbD can lead to a traceable and rigorous method of product development, which ultimately benefits both the industry and the consumers.[52]

**Quality Target Product Profile (QTPP):**

The QTPP is documentation that defines the desired quality attributes or objectives of the product. It has been accepted by the FDA to be done first to allow for subsequent goal setting and transparency in the aims of the development.[54] Typical descriptors listed in the QTPP documentation include concentration, intended usage and purity to focus on efficacy and safety. Namjoshi *et al.*[47] highlighted many useful descriptors such as solubility, polymorphism and volatility in the study of topical creams.

**Critical Quality Attributes (CQA) Identification:**

Identification of the CQA and critical materials attributes (CMA) is the determination of the factors that may alter the process if left uncontrolled. CQA documentation holds a requirement to detail functional operating ranges for each critical parameter so that consistency can be achieved in output parameters such as crystal size, crystal shape and yield. CMA refers to the tolerance in quality of the raw materials. This stage is imperative and shows the importance of marginal gains in optimisation and reproducibility. [34, 47]

**Risk Assessment:**

To determine the critical quality attributes (CQA) that require control during the manufacturing process, a risk assessment is conducted by evaluating the severity and likelihood of potential damages.[55] Prior knowledge, literature, and screening reactions are used to assist with this step, which leads to the identification of critical process parameters (CPP) that must be controlled during manufacture in order to assure the CQAs are achieved.[56] Several methods, such as failure mode effect analysis (FMEA)[55], hazard analysis and critical control points (HACCP)[57], and statistical approaches[58], can be used for this risk assessment. [34, 52] Although these methods vary in complexity, HACCP is generally the preferred option due to its proven effectiveness in controlling process risks.

**Experimental Design Space:**

Experimental design spaces are essential for systematically exploring the interactions between variables. The International Conference on Harmonization (ICH) Q8 guidelines[52] detail regulations regarding design spaces (a multidimensional combination and interaction of process parameters and input variables), which once established must not be deviated from.[59] Traditional iterative methods, such as one variable at a time (OVAT), are conducted by changing one factor at a time while maintaining all other factors constant. However, this approach is time-consuming and wasteful for multivariate systems as it involves many experiments and can miss three-dimensional interactions. On the other hand, statistical methods such as design of experiment (DoE), which were developed as

tools for early work in agriculture[60][61], offer a more efficient way to explore design spaces.[62]
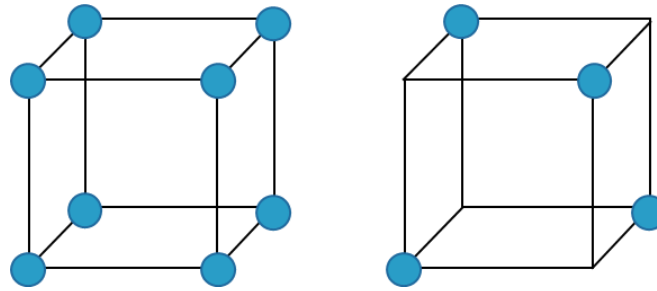


Figure 11. Schematic showing full factorial (left) and fractional factorial (right) design space.

DoE creates an experimental plan that aims to reduce the number of physical experiments while still testing all variables and their weighting on the process outcome. For example, factorial design dictates the number of experiments based on the number of increments and variables.[55] Fractional factorial design, which only executes a select few experimental points, is regarded as the most efficient DoE method. The difference between the factorial designs is illustrated in Figure 11. Response surface methodology (RSM)[63] is an adaptation of fractional factorial design that includes axial points and a zero point and offers an increased experimental scope due to the increase in data points. Taguchi design, is a more detailed method that involves four levels around each data point from a factorial design, making it a robust tool for optimization and reducing variation.[64]

When executing experiments, it is important to perform them randomly to eliminate trends caused by uncontrolled variables such as general degradation or equipment errors. However, if this leads to increased experimental time, blocking can be used where experiments are grouped by variable change. The aim of using these methods is to ensure that the results achieved are reproducible and can be validated.

**Control Strategy & Monitoring:**

Process Analytical Technology (PAT) is used for control strategy and continuous improvement of manufacturing processes.[65] In-line monitoring, integrated into the manufacturing process, offers real-time data without the wastage of sampling but requires expensive implementation and a large equipment footprint. Online

monitoring requires non-destructive analysis of sampled material, at-line requires physical removal of a sample for destructive analysis nearby, and offline is the traditional method of analysis in a centralized laboratory, despite being often destructive and inflexible for micro projects. Reviews[66] and audits[67] into PAT and the development of bench scale analysis tools are ongoing.[68]

PAT has been used for online monitoring of particle size, shape, and count in crystallisation and granulation through techniques such as focused beam reflectance measurement (FBRM).[69] Other PAT methods include acoustic resonance spectroscopy[70], air-coupled excitation vibrational analysis[71], and near-infrared spectroscopy[72] which have been used for tablet property analysis. In addition, laser-induced breakdown spectroscopy has been applied for the detection of salt in liquid phases.[73] In-situ optical microscopy has also been utilised for monitoring biomass concentration and system changes, albeit with a limited field of view.[74]

## 1.2.2. Quality by Digital Design (QbDD)

QbD has been a prominent framework in the pharmaceutical industry, emphasizing a systematic approach to ensure product quality. With the development of advanced digital technology and improved accessibility of models, such as robotic arms, powerful computation and data analysis software respectively, an extended version of QbD known as Quality by Digital Design (QbDD) has emerged. QbDD incorporates the integration of digital technologies into the QbD framework, which sits predominantly within the modelling and design space sections of Figure 12. This includes the utilisation of statistical or mechanical models developed through machine learning (ML), data mining, or simulation techniques. Moreover, the concept of digital twins, which are virtual replicas of physical processes or systems, can play a crucial role in scaling up the QbDD development framework.
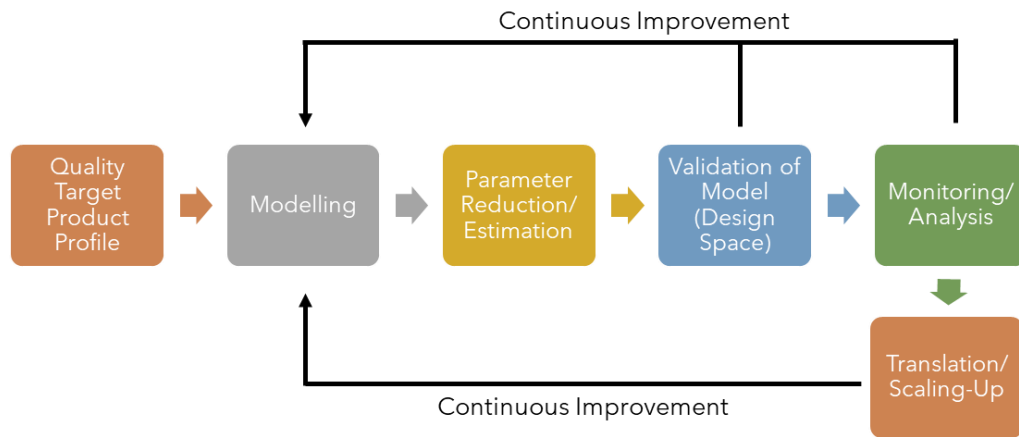
*Figure 12. Proposed Quality by Digital Design framework.*

One significant technological advancement that has facilitated QbDD is the Internet of Things (IoT)[40], which enables devices like PAT to collect and share data online and in real time, making information more accessible to gain insights and inform decisions. The IoT is part of the larger Industry 4.0[41] evolution, which aims to integrate cloud platforms, IoT, and Cyber-Physical Systems (CPS) to enhance manufacturing processes. CPS allows for the automation and control of manufacturing steps by computers, leading to a move towards autonomous design in manufacturing. This move has come in the form of the start of Industry 5.0.[43, 75]

## 1.3. Optimisation Algorithms

A key industrial digital technology is machine learning (ML), as referred to in Figure 9, which are mathematical/ algorithm-based models that are used to predict or classify based on prior knowledge. Training data can be collected from experimental measurements or literature sources to identify predictive patterns in complex data.[76] Optimisation algorithms are used to predict the best solution or best next experiment to a defined problem so as to either minimise or maximise the objective function. With the advancement of ML techniques, mathematical-based optimisation algorithms can be integrated into process development and applied to complex problems.

**Genetic Algorithm (GA):**

Genetic algorithms (GA) are based upon Darwin's theory of evolution in which the survival of the fittest occurred. GA is one of the first optimisation algorithms to employ a population-based approach and follows the main sequence of selection, reproduction, crossover and mutation.[77] [78] Potential solutions to the optimisation problem are referred to as chromosomes. Figure 13 details the line-by-line instruction for the algorithm.

---

**Input:**
    *Population Size, n*
    *Maximum number of iterations, MAX*
**Output:**
    *Global best solution, $Y_{bt}$*

---

**begin**
    *Generate initial population of n chromosomes $Y_i$ $(i=1,2,....,n)$*
    *Set iteration counter $t=0$*
    *Compute the fitness value of each chromosomes*
    **while** $(t < MAX)$
        *Select a pair of chromosomes from initial population based on fitness*
        *Apply crossover operation on selected pair with crossover probability*
        *Apply mutation on the offspring with mutation probability*
        *Replace old population with newly generated population*
        *Increment the current iteration $t$ by 1.*
    **end while**
    *return the best solution, $Y_{bt}$*
**end**

---

*Figure 13. Classical genetic algorithm sequence, taken from reference.[79]*

The selection step refers to the algorithm choosing the chromosomes based on their fitness and likeliness to be chosen as parents for the next iteration of the population, named a generation. These chromosomes are then reproduced to form a new iteration of the population. Crossover is the creation of offspring with combined genetic information from the parents, this introduces diversity into the population. Further diversity is caused by mutation, random changes to the genetics of the offspring, thus adding exploration to the optimisation search space.[78] GAs can handle large optimisation search spaces, can be used for many-objective problems and can be used for nonlinear objective problems. They have been successfully implemented in pharmaceutical distribution network design[80], facility layout design[81] and optimisation of a modelled crystallisation process.[82] Despite the advantages, GAs come with some limitations such as premature

convergence and the requirement for effective tuning of selection, crossover and mutation parameters.[79]

## Differential Evolution (DE):

Differential evolution (DE) belongs to the evolutionary class of optimisation algorithms and follows a sequence of mutation, crossover and selection.[83] [84] DE operates on a population of potential solutions to the optimisation problem, named vectors.
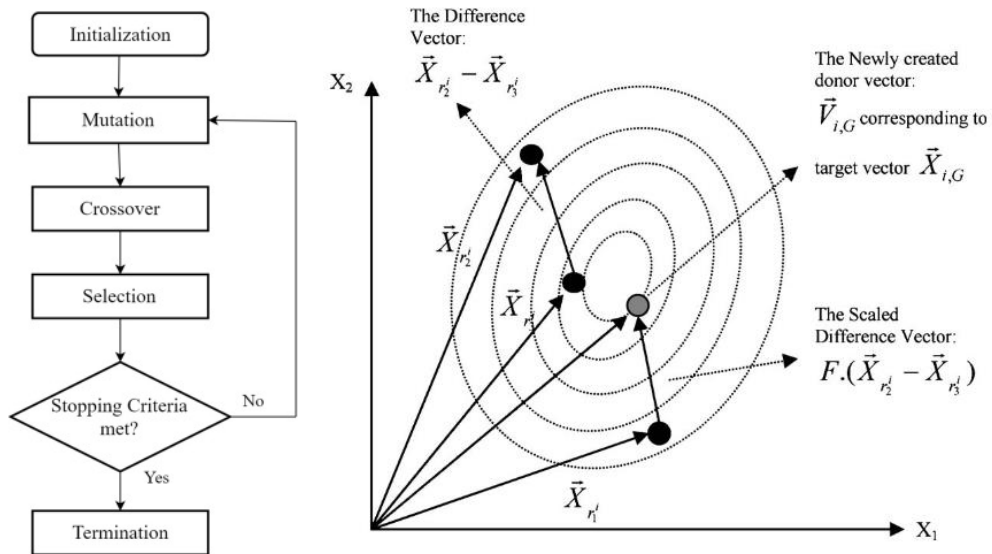


*Figure 14. Differential evolution sequence (left) and graphical representation of a DE mutation scheme within a 2D search space (right), taken from references.[85] [86]*

Mutation occurs by disrupting the vectors and adding a scaled difference vector between different potential solutions, illustrated in Figure 14, thus allowing diversity and exploration within the search space. Crossover then couples mutated and initial potential solutions. Selection is then done to compare and choose the best-fitted solutions. DE has similar advantages as GA but also exhibits robustness for convergence towards global optimisation.[86] DE algorithms have been applied to the medical industry in leukocyte segmentation in tissue images[87], drug synergy prediction[88] and drug product formulation.[89]

**Covariance Matrix Adaptation Evolution Strategy (CMA-ES):**

The covariance matrix adaption evolution strategy (CMA-ES) is an optimisation algorithm that is an evolution strategy classification and operates by maintaining and adapting a covariance matrix of the potential solutions.[90]
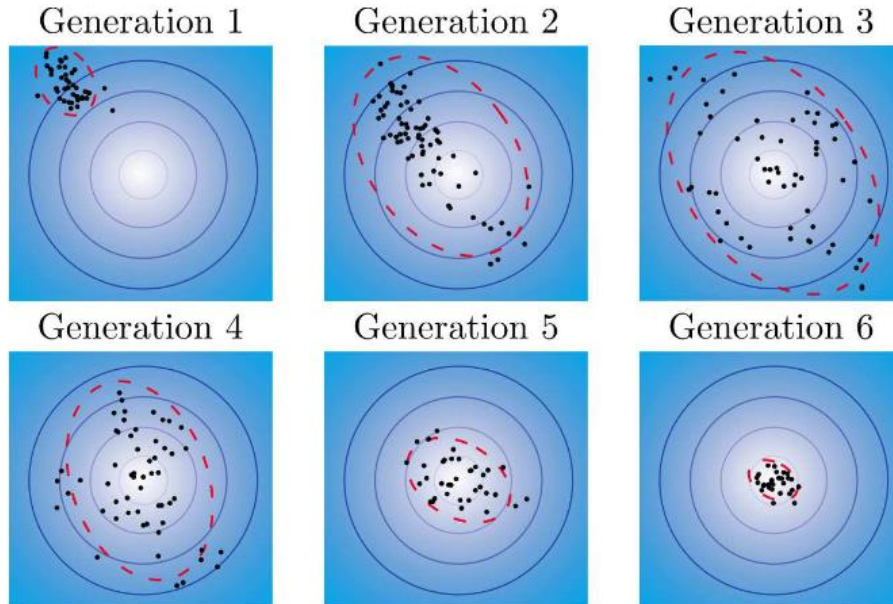


*Figure 15. Illustration of the eclipsing phenomenon of CMA-ES, taken from reference.[91]*

As shown in Figure 15 the algorithm dynamically adapts the sampling distribution and updates the covariance matrix to simultaneously explore potential solutions and converge towards a global optimum.[92] This is performed via the two mechanisms of generation and selection, based on a fitness ranking, of new potential solutions. CMA-ES is a robust, cheap and scalable algorithm that works well for multimodal and multi-objective problems.[92] The medical industry has benefited from CMA-ES algorithms with research on analytical chromatography[93] and radiographs.[94]

**Nelder-Mead:**

The Nelder-Mead algorithm is a direct search method that simply operates by iteratively transforming a simplex to explore the search space.[95] The simplex is a polygon of n+1 vertices in n dimensions. The potential solutions, at each vertex, are assessed and the worst one is replaced by updating the simplex following a series

of operations. Figure 16 shows how the simplex be altered by reflection, expansion, contraction and shrinkage.
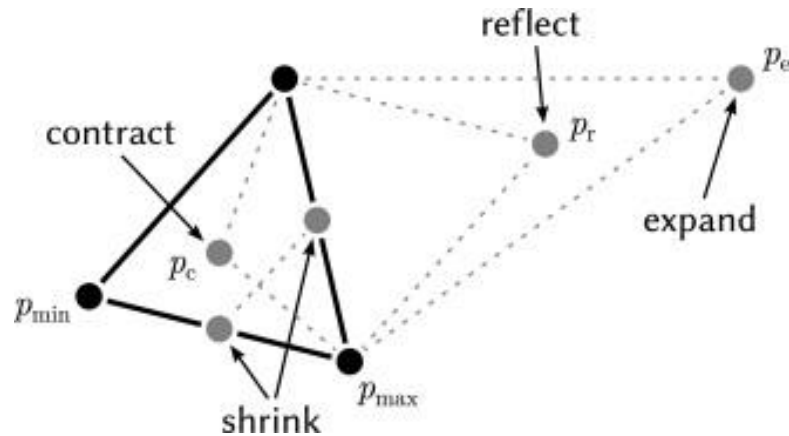


*Figure 16. Illustration of the Nelder-Mead simplex and the four operations, taken from reference.[96]*

Despite its simple and direct approach, Nelder-Mead has some limitations such as performance reduction for high-dimensional search spaces and the risk of convergence in local optima.[97] Nelder-Mead algorithms have been employed in medical research, where some examples are relating to microreactors[98], depression detection[99] and non-invasive drug bioavailability testing from saliva.[100]

**Pattern Search:**

Pattern search works by iteratively exploring the search space and evaluating the objective function at these points.[101] The direction of searching is dictated by a pattern, from a predefined series of vectors, where the algorithm follows the best potential solution. Two types of 'move' can be performed by the algorithm to allow for exploration and exploitation, as shown in Figure 17.
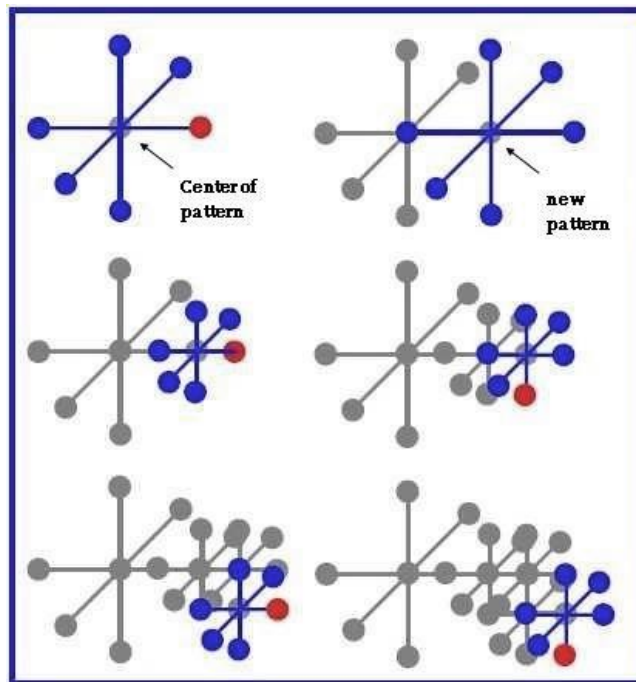
*Figure 17. Illustration of the Pattern Search algorithm, taken from reference.[102]*

Pattern search algorithms can optimise non-smooth and non-convex objective functions but are limited by their efficiency in convergence due to the behaviour associated with choosing the step size.[103] Examples of the use of pattern search in medical research include pharmaceutical size control in a batch dryer[104] and pharmacophoric study in drug design.[105] [106]

**Bayesian:**

Bayesian optimisation works by constructing a probabilistic model of the objective function and employs an acquisition function to iteratively suggest the next point of evaluation, or experiment, as illustrated in Figure 18. Gaussian processes are often used to construct the probabilistic surrogate model. The acquisition function looks to both reduce error and find the true optimum, therefore, there is a balance between exploration and exploitation. The most popular acquisition functions are maximum probability of improvement (MPI), expected improvement (EI) and upper confidence bound (UCB).
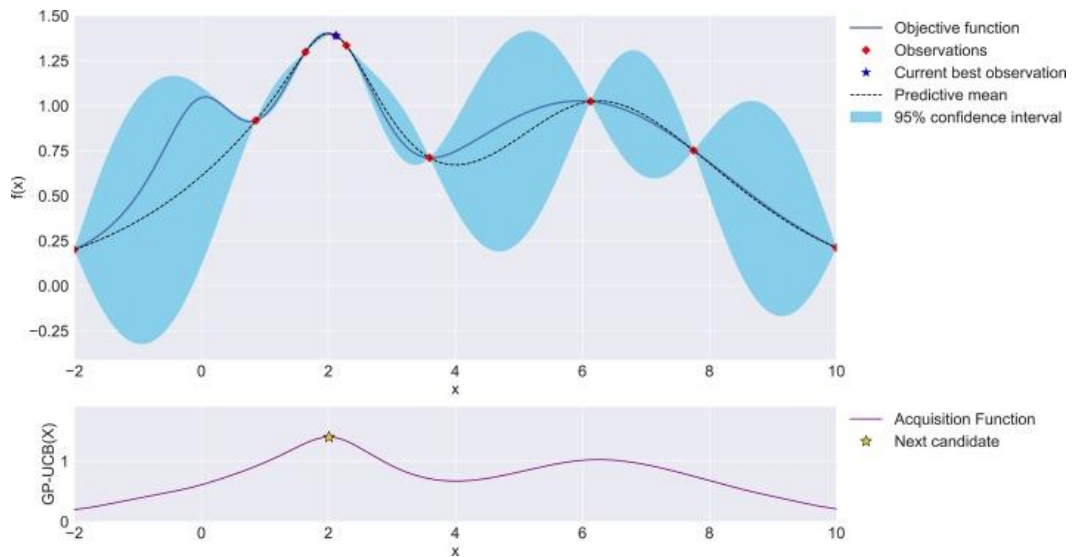
*Figure 18. Graphical illustration of Bayesian optimisation using an EI acquisition function, taken from reference.[107]*

Bayesian methods can be used to incorporate prior knowledge and uncertainty into an optimisation problem but generally lose performance at high dimensionality[108, 109]. Bayesian optimisation has been used across many research focuses and industries such as robotics[110], machine learning[111, 112] and healthcare[113, 114]. Previous studies[114] utilising Bayesian optimisation in pharmaceutical crystallisation are promising but additional study needs to be focused on adapting to multiple objectives.

## Other algorithms of interest:

Particle swarm optimisation and firefly algorithm, both swarm methods, work by mimicking the natural swarm behaviour of animals found in nature passing on shared knowledge. Simulated annealing, a trajectory-based algorithm, works by mimicking the physical process of where a material is heated and cooled to its original form in a lower energy state i.e., finding the system energy minimum.[115]

## 1.4. Robotic & Autonomous Systems

The integration of digital design into the final manufacturing stage is a rapidly evolving trend facilitated by CPS and its ability to control manufacturing processes, as highlighted in Figure 9. The incorporation of automation in operational

processes reduces reliance on human intervention, resulting in improved consistency and efficiency. However, the adoption of automation poses ethical considerations and requires substantial investment.

Burger *et al.*[116] introduced a robotic platform capable of identifying photocatalyst mixtures with significantly enhanced activity compared to previous materials. The robot exhibited autonomous mobility, operated machinery, accurately measured samples, and employed a Bayesian optimisation algorithm to explore the design space. By automating the design process, a broader range of variables could be considered, surpassing the limitations imposed by traditional DoE methods that typically restrict the scope to minimize physical experimentation.

There are many formats and vendors of robotics but generally, they can be classified as XYZ[117, 118], fixed-arm[119] or moving base[120]. There is also an introduction of collaborative robotics with force feedback allowing the safe operation in close proximity to human operators.[121] The use of XYZ robots and robotic arms have been used in drug development for nuclear magnetic resonance (NMR) and high-powered liquid chromatography (HPLC) dispensing,[122] solvent extraction[123] and prototype chemical synthesis.[124]

# 2. Aims and Objectives.

Standardised workflows have been widely used and developed within industrial sectors. The pharmaceutical industry has implemented its own workflows but this tends to be localised within individual companies and even sites within an organisation can collect data in different ways. Hence there is an absence of a standardised workflow accompanied with the prospect of a standardised crystallisation parameter database. Reliable, material reductive and fast collection of crystallisation data (such as solubility, fouling, nucleation and growth kinetics) will play a key role in cheaper and more efficient medicine manufacturing. This thesis aims to apply QbDD principles to intensify and accelerate a workflow for particle engineering and crystallisation process design that could generate such a database. This aim can be split into five objectives that align with the following individual research chapters:

**1. Developing a model-driven workflow for the digital design of small-scale batch cooling crystallisation with the antiviral lamivudine.** This chapter aims to take general lab practices, research methods and previous workflows and develop a step-by-step logical flow for use with any API. The main difference to other workflows[33] here will come two-fold: it will focus on unseeded crystallisation all at a small scale and include the introduction of model-driven experimental planning to accelerate experimental timelines. The main objective of this chapter is to take QbD principles (as set out in Section 1.2.1.) and advance them to QbDD principles (set out in Section 1.2.2.) to allow for predictive and quick and effective collection of crystallisation parameters (the importance of which was set out in Section 1.1.).

**2. High-throughput screening for large-scale data collection to inform medicine manufacture of aspirin.** This chapter aims to present another case study of using the workflow to collect both quantitative and qualitative thermodynamic and kinetic parameter estimation data.

**3. Adaptive Bayesian optimisation of process conditions for small-scale batch cooling crystallisation across fast and slow kinetic parameters: a comparative study.** This chapter aims to apply two Python optimisation libraries to address the highlighted need (from Chapters 4 and 5) for smarter model guided decision-

making experimental planning focused on crystallisation kinetic parameter estimation. The main objective of this chapter is to take various optimisation algorithms (outlined in Section 1.3.), apply them to a crystallisation problem and compare each one's relative performance.

**4. Utilisation of the model-driven workflow for the digital design of small-scale batch cooling crystallisation of a broader scope of the chemical space.** This chapter aims to take the latest image analysis algorithm accompanied by the best (as shown in Chapter 6) experimental planning optimisation algorithm to collect crystallisation data on 5 APIs. The main objective of this chapter is to show the generalisability of the workflow to allow for future work into robotic and autonomous integration of the logic (as highlighted in Section 1.4.).

**5. Integration of the model-driven workflow into an industrial pharmaceutical facility (Pfizer): Supporting Process Development of API Crystallisation.** This chapter aims to expand the scope of the workflow into seeded and antisolvent modes of crystallisation (outlined in Section 1.1.5.) and validate the workflow against scale-up experiments and previous industrial processes. This chapter will also show the integration of the workflow into an industrial setting (the importance of which was shown in Section 1.1.1.)

# 3. Methods and Equipment.

The CAD drawings used in this chapter can be found at <u>tpicks95/Chapter3 (github.com)</u>.

## 3.1. Crissy Platform (Zinsser Analytics)

The Crissy platform, more colloquially referred to as the Zinsser, is an XYZ robotic platform that is capable of vial handling, capping and decapping, powder handling, liquid dosing, heating and cooling, image collection and filtration. For this project, it was used for preparing vials to be removed from the Zinsser for analysis elsewhere.
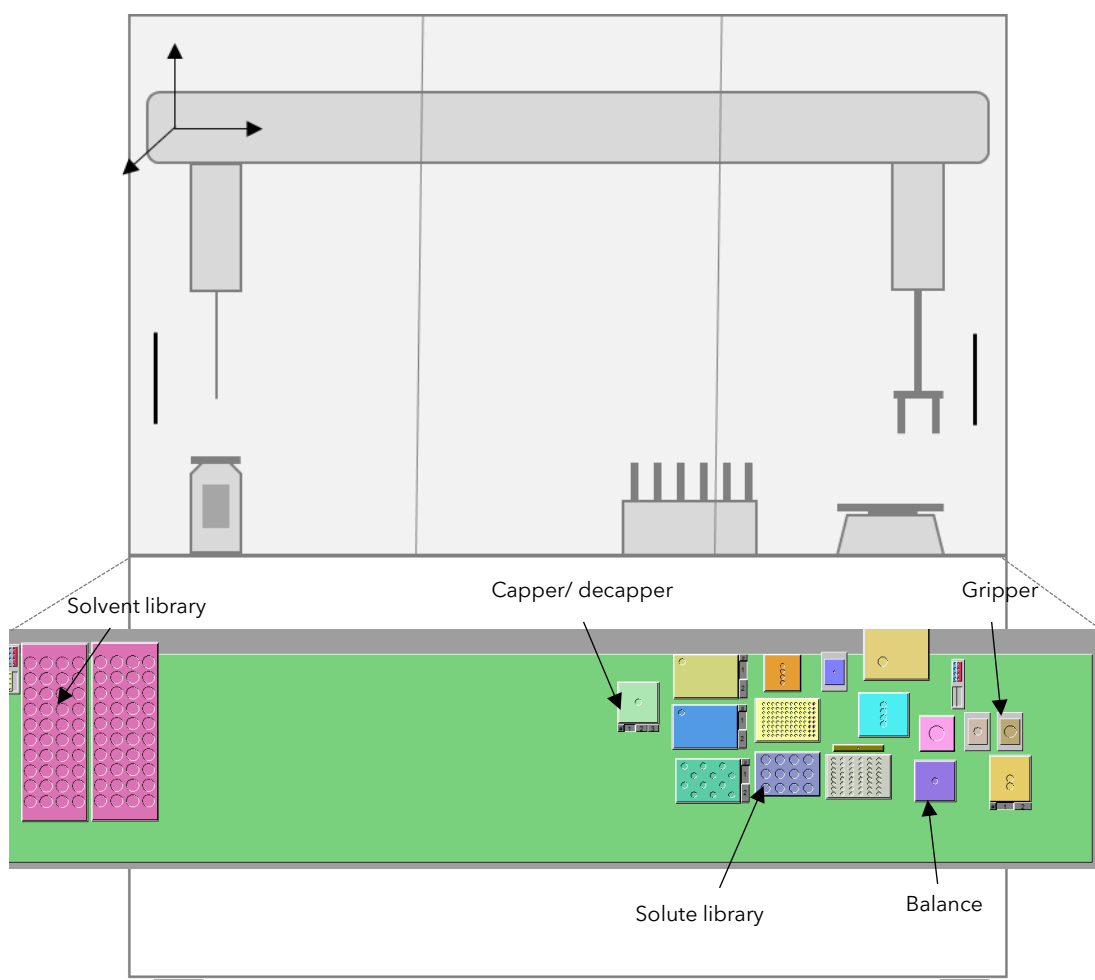


*Figure 19. Schematic diagram showing the Crissy Platform and the workspace layout for CrystallinePrep methodology (Zinsser Method Runner).*

The CrystallinePrep method (software) developed externally by the vendors for this project follows a stepwise sequence of instructions and issues relevant commands to the Crissy platform (hardware). A simplified sequence is as follows:

1. Initialise the solvent dosing arm (left) and powder handling arm (right) to starting positions.
2. Import vial dosing information from a csv file detailing what solute and solvent are required and the quantity of each. This information is stored temporarily in a SQL database.
3. The right-hand side arm picks up the gripper tool, allowing for a capped vial to be picked up, decapped and then moved to the balance.
4. The right-hand side arm picks up a powder pipette tip and doses the API into the vial. Dosing is volume based but as the mass is recorded the SQL database is automatically updated to provide information on the density of the powder, so each powder addition iteration improves until the required quantity is achieved.
5. The powder-dosed vial is moved to a holding vial.
6. The left-hand side arm, using a needle and syringe, doses the relevant solvent into the vial. Solvent addition is adjusted, from the SQL database, so that the initial required concentration is still achieved.
7. The solid and solvent dosed cap is returned to the balance to get the mass of solvent addition.
8. The vial is then recapped and then steps 3-8 of the sequence repeat for all other vials.
9. The final dosage data of each vial is returned as an export csv file.

## 3.2. Crystalline (Technobis)[125]

The Crystalline is a benchtop crystallisation reactor composed of eight individual chambers each with the ability to run its own program. Programs are designed by the user to set controllers for temperature, heating/ cooling rate, duration, sampling intervals and stirring speed. The glass vials used in the reactors have a maximum volume of 8 mL and a recommended working volume of 7 mL and can be equipped with multiple different types of lids. The lid types are as follows: Crystalline cap (standard), Crystalline cap (evaporation), Crystalline cap (solvent addition) and

Crystalline cap (basic). All caps except the basic cap utilise overhead stirring of either a short 3-blade impeller, long 3-blade impeller, long 3-blade double impeller or hook stirrer. Basic caps rely on magnetic PTFE stir bars in the bottom of the reactor vial.
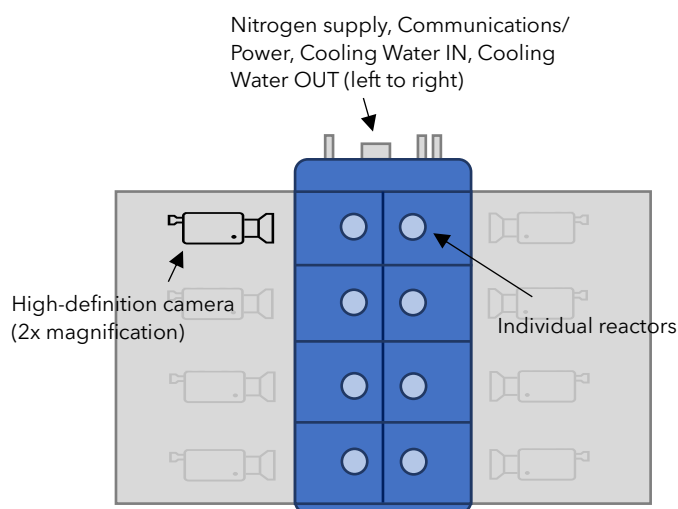


*Figure 20. Schematic diagram showing the Crystalline.*

Two main types of experiments were performed on the Crystalline equipment:

1. Polythermal – linear heating and cooling between a minimum and maximum temperature done in triplicate. The heating rate was either fixed or varied between repeated cycles.
2. Isothermal – heating to dissolution, followed by rapid cooling to the crystallisation temperature and then an extended hold at a specified temperature. These were done either in triplicate or quintuplicate.

Crystalline has in-built process analytical technology (PAT) capabilities for turbidity measurement and image collection. Segmentation image analysis is utilised to get particle shape and size data from each sampling point. In-house machine learning algorithms, based on a Detectron 2 model, have been developed to extract kinetic parameters from the isothermal experiments described above.

## 3.3. EasyMax (Mettler Toledo)[126]

The EasyMax is a bench-top stirred tank reactor (STR) that consists of two glass vessels with a working volume of 100 mL each. Each reactor can be used in parallel.

Experiments can either be set up using the control panel or using IControl software. Experiments are designed by the user to set controllers for temperature, heating/ cooling rate, duration and stirring speed. A temperature probe can be inserted into the reaction mixture to ensure accurate control between jacket temperature and solution temperature. A range of PAT probes, such as FBRM, IR, PVM and Blaze[127] can be inserted into the vessel. For this project Blaze was used.
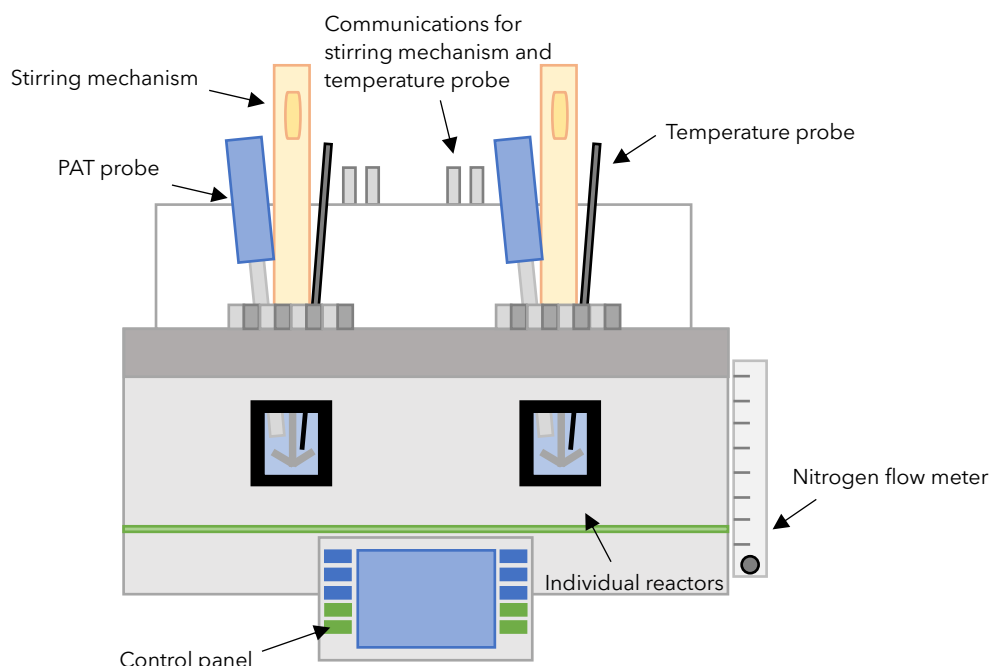


*Figure 21. Schematic diagram showing EasyMax.*

The two main experiments that were performed on the EasyMax equipment:

1. Isothermal – heating to dissolution, followed by rapid cooling to the crystallisation temperature and then an extended hold at a specified temperature. These were done in quintuplicate.
2. Cooling/ antisolvent hybrid crystallisation process – heating to dissolution, followed by slow cooling, the addition of seed, the addition of antisolvent and an extended low-temperature hold.

## 3.4. D8 Discover (Bruker)[128]

Powder pattern data was collected, for polymorph confirmation, using a Bruker D8 Discover diffractometer with the below experiment setup:

Table 2. Experimental set-up for XRPD in the Bruker D8 Discover.

| **Emission:** | |
|---|---|
| **Geometry** | Debye-Scherrer Transmission |
| **Source radiation** | Cu K$\alpha_{1,2}$ 1.540596 Å (line focus) |
| | |
| **Operating Voltage** | 40 kV |
| **Operating Current** | 40 mA |
| **Monochromator** | Primary focusing Goebel Mirror |
| **Optics** | 1 mm Anti-divergence slit<br>10.5 mm opening<br>2.5° Soller Slits |
| **Rotary absorber** | None |
| **Collimation** | None |
| **Sample:** | |
| **Treatment** | Lightly Ground |
| **Stage** | Transmission flat plate |
| **Rotation** | Sample averaging by xy displacement |
| **Sample Holder** | 40-well transmission flat plate with sample suspended on 7.5 microns Kapton film |
| **Temperature** | Ambient |
| **Anti Scatter Screen** | Beamstop |
| **Detection:** | |
| **Optics** | 2.5° Axial Soller Slits<br>Ni low-beta filter - 1.56% absorption<br>Secondary slit size 9.5mm<br>2.45° detector opening |
| **Detector** | SSD136_2 (1D) |

All powder patterns were visualised and analysed using DIFFRAC.EVA version 5.2.0.5 and then the raw data were exported as .xy files so that they could be plotted.

## 3.5. Robotic Integration of the DataFactory Platform

In the context of developing a robotic and autonomous platform for collecting crystallisation parameters, it was necessary to make specific modifications to hardware components to enable seamless physical hand-offs between different instruments.

### 3.5.1. 3D Printed Caps
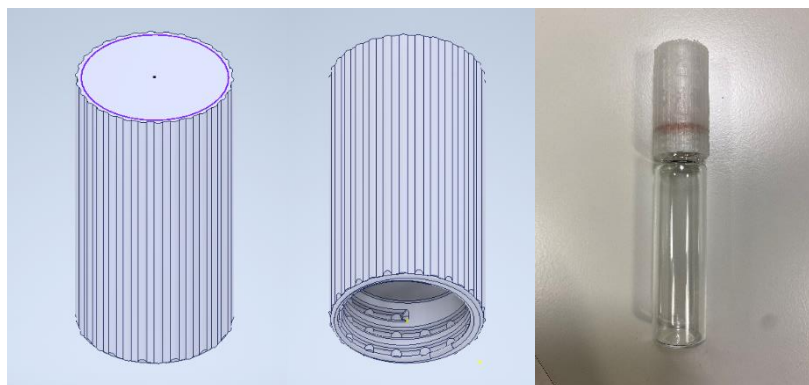


*Figure 22. Computer-aided design (CAD) of the top side and underside of the tall cap (left) and a picture of a vial equipped with a polypropylene printed cap (right).*

The Crystalline reactors are situated deep within the outer case, making extracting vials from the hardware challenging without requiring intricate dexterity. To address this issue, taller caps were designed and fabricated so that a Kuka cobot could remove them. The caps were constructed from polypropylene to ensure chemical compatibility and a small rubber seal was inserted into the top of the threaded section to minimise solvent loss. The caps were produced with an Ultimaker 2+ Connect 3D printer, utilising a 0.2 mm layer height, grid infill, and 20% infill, which required approximately 60 minutes per cap and incurred a raw material cost of £0.16 per cap.

### 3.5.2. Computer Numerical Control (CNC) Machined Gripper Fingers
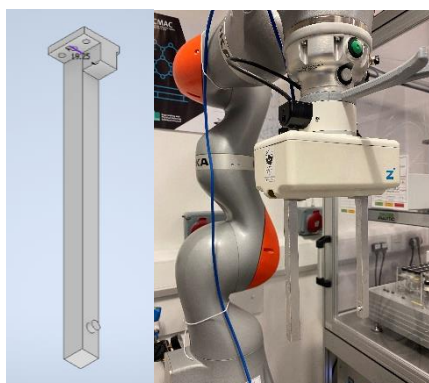


*Figure 23. CAD drawing of the gripper finger (left) and a picture showing the gripper attached to a Zimmer Gripper connected to the Kuka IIWA (right).*

The gripper fingers were developed to emulate the gripper mechanism of the Crissy platform (Zinsser) to facilitate the handling of identical trays and vials. To

achieve this, an external contractor (Psyber-Tech 3D Printing) was commissioned to utilise CNC machining technology to manufacture the grippers from aluminium. These grippers were subsequently integrated with the Kuka IIWA via a Zimmer motor, allowing them to pick and place entire trays of vials and individual vials using a specialized Zinsser tool.

# 4. Developing a model-driven workflow for the digital design of small-scale batch cooling crystallisation with the antiviral lamivudine.

## Abstract

We present a workflow that uses digital tools to optimise the experimental approach and maximise the efficiency in achieving the required process parameters for a desired set of crystallisation responses, kinetics and objectives. Model-driven small-scale experiments can contribute to reducing time and material waste in the development of pharmaceutical crystallisation processes. The workflow presented here guides the development of a small-scale batch cooling crystallisation process via solubility measurements, particle shape and size determination, form identification and preliminary kinetic parameter estimation to make crystals that satisfy quality target parameters (for shape, size and solubility) for a given active pharmaceutical ingredient (API). The case study herein follows the development of a crystallisation process for lamivudine, an API used in the preventative treatment of human immunodeficiency virus (HIV). This work identifies ethanol as a suitable solvent, meeting the acceptable solubility parameters for industrially relevant processes and yielded the biorelevant form, form I. The target kinetic parameters that were measured included induction time, growth rate and nucleation rate for lamivudine in ethanol under a range of conditions as guided by experimental planning models. Data was collected as part of the development of a DataFactory platform in which experimental optimisation can be autonomously implemented and all measurements stored in a crystallisation parameter database. This database will have further value in informing model development and continuous crystallisation process design and optimisation. The model objective-driven development workflow identified the following conditions, 19.9 °C, 600 RPM and supersaturation of 1.70, as achieving the desired objective successfully in 80

polythermal and 28 isothermal experiments. Integration of the workflow alongside the optimisation algorithm within the automated DataFactory system will enable fully autonomous, rapid data collection for small-scale API crystallisation. Such autonomous systems could play vital roles in pharmaceutical development and manufacturing driving towards more efficient and sustainable practices via digital transformation.

## 4.1. Introduction

Crystallisation is widely used as a fundamental purification step in the primary processing of active pharmaceutical ingredients (API) and many other molecular materials that also imparts the physical and bulk properties to the material relevant to subsequent processing[3]. The fundamental rate processes controlling crystallisation outcomes i.e. primary/secondary nucleation, crystal growth, agglomeration, phase transformations, and impurity rejection can all be measured using a range of established techniques[33,129] but remain difficult to predict for any given combination of API solute, solvent system, equipment geometry and process parameters. Hence process development can involve extended experimental efforts to select the composition and process conditions under which desirable outcomes can be achieved and transfer this across scales. Thus, approaches to enable rapid, optimised selection of conditions that provide the thermodynamic and kinetic control to achieve desirable size, shape and form outcomes early in the development cycle are of considerable interest.

Lamivudine, a nucleoside reverse transcriptase inhibitor used as an antiviral medication to treat and prevent human immunodeficiency virus (HIV) and hepatitis B virus (HBV)[130,131], is also being investigated as a potential treatment for COVID-19[132]. Lamivudine is currently produced using cooling crystallisation in methanol and (-) Binol[133]. There is limited prior literature describing the solubility of lamivudine and no kinetic parameters have been published. Past studies have shown that the dissolution rate of lamivudine is not substantially affected by changes in pH and that, subsequently, lamivudine exhibits good bioavailability in clinical usage[134]. Good bioavailability is determined by the biopharmaceutical classification system (BCS) assignment of group III, noting it is on the border of group I for permeability[135]. Lamivudine has been shown to produce three crystalline forms, form I, which is a 0.2 hydrate that exhibits needle geometry, and two anhydrous forms, form II, and III which typically adopt bipyramidal[136] and needle morphologies, respectively[137]. Existing patents state that form I and form II are preferred in solid oral dosing[138]. Therefore, the objective of the crystallisation is the purification and management of physical attributes to ease downstream processing.

The DataFactory project at CMAC, The University of Strathclyde, involves the development of an autonomous data collection platform primarily focusing on API solubility and kinetic parameters. The work presented here lays the foundation for a robotic workflow that can be carried out without human supervision. The core experimental process will follow the steps discussed in this paper but be carried out robotically, i.e. by a Kuka KMR iiwa similar to the one used by Burger *et al*[116]. When autonomous around-the-clock experiments are possible, the DataFactory will have the capacity to output data up to four times faster than human-led laboratory work due to operational hours increasing from 8-12 hours per weekday to 24 hours every day. A crystallisation parameter database is being developed to support the development of a suite of predictive tools in the form of a crystallisation classification system (CCS) that will accelerate the development of robust, sustainable crystallisation processes. This platform aligns with the wider need for more structured data management and curation adhering to findable, accessible, interoperable, and reusable (FAIR)[139] data principles to support digital transformation in pharmaceutical and other process industries.

In this work, the target workflow builds on the stages and guidelines of previous process development workflows for pharmaceutical processes[33]. This workflow integrates predictive models, automation and robotics to improve overall experimental efficiency and speed. The workflow also establishes and tests the logic necessary to enable an autonomous robotic data-collection platform to undertake the key tasks towards the selected experimental objective. Such autonomous platforms have the potential to accelerate development timescales and allow rapid process design[140]. Autonomous development systems that can intelligently vary experimental conditions to achieve a target outcome based on real-time data also promote sustainability by reducing human error, minimizing waste and optimising resource utilisation, contributing towards a more efficient and sustainable pharmaceutical manufacturing sector[141].

## 4.2. Experimental

In this study, images taken during the crystallisation experiments were collected using Technobis's Crystalline Reactor system at a rate of one image every five

seconds. Each of the eight reactor vials heats and cools following its own temperature profile, and every reactor has an HD camera focused on the lower part of the vial. X-ray powder diffraction (XRPD) patterns were collected using a Bruker D8 Discover, and the data was visualized using DIFFRAC.EVA[142] software from Bruker. Thermogravimetric analysis (TGA) and differential scanning calorimetry (DSC) were performed simultaneously using a Netszch STA 449 F1 Jupiter and data was analysed using Proteus Analysis[143]. Solid and solvent dosing was carried out using a Zinsser Crissy GGXXL robotic platform, which recorded solid and liquid dosed masses with a precision of +/- 0.005 mg. The Zinsser platform was used to dose amounts in the range of 5 to 1000 mg of solid and 1 to 8 mL of solvent. Raman scans were collected using a Tornado Focused Non-Contact probe from a stirred suspension in 8 mL glass vials.

The optimisation of crystallisation parameters to achieve the desired outcome was conducted in Modde 12.1 (Sartorius)[144] following a full factorial design of experiment (DoE) plan with a multiple linear regression model (MLR). This model allows a 3-dimensional optimisation space *i.e.*, 2 objectives and a combination of input parameters. Additional optimisation objectives can be added as constraints but will not explicitly feature as objectives of the optimisation algorithm. The experimental section and subsequent optimisation were performed in tandem as batch offline iterations. The initial DoE plan had high exploration across the whole design space and covered the corners, edges and centre points. Iteratively, the DoE experimental coverage got smaller to focus on exploiting the true optimum. Termination criteria for the optimisation loop was +/- 0.01 supersaturation and +/- 1 °C between MLR recommended experiments.

## 4.3. Workflow

In this section, each stage of the workflow, corresponding to the boxes in Figure 24, is described in detail. The challenges associated with the application of this workflow are also discussed for each step.
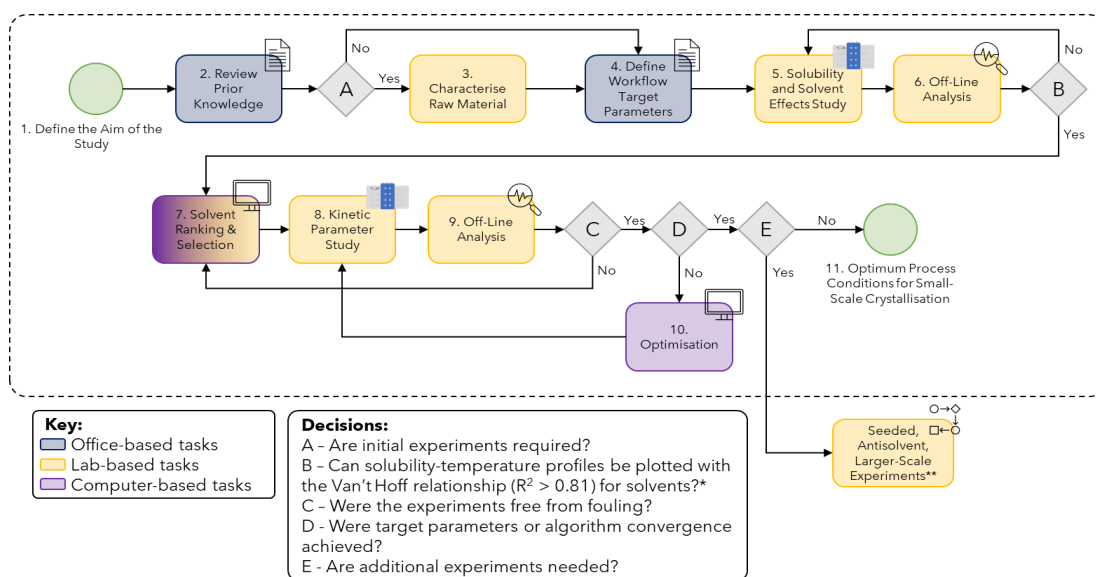
Figure 24. Workflow for small-scale batch cooling crystallisation data collection (for more details refer to Table 5 in the Appendix of Chapter 4).

*Minimum of 3-4 data points performed in triplicate must be used for reliable estimates of $R^2$ values. For some solvent systems, this may not be achievable as qualitative solubility can still be used to eliminate potential solvents.

**Although not explicitly discussed in the case study of this paper, if the target parameters for the study are not met then further study can be done to explore seeded, antisolvent and larger-scale crystallisation.

### 4.3.1. General Considerations and Challenges

The following section discusses the challenges associated with the generalisability of each step of the workflow. Each subsection corresponds to a step in the workflow.

### 4.3.1.1. Define the Aim of the Study

For the data collected using this workflow to be of use in the establishment of a CCS with broad applicability, the experiments need to span a varied chemical space in terms of physical and molecular descriptors. Consideration of scale and scope of experiments would benefit from standardisation so consistent information types and quality can inform future model development.

### 4.3.1.2. Review Prior Knowledge

Crystallisation and solubility databases are not freely available and/or do not present sufficiently consistent data across a wide enough chemical and process space, thus limiting their utility in this context. Reference solubility data may be very

limited and/or difficult to gather. Furthermore, APIs passing through the workflow may be newly developed or legally embargoed and thus very limited prior data may be available particularly if this workflow was applied to a commercial setting. FAIR[139] data principles should be adopted during data collection, storage and curation to enable the value of all data collected to be realised during and after the primary project.

### 4.3.1.3. Characterise Raw Material

It is important to understand the characteristics of the raw material before any processing to allow for referencing after crystallisation experiments have been conducted. Choosing the correct analytical methods for raw material analysis poses unique challenges depending on what equipment is available to the researcher.

### 4.3.1.4. Define Workflow-Specific Target Parameters

Target solubility[145], shape and size[146] values can be estimated using developability[147] and manufacturability[148] considerations for different APIs. The target kinetic values, including nucleation and growth, needed to achieve these outcomes are not available without more detailed knowledge of the rate processes involved for each system. Another challenge associated with setting target parameters is determining the relative importance of each parameter – particularly when introducing data filters and machine learning (ML) models. The relative importance of parameters is needed as these parameters will need to be weighted accordingly in multi-objective optimisation.

