Automated, computational approaches to kinetic model and