### 4.3.1.5. Solubility and Solvent Effects Study (Polythermal Global Search)

Numerical values for solubility cannot always be measured due to equipment upper and lower constraints when dosing very high and low concentrations, *e.g.,* the lowest concentration dose possible may still not dissolve at the highest temperature. In this case, qualitative solubility observations, for example, either anti-solvent or too soluble, are sufficient for decision-making.

If prior solubility data is not available for a given system, initial experiments are conducted systematically or guided by predictive models[149]. When conducting experiments systematically, exploring concentrations at the lower and upper bounds of the target range allows us to generate qualitative solubility classifications

that guide subsequent experimental concentrations. For some APIs, the workflow-specific target parameters may not be met using a single solvent from the library chosen. In this scenario, a binary solvent screen is recommended as intermolecular interactions in binary solvent systems can alter molecular solubilities.[150]

To assess the uncertainty of the solubility data at this point, three tests could be performed:

1. Was the experiment performed in triplicate to estimate the average mean data point?
2. Does the data fit to the Van't Hoff relationship with a $R^2$ exceeding a set point?
3. Was the heating rate appropriate for the difference in isothermal and kinetic solubility data collection?

### 4.3.1.6. Off-Line Analysis

XRPD is recommended as the main technique for polymorph determination as it is the gold standard for crystallisation polymorph fingerprinting and can identify amounts (typically down to 5%) of physical impurities[151]. An alternative and widely available fingerprinting technique for polymorph determination is Raman spectroscopy which can provide information on solid form changes as well as yield. While both techniques are non-destructive, Raman can also be implemented in situ via optical probes removing the need for sampling. At the same time, the application of Raman can be limited by low signal-to-noise ratio and interference from solvent signal and fluorescence[152]. In some cases, the desired polymorph may not have been isolated at this stage. If this is the case, it may be necessary to revisit the solvent library and perform a wider form screen with the inclusion of binary solvent systems.

### 4.3.1.7. Solvent Ranking & Selection

This step is typically a manual human intervention step, and, therefore, poses challenges for autonomous implementation. Useful solvent selection tools are available to aid in this key decision step. These tools include, for example, the GSK Solvent Selection Guide 2009[153] which classifies solvents as having one of the following: few issues, some issues or major issues (in terms of solvent favourability

for use in industrial processes). These classifications are determined using physical properties, mainly melting point and boiling point, and a ranking of waste, environmental, health, safety, stability and life cycle impacts. The solvent ranking and selection step is important for the goal of the workflow as it can allow for the identification of a greener and easier-to-process solvent. Ultimately, experiments in this step could be limited to only those required for refining solubility predictions across a wide range of solvents for the API of interest.

### 4.3.1.8. Kinetic Parameter Study (Isothermal Local Search)

As the goal of this step is to determine estimates of the kinetic responses under representative process conditions, several approaches can be used to derive these from small-scale experiments[154].

The reliable measurement of the induction time of a solute in a given solution also requires multiple data points in order to properly sample the probability distribution[154]. The number of experiments and/or cycles required to address this challenge can therefore be expensive in terms of time and data storage. In this study, five cycles were used to reduce measurement uncertainty. Fewer cycles would have resulted in a higher degree of uncertainty while more cycles would have risked exceeding the computational and temporary data storage limits available for a given experiment.

### 4.3.1.9. Off-Line Analysis

Refer to Section 4.3.1.6. If unexpected fouling has occurred at this stage, certain countermeasures such as altering heating rate, cooling rate, dissolution temperature and solvent choice can be explored.[155]

### 4.3.1.10. Optimisation

Selecting the optimal composition and operating conditions to achieve desired crystallisation outcomes with respect to yield, particle size, form and/or purity involves optimisation of several different factors and associated uncertainties. At this stage, it needs to be decided whether single or multi-objective optimisation will suffice for the aim of the workflow, as both come with separate challenges. Single factor, single response optimisation is straightforward to implement but does not explore the impact of latent variables (such as how both temperature and

supersaturation alter the growth rate simultaneously) or conflicting optimisation objectives (such as how induction time decreases as nucleation rate increases with supersaturation). On the other hand, in multi-objective optimisation, each target parameter could be an optimisation objective resulting in the challenges and complexities of a multi-dimensional optimisation problem. Going above 20 dimensions is often regarded as detrimental to many optimisation models[108,109]. Multi-objective optimisation problems come with the additional requirement of data normalisation and the additional challenge of weighting target parameters appropriately to align to the aim of the workflow. If this optimisation approach does not achieve the desired shape, size and/or form objectives, then seeded-cooling and antisolvent crystallisation routes should be explored[33] (these methods are beyond the scope of the current case study and workflow).

## 4.4. Results & Discussion: Workflow Case Study of Lamivudine

In this section the individual steps of the workflow, as described in Table 5 of the Appendix of Chapter 4, are presented using lamivudine as a test case. The sections for the workflow case study are named 4.4.2.X where X refers to a specific box or decision from Figure 24.

### 4.4.1. Materials

Lamivudine (90 g, CAS ID: 134678-17-4) was purchased from Molekula Ltd as an off-white powder and Form II. Lamivudine is a suspected teratogen. Therefore, dosing was performed in a fume hood and samples were transferred in capped vials. The solvents used for the cooling crystallisation were purchased from Sigma Aldrich, Fisher Scientific, VWR, and Alfa Aesar or prepared in-house (deionised water).
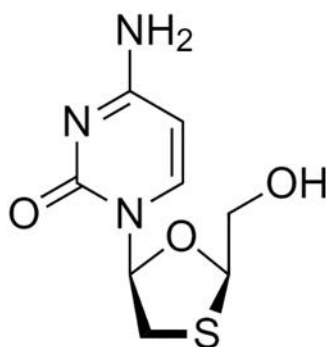
*Figure 25. Molecular structure of lamivudine.*

Lamivudine was chosen as the API of interest due to its useful pharmaceutical applications, and, in particular, its antiviral properties. There is also limited data in the literature for lamivudine solubility and crystallisation kinetics[156]. Contextual data was collated for lamivudine from the British National Formulary (BNF) and National Health Service of dictionary of medicines and devices (NHS dm+d)[157]. The BNF recommends a dosage of 150 mg twice daily when used as a preventative treatment for HIV and HBV. The BNF and NHS dm+d shows current suppliers include Teva, Alliance and Milpharm. The developability classification[147] of lamivudine was calculated using BASF Zoomlab[158] and assigned a classification of group I, meaning that solubility and permeability are not expected to influence bioavailability.

The solvent library was chosen using the International Conference on Harmonization (ICH) residual solvent guidelines, safety, health and environment (SHE) ranking, cost, molecular descriptors, a principal component analysis (PCA) study in the literature[159] and a clusterSim study in the literature[160]. Relative locations of solvents on a PC1 vs. PC2 plot[159] of 272 solvents were used to ensure the diversity of solvents. To do this, solvents were selected from all four quadrants of the plot published by Diorazio *et al*. An online tool, from the same study, allows users to visualise solvents in different ICH and SHE classifications in their relevant quadrant of the PCA. ClusterSim[160] was then used to identify solvents with similar properties for our study. The solvents chosen from the PCA tool[159] were selected using a multidimensional scaling (MDS) plot which also showed similar solvents that sit within the same cluster[160]. While we ensured a diverse group of solvents was present in our selected solvents for this work, also choosing solvents from the same

clusters allows us, and future users of the data provided, to compare experimental results within solvent families (i.e. alcohols) not just between solvent families to investigate any effects on crystallisation rate processes.

### 4.4.2. Key Workflow Tasks

### 4.4.2.1. Define the Aim of the Study

For this case study, the aim was to collect quantitative and qualitative solubility and solvent effects data and then subsequently optimise the small-scale batch cooling crystallisation process based on these results. The optimisation objectives were induction time and growth rate in response to supersaturation and isothermal temperature. Solvent choice in the optimisation was constrained with respect to solubility and shape and size parameters.

### 4.4.2.2. Review Prior Knowledge

Experimental data for lamivudine has previously been reported for a variety of techniques such as Raman[161], IR[161,162], DSC/TGA[161,162], NMR[130], SEM[162] and XRPD[156,163]. This data is used to validate our methods and confirm that the desired form is present. There are entries for each of the forms in the crystal structural database (CSD). The only lamivudine solubility publication available in the literature found by the authors was a study by Jozwiakowski *et al.* that showed that lamivudine has low solubility in most organic solvents except for ethanol, methanol and aqueous solvents.[156] The same study showed that crystallisation of lamivudine in methanol and water yielded needle habit.

Previous manufacturing routes consist of purification by crystallisation in methanol. This purification method allowed for the separation of the pharmaceutically active (-)-cis isomer[164].

### 4.4.2.A. Are initial experiments required?

Yes, analysis of raw materials was required (see step 4.4.2.3 below for details).

### 4.4.2.3. Characterise Raw Material

The experimental DSC data for the lamivudine raw material showed a single peak at a melting endotherm of 178.6 °C (see Figure 28 in the Appendix of Chapter 4), and the TGA data showed no mass loss during heating (Figure 29 in the Appendix

of Chapter 4). When compared to reference data, this data shows that the raw material is form II. XRPD also confirmed the presence of the most thermodynamically stable form, form II.[137] Later data collected in this study showed that Form I, on the other hand, had a crystallisation exotherm at 139.1 °C (see Figure 28 in the Appendix of Chapter 4) and melting endotherm at 178.8 °C with TGA data showing approximately 2% mass loss (Figure 29 in the Appendix of Chapter 4). This mass loss most likely corresponds to the removal of the 0.2 stoichiometric water molecule.

### 4.4.2.4. Define Workflow Specific Target Parameters

The target parameters for the solubility and solvent effects study were a lamivudine solubility of 0.005 g/g at low temperature (5 – 10 °C) and 0.05-0.25 g/g[145] at an elevated temperature (10 °C below boiling point) with an aspect ratio of above 0.5[146]. These physical constraints were chosen to ensure a well-suspended slurry (i.e., the elevated temperature concentration cannot be so high that a paste results at low temperatures), to promote a high crystallisation yield (i.e., the low temperature concentration should be sufficiently low to crystallise most of the API but not so low to promote potential inclusion of impurities)[145], and, finally, to encourage powder flowability of the resulting drug product (aspect ratio >0.5)[146]. At this stage of the workflow, yield and crystal shape are not optimised but, instead, used as constraints (e.g., only solvent systems that give crystals with an aspect ratio exceeding 0.5 are accepted) to subsequently rank possible solvent choices.

The optimisation objectives of the kinetic parameter estimation had target values of an induction time of 1 hour and a growth rate of 1 µm/min. These target parameter values correspond to typical values that lead to operable conditions in larger-scale crystallisation processes.[165] Additional parameter constraints were included in the kinetic parameter estimation including a d90 size distribution of 100 µm to 250 µm so that the resultant powder would be free-flowing [146].

From a bioavailability and drug product perspective, there was no targeted preference between form I and form II as the both forms have similar oral bioavailability and are both used in commercial formulations.[166] Form I and form III exhibited needle geometry, and thus would not satisfy the aspect ratio target parameter. Form III is also currently not used as drug product. Form II, therefore, is

the most desirable crystal due to increased thermal stability and flow properties for downstream processing.

### 4.4.2.5. Solubility and Solvent Effects Study (Polythermal Global Search)

Known masses of lamivudine were weighed into screw-top 8 mL vials with known volumes of solvent added automatically using the Zinsser Crissy platform, and a 10 mm PTFE magnetic stirrer bar was included in the vials prior to dosing. The vials were then placed in the Crystalline platform where each reactor had the following temperature profile applied:

1. Heat to 10 °C below the solvent boiling point or 90 °C (whichever is lowest) at a rate of 0.5 °C/min and hold for 10 minutes.
2. Cool to 5 °C at a rate of 0.5 °C/min and hold for 10 minutes.
3. Repeat the cycle 2 more times.

The heating rate in the above temperature profile was chosen to reduce the uncertainty of the dissolution temperature reported (refer to Figure 30 in the Appendix of Chapter 4). The maximum temperature was set to ensure that boiling solvents were not handled, thus making the experiment safe. The slow cooling rate was also used to reduce the error in determining the cloud point, and each sample vial underwent this temperature cycle three times to check for anomalous results. The temperature profile was the same for all experiments and the stir rate was fixed at 600 RPM throughout (refer to Figure 32 in the Appendix of Chapter 4 for why this stir rate was chosen).

The clear point, or the time at which full dissolution occurs, was defined as the temperature where no particles were observed in the collected images. In the literature, transmission data is used to identify the clear point[33]. Transmissivity, however, often gives a clear point below the true value as transmission reached 100% even when a few particles are still present in the images.

Similarly, the cloud point, or the time at which primary nucleation first occurs, was defined as the temperature at which particles were first observed in the collected images rather than when transmissivity goes below 100% or a similar threshold. The meta-stable zone width (MSZW) is the difference between the clear and cloud points as defined by the image data.
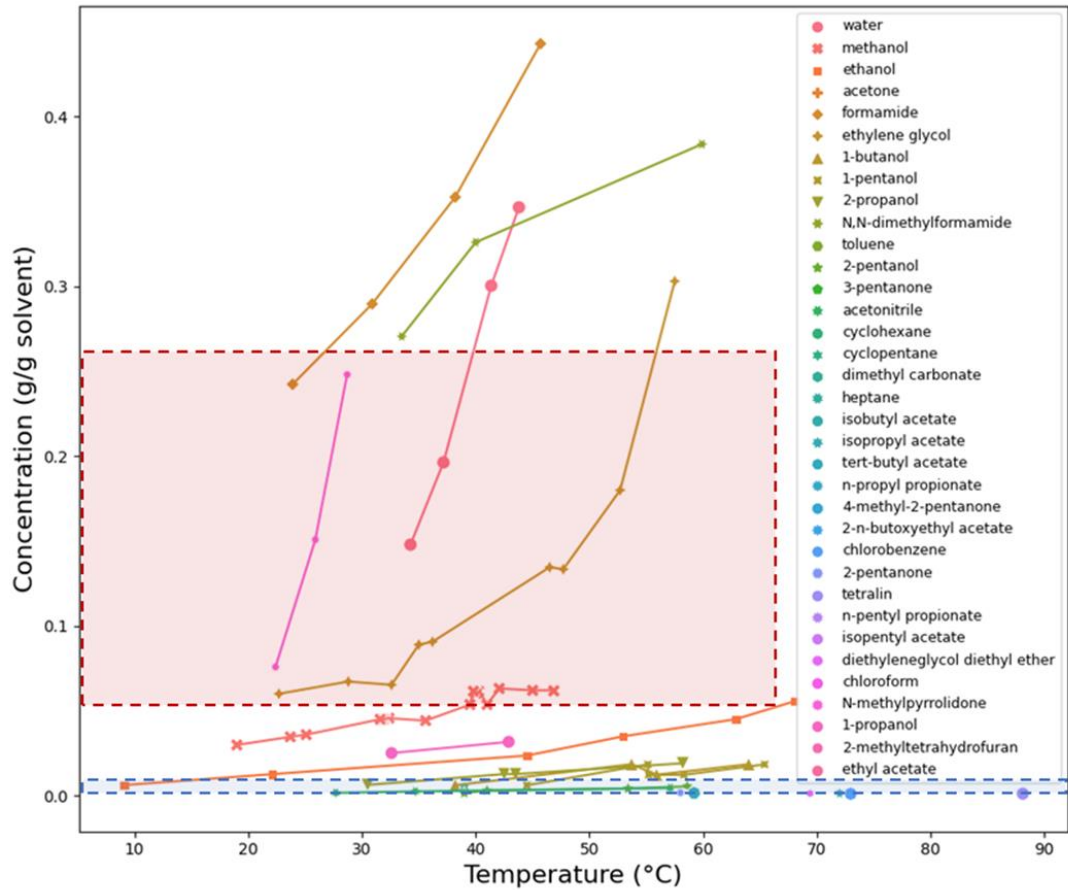
*Figure 26. The temperature-dependant solubility profile of lamivudine in 20 of the 31 solvents. The blue highlighted area shows solubility target parameters for low temperatures and the red highlighted area for high temperatures.*

The solubility and solvent effects study determined that 11 of the 31 solvents were antisolvents (lamivudine solubility < 0.005 g/g at elevated temperature) due to dissolution not being observed for vials with 0.005 g/g concentration of lamivudine. This concentration represents the lower limit of the Zinsser dosing platform's capabilities. The solubility profiles (Figure 26) of lamivudine in the remaining 20 solvents showed that in 4 solvents, lamivudine was too soluble (lamivudine solubility > 0.25 g/g at elevated temperature), in 13 solvents, lamivudine was not sufficiently soluble (lamivudine solubility < 0.005 g/g at elevated temperature) and in the 3 remaining solvents, the solubility of lamivudine satisfied the solubility constraints. These 3 solvents were ethylene glycol, methanol and ethanol.

### 4.4.2.6. Off-Line Analysis

All powder product from the crystallisation experiments was filtered, and then XRPD was done to confirm which form was present. By comparing collected data to reference data, powder patterns (Figure 31 in the Appendix of Chapter 4) showed that form II recrystallised from ethanol and isopentyl acetate. 1-pentanol and chlorobenzene recrystallised only trace amounts, and large glass-like particles that diffract poorly grew in formamide precluding form identification. Crystallisation in all other solvents yielded form I, the undesired 0.2 hydrate form. The 0.2 stoichiometric amount of water in form I is likely attributed to residual water in solvents.

Small differences were observed in the powder patterns for the crystals grown in methanol, water and 1-pentanol. TGA and Raman was used to confirm that form I was present.

No major fouling or agglomeration was observed under the crystallisation conditions tested in any of the solvents used with the exception of 4-methyl-2-pentanone in which aggregates of needle crystals formed.

### 4.4.2.B. Can solubility-temperature profiles be plotted with the Van't Hoff relationship ($R^2 > 0.81$) for solvents?

Yes, refer to Table 6 in the Appendix of Chapter 4. It is important to note at this stage that the thermodynamic model for lamivudine in the solvent systems is an initial estimate aligned to a defined uncertainty of repeating data points in triplicate and with a suitable fit of the data whilst using a kinetic/ dynamic heating rate.

### 4.4.2.7. Solvent Ranking & Selection

The 3 successful crystallisation solvents, in terms of solubility constraints (solubility of 0.005 g/g at low temperature and 0.05-0.25 g/g[145] at an elevated temperature), were methanol, ethanol and ethylene glycol. While crystallisation of lamivudine in ethylene glycol had a steep temperature dependence that would theoretically result in a high yield, ethylene glycol was still eliminated due to the solvent's high viscosity and consequently slow crystallisation kinetics. Nucleation was not

observed for lamivudine in ethylene glycol despite cooling more than 55 °C below the dissolution point (i.e., the supersaturation of lamivudine will greatly exceed the normal threshold for inducing nucleation). Methanol was also eliminated as crystals were form I with needle geometry, and the resulting aspect ratio did not satisfy the aspect ratio target value.

Ultimately out of 31 solvents evaluated for crystallisation of lamivudine, ethanol was the only system to satisfy the target parameters for shape and solubility. Specifically, these experiments showed that the form was II, the crystal shape was bipyramidal, and the solubility exceeded 50 g/L at elevated temperature meeting all the objectives for this step.

### 4.4.2.8. Kinetic Parameter Study (Isothermal Local Search) & 4.2.9. Off-Line Analysis

Similar to the solubility and solvent effects study, known masses of lamivudine were weighed into screw-top 8 mL vials with 2 mL volumes of ethanol added using the Zinsser platform, and a 10 mm PTFE magnetic stirrer bar was included in the vials prior to dosing. The solubility profile for lamivudine in ethanol from the solubility study was used to calculate the concentrations required for the relevant supersaturation. The vials were then placed in the Crystalline platform where each reactor had the following temperature profile:

1.  Heat to 68 °C at a rate of 1 °C/min and hold for 10 minutes.
2.  Cool to the experimental temperature at a rate of 10 °C/min with no stirring.
3.  Hold at the isothermal point for 6 hours.
4.  Repeat the cycle 2 more times to collect a total of 3 data points.

The fast heating rate in the above temperature profile was chosen given that the dissolution point did not need to be measured; it was only necessary to ensure that dissolution was achieved. The maximum temperature was set at 68 °C to ensure that ethanol did not reach its boiling point and that solvent loss from evaporation was minimised. A fast crash cooling rate was used to achieve the isothermal temperature of interest as fast as possible to reduce the risk of nucleation occurring during the cooling ramp. The isothermal regime was held for 6 hours to capture crystallisations with longer induction times. Each sample vial underwent this

temperature cycle three times to check for anomalous results and to get a median estimate for induction time. Induction times are reported as median values of these three measurements to reduce the impact of outliers. The temperature profile was the same for all experiments to minimise the variability across experiments. The stir rate was fixed at 600 RPM unless otherwise stated (refer to Figure 31 in the Appendix of Chapter 4).

Induction time, defined as the time between reaching isothermal conditions and the point at which crystals first start to nucleate, was determined using image data in this study, specifically using Helmli's mean. Helmli's mean[167] is an image feature used in computer vision that has been found to be a good indicator of the presence of particles in in-line microscopy[168]. Here, Helmli's mean is calculated with a window size of 5 pixels and thus termed $HELM_5$. Induction time is estimated by measuring the time from temperature crash to where $HELM_5$ rises above a threshold. Due to differences in particle habit, lighting, and solubilities, the threshold is not a fixed value but is chosen based on the range of $HELM_5$ values obtained in each experiment. The threshold $HELM_5$ value is taken as the $5^{th}$ percentile of the range of values $HELM_5$ obtained over the course of the experiment. Induction time estimation using $HELM_5$ has been shown to have improved accuracies over other approaches[168][169].

Finally, in this study image data was also used to determine the growth rate (the change in the particle size, over time). These parameters were determined using data spanning from slightly before the induction time (to account for potential errors in induction time estimate) until the time when the images became too crowded to identify single particles. Images were analysed to detect individual particles using the deep learning model "Mask R-CNN"[170]. This model was adept at identifying individual objects in crowded situations and at multiple scales making it well-suited to the task of particle characterisation[171]. Versions of the Mask R-CNN are available pre-trained on standard datasets which can be further trained ("fine-tuned") to apply to new tasks. This transfer learning reduces the training time for new applications.

200 images from the Crystalline platform were chosen from various experiments to form a training set (which included both form I and form II of lamivudine amongst

other APIs) based on their varied lighting conditions, particle habit, and particle solid loading. These images were manually annotated with annotations indicating the position and size of particles. These annotated images were supplied to a training algorithm to teach the Mask R-CNN model to detect particles on Crystalline images. The trained model was then applied to new images resulting in predicted particle locations, and a given confidence score. Predictions of confidence score less than 50% were discarded. The remaining particles were sized by fitting a rotated rectangle to the detected outline, minimising its area. The long side of this rectangle is referred to as particle length. Lengths were aggregated for an image and compiled in a histogram plot to determine the number-weighted-particle size distribution (PSD). Repeating across all images resulted in one PSD for every 5 seconds of the experiment during nucleation and growth. Quantiles and mean values of the PSDs were tracked over time. A linear fit was made to the mean size over time, and the gradient of the line gave an estimate of growth rate (see Table 3).

An initial full factorial design screening experimental plan[60] was created (Table 3). The experimental bounds on isothermal temperature and supersaturation (SS = C/C*, where SS is the supersaturation, C is the solution concentration and C* is the equilibrium concentration at the given temperature) were dictated only by physical limitations of solvent operating temperatures and industrial process solubility requirements to explore as large a design space as possible.

*Table 3. Process parameters and measured median induction time and mean growth rate for initial kinetic parameter screening of lamivudine in ethanol with a fitted MLR model. Standard deviations of 0.00 refer to where the image analysis only segmented one growth phase.*

| Run Order | Isothermal temperature (°C) | Supersaturation | Median induction time (s) | The standard deviation** of induction time (s) | Mean growth rate (µm/s)* | The standard deviation** of growth rate (µm/s) |
|---|---|---|---|---|---|---|
| | | | **Initial Screening** | | | |
| 1 | 10 | 1.93 | 2778 | 374 | 0.0355 | 0.0052 |
| 2 | 25 | 1.67 | 3495 | 1073 | 0.0001 | 0.0213 |
| 3 | 25 | 1.56 | 3737 | 1496 | 0.0153 | 0.0200 |
| 4 | 40 | 1.84 | 1585 | 1555 | 0.0185 | 0.0173 |
| 5 | 10 | 1.42 | 11380 | 2596 | 0.0263 | 0.0000 |
| 6 | 25 | 1.63 | 1591 | 1008 | 0.0287 | 0.0238 |
| 7 | 40 | 1.39 | 10017 | 8011 | | |
| | | **MLR Output: SS 1.70, Temp 28.5 °C** | | | | |

*Growth rate values are missing due to the image analysis being unable to detect particle growth likely due to the small sample size of images collected at the point of nucleation before the images became too convoluted.*

**Standard deviation reported is the uncertainty between the extracted parameters from each vial done in triplicate.*

An MLR model was fitted to the initial results and the optimizer feature in Modde 12.1 was run with the stated target values for induction time (1 hour) and growth rate (1 µm/min). The MLR model trained on the initial screening results returned process conditions of a supersaturation of 1.70 and an isothermal temperature of 28.5 °C. Similarly, a partial least squares (PLS) model was fitted to the data where an optimisation returned similar values of a supersaturation of 1.69 and an isothermal temperature of 29.5 °C.

### 4.4.2.C. Were the experiments free from fouling?

Yes, there was no significant fouling experimentally observed with ethanol as the chosen solvent system (visually observed in Section 4.4.2.8).

### 4.4.2.D. Were target parameters or algorithm convergence achieved?

No, there was no measurement at the MLR recommended supersaturation and temperature therefore experimental optimisation was required.

### 4.4.2.10. Optimisation

The subsequent (from the MLR model in Section 4.4.2.8) full factorial experimental plans had a higher focus on exploitation over exploration, and, therefore, an increasingly narrower range was used for the supersaturation (+/- 0.27 to 0.09) and isothermal temperature (+/- 15 to 2 °C). The experimental plan followed by the fitting of the MLR model was done iteratively. As there was minimal change in supersaturation and temperature values predicted by the third and fourth MLR models (+/- 0.01 for supersaturation +/- 1.0 °C for temperature), it was determined that convergence was achieved on the fourth MLR-trained model which returned process conditions of a supersaturation of 1.70 and a temperature of 19.9 °C.

*Table 4. Process parameters and measured median induction time and mean growth rate of kinetic parameter optimisation iterations of lamivudine in ethanol with fitted MLR models. Standard deviations of 0.00 refer to where the image analysis only segmented one growth phase.*

| Run Order | Isothermal temperature (°C) | Supersaturation | Median induction time (s) | The standard deviation of induction time (s) | Mean growth rate (µm/s)* | The standard deviation of growth rate (µm/s) |
|---|---|---|---|---|---|---|
| **Optimisation iteration 1** | | | | | | |
| 1 | 28.5 | 1.78 | 1524 | 81 | | |
| 2 | 30 | 1.77 | 1310 | 248 | | |
| 3 | 28.5 | 1.77 | 615 | 464 | 0.0186 | 0.0088 |
| 4 | 27 | 1.89 | 2000 | 374 | 0.0413 | 0.0309 |
| 5 | 28.5 | 1.8 | 2290 | 2023 | 0.0147 | 0.0000 |
| 6 | 30 | 1.68 | 4536 | 957 | 0.0740 | 0.0193 |
| 7 | 27 | 1.73 | 2460 | 895 | 0.0387 | 0.0000 |
| **MLR Output: SS 1.74, Temp 19.6 °C** | | | | | | |
| **Optimisation Iteration 2** | | | | | | |
| 1 | 19.5 | 1.75 | 3901 | 301 | 0.0875 | 0.0778 |
| 2 | 18 | 1.74 | 1050 | 199 | 0.0440 | 0.0000 |
| 3 | 18 | 1.64 | 2305 | 3314 | 0.0472 | 0.0414 |
| 4 | 19.5 | 1.7 | 1745 | 564 | 0.0136 | 0.0196 |
| 5 | 21 | 1.81 | 2715 | 1483 | 0.0497 | 0.0265 |
| 6 | 21 | 1.68 | 6080 | 7293 | 0.0287 | 0.0000 |
| 7 | 19.5 | 1.72 | 1070 | 458 | 0.0279 | 0.0030 |
| **MLR Output: SS 1.70, Temp 19.5 °C** | | | | | | |
| **Optimisation iteration 3** | | | | | | |
| 1 | 21.5 | 1.68 | 2999 | 2132 | 0.0053 | 0.0014 |
| 2 | 21.5 | 1.62 | 2021 | 470 | 0.0015 | 0.0007 |
| 3 | 17.5 | 1.77 | 1986 | 803 | 0.0059 | 0.0005 |
| 4 | 19.5 | 1.72 | 1787 | 367 | 0.0028 | 0.0007 |
| 5 | 19.5 | 1.74 | 1506 | 986 | 0.0030 | 0.0018 |
| 6 | 17.5 | 1.6 | 11552 | 2063 | 0.0020 | 0.0017 |
| 7 | 19.5 | 1.71 | 1320 | 3079 | 0.0089 | 0.0077 |
| **MLR Output: SS 1.70, Temp 19.9 °C** | | | | | | |

*Growth rate values are missing due to the image analysis being unable to detect particle growth likely due to the small sample size of images collected at the point of nucleation before the images became too convoluted.*

All samples from each iteration of the optimisation loop were fingerprinted by XRPD (Figure 33 in the Appendix of Chapter 4) where results confirmed that all runs gave the desired form, form II. No fouling or aggregation was observed through image or observational analysis. The D90 of crystals detected by image analysis for the final run (run 7, optimisation iteration 3) was 105 µm. The aspect ratio was 0.62. Therefore, the kinetic estimation section of the workflow was completed as all target parameters achieved values within the designated range.

## 4.4.2.D. Were target parameters or algorithm convergence achieved? (Revisited)

Yes, convergence was achieved as determined by the termination criteria.

## 4.4.2.E. Are additional experiments needed?

No. A suitable process was developed using cooling crystallisation; therefore seeded, antisolvent and scaled-up experiments were not required to meet the objectives for this study.

## 4.4.2.11. Optimum Process Conditions for Small-Scale Crystallisation

In this work, we have identified experimental conditions that yield lamivudine crystals with desirable physical attributes for pharmaceutical manufacture, specifically crystals with the biorelevant polymorphic form and an aspect ratio near 1:1. The process conditions required for these crystal attributes were isothermal crystallisation from ethanol at 19.9 °C, 600 RPM and supersaturation of 1.70. The resulting crystals had a D90 of 105 µm and a bipyramidal crystal habit. The solubility of lamivudine in ethanol also satisfied all target constraints with a value of 10 g/L at low temperatures and a value greater than 50 g/L at elevated temperatures. This approach also enabled kinetic parameter estimations for lamivudine from ethanol under various process conditions (Table 3 and Table 4).
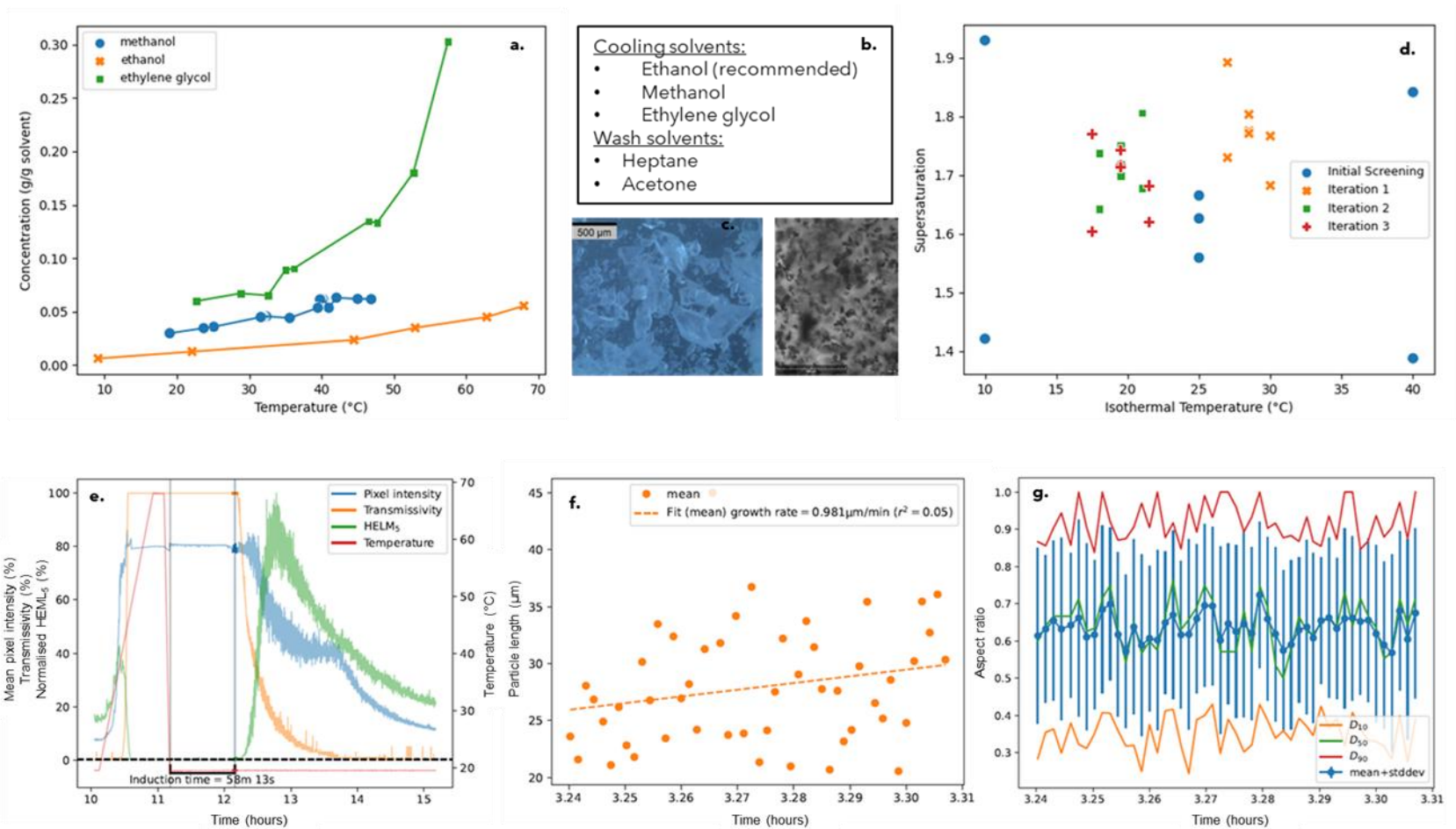
*Figure 27. DataFactory dashboard showing thermodynamic and kinetic crystallisation parameters for the proposed optimum lamivudine recrystallisation – solubility profiles (a), solvent classification (b), microscopy and crystalline images (c), DoE design space (d) and induction time (e), growth rate (f) and aspect ratio over time (g) for the optimum run.*

## 4.5. Conclusions

Following workflows such as the one presented in this work can provide direction and reduce cross-researcher variability in experimental work, thus promoting reliable, reproducible and robust data generation. This work has demonstrated the usefulness of the proposed workflow to go from API selection to an optimised system for small-scale batch cooling crystallisation. By following the workflow, we have determined an alternative method to the current crystallisation process of lamivudine, specifically recrystallisation from methanol that results in the more desirable bipyramidal habit of form II. This case study yielded experimental parameters for lamivudine crystallisation that satisfied our target parameters for an operable isothermal crystallisation process from ethanol at 19.9 °C, 600 RPM and supersaturation of 1.70. Crystalline images and XRPD data confirmed that the resultant crystals were form II, the most stable form with desired downstream processing properties.

This work also shows that a solubility and kinetic dataset could be generated using only 53 g of lamivudine and 1169 mL of solvents compared to previous methods that used 252 g of solute and 1448 mL of solvents[33]. While less material was used in the work presented here, the reduction in material did not result in reduced information content as we collected approximately 7500 high-definition images with extracted thermodynamic and kinetic parameters and solution and solid-state data. Although droplet studies would result in significantly less material use, only solubility and growth estimates could be achieved and as such more complex and agitated thermodynamic and kinetic measurements would not be possible. Furthermore, our contributions to the solubility and kinetic parameter literature data for lamivudine can also be used to inform future wet lab experiments, train and test improved solubility models and build kinetic population balance models for process development.

Workflows such as this enable the future development of smarter experimental planning tools and predictive models, whereby model-driven experimental measurements can also be used to continuously improve models. Incorporating digital tools could help improve efficiency in crystallisation process development by reducing the amount of time and materials needed for the experimental sections

of the workflow. This work has shown that the individual components, such as data analytics, optimisation and hardware, can be integrated and linked together in the logical, step-by-step manner necessary for achieving successful closed-loop automation.

This workflow can be applied to other APIs to identify industry-relevant, robust crystallisation conditions to inform the design of pharmaceutical manufacturing processes. Furthermore, applying this workflow to a wide range of APIs will enable the creation of a solubility and kinetics crystallisation parameter database that spans a varied chemical and crystallisation process space.

The availability of quantitative data on solubility and kinetic parameters in a wide range of APIs could aid laboratory-based researchers in optimising crystallisation processes to improve yield and crystal quality[172] via the application of ML models.[146] [149] [173] Such FAIR data collection for solubility and kinetic parameters can lead to the development of a crystallisation parameter database in a structured format developing benefits analogous to the CSD or the protein data bank (PDB). Workflows such as this will allow us to generate the standardised data that will be an invaluable resource for the predictive design of molecular crystallisation processes.

## Appendix

This appendix contains additional information for context and details pertaining to the workflow design, material characterisation, experimental method development and extracted data tables.

Table 5. Details of each key task and decision in the workflow, listing all outputs and equipment or tools required to carry out the step.

| Stage | Description | Output | Equipment/ Tools |
|---|---|---|---|
| 1. Define the Aim of the Study | Determine the goal/ basis for undertaking the research i.e., full quantitative/ fast qualitative dataset. | A clear aim for the work to be carried out | User |
| 2. Review Prior Knowledge | Collate any past experimental or calculated data or prior knowledge | An understanding of previous studies and what analytical methods are available | Lab notebooks, published work |
| *Decision A* | *Are initial experiments required?* | YES – progress to 'Experimental Input' NO – progress to 'Review Workflow Specific Target Parameters' | Manual user check |
| 3. Characterise Raw Material | Execute experiments to select analytical methods, analyse raw materials and generate reference data for further characterisation and phase ID | Polymorph, physical properties, molecular analysis of raw material | XRPD, STA, Raman, UV-Vis, IR |
| 4. Define Workflow-Specific Target Parameters | Set objective based on crystallisation process screening objectives | Target particle (size, shape) and process (solubility, kinetics, yield) attributes for operable crystallisation process outcomes | Literature, Quality Target Product Profile (QTPP) |
| 5. Solubility and Solvent Effects Study (Polythermal Global Search) | Conduct polythermal experiments (3 cycles) of the API at varying concentrations in a wide range of solvents | Solubility profiles, solvent effects, indications of fouling, kinetic estimates (MSZW), yield estimates | Crystalline (Technobis), 8 mL vials, stirrers, Zinsser Crissy Platform (Zinsser Analytics) |
| 6. & 9. Off-Line Analysis | Conduct experiments with other analytical techniques | Solid form determination, thermodynamic data | XRPD, STA, Raman, Microscopy, Solubility |
| *Decision B* | *Can solubility-temperature profiles be plotted with the Van't Hoff relationship ($R^2 > 0.81$) for solvents?** | YES – progress to next step NO – loop back to secure more data points | Manual user inspection, R-Value filters (coded) |
| 7. Solvent Ranking & Selection | Rank solvents based on target parameters and top solvent progressed to next stage. Collect extra solubility points to give a more accurate temperature-solubility profile. | Solvent choice with a solubility-temperature profile of 6-8 experimental data points | Crystalline (Technobis), 8 mL vials, stirrers, Zinsser Crissy Platform (Zinsser Analytics) |
| 8. Kinetic Parameter Study (Isothermal Local Search) | Conduct isothermal kinetic parameter estimation experiments (3-5 cycles). | Nucleation rate, growth rate, induction time, aspect ratio | Crystalline (Technobis), 8 mL vials, stir bars, Zinsser Crissy Platform (Zinsser Analytics), Image analysis |
| *Decision C* | *Were the experiments free from fouling?* | YES – progress to next step NO – change solvent and loop back | User visual checks |

| | | | |
|---|---|---|---|
| *Decision D* | *Were target parameters or algorithm convergence achieved?* | YES – progress to next step<br>NO – loop back via optimisation | Manual user check |
| 10. Optimisation | Run optimisation algorithms (Multiple Linear Regression/ Partial Least Squares).** | Experimental plan for next best 8-16 experiments | Modde 12.1 software |
| *Decision E* | *Are additional experiments needed?* | YES – call out to additional workflows where seeding, antisolvent and larger scale can be explored***<br>NO – progress to next step | Manual user check |
| 11. Optimum Process Conditions for Small-Scale Crystallisation | Record and pass along conditions from this workflow to complete API process development. | Optimum process conditions for small-scale batch crystallisation | Manual user documentation |
| Update Crystallisation Parameter Database**** | Ensure data is stored in crystallisation parameter database with all conditions, associated responses and outputs from the workflow. | Structured data storage: solute, solvent, concentration (all), dissolution temperature (solubility), induction time, aspect ratio, nucleation rate, growth rate (kinetics) | SQL, Knowledge graph, ontology |

*\*Minimum of 3-4 data points done in triplicate must be used for reliable estimates of $R^2$ values. Some solvent systems this may not be achievable as qualitative solubility can be used.*

*\*\*In current software implementation allows only 2 simultaneous optimisation objectives, sufficient for this study.*

*\*\*\*Although not explicitly discussed in the case study of this paper, if the target parameters for the study are not met then further study can be done to explore seeded, antisolvent and larger-scale crystallisation.*

*\*\*\*\*Although not represented in the graphical workflow diagram, it is important to note that all data from the experimental sections and the offline analysis is stored in a standardised data format.*
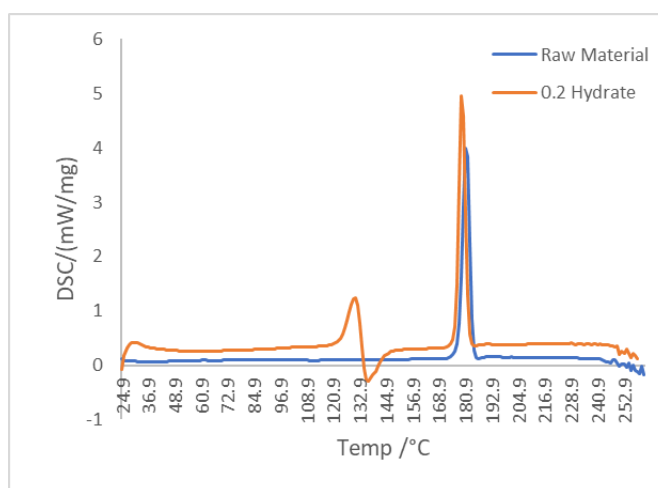


*Figure 28. DSC data for raw material (Form II) and the 0.2 hydrate (Form I).*
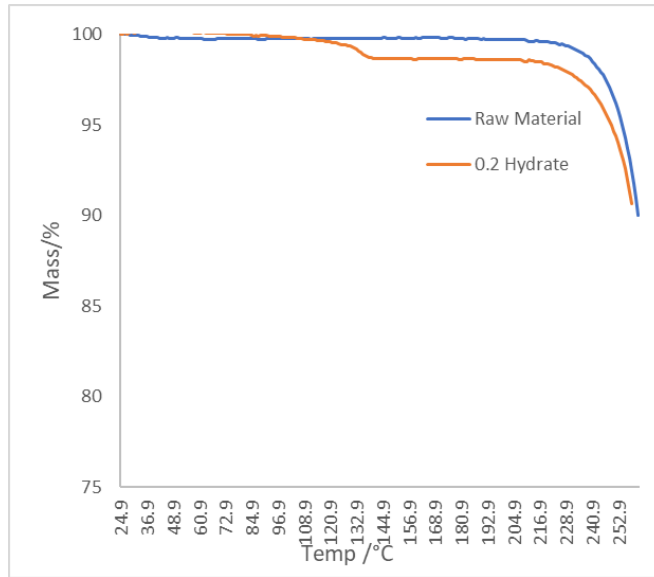
*Figure 29. TGA data for raw material (Form II) and the 0.2 hydrate (Form I).*
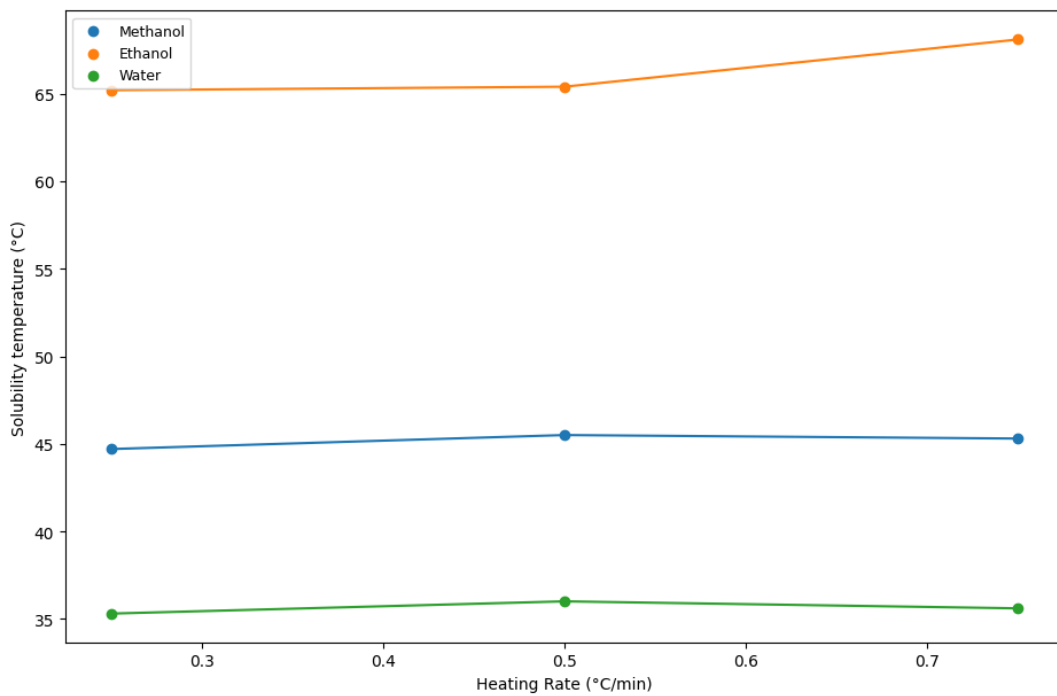


*Figure 30. Dissolution temperature measurement of lamivudine in 3 chosen solvents (methanol, ethanol and water) at different heating rates. The figure shows a large stepwise change for ethanol at the fastest heating rate and minimal differences across all other systems and heating rates. Therefore, for efficiency, whilst still maintaining accurate measurements, a heating rate of 0.5 °C/min was used for all experiments.*

*Table 6. Solubility parameter coefficients (m, c) for lamivudine across all solvents in which quantitative data was collected. Van't Hoff relationship was applied where concentration was expressed as lnX (g/g solvent) and temperature expressed as 1/T (K$^{-1}$) and linear regression applied.*

| Solvent | m | c | R$^2$ |
|---|---|---|---|
| 1-butanol | -4520.92 | 9.510173 | 0.811304 |
| 1-pentanol | -5465.74 | 12.25163 | 0.729647 |
| 1-propanol | -2190.75 | 3.480024 | 1* |
| 2-pentanol | -5420.37 | 11.1428 | 0.821315 |
| 2-propanol | -3951.18 | 8.04683 | 0.950545 |
| N,N-dimethylformamide | -1244.04 | 2.792469 | 0.90898 |
| N-methylpyrrolidone | -16807.2 | 54.29446 | 0.999524 |
| acetonitrile | -3291.86 | 4.602571 | 0.976261 |
| ethanol | -3428.49 | 7.130674 | 0.992587 |
| ethylene glycol | -4334.9 | 11.63974 | 0.920213 |
| formamide | -2612.22 | 7.363137 | 0.996196 |
| methanol | -2657.61 | 5.603299 | 0.928051 |
| water | -8918.68 | 27.11113 | 0.993517 |

*Only two data points were collected due to the low solubility of lamivudine.
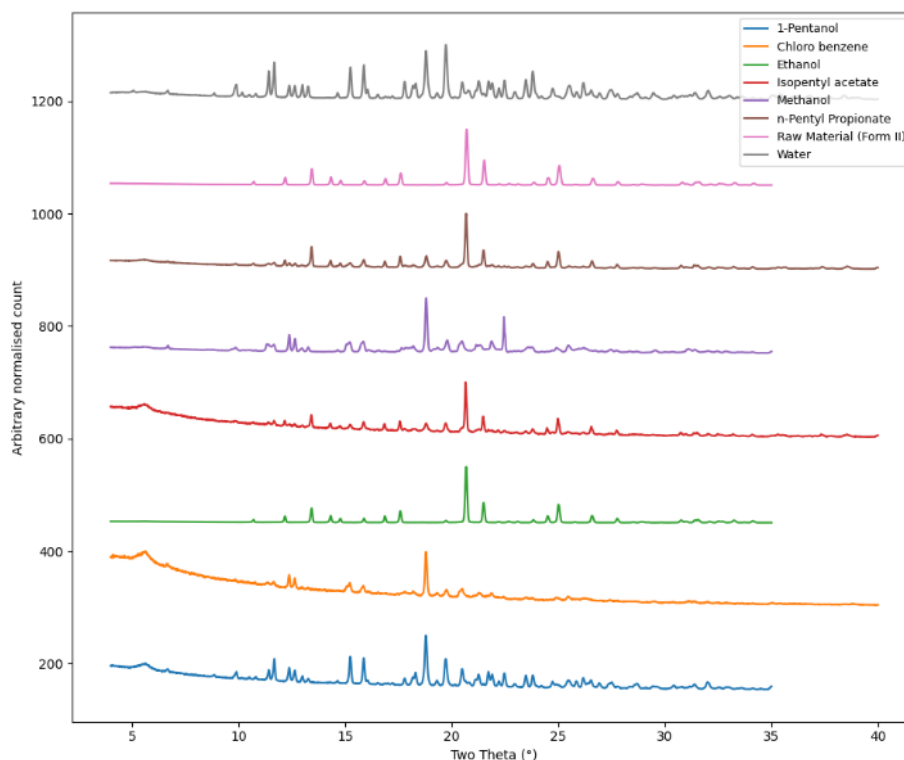


*Figure 31. X-ray powder diffraction patterns of lamivudine from various solvents, offset to allow for polymorph comparison.*
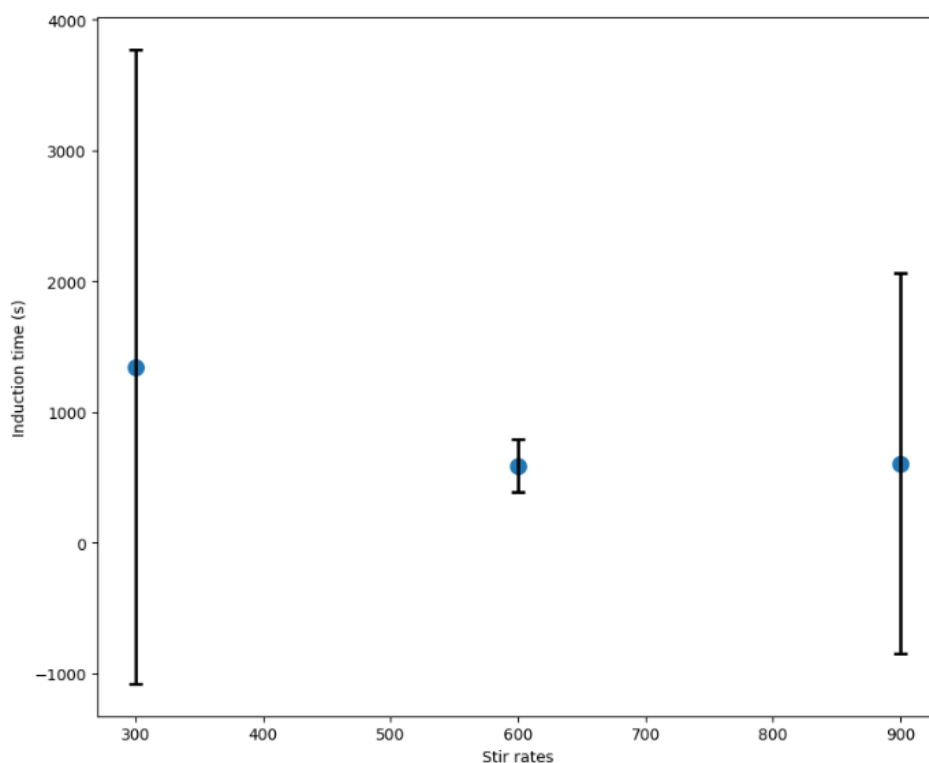
*Figure 32. Induction time measurement (8 replicates) of lamivudine in ethanol at different stir rates. The figure shows large statistical differences in induction time for 300 and 900 RPM likely due to slow mixing and magnetic stir bar bumping respectively. Therefore, a fixed stir rate of 600 RPM was used for all experiments in this study.*
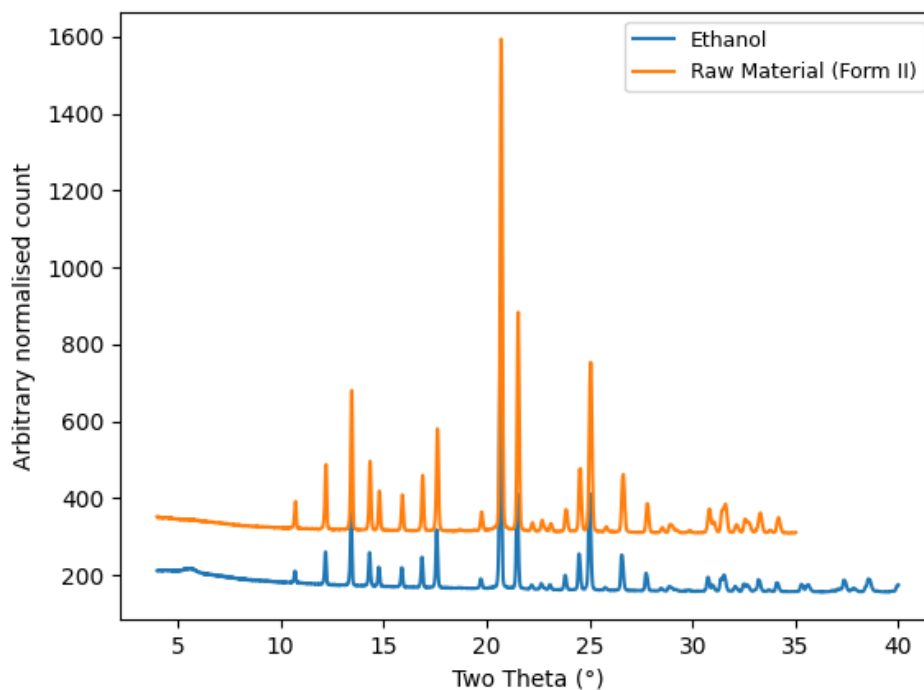


*Figure 33. X-ray powder diffraction patterns of lamivudine from ethanol (taken from the kinetic parameter estimation experimental) showing that the recovered solid form was form II*

# 5. High-throughput screening for large-scale data collection to inform medicine manufacture of aspirin.

*This chapter is the peer-reviewed published work found here:*

## Abstract

Using small-scale crystallisation to inform downstream processes, we can reduce time and material costs in medicine manufacturing. This work introduces a preliminary workflow for information-rich data collection of crystallisation parameters including solubility, induction time, growth rate, secondary nucleation rate, particle shape and size. The results for aspirin are presented here. Highlights include the identification of 24 potential alternative crystallisation solvents for manufacturing aspirin, all of which yield the biorelevant polymorph. Automation of this workflow will enable the use of robotics to further reduce time and material usage when conducting crystallisation experiments for future APIs.

## 5.1. Introduction

Crystallisation is an integral part of the synthesis and manufacturing processes of APIs. Information-rich, high-throughput data collection for small-scale crystallisation experiments can rapidly guide medicine manufacturing by informing digital models. The DataFactory at the CMAC Future Manufacturing Research Hub is establishing an automated crystallisation parameter collection platform that incorporates the following data to inform medicine manufacture. To enable this automation of crystallisation experiments, we present a workflow that guides consistent crystallisation data collection for a range of APIs and solvents under various process conditions. The data collected for aspirin is presented here.

Existing patents for the crystallisation of aspirin in manufacturing[174] use acetone and other organic solvents such as benzene. However, the use of benzene comes with significant health risks. In this study, we identified greener, safer and more sustainable crystallisation solvents for aspirin with industrially-relevant solubilities while also introducing a preliminary workflow to guide future autonomous data collection.

## 5.2. Materials and Methods

Aspirin was purchased from Alfa Aesar. Solvents were purchased from Sigma Aldrich, Fisher Scientific, VWR and Alfa Aesar. For solubility screening, vials of known concentrations of aspirin and solvent were thermocycled between 5 °C and 90 °C (or 10 °C below solvent boiling point if the boiling point of the solvent used was below 100 °C), with a heating rate of 0.5 °C/min and a stirring rate of 600 rpm in the Crystalline (Technobis). For kinetics estimation, vials of known concentration were heated to dissolution and then crash-cooled (5 °C/min) to an isothermal hold with a stirring rate of 600 rpm. Crystalline images, collected at a frequency of 5s, were analysed using a machine learning algorithm.

## 5.3. Results & Discussion

The simplified workflow used in this study (Figure 34) guides experimental work using existing knowledge and model predictions followed by the integration of experimental analysis into a parameter database that guides future experiments.
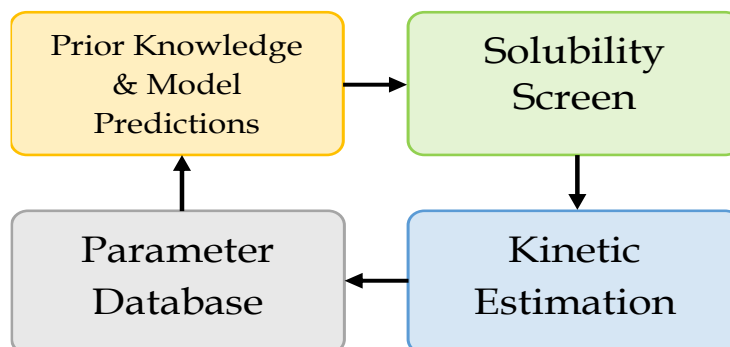


*Figure 34. Simplified workflow for crystallisation data collection*

This workflow guided the small-scale crystallisation experiments for aspirin. A representative group of solvents was chosen using principal component analysis of laboratory-grade solvents[159].

The solubility profiles (Figure 35) of aspirin showed that 24 of the solvents investigated have a suitable solubility range for cooling crystallisation in a manufacturing process[145].
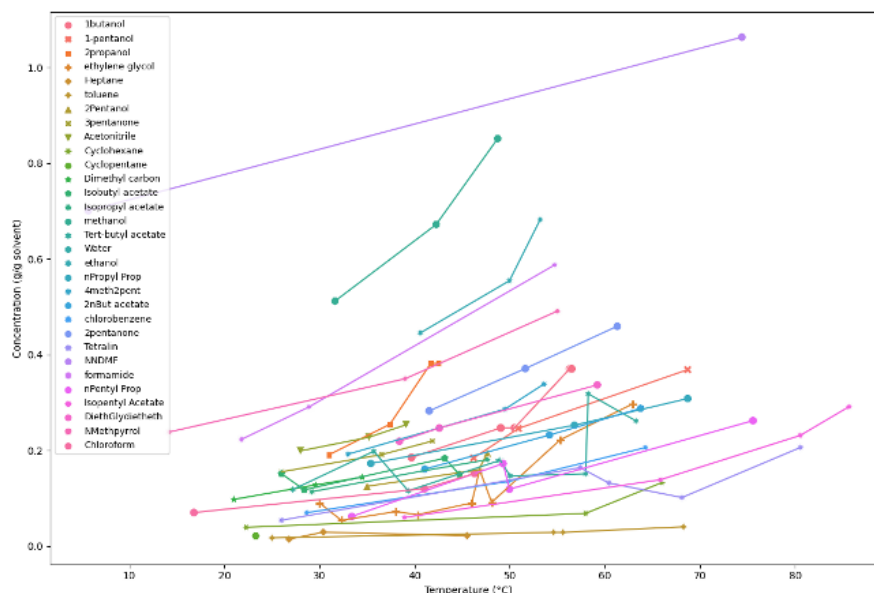


*Figure 35. Solubility- temperature profiles for aspirin in 31 solvents*

Isolation and subsequent analysis of crystals by XRPD confirmed all successful crystallisations yielded only the biorelevant polymorph. Isopentyl acetate was then used for kinetics estimation due to its favourable aspirin solubility and relative safety. In isopentyl acetate, a supersaturation of 1.22 and a temperature of 28.9 °C gave an optimal crystallisation process with target values of induction time of 1 h, growth rate of 1 µm/min and secondary nucleation rate of 1 count/sec.

These small-scale experiments (2-8 mL) can be compared to a study by Maia *et al.,* where samples were circa 50 – 75 mL and heated at 0.01 °C/min. Aspirin solubility in ethanol in the literature[175] and aspirin solubility collected by the methodology here yielded a value difference of only 1.4% (Table 7). This comparison suggests the reliability of the method used here despite each data point being collected 50 times quicker and with a 10-fold reduction in material. Similar comparisons were observed for other APIs when data was available.

*Table 7. Comparison of solubility collection methods at 30 °C for aspirin in ethanol.*

| Solvent | Small-scale[a] (g/g solvent) | Mid-scale[b] (g/g solvent) | Difference[c] (%) |
|---------|---------|---------|---------|
| Ethanol | 0.295 | 0.299 | 1.36 |

[a] Values are extracted by linear regression using the Van't Hoff relationship.

[b.175]

[c] the difference was calculated in respect to small scale.

## 5.4. Conclusions

Solubility data for aspirin in 31 solvents was collected rapidly (~ 3 weeks) using only 62 g of API. The method was found to be reliable and on par with a method previously described in the literature. Kinetic parameter estimations were also achieved using only 9 g of API. These values will allow for the parameterisation of digital twin models which will, in turn, inform larger-scale medicine manufacturing. The workflow used here will guide autonomous robotic data collection for a range of APIs and solvents as part of the CMAC DataFactory.

## Appendix

This appendix contains additional information for context and details pertaining to extracted data tables.

*Table 8. Solubility parameter coefficient (m, c) for aspirin across all solvents where quantitative data was collected. Van't Hoff relationship was applied where concentration was expressed as lnX (g/g solvent) and temperature expressed as 1/T (K$^{-1}$) and linear regression applied.*

| Solvent | m | c | R$^2$ |
|---|---|---|---|
| 1-butanol | -4323.63 | 12.07379 | 0.929609 |
| 1-pentanol | -3046.5 | 7.939335 | 0.935387 |
| 2-methyltetrahydrofuran | 16480.7 | -57.6657 | 1* |
| 2-n-butoxyethyl acetate | -2723.42 | 6.845822 | 0.998423 |
| 2-pentanol | -2849.71 | 7.162042 | 0.844134 |
| 2-pentanone | -2590.8 | 6.973981 | 0.998468 |
| 2-propanol | -6058.57 | 18.22854 | 0.966292 |
| 3-pentanone | -2021.86 | 4.890497 | 0.993043 |
| 4-methyl-2-pentanone | -2643.27 | 6.977208 | 0.987024 |
| acetonitrile | -2011.45 | 5.058651 | 0.986104 |
| chlorobenzene | -3085.55 | 7.558764 | 0.999814 |
| chloroform | -2292.44 | 5.238838 | 0.979163 |
| cyclohexane | -2376.4 | 4.763598 | 0.832443 |
| diethyleneglycol diethyl ether | -2091.22 | 5.206969 | 0.995441 |
| dimethyl carbonate | -2632.37 | 6.626194 | 0.997133 |
| ethanol | -3209.63 | 9.405387 | 0.930932 |
| ethyl acetate | -2728.73 | 6.75483 | 0.991284 |
| ethylene glycol | -4732.85 | 12.69255 | 0.727178 |
| formamide | -2808.14 | 8.038103 | 0.998204 |
| heptane | -960.549 | -0.73169 | 0.081151 |
| isobutyl acetate | -1064.09 | 1.543915 | 0.356205 |
| isopentyl acetate | -3713.1 | 9.052439 | 0.992751 |
| isopropyl acetate | -2454.24 | 5.948659 | 0.996628 |
| methanol | -2866.41 | 8.722827 | 0.987047 |
| N,N-dimethylformamide | -588.01 | 1.75283 | 1** |
| N-methylpyrrolidone | -1636.35 | 4.242122 | 0.984581 |
| n-pentyl propionate | -3535.9 | 8.894414 | 0.881268 |
| n-propyl propionate | -1830.56 | 4.17541 | 0.999887 |
| tert-butyl acetate | -1886.39 | 4.146686 | 0.41835 |
| tetralin | -2314.87 | 4.878595 | 0.75667 |
| toluene | -1894.82 | 2.272772 | 0.973644 |

*Only 2 data points were recorded, the trend was opposite to what was expected (therefore can be ignored).

**Only 2 data points were recorded.

# 6. Comparative Study on Adaptive Bayesian Optimization for Batch Cooling Crystallization for Slow and Fast Kinetic Regimes.

## Abstract

Crystallisation kinetic parameter estimation is important for the classification, design and scale-up of pharmaceutical manufacturing processes. This study investigates the impact of supersaturation and temperature on induction time, nucleation rate, and growth rate for the compounds lamivudine (slow kinetics) and aspirin (fast kinetics). Adaptive Bayesian optimisation (AdBO) has been used to predict experimental conditions that achieve target crystallisation kinetic values for each of these parameters of interest. The use of AdBO to guide the choice of the experimental conditions reduced material usage up to five-fold when compared to a more traditional statistical Design of Experiments (DoE) approach. The reduction in material usage demonstrates the potential of AdBO to accelerate process development as well as contribute to Net-Zero and green chemistry strategies. Implementation of AdBO can lead to reduced experimental effort and increase efficiency in pharmaceutical crystallisation process development. The integration of AdBO into the experimental development workflows for crystallisation development and kinetic experiments offers a promising avenue for advancing the field of autonomous data collection exploiting digital technologies and the development of sustainable chemical processes.

## 6.1. Introduction

Crystallisation is a key step in the manufacture of high-quality drug products and serves as a purification step to remove impurities from the crude product.[176] Understanding and controlling the crystallisation of pharmaceuticals allows us to design manufacturing processes that yield the desired particle shape, size, and polymorph without impacting the yield, purity and quality of the final product. For example, differences in particle shape and size affect downstream processing as well as the effectiveness of dissolution for human absorption.[177]

As particle shape and size are dictated by the kinetics of crystallisation, these properties can be controlled by altering key process parameters that dictate supersaturation. Controlling particle shape and size relies on controlling nucleation, growth and agglomeration rate processes. Primary nucleation, which is the initial formation of new crystals in solution and can be described by induction time, is inherently difficult to control deterministically as it displays a stochastic character.[154] Secondary nucleation relates to the formation of new nuclei from the attrition of existing crystals, and can be controlled by changes in solid loading, particle size, shear rate, mixing and supersaturation.[178] A pharmaceutical crystallisation process requires nucleation and growth rates that can be maintained within ranges that facilitate robust and controlled operation. Excessive nucleation rates lead to too many fines and/or fouling and excess particle growth can lead to unfavourable particle shapes, agglomeration, and reduced purification performance with the potential inclusion of impurities. Agglomeration, the aggregation of two or more particles to form larger particles, can change the final product's physical and chemical properties, and while it can be used to aid in filtering, agglomeration must be controlled to ensure consistency and tablet content uniformity.[179, 180] Conversely, very low nucleation and growth rates can lead to low yield and long, uneconomic processing times.

Various methods have been applied to develop controlled crystallisation processes ranging from chemical intuition[181], methodical parameter investigation[182] or machine learning approaches proposed in this work. DoE[61] is a powerful statistical tool used to map the design space, fit models to the data, interpolate between known values and forms a useful component of quality by design (QbD) experimental planning. The use of statistical DoE for process understanding was

introduced initially in the agriculture industry[60, 61] and in recent years has seen wide application in biotechnological processing,[183, 184] drug discovery and medicines manufacture[55, 185-187]. DoE experimental planning aims to use the least number of experiments to explore the effects of changing a given number of variables across the whole design space. However, even with DoE methods, exploring large numbers of process variables still requires many experiments. For example, a full factorial design exploring 5 variables with 3 increments would result in a plan of 243 experiments which, when using expensive active pharmaceutical ingredients (APIs), may not be necessitating using other design methods.

In this work, we explore Bayesian optimisation[188] (BO) as an alternative method to DoE for finding global or local minima or maxima in functions of interest. BO constructs a probabilistic model of the objective function, specifically the difference between target and experimental crystallisation kinetic parameters. It employs an acquisition function to iteratively suggest the next point of evaluation, or experiment, to either reduce uncertainty in the model by further exploration and or determine the global optimum by exploiting known values. Previous studies[114] utilising BO in pharmaceutical crystallisation are promising but highlight the need for methods to accommodate the multiple objectives required for suitable process design. For example, Bayesian approaches for optimisation of chemical reactions were shown to be efficient, taking a few hours of lab work compared to Super modified simplex algorithm (10+ hours per variable) and grid searches (600+ hours per variable).[189]

This work demonstrates the application of algorithms to optimise crystallisation process conditions to achieve desired pharmaceutical crystallisation kinetic parameters. Methods including DoE and BO can be used to target specific values for pharmaceutical crystallisation parameters, and both methods are shown to be significantly more efficient than grid-search approaches. An adaptive (a varying exploration and exploitation model) BO experimental planner showed further improved performance over DoE and fixed BO methods. The exploration of two API case studies shows the algorithmic approach proposed may be generalisable to other APIs. Implementing improved decision-making algorithms such as these could reduce time and material use and contribute to a greener and more sustainable approach to process development in medicine manufacturing.

## 6.2. Materials

Lamivudine was purchased from Molekula Ltd. as an off-white powder and aspirin was purchased from Alfa Aesar as a white solid. Both APIs have a purity exceeding 99% and did not undergo further purification. Ethanol and ethyl acetate were purchased from VWR and have purity exceeding 99.97% so did not undergo further purification.

## 6.3. Experimental Methods and Optimisation

The optimisation loop included the following steps: vial dosing, crystallisation, analysis of the resulting image data, and further data analysis by an optimisation algorithm to recommend the next best round of experiments. A schematic to represent this logical flow is shown in Figure 36.
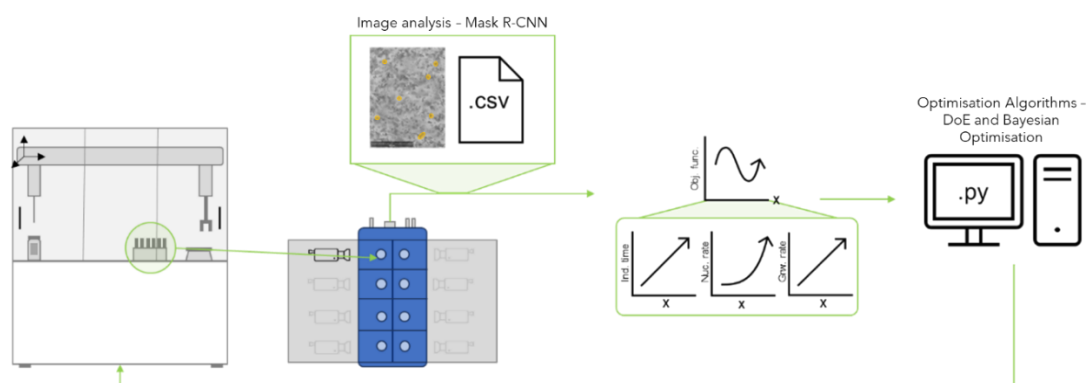


*Figure 36. Schematic diagram of the experimental set-up and optimisation loop consisting of the Zinsser Crissy platform, the Technobis Crystalline, data analysis, and optimisation algorithms (left-to-right).*

### 6.3.1. Vial Dosing, Crystallisation and Image Analysis

Sample preparation for crystallisation experiments was carried out using a Zinsser Analytics Crissy platform[190], an XYZ robot that doses both powders and liquids. Crystallisation experiments were conducted using a Technobis Crystalline platform[125], a parallel reactor system that can perform eight separate heating, cooling, and stirring procedures with in-built sample imaging at the 2 – 7 mL scale.

The experimental procedure was consistent for all experiments and involved the following steps:

1. Heat the solution to a temperature 10 °C below the solvent's boiling point at a rate of 0.5 °C/min.
2. Maintain the elevated temperature for 10 minutes to ensure complete dissolution.
3. Cool the solution to the desired isothermal temperature at a rate of -10 °C/min, with no stirring at this stage.
4. Keep the solution at the isothermal temperature for three hours.
5. Repeat steps 1-4 for a total of 5 cycles.

The stir rate was fixed at 600 rpm throughout the experiment, except where specified.

Images were captured every five seconds, and an in-house convolutional neural network (CNN) image analysis algorithm [191] was used to extract kinetic parameters. X-ray powder diffraction (XRPD) patterns were collected using a Bruker D8 Discover[128], and the data were visualized using DIFFRAC.EVA[192] and Matplotlib[193] (Python). Solubility profiles were generated for each API in a solvent selected from previous work (Chapter 4 and Chapter 5).

## 6.3.2. Optimisation: Input Parameter Bounds, Target Parameter Objectives and Approaches

The objective of this optimisation problem was to minimise the difference between the experimentally measured values for the kinetic parameters of interest and the associated target values. Table 9 presents the bounds on the input parameters for supersaturation and isothermal experimental temperature and the target objectives for induction time, nucleation rate and growth rate. As this optimisation problem has multiple objectives it can be assumed that the optimum process conditions will sit on a Pareto front[194] so the objective function value will be used as a quantifier for optimisation performance.

Table 9. The parameters and objectives for the optimisation problem.

| API | Input parameter bounds | | Target parameter objectives | | |
|---|---|---|---|---|---|
| | Super-saturation | Temperature (°C) | Induction time (s) | Nucleation rate (#/image/s) | Growth rate (µm/s) |
| Lamivudine | 2 – 3 | 5 – 50 | 3600 | 0.1 | 0.01 |
| Aspirin | 1.05 – 2 | 5 – 50 | 3600 | 0.1 | 0.05 |

The input parameter bounds were varied for each API to accommodate the different sizes of their metastable zone width (MSZW). Lamivudine showed a broad MSZW (> 30 °C) and therefore was deemed unlikely to observe any nucleation at low supersaturations within the time constraints of the experiment. Aspirin displayed a narrow MSZW (mean of 16 °C) and thus nucleation was likely to be feasible at low supersaturations (generally below 1.2). As a larger MSZW generally allows higher supersaturations to be achieved before primary nucleation occurs and nucleation is known to dominate growth at high supersaturations[195], it was mechanistically assumed that, comparatively, nucleation would dominate for lamivudine crystallisation but growth would dominate for aspirin crystallisation. Thus, while the nucleation rate target was held constant for both systems, the growth rate target for aspirin was set to 5x the target for lamivudine. Numerical values for rate targets were based on initial experiments.

**6.3.2.1. Optimisation Approach 1: Design of Experiment (DoE) with Surface Minimisation**

For the DoE optimisation, an initial DoE screening of 28 experiments was followed by successive rounds of smaller screens (7 experiments/iteration) centred at the next predicted optimum. We refer to this method as adaptive due to the iterative update of the objective function and the adaption of this new objective function surface to guide the next round of experiments. This cycle was repeated until the termination of a change in temperature of less than 2 °C and a change in supersaturation of less than 0.02 between previous and next recommended experiments was achieved.

The initial experimental screen was performed by employing a full-factorial design, consisting of five supersaturation levels, five temperature levels, and three central

points, resulting in 28 experiments. The initial DoE plan for lamivudine (Table 10) was centred around a supersaturation of 2.4 and an isothermal temperature of 20 °C and for aspirin (Table 10), the DoE plan was centred around a supersaturation of 1.5 and a temperature of 20 °C.

Table 10. Adaptive design of experiment plan for optimisation iterations. *If iterations continue past 3 then the same exploitative (small scope of experimental design space) plan is followed.

|  | Iteration | | | |
| --- | --- | --- | --- | --- |
|  | Initial | 1 | 2 | 3* |
| **Supersaturation** | +/- 0.4 | +/- 0.2 | +/- 0.1 | +/- 0.05 |
| **Temperature** | +/- 10 °C | +/- 10 °C | +/- 5 °C | +/- 2 °C |

For the subsequent optimisation, an objective function surface could not be fitted directly to the kinetic parameters measured experimentally for multiple reasons. Firstly, as nucleation is a stochastic process dozens of experiments are required to sample the probability distribution of induction time values.[154] Secondly, there is an inherent measurement uncertainty associated with the nucleation and growth rate parameters obtained via image analysis as only a subset of sample particles are sampled. Consequently, multiple data points are required to reduce uncertainties associated with these kinetic parameters.

As the aim of the optimisation was to achieve convergence in as few experiments as possible, equations that describe the kinetics of crystallisation were fit to the experimental data and used to smooth the surface of the objective. To smooth the surface of the objective function with domain knowledge, experimental data were plotted for each input parameter (temperature and supersaturation) with respect to each objective (induction time, nucleation rate and growth rate) and equations 1-5 (below)[196] [197] were fit to the data:

Equation 8. $\quad t_{ind}(SS) = A * e^{-b*SS}$

Equation 9. $\quad R_{growth}(SS) = a * SS + b$

Equation 10. $\quad R_{growth}(T) = a * T + b$

Equation 11. $\quad R_{nuc}(SS) = A * e^{b*SS}$

$$\text{Equation 12.} \qquad R_{nuc}(T) = a * T + b$$

where $t_{ind}$ is induction time, $R_{growth}$ is growth rate, $R_{nuc}$ is nucleation rate at a given supersaturation *SS (where SS = 1 at equilibrium solubility)* and temperature *T* and *a* and *b* are fitted parameter coefficients. The growth rate objective was weighted by a factor of ten, this meant that all parameters sat within the same order of magnitude and thus on a comparable scale. There is no direct domain relationship between induction time and temperature, so this relationship was assumed as negligible and therefore parameters were not fitted. The parameters *A, a*, and *b* were obtained from the fitted equations, and the resulting fitted functions for induction time, nucleation rate and growth rate were then mathematically manipulated[198] so that the minimum of each function occurs at the objective target. The functions transformations used are given below with Equation 13 being used to transform equations that feature exponential relationships between the objective and the input parameters (Equation 8 and Equation 11), and Equation 14 being used to transform equations that feature linear relationships (Equation 9, Equation 10 and Equation 12):

$$\text{Equation 13.} \qquad D_x\big(P_{fitted}(x)\big) = e^{|log(P_{target})-log(P_{fitted}(x))|}$$

$$\text{Equation 14.} \qquad D_x\big(P_{fitted}(x)\big) = \big|\big(P_{target}\big) - \big(P_{fitted}(x)\big)\big|$$

where $P_{target}$ is the target value for a given parameter, $P_{fitted}$ is the fitted equation for a given parameter evaluated at the input value for either supersaturation, *SS,* or temperature, *T,* both here represented by *x*, and the difference between the target value $P_{target}$ and fitted value $P_{fitted}(x)$ for a given target parameter is defined as the function $D_x(P_{fitted}(x))$.

The objective function for a given supersaturation and temperature, *f(SS,T)*, could then be defined as follows:

$$\text{Equation 15.} \qquad f(SS,T) = D_{SS}\big(t_{ind}(SS)\big) + D_{SS}\big(R_{nuc}(SS)\big) + D_T\big(R_{nuc}(T)\big) +$$
$$D_{SS}(R_{growth}(S)) + D_T(R_{growth}(T))$$

The minimum of the objective surface (within the parameter upper and lower bounds) was then calculated using multiple approaches: a genetic algorithm (GA), differential evolution (DE), covariance matrix adaptation evolution strategy (CMA-ES), Nelder-Mead or pattern search approach. These algorithms were implemented in Python using the PyMOO[199] library. The means of the values of supersaturation and temperatures corresponding to the predicted minimum from each algorithm were used as the centre point for the next round of experiments. A smaller two-level full factorial DoE (Table 10) was then performed at this supersaturation and temperature. Following this round of experiments, the fitted functions were updated, the objective function surface recalculated and a new average minimum predicted for the next round of experiments. This loop was repeated until the termination criteria were met. The subsequent DoE plans can be seen in Table 10. Using multiple algorithms to predict the next best experiment allowed us to remove outliers (see Table 13 and Table 14).

### 6.3.2.2. Optimisation Approach 2: Bayesian Optimisation

Three centre points were taken as initial values and the difference between the target objectives and the experimental values for the parameters of interest were included in the objective function using the equation below:

$$\textit{Equation 16.} \qquad D_{SS,T}\big(P_{exp@SS,TT}\big) = \big|log(P_{target}) - log(P_{exp@SS,T})\big|$$

where $P_{target}$ is the target value for a given parameter, $P_{exp@SS,T}$ is the experimental value for a given parameter at given supersaturation, $SS$, and temperature, $T$, inputs, and the difference between the log target value, $P_{target}$, and log experimental value, $P_{exp@SS,T}$, for a given supersaturation and temperature is defined as the function $D_{SS,T}(P_{exp@SS,T})$ where P is the parameter of interest. As stated earlier in the construction of Equation 15, the relationship between induction and temperature can be assumed to be negligible, the same can be applied in Equation 17 also.

The objective function assessed with kinetic parameters measured at a given supersaturation and temperature, $f(SS,T)$, compared with the target kinetic parameters was then defined as follows:

*Equation 17.*     $$f(SS, T) = D_{SS,T}(t_{ind}) + D_{SS,T}(R_{nuc}) + D_{SS,T}(R_{growth})$$

A Bayesian Optimisation algorithm was then implemented to determine the next experimental point to trial within the lower and upper bounds of the parameters. The algorithm was implemented in Python using the GPyOpt[200] library using a Gaussian process probabilistic model and expected improvement acquisition function (see Item 1 in the Appendix of Chapter 6).

A Bayesian model with acquisition jitter (a scalar ratio between exploration: exploitation) of 0.001, 0.1, 1 and 10 as well as an adaptive dynamic model was performed. The criterion for changing the exploration/ exploitation trade off was if the objective function value falls below 10% of the maximum objective function value, the acquisition jitter is set to 1. Then, if the objective function value falls below 5% of the maximum, the acquisition jitter is assigned a value of 0.1. In summary this approach starts by selecting experiments across the whole design space and then adapts its search purpose to focus on finding the true optimum solution. This discrete-value adaptive approach makes sense for a crystallisation given the potential for many local optima and that experiments are done in batch. A continuous-value adaptive approach to changing the acquisition jitter could be implemented on fast moving experiments such as flow chemistry. Similar to the DoE optimisations, the termination criteria for convergence were a change in temperature of less than 2 °C and a change in supersaturation of less than 0.02 between recommended experiments. Unlike the DoE optimisations, no domain knowledge in the form of physical equations that describe the system was needed as Gaussian processes can adapt the function used allowing good fit to data with many minia and large uncertainties.

## 6.4. Results and Discussion

Lamivudine recrystallising from ethanol typically has relatively slow kinetics (i.e., large MSZW and induction times in the hours (SS ~ 2)) whereas aspirin in ethyl acetate typically has fast kinetics (i.e., narrow MSZW and induction times in the minutes (SS ~ 2)). Comparing the performance of the DoE and BO optimisation methods for APIs with different inherent kinetic profiles can provide evidence for the generalisability of application of these methods.

### 6.4.1. Design of Experiment with surface minimisation:

Overall, the DoE approach allowed for sufficient data collection of measured kinetic parameters and, after applying domain knowledge, allowed for visualisation of the data on a 3D surface. The minimum point of the surface (the only optimum value) was then found each time by search-based algorithms (refer to Section 6.3.2.1.).

The initial DoE screen consisting of 28 experiments investigated how supersaturation and temperature impacted each of the three objectives: induction time, nucleation rate and growth rate. The minimum of the 3D objective function surface corresponded to a supersaturation of 2.88 and temperature 26.9 °C for lamivudine and a supersaturation of 1.21 and temperature 45.0 °C for aspirin. As discussed earlier, various approaches were used to identify the minimum, and the values determined by GA, DE, CM-AES and pattern search consistently agreed (Table 13 and Table 14 in the Appendix of Chapter 6). By contrast, the Nelder Mead values were discarded when calculating the mean value in early iterations as this algorithm encountered calculation errors and only predicted the minimum at boundary conditions. The results over seven iterations of the optimisation and experimental loop are shown in Table 11 where the median values of supersaturation and isothermal temperature predicted across all algorithms are presented.

*Table 11. Predicted optimum supersaturations and isothermal experimental temperatures for lamivudine and aspirin over multiple iterations, with the objective function also presented.*

| | Lamivudine | | | Aspirin | | |
|---|---|---|---|---|---|---|
| Iteration | Temperature (°C) | SS | Obj. Function Value | Temperature (°C) | SS | Obj. Function Value |
| Initial Screen | 26.9 | 2.88 | 0.86 | 45.0 | 1.21 | 1.00 |
| 2 | 26 | 2.63 | 0.81 | 50.0 | 1.18 | 0.90 |
| 3 | 24.9 | 2.56 | 0.73 | 22.62 | 1.12 | 0.84 |
| 4 | 22.4 | 2.28 | 0.72 | 19.13 | 1.14 | 0.72 |
| 5 | 22.97 | 2.31 | 0.68 | 12.97 | 1.14 | 0.68 |
| 6 | 23.73 | 2.36 | 0.61 | 5.98 | 1.16 | 0.57 |
| 7 | 24.12 | 2.36 | 0.61 | 5.0 | 1.16 | 0.57 |

For both APIs, the termination criteria (see section 6.3.2.1.) were achieved after the seventh iteration of the algorithm, equating to 70 experiments. These optimisation approaches were consistent in predicting relatively similar values of supersaturation and temperature for the global minimum of the objective function for lamivudine. However, for aspirin, a series of supersaturation and temperature values resulted in the global minimum 'valley' as seen in Figure 37.
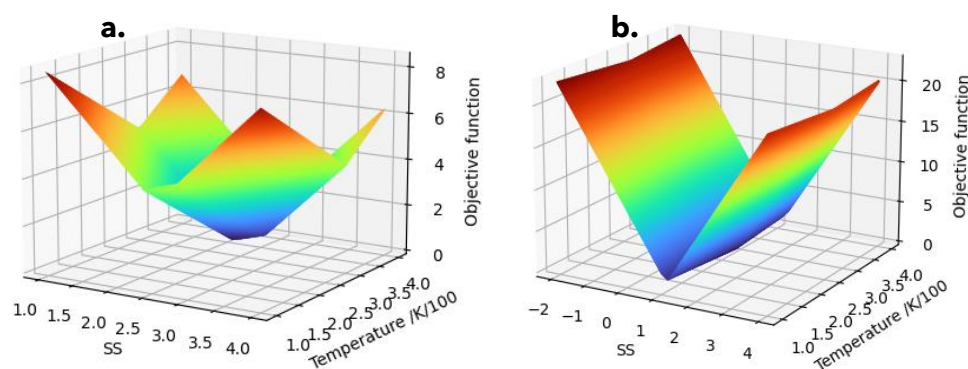


*Figure 37. Surface plot of the objective function resulting from the adaptive DoE optimisation experiments for lamivudine (a) and aspirin (b).*

The surface plot for aspirin has minimal dependence on temperature and is much more dependent on supersaturation such that small changes (+/- 0.1) in the supersaturation had more impact on the crystallisation kinetics than changes in the temperature. The surface plot for lamivudine shows a comparable impact on crystallisation kinetics caused both by supersaturation and temperature.

DoE approaches are typically less suited to handle high levels of noise.[201] The fitting of physical equations to smooth the objective function surface helped address this issue of noise[202] associated with the stochastic nature of crystallisation nucleation kinetics, and, arguably, significantly reduces the number of experiments required to generate a smooth enough surface to find a minimum. This in-part mechanistic model parameter fitting allowed for a 2$^{nd}$ order polynomial kinetic model to be defined for each API.

## 6.4.2. Bayesian optimisation:

BO was applied using five different acquisition jitter level models to evaluate the performance from different exploration and exploitation weightings. These were: highly exploitation-focused, exploitation-focused, exploration-focused, a model that balanced exploration vs. exploitation and an adaptive model which started as exploration-focused and moved to exploitation-focused as the objective function value decreased (see section 6.3.2.2 for further details). The three centre points of the domain space of interest (SS of 2.4 and temperature of 20 °C for lamivudine and SS of 1.5 and temperature of 20 °C for aspirin), and the related extracted kinetic data from the initial DoE plan, were used as initial values for the Bayesian model.

As expected, exploitation-focused models i.e., acquisition jitter of 0.001 to 0.1, terminated quickly but still had high values (relative to other models) for the objective function for both lamivudine and aspirin indicating that optimisation likely found a local rather than global minimum (Table 12). The model which balanced the focus between exploitation and exploration, i.e. acquisition jitter of 1, terminated for both lamivudine and aspirin after 6 experiments again with high objective function values indicative of local minima (Table 12). The exploration-focused model, i.e., acquisition jitter of 10, achieved significantly lower objective function values for both APIs compared with the exploitation-focused model and the model that balanced exploration and exploitation models. However, the

exploration-focused model only terminated after 31 experiments for lamivudine and 25 experiments for aspirin (Table 12).

*Table 12. Predicted optimum supersaturations and isothermal experimental temperatures for lamivudine and aspirin across different exploration/ exploitation Bayesian models.*

| Acquisition jitter model | Lamivudine | | | | Aspirin | | | |
|---|---|---|---|---|---|---|---|---|
| | Temperature (°C) | SS | Obj. Function Value | No. of Exp. | Temperature (°C) | SS | Obj. Function Value | No. of Exp. |
| 0.001 | 20 | 2.43 | 1.90 | 8 | 20 | 1.49 | 1.98 | 6 |
| 0.1 | 19.82 | 2.58 | 1.76 | 6 | 19.52 | 1.50 | 1.98 | 8 |
| 1 | 20.95 | 2.63 | 1.38 | 6 | 20 | 1.51 | 1.98 | 6 |
| 10 | 29.8 | 2.57 | 0.86 | 31 | 14.39 | 1.36 | 0.40 | 25 |
| Adaptive | 17.9 | 2.86 | 0.40 | 32 | 18.17 | 1.27 | 0.40 | 15 |

To integrate the advantages of exploration-focused models (lower objective function values indicative of a global minimum) and exploitation-focused models (faster convergence), an adaptive Bayesian model was explored. This acquisition jitter for this model started high (10), i.e., exploration-focused, and was reduced stepwise with steps down in value triggered by reductions in objective function values until the termination criteria was met. In other words, an objective function value of between 5 and 10% of the maximum value (this was 8% after 13 experiments for lamivudine and 6% after 14 experiments for aspirin) triggered the change of acquisition jitter from 10 of 1. Termination criteria was met for aspirin after 1 experiment at an acquisition jitter of 1. In the lamivudine optimisation, a reduction in the objective function value to a value of 3% of the maximum triggered the acquisition jitter to change from 1 to 0.1 after a further 19 experiments. The results from next experiment following this change met termination criteria. If the problem was more complex or the termination criteria more stringent then the acquisition jitter could be further reduced, however this was unnecessary here. In total termination criteria were satisfied after 32 experiments for lamivudine and 15 experiments for aspirin (Table 12). This difference in number of experiments required is indicative of the differences between fast and slow kinetic regimes for the two APIs. Aspirin has a smaller 'sweet spot' and a large gradient into the

minimum due to the large impact on crystallisation kinetics from small changes in process conditions particularly SS and as such the BO algorithm can find the optimum more efficiently.
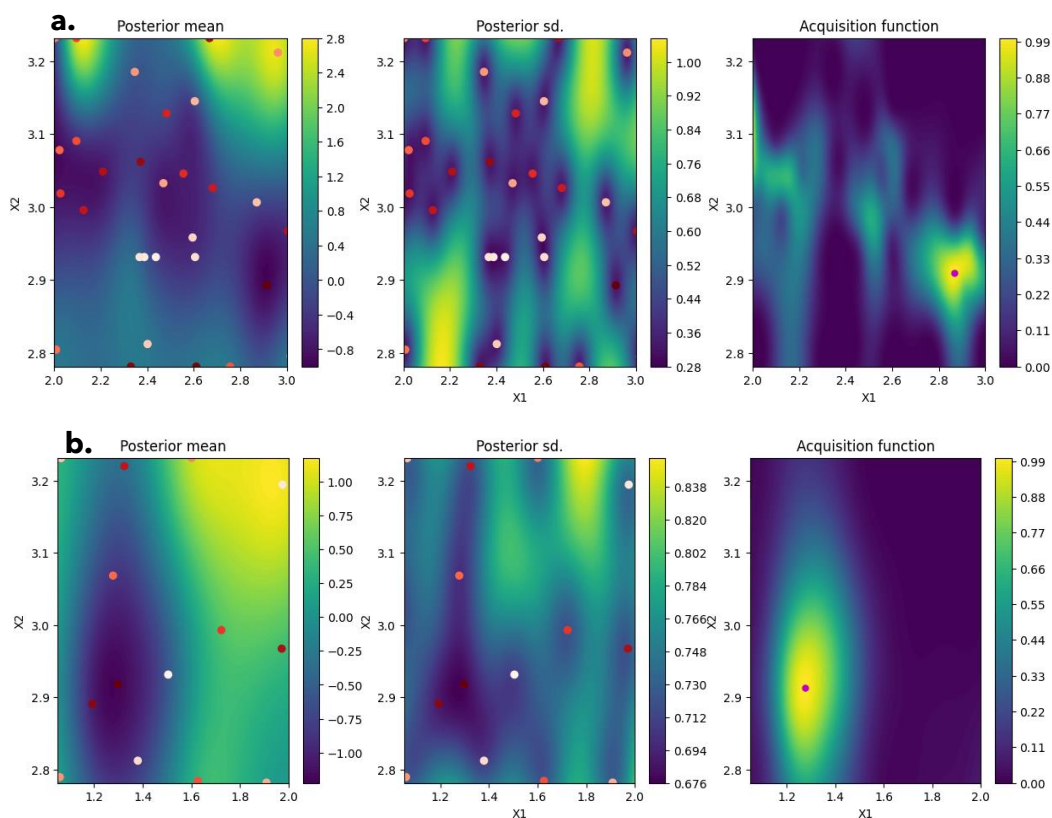


*Figure 38. Plot of posterior mean, posterior standard deviation (sd), and acquisition function for lamivudine (a) and aspirin (b) for the final iteration of the adaptive exploration/ exploitation model. The X and Y axis, labelled as X1 and X2 correspond to the supersaturation and temperature (respectively) input values at which the posterior mean, posterior standard and acquisition function are evaluated.*

Figure 38 shows the 2D heat maps for the objective function value (here labelled posterior mean referring to the mean of the posterior distribution[203] output in BO), the uncertainty associated with these values (posterior standard deviation), and the acquisition function (an expected improvement algorithm used to weight the optimisation towards exploration or exploitation). While some parts of the design space still have areas with higher levels of uncertainty (i.e., posterior standard deviation ~0.8 and above), these areas also correlate to points where the value of the objective function (i.e., posterior mean values) is also predicted to be high.

Thus, these areas are unlikely to correspond to the global minimum, i.e., objective of the optimisation.

### 6.4.3. Comparison of the optimisation methods

It is worth focussing on the results from AdBO as this approach either performed as well as or better than other BO configurations tested and shows most promise as a general experimental optimisation approach. As the objective functions for BO and DoE approaches investigate the absolute difference between the experimental or fitted outcome and the target values for each induction time, nucleation rate, and growth rate, we can compare the objective function values for both methods. All target parameters for induction time, nucleation rate and growth rate were satisfied for both lamivudine and aspirin with experimental validation (Figure 45 in the Appendix of Chapter 6 – with reference to reliability of each measurement), and XRPD (Figure 46 and Figure 47) was run on each crystal sample recovered to confirm the target polymorphic form for both lamivudine and aspirin was produced.

The BO methodology for optimising kinetic parameters required fewer experiments when compared with the DoE approach (Figure 39). Specifically, AdBO optimisation of lamivudine crystallisation used 50 % less material and required 54 % less experiments than DoE methodology. Further improvements were seen for aspirin crystallisation where AdBO used 75 % less material and required 79 % less experiments than the DoE optimisation. This improvement in experimental efficiency is predominantly due to the fact that the AdBO method does not require an initial screen of experiments.

As shown in Figure 39, the AdBO method reaches the lowest objective function value and satisfies the termination criteria in the least number of experiments suggesting that the AdBO method can outperform both the fixed acquisition jitter Bayesian models and the DoE method. While the DoE method achieved a relatively low objective function through initial screening, the AdBO method continues to drive the trends in objective function even lower. Notably, the AdBO model demonstrates a significant improvement over the BO10 (Bayesian with an acquisition jitter of 10) for aspirin, achieving a low objective function in 10 fewer experiments. Furthermore, AdBO's faster reduction of the objective function and satisfaction of achieving the termination criteria comes at no additional

computational cost as all the algorithms employed in this study exhibited comparable average execution times (from 3 to 4 s), with no statistically significant differences observed (Figure 44 in the Appendix of Chapter 6).

The inherent noise in the objective function also presented no noticeable challenge for the AdBO optimisation. The AdBO method, as applied, was 'blind' to the physics of the experiment in that no domain knowledge was required to achieve a fit for the objective function. This lack of reliance on equations that describe the physics of the system provides us with a potentially generalisable method that can handle the inherent stochastic nature of induction time and the higher levels of noise associated with fewer numbers of experimental points.[204] These results also suggest that the AdBO method may have potential application to optimise other parameters for which the physics of the system is complex or poorly understood as well as other physical processes beyond crystallisation.
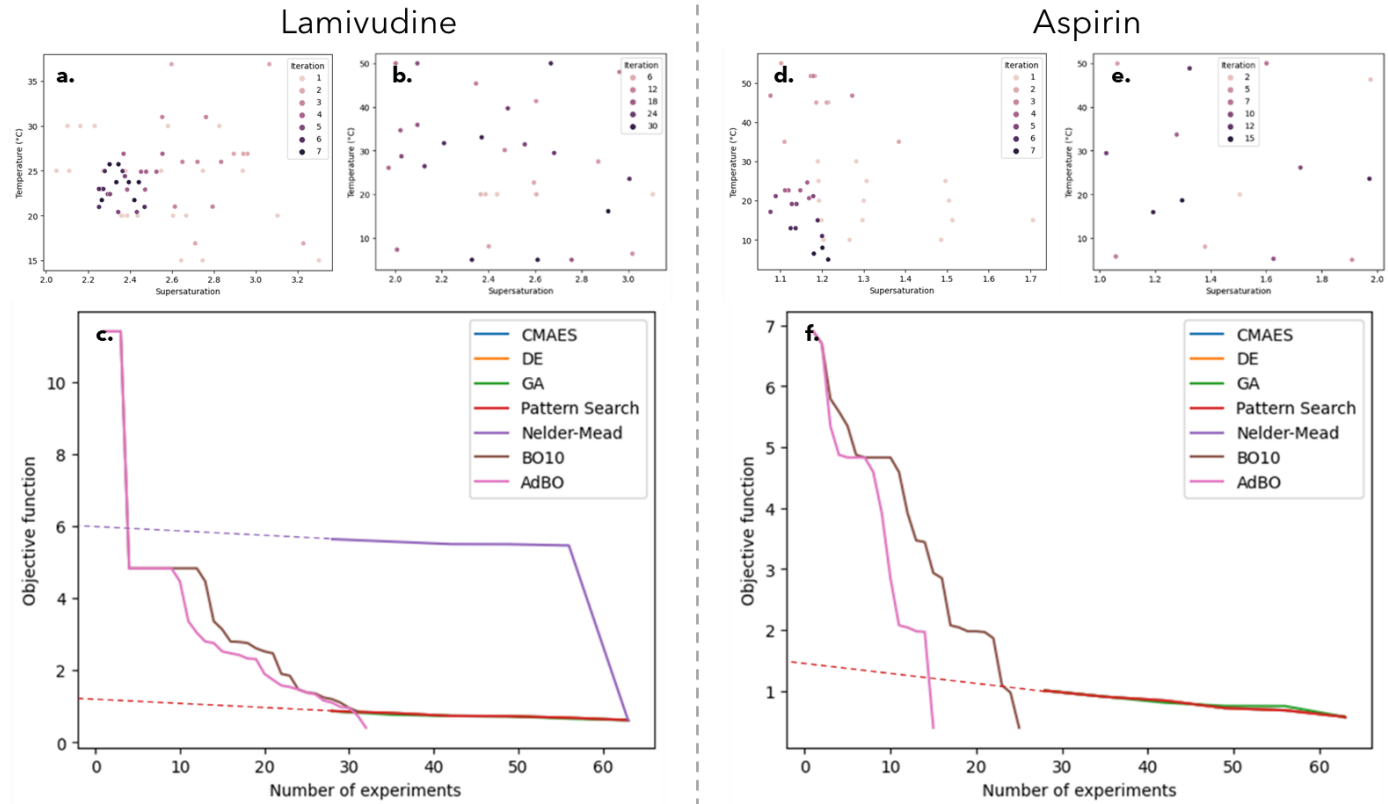
*Figure 39. Comparison between different optimisation approaches, DoE for lamivudine (a) and aspirin (d) and AdBO for lamivudine (b\*) and aspirin (e\*) and algorithm performance in reducing the objective function for process condition optimisation for small-scale batch cooling crystallisation of lamivudine (c\*\*) and aspirin (f\*\*). The dashed lined represents an extrapolation back to experiment 0, as the DoE methods required a 28 initial screening experimental plan.*

\*The numbering on the legends for these figures were not sequential as the AdBO method was a 1 experiment/ iteration approach and thus would make the legend too large to display.

\*\*Some algorithms performed exactly the same as each other and hence not all lines can be displayed on the graph due to overlapping.

Both approaches (DoE and AdBO) discussed in this paper have also shown significant improvements over a grid search with Bayesian methods requiring the least experimental work. A grid search of the design space for lamivudine and aspirin would require 1125 and 1069 experiments respectively based on the increments of the termination criteria and the span of the design space for each variable for each API.

## 6.5. Conclusions

This study demonstrated the successful application of two optimisation methods for crystallisation kinetic parameters of two APIs with the approaches implemented in the Python libraries, PyMOO and GPyOpt. By minimising the total sum of the differences between target and experimental values, these algorithms have successfully achieved the desired target values for induction time, secondary nucleation rate, and growth rate that relate to attainable conditions for a viable industrial process design. Two case studies of lamivudine and aspirin were explored to assess the effectiveness of the algorithms for APIs that display widely differing crystallisation kinetics. The DoE and AdBO methods identified the Pareto optimal process conditions, specifically supersaturation and temperature, essential for optimising crystallisation processes. Notably, both methods yielded low objective function values (0.61 for DoE lamivudine and 0.57 for DoE aspirin and 0.40 for AdBO both lamivudine and aspirin with respect to initial values upwards of 11 for lamivudine and 7 for aspirin). The savings, in terms of time and material (Section 6.4.3.), of using AdBO over DoE methods can be estimated as between a 15 – 80 kWh reduction in energy and specifically £20,000/kg for lamivudine and £60/kg for aspirin.

The effective application of the AdBO method to these two APIs is promising however further study across a wider range of APIs is required to confirm the generalisability of this approach to the wide range of physicochemical properties presented by new pharmaceutical molecules. It will be important to address known limitations of the method, such as increasing the dimensionality of this problem beyond 20 inputs/outputs (e.g., stirring rate, rate of antisolvent addition, heating/cooling rate, morphology) could reduce algorithmic performance.[108, 109] The

incorporation of these more complex parameters will have a complex relationship with nucleation rate and growth rate and as such convergence of the algorithm may require further experimentation or an initial training data set to achieve high algorithmic performance. Furthermore, in this work we relied on prior knowledge of the MSZW in specific solvents system tested when constraining the design space search. However, the approach can be phased to explore potential process conditions to evaluate MSZW under different conditions to then feed into more detailed kinetic parameter studies.

Even with these limitations, this study clearly shows that using BO to guide experimental design allows for a faster, targeted and more sustainable approach to API crystallisation kinetics data collection in pharmaceutical development. Furthermore, the AdBO model could transform automated crystallisation data collection to autonomous experimental design enabling smart experimentation for crystallisation kinetics to further enhance R&D productivity and process understanding. This implementation of BO has the potential to translate to meaningful cost savings for materials, resource, energy utilisation and chemical waste and to accelerate development timelines. This method could also be expanded to investigate the optimisation of other numerical parameter objectives such as solute concentration, aspect ratio and yield and applied to optimisation of drug substant filtration[205] and flow[146].

# Appendix

This appendix contains additional information for context and details pertaining to the algorithm code, algorithm results, computational time used and extracted data figures and tables.

```python
def f(x,y):
    return x**2 + x + y**2 + y
GPyOpt.methods.BayesianOptimization(f = f,
                    domain=bounds,
                    model_type='GP',
                    acquisition_type ='EI',
                    acquisition_jitter = 0.001 to 10,
                    X=xy_init,
                    Y=z_init)
```

*Item 1. Bayesian optimisation code implemented in GPyOpt – using a simple 2D parabola function to aim for minimums on the surfaces of the initial data, domains set out as the input parameter bounds, a Gausian process probabilistic model, an expected improvement acquisition function, a varied exploration/ exploitation acquisition jitter and experimental data inputs of X (supersaturation (x) and temperature (y)) and the measured inputs of Y (crystallisation kinetics (z)).*

*Table 13. Individual performance for next best experiment from each algorithm for DoE (PyMOO) for lamivudine.*

| Iteration | Algorithm | Results |
|---|---|---|
| **Initial Screen** | GA | Supersaturation - 2.88 <br> Temperature - 27.06 °C |
| | DE | Supersaturation - 2.88 <br> Temperature - 26.92 °C |
| | CMAES | Supersaturation - 2.88 <br> Temperature - 26.92 °C |
| | Nelder Mead | Supersaturation - 2.88 <br> Temperature - -273.15 °C |
| | Pattern Search | Supersaturation - 2.88 <br> Temperature - 26.92 °C |
| **Supersaturation of 2.88 +/- 0.2 and temperature of 26.9 +/- 10 °C will be used.** | | |
| 2 | GA | Supersaturation - 2.63 <br> Temperature - 25.65 °C |
| | DE | Supersaturation - 2.63 <br> Temperature - 25.97 °C |

| | | | |
|---|---|---|---|
| | CMAES | Supersaturation - 2.63<br>Temperature - 25.97 °C | |
| | Nelder Mead | Supersaturation - 2.62<br>Temperature - -273.15 °C | |
| | Pattern Search | Supersaturation - 2.63<br>Temperature - 25.97 °C | |

**Supersaturation of 2.63 +/- 0.1 and temperature of 26.0 +/- 5 °C will be used.**

| | | | |
|---|---|---|---|
| 3 | GA | Supersaturation - 2.51<br>Temperature - 24.46 °C | |
| | DE | Supersaturation - 2.56<br>Temperature - 24.87 °C | |
| | CMAES | Supersaturation - 2.56<br>Temperature - 24.87 °C | |
| | Nelder Mead | Supersaturation - 2.56<br>Temperature - -273.15 °C | |
| | Pattern Search | Supersaturation - 2.56<br>Temperature - 24.87 °C | |

**Supersaturation of 2.56 +/- 0.05 and temperature of 24.9 +/- 2 °C will be used.**

| | | | |
|---|---|---|---|
| 4 | GA | Supersaturation - 2.3<br>Temperature - 22.18 °C | |
| | DE | Supersaturation - 2.28<br>Temperature - 22.4 °C | |
| | CMAES | Supersaturation - 2.28<br>Temperature - 22.4 °C | |
| | Nelder Mead | Supersaturation - 2.28<br>Temperature - -273.15 °C | |
| | Pattern Search | Supersaturation - 2.28<br>Temperature - 22.4 °C | |

**Supersaturation of 2.28 +/- 0.05 and temperature of 22.4 +/- 2 °C will be used.**

| | | | |
|---|---|---|---|
| 5 | GA | Supersaturation - 2.31<br>Temperature - 22.97 °C | |
| | DE | Supersaturation - 2.31<br>Temperature - 22.97 °C | |
| | CMAES | Supersaturation - 2.31<br>Temperature - 22.97 °C | |
| | Nelder Mead | Supersaturation - 2.31<br>Temperature - -273.15 °C | |
| | Pattern Search | Supersaturation - 2.31<br>Temperature - 22.97 °C | |

**Supersaturation of 2.31 +/- 0.05 and temperature of 22.97 +/- 2 °C will be used.**

| | 6 | GA | Supersaturation - 2.36<br>Temperature - 23.73 °C |
|---|---|---|---|
| | | DE | Supersaturation - 2.36<br>Temperature - 23.73 °C |
| | | CMAES | Supersaturation - 2.36<br>Temperature - 23.73 °C |
| | | Nelder Mead | Supersaturation - 2.36<br>Temperature - 23.73 °C |
| | | Pattern Search | Supersaturation - 2.36<br>Temperature - 23.73 °C |
| **Supersaturation of 2.36 +/- 0.05 and temperature of 23.73 +/- 2 °C will be used.** | | | |
| | 7 | GA | Supersaturation - 2.36<br>Temperature - 24.12 °C |
| | | DE | Supersaturation - 2.36<br>Temperature - 24.12 °C |
| | | CMAES | Supersaturation - 2.36<br>Temperature - 24.12 °C |
| | | Nelder Mead | Supersaturation - 2.36<br>Temperature - 24.11 °C |
| | | Pattern Search | Supersaturation - 2.36<br>Temperature - 24.12 °C |
| **Termination criteria of +/- 2 °C & 0.02 SS has been met** | | | |

Table 14. Individual performance for next best experiment from each algorithm for DoE (PyMOO) for aspirin.

| Iteration | Algorithm | Results |
|---|---|---|
| **Initial Screen** | GA | Supersaturation - 1.21<br><br>Temperature - 45.16 °C |
| | DE | Supersaturation - 1.21<br><br>Temperature - 45.04 °C |
| | CMAES | Supersaturation - 1.21<br><br>Temperature - 45.04 °C |
| | Nelder Mead | Supersaturation - ERROR<br><br>Temperature - ERROR |
| | Pattern Search | Supersaturation - 1.21<br><br>Temperature - 45.04 °C |

| | | |
|---|---|---|
| **Supersaturation of 1.21 +/- 0.2 and temperature of 45 +/- 10 °C will be used. Minimum SS to be trialled is 1.11.** | | |
| 2 | GA | Supersaturation - 1.18 |
| | | Temperature - 52.06 °C |
| | DE | Supersaturation - 1.18 |
| | | Temperature - 51.75 °C |
| | CMAES | Supersaturation - 1.18 |
| | | Temperature - 51.75 °C |
| | Nelder Mead | Supersaturation - ERROR |
| | | Temperature - ERROR |
| | Pattern Search | Supersaturation - 1.18 |
| | | Temperature - 51.75 °C |
| **Supersaturation of 1.18 +/- 0.1 and temperature of 51.75 +/- 5 °C will be used. (Incorrect DoE plan was used here)** | | |
| 3 | GA | Supersaturation - 1.12 |
| | | Temperature - 22.62 °C |
| | DE | Supersaturation - 1.11 |
| | | Temperature - 22.62 °C |
| | CMAES | Supersaturation - 1.11 |
| | | Temperature - 22.62 °C |
| | Nelder Mead | Supersaturation - ERROR |
| | | Temperature - ERROR |
| | Pattern Search | Supersaturation - 1.11 |
| | | Temperature - 22.62 °C |
| **Supersaturation of 1.12 +/- 0.05 and temperature of 22.62 +/- 2 °C will be used.** | | |
| 4 | GA | Supersaturation - 1.13 |
| | | Temperature - 19.13 °C |
| | DE | Supersaturation - 1.14 |
| | | Temperature - 19.13 °C |
| | CMAES | Supersaturation - 1.14 |
| | | Temperature - 19.13 °C |

|  |  |  |
|---|---|---|
|  | Nelder Mead | Supersaturation - ERROR |
|  |  | Temperature - ERROR |
|  | Pattern Search | Supersaturation - 1.14 |
|  |  | Temperature - 19.13 °C |

**Supersaturation of 1.14 +/- 0.05 and temperature of 19.13 +/- 2 °C will be used.**

|  |  |  |
|---|---|---|
| 5 | GA | Supersaturation - 1.12 |
|  |  | Temperature - 12.94 °C |
|  | DE | Supersaturation - 1.14 |
|  |  | Temperature - 12.97 °C |
|  | CMAES | Supersaturation - 1.14 |
|  |  | Temperature - 12.97 °C |
|  | Nelder Mead | Supersaturation - ERROR |
|  |  | Temperature - ERROR |
|  | Pattern Search | Supersaturation - 1.14 |
|  |  | Temperature - 12.97 °C |

**Supersaturation of 1.14 +/- 0.05 and temperature of 12.97 +/- 2 °C will be used.**

|  |  |  |
|---|---|---|
| 6 | GA | Supersaturation - 1.16 |
|  |  | Temperature - 5.98 °C |
|  | DE | Supersaturation - 1.16 |
|  |  | Temperature - 5.98 °C |
|  | CMAES | Supersaturation - 1.16 |
|  |  | Temperature - 5.98 °C |
|  | Nelder Mead | Supersaturation - 1.16 |
|  |  | Temperature - 5.98 °C |
|  | Pattern Search | Supersaturation - 1.16 |
|  |  | Temperature - 5.98 °C |

**Supersaturation of 1.16 +/- 0.05 and temperature of 5.98 +/- 2 °C will be used.**

|  |  |  |
|---|---|---|
| 7 | GA | Supersaturation - 1.16 |
|  |  | Temperature - 5 °C |

| | |
|---|---|
| DE | Supersaturation - 1.16 |
| | Temperature - 5 °C |
| CMAES | Supersaturation - 1.16 |
| | Temperature - 5 °C |
| Nelder Mead | Supersaturation - 1.16 |
| | Temperature - 5 °C |
| Pattern Search | Supersaturation - 1.16 |
| | Temperature - 5 °C |
| **Termination criteria of +/- 2 °C & 0.02 SS has been met** | |



*Figure 40. Plot of posterior mean, posterior standard deviation, and acquisition function for lamivudine (a) and aspirin (b) for the final iteration of the acquisition jitter = 0.001.*
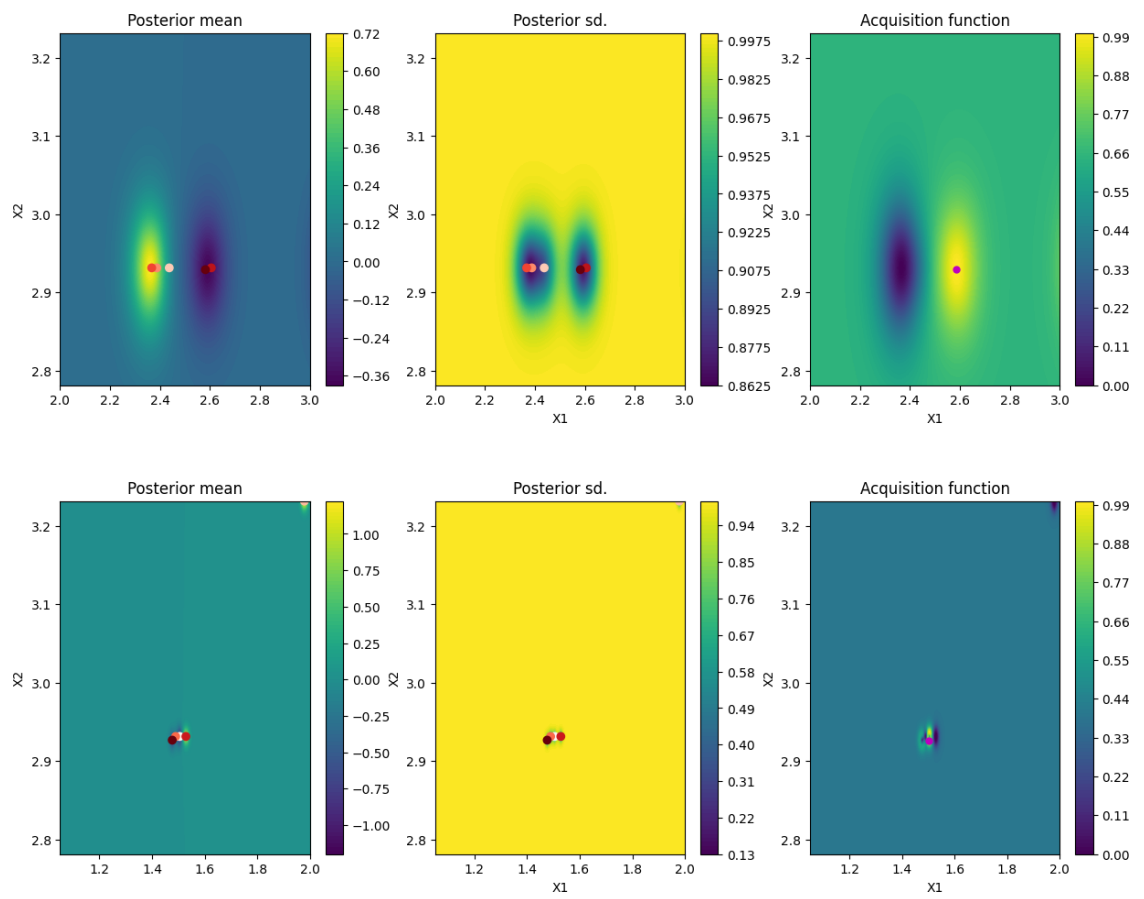
*Figure 41. Plot of posterior mean, posterior standard deviation, and acquisition function for lamivudine (a) and aspirin (b) for the final iteration of the acquisition jitter = 0.1.*
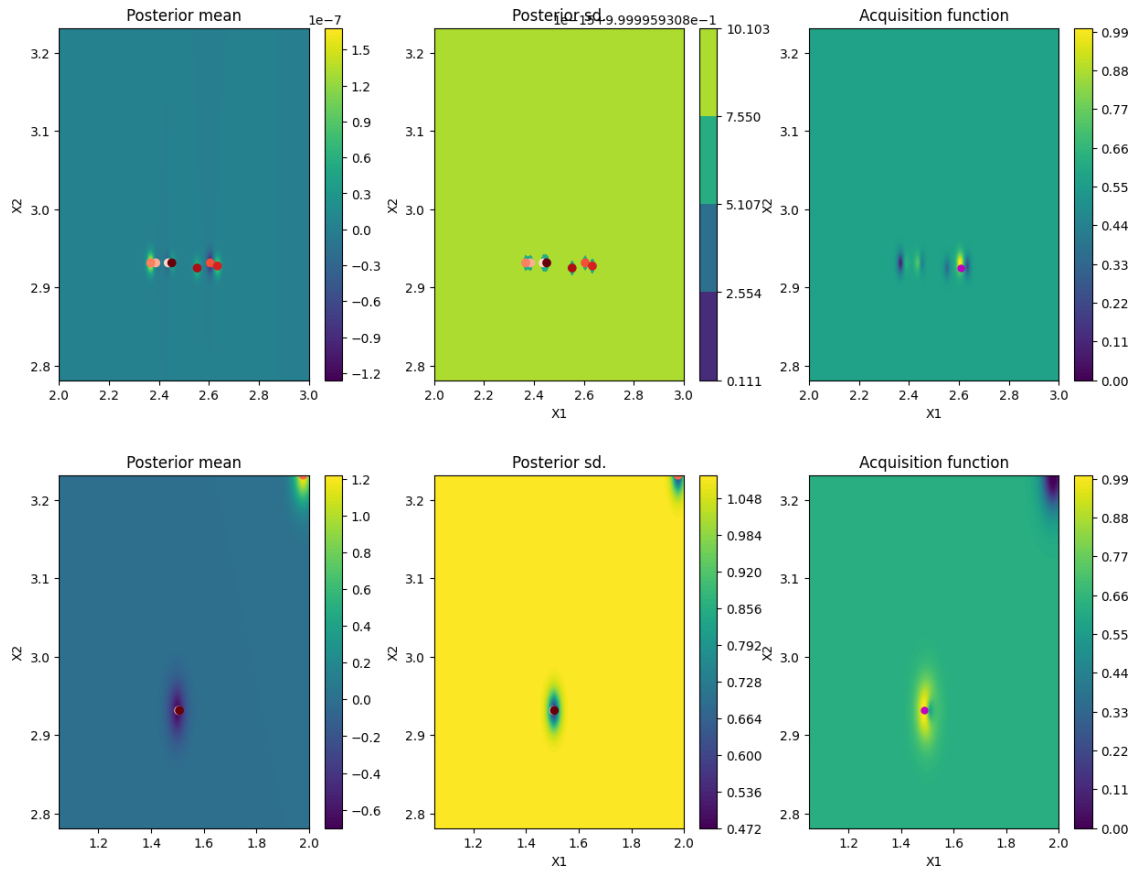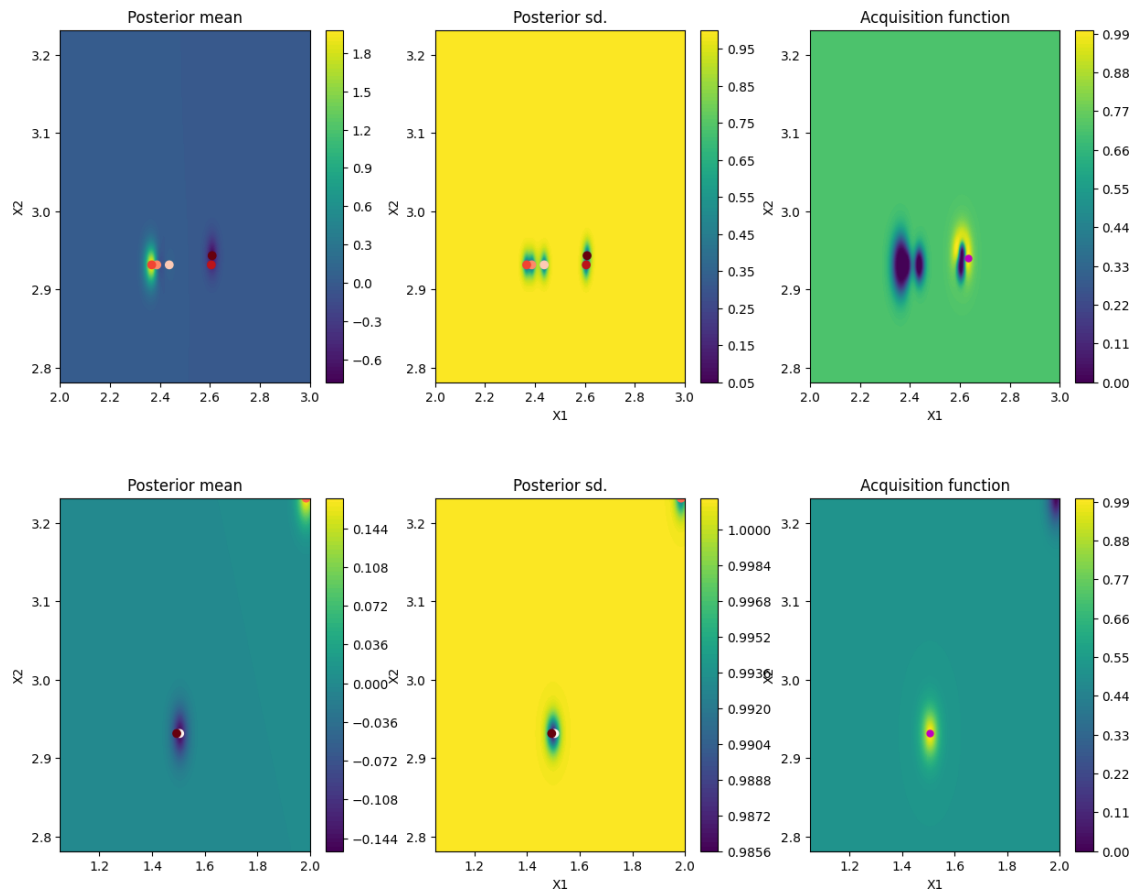
*Figure 42. Plot of posterior mean, posterior standard deviation, and acquisition function for lamivudine (a) and aspirin (b) for the final iteration of the acquisition jitter = 1.*
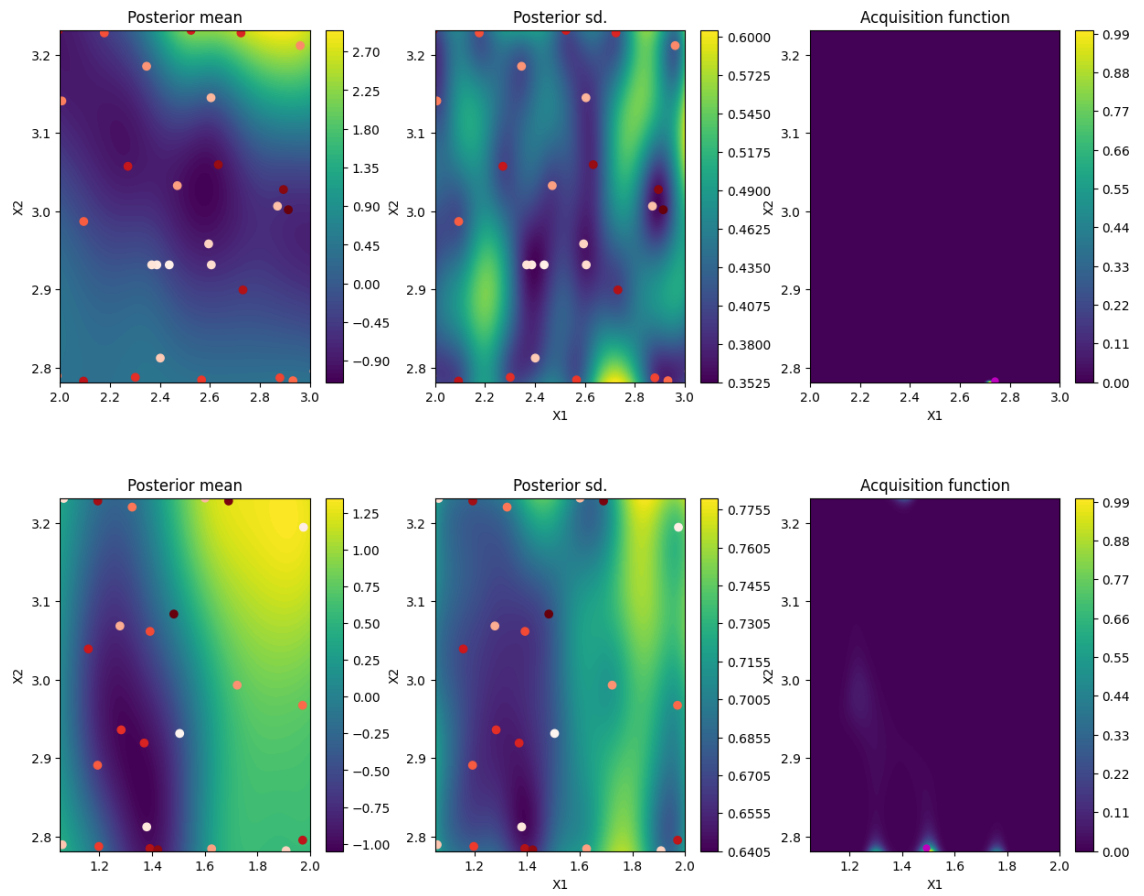
*Figure 43. Plot of posterior mean, posterior standard deviation, and acquisition function for lamivudine (top) and aspirin (bottom) for the final iteration of the acquisition jitter = 10.*
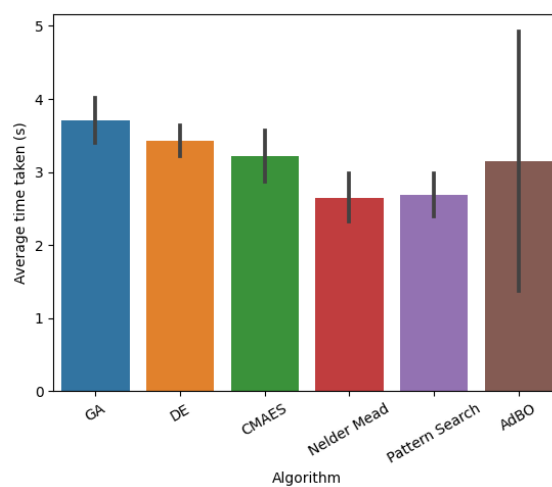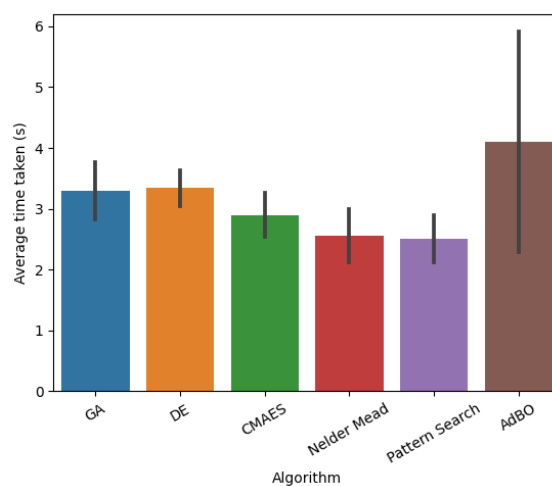
*Figure 44. Computational time required to perform each optimisation algorithm, averaged over every iteration, for lamivudine (a) and aspirin (b). There is a general trend that Nelder Mead and Pattern Search performed the fastest but there are no statistically significant differences. Additionally, as all algorithm times were below an average of 5 s it can be deemed that due to the large time expense associated to the practical crystallisation experiment that algorithm time is trivial in the whole process loop.*
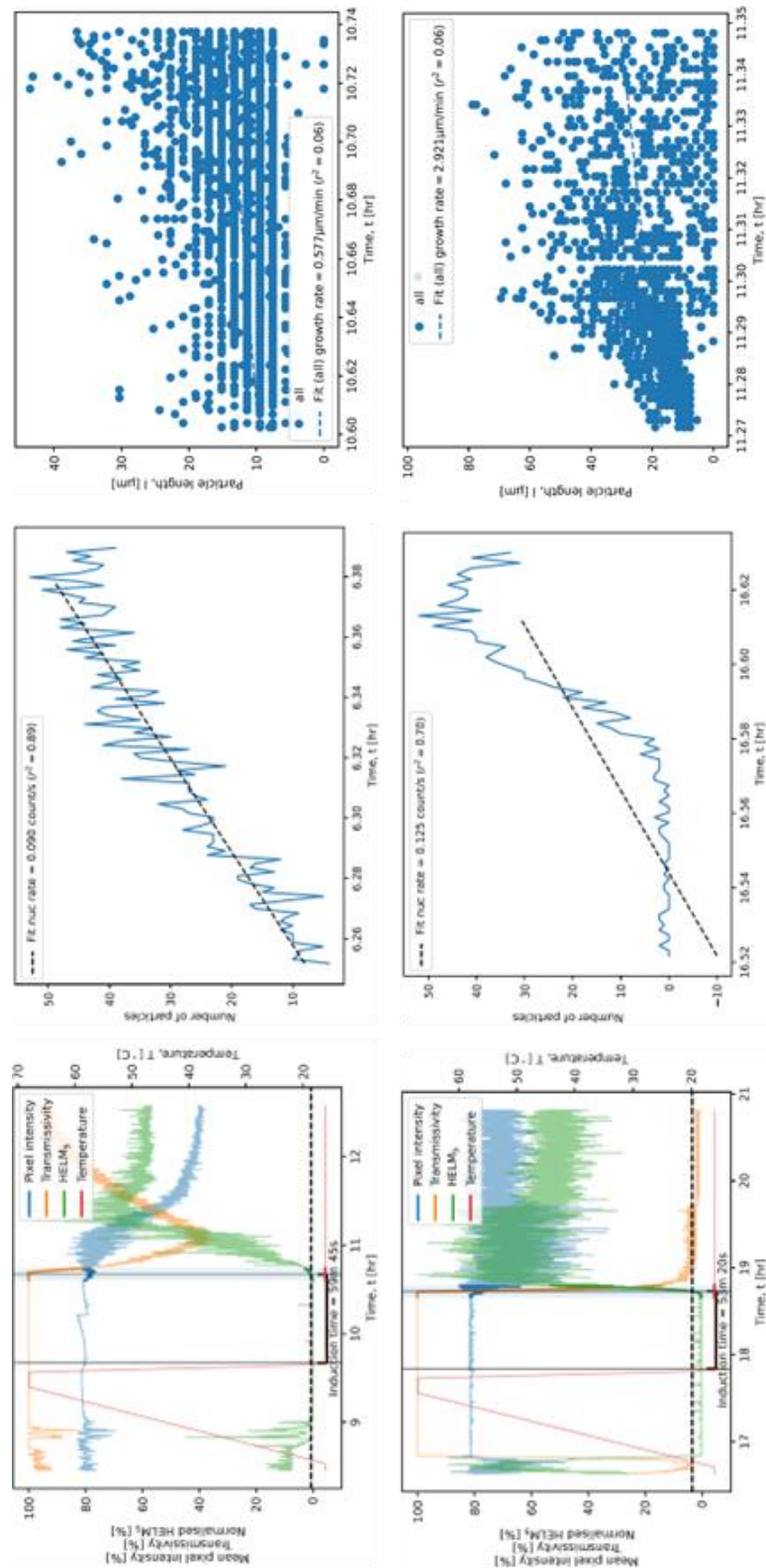
*Figure 45. Induction time plot (1a), nucleation rate plot (1b), and growth rate plot fitted to all bins (1c) for optimised run of lamivudine and induction time plot (2a), nucleation rate plot (2b), and growth rate plot fitted to all bins (2c) for optimised run of aspirin. The low $R^2$ for the growth rate graphs is due to fitting the growth rate to all data rather than a particular bin size to increase population size.*
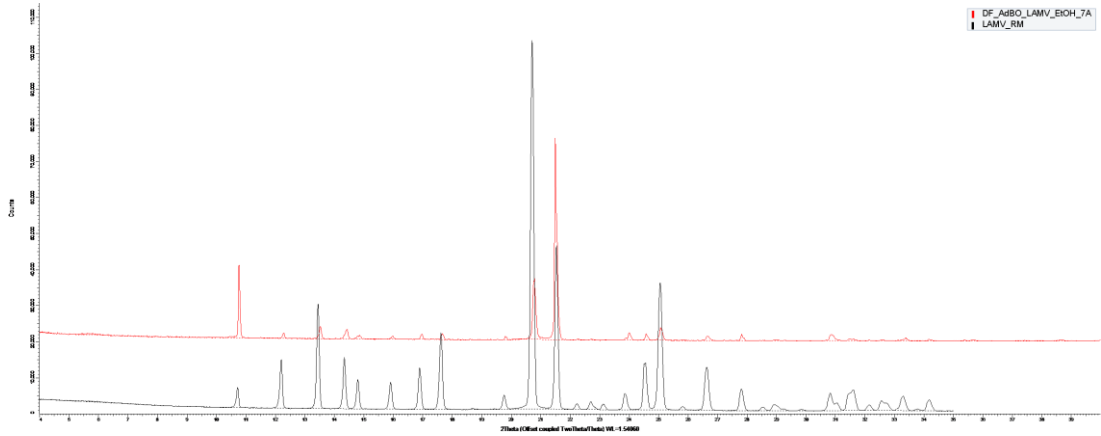
*Figure 46. XRPD pattern for recrystallised lamivudine overlayed against lamivudine raw material.*
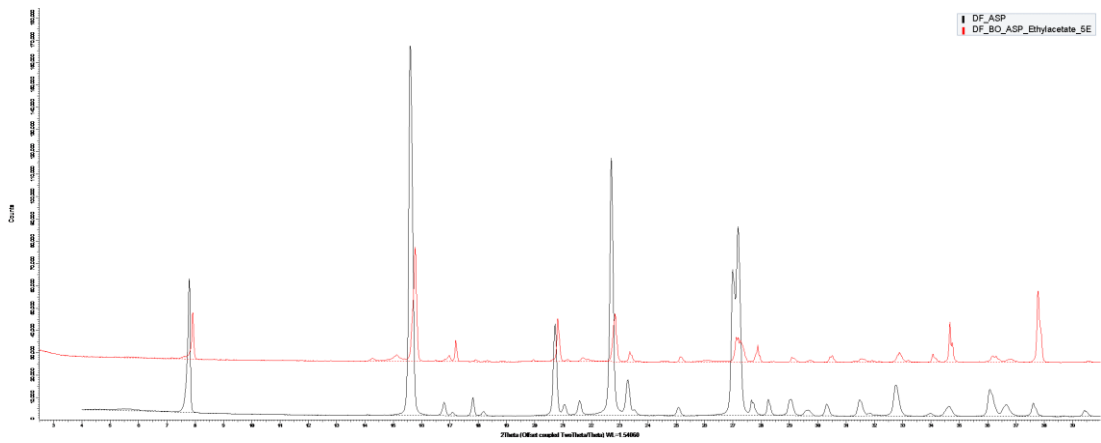


*Figure 47. XRPD pattern for recrystallised aspirin overlayed against aspirin raw material.*

# 7. Utilisation of the model-driven workflow for the digital design of small-scale batch cooling crystallisation of a broader scope of the chemical space.

## Abstract

This study presents five case studies of the model-driven workflow for the digital design of small-scale batch cooling crystallisation. The workflow allowed for the efficient collection of comprehensive thermodynamic data for ibuprofen, ascorbic acid, salicylic acid, benzoic acid and D-mannitol. The subsequent solvent system ranking and selection then allowed for the local kinetic parameter estimation and optimisation for all five APIs in a chosen solvent. The five APIs demonstrated a breadth of molecular and physical properties which exemplified the generalisability of the workflow.

## 7.1. Introduction

The use of workflows has demonstrated an increase in efficiency for data collection that is findable, accessible, interoperable, and reusable (FAIR).[139]
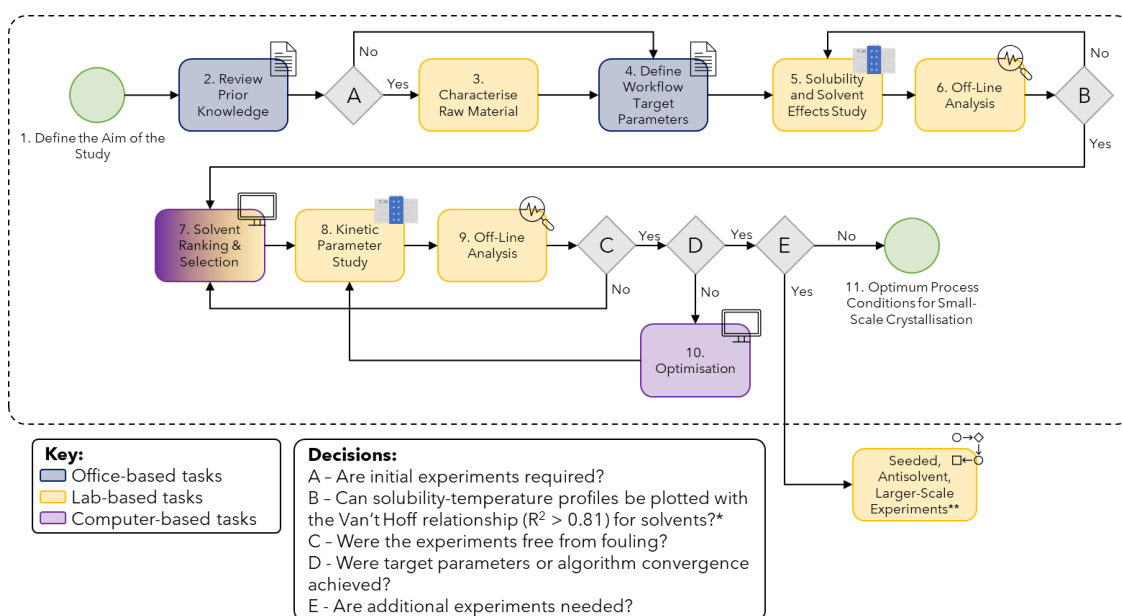


*Figure 48. Workflow for small scale batch cooling batch crystallisation parameter data collection, figure taken from Chapter 4.*

As this work, presented in this chapter, is a continuation of previous studies (Chapter 4 and 5) the same workflow for crystallisation parameter data collection was used. Additional advancements have been made to the data filters, optimisation decision-making loop (Chapter 6) and image analysis[191] allowing for even more efficient data collection for the five active pharmaceutical ingredients (APIs) presented in this chapter.

The solutes were chosen based on relevance (i.e. NHS demands and production volumes) and the goal to include a range of molecular attributes across the thesis. Five APIs were chosen for this study to draw comparisons and for a breadth of chemical structure to further test the generalisability of the workflow. For example, salicylic acid, an anti-inflammatory[206], and benzoic acid, an antimicrobial[207], have similar molecular structures differing by an *otho*-hydroxy group. Ascorbic acid, vitamin C[208], and D-mannitol, a diuretic[209] and common excipient, have very different molecular structures. Additionally, this study features ibuprofen, a common anti-inflammatory[210], which features both branched hydrocarbon chains

and an aromatic ring. The similarities and differences of the five APIs can also be seen in Table 15.

Table 15. LogP, pKa and molecular weight for the five APIs used in this study.[211]

| API | LogP | pKa | Molecular weight (g/mol) |
|---|---|---|---|
| Ibuprofen | 3.97 | 4.45 | 206.28 |
| Ascorbic acid | -1.85 | 4.70 | 176.12 |
| Salicylic acid | 2.26 | 2.78 | 138.12 |
| Benzoic acid | 1.87 | 4.21 | 122.12 |
| D-mannitol | -3.10 | 13.5 | 182.17 |

There is a large spread of values for both LogP (D-mannitol is the most hydrophilic and ibuprofen is the most hydrophobic) and for pKa (salicylic acid is the most acidic and D-mannitol is the least) within the five APIs. There is also a range of molecular weights (benzoic acid is the lowest and ibuprofen is the highest) with some molecules having similar weights.

## 7.2. Materials, Equipment and Methods

### 7.2.1. Materials

The materials used in this study are detailed in Table 16 and Table 17 and none underwent further purification.

Table 16. Detailed information regarding the solutes used in this study.

| Chemical name | Source | Grade (%) | DCS Group* |
|---|---|---|---|
| Ibuprofen | Molekula | >98 | I |
| Ascorbic acid | Alfa Aesar | 98+ | III |
| Salicylic acid | Alfa Aesar | 99 | n/a |
| Benzoic acid | Alfa Aesar | 99 | n/a |
| D-Mannitol | Alfa Aesar | 97+ | III |

*Some APIs are mainly manufactured as topical treatments and therefore do not have a developability classification system (DCS) group.

The solvent choice aimed to incorporate a wide range of functional groups to have physical and chemical properties diversity. Refer to Chapter 4 for a more comprehensive explanation of how the solvent library was chosen.

*Table 17. Detailed information regarding the solvents used in this study grouped by functional group.*

| Chemical name | Source | Grade (%) |
|---|---|---|
| Methanol | VWR | 99.95 |
| Ethanol | VWR | 99.95 |
| IPA [Propan-2-ol] | Alfa Aesar | 99+ |
| Pentan-2-ol | Alfa Aesar | 99 |
| Butan-1-ol | Alfa Aesar | 99 |
| Acetone | Thermo Scientific | 99+ |
| Pentan-2-one | Thermo Scientific | 99 |
| Pentan-3-one | Alfa Aesar | 99 |
| MIBK [Methyl iso-Butyl Ketone or 4-Methyl-2-pentanone] | Alfa Aesar | 99 |
| NMP [N-Methylpyrrolidone] | Thermo Scientific | 99+ |
| Ethyl acetate | VWR | 99.9 |
| iso-Propyl Acetate | Alfa Aesar | 99+ |
| iso-Butyl Acetate | Alfa Aesar | 98 |
| tert-Butyl Acetate | Alfa Aesar | 99 |
| Isopentyl Acetate [Isoamyl acetate] | Alfa Aesar | 99 |
| Butyl Cellosolve Acetate [2-n-Butoxyethyl acetate] | Thermo Scientific | 98 |
| Formamide | Alfa Aesar | 99 |
| DMF [N,N-Dimethylformamide] | Alfa Aesar | 99 |
| 2-methyl THF | Thermo Scientific | 99+ |
| n-Propyl Propionate | Alfa Aesar | 99 |
| n-Pentyl Propionate | Sigma Aldrich | >99 |
| Acetonitrile | Alfa Aesar | 99 |
| Chlorobenzene | Alfa Aesar | 99 |
| Chloroform | Alfa Aesar | 99.8+ |
| Cyclohexane | Alfa Aesar | 99+ |
| Cyclopentane | Alfa Aesar | 95 |
| Diethyleneglycol diethyl ether | Alfa Aesar | 99 |
| Dimethyl Carbonate | Alfa Aesar | 99 |
| Ethylene Glycol [1,2-Ethanediol] | Alfa Aesar | 99 |
| n-Heptane | Alfa Aesar | 95 |
| Tetralin | Alfa Aesar | 97 |
| Toluene | Alfa Aesar | 99 |
| Water | Milli-Q (Merck) | Ultra-Pure |

### 7.2.2. Equipment

In this study, images taken during the crystallisation experiments were collected using Technobis's Crystalline Reactor system[125] at a rate of one image every five seconds. Each of the eight reactor vials heats and cools following its own temperature profile, and every reactor has an HD camera focused on the lower part of the vial. X-ray powder diffraction (XRPD) patterns were collected using a Bruker D8 Discover, and the data was visualized using DIFFRAC.EVA[142] software from Bruker. Solid and solvent dosing was carried out using a Zinsser Crissy GGXXL robotic platform, which recorded solid and liquid dosed masses with a precision of +/- 0.005 mg. The Zinsser platform was used to dose amounts in the range of 5 to 1000 mg of solid and 1 to 8 mL of solvent. The Quintix (Sartorius)[212] analytical balance was used for the kinetic parameter study.

### 7.2.3. Methods

#### 7.2.3.1. Solubility and Solvent Effects Study (Polythermal Global Search)

The stir rate was set at 600 rpm for the template profile unless stated otherwise.

1. Heat the solution to 90 °C, or 10 °C below the solvent boiling point at a heating rate of 0.5 °C/min and hold at the elevated temperature for 10 minutes.
2. Cool the solution to 5 °C at a cooling rate of 0.5 °C/min and hold at the lowered temperature for 10 minutes.
3. Repeat the cycle for a total of 3 times.

The clear point was defined as the temperature at which no particles are present within the images collected, similarly, the cloud point was the temperature at which crystals first started to form. Crystal product was filtered under vacuum and dried under nitrogen to allow for XRPD patterns to be collected.

#### 7.2.3.2. Kinetic Parameter Study (Isothermal Local Search)

The stir rate was set at 600 rpm for the template profile unless stated otherwise.

1. Heat the solution to 90 °C, or 10 °C below the solvent boiling point at a heating rate of 0.5 °C/min and hold at the elevated temperature for 10 minutes.

2. Cool the solution to the experimental temperature at a cooling rate of 10 °C/min, with the stirring rate set to 0 rpm.
3. Hold at the isothermal experimental temperature for 2 hours.
4. Repeat the cycle for a total of 5 cycles.

Images, collected at a frequency of 1 image per 5 seconds, were analysed using a convolutional neural network (CNN) image analysis algorithm.[191] Induction time, defined as the time from reaching isothermal conditions to the detection of 5 particles or more, growth rate (change in particle size over time) and secondary nucleation rate (change in number of particles per image over time) were subsequently extracted. If measured kinetic parameter values were not achievable then quantitative values were assigned to allow for the optimisation algorithm to continue. Qualitative assignments were done in such a way that if nucleation was not observed then the assigned parameters were five times below the target parameters or if the solute did not dissolve (i.e., too high of a concentration) then the parameters were five times above the target parameter.

The crystal product was filtered under vacuum and dried under nitrogen to allow for XRPD patterns to be collected.

The experimental planner for the optimisation loop associated with the kinetic parameter study was an application of the GPyOpt[200] Bayesian Optimisation (BO) model with an adaptive exploration/ exploitation acquisition jitter. The general objective function (f(x)) of the optimisation was:

$$\textit{Equation 18.} \quad f(SS,T) = D_{SS,T}(Response)$$

where $D_{SS,T}$ is the difference between the target and measured objective at a given supersaturation (SS) and temperature (T). A full derivation of this objective function can be found in Section 6.3.2.2. using Equation 16 and Equation 17.

The BO model employed a Gaussian process (GP) model and expected improvement (EI) acquisition with acquisition jitter between 0.1 and 10. More details about the application of the adaptive Bayesian optimisation (AdBO) model can be found in Chapter 6. The termination criteria for convergence were a temperature of +/- 2 °C & supersaturation of +/- 0.02 between recommended experiments.

## 7.3. Results and Discussion

The work laid out below is an application of the workflow used in Chapter 4 to several APIs and thus each section header refers to a specific box of the workflow (Figure 24).

### 7.3.1. Define the Aim of the Study

The aim of this study was to collect quantitative and qualitative solubility and solvent effects data on a wider range of APIs from Chapter 4 and 5 and then subsequently optimise the small-scale batch cooling crystallisation process in a chosen solvent system. This study aims to show the generalisability of the workflow into different APIs.

### 7.3.2. Review Prior Knowledge

Experimental solubility data has been previously reported for ibuprofen[213-215], ascorbic acid[216], salicylic acid[214, 217-222], benzoic acid[214, 223-228] and D-mannitol[229]. This data was used to guide initial experiments and validate our methods.

### 7.3.A. Decision A: Are initial experiments required?

Yes, analysis of raw material was required (see section 7.3.3. below for details).

### 7.3.3. Characterise Raw Material

XRPD was used to determine the form of the raw material for all APIs to allow for fingerprinting of experimentally crystallised material against the most stable form (see Figure 52 in the Appendix of Chapter 7). No other raw material analysis was performed.

### 7.3.4. Define Workflow Specific Target Parameters

The target parameters for the solubility and solvent effects study were an API solubility of 0.005 g/g at low temperature and 0.05-0.25 g/g[145] at an elevated temperature with an aspect ratio of above 0.5[146]. These physical properties were chosen as they ensure a well-suspended slurry whilst retaining a high crystallisation yield with good powder flowability of the resulting drug product. At this stage of the workflow, yield and crystal shape are not optimised but, instead, used as constraints (e.g., only solvent systems that give crystals with an aspect ratio exceeding 0.5 are accepted) to subsequently rank possible solvent choices.

The optimisation objectives of the kinetic parameter estimation had target values as shown in Table 18. These target parameter values correspond to typical values that lead to operable conditions in larger-scale crystallisation processes.[165] Additional parameter constraints were included in the kinetic parameter estimation including a d90 size distribution between 100 μm and 250 μm and a mean aspect ratio above 0.5 so that the resultant powder would be free-flowing[146].

Table 18. Target parameters for induction time, nucleation rate and growth rate (optimisation responses) for the kinetic parameter study.

| API | Target induction time (s) | Nucleation rate (#/image/s) | Growth rate (μm/s) |
| --- | --- | --- | --- |
| Ibuprofen | 3600 | 0.1 | 0.1 |
| Ascorbic Acid | 3600 | 0.1 | 0.01 |
| Salicylic Acid | 3600 | 0.1 | 0.05 |
| Benzoic Acid | 3600 | 0.1 | 0.1 |
| D-Mannitol | 3600 | 0.1 | 0.1 |

The same target values were not used for all APIs due to feasibility of what is achievable within the metastable zone width (MSZW) of the API-solvent system.

### 7.3.5. Solubility and Solvent Effects Study (Polythermal Global Search)

The results from the solubility and solvent effects study can be summarised for all APIs in Figure 49 where the use of colour classifications allows for quick visualisation.

| Solvent\ Solute | Ibuprofen | Ascorbic Acid | Salicylic Acid | Mannitol | Benzoic Acid |
|---|---|---|---|---|---|
| 2-methyl THF | TS | AS | TS | AS | TS |
| Acetone | TS | AS | TS | AS | TS |
| Acetonitrile | TS | Y-A | Y-G | AS | Y-A |
| Butan-1-ol | TS | Y-A | Y-A | AS | TS |
| Butyl Cellosolve Acetate (2-n-Butoxyethyl acetate) | TS | AS | Y-A | AS | Y-A |
| Chlorobenzene | TS | AS | Y-G | AS | Y-G |
| Chloroform | TS | AS | Y-G | AS | TS |
| Cyclohexane | Y-A | OO | N | AS | Y-G |
| Cyclopentane | Y-A | AS | PM | AS | Y-G |
| Diethyleneglycol diethyl ether | TS | AS | TS | AS | TS |
| Dimethyl Carbonate | TS | AS | Y-A | AS | Y-A |
| DMF [N,N-Dimethylformamide] | TS | CC | TS | WS | TS |
| Ethanol | TS | DNN | Y-A | AS | TS |
| Ethyl acetate | TS | AS | TS | AS | TS |
| Ethylene Glycol [1,2-Ethanediol] | Y-G | DNN | DNN | WS | WS |
| Formamide | Y-A | CC | DNN | DNN | TS |
| IPA [Propan-2-ol] | TS | Y-A | TS | AS | TS |
| iso-Butyl Acetate | TS | AS | Y-A | AS | Y-A |
| Isopentyl Acetate [Isoamyl acetate] | TS | AS | Y-A | AS | Y-A |
| iso-Propyl Acetate | TS | AS | Y-A | AS | Y-A |
| Methanol | TS | Y-G | TS | AS | TS |
| MIBK [Methyl iso-Butyl Ketone or 4-Methyl-2-pentanone] | TS | AS | Y-A | AS | Y-A |
| n-Heptane | Y-A | AS | OO | AS | Y-A |
| NMP [N-Methylpyrrolidone] | TS | DNN | N | WS | TS |
| n-Pentyl Propionate | TS | AS | Y-A | AS | Y-A |
| n-Propyl Propionate | TS | AS | Y-A | AS | Y-A |
| Pentan-2-ol | TS | Y-A | Y-A | AS | TS |
| Pentan-2-one | TS | AS | Y-A | AS | TS |
| Pentan-3-one | TS | AS | Y-A | AS | TS |
| tert-Butyl Acetate | TS | AS | Y-A | AS | Y-A |
| Tetralin | TS | AS | Y-G | AS | Y-G |
| Toluene | TS | N | Y-G | AS | Y-A |
| Water | AS | TS | WS | WS | WS |

| | |
|---|---|
| AS | Antisolvent[1] |
| N | Not run |
| PM | Phobic Material |
| Y-G | **Yes - Good** (in solubility range[2], no needles) |
| Y-A | **Yes - Acceptable** (near desired solubility[3] range limits) |
| WS | in solubility range[2] OR near[3] but **Wrong Shape** (i.e. needles) |
| DNN | in solubility range[2] OR near[3] but **Did Not Nucleate** |
| OO | in solubility range[2] OR near[3] but **Oiled Out** |
| CC | **Colour change** |
| TS | **Too Soluble** |

$T_1 = 50\,°C$ and $T_2 = 5\,°C$ ($T_1$ for cyclopentane is 39°C)

[1] Solubility < 5 g/kg at $T_1$

[2] Desired solubility range: at temp $T_1$, solubility is 50-250 g/kg AND at temp $T_2$ solubility is > 5 g/kg

[3] Near desired solubility range: at temp $T_1$, solubility is 5-50 g/kg OR 250-500 g/kg AND at temp $T_2$ solubility is > 5 g/kg

*Solubility units of g/kg refer to gAPI/kgSOLVENT*

*Figure 49. Rubric classifying all APIs in all the solvents for the solubility study.*

The 'too soluble' pink classification resulted in an inoperable process where the high concentration (more than 0.5 g/g solvent at elevated temperature) meant a thick slurry at low temperatures which did not allow for effective mixing. On the other side of the spectrum the 'antisolvent' blue classification was where the API had a very low concentration (less than 0.005 g/g solvent at elevated temperature) which would not be a feasible solvent choice due to excessive solvent consumption and low yield. The 'phobic material' grey classification is where API sticking to the glass walls of the reactor was observed. The 'wrong shape', 'did not nucleate' and 'oiled out' purple classifications were where the solubility target parameters were met or near to the range, but a commonly seen but undesirable issue was observed. The 'colour change' black classification was where a colour change was observed within the vial, often going from a colourless solution to yellow likely due to high concentrations. The 'yes – acceptable' amber classification was where no fouling, desired crystal habit and near target parameter solubility was observed. Finally, the 'yes – good' green classification is where all target parameters for solubility and shape have been met, meaning it is the most desired classification. More details on the ranking and selection of the solvent system can be found in Section 3.7.

Ibuprofen was too soluble (solubility > 0.5 g/g solvent at elevated temperature) in most organic solvents, this resulted in thick slurries at low temperatures which did not allow for stirring. This issue was the reason for poor $R^2$ values (Figure 53 and Table 22 in the Appendix of Chapter 7) for some of the solvents. Dissolution of ibuprofen in ethylene glycol aligned with target parameters (solubility of 0.005 g/g at low temperature and 0.05-0.25 g/g solvent[145] at an elevated temperature) however, had a MSZW that exceeded 36 °C which approximately is a supersaturation of 50, a value that greatly exceeds what would be feasible for higher temperature operations. Solvent systems that were near the target parameters included cyclohexane, cyclopentane, formamide and n-heptane which were all above the range limits. Recoverable crystal product all had a similar habit of plates.

Ascorbic acid solubility was qualitatively assigned as in antisolvent (solubility < 0.005 g/g solvent at elevated temperature) in most solvents where this concentration represents the lower dosing of the Zinsser limit combined with the highest operating temperature of the solvent in the Crystalline. On the other hand,

ascorbic acid was too soluble in water, resulting in thick hard to control slurries. Some solvent systems such as acetonitrile, 1-butanol, 2-propanol and 2-pentanol were near the target parameters on the lower side. Oiling out was observed for cyclohexane and colour changes were observed for DMF and formamide. The dissolution of ascorbic acid in methanol met target solubility parameters. Recoverable crystal product all had a similar habit of plates.

Salicylic acid had high solubility in most organic solvents where some were classified as too soluble or near the target parameters (these were on the high side). Qualitative observations of salicylic acid in cyclopentane showed sticking to the vial walls signifying a phobic interaction between the API/ solvent system. Oiling out was observed for salicylic acid in n-heptane. The solubility target parameters were achieved for salicylic acid in acetonitrile, chlorobenzene, chloroform, tetralin and toluene. All solvent systems gave a similar crystal habit of blocks/ rods except for water which exhibited needles.

Benzoic acid also had high solubility in most organic solvents with some too soluble and some near target parameters (slightly above the target range). The target parameters for solubility were met for benzoic acid in chlorobenzene, cyclohexane, cyclopentane and tetralin. Long plates/ blocks crystal habit was observed for all solvent systems except for ethylene glycol and water which exhibited needles.

D-mannitol was qualitatively assigned as in antisolvent for all organic solvent systems except DMF, ethylene glycol, formamide and NMP where they had solubility near the target parameters but on the low side. These solvents typically had higher solubility for all APIs. Mannitol had high solubility in water (near the target parameters on the high side) and exhibited needle crystal habit, like the forementioned four organic solvents.

It can be qualitatively observed, from Figure 49, that salicylic acid and benzoic acid have similar solubility classifications with some increases in solubility (e.g., acetonitrile and toluene). The addition of the hydroxy group to the base benzoic acid molecule being the likely reason for the increase. Ascorbic acid and D-mannitol had qualitatively similar solubilities where they mainly had very low affinity for organic solvents and high solubility in the more polar solvents (e.g., water).

Ibuprofen stands out from the other solvents due to having very high solubility in mostly all organic solvents and very minimal solubility in water.

### 7.3.6. Off-Line Analysis

All APIs were fingerprinted by XRPD and showed that all recoverable crystal product was the same as the starting material (see Figure 58 to Figure 62 in the Appendix of Chapter 7), the most stable form, except for 2 discrepancies. D-mannitol in NMP and N,N-DMF had the same XRPD pattern that differed to the starting material (see Figure 63 in the Appendix of Chapter 7); however, these were not solvents of interest, so no further characterisation was done.

### 7.3.B. Decision B: Can solubility-temperature profiles be plotted with the Van't Hoff relationship ($R^2 > 0.81$) for solvents?

Yes, refer to Table 22 in the Appendix of Chapter 7. There were a few discrepancies e.g., ascorbic acid in ethylene glycol, but these were not solvents of interest so were qualitatively ignored from further study. Refer back to Table 5 in the Appendix of Chapter 4 for more details on the uncertainty checks.
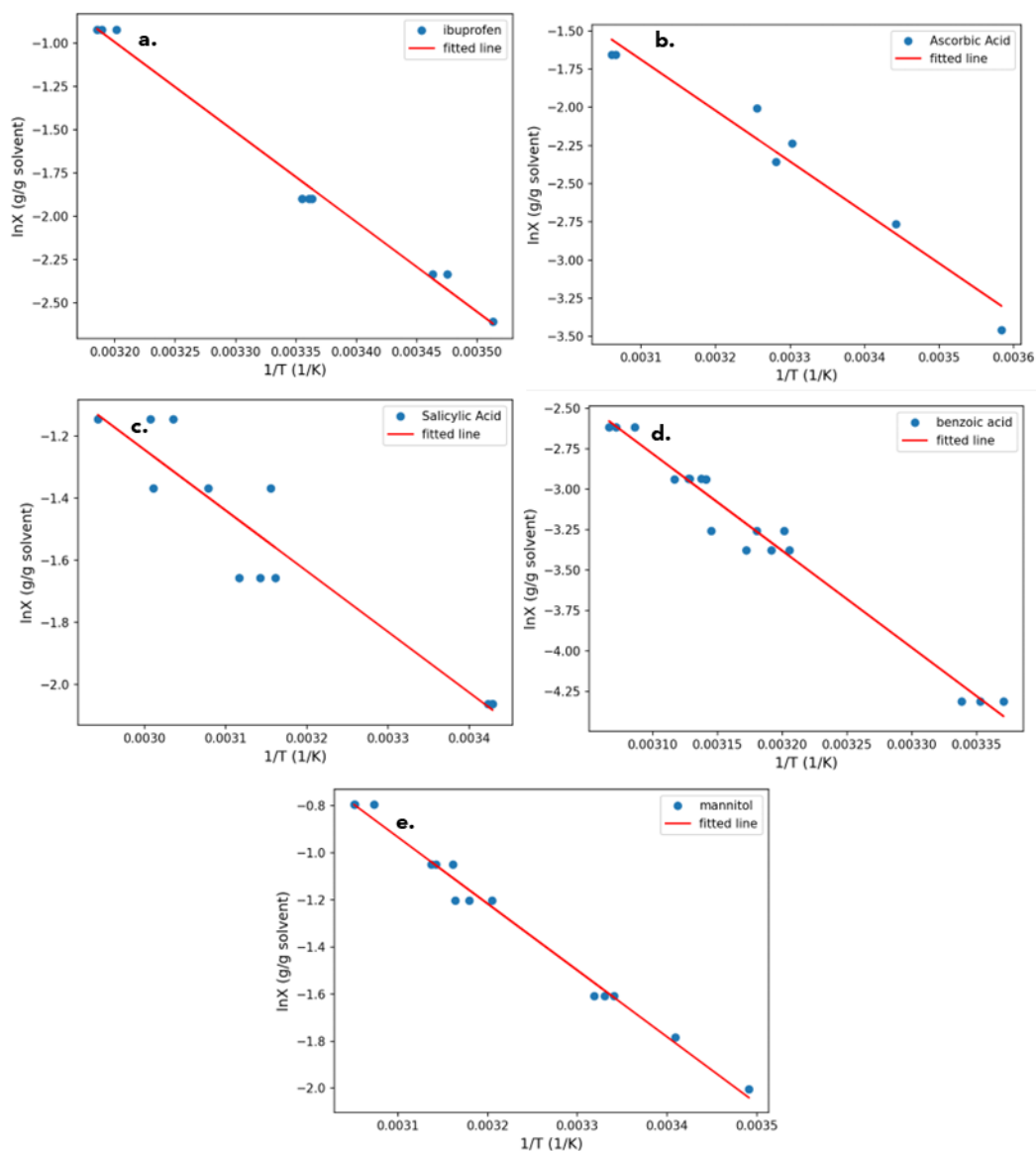
*Figure 50. Van't Hoff plots for ibuprofen in 0.95:0.05 (v/v) heptane: 2-pentanol (a), ascorbic acid in methanol (b), salicylic acid in acetonitrile (c), benzoic acid in cyclohexane (d) and D-mannitol in water (e). Solubility data points were done in triplicate with no outliers removed as the linear regression met the criteria of $R^2 > 0.81$.*

## 7.3.7. Solvent Ranking & Selection

For ibuprofen the best, in terms of solubility parameters (solubility of 0.005 g/g at low temperature and 0.05-0.25 g/g[145] at an elevated temperature) and shape parameters (aspect ratio greater than 0.5[146] i.e., plates), single solvent system was ethylene glycol. The high supersaturation required for nucleation however makes this an unfeasible process solvent choice. The next best choice was n-heptane however the system did not have a steep temperature dependence. This would theoretically result in a low yield. The introduction of a second solvent, to make a

binary mixture, was explored where various compositions of alcohols were added. 95:5 (v/v) heptane: 2-pentanol resulted in the steepest temperature dependence within the solubility constraints. This aligned with SAFT-γ Mie group-contribution solubility predictions.[230]

For ascorbic acid, the best solvent system was methanol due to the good solubility compared to the generally low solubility in most other systems.

For salicylic acid, the best solvent system was acetonitrile due to the good solubility, yielded good crystal habit and had a steep temperature dependence. However downward reduction with subsequent cycles (Figure 50c) in the solubility temperature was observed indicating potential degradation at higher temperatures.

For benzoic acid, the best solvent system was cyclohexane due to the good solubility, yielded good crystal habit and had the steepest temperature dependence of the four potential solvent choices.

For D-mannitol, the chosen solvent system was water as was near the target parameters in terms of solubility and was the most stable polymorph (this ruled out NMP and N,N-DMF). Water was chosen over ethylene glycol and formamide due to being a greener and more sustainable solvent choice. Persistent needles of D-mannitol recrystallised in water meant that the solvent choice did not meet the target value for the shape parameters therefore the introduction of optimising for shape in later study was required. A quantitative assignment of 0.75 for the aspect ratio optimisation objective was added to the objective function for D-mannitol. This assignment was a mid-point target value between 0.5 (shape target parameter from the solubility study of which to exceed) and 1.0 (maximum aspect ratio possible).

### 7.3.8. Local Search: Kinetic Parameter Study

The design space for each of the API/ solvent systems were determined based upon what was feasible within the MSZW where fast induction was expected for all systems except ascorbic acid in methanol. Table 19 shows the design space and the first experiment, the mid-point for both variables of the design space, and the resultant objective function value from the measured kinetic parameters.

Nucleation was only experimentally observed for salicylic acid in acetonitrile therefore quantitative values were qualitatively assigned for the other systems to allow for the Bayesian optimisation algorithm to suggest the next experiment.

*Table 19. Initial 1-experiment screening results for the kinetic parameter estimation for all APIs. SS refers to supersaturation.*

| API | Design Space | | First Experiment | | Obj. Function Value |
|---|---|---|---|---|---|
| | SS | Temperature (°C) | SS | Temperature (°C) | |
| **Ibuprofen** | 1.05 – 2 | 5 – 50 | 1.48 | 27.5 | 4.83 |
| **Ascorbic Acid** | 2 – 3 | 5 – 50 | 2.5 | 27.5 | 6.00 |
| **Salicylic Acid** | 1.05 – 2 | 5 – 50 | 1.48 | 27.5 | 5.51 |
| **Benzoic Acid** | 1.05 – 2 | 10 – 50 | 1.48 | 27.5 | 3.50 |
| **D-Mannitol** | 1.05 – 2 | 5 – 50 | 1.48 | 27.5 | 3.67 |

High objective function values, when compared to previous implementations of this algorithm (Chapter 6), and that qualitative assignments were required shows that the initial experiments were far from the optimum process conditions.

### 7.3.9. Off-Line Analysis

All APIs were fingerprinted by XRPD and showed that all recoverable crystal product was the same as the starting material, the most stable form (see Figure 64 to Figure 68 in the Appendix of Chapter 7).

### 7.3.C. Decision C: Were the experiments free from fouling?

Yes, there was no significant fouling experimentally observed for any of the APIs with their chosen solvent system (visually observed in Section 7.3.8.).

### 7.3.D. Decision D: Were target parameters or algorithm convergence achieved?

No, as only one experiment had been performed in the initial study it was inconclusive whether the optimum process conditions had been found. Experimental optimisation was required.

## 7.3.10. Optimisation

Following the initial iteration of the kinetic parameter study an experimental-optimisation loop was started that terminated with the following results:

*Table 20. Results of the experiment-optimisation loop that was terminated due to successful convergence of a temperature of +/- 2 °C & supersaturation of 0.02 between recommended experiments.*

| API | Number of experiments | Obj. Function Value | Reduction in the objective function value (%)* |
|---|---|---|---|
| Ibuprofen | 19** | 0.38 | 93.5 |
| Ascorbic Acid | 21 | 3.30 | 74.2 |
| Salicylic Acid | 35 | 2.26 | 82.6 |
| Benzoic Acid | 10 | 1.27 | 81.4 |
| D-Mannitol | 30 | 1.41 | 84.6 |

*Reduction is calculated as the minimum objective function value/ maximum objective function value (as a percentage).

**14 experiments due to the AdBO acquisition jitter of 10, 5 experiments due to the acquisition jitter of 1 due to the reduction in the objective function triggering the adaptive model.

A good reduction in the objective function was calculated for all five APIs with excellent reduction especially for ibuprofen, sufficiently enough to trigger a more exploitative Bayesian optimisation model. The AdBO model performed well for all APIs and arrived at the termination criteria generally quickly (between 10 and 30 experiments) when compared to the ADoE methods in Chapter 6. The optimisation design planner performed best, in terms of number of experiments required, for ibuprofen and benzoic acid, but required the greatest number of experiments for salicylic acid. This is likely accounted for by the very fast nucleation of salicylic acid in acetonitrile causing there to be many local optimums at low supersaturations i.e., supersaturation impacted the measured kinetic parameters a lot more than temperature.
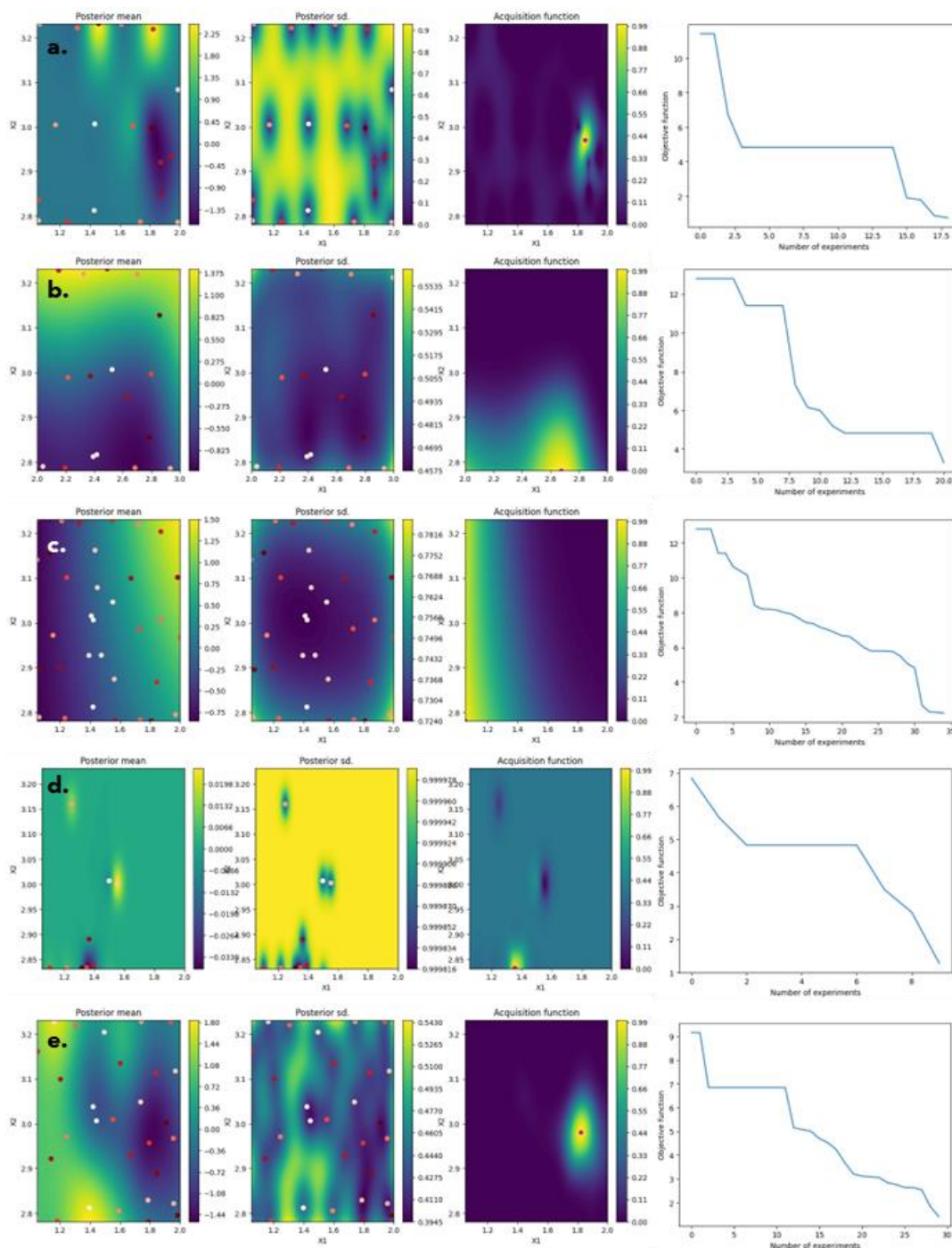
*Figure 51. Plot of posterior mean, posterior standard deviation, acquisition function and objective function value vs number of experiments for ibuprofen (a), ascorbic acid (b), salicylic acid (c), benzoic acid (d) and D-mannitol (e) after the experimental optimisation had terminated. (X1 refers to supersaturation and X2 refers to temperature (K/100)).*

It can be seen in Figure 51(left) that there is a global minimum for each of the APIs and that the algorithm has generally exploited that area with more experiments. It can be observed from Figure 51(second left) that there has been reduction in the uncertainty across the design space, especially well for ibuprofen, ascorbic acid,

salicylic acid and D-mannitol but poor reduction for benzoic acid. The poor performance in reducing the uncertainty of the design space for benzoic acid is likely contributed to only 10 experiments being performed before termination. However, the reduction in objective function (Figure 51(right)) and that the algorithm exploited and converged still suggests that the minimum was found. The acquisition function heatmap (Figure 51(second right)) was plotted using an acquisition jitter of $1E^{-7}$ to visualise the minimum i.e., optimum process conditions solution.

## 7.3.D. Decision D: Were target parameters or algorithm convergence achieved? (Revisited)

Yes, convergence was achieved as determined by the termination criteria of a temperature of +/- 2 °C & supersaturation of 0.02 between recommended experiments.

## 7.3.E. Decision E: Are additional experiments needed?

No. A suitable process was developed using cooling crystallisations for all APIs; therefore seeded, antisolvent and scaled-up experiments were not required to meet the objectives for this study.

## 7.3.11. Optimum Process Conditions for Small-Scale Crystallisation (Summary)

The workflow has allowed for the identification of experimental conditions for optimum process attributes (good solubility to balance between easy-to-handle slurries and not high solvent waste) and crystal attributes (desired polymorph and shape and size).

*Table 21. Optimum process conditions for small-scale crystallisation for the 5 APIs with relevant measured kinetic parameters (all processes had a stir rate of 600 RPM).*

| API | Solvent choice | Temp. (°C) | SS | Induction time (s)* | Nucleation rate (#/image/s)* | Growth rate (μm/s)* | Aspect ratio* |
|---|---|---|---|---|---|---|---|
| **Ibuprofen** | 0.95:0.05 (v/v) heptane: 2-pentanol | 23.92 | 1.85 | 2476 (3600) | 0.080 (0.100) | 0.076 (0.100) | n/a |
| **Ascorbic acid** | Methanol | 5** | 2.67 | 4410 (3600) | 0.011 (0.100) | 0.004 (0.010) | n/a |
| **Salicylic acid** | Acetonitrile | 5** | 1.05** | 2683 (3600) | 0.075 (0.100) | 0.257 (0.050) | n/a |
| **Benzoic acid** | Cyclohexane | 10.14 | 1.35 | 5152 (3600) | 0.123 (0.100) | 0.049 (0.100) | n/a |
| **D-manntiol** | Water | 24.99 | 1.82 | 6098 (3600) | 0.096 (0.100) | 0.071 (0.100) | 0.45 (0.75) |

*Target values are in brackets.

**These are bounds of the design space.

Taking into consideration the desired solvent choice from Section 7.3.7. and following the kinetic study (Section 7.3.8.) and optimisation (Section 7.3.10.) loop the final process conditions for small-scale batch cooling crystallisation can be stated, as found in Table 21. Good optimisation performance can be observed across most APIs for converging on the target induction time. For example, ascorbic acid had only a 23% difference from the target value. There was also great optimisation convergence performance for nucleation rate across most APIs however reduced performance was observed for ascorbic acid. D-mannitol had a 4% difference from the target value for nucleation rate, which is an example of great convergence. The lowest performing response was growth rate where differences between measured and target values ranged from 24% (ibuprofen) and 400% (salicylic acid). This can be attributed to competing objectives meaning that low objective function values for induction time and nucleation rate allow a larger objective function value for growth rate whilst still maintaining an overall low objective function value. Additionally, there are problems associated with assigning appropriate growth rate target parameters without a larger design of experiment (DoE) initial screening. It was observed throughout the study that growth rate varied the most between APIs.

As mentioned in Section 7.3.7. persistent needles were observed for D-mannitol so therefore an additional fourth objective was introduced into the optimisation

problem. An aspect ratio (defined as width/ length) of 0.75 was quantitatively assigned, as that value was the mid-point of the target range with the focus to getting greater than 0.5. Despite not achieving 0.5, the final value of 0.45 was a significant improvement over measured values in the study being as low as 0.126.

Finally, it can be observed in Table 21 that the optimum values for temperature for ascorbic acid and for supersaturation and temperature for salicylic acid sit on the bounds of the design space. This suggests that the true optimum was not found and that a reduced objective function value sits outside of the bounds. The boundaries of the design space were set in accordance with what is achievable with the equipment available i.e., accuracy of balances and temperature controllers. The true optimum supersaturation for salicylic acid of likely being less than 1.05 would be impractical to experimentally perform due to error associated with the solubility-temperature profile and error of the Crystalline temperature controller propagating which could result in undersaturated solutions.

## 7.4. Conclusions

Five case studies have been demonstrated of the workflow without amendment to the logic set out in the introduction. This shows great generalisability of the workflow across a range, in terms of chemical properties, of APIs giving confidence in the logical flow of the workflow. This work demonstrated that both quantitative and qualitative thermodynamic data can be collected fast and effectively, as the data collection was done in circa. 8 months. This thermodynamic data was then used to inform solvent selection for kinetic parameterisation which was concluded in between 10 and 35 experiments (circa. 1 month). Process crystallisation routes have been proposed for all five APIs that give both good solubility and shape attributes and desired kinetic parameters at the small scale.

The integration of a fourth objective into the kinetic parameter optimisation for D-mannitol demonstrates the modulus approach to the AdBO experimental planning model. This will allow for the introduction of new factors and new responses as the workflow is applied to the data collection of more APIs.

This work has further reinforced that the individual components, specifically the image analysis and Bayesian experimental planner, work as standalone tools and when integrated into a workflow. The tested logic of the workflow, through the five API case studies, can in the future be applied to robotic integration and autonomous systems. There is still a need to extend this study to more systems with greater molecular flexibility and polymorphic and solvate formation propensities to understand how the approach behaves in a potentially more complex and multi-objective response surface.

## Appendix

This appendix contains additional information for context and details pertaining to material characterisation and extracted data figures and tables.
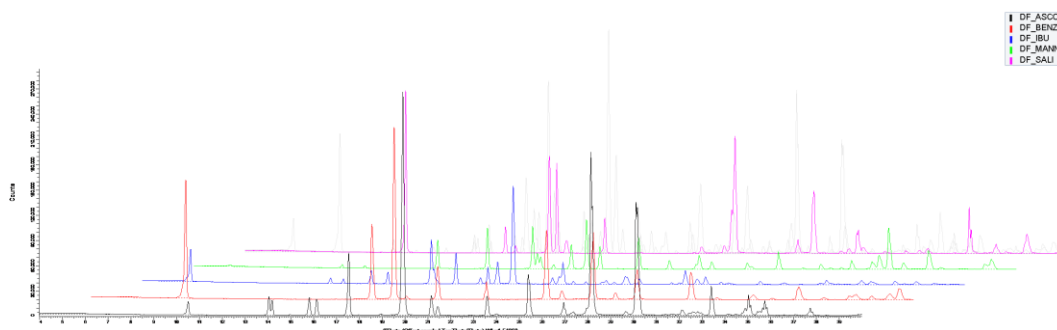


*Figure 52. XRPD patterns of raw material for all five APIs.*

*Table 22. Van't Hoff solubility linear regression coefficients and $R^2$ for 5 APIs in various solvents.*

| Solvent | m | c | $R^2$ |
|---|---|---|---|
| **Ibuprofen** | | | |
| 1-butanol | -7893.94 | 25.29 | 1.00* |
| 2-n-butoxyethyl acetate | -1543.61 | 4.30 | 0.96 |
| 2-pentanol | -1407.53 | 3.99 | 0.79 |
| 2-pentanone | -1533.61 | 4.68 | 0.86 |
| 2-propanol | -890.29 | 2.54 | 0.90 |
| 3-pentanone | -5220.92 | 16.43 | 0.99 |
| 4-methyl-2-pentanone | -840.19 | 2.51 | 0.33 |
| N,N-dimethylformamide | -5563.39 | 18.33 | 0.82 |
| N-methylpyrrolidone | -1692.82 | 4.96 | 0.86 |
| Water | 28092.60 | -89.21** | 1.00* |
| acetonitrile | -3283.29 | 10.17 | 0.97 |
| chlorobenzene | -4092.40 | 12.32 | 0.98 |
| chloroform | -2527.21 | 7.35 | 1.00 |
| cyclohexane | -2536.95 | 7.43 | 0.80 |
| cyclopentane | -675.80 | 1.49 | 0.96 |
| diethyleneglycol diethyl ether | -1447.67 | 4.40 | 0.82 |
| dimethyl carbonate | -1047.68 | 2.79 | 0.65 |
| ethanol | -995.41 | 3.04 | 0.53 |
| ethylene glycol | -12113.81 | 35.75 | 0.99 |
| formamide | -32159.63 | 100.89 | 0.96 |
| heptane | -3398.60 | 9.84 | 0.91 |
| heptane/ 1-butanol 0.95 (v/v) | -5011.60 | 15.06 | 1.00 |

| | | | |
|---|---|---|---|
| heptane/ 1-propanol 0.95 (v/v) | -4845.38 | 14.59 | 0.99 |
| heptane/ 2-pentanol 0.95 (v/v) | -5162.40 | 15.53 | 0.99 |
| heptane/ ethanol 0.95 (v/v) | -5490.33 | 16.83 | 1.00 |
| isobutyl acetate | -707.50 | 1.93 | 0.20 |
| isopentyl acetate | -3774.28 | 11.45 | 0.81 |
| isopropyl acetate | -1959.13 | 6.00 | 0.76 |
| methanol | -266.42 | 0.65 | 0.02 |
| n-pentyl propionate | -2215.68 | 6.49 | 1.00 |
| n-propyl propionate | -858.39 | 2.46 | 0.34 |
| tert-butyl acetate | -9121.64 | 28.65 | 0.83 |
| tetralin | -3027.84 | 8.90 | 0.97 |
| toluene | -938.93 | 2.73 | 0.20 |
| **Ascorbic Acid** | | | |
| 1-butanol | -2445.65 | 3.50 | 0.96 |
| 2-pentanol | -3839.69 | 8.42 | 0.92 |
| 2-propanol | -3154.17 | 5.99 | 0.90 |
| 3-pentanone | -6722.14 | 14.65 | 1.00* |
| N-methylpyrrolidone | -4731.77 | 13.69 | 1.00 |
| Water | -3097.26 | 9.02 | 0.99 |
| acetonitrile | -14280.26 | 38.65 | 1.00* |
| ethanol | -1120.67 | 0.41 | 0.98 |
| ethylene glycol | -593.61 | -1.13 | 0.01 |
| methanol | -3481.07 | 9.15 | 0.95 |
| **Salicylic Acid** | | | |
| 1-butanol | -1989.77 | 5.52 | 1.00 |
| 2-n-butoxyethyl acetate | -1083.24 | 2.40 | 0.89 |
| 2-pentanol | -1419.52 | 3.72 | 1.00 |
| 2-pentanone | -1025.49 | 2.54 | 1.00 |
| 2-propanol | -2011.09 | 5.82 | 1.00 |
| 3-pentanone | -449.82 | 0.62 | 0.34 |
| 4-methyl-2-pentanone | -1325.08 | 3.36 | 1.00 |
| Water | -4839.64 | 10.32 | 0.94 |
| acetone | -4601.64 | 14.24 | 0.81 |
| acetonitrile | -2670.00 | 6.93 | 0.60 |
| chlorobenzene | -1836.66 | 3.84 | 0.99 |
| chloroform | -1526.73 | 2.43 | 1.00 |
| diethyleneglycol diethyl ether | -926.81 | 2.35 | 1.00* |
| dimethyl carbonate | -1690.81 | 3.86 | 0.65 |
| ethanol | -944.79 | 2.37 | 0.82 |
| ethylene glycol | -2925.64 | 7.75 | 0.93 |
| formamide | -1250.14 | 2.51 | 0.96 |
| isobutyl acetate | -1143.45 | 2.51 | 0.98 |
| isopentyl acetate | -2136.80 | 5.48 | 0.94 |

| | | | |
|---|---|---|---|
| isopropyl acetate | -2218.09 | 5.87 | 0.99 |
| methanol | -1538.58 | 4.67 | 0.98 |
| n-pentyl propionate | -1152.37 | 2.40 | 0.93 |
| n-propyl propionate | -1259.20 | 2.86 | 0.92 |
| tert-butyl acetate | -1266.07 | 2.79 | 1.00 |
| tetralin | -674.06 | -0.17 | 0.17 |
| toluene | -3008.13 | 7.30 | 0.94 |
| **Benzoic Acid** | | | |
| 1-butanol | -2888.22 | 8.57 | 0.98 |
| 2-n-butoxyethyl acetate | -2477.48 | 6.69 | 1.00 |
| 2-pentanol | -4304.69 | 13.29 | 0.97 |
| 2-pentanone | -1746.69 | 4.83 | 0.99 |
| 2-propanol | -3228.33 | 10.10 | 0.92 |
| 3-pentanone | -2412.66 | 6.96 | 1.00 |
| 4-methyl-2-pentanone | -2415.76 | 6.72 | 1.00 |
| N,N-dimethylformamide | -35372.86 | 125.90 | 1.00* |
| Water | -5715.39 | 13.90 | 1.00* |
| acetonitrile | -2666.90 | 7.21 | 0.95 |
| chlorobenzene | -3746.24 | 10.20 | 1.00 |
| chloroform | -8240.19 | 25.48 | 1.00 |
| cyclohexane | -5743.04 | 15.02 | 0.98 |
| cyclopentane | -4065.50 | 10.09 | 1.00 |
| diethyleneglycol diethyl ether | -1908.89 | 5.40 | 0.98 |
| dimethyl carbonate | -3363.36 | 9.38 | 1.00 |
| ethanol | -2777.68 | 8.70 | 0.90 |
| ethyl acetate | -2624.75 | 7.47 | 1.00* |
| ethylene glycol | -7044.39 | 20.96 | 0.88 |
| formamide | -3543.18 | 10.36 | 0.99 |
| heptane | -4431.57 | 10.65 | 1.00 |
| isobutyl acetate | -2797.48 | 7.65 | 0.99 |
| isopentyl acetate | -2782.60 | 7.50 | 1.00 |
| isopropyl acetate | -2601.62 | 7.15 | 1.00 |
| methanol | -2360.73 | 7.57 | 0.99 |
| n-pentyl propionate | -2825.43 | 7.49 | 1.00 |
| n-propyl propionate | -2538.25 | 6.84 | 1.00 |
| tert-butyl acetate | -3118.98 | 8.49 | 1.00 |
| tetralin | -4267.11 | 11.42 | 1.00 |
| toluene | -3779.33 | 10.36 | 1.00 |
| **D-Mannitol** | | | |
| N,N-dimethylformamide | -5741.31 | 13.41 | 0.99 |
| N-methylpyrrolidone | -4634.52 | 10.40 | 1.00 |
| Water | -2783.47 | 7.69 | 1.00 |
| ethylene glycol | -6708.56 | 16.90 | 0.99 |

| | | | |
|---|---|---|---|
| formamide | -4821.44 | 11.86 | 1.00* |
| methanol | -4654.67 | 8.95 | 0.90 |

*R$^2$ values of 1.00 were achieved as only 2 data points were measured.
**Large negative c coefficients were due to the Van't Hoff relationship having the opposite trend to what is expected due to extremely noisy data.
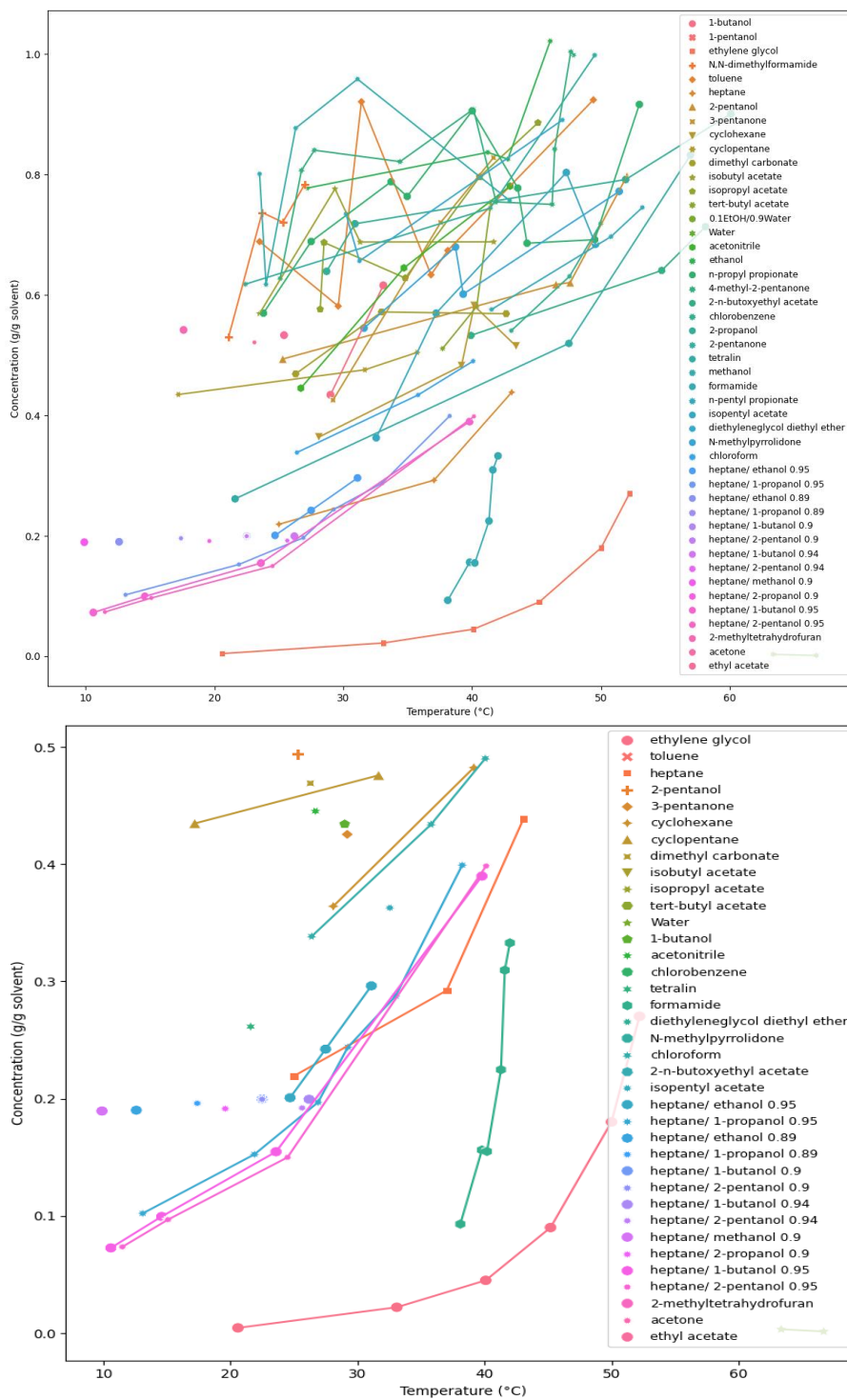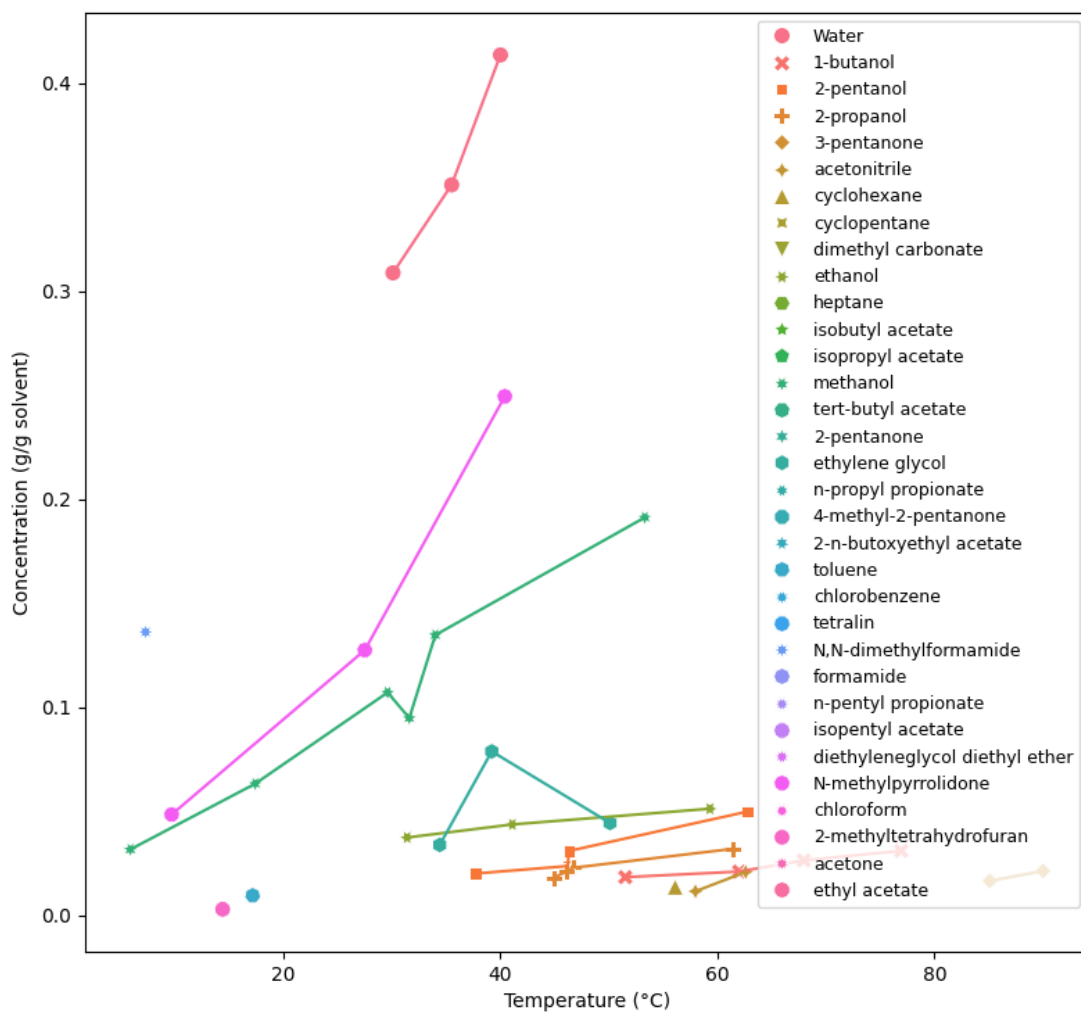
*Figure 53. Solubility-temperature profile for ibuprofen in various solvents where quantitative solubilities recorded over 0.5 g/g solvent were qualitatively assigned as too soluble and removed from the figure d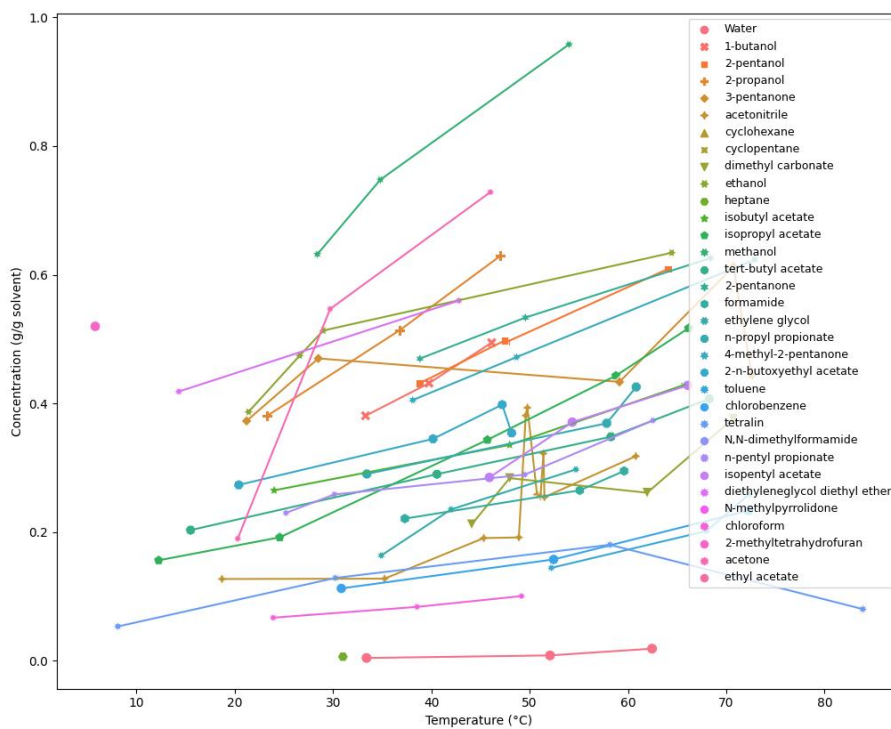ue to high uncertainty causing a poor representation of data (top) and solubility-temperature profile for ibuprofen in various solvents with no data removed (bottom).*

127

*Figure 54. Solubility-temperature profile for ascorbic acid in various solvents. Note: quantitative solubilities recorded over 0.5 g/g solvent were qualitatively assigned as too soluble and removed from the figure due to high uncertainty causing a convoluted representation of data.*
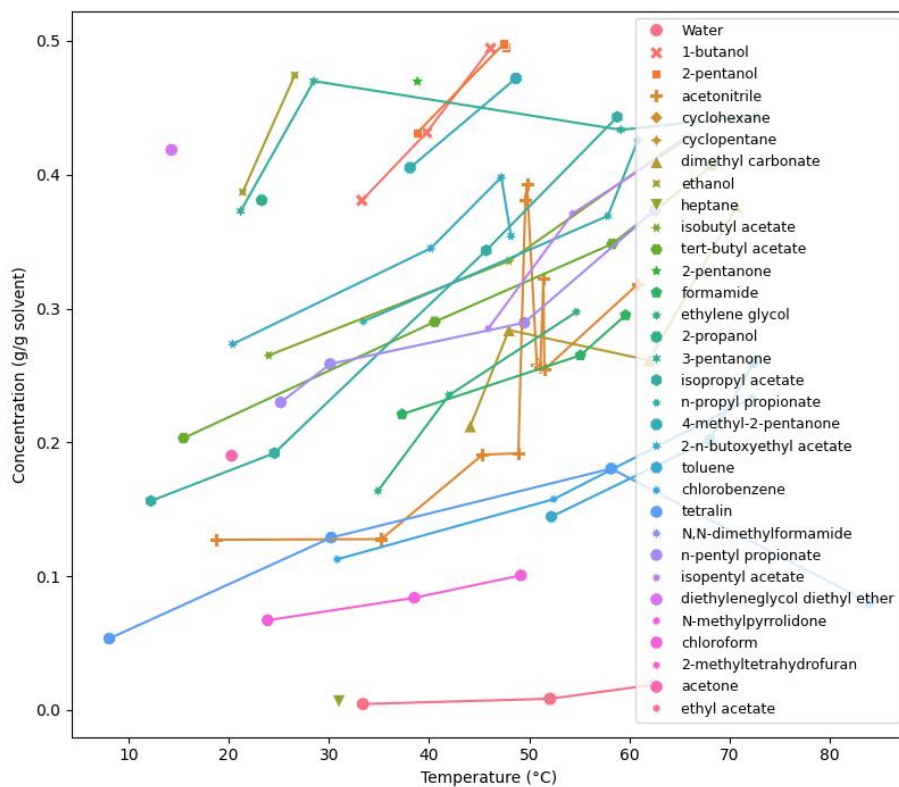
*Figure 55. Solubility-temperature profile for salicylic acid in various solvents where quantitative solubilities recorded over 0.5 g/g solvent were qualitatively assigned as too soluble and removed from the figure due to high uncertainty causing a poor representation of data (top) and solubility-temperature profile for salicylic acid in various solvents with no data removed (bottom).*
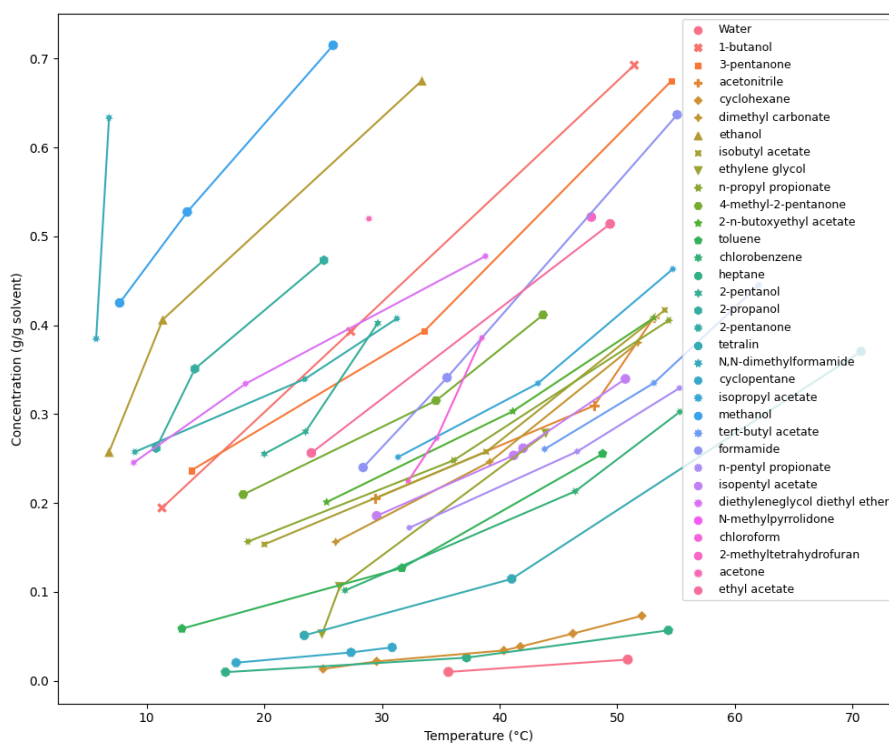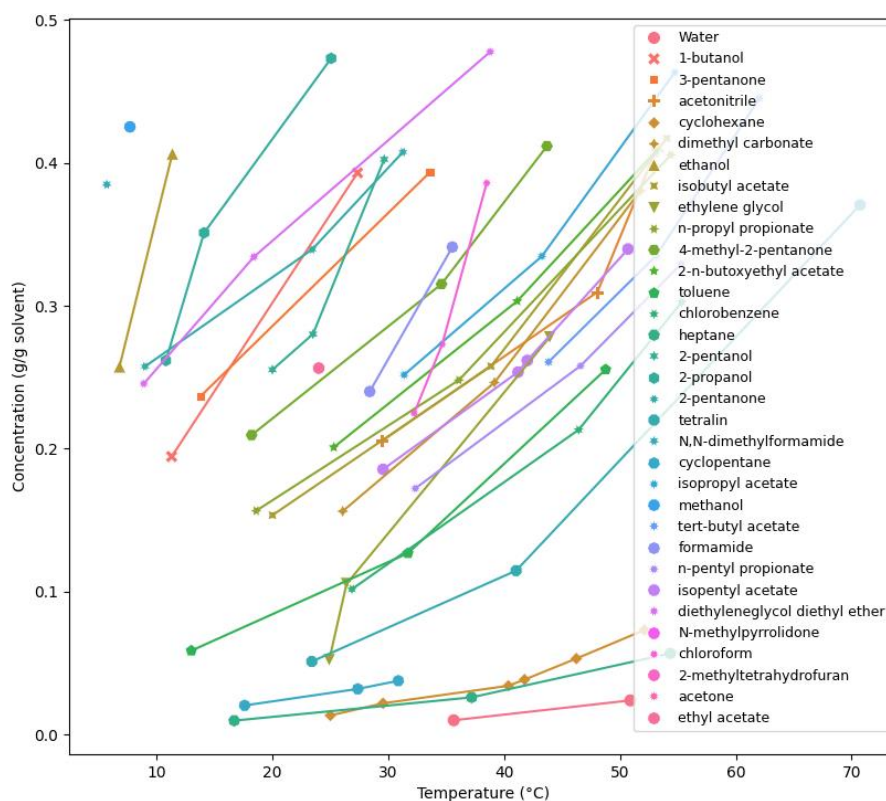
Figure 56. Solubility-temperature profile for benzoic acid in various solvents where quantitative solubilities recorded over 0.5 g/g solvent were qualitatively assigned as too soluble and removed from the figure due to high uncertainty causing a poor representation of data (top) and solubility-temperature profile for benzoic acid in various solvents with no data removed (bottom).
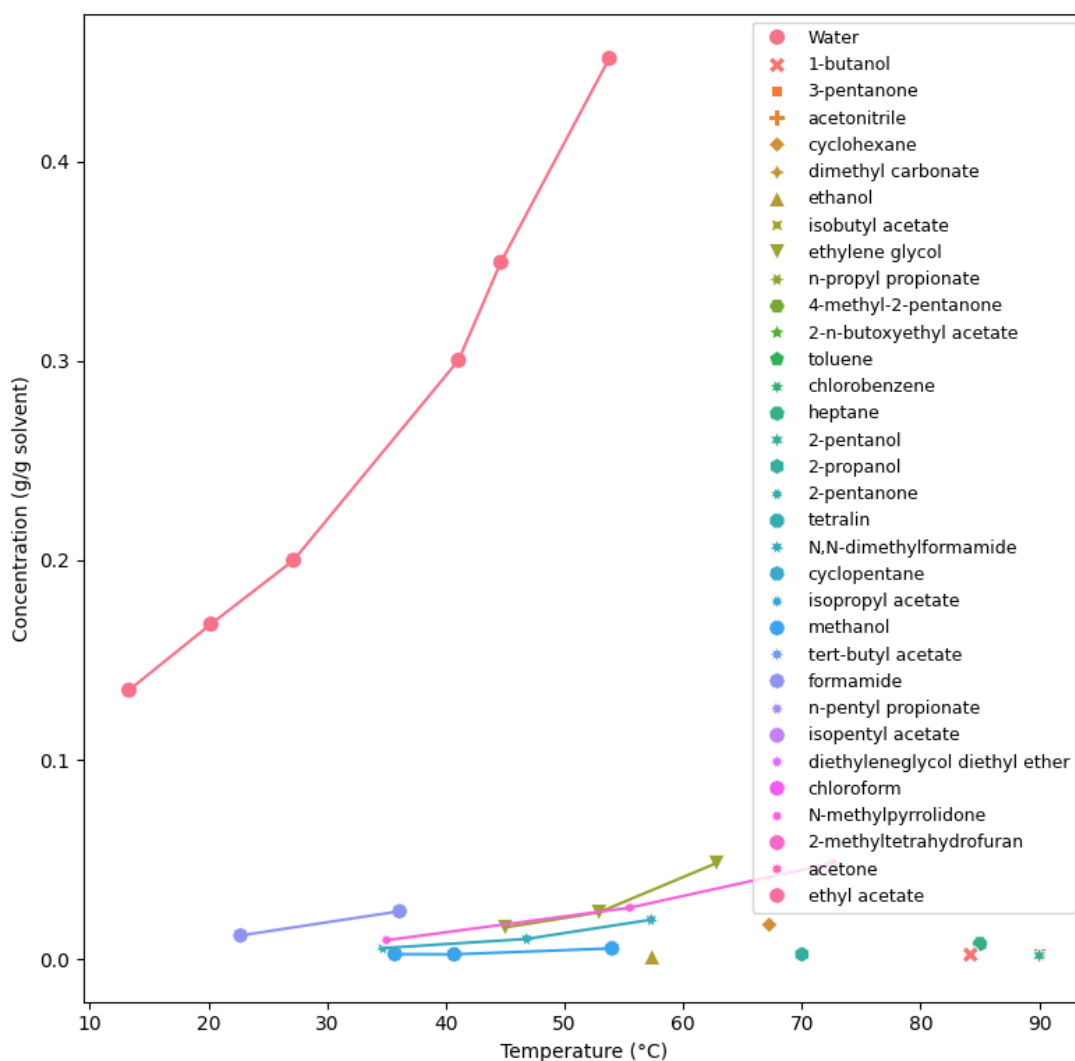
*Figure 57. Solubility-temperature profile for D-mannitol in various solvents. Note: quantitative solubilities recorded over 0.5 g/g solvent were qualitatively assigned as too soluble and removed from the figure due to high uncertainty causing a convoluted representation of data.*
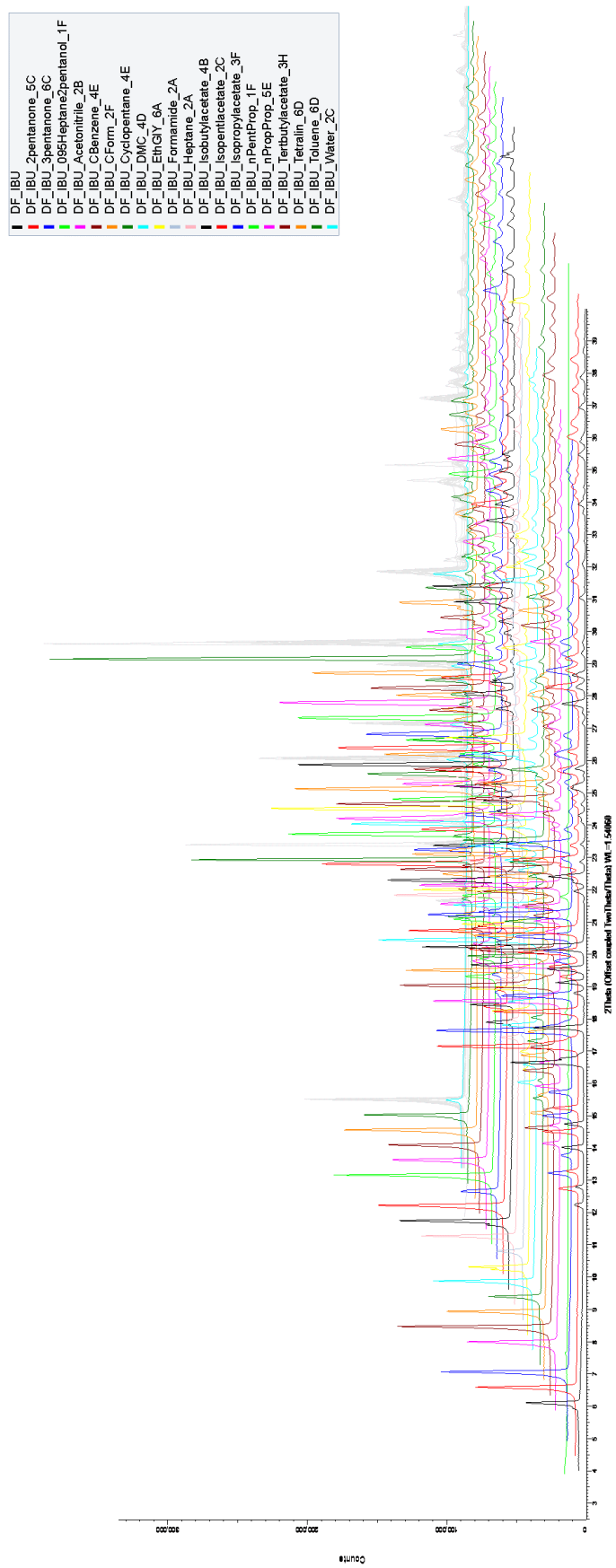
*Figure 58. XRPD patterns for the solubility and solvent effects study for ibuprofen in various solvents.*
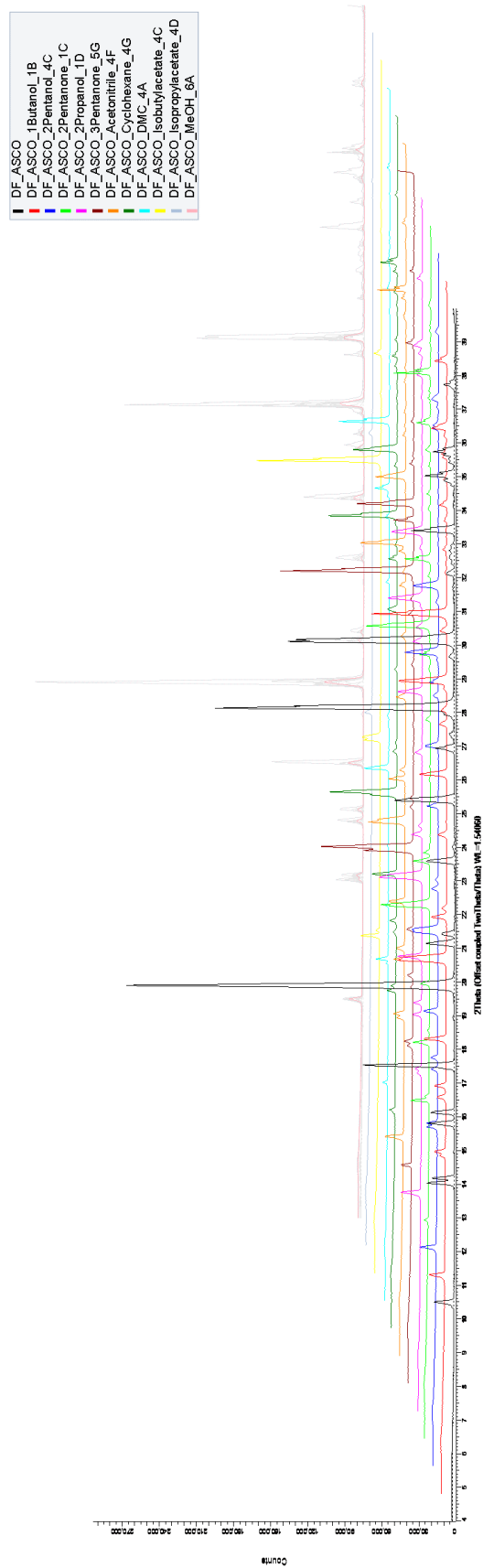
*Figure 59. XRPD patterns for the solubility and solvent effects study for ascorbic acid in various solvents.*
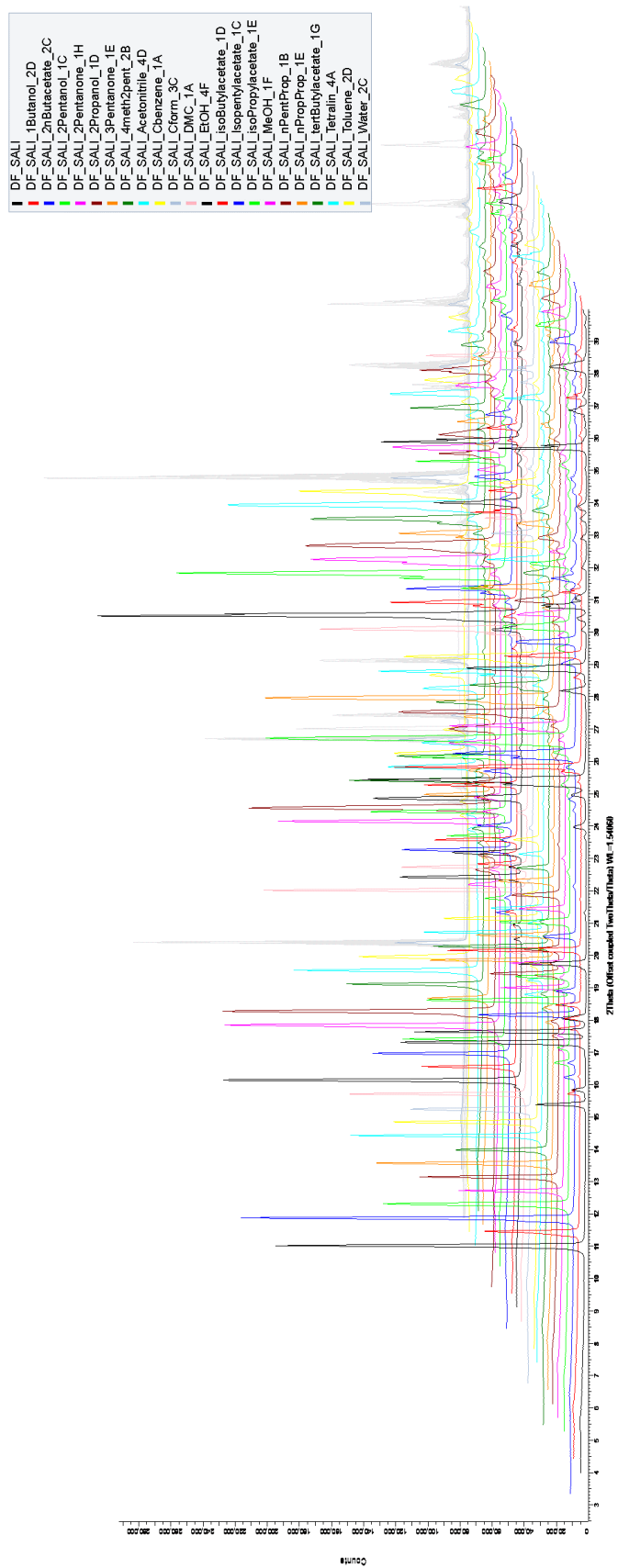
*Figure 60. XRPD patterns for the solubility and solvent effects study for salicylic acid in various solvents.*

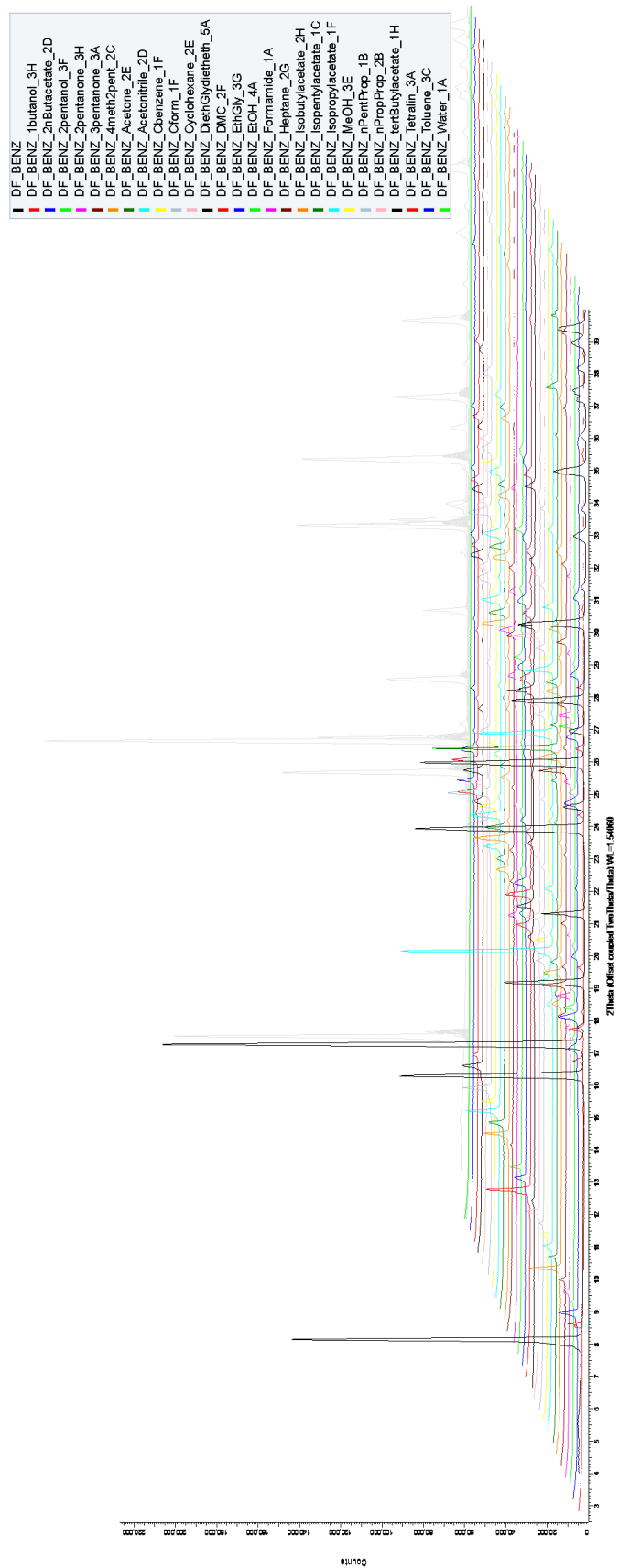*Figure 61. XRPD patterns for the solubility and solvent effects study for benzoic acid in various solvents.*
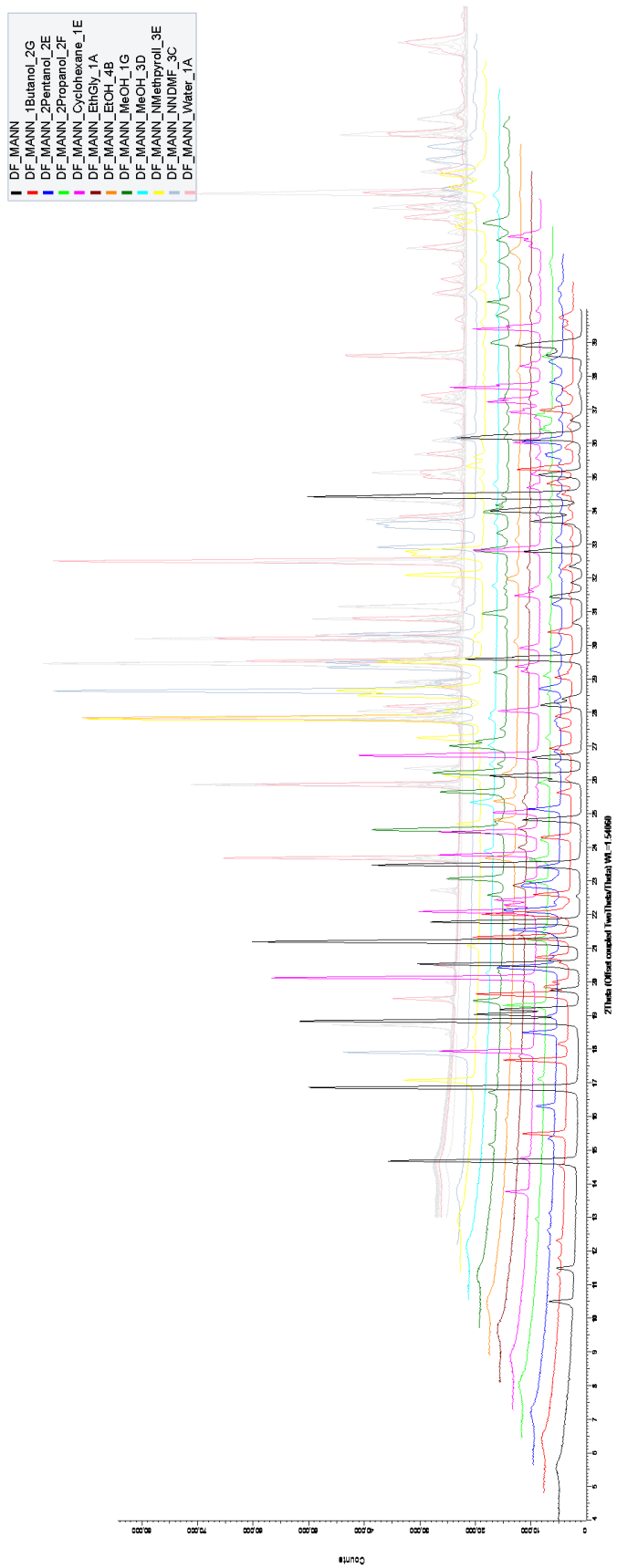
*Figure 62. XRPD patterns for the solubility and solvent effects study for D-mannitol in various solvents.*
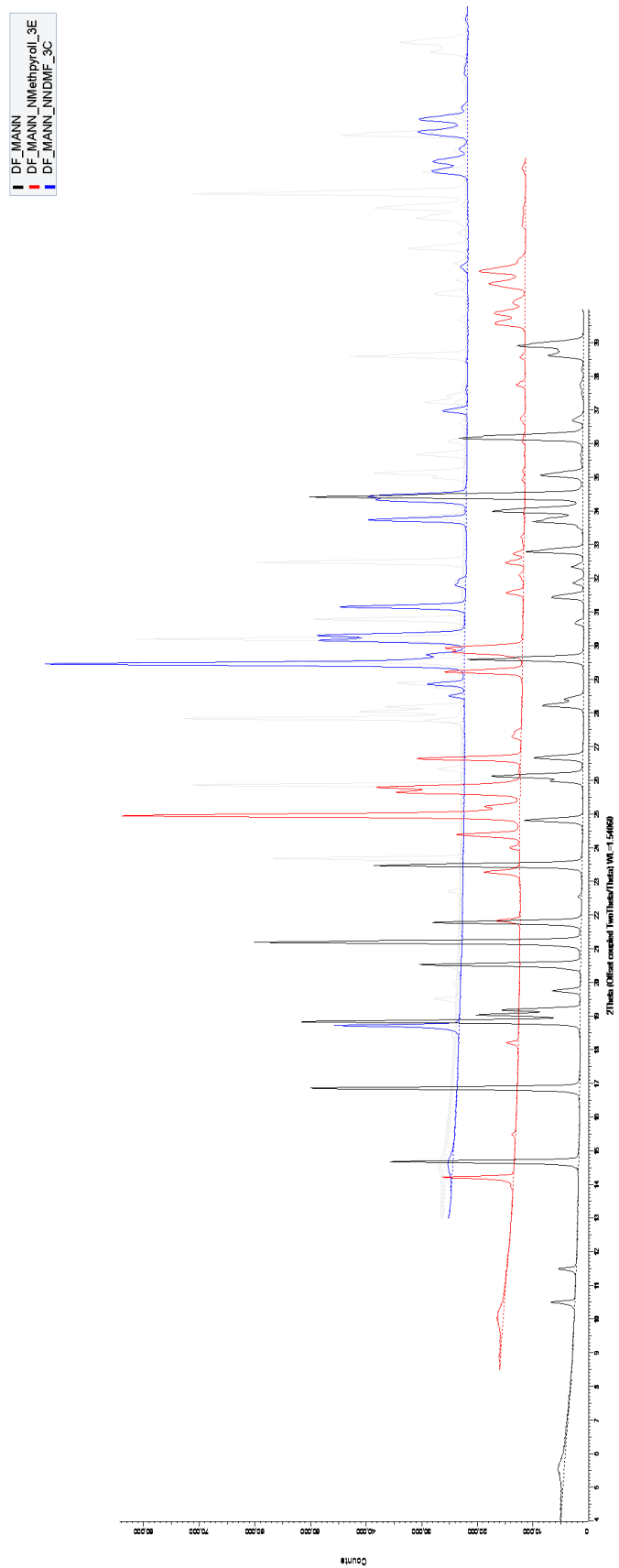
*Figure 63. XRPD patterns for the solubility and solvent effects study for D-mannitol in NMP and N,N-DMF fingerprinted against the raw material to highlight that a different form was present.*
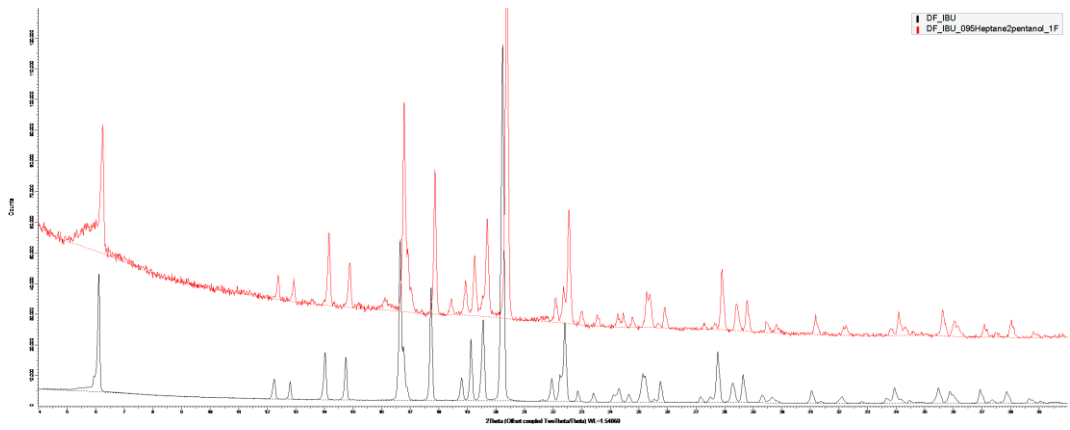
*Figure 64. XRPD patterns for the kinetic parameters study for ibuprofen, y-offset to allow for fingerprinting.*
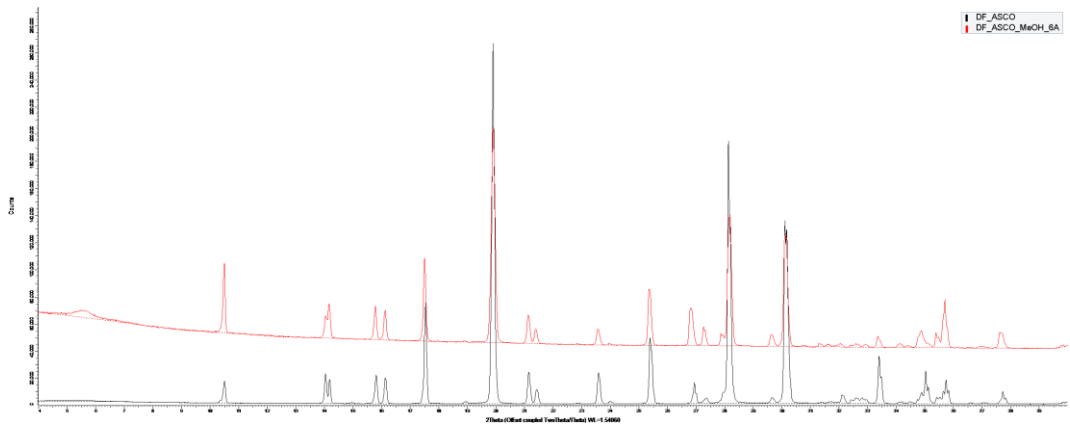


*Figure 65. XRPD patterns for the kinetic parameters study for ascorbic acid, y-offset to allow for fingerprinting.*



*Figure 66. XRPD patterns for the kinetic parameters study for salicylic acid, y-offset to allow for fingerprinting.*

*Figure 67. XRPD patterns for the kinetic parameters study for benzoic acid, y-offset to allow for fingerprinting.*
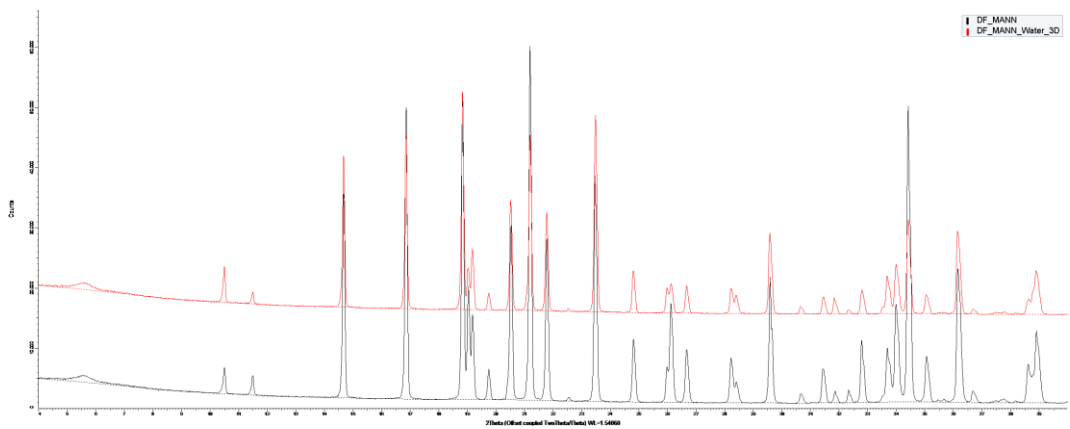


*Figure 68. XRPD patterns for the kinetic parameters study for D-mannitol, y-offset to allow for fingerprinting.*

# 8. Integration of a Model-Driven Workflow into an Industrial Pharmaceutical Facility: Supporting Process Development of API Crystallisation.
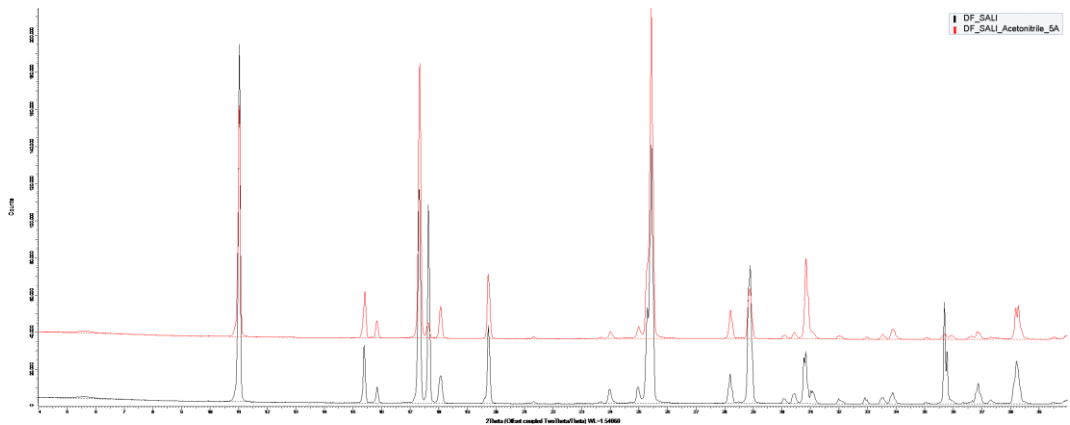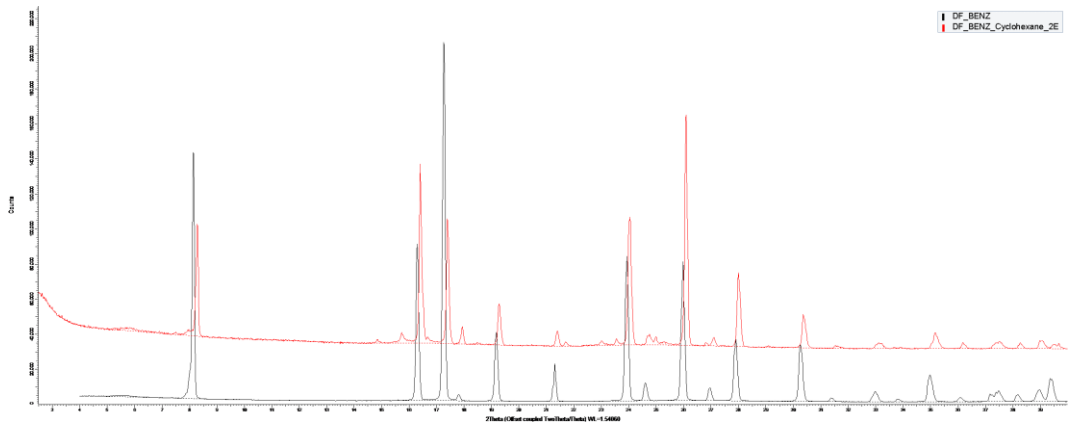
## Abstract

This study develops and evaluates a knowledge-driven workflow for a material- and resource-sparing approach to crystallisation process development. By following this workflow, thermodynamic and isothermal kinetic data for the cooling crystallization of (3S,5R)-3-(aminomethyl)-5-methyl-octanoic acid (PD-299685) within an 8-week timeframe, were obtained. Moreover, the workflow was expanded to include isothermal kinetic parameters from a scaled-up cooling crystallisation, as well as antisolvent and seeded crystallisation of PD-299685. The systematic and standardised data collection facilitated by this workflow enabled the design and optimisation of the PD-299685 crystallisation process. The proposed industrial crystallisation route for PD-299685 combines cooling and antisolvent techniques, offering a wide metastable zone width to facilitate speck-free filtration and effective seeding. This approach allows for excellent control over product quality, resulting in particles with a desired aspect ratio of 0.766 and a d(v,90) value of 234 μm. These parameters align with the proposed API material target specifications for solid oral dosage form quality, specifically oral bioavailability and content uniformity, and efficient drug product manufacture.

## 8. 1. Introduction

The use of crystallisation in the pharmaceutical industry offers several advantages, including purification and particle engineering. It serves as a purification technique to eliminate impurities, ensuring that drug products comply with regulatory requirements.[3] [1] Additionally, crystallisation can be employed for particle engineering by modifying solid-state characteristics such as polymorphism[15] and physical properties such as shape and size. These differences in properties can impact bioavailability, downstream processing and drug product manufacturing.[146]

(3S,5R)-3-(aminomethyl)-5-methyl-octanoic acid (PD-299685) had been clinically evaluated for the reduction of hot flashes associated with menopause as a replacement for hormone replacement therapy.[231] PD-299685 as the final product comes from a complex many-step synthesis where previous crystallisation routes included cooling crystallisation in 50:50 water/isopropanol or 50:50 water/ethanol by volume.[232] Given that the final crystallization had been the object of previous development efforts, this study set out to assess whether novel, workflow-based approaches to crystallization development would lead either to a more efficient process compared to that previously published, or the faster identification of a comparable process.

Typical workflows for developing crystallisation processes consist of solubility and metastable zone width (MSZW) measurements to find a suitable solvent that primarily gives a high yield within safe operating temperatures and the desired polymorph. Once a solvent system and crystallization type is chosen then various experiments can be performed, generally at 50 – 100 mL scale, to determine most kinetic parameters and whether seeding or other process controls, such as milling, are necessary to control additional product attributes such as purity or particle size. Process optimisation is usually carried out following known domain relationships and chemistry intuition, or by Design of Experiments (DoE)[48]. A step-by-step workflow to assist with consistent data collection for the early stages of process understanding has been developed, as shown in Figure 69. Execution of the workflow produces a comprehensive dataset of thermodynamic and kinetic parameters for cooling crystallisation via isothermal experiments. The workflow consists of:

- Aim setting.
- Collating of prior knowledge.
- *Decision a determines whether enough experimental raw material data is already known.*
- Setting target parameters.
- A solubility study.
- Choosing a solvent system.
- *Decision B is whether a Van't Hoff solubility line can be plotted for the solvents and solvent mixtures trialled, based upon predictions, with an $r^2$ accuracy exceeding 0.90.*
- A kinetic parameter study is then completed varying the supersaturation and temperature looking for the effects on nucleation rate, growth rate and induction time.
- *Decision C allows for the user to change the solvent system if any unexpected fouling occurred.*
- *Decision D forms part of an optimisation loop until the target parameters set earlier are achieved.*
- *Decision E, the main extension from previous reported work, allows for a call-out to seeded, antisolvent, validation and a crystallisation process design.*
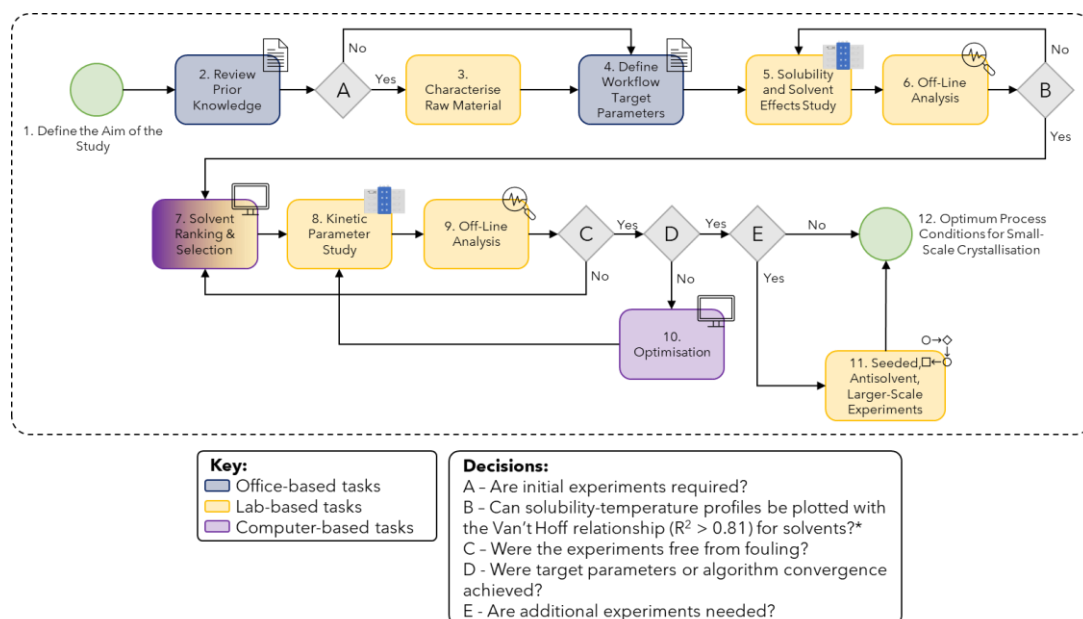
*Figure 69. Schematic diagram of the workflow for crystallisation data collection, adapted from Chapter 4.*

*Minimum of 3-4 data points must be used for reliable estimates of $R^2$ values. For some solvent systems, this may not be achievable as qualitative solubility can still be used to eliminate potential solvents.

Advancements in data-driven algorithms such as solubility predictive models, enhancements in machine learning image analysis and availability of scale-up predictive models for unit operations such as filtration and drying[205] have been deployed to enhance the workflow. With the extension into seeded and antisolvent crystallisation and validation from 2 mL to 50-100 mL scale, a new crystallisation of PD-299685 was designed and optimised.

With the major advancements in small-scale crystallisation hardware, the overarching aim of this chapter was to successfully develop a batch crystallisation process that supersedes past methods, using much less material and time.

## 8.2. Materials, Equipment and Methods

### 8.2.1. Materials

PD-299685 (PRD:114276/102), (3S,5R)-3-(aminomethyl)-5-methyl-octanoic acid, was supplied by Pfizer R&D UK.
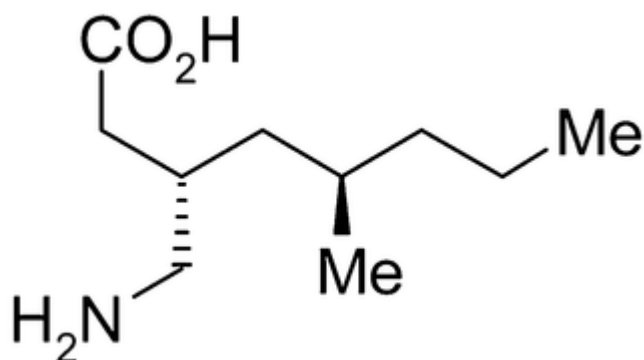
*Figure 70. Molecular structure of (3S,5R)-3-(aminomethyl)-5-methyl-octanoic acid.*

The pure solvents were chosen to ensure functional diversity while following solvents guidelines by the International Conference on Harmonization (ICH)[52] i.e., avoiding class 1.

## 8.2.2. Equipment

Small-scale crystallisation experiments were performed in the Crystalline (Technobis)[125] where in-built process analytical technology (PAT) consists of light transmissivity and particle viewing via high-definition image collection. Images were processed either in proprietary software or by a convolutional neural network (CNN) image analysis algorithm[191]. Validation crystallisation experiments were performed in the EasyMax (Mettler Toledo)[126] fitted with a Blaze probe (BlazeMetrics)[127], providing both microscopy and turbidity capabilities.

X-ray powder diffraction (XRPD) patterns were collected using a D8 Endeavor (Bruker)[233] and the data was visualised in DIFFRAC.EVA software[192]. Off-line dry powder size and shape measurements were collected using Morphologi G3 (Malvern Panalytical)[234]. Differential scanning calorimetry (DSC) was performed on the DSC3+ (Mettler Toledo)[235]. Microscopy images were collected using an Eclipse Ci POL (Nikon)[236] equipped with a Micropix camera[237].

## 8.2.3. Methodology
### 8.2.3.1. Workflow

#### 8.2.3.1.1. Solubility and Solvent Effects Study (Polythermal Global Search)

In this section, the solubility and effect of solvents of PD-299685 were investigated by performing a series of experiments using a Crystalline (Technobis) platform.

Known masses of PD-299685 were weighed into 8 mL screw-top vials where known solvent volumes were added; each vial contained a 10 mm PTFE magnetic stirrer bar. The vials were then subjected to a temperature program consisting of three repeated cycles, with each cycle is as follows:

1. Heat at a rate of 0.5 °C/min to 10 °C below the solvent boiling point or 90 °C (whichever is lowest) and hold at this temperature for 10 minutes.
2. Cool at a rate of 0.5 °C/min to 5 °C and hold at this temperature for 10 minutes.
3. The above temperature cycle was repeated two more times.

The stir rate was fixed at 900 RPM for all the above steps.

Image analysis was carried out to extract the point of dissolution and the primary nucleation threshold for each vial, and subsequently, the metastable zone width (MSZW) was calculated. Solid state analysis, via XRPD, was performed on samples where deviations were observed from the expected crystal morphology. The data obtained from the experiments were analysed to understand the effect that different solvent systems have both on the crystal and the process.

Initial concentration values and additional solvent systems to trial were predicted using COSMO-RS[238] and a group contribution UNIFAC model, implemented in Dynochem (Scale-up systems)[239]. The model has the underlying theory that molecules of solute and solvent are broken down into their constituent functional groups and interactions can then be parameterised. Predicted and measured solubility values were done in an iterative process. Measured data on single solvent systems allowed for refinement of the model for single solvent systems and also predictions in binary and ternary solvent systems.

### 8.2.3.1.2. Kinetic Parameter Study (Isothermal Local Search)

The impact of varying process conditions on the extracted kinetic parameters of PD-299685 at isothermal temperatures from Crystalline (Technobis) data was investigated. Known masses of PD-299685 were weighed into 8 mL screw-top vials where 7 mL of 55:45 (v/v) water/ 1-propanol was then added, each vial contained a stirrer. The vials were then subjected to a temperature program consisting of five repeated cycles, each cycle is as follows:

1. The vials were heated at a rate of 1 °C/min to 60 °C and held at this temperature for 10 minutes.
2. The vials were rapidly cooled at a rate of 10 °C/min to the isothermal temperature of interest, with no stirring.
3. The vials were held at this temperature for 3 hours.
4. The above temperature cycle was repeated four more times.

The agitator geometry was varied across this study to assess magnetic flea, hook stirrer, 3-blade pitched impeller and double 3-blade pitched impeller as different stirring methods. The stir rate was fixed at 900 RPM.

Image analysis of the Crystalline image data was done using a CNN algorithm[191] to extract the induction time, the growth rate, the aspect ratio and an arbitrary secondary nucleation rate from the isothermal hold. Samples were filtered, washed with acetone, and dried using a standard vacuum filtration setup. The data obtained from the experiments were analysed to understand a good spread of the design space between process conditions and crystallisation kinetic parameters.

### 8.2.3.1.3. Additional Experiments – Seeded, Antisolvent, Easy-Max Validation

The impact of other modes of crystallisation and the implications of scaling up were investigated. The same methodology and analysis as detailed in Section 8.2.3.1.2 were used for the kinetic parameter estimations of seeded and antisolvent crystallisations. The only minor discrepancy was that for seeded crystallisation experiments, a known mass of PD-299685 was added during the isothermal holds.

For the validation experiments, an EasyMax (Mettler Toledo) was equipped with a 50 mL glass one-piece vessel. A known mass of PD-299685 was added to the reactor vessel and then 45 mL of 55:45 (v/v) water/ 1-propanol was added, with agitation using an overhead PTFE half-moon stirrer. The experiment used the same temperature program as detailed in Section 8.2.3.1.2.

### 8.2.3.2. Process Design

In this section, both the Crystalline (Technobis) and EasyMax (Mettler Toledo) were used with the following base process profile:

1. Heated to 75 °C at a rate of 1 °C/min and held for 30 minutes.
2. Cooled to 60 °C at a rate of 0.2 °C/min.

3. Seed material was added and then held for 30 minutes.

4. Antisolvent was added over 45 minutes and held for 30 minutes.

5. Cooled to 5 °C at a rate of 0.2 °C/min.

6. Additional specific method section as detailed in Section 8.2.3.2.1 and 8.2.3.2.2.

This base profile was determined as a product of the workflow based upon a covariance analysis of the kinetic parameters which allowed prediction of process parameters to get the desired product and performance. The predicted process, that complies with operational contracts and objectives set, was developed in Crystalline and then validated in the EasyMax equipped with a Blaze probe.

### 8.2.3.2.1. Crystalline

For the optimized process run at screening scale, 350.9 mg of PD-299685 was added to an 8 mL vial and 4.67 mL of 55:45 water/ 1-propanol was added. An overhead 3-blade pitched impeller cap was used to seal the vial. The vial was transferred to the Crystalline (Technobis) and underwent the above process profile where 7.1 mg of seed and 2.33 mL of water (antisolvent) were added. After the final cool, the vial was isothermally held for 10 hours. The stir rate was fixed at 900 rpm throughout the whole process. The API was filtered using a Whatman Autovial 12 (Cytiva) filter and washed with 2x2 mL of cold acetone and dried in a vacuum oven at 40 °C.

### 8.2.3.2.2. EasyMax

Referring to Table 23 for specific details, PD-299685 was added to a two-piece 100 mL vessel and 66.67 mL of 55:45 water/ 1-propanol was added. A pitched 4-blade metal stirrer was used at an agitation rate of 300 RPM throughout, except for when high-shear wet milling (HSWM) was in operation. The reaction mixture underwent the base process profile detailed above with 33.33 mL of water (antisolvent) added. After the final cool, the process underwent specific methodology per iteration. The slurry was wet milled. The API was filtered through a 21 mm diameter Buchner funnel and washed with 2x10 mL cold acetone and dried in a vacuum oven at 40 °C.

*Table 23. Experimental details for End Process Design performed in the EasyMax for 4 iterations.*

| Mass of API (g) | Mass of seed (mg) | Additional specific method details | HSWM details |
|---|---|---|---|
| 4.992 | 99.67 | Isothermally held at 5 °C for 10 hours after the final cool | IKA T 25 Ultra Turrax equipped with a S25KD-25F dispersing tool for 2 minutes at 5000 rpm |
| 5.013 | 101.8 | Milled after the final cool and the slurry underwent a thermocycling of heating to 20 °C at 1 °C/min, held for 30 minutes, cooled to 5 °C at 0.2 °C/min three times. Isothermally held for 8 hours. | IKA T25 Ultra Turrax equipped with a S25KD-25F dispersing tool for 2 minutes at 5000 rpm |
| 5.055 | 102.8 | Isothermally held at 5 °C for 14 hours after final cool, milled, and isothermally held for 1 hour. | IKA T25 Ultra Turrax equipped with a S25KD-25F dispersing tool for 2 minutes at 3000 rpm |
| 5.061 | 99.7 | Isothermally held at 5 °C for 12 hours after final cool, milled, and isothermally held for 1 hour. | IKA MagicLab equipped with fine teeth for a single pass at 14,600 rpm |

## 8.3. Results & Discussion

### 8.3.1. Workflow

*The sections described below align to boxes within the workflow (Figure 69) where the section number of 8.3.1.X. with X referring to the specific numbered box.*

#### 8.3.1.1. Define the Aim of the Study

The overall aim of the study was to collect thermodynamic and kinetic parameter data on PD-299685 at small-scale to be able to select the appropriate crystallisation method and then subsequently design a viable process that gives adequate product quality, is safe, economical and environmentally sustainable in a reasonable cycle time.

The aim of the solubility and solvent effects study was to collect quantitative and qualitative data on the thermodynamic crystallisation behaviour and to evaluate the UNIFAC solubility prediction[240] regression model. The aim of the kinetic parameter

study focused on collecting quantitative kinetic data using a range of agitation methods, volumes, and crystallisation techniques. Additional experimentation was aimed at expanding and validating the workflow by using various stirrer types (to evaluate mixing sensitivity) and particle size and shape analysis methods (to simplify data analysis). The investigation of the seeded, antisolvent, and validation experiments aimed to explore the design space and enhance the capabilities of the workflow. This research built upon prior work (Chapter 4 and 5) developing efficient process development workflows, offering an extension and enhancement of its application.

The collected data, both quantitative and qualitative, were utilised to inform a novel crystallisation process for the API. The process was optimised to align with drug product attributes, highlighting the industrial advantages of the workflow's approach. The study aimed to show that the use of workflows allows for efficient and effective process development direction for any given API.

### 8.3.1.2. Review Prior Knowledge & Literature Search

The physical properties of PD-299685 have been previously assessed by optical microscopy, DSC, thermogravimetric analysis (TGA), XRPD and vapour sorption/ desorption analysis. PD-299685 form A (the most stable polymorph) was classified as non-hygroscopic and highly crystalline. Experimental data showed a single melting endotherm at 177 °C, <0.5 % wt % moisture loss and that the crystal exhibited needle geometry. Additional polymorphs were identified, including form B (from rapid evaporation of acetone), form C (methanol solvate) and form D (monohydrate). Gravimetric solubility data was available for PD-299685 in pure and binary solvent systems, previously measured during screening. PD-299685 had low solubility (3.4 mg/mL at room temperature) in water at neutral pH, showed synergetic effects in organic/aqueous and organic/organic solvent mixtures, and had a high solubility in acidic and alkaline media.[241] The manufacturing process developed previously, using more traditional and empirical methods, involved an 11-step linear synthesis, which was followed by a cooling recrystallization in a mixture of 50:50 vol/vol EtOH: water to get a yield of 87%.[232] The resulting drug product was made into immediate-release tablets with doses of 5-20 mg using direct compression. However, a drawback of the previous process was the final

broad particle size distribution (PSD), with a d90 value exceeding 800 μm. This could potentially impact drug product content uniformity, particularly at lower doses. HSWM was therefore required to reduce the PSD to target size for tabletting.

### 8.3.1.A. Decision A: Are initial experiments required?

Yes, analysis of raw materials was required (see Section 8.3.1.3. below for details).

### 8.3.1.3. Characterise Raw Material

The starting material for the recrystallisation was form A, as shown by a fingerprinting comparison of the XRPD patterns to previous studies. The classification was further confirmed based on the onset of a melting endotherm observed at 177 °C in the DSC data.

### 8.3.1.4. Define Workflow-Specific Target Parameters

For the solubility and solvent effects study of PD-299685, the objective target parameters were to achieve a solubility in the range of 50-250 g/L solvent[145], which translates to a solvent-to-API ratio of 4-20 L/kg. The solvent system with the steepest solubility vs temperature profile was the most preferred. The desired physical properties of PD-299685 crystals included an aspect ratio greater than 0.5[146] and a size (d90) of approximately 100 - 350 μm. Form A was preferred due to its stability and lack of solvate character.

For the kinetic parameter study, the primary objective target was to collect data over the design space to determine induction time, growth rate, and nucleation rate under different process conditions, such as supersaturation and temperature.

The target parameters for the end crystallisation process design and optimization were yield, particle shape, and particle size. The study aimed to maximize yield while maintaining an aspect ratio greater than 0.5[146] and a particle size (d90) of 100 – 350 μm.

### 8.3.1.5. Solubility and Solvent Effects Study (Polythermal Global Search)

### 8.3.1.5.1. Iteration One

The first iteration of the solubility and solvent effects study consisted of collecting data for PD-299685 in pure solvents. As shown in Figure 71 (a) the solubility of PD-

299685 was very low in all organic solvents and moderate in water. To improve the coverage of the UNIFAC model and to provide data on synergetic effects in solvent mixtures, the solubility of PD-299685 was explored in 50:50 (v/v) binary mixtures of water and organic solvents. The range of solvent mixtures was limited due to the poor miscibility of some organic solvents with water. The increased solubility caused by synergetic effects can be seen in Figure 71 (b). Observed crystal habit was block-like for most solvent systems with a small amount of fines generation or aggregation, which was likely caused by the magnetic stir bar grinding particles.
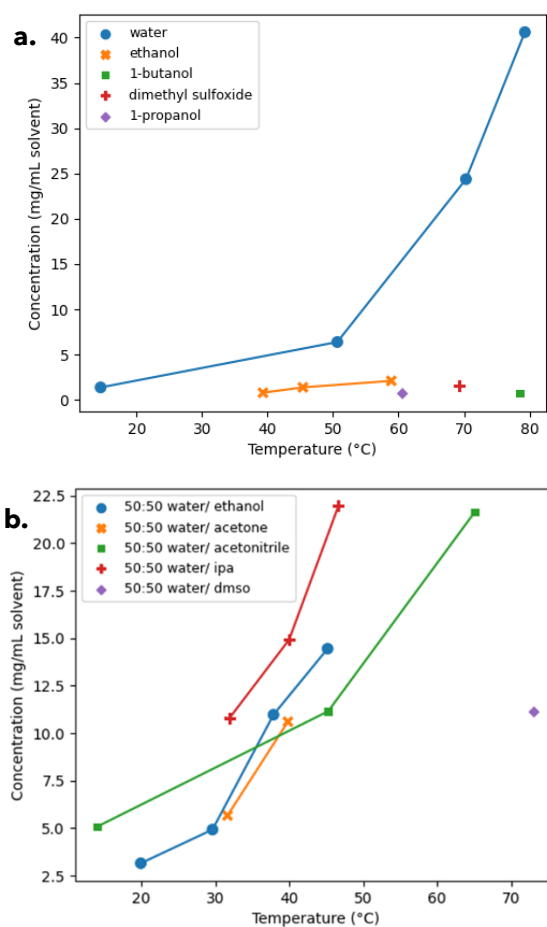


*Figure 71. Approximate solubility-temperature profile of experimental results of PD-299685 in various pure solvents (a) and 50:50 binary solvent mixtures (b). NOTE: join-the-dot lines were used as the solubility-temperature profiles were estimates based on limited data points.*

The data from this iteration was used to fit a Dynochem UNIFAC model, and binary solubility prediction was run at 5 °C, 25 °C, 50 °C and 70 °C for the covered solvents. Highly synergetic effects were observed for water/alcohol and water/acetonitrile mixtures as shown in Figure 80 (in the Appendix of Chapter 8). The higher predicted

solubility of PD-299685 at high temperatures, compared to pure and 50:50 (v/v) solvents, allows for more economic solvent and vessel usage, whereas minimal change at low temperatures enhances a predicted cooling crystallisation yield to more than 98%.

The best binary solvent compositions, due to highest synergistic solubility, for further study from this iteration were 40:60 (w/w) ethanol/ water, 40:60 (w/w) IPA/ water and 40:60 (w/w) 1-propanol/ water which equates to volume fractions of 45:55. Acetonitrile/ water has good synergy (where the solubility exceeds that of the single solvents) at 30:70 (w/w), equating to 33:66 (v/v), and was also further explored. Acetone/ water mixtures were eliminated from the study due to low solubility at the upper operating temperature of acetone due to its low boiling point. Ethanol/ glycerol had the highest predicted synergy for an organic/ organic solvent mixture but with still very low solubility so was also eliminated from the study.

### 8.3.1.5.2. Iteration Two

The recommended solvent systems from the above iteration were trialled experimentally and the data observed was fed back into a UNIFAC model.
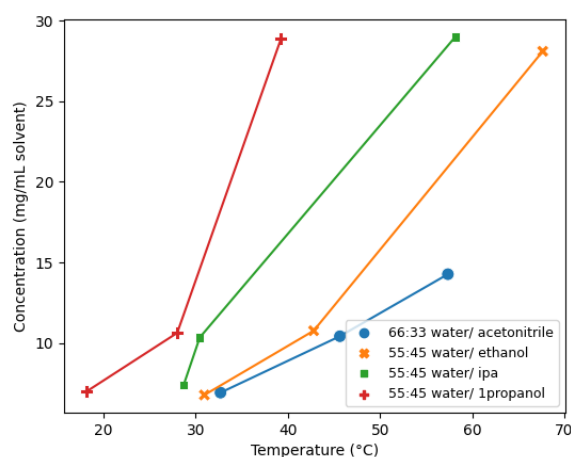


*Figure 72. Approximate solubility-temperature profile of the experimental results of PD-299685 in the UNIFAC recommended solvent systems. NOTE: join-the-dot lines were used as the solubility-temperature profiles were estimates based on limited data points.*

The UNIFAC regression model fitted with the new experimental data returned recommended solvent systems the same as the previous iteration, therefore termination criteria for the solubility and solvent effects study had been met.

Section 8.3.1.7. discusses the classification, ranking and selection of a solvent system.

### 8.3.1.5.3. Iteration Three

As an extension to the general workflow, ternary solvents were also explored. As 55:45 (v/v) water/1-propanol gave the steepest solubility curve that was chosen as the base binary mixture. Various UNIFAC model predictions were done with the water/1-propanol solvent mixture plus an additional miscible solvent. The general trends showed that the addition of a third solvent had a limited impact on the solubility except with acetonitrile where the addition of large volumes of solvent slightly reduces the overall solubility of PD-299685, at low temperatures, in the ternary solvent system. These predictions for PD-299685 in water/1-propanol and acetonitrile were validated in the laboratory and it was confirmed the addition of acetonitrile reduced the solubility but by a minimal amount. The importance of this in designing an antisolvent crystallisation is discussed in Section 8.3.1.7.

### 8.3.1.6. Off-Line Analysis
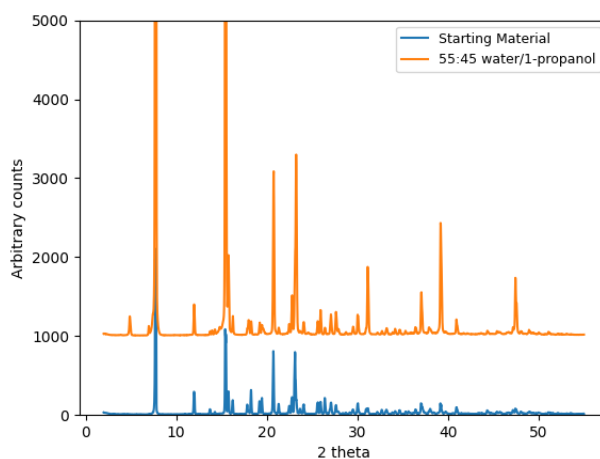


*Figure 73. X-ray powder diffraction patterns for PD-299685 recrystallised from 55:45 (v/v) water/ 1-propanol overlayed against reference Form A/ starting material.*

The XRPD pattern (Figure 73) shows that the recrystallised material using the chosen solvent system gives the most stable polymorph, Form A. This is confirmation to show that increasing the water content in the solvent system, compared to previous crystallisation methods[232], does not favour the formation of the monohydrate polymorph form.

**8.3.1.B Decision B: Can solubility-temperature profiles be plotted with the Van't Hoff relationship (R2 > 0.81) for solvents?**

Measured solubility profiles for the predicted best solvent systems for single and binary solvent systems, based on target parameters, were achieved with above 0.81 $R^2$ values for the line of fit. Progression onto the next stages of the workflow was permitted.

**8.3.1.7. Solvent Ranking & Selection**

From the solubility and solvent effects study (Section 8.3.1.5), the following classifications can be made for PD-299685:

- 1-butanol, 1-propanol, 2-methyl THF, acetone, acetonitrile, ethyl acetate, heptane, isopropanol and isopropyl acetate are antisolvents.
- Ethanol and DMSO are practically unusable pure solvent systems due to low solubility.
- Water/ organic mixtures gave good synergistic solubility.
- Acetone would be an ideal wash solvent for filtration and drying due to low API solubility and low boiling point.

**8.3.1.7.1. Cooling Crystallisation Solvent:**

Solubility within the target parameters was observed for water/ organic binary solvent systems due to synergy meaning that these solvent systems can be classified as good cooling crystallisation solvents. The solubility profile (Figure 72) for 55:45 (v/v) water/1-propanol had the steepest gradient and therefore highest predicted yield from cooling crystallisation. Furthermore, water/1-propanol had the highest operating temperature out of the potential solvent systems meaning that higher temperatures and thus higher throughput can be achieved. No considerable operational and handling issues were observed at low temperature. No fouling was observed through image analysis and direct observation therefore this solvent mixture was chosen as the cooling crystallisation solvent.

*Figure 74. Solubility-temperature profile for PD-299685 in 55:45 water/ 1-propanol with the data fitted with the Van't Hoff relationship.*

Additional solubility measurements were carried out to allow for a better fit to the data over a range of temperatures. A thermal stability trial was also conducted where a highly concentrated sample was held for 24 hours at 75, 80, 82.5 and 85 °C and aliquots were sampled using HPLC. Negligible chemical degradation was observed for all temperatures after 2 hours, however a new compound with molecular weight of 170.2 g/mol was observed after 5 hours for vials held at 80 °C and above. Therefore, a maximum temperature of 75 °C will be used for PD-299685 in 55:45 (v/v) water/1-propanol to allow an extended hold for dissolution if scaled up to plant.

### 8.3.1.7.2. Antisolvent Crystallisation Solvent:

Three possible antisolvent processes were explored in Section 8.3.1.5., which were the addition of water, the addition of 1-propanol and the addition of acetonitrile. As seen in Figure 80 (of the Appendix of Chapter 8) the predicted solubility was lower the closer the solvent mixture was to pure solvent. The addition of 1-propanol to PD-299685 in 55:45 (v/v) water/ 1-propanol (to change the 1-propanol content to 63%) reduced the solubility as did the addition of water (to change the water content to 70%). The addition of water reduced the solubility of PD-299685 more as the final solvent composition is closer to the pure solvent and therefore had the least synergy of the two solvent system options. As presented in Iteration Three, of

155

Section 8.3.1.5. the addition of acetonitrile reduced the solubility but to minimal effect compared to the addition of water. It is also worth noting that ternary solvents are not preferred due to increased complexity of solvent recovery and increased specification burden. Therefore 70:30 (v/v) water/1-propanol was the chosen final composition for antisolvent crystallization for the kinetic parameter study.



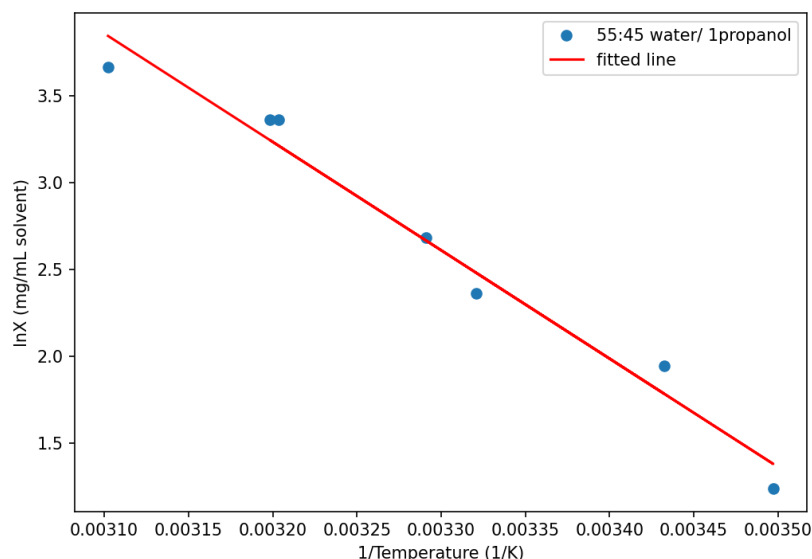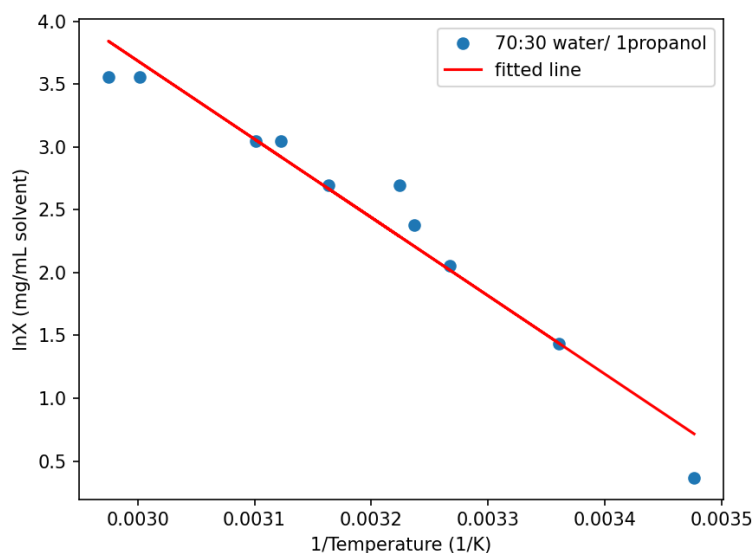*Figure 75. Solubility-temperature profile for PD-299685 in 70:30 water/ 1-propanol with the data with the Van't Hoff relationship. Note that the triplicate data and fitted line check for uncertainty removed 2 data points from the dataset for this figure. Also not all data points were able to be collected in triplicate due to poor nucleation at low concentrations.*

An additional series of measurements was conducted for PD-299685 in 64:36 (v/v) water/ 1-propanol to fit solubility for the halfway point of antisolvent addition. No fouling was observed through image analysis and observational analysis.

The XRPD pattern (Figure 81) confirms that the increase in water content in the solvent composition did not cause the formation of the monohydrate polymorph. The most stable form, A, was produced. In addition, this solvent composition was stress-tested by holding a slurry of PD-299685 in 70:30 water/ 1-propanol at 5 °C for more than 5 days where no solid form transformation in the product was observed.

### 8.3.1.8. Kinetic Parameter Study (Isothermal Local Search) & 8.3.1.9. Off-Line Analysis

A design of experiment plan was created with a 2-dimensional design space of varied isothermal temperature and supersaturation. The plan was based on a correlation Latin hypercube design[242], a statistical method for designing a near-random experimental plan, (Table 25 in the Appendix of Chapter 8) adapted so that the experiments sat on a Pareto front of feasible nucleation within the MSZW.

Primary nucleation was not observed for any of the vials ran at a temperature of 16 °C and supersaturation of 3.2 likely due to the low concentration of the vial. This qualitatively showed that high concentration or seeding are required at low temperatures. A comparison of different stirrers on the effect of crystallisation was conducted on vial 7 of each experiment (temperature of 28 °C, supersaturation of 2.91) where the magnetic stirrer showed attrition, the hook-shaped stirrer showed reduced mixing and the 3-blade pitched stirrer showed good mixing. The differences in quality of crystal generation were negligible between the single and the double 3-blade pitched impellers therefore the single was chosen for further study in Section 8.3.2. as it is more commonly used in plant-scale reactors.
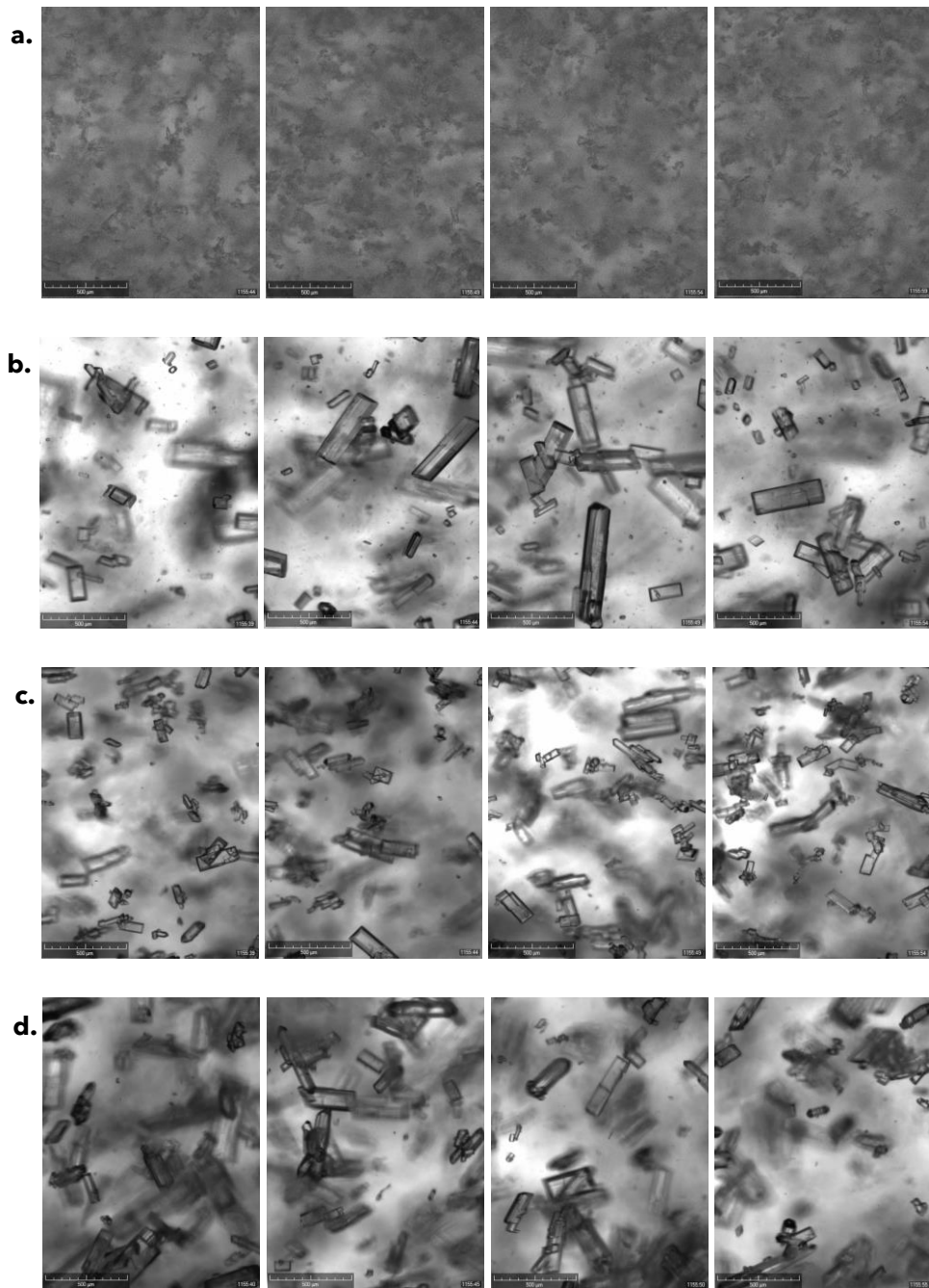
*Figure 76. Crystalline (Technobis) images from the end of the final hold cycle for vial 7 agitated by magnetic flea (a), 3-blade pitched (b), 3-blade double pitched (c) and hook (d) stirrers.*

The qualitative observations between the different stirring methods were validated using microscopy and Morphologi data which showed that the particles crystallised in the vial with a magnetic flea had a d(v,90) approximately 3.5 times smaller compared to overhead stirring methods.
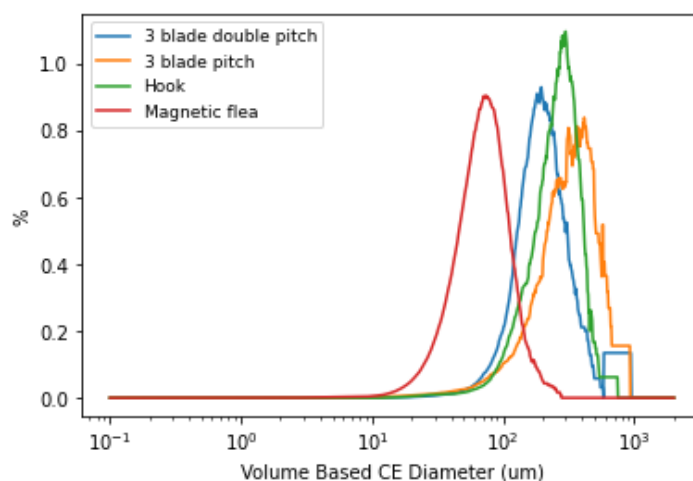
*Figure 77. Volume-based particle size distribution (PSD) from Morphologi for all Crystalline stirring methods. NOTE; the difference in smoothing is due to varying number of total particles.*

The kinetic parameters extracted, using CNN image analysis, were visualised using a covariance matrix due to the multiple dimensionalities of the design and measured variable space.

The agitation methods were assigned a discrete eigenvalue to allow for numerical comparisons between continuous and categorical variables across the design space. It can be seen from the co-variance matrix (Figure 82 (a)) that changing between the magnetic flea and the 3-blade pitched impellor, had a large impact on the measured kinetic parameters. It can qualitatively be concluded that the use of the magnetic flea in this study ground the particles and created clouds of fines which created more particle surface area for faster desaturation.

The data presented in Figure 82 (b) (in the Appendix of Chapter 8) indicates that there was no statistically significant difference in terms of impact for type of overhead agitation (hook, single 3-blade pitch, double 3-blade pitch) on nucleation, although - as mentioned in an earlier section - the hook impeller gave rise to a poor mixing environment.

PD-299685 can be classified as a slow nucleating and fast growing API across varied process conditions and when using an overhead stirrer, the most representative of a plant scale process, yields large crystals. The d90 of crystals from the isothermal study were larger than 500 μm which exceeded the required particle size required for API formulation in this instance. The general crystal habit was blocks, which were

acceptable for downstream processing. Other morphologies, such as needles, are less desirable from a process perspective given that particles exhibiting this shape generally have poor bulk properties, including for instance flow,

The information gathered in this stage was foundational for designing an end crystallisation process mainly in terms of the impact of agitation and mixing and knowing the impact of temperature and supersaturation on nucleation and growth.

### 8.3.1.C Decision C: Were the experiments free from fouling?

No significant fouling occurred, from visual observation, of the API in the chosen solvent system, progression onto the next stage of the workflow was permitted. This would be particularly important when considering a continuous process.

### 8.3.1.D. Decision D: Were target parameters or algorithm convergence achieved?

The target parameters for the solubility study were satisfied in respect to choosing a solvent system with appropriate solubility giving crystal habit with a desired aspect ratio. The target parameter for size was not satisfied however, and mechanical interventions, for example HSWM, would be required. The target parameters for the kinetic study were all satisfied. Progression onto the next stage of the workflow was permitted with being mindful that crystal size control would require milling, similar to the previous recrystallisation process[232].

### 8.3.1.10. Optimisation

As target parameters were met (as outlined in Section 8.3.1.D) the optimisation step was not required.

### 8.3.1.E. Decision E: Are additional experiments needed?

As the aims of the study were to expand the workflow, into seeding, scale-up validation and process design, then further experimentation was required. This decision point formed a loop.

### 8.3.1.11. Additional Experiments: Seeded, Antisolvent, Easy-Max Validation

### 8.3.1.11.1. Seeded

Seeded experiments were performed following a space-filling Latin hypercube DoE plan, as explained in Section 8.3.1.8. (Table 26 in the Appendix of Chapter 8) at lower supersaturations than the previous section. The 3-pitched blade stirrer was used in line with findings reported in Section 8.3.1.8. Due to the difficulty in accurately dosing equal small quantities of seed into the Crystalline during the hold period, it was decided to pause the image analysis and add circa. 5-10 mg and the difference in vial weight was recorded, before then restarting the image collection.

The covariance matrix (Figure 83 (a)) shows that an increase in seed mass caused both growth and nucleation rates to increase where the growth rate dominated over the nucleation rate. Nucleation however dominated over growth at higher temperatures and higher supersaturations, as expected. Microscopy of samples taken of vial 3 (supersaturation of 1.60, temperature of 35 °C) showed many small crystals with some large agglomerates compared to vial 7 (supersaturation of 1.55, temperature of 28 °C) which showed dispersed square crystals over 500 μm.

### 8.3.1.11.2. Antisolvent

The same experimental plan as previously used for the local kinetic search (Section 8.3.1.8.) was performed here but using only the overhead 3-blade pitched stirrer. The data from that section was used alongside data gathered for the addition of half of the antisolvent and the addition of all the antisolvent. I.e., an isothermal study was conducted in 55:45 (v/v), 64:36 (v/v) and 70:30 (v/v) water/ 1-propanol.

It can be seen from the covariance matrix (Figure 83 (b)) that the addition of antisolvent had a positive impact on nucleation and a negative impact on growth. This was ideal for limiting the d90 size of the final crystals. There was also a largely positive relationship between the addition of antisolvent at higher temperatures on nucleation with minimal impact on growth. Therefore, nucleation would largely dominate over growth at high temperatures and higher antisolvent composition.

### 8.3.1.11.3. Easy-Max Validation

The same experimental plan that was conducted in the local kinetic search (Section 8.3.1.8) was performed using a 50 mL vessel in the EasyMax with an overhead half-moon PTFE stirrer. The crystallisation kinetic parameter data for the 3-blade pitched impellor in the Crystalline and the data from the EasyMax were compared by covariance matrix as shown in Figure 83 as these two stirrers have the most similar geometry and tip speed.
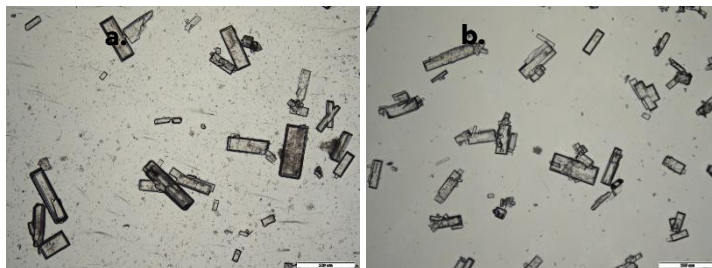


*Figure 78. Microscopy images taken at 4x magnification of PD-299685 recrystallised from 55:45 (v/v) water/1-propanol using the Crystalline with overhead 3-blade pitched impeller (a) and EasyMax with half-moon PTFE stirrer (b).*

It can be seen from the covariance matrix (Figure 83 (c)) that there are major variances between the Crystalline and EasyMax nucleation and growth data which can be accounted for by differences in PAT and reactor geometry. The Blaze probe used for the EasyMax image collection had a much smaller field of view (FOV) compared to the Crystalline cameras, and therefore it was hard to track nucleation and growth. Despite the large variances, microscopy images (Figure 78) showed good qualitative similarities in final crystal size between EasyMax and Crystalline when using the 3-blade pitched double impellor.
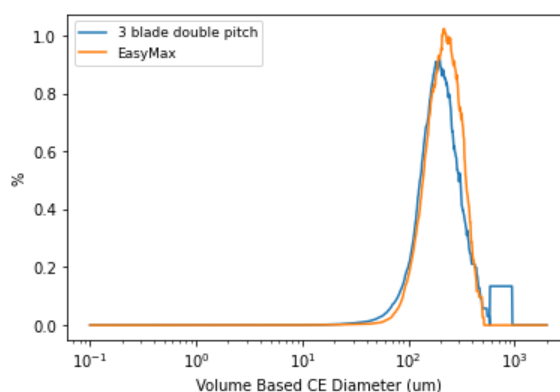
*Figure 79. Volume-based particle size distribution (PSD) from Morphologi for EasyMax overhead PTFE stirrer and Crystalline overhead 3-blade double pitched for SS=2.91 and Temperature=28°C.*

The PSDs as shown in Figure 79 also showed good similarity, where the d(v,90) was 384 µm and 328 µm for the Crystalline 3-blade double pitched and EasyMax PTFE stirrer respectively. Therefore, it can be concluded that for the best estimations of scaling up from Crystalline to EasyMax, the 3-blade double pitched should be used at the millilitre scale. However alternative PAT should be used for online measurements of kinetic parameters for a fast growing compound as the one used here such as imaging with a wider field of view (FOV). However, this is compound specific and the set-up discussed in this paper would most times suffice.

### 8.3.1.E. Decision E: Are additional experiments needed? (revisited)

After completion of the above experiments required to expand the original workflow study the decision criterion was met and progression onto the next stage of the workflow was permitted.

### 8.3.2. Process Design and Optimisation

After completion of the workflow there was a wealth of data collected, both quantitative and qualitative, on the thermodynamics and kinetics of the system. This allowed the design and optimisation of a commercially viable crystallisation process.

As shown in Section 8.3.1.5., the solubility and solvent effects study, the solvent that gave the highest solubility and temperature dependence was 55:45 water/ 1-propanol with a predicted yield of over 90% for a cooling crystallisation only. The addition of water for a final solvent composition of ~70:30 reduced the solubility at

low temperatures further pushing the predicted yield to 95%. Therefore, a cooling and antisolvent hybrid crystallisation was proposed as the ideal mode of crystallisation. The local kinetic parameter study (Section 8.3.1.8.) showed that the use of an overhead blade impeller would yield crystals to align with shape and size target parameters set out and would be most like the actual manufacturing vessel. Analysis of covariance between isothermal experiments in Crystalline and EasyMax showed good comparisons between the different scales. This enabled process understanding of kinetic parameters at millilitre scale before process optimisation. Due to past issues of large particle size, the proposed crystallisation route should allow for nucleation to dominate over growth. Analysis of covariance in the seeded and antisolvent experiments in Section 8.3.1.11. showed that this could be achieved where seeding was done at higher temperatures and using lower seed mass. Furthermore, a fast antisolvent addition at higher temperatures should also allow for more nucleation. A cooling rate of 0.2 °C/min was determined due to heat transfer constraints associated with plant-sized equipment.

### 8.3.2.1. Crystalline

The API was recrystallised with a yield of 99%, d90 of 759 μm and a mean aspect ratio of 0.718 under conditions described in Section 8.2.3.2.1. Control over particle sizing was not achieved by purely crystallisation process control. Yield was abnormally high likely due to the excellent purity of the starting material and also the narrow geometry of the Crystalline vials allowing for fast depletion of supersaturation. This initial experiment however showed that the crystallisation process worked well but needs the integration of HSWM.

### 8.3.2.2. EasyMax

*Table 24. Experimental results and comments from the four EasyMax iterations of the final process design.*

| Run | d(V,90) (μm) | Mean aspect ratio | Yield (%) | Comment(s) |
|:---:|:---:|:---:|:---:|:---:|
| **1** | 476 | 0.688 | 91 | HSWM reduced d(v,90) from 664 μm, slow filtration |
| **2** | 600 | 0.743 | 92 | Hold and thermocycling after HSWM allowed for excessive growth, narrow PSD* (less fines), slow filtration |

| | | | | |
|---|---|---|---|---|
| **3** | 523 | 0.725 | 92 | Hold after HSWM allowed recovery of yield but also growth, slow filtration |
| **4** | 234 | 0.766 | 92 | HSWM gave narrow PSD*, fast filtration |

*Refers to Figure 84.

Iteration 4 of the end process design yielded the best process in terms of crystal attributes as met the desired target parameters of d90 of 100 – 350 μm and an aspect ratio greater than 0. 5. The size and shape of the crystals recovered would allow for a fast free flowing bulk powder that is also suitable for tabletting in the correct dosage. Using the IKA MagicLab gave a higher degree of control over exposure to the mill teeth of the HSWM compared to the T-25 due to the material showing poor wettability presumably due to mild having hydrophobic character.

## 8.4. Conclusions

The use of a structured workflow allowed for quick collection of data where a thermodynamic model and kinetic model were developed for PD-299685 in circa 8 weeks. Additionally, validation between stirring methods, scale-up and different modes of crystallisation was done as part of this study. This showed that the Crystalline with an overhead stirrer gave the best estimate of how the crystallisation would occur in the EasyMax. Scaling down to the millilitre scale and using offline analysis such as Morphologi provided the best comparisons. Further work however needs to be done to incorporate comparable online PAT at difference scales. Smart decision-making alternating between experimental data and solubility prediction algorithms allowed for a fast thermodynamic model to be developed and applied to inform solvent selection for an industrial crystallisation process. A space-filling Latin hypercube DoE were employed to plan the least number of experiments to make a kinetic model for cooling, seeded and antisolvent crystallisations. These models were then analysed using covariance matrices to determine the ideal supersaturation and temperature for specific process steps in the crystallisation process. The optimum crystallisation process developed was a cooling, seeded and antisolvent hybrid of PD-299685 in 55:45 (v/v) water: 1-propanol with additional water as antisolvent. Dissolution was at 75 °C, seeding and addition of antisolvent was done at 60 °C, and then the mixture was cooled to 5 °C before milling. The process proposed gives higher control and product quality over previous

processes[232] as had an increased MSZW to allow for speck-free filtration and seeding. Additionally, due to the use of antisolvent addition, the yield has been increased by 5%.

# Appendix

This appendix contains additional information for context and details pertaining to extracted data figures, material characterisation, experimental plans and covariance matrix analysis.
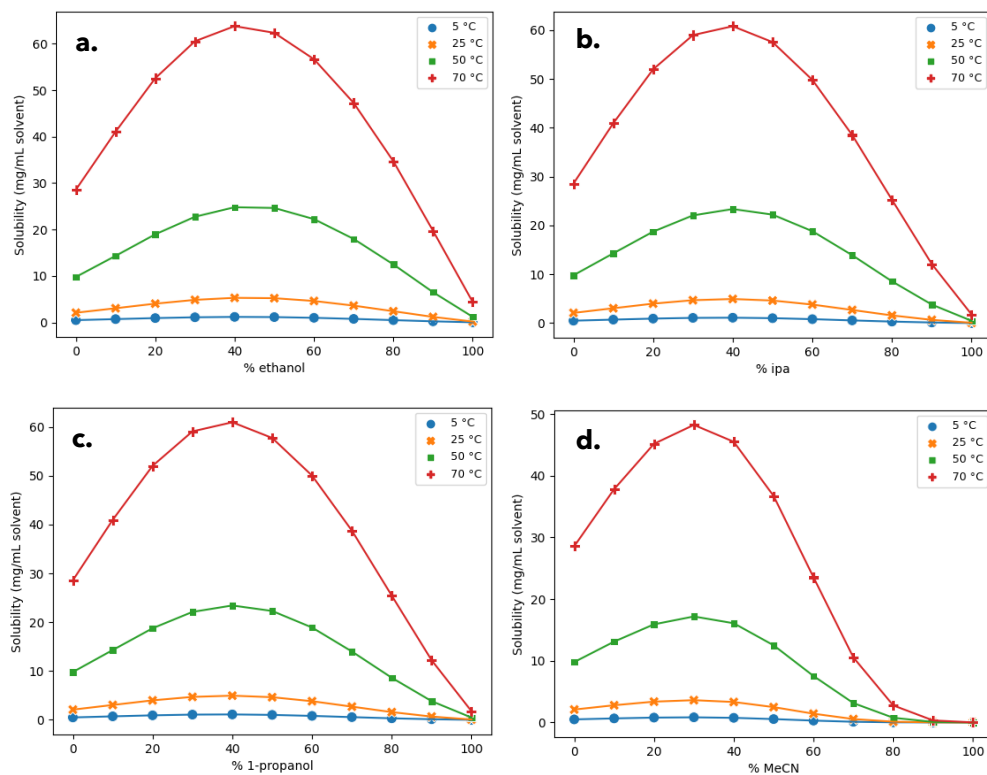


*Figure 80. Predicted solubility profiles of varying mass compositions of binary solvent mixtures at 5, 25, 50 and 70 °C of ethanol/ water (a), IPA/ water (b), 1-propanol/ water (c) and acetonitrile/ water (d).*
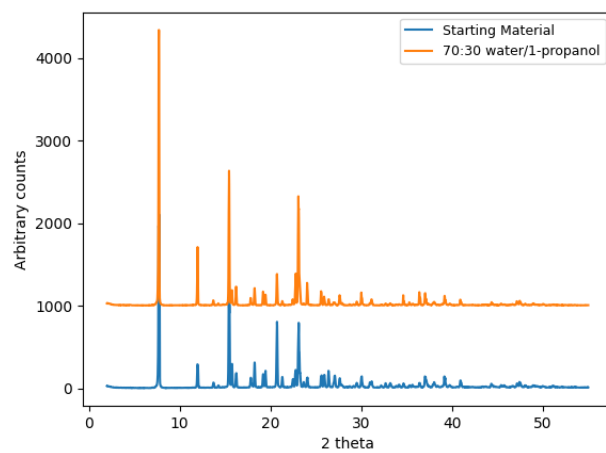
*Figure 81. X-ray powder diffraction patterns for PD-299685 recrystallised from 70:30 (v/v) water/ 1-propanol overlayed against reference Form A/ starting material.*

*Table 25. Design of Experiment (DoE) plans for the kinetic parameter study.*

| Vial | Isothermal temperature (°C) | SS |
|---|---|---|
| 1 | 36 | 1.84 |
| 2 | 16 | 3.2 |
| 3 | 35 | 2.26 |
| 4 | 24 | 3.1 |
| 5 | 21 | 3.2 |
| 6 | 31 | 2.64 |
| 7 | 28 | 2.91 |

Table 26. Design of Experiment (DoE) plans for the seeded kinetic parameter study. Note: supersaturation for each vial increases for the seeded experiments due to the addition of seed from the previous cycle.

| Vial | Isothermal temperature (°C) | SS |
|---|---|---|
| 1 | 36 | 1.2 |
| 2 | 16 | 1.4 |
| 3 | 35 | 1.41 |
| 4 | 24 | 1.08 |
| 5 | 21 | 1.45 |
| 6 | 31 | 1.28 |
| 7 | 28 | 1.33 |
| 1 | 36 | 1.25 |
| 2 | 16 | 1.45 |
| 3 | 35 | 1.51 |
| 4 | 24 | 1.23 |
| 5 | 21 | 1.62 |
| 6 | 31 | 1.35 |
| 7 | 28 | 1.41 |
| 1 | 36 | 1.3 |
| 2 | 16 | 1.63 |
| 3 | 35 | 1.6 |
| 4 | 24 | 1.31 |
| 5 | 21 | 1.64 |
| 6 | 31 | 1.43 |
| 7 | 28 | 1.55 |

**a.**
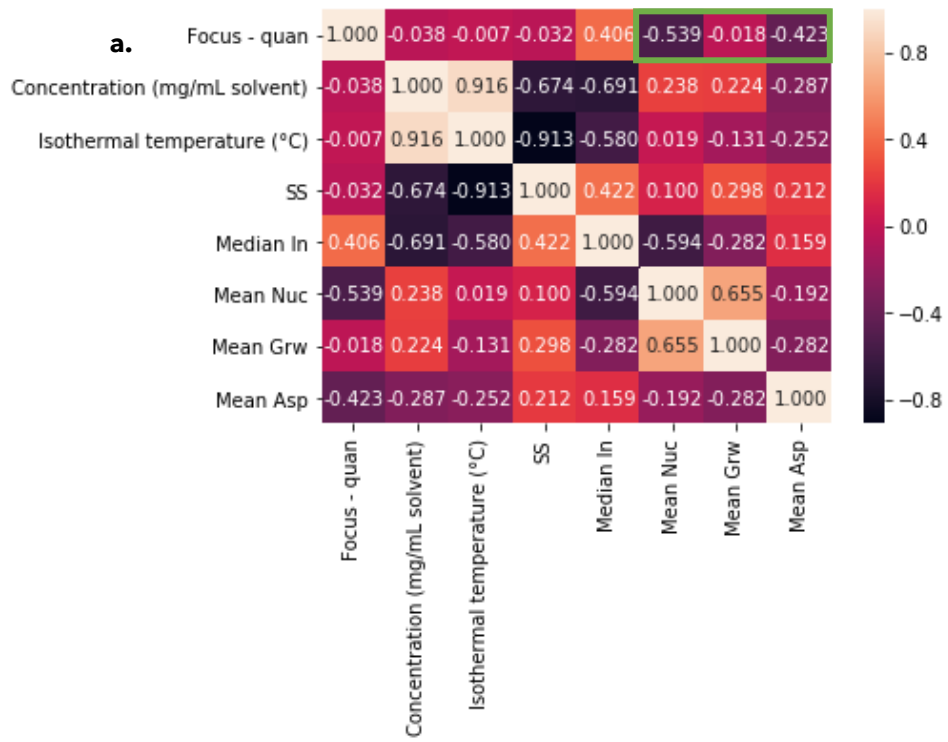
*The figure continues on the next page.*
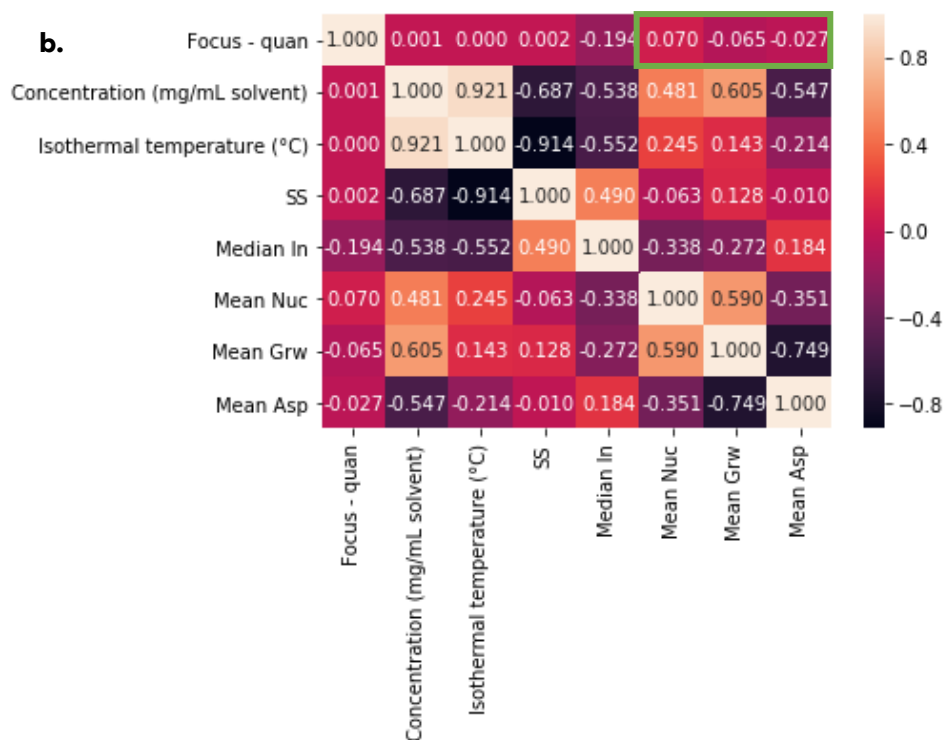
*Figure 82. The covariance matrices of input and output variables\* across the kinetic parameter study for two different agitation methods, magnetic flea (assigned as 0) and the 3-blade pitched impellor (assigned as 1) (a) and for the 3-blade pitched impellor (assigned as 3), 3-blade double pitched impellor (assigned as 6) and the hook stirrer (assigned as 2)\*\* (b). The green box resembles the main point of interest.*

\*Abbreviations in the legend are as follows: Focus – quan refers to a quantitative assignment of the qualitative focus of the analysis (see the comment on eigenvalues below), SS is supersaturation, In is induction time, Nuc is nucleation rate, Grw is growth rate and Asp is aspect ratio.

\*\*As the focus of the covariance analysis was looking at differences between stirrer types, these categoric variables needed a quantitative value eigen value which was assigned to reflect the number of corners/ blades in the stirrer.

**a.**



**b.**
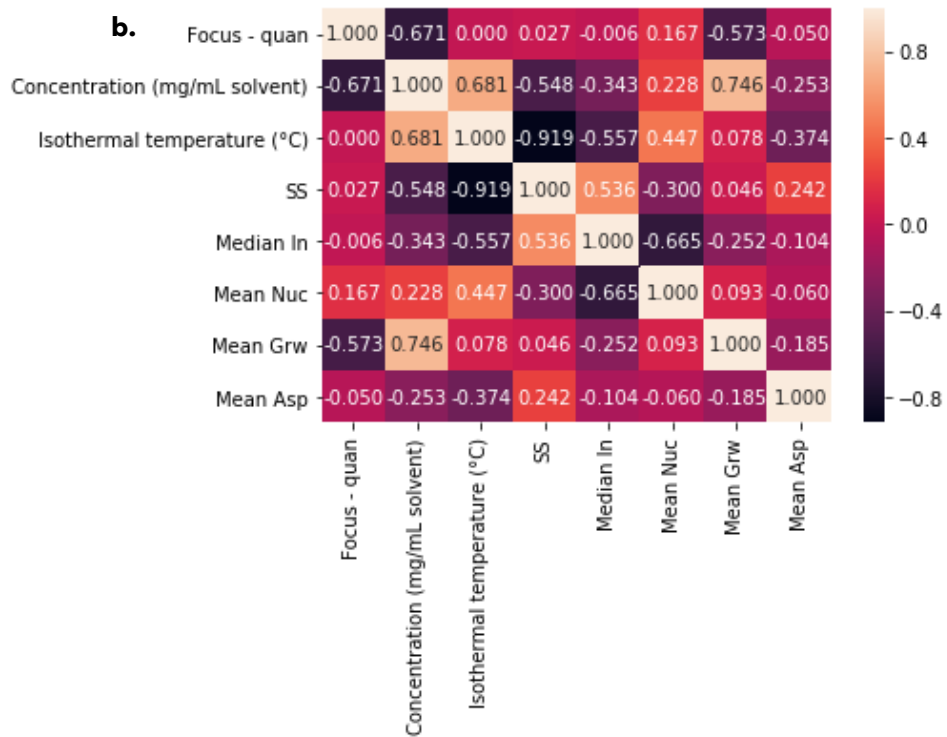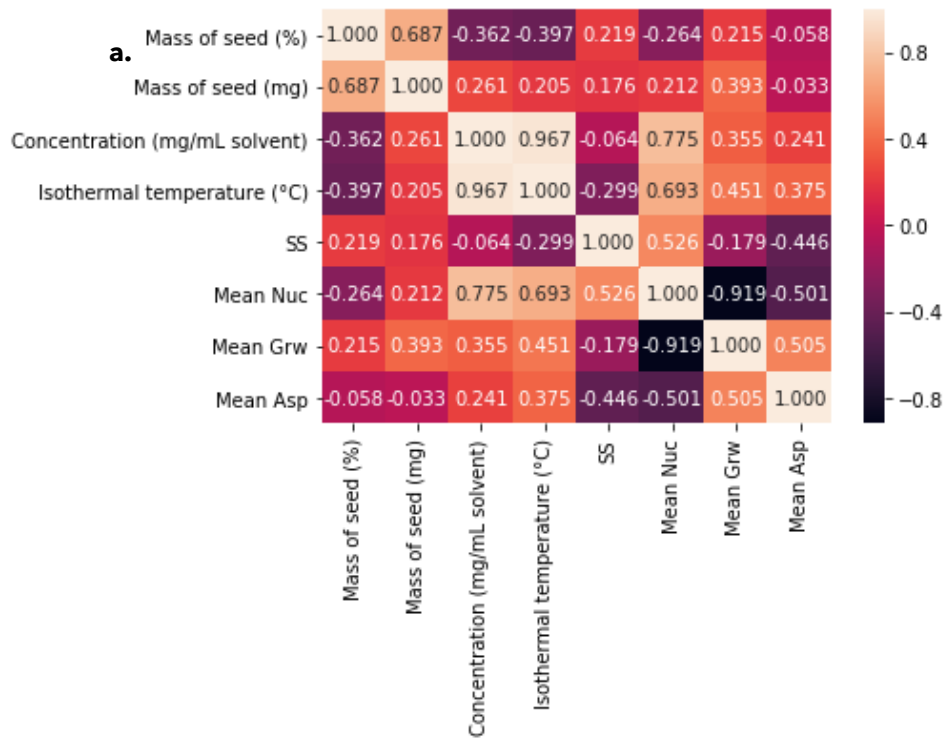
*The figure continues on the next page.*
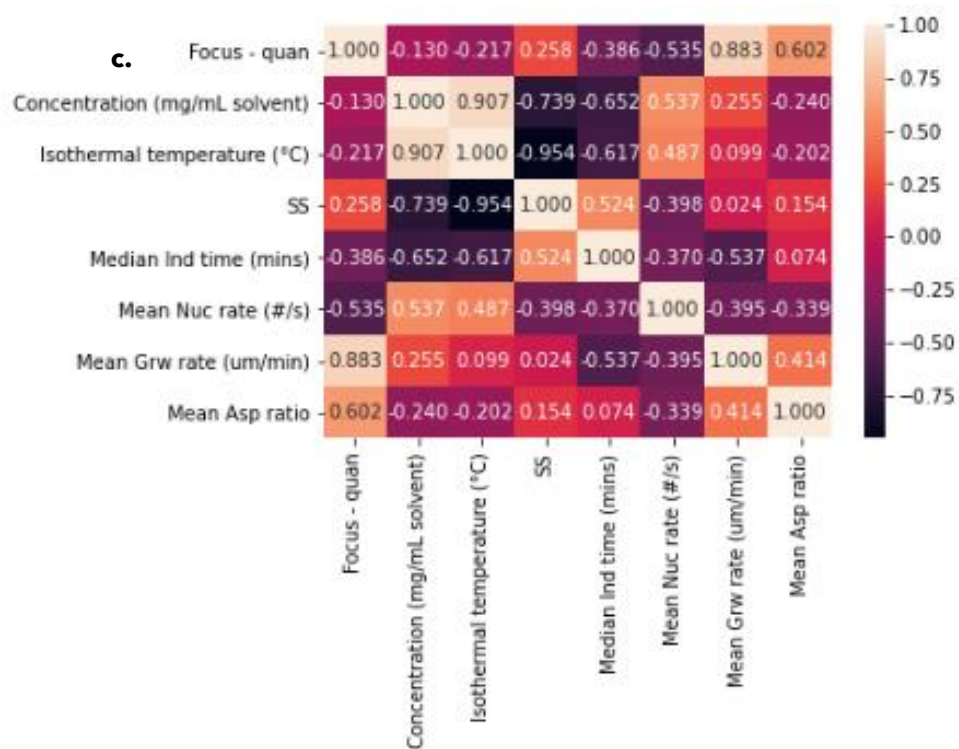
*Figure 83. The covariance matrixes of varied and measured variables across the kinetic parameter study for seeded isothermal experiments (a) for three different antisolvent compositions, initial (assigned as 0), halfway (assigned as 0.25) and final (assigned as 0.5)\* (b) and study for Crystalline and EasyMax isothermal experiments (c).*

\*Eigenvalues were assigned according to the part of antisolvent added relative to the initial starting volume.
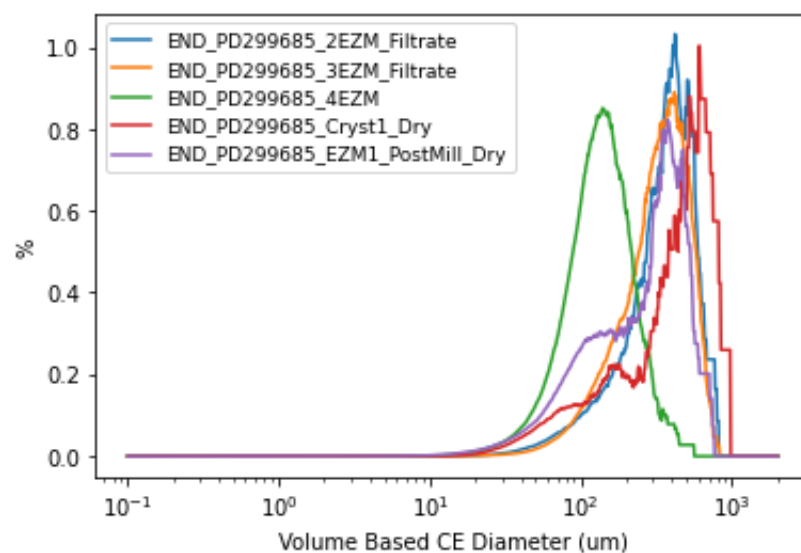
*Figure 84. Volume-based PSD of the various end process crystallisation of PD-299685 from the Crystalline and EasyMax.*

# 9. Conclusions.

## 9.1. General conclusions

This thesis has demonstrated the usefulness of a standardised workflow applied to pharmaceutical crystallisation data collection. The workflow incorporated QbD principles and was accelerated to advance the workflow to incorporate QbDD principles. The workflow was subsequently validated giving the opportunity for the work in this thesis to be integrated into industrial pharmaceutical research with confidence.

The main conclusions, aligned to each research chapter, can be summarised as follows:

**1. Developing a model-driven workflow for the digital design of small-scale batch cooling crystallisation with the antiviral lamivudine.** A logical 11-step workflow was developed (main sections including material characterisation, solubility studies, kinetic parameter estimation and solid-state analysis). QbD principles such as using prior knowledge, setting target parameters and DoE experiments allowed for the process optimisation for the crystallisation of lamivudine. Data collection was done in 80 polythermal and 28 isothermal experiments using only 53 g of API exemplifying vast improvements in efficiency and reduction in material wastage. Despite the challenges associated with aspects of the workflow e.g., hard-to-review prior knowledge, bi-objective only optimisation or equipment costs and constraints the anti-viral case study highlighted areas within the workflow for digitalisation into QbDD principles.

**2. High-throughput screening for large-scale data collection to inform medicine manufacture of aspirin.** The workflow was demonstrated through a second case study, the crystallisation of aspirin, which took circa 4 weeks and 71 g of API. The structured data collection showed both efficiency and generalisability of the workflow and suggests a successful application in industrial pharmaceutical research and manufacture.

**3. Adaptive Bayesian optimisation of process conditions for small-scale batch cooling crystallisation across fast and slow kinetic parameters: a comparative study.** DoE (implemented in PyMOO[199]) and Bayesian (GPyOpt[200]) methods have significant improvements over grid search methods at finding the optimum process conditions to satisfy the target crystallisation kinetic parameters. AdBO outperformed all other optimisation algorithms in reducing the objective function value for both case studies. Furthermore, AdBO required much fewer experiments. This time, and subsequently material, reduction translates to an approximate saving of 15-80 kWh for a general lab electricity requirement and up to £20,000/ kg for an in-demand anti-viral API (lamivudine).

**4. Utilisation of the model-driven workflow for the digital design of small-scale batch cooling crystallisation of a broader scope of the chemical space.** Five API case studies were used to exemplify the integration of the latest Python Crystalline image analysis (PyCIA) model and the AdBO experimental planner into the workflow. No adaptations to the workflow were required across the breadth of chemical space thus showing good generalisability of the workflow for future case studies. This suggests that the translation of the logic of the workflow, or specific sections, to a robotic smart lab would be achievable without change. A modulus approach to the optimisation algorithm was demonstrated through the inclusion of a fourth objective to avoid persistent needles for D-mannitol. The ability to introduce new objectives into the optimisation problem will allow for the adaptation of this work into other areas of research such as process chemistry, drug product development and beyond.

**5. Integration of the model-driven workflow into an industrial pharmaceutical facility (Pfizer): Supporting Process Development of API Crystallisation.** The workflow was validated and expanded using an industry-relevant API case study. The comparison between different stirring methods and crystalliser volume validated that the workflow provided key information for thermodynamic and kinetic models and crystallisation process design. Seeded and antisolvent experiments were performed to allow for the design of a hybrid antisolvent and cooling process crystallisation of PD-299685 that aligned with previous shape and size parameters and had improved yield. The integration of the workflow into an industrial pharmaceutical research facility allowed for vast time and material

savings over more traditional approaches. Increases in efficiency, as demonstrated here, are of paramount importance in industry where time constraints and costs are a lot more stringent.

# References.

1.      Khanna, I., Drug discovery in pharmaceutical industry: productivity challenges and trends. *Drug discovery today* **2012,** *17* (19-20), 1088-1102.

2.      Black, S.; Muller, F., On the Effect of Temperature on Aqueous Solubility of Organic Solids. *Org. Process Res. Dev* **2010,** *14* (3), 661-665.

3.      Vedantam, S.; Ranade, V. V., Crystallization: Key thermodynamic, kinetic and hydrodynamic aspects. *Sadhana* **2013,** *38* (6), 1287-1337.

4.      Sadeghi, M.; Rasmuson, Å. C., On the estimation of crystallization driving forces. *CrystEngComm* **2019,** *21* (34), 5164-5173.

5.      Doran, P. M., Unit Operations. *Bioprocess engineering principles* **2013**, 445-595.

6.      De Yoreo, J. J., Casting a bright light on Ostwald's rule of stages. *Proceedings of the National Academy of Sciences* **2022,** *119* (7).

7.      Karthika, S.;  Radhakrishnan, T. K.; Kalaichelvi, P., A Review of Classical and Nonclassical Nucleation Theories. *Crystal Growth & Design* **2016,** *16* (11), 6663-6681.

8.      Evans, J. S., Polymorphs, Proteins, and Nucleation Theory: A Critical Analysis. *Minerals* **2017,** *7*, 62.

9.      Garten, V. A.; Head, R. B., Crystalloluminescence and the nature of the critical nucleus. *Philosophical Magazine* **1963,** *8* (95), 1793-1803.

10.     Otpushchennikov, N., Determination of the size of the elementary crystal nucleus from acoustic measurement. *Soviet Phys. Cryst* **1962,** *7*, 237-240.

11.     Ostwald, W., Studien über die Bildung und Umwandlung fester Körper. *1. Abhandlung: Übersättigung und Überkaltung* **1897,** *22U* (1), 289-330.

12.     Garside, J., The concept of effectiveness factors in crystal growth. *Chemical Engineering Science* **1971,** *26* (9), 1425-1431.

13.     Voorhees, P. W., The theory of Ostwald ripening. *Journal of Statistical Physics* **1985,** *38* (1-2), 231-252.

14.     Lafferrère, L.;  Hoff, C.; Veesler, S., In Situ Monitoring of the Impact of Liquid−Liquid Phase Separation on Drug Crystallization by Seeding. *Cryst. Growth Des* **2004,** *4* (6), 1175-1180.

15.	Mangin, D.; Puel, F.; Veesler, S., Polymorphism in Processes of Crystallization in Solution: A Practical Review. *Organic Process Research & Development* **2009,** *13* (6), 1241-1253.

16.	Kadam, K.; Chavan, R., Influence of solid-state chemistry in drug substances in pharmaceutical products. A review. *International Journal of Chemical Studies* **2016,** *4*, 102-105.

17.	López-Mejías, V.; Kampf, J. W.; Matzger, A. J., Nonamorphism in Flufenamic Acid and a New Record for a Polymorphic Compound with Solved Structures. *Journal of the American Chemical Society* **2012,** *134* (24), 9872-9875.

18.	Yu, L., Amorphous pharmaceutical solids: preparation, characterization and stabilization. *Advanced drug delivery reviews* **2001,** *48* (1), 27-42.

19.	Hancock, B. C.; Parks, M., What is the true solubility advantage for amorphous pharmaceuticals? *Pharmaceutical research* **2000,** *17*, 397-404.

20.	Chekal, B. P.; Campeta, A. M.; Abramov, Y. A.; Feeder, N.; Glynn, P. P.; McLaughlin, R. W.; Meenan, P. A.; Singer, R. A., The Challenges of Developing an API Crystallization Process for a Complex Polymorphic and Highly Solvating System. Part I. *Organic process research & development.* **2009,** *13*, 1327-1337.

21.	Berge, S. M.; Bighley, L. D.; Monkhouse, D. C., Pharmaceutical salts. *Journal of pharmaceutical sciences* **1977,** *66* (1), 1-19.

22.	Trescot, A., Extended-release morphine sulfate in treatment of severe acute and chronic pain. *Journal of Pain Research* **2010**, 191.

23.	Gadade, D. D.; Pekamwar, S. S., Pharmaceutical cocrystals: Regulatory and strategic aspects, design and development. *Advanced pharmaceutical bulletin* **2016,** *6* (4), 479-494.

24.	Sekhon, B. S., Drug-drug co-crystals. *Daru* **2012,** *20* (1), 45-45.

25.	Thipparaboina, R.; Kumar, D.; Chavan, R. B.; Shastri, N. R., Multidrug co-crystals: towards the development of effective therapeutic hybrids. *Drug Discov Today* **2016,** *21* (3), 481-490.

26.	Bashimam, M.; El-Zein, H., Pharmaceutical cocrystal of antibiotic drugs: A comprehensive review. *Heliyon* **2022,** *8* (12), e11872.

27.	Muller, F. L.; Fielding, M.; Black, S., A Practical Approach for Using Solubility to Design Cooling Crystallisations. *Organic Process Research & Development* **2009,** *13* (6), 1315-1321.

28.     Black, S. N.;  Quigley, K.;  Parker, A., A Well-Behaved Crystallisation of a Pharmaceutical Compound. *Org. Process Res. Dev* **2006,** *10* (2), 241-244.

29.     Jia, S.;  Yang, P.;  Gao, Z.;  Li, Z.;  Fang, C.; Gong, J., Recent progress in antisolvent crystallization. *CrystEngComm* **2022,** *24* (17), 3122-3135.

30.     Barrett, P.;  Smith, B.;  Worlitschek, J.;  Bracken, V.;  O'Sullivan, B.; O'Grady, D., A Review of the Use of Process Analytical Technology for the Understanding and Optimization of Production Batch Crystallization Processes. *Org. Process Res. Dev* **2005,** *9* (3), 348-355.

31.     Besenhard, M. O.;  Neugebauer, P.;  Scheibelhofer, O.; Khinast, J. G., Crystal Engineering in Continuous Plug-Flow Crystallizers. *Cryst. Growth Des* **2017,** *17* (12), 6432-6444.

32.     Cui, Y.;  O'Mahony, M.;  Jaramillo, J. J.;  Stelzer, T.; Myerson, A. S., Custom-Built Miniature Continuous Crystallization System with Pressure-Driven Suspension Transfer. *Org. Process Res. Dev* **2016,** *20* (7), 1276-1282.

33.     Brown, C. J.;  McGlone, T.;  Yerdelen, S.;  Srirambhatla, V.;  Mabbott, F.; Gurung, R.;  L. Briuglia, M.;  Ahmed, B.;  Polyzois, H.;  McGinty, J.;  Perciballi, F.; Fysikopoulos, D.;  MacFhionnghaile, P.;  Siddique, H.;  Raval, V.;  Harrington, T. S.; Vassileiou, A. D.;  Robertson, M.;  Prasad, E.;  Johnston, A.;  Johnston, B.;  Nordon, A.;  Srai, J. S.;  Halbert, G.;  ter Horst, J. H.;  Price, C. J.;  Rielly, C. D.;  Sefcik, J.; Florence, A. J., Enabling precision manufacturing of active pharmaceutical ingredients: workflow for seeded cooling continuous crystallisations. *Molecular systems design & engineering* **2018,** *3* (3), 518-549.

34.     Sangshetti, J. N.;  Deshpande, M.;  Zaheer, Z.;  Shinde, D. B.; Arote, R., Quality by design approach: Regulatory need. *Arabian journal of chemistry* **2017,** *10* (S2), S3412-S3425.

35.     Belič, A.;  Škrjanc, I.;  Božič, D. Z.; Vrečer, F., Tableting process optimisation with the application of fuzzy models. *Int J Pharm* **2010,** *389* (1), 86-93.

36.     Asteasuain, M.; Brandolin, A., Modeling and optimization of a high-pressure ethylene polymerization reactor using gPROMS. *Computers & chemical engineering* **2008,** *32* (3), 396-408.

37.     Elzomor, M.; Parrish, K., Investigating Building Construction Process and Developing a Performance Index. *Procedia engineering* **2016,** *145*, 211-218.

38.     Voutchkov, I.; Jaworski, B.; Vitanov, V. I.; Bedford, G. M., An integrated approach to friction surfacing process optimisation. *Surface & coatings technology* **2001,** *141* (1), 26-33.

39.     Juran, J. M., *Juran on Quality by Design: The New Steps for Planning Quality into Goods and Services*. Penton Media, Inc., Penton Business Media, Inc. and their subsidiaries: **1992**; Vol. 241, p 60.

40.     Buyya, R.; Vahid Dastjerdi, A., *Internet of Things: Principles and Paradigms*. San Francisco: Elsevier Science & Technology: San Francisco, **2016**.

41.     Sanders, A.; Elangeswaran, C.; Wulfsberg, J., Industry 4.0 implies lean manufacturing: Research activities in industry 4.0 function as enablers for lean manufacturing. *Journal of industrial engineering and management* **2016,** *9* (3), 811-833.

42.     Arden, N. S.; Fisher, A. C.; Tyner, K.; Yu, L. X.; Lee, S. L.; Kopcha, M., Industry 4.0 for pharmaceutical manufacturing: Preparing for the smart factories of the future. *Int. J. Pharm. (Amsterdam, Neth.)* **2021,** *602*, 120554.

43.     Xu, X.; Lu, Y.; Vogel-Heuser, B.; Wang, L., Industry 4.0 and Industry 5.0— Inception, conception and perception. *Journal of Manufacturing Systems* **2021,** *61*, 530-535.

44.     https://knowhow.distrelec.com/manufacturing/is-your-business-ready-for-industry-5-0/ Is Your Business Ready for Industry 5.0?

45.     Nagaich, U., Pharmaceutical "quality by design" approach. *J Adv Pharm Technol Res* **2018,** *9* (1), 1-1.

46.     Venkatasubramanian, V.; Morris, K. R., Drinking from a fire hose: the new FDA framework for PAT and Quality by Design. *Pharmaceutical technology Europe* **2008,** *20* (9), 40.

47.     Namjoshi, S.; Dabbaghi, M.; Roberts, M. S.; Grice, J. E.; Mohammed, Y., Quality by Design: Development of the Quality Target Product Profile (QTPP) for Semisolid Topical Products. *Pharmaceutics* **2020,** *12* (3), 287.

48.     Owen, M. R.; Luscombe, C.; Lai; Godbert, S.; Crookes, D. L.; Emiabata-Smith, D., Efficiency by Design: Optimisation in Process Research. *Org. Process Res. Dev* **2001,** *5* (3), 308-323.

49.     Rosas, J. G.; Blanco, M.; González, J. M.; Alcalá, M., Quality by design approach of a pharmaceutical gel manufacturing process, part 2: Near infrared

monitoring of composition and physical parameters. *J Pharm Sci* **2011,** *100* (10), 4442-4451.

50.　　Maltesen, M. J.;　Bjerregaard, S.;　Hovgaard, L.;　Havelund, S.; van de Weert, M., Quality by design – Spray drying of insulin intended for inhalation. *Eur J Pharm Biopharm* **2008,** *70* (3), 828-838.

51.　　Zhou, D.; Qiu, Y., Understanding drug properties in formulation and process design of solid oral products. *Journal of validation technology* **2010,** *16* (4), 74.

52.　　www.ich.org.uk ICH Harmonisation for better health.

53.　　Ter Horst, J. P.;　Turimella, S. L.;　Metsers, F.; Zwiers, A., Implementation of Quality by Design (QbD) Principles in Regulatory Dossiers of Medicinal Products in the European Union (EU) Between 2014 and 2019. *Therapeutic Innovation & Regulatory Science* **2021,** *55* (3), 583-590.

54.　　Kimball, M., Manufacturing Topical Formulations: Scale-up from Lab to Pilot Production. In *Handbook of Formulating Dermal Applications*, **2016**; pp 167-232.

55.　　Shah, B.;　Khunt, D.;　Bhatt, H.;　Misra, M.; Padh, H., Intranasal delivery of venlafaxine loaded nanostructured lipid carrier: Risk assessment and QbD based optimization. *Journal of drug delivery science and technology*. **2016,** *33*, 37-50.

56.　　Ghosh, I.;　Schenck, D.;　Bose, S.;　Liu, F.; Motto, M., Identification of critical process parameters and its interplay with nanosuspension formulation prepared by top down media milling technology - A QbD perspective. *Pharm Dev Technol* **2013,** *18* (3), 719-729.

57.　　www.fda.gov HACCP Principles & Application Guidelines.

58.　　Yang, H.;　Zhang, L.; Galinski, M., A probabilistic model for risk assessment of residual host cell DNA in biological products. *Vaccine* **2010,** *28* (19), 3308-3311.

59.　　Rathore, A. S., Roadmap for implementation of quality by design (QbD) for biotechnology products. *Trends Biotechnol* **2009,** *27* (9), 546-553.

60.　　Montgomery, D. C., *Design and analysis of experiments*. John wiley & sons: **2017**.

61.　　Barrentine, L. B., *An introduction to design of experiments: a simplified approach*. Quality Press: **1999**.

62.　　Jacyna, J.;　Kordalewska, M.; Markuszewski, M. J., Design of Experiments in metabolomics-related studies: An overview. *Journal of pharmaceutical and biomedical analysis*. **2019,** *164*, 598-606.

63.     Delgado-Nieblas, C. I.;  Zazueta-Morales, J. J.;  Ahumada-Aguilar, J. A.; Aguilar-Palazuelos, E.;  Carrillo-López, A.;  Jacobo-Valenzuela, N.; Telis-Romero, J., Optimization of an Air-Drying Process to Obtain a Dehydrated Naranjita (Citrus Mitis B.) Pomace Product With High Bioactive Compounds and Antioxidant Capacity. *Journal of food process engineering* **2017,** *40* (1).

64.     Garai, L., Improving HPLC Analysis of Vitamin A and E: Use of Statistical Experimental Design. *Procedia Computer Science* **2017,** *108*, 1500-1511.

65.     Glassey, J.;  Gernaey, K. V.;  Clemens, C.;  Schulz, T. W.;  Oliveira, R.; Striedner, G.;  Mandenius, C.-F., Process analytical technology (PAT) for biopharmaceuticals. *Biotechnology journal.* **2011,** *6* (4), 369-377.

66.     Chen, Z.;  Lovett, D.; Morris, J., Process analytical technologies and real time process control a review of some spectroscopic issues and challenges. *Journal of process control* **2011,** *21* (10), 1467-1482.

67.     Wu, L.; Vogt, F. G., A review of recent advances in mass spectrometric methods for gas-phase chiral analysis of pharmaceutical and biological compounds. *J Pharm Biomed Anal* **2012,** *69*, 133-147.

68.     Zhou, Y.;  Song, J.-Z.;  Choi, F. F.-K.;  Wu, H.-F.;  Qiao, C.-F.;  Ding, L.-S.; Gesang, S.-L.; Xu, H.-X., An experimental design approach using response surface techniques to obtain optimal liquid chromatography and mass spectrometry conditions to determine the alkaloids in Meconopsi species. *Journal of Chromatography A* **2009,** *1216* (42), 7013-7023.

69.     Sistare, F.;  St Pierre Berry, L.; Mojica, C. A., Process Analytical Technology: An Investment in Process Knowledge. *Organic process research & development.* **2005,** *9* (3), 332-336.

70.     Medendorp, J.; Lodder, R. A., Acoustic-resonance spectrometry as a process analytical technology for rapid and accurate tablet identification. *AAPS PharmSciTech* **2006,** *7* (1), E175-E183.

71.     Akseli, I.; Cetinkaya, C., Air-coupled non-contact mechanical property determination of drug tablets. *Int J Pharm* **2008,** *359* (1), 25-34.

72.     Rosa, S. S.;  Barata, P. A.;  Martins, J. M.; Menezes, J. C., Development and validation of a method for active drug identification and content determination of ranitidine in pharmaceutical products using near-infrared reflectance spectroscopy: A parametric release approach. *Talanta* **2008,** *75* (3), 725-733.

73.     St-Onge, L.; Kwong, E.; Sabsabi, M.; Vadas, E. B., Rapid analysis of liquid formulations containing sodium chloride using laser-induced breakdown spectroscopy. *J Pharm Biomed Anal* **2004,** *36* (2), 277-284.

74.     Guez, J. S.; Cassar, J. P.; Wartelle, F.; Dhulster, P.; Suhr, H., Real time in situ microscopy for animal cell-concentration monitoring during high density culture in bioreactor. *J Biotechnol* **2004,** *111* (3), 335-343.

75.     Maddikunta, P. K. R.; Pham, Q.-V.; Prabadevi, B.; Deepa, N.; Dev, K.; Gadekallu, T. R.; Ruby, R.; Liyanage, M., Industry 5.0: A survey on enabling technologies and potential applications. *Journal of Industrial Information Integration* **2022,** *26*, 100257.

76.     www.analyticsvidhya.com Commonly used Machine Learning Algorithms (with Python and R Codes).

77.     Srinivas, M.; Patnaik, L. M., Genetic algorithms: a survey. *Computer* **1994,** *27* (6), 17-26.

78.     Mirjalili, S., Genetic Algorithm. In *Studies in Computational Intelligence*, Springer International Publishing: **2019**; pp 43-55.

79.     Katoch, S.; Chauhan, S. S.; Kumar, V., A review on genetic algorithm: past, present, and future. *Multimedia Tools and Applications* **2021,** *80* (5), 8091-8126.

80.     Izadi, A.; Kimiagari, A. M., Distribution network design under demand uncertainty using genetic algorithm and Monte Carlo simulation approach: a case study in pharmaceutical industry. *Journal of Industrial Engineering International* **2014,** *10* (1), 1.

81.     Hamamoto, S., Development and validation of genetic algorithm-based facility layout a case study in the pharmaceutical industry. *International Journal of Production Research* **1999,** *37* (4), 749-768.

82.     Velásco-Mejía, A.; Vallejo-Becerra, V.; Chávez-Ramírez, A.; Torres-González, J.; Reyes-Vidal, Y.; Castañeda-Zaldivar, F., Modeling and optimization of a pharmaceutical crystallization process by using neural networks and genetic algorithms. *Powder Technology* **2016,** *292*, 122-128.

83.     Storn, R. In *On the usage of differential evolution for function optimization*, Proceedings of North American Fuzzy Information Processing, IEEE **1996**.

84.     Price, K. V., Differential Evolution. In *Handbook of Optimization*, Springer Berlin Heidelberg: **2013**; pp 187-214.

85.    Pant, M.;  Zaheer, H.;  Garcia-Hernandez, L.; Abraham, A., Differential Evolution: A review of more than two decades of research. *Engineering Applications of Artificial Intelligence* **2020,** *90*, 103479.

86.    Das, S.; Suganthan, P. N., Differential Evolution: A Survey of the State-of-the-Art. *IEEE Transactions on Evolutionary Computation* **2011,** *15* (1), 4-31.

87.    Saraswat, M.;  Arya, K.; Sharma, H., Leukocyte segmentation in tissue images using differential evolution algorithm. *Swarm and Evolutionary Computation* **2013,** *11*, 46-54.

88.    Kaur, M.;  Singh, D.; Kumar, V., Drug synergy prediction using dynamic mutation based differential evolution. *Current Pharmaceutical Design* **2021,** *27* (8), 1103-1111.

89.    Eswari, J. S.; Venkateswarlu, C., Multiobjective Pareto optimisation of pharmaceutical product formulation using radial basis function network and non-dominated sorting differential evolution. *International Journal of Biomedical Engineering and Technology* **2020,** *33* (2), 95-122.

90.    Hansen, N.; Ostermeier, A., Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation* **2001,** *9* (2), 159-195.

91.    Tan, U.;  Rabaste, O.;  Adnet, C.; Ovarlez, J.-P. In *On the Eclipsing Phenomenon with Phase Codes*, 2019 International Radar Conference (RADAR) IEEE **2019**.

92.    Loshchilov, I.; Hutter, F., CMA-ES for Hyperparameter Optimization of Deep Neural Networks. **2016**.

93.    Jebalia, M.;  Auger, A.;  Schoenauer, M.;  James, F.; Postel, M. In *Identification of the Isotherm Function in Chromatography Using CMA-ES*, 2007 IEEE Congress on Evolutionary Computation IEEE: **2007**.

94.    Wen, T.;  Mihail, R. P.; Vidal, F. P., 3D-2D Registration Using X-Ray Simulation and CMA-ES. In *Applications of Evolutionary Computation*, Springer International Publishing: **2021**; pp 453-468.

95.    Nelder, J. A.; Mead, R., A Simplex Method for Function Minimization. *The Computer Journal* **1965,** *7* (4), 308-313.

96.    Cheng, J. Y.; Mailund, T., Ancestral population genomics using coalescence hidden Markov models and heuristic optimisation algorithms. *Computational Biology and Chemistry* **2015,** *57*, 80-92.

97.     Singer, S.; Nelder, J., Nelder-mead algorithm. *Scholarpedia* **2009,** *4* (7), 2928.

98.     Fath, V.; Kockmann, N.; Otto, J.; Röder, T., Self-optimising processes and real-time-optimisation of organic syntheses in a microreactor system using Nelder–Mead and design of experiments. *Reaction Chemistry & Engineering* **2020,** *5* (7), 1281-1299.

99.     Rajawat, A. S.; Rawat, R.; Barhanpurkar, K.; Shaw, R. N.; Ghosh, A., Depression detection for elderly people using AI robotic systems leveraging the Nelder–Mead Method. In *Artificial Intelligence for Future Generation Robotics*, Elsevier: **2021**; pp 55-70.

100.    Idkaidek, N. M., Comparative assessment of saliva and plasma for drug bioavailability and bioequivalence studies in humans. *Saudi Pharmaceutical Journal* **2017,** *25* (5), 671-675.

101.    Hooke, R.; Jeeves, T. A., "Direct Search'' Solution of Numerical and Statistical Problems. *Journal of the ACM* **1961,** *8* (2), 212-229.

102.    Zewail, I.; Saad, W.; Shokair, M.; El_dolil, S. A., Maximization of total throughput using pattern search algorithm in underlay cognitive radio network. *Menoufia Journal of Electronic Engineering Research* **2017,** *26* (2), 307-319.

103.    Torczon, V., On the Convergence of Pattern Search Algorithms. *SIAM Journal on Optimization* **1997,** *7* (1), 1-25.

104.    Simon, l.; Schongut, M.; Stepanek, F.; Hungerbuhler, K., Optimal Particle Size Distribution Control in a Pharmaceuticals Batch Dryer Using a Hybrid Genetic Algorithm-Pattern Search Optimization Strategy. *Scientia Pharmaceutica* **2010,** *78* (3), 566.

105.    Feuilleaubois, E.; Fabart, V.; Doucet, J. P., Implementation of the Three-Dimensional-Pattern Search Problem on Hopfield-like Neural Networks. *SAR and QSAR in Environmental Research* **1993,** *1* (2-3), 97-114.

106.    Takahasi, Y.; Akagi, T.; Sasaki, S.-i., Three-dimensional pharmacophoric pattern search using COMPASS: Nootropic agents. *Tetrahedron Computer Methodology* **1990,** *3* (1), 27-35.

107.    Joy, T. T.; Rana, S.; Gupta, S.; Venkatesh, S., A flexible transfer learning framework for Bayesian optimization with convergence guarantee. *Expert Systems with Applications* **2019,** *115*, 656-672.

108.    Clayton, A. D.; Pyzer-Knapp, E. O.; Purdie, M.; Jones, M. F.; Barthelme, A.; Pavey, J.; Kapur, N.; Chamberlain, T. W.; Blacker, A. J.; Bourne, R. A., Bayesian Self-Optimization for Telescoped Continuous Flow Synthesis. *Angewandte Chemie International Edition* **2023,** *62* (3).

109.    Moriconi, R.; Deisenroth, M. P.; Sesh Kumar, K. S., High-dimensional Bayesian optimization using low-dimensional feature spaces. *Machine Learning* **2020,** *109* (9-10), 1925-1943.

110.    Martinez-Cantin, R.; De Freitas, N.; Doucet, A.; Castellanos, J. A., Active policy learning for robot planning and exploration under uncertainty. **2008,** *3*, 321-328.

111.    Hoffman, M. W.; Shahriari, B.; Nando de, F., Exploiting correlation and budget constraints in Bayesian multi-armed bandit optimization. *arXiv.org* **2013**.

112.    Joy, T. T.; Rana, S.; Gupta, S.; Venkatesh, S. In *Hyperparameter tuning for big data using Bayesian optimisation*, 2016 23rd International Conference on Pattern Recognition (ICPR) IEEE **2016**.

113.    Olofsson, S.; Mehrian, M.; Calandra, R.; Geris, L.; Deisenroth, M. P.; Misener, R., Bayesian Multiobjective Optimisation With Mixed Analytical and Black-Box Functions: Application to Tissue Engineering. *IEEE Transactions on Biomedical Engineering* **2019,** *66* (3), 727-739.

114.    Luna, M. F.; Martínez, E. C., Bayesian optimization of crystallization processes to guarantee end-use product properties. *Latin American Applied Research* **2020,** *50* (2), 109-114.

115.    Roni, M. H. K.; Rana, M. S.; Pota, H. R.; Hasan, M. M.; Hussain, M. S., Recent trends in bio-inspired meta-heuristic optimization techniques in control applications for electrical systems: a review. *International Journal of Dynamics and Control* **2022,** *10* (3), 999-1011.

116.    Burger, B.; Maffettone, P. M.; Gusev, V. V.; Aitchison, C. M.; Bai, Y.; Wang, X.; Li, X.; Alston, B. M.; Li, B.; Clowes, R.; Rankin, N.; Harris, B.; Sprick, R. S.; Cooper, A. I., A mobile robotic chemist. *Nature* **2020,** *583* (7815), 237-241.

117.    Bose, B.; Kalra, A. K.; Thukral, S.; Sood, A.; Guha, S. K.; Anand, S. In *Tremor compensation for robotics assisted microsurgery*, Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE **1992**.

118.     Elhawary, H.; Zivanovic, A.; Rea, M.; Davies, B. L.; Besant, C.; Young, I.; Lamperth, M. U., A Modular Approach to MRI-Compatible Robotics. *IEEE Engineering in Medicine and Biology Magazine* **2008,** *27* (3), 35-41.

119.     Bischoff, R.; Kurth, J.; Schreiber, G.; Koeppe, R.; Albu-Schäffer, A.; Beyer, A.; Eiberger, O.; Haddadin, S.; Stemmer, A.; Grunwald, G. In *The KUKA-DLR Lightweight Robot arm-a new reference platform for robotics research and manufacturing*, ISR 2010 (41st international symposium on robotics) and ROBOTIK 2010 (6th German conference on robotics) VDE **2010**; pp 1-8.

120.     Chen, F.; Selvaggio, M.; Caldwell, D. G., Dexterous Grasping by Manipulability Selection for Mobile Manipulator With Visual Guidance. *IEEE Transactions on Industrial Informatics* **2019,** *15* (2), 1202-1210.

121.     Taesi, C.; Aggogeri, F.; Pellegrini, N., COBOT Applications—Recent Advances and Challenges. *Robotics* **2023,** *12* (3), 79.

122.     www.theengineer.co.uk The benefits of automation in pharmaceutical manufacturing.

123.     Alexovič, M.; Dotsikas, Y.; Bober, P.; Sabo, J., Achievements in robotic automation of solvent extraction and related approaches for bioanalysis of pharmaceuticals. *J Chromatogr B Analyt Technol Biomed Life Sci* **2018,** *1092*, 402-421.

124.     Schneider, G., Automating drug discovery. *Nature Reviews Drug Discovery* **2018,** *17* (2), 97-113.

125.     https://www.crystallizationsystems.com/crystalline Crystalline (Technobis)

126.https://www.mt.com/gb/en/home/products/L1_AutochemProducts/chemical-synthesis-reactor-systems/EasyMax-Synthesis-Reactor.html     EasyMax (Mettler Toledo)

127.     https://www.blazemetrics.com/ BlazeMetrics Single Probe - Multiple Measurement Technologies

128.     https://www.bruker.com/en/products-and-solutions/diffractometers-and-scattering-systems/x-ray-diffractometers/d8-discover-family.html     Bruker D8 DISCOVER FAMILY

129.     Ring, S. G.; Colonna, P.; I'Anson, K. J.; Kalichevsky, M. T.; Miles, M. J.; Morris, V. J.; Orford, P. D., The gelation and crystallisation of amylopectin. *Carbohydrate research* **1987,** *162* (2), 277-293.

130.    Harris, R. K.;  Yeung, R. R.;  Lamont, R. B.;  Lancaster, R. W.;  Lynn, S. M.;  Staniforth, S. E., 'Polymorphism' in a novel anti-viral agent: Lamivudine. *Journal of the Chemical Society, Perkin Transactions 2* **1997,** (12), 2653-2660.

131.    Lau, D. T. Y.;  Lau, D., Long-Term Therapy of Chronic Hepatitis B With Lamivudine. *Hepatology :* **2000,** *32* (4), 828-834.

132.    García-Trejo, J. J.;  Ortega, R.; Zarco-Zavala, M., Putative repurposing of lamivudine, a nucleoside/nucleotide analogue and antiretroviral to improve the outcome of cancer and COVID-19 patients. *Frontiers in Oncology* **2021,** *11*.

133.    Roy, B. N.;  Singh, G. P.;  Shrivastava, D.;  Jadhav, H. S.;  Aher, U. P.; Deokar, S. C. An improved process for the manufacture of lamivudine (Patent No. WO2011141805A2).

134.    Prakash, K.;  Narayana Raju, P.;  Shanta Kumari, K.; Lakshmi Narasu, M., Solubility and Dissolution Rate Determination of Different Antiretroviral Drugs in Different pH Media Using UV Visible Spectrophotometer. *E-Journal of Chemistry* **2008,** *5* (s2), 1159-1164.

135.    Strauch, S.;  Jantratid, E.;  Dressman, J. B.;  Junginger, H. E.;  Kopp, S.;  Midha, K. K.;  Shah, V. P.;  Stavchansky, S.;  Barends, D. M., COMMENTARY: Biowaiver Monographs for Immediate Release Solid Oral Dosage Forms: Lamivudine. *Journal of Pharmaceutical Sciences* **2011,** *100* (6), 2054-2063.

136.    Harris, R. K.;  Yeung, R. R.;  Lamont, R. B.;  Lancaster, R. W.;  Lynn, S. M.;  Staniforth, S. E., 'Polymorphism' in a novel anti-viral agent: Lamivudine. *Perkin transactions. 2* **1997,** (12), 2653-2660.

137.    Renu, C.;  Poonam, A.; Swati, B., Polymorphic Forms of Lamivudine: Characterization, Estimation of Transition Temperature, and Stability Studies by Thermodynamic and Spectroscopic Studies. *ISRN Thermodynamics* **2012**.

138.    Reddy, B. P.;  Reddy, M. S.;  Reddy, P. V.; Kumari, M. V. Solid oral dosage forms of lamivudine (U.S. Patent No. US8481554B2).

139.    Wilkinson, M. D.;  Dumontier, M.;  Aalbersberg, I. J.;  Appleton, G.;  Axton, M.;  Baak, A.;  Blomberg, N.;  Boiten, J.-W.;  Da Silva Santos, L. B.;  Bourne, P. E.;  Bouwman, J.;  Brookes, A. J.;  Clark, T.;  Crosas, M.;  Dillo, I.;  Dumon, O.;  Edmunds, S.;  Evelo, C. T.;  Finkers, R.;  Gonzalez-Beltran, A.;  Gray, A. J. G.;  Groth, P.;  Goble, C.;  Grethe, J. S.;  Heringa, J.;  'T Hoen, P. A. C.;  Hooft, R.;  Kuhn, T.;  Kok, R.;  Kok, J.;  Lusher, S. J.;  Martone, M. E.;  Mons, A.;  Packer, A. L.;  Persson, B.;  Rocca-Serra, P.;

Roos, M.; Van Schaik, R.; Sansone, S.-A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M. A.; Thompson, M.; Van Der Lei, J.; Van Mulligen, E.; Velterop, J.; Waagmeester, A.; Wittenburg, P.; Wolstencroft, K.; Zhao, J.; Mons, B., The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **2016,** *3* (1), 160018.

140.    Knoll, S.; Jusner, C. E.; Sagmeister, P.; Williams, J. D.; Hone, C. A.; Horn, M.; Kappe, C. O., Autonomous model-based experimental design for rapid reaction development. *Reaction Chemistry & Engineering* **2022,** *7* (11), 2375-2384.

141.    Guenat, S.; Purnell, P.; Davies, Z. G.; Nawrath, M.; Stringer, L. C.; Babu, G. R.; Balasubramanian, M.; Ballantyne, E. E. F.; Bylappa, B. K.; Chen, B.; De Jager, P.; Del Prete, A.; Di Nuovo, A.; Ehi-Eromosele, C. O.; Eskandari Torbaghan, M.; Evans, K. L.; Fraundorfer, M.; Haouas, W.; Izunobi, J. U.; Jauregui-Correa, J. C.; Kaddouh, B. Y.; Lewycka, S.; Macintosh, A. C.; Mady, C.; Maple, C.; Mhiret, W. N.; Mohammed-Amin, R. K.; Olawole, O. C.; Oluseyi, T.; Orfila, C.; Ossola, A.; Pfeifer, M.; Pridmore, T.; Rijal, M. L.; Rega-Brodsky, C. C.; Robertson, I. D.; Rogers, C. D. F.; Rougé, C.; Rumaney, M. B.; Seeletso, M. K.; Shaqura, M. Z.; Suresh, L. M.; Sweeting, M. N.; Taylor Buck, N.; Ukwuru, M. U.; Verbeek, T.; Voss, H.; Wadud, Z.; Wang, X.; Winn, N.; Dallimer, M., Meeting sustainable development goals via robotics and autonomous systems. *Nature Communications* **2022,** *13* (1).

142.    https://www.bruker.com/content/bruker/int/en/products-and-solutions/diffractometers-and-x-ray-microscopes/x-ray-diffractometers/diffrac-suite-software/diffrac-eva.html Bruker DIFFRAC.EVA: software to evaluate X-ray diffraction data (Version 5.2)

143.    https://analyzing-testing.netzsch.com/en/products/software/proteus Netzsch Proteus

144.    https://www.sartorius.com/en/products/process-analytical-technology/data-analytics-software/support/knowledge-base/modde-121-550674 Sartorius Modde version 12.1

145.    Muller, F. L.; Fielding, M.; Black, S., A Practical Approach for Using Solubility to Design Cooling Crystallisations. *Org. Process Res. Dev* **2009,** *13* (6), 1315-1321.

146.    Deng, T.; Garg, V.; Pereira Diaz, L.; Markl, D.; Brown, C.; Florence, A.; Bradley, M. S. A., Comparative studies of powder flow predictions using milligrams

of powder for identifying powder flow issues. *International journal of pharmaceutics.* **2022,** *628*, 122309.

147. Butler, J. M.; Dressman, J. B., The Developability Classification System: Application of Biopharmaceutics Concepts to Formulation Development. *Journal of pharmaceutical sciences.* **2010,** *99*, 4940-4954.

148. Leane, M.; Pitt, K.; Reynolds, G., A proposal for a drug product Manufacturing Classification System (MCS) for oral solid dosage forms. *Pharmaceutical Development and Technology* **2015,** *20* (1), 12-21.

149. Vassileiou, A. D.; Robertson, M.; Wareham, B. G.; Soundaranathan, M.; Ottoboni, S.; Florence, A. J.; Hartwig, T.; Johnston, B. F., A Unified AI Framework for Solubility Prediction Across Organic Solvents. *Digital Discovery* **2023**.

150. Qiu, J.; Albrecht, J.; Janey, J., Synergistic Solvation Effects: Enhanced Compound Solubility Using Binary Solvent Mixtures. *Organic Process Research & Development* **2019,** *23* (7), 1343-1351.

151. Boetker, J.; Rades, T.; Rantanen, J.; Hawley, A.; Boyd, B. J., Structural Elucidation of Rapid Solution-Mediated Phase Transitions in Pharmaceutical Solids Using in Situ Synchrotron SAXS/WAXS. *Molecular Pharmaceutics* **2012,** *9* (9), 2787-2791.

152. Vankeirsbilck, T.; Vercauteren, A.; Baeyens, W.; Van der Weken, G.; Verpoort, F.; Vergote, G.; Remon, J. P., Applications of Raman spectroscopy in pharmaceutical analysis. *Trends in analytical chemistry* **2002,** *21*, 869-877.

153. Alder, C. M.; Hayler, J. D.; Henderson, R. K.; Redman, A. M.; Shukla, L.; Shuster, L. E.; Sneddon, H. F., Updating and further expanding GSK's solvent sustainability guide. *Green Chemistry* **2016,** *18* (13), 3879-3890.

154. Brandel, C.; Ter Horst, J. H., Measuring induction times and crystal nucleation rates. *Faraday Discussions* **2015,** *179*, 199-214.

155. Schoenitz, M.; Grundemann, L.; Augustin, W.; Scholl, S., Fouling in microstructured devices: a review. *Chemical Communications* **2015,** *51* (39), 8213-8228.

156. Jozwiakowski, M. J.; Nguyen, N.-A. T.; Sisco, J. M.; Spancake, C. W., Solubility Behavior of Lamivudine Crystal Forms in Recrystallization Solvents. *Journal of pharmaceutical sciences.* **1996,** *85* (2), 193-199.

157. https://services.nhsbsa.nhs.uk/dmd-browser/ NHS NHS (dm+d)

158. https://www.basf.com/global/en/who-we-are/innovation/our-innovations/health_and_nutrition/zoomlab.html ZoomLab

159. Diorazio, L. J.; Hose, D. R. J.; Adlington, N. K., Toward a More Holistic Framework for Solvent Selection. *Organic Process Research & Development* **2016,** *20* (4), 760-773.

160. Johnston, A.; Bhardwaj-Miglani, R.; Gurung, R.; Vassileiou, A. D.; Florence, A. J.; Johnston, B. F., Combined Chemoinformatics Approach to Solvent Library Design Using clusterSim and Multidimensional Scaling. *Journal of Chemical Information and Modeling* **2017,** *57* (8), 1807-1815.

161. Du, Y.; Zhang, H.; Xue, J.; Tang, W.; Fang, H.; Zhang, Q.; Li, Y.; Hong, Z., Vibrational spectroscopic study of polymorphism and polymorphic transformation of the anti-viral drug lamivudine. *Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy* **2015,** *137*, 1158-1163.

162. Jozwiakowski, M. J.; Nguyen, N.-A. T.; Sisco, J. M.; Spancake, C. W., Solubility Behavior of Lamivudine Crystal Forms in Recrystallization Solvents. *J Pharm Sci* **1996,** *85* (2), 193-199.

163. Mazivila, S. J.; Castro, R. A. E.; Leitão, J. M. M.; Esteves da Silva, J. C. G., At-line green synthesis monitoring of new pharmaceutical co-crystals lamivudine:theophylline polymorph I and II, quantification of polymorph I among its APIs using FT-IR spectroscopy and MCR-ALS. *Journal of pharmaceutical and biomedical analysis*. **2019,** *169*, 235-244.

164. Bhairab Nath Roy, G. P. S., Dhananjai Shrivastava, Harishchandra Sambhaji Jadhav, Umesh Parashram Aher, Sharad Chandrabhan Deokar An improved process for the manufacture of lamivudine. **2011**.

165. Maldonado, D. A.; Vassileiou, A.; Johnston, B.; Florence, A. J.; Brown, C. J., Data mining crystallization kinetics. *Digital Discovery* **2022,** *1* (5), 621-635.

166. Bandi Parthasaradhi Reddy, M. S. R., Pothireddy Venkateswar Reddy, Muppidi Vanaja Kumari Solid oral dosage forms of lamivudine (U.S. Patent No. US8481554B2). **2009**.

167. Helmli, F. S.; Scherer, S., Adaptive shape from focus with an error estimation in light microscopy. *ISPA 2001 : proceedings of the 2nd International Symposium on Image and Signal Processing and Analysis : in conjunction with 23nd [sic] Int'l*

*Conference on Information Technology Interfaces : Pula, Croatia, June 19-21, 2001* **2001**, 405.

168.    Doerr, F. J. S.;  Brown, C. J.; Florence, A. J., Direct Image Feature Extraction and Multivariate Analysis for Crystallization Process Characterization. *Crystal Growth & Design* **2022,** *22* (4), 2105-2116.

169.    Han, T.;  Tang, Y.;  Yang, X.;  Lin, Z.;  Zou, B.; Feng, H., Change Detection for Heterogeneous Remote Sensing Images with Improved Training of Hierarchical Extreme Learning Machine (HELM). *Remote Sensing* **2021,** *13* (23), 4918.

170.    He, K.;  Gkioxari, G.;  Dollar, P.; Girshick, R., Mask R-CNN. **2018**.

171.    Boyle, C.;  Brown, C.;  Sefcik, J.; Cardona, J. In *Improved Particle Characterisation from in-Line PAT: Comparison of Deep Learning and White-Box Methods*, American Institute of Chemical Engineers: PAT and Process Monitoring in Crystallization Development and Manufacturing, **2022**.

172.    Chen, H.;  Chiang, R. H.; Storey, V. C., Business intelligence and analytics: From big data to big impact. *MIS quarterly* **2012**, 1165-1188.

173.    Xiouras, C.;  Cameli, F.;  Quilló, G. L.;  Kavousanakis, M. E.;  Vlachos, D. G.; Stefanidis, G. D., Applications of Artificial Intelligence and Machine Learning Algorithms to Crystallization. *Chemical Reviews* **2022,** *122* (15), 13006-13042.

174.    Hamer William Edward, P. G. V. Aspirin crystallisation (Patent No. US2890240A). **1957**.

175.    Maia, G. D.; Giulietti, M., Solubility of Acetylsalicylic Acid in Ethanol, Acetone, Propylene Glycol, and 2-Propanol. *Journal of Chemical & Engineering Data* **2008,** *53* (1), 256-258.

176.    Ulrich, J.; Stelzer, T., Crystallization. *Kirk-Othmer Encyclopedia of Chemical Technology* **2000**, 1-63.

177.    Champion, J. A.;  Katare, Y. K.; Mitragotri, S., Particle shape: a new design parameter for micro-and nanoscale drug delivery carriers. *Journal of controlled release* **2007,** *121* (1-2), 3-9.

178.    Denk Jr, E. G.; Botsaris, G. D., Fundamental studies in secondary nucleation from solution. *Journal of Crystal Growth* **1972,** *13*, 493-499.

179.    Peña, R.;  Jarmer, D. J.;  Burcham, C. L.; Nagy, Z. K., Further Understanding of Agglomeration Mechanisms in Spherical Crystallization Systems: Benzoic Acid Case Study. *Crystal Growth & Design* **2019,** *19* (3), 1668-1679.

180. Lewis, A.; Seckler, M.; Kramer, H.; van Rosmalen, G., Agglomeration. **2015**, 130-150.

181. Marton, F.; Fensham, P.; Chaiklin, S., A Nobel's eye view of scientific intuition: discussions with the Nobel prize-winners in physics, chemistry and medicine (1970-86). *International Journal of Science Education* **1994,** *16* (4), 457-473.

182. Stennikov, V. A.; Barakhtenko, E. A.; Sokolov, D. V., A methodological approach to the determination of optimal parameters of district heating systems with several heat sources. *Energy* **2019,** *185*, 350-360.

183. Kumar, V.; Bhalla, A.; Rathore, A. S., Design of experiments applications in bioprocessing: Concepts and approach. *Biotechnology progress.* **2014,** *30* (1), 86-99.

184. Mandenius, C.-F.; Brundin, A., Bioprocess optimization using design-of-experiments methodology. *Biotechnology progress.* **2008,** *24* (6), 1191-1203.

185. Li, W.; Rasmussen, H. T., Strategy for developing and optimizing liquid chromatography methods in pharmaceutical development using computer-assisted screening and Plackett–Burman experimental design. *J Chromatogr A* **2003,** *1016* (2), 165-180.

186. Panigrahi, K. C.; Jena, J.; Jena, G. K.; Patra, C. N.; Rao, M. E. B., QBD-based systematic development of Bosentan SNEDDS: Formulation, characterization and pharmacokinetic assessment. *Journal of drug delivery science and technology.* **2018,** *47*, 31-42.

187. Karageorgou, E.; Samanidou, V., Youden test application in robustness assays during method validation. *Journal of Chromatography A* **2014,** *1353*, 131-139.

188. Garnett, R., *Bayesian optimization*. Cambridge University Press: 2023.

189. Parrott, A. J.; Bourne, R. A.; Akien, G. R.; Irvine, D. J.; Poliakoff, M., Self-Optimizing Continuous Reactions in Supercritical Carbon Dioxide. *Angewandte Chemie International Edition* **2011,** *50* (16), 3788-3792.

190. https://www.zinsserna.com/crissy.htm Zinsser Analytics

191. Boyle, C.; Brown, C.; Sefcik, J.; Uyttersprot, J.-S.; Calderon de Anda, J.; Burcham, C. L.; Nazermifard, N.; Svovoda, V.; Cardona, J., Evaluation of Deep-Learning-based Image Analysis for *in-situ* Pharmaceutical Particle Characterisation (manuscript in progress).

192. https://www.bruker.com/content/bruker/int/en/products-and-solutions/diffractometers-and-x-ray-microscopes/x-ray-diffractometers/diffrac-suite-software/diffrac-eva.html Bruker DIFFRAC.EVA: software to evaluate X-ray diffraction data Version 5.2

193. https://matplotlib.org/ MatplotLib Development Team

194. Lotov, A. V.; Miettinen, K., Visualizing the Pareto Frontier. In *Multiobjective Optimization*, Springer Berlin Heidelberg: **2008**; pp 213-243.

195. https://www.mt.com/gb/en/home/applications/L1_AutoChem_Applications/L2_Crystallization/Supersaturation_Application.html Mettler Toledo

196. Fan, S.; Gu, X.; Zhou, X.; Duan, X.; Li, H., Determination of nucleation kinetics from the induction time of 1,1-diamino-2,2-dinitroethylene (FOX-7) in DMSO/Water. *Energetic materials frontiers* **2021,** *2*, 62-68.

197. J. McGinty, N. Yazdanpanah, C. Price, J. H. ter Horst, and J. Sefcik, in The Handbook of Continuous Crystallization, ed. N. Yazdanpanah and Z. K. Nagy, The Royal Society of Chemistry, **2020**, pp. 1-50..

198. Salvy, B.; Zimmermann, P., GFUN. *ACM Transactions on Mathematical Software* **1994,** *20* (2), 163-177.

199. Blank, J.; Deb, K., Pymoo: Multi-Objective Optimization in Python **2020,** *8*, 89497-89509.

200. http://github.com/SheffieldML/GPyOpt GPyOpt GPyOpt: A Bayesian Optimization framework in python

201. Franceschini, G.; Macchietto, S., Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science* **2008,** *63* (19), 4846-4872.

202. Baake, E.; Baake, M.; Bock, H. G.; Briggs, K. M., Fitting ordinary differential equations to chaotic data. *Physical Review A* **1992,** *45* (8), 5524-5529.

203. Bernardo, J. M., Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **1979,** *41* (2), 113-128.

204. Biedermann, A.; Taroni, F., Bayesian networks and probabilistic reasoning about scientific evidence when there is a lack of data. *Forensic science international* **2006,** *157* (2-3), 163-167.

205. Ottoboni, S.; Wareham, B.; Vassileiou, A.; Robertson, M.; Brown, C. J.; Johnston, B.; Price, C. J., A Novel Integrated Workflow for Isolation Solvent Selection Using Prediction and Modeling. *Organic Process Research & Development* **2021,** *25* (5), 1143-1159.

206. Ekinci, D.; Şentürk, M.; Küfrevioğlu, Ö. İ., Salicylic acid derivatives: synthesis, features and usage as therapeutic tools. *Expert opinion on therapeutic patents* **2011,** *21* (12), 1831-1841.

207. Del Olmo, A.; Calzada, J.; Nuñez, M., Benzoic acid and its derivatives as naturally occurring compounds in foods and as additives: Uses, exposure, and controversy. *Critical Reviews in Food Science and Nutrition* **2017,** *57* (14), 3084-3103.

208. Gershoff, S. N., Vitamin C (ascorbic acid): new roles, new requirements? *Nutrition reviews* **1993,** *51* (11), 313-326.

209. Zhao, Y.-y.; Xie, R.-m.; Chao, X.; Zhang, Y.; Lin, R.-c.; Sun, W.-j., Bioactivity-directed isolation, identification of diuretic compounds from Polyporus umbellatus. *Journal of ethnopharmacology* **2009,** *126* (1), 184-187.

210. Bushra, R.; Aslam, N., An overview of clinical pharmacology of Ibuprofen. *Oman medical journal* **2010,** *25* (3), 155.

211. https://pubchem.ncbi.nlm.nih.gov/ Pubchem

212. https://www.sartorius.com/en/products/weighing/laboratory-balances/quintix-secura Sartorius

213. Stovall, D. M.; Givens, C.; Keown, S.; Hoover, K. R.; Rodriguez, E.; Acree, W. E.; Abraham, M. H., Solubility of crystalline nonelectrolyte solutes in organic solvents: Mathematical correlation of ibuprofen solubilities with the Abraham solvation parameter model. *Physics and Chemistry of Liquids* **2005,** *43* (3), 261-268.

214. Yalkowskyx, S.; Valvani, S.; Roseman, T., Solubility and partitioning VI: octanol solubility and octanol–water partition coefficients. *Journal of pharmaceutical sciences* **1983,** *72* (8), 866-870.

215. Bustamante, P.; Pena, M.; Barra, J., The modified extended Hansen method to determine partial solubility parameters of drugs containing a single hydrogen bonding group and their sodium derivatives: benzoic acid/Na and ibuprofen/Na. *International journal of pharmaceutics* **2000,** *194* (1), 117-124.

216.    Shalmashi, A.; Eliassi, A., Solubility of (+)-Ascorbic Acid in Water, Ethanol, Methanol, Propan-2-ol, Acetone, Acetonitrile, Ethyl Acetate, and Tetrahydrofuran from (293 to 323) K. *Journal of Chemical & Engineering Data* **2008,** *53* (6), 1332-1334.

217.    Barra, J.;  Peña, M.-A.; Bustamante, P., Proposition of group molar constants for sodium to calculate the partial solubility parameters of sodium salts using the van Krevelen group contribution method. *European journal of pharmaceutical sciences* **2000,** *10* (2), 153-161.

218.    Jouyban, A., *Handbook of solubility data for pharmaceuticals*. CRC press: **2009**.

219.    De Fina, K. M.;  Sharp, T. L.;  Roy, L. E.; Acree, W. E., Solubility of 2-hydroxybenzoic acid in select organic solvents at 298.15 K. *Journal of Chemical & Engineering Data* **1999,** *44* (6), 1262-1264.

220.    Matsuda, H.; Kaburagi, K.; Matsumoto, S.; Kurihara, K.; Tochigi, K.; Tomono, K., Solubilities of salicylic acid in pure solvents and binary mixtures containing cosolvent. *Journal of Chemical & Engineering Data* **2009,** *54* (2), 480-484.

221.    Nordström, F. L.; Rasmuson, Å. C., Solubility and melting properties of salicylic acid. *Journal of Chemical & Engineering Data* **2006,** *51* (5), 1668-1671.

222.    Shalmashi, A.; Eliassi, A., Solubility of salicylic acid in water, ethanol, carbon tetrachloride, ethyl acetate, and xylene. *Journal of Chemical & Engineering Data* **2008,** *53* (1), 199-200.

223.    Beerbower, A.;  Wu, P.; Martin, A., Expanded solubility parameter approach I: naphthalene and benzoic acid in individual solvents. *Journal of pharmaceutical sciences* **1984,** *73* (2), 179-188.

224.    Qingzhu, J.;  Peisheng, M.;  Shouzhi, Y.;  Qiang, W.;  Chang, W.; Guiju, L., Solubilities of benzoic acid, p-methylbenzoic acid, m-methylbenzoic acid, o-methylbenzoic acid, p-hydroxybenzoic acid, and o-nitrobenzoic acid in 1-octanol. *Journal of Chemical & Engineering Data* **2008,** *53* (6), 1278-1282.

225.    Li, D.-Q.;  Liu, D.-Z.; Wang, F.-A., Solubilities of terephthalaldehydic, p-toluic, benzoic, terephthalic, and isophthalic acids in N-methyl-2-pyrrolidone from 295.65 K to 371.35 K. *Journal of Chemical & Engineering Data* **2001,** *46* (1), 172-173.

226.    Perlovich, G. L.; Bauer-Brandl, A., Thermodynamics of solutions I: Benzoic acid and acetylsalicylic acid as models for drug substances and the prediction of solubility. *Pharmaceutical research* **2003,** *20*, 471-478.

227.    Restaino, F.; Martin, A., Solubility of benzoic acid and related compounds in a series of n-alkanols. *Journal of Pharmaceutical Sciences* **1964,** *53* (6), 636-639.

228.    Tully, G.; Hou, G.; Glennon, B., Solubility of benzoic acid and aspirin in pure solvents using focused beam reflective measurement. *Journal of Chemical & Engineering Data* **2016,** *61* (1), 594-601.

229.    Findlay, A., CXIII.—The solubility of mannitol, picric acid, and anthracene. *Journal of the Chemical Society, Transactions* **1902,** *81*, 1217-1221.

230.    Watson, O. L.; Jonuzaj, S.; Mcginty, J.; Sefcik, J.; Galindo, A.; Jackson, G.; Adjiman, C. S., Computer Aided Design of Solvent Blends for Hybrid Cooling and Antisolvent Crystallization of Active Pharmaceutical Ingredients. *Organic Process Research & Development* **2021,** *25* (5), 1123-1142.

231. https://www.pfizer.com/sites/default/files/investors/presentations/MackayNeuro_112906_part2.pdf Pfizer R&D

232.    Murtagh, L.; Dunne, C.; Gabellone, G.; Panesar, N. J.; Field, S.; Reeder, L. M.; Saenz, J.; Smith, G. P.; Kissick, K.; Martinez, C.; Van Alsten, J. G.; Evans, M. C.; Franklin, L. C.; Nanninga, T.; Wong, J., Chemical Development of an α2δ Ligand, (3S,5R)-3-(Aminomethyl)-5-methyloctanoic Acid. *Organic Process Research & Development* **2011,** *15* (6), 1315-1327.

233.    https://www.bruker.com/en/products-and-solutions/diffractometers-and-x-ray-microscopes/x-ray-diffractometers/d8-endeavor.html Bruker

234.    https://www.malvernpanalytical.com/en/support/product-support/morphologi-range/morphologi-g3 Morphologi G3 (Malvern Panalytical)

235. https://www.mt.com/gb/en/home/products/Laboratory_Analytics_Browse/TA_Family_Browse/ta-instruments/thermal-analysis-system-DSC-3-plus.html Thermal Analysis System DSC 3+ (Mettler Toledo)

236.    https://www.microscope.healthcare.nikon.com/products/polarizing-microscopes/eclipse-ci-pol Eclipse Ci POL (Nikon)

237.    https://micropix.co.uk/ Micropix

238.    Klamt, A.;  Jonas, V.;  Bürger, T.; Lohrenz, J. C. W., Refinement and Parametrization of COSMO-RS. *The Journal of Physical Chemistry A* **1998,** *102* (26), 5074-5085.

239.    https://www.scale-up.com/dynochem Scale-Up Systems

240.    Gracin, S.;  Brinck, T.; Rasmuson, Å. C., Prediction of Solubility of Solid Organic Compounds in Solvents by UNIFAC. *Industrial & Engineering Chemistry Research* **2002,** *41* (20), 5114-5124.

241.    Pfizer R&D, *Pd-299685 API Co-Development Plan (Internal)*; **2006**.

242.    Park, J.-S., Optimal Latin-hypercube designs for computer experiments. *Journal of statistical planning and inference* **1994,** *39* (1), 95-111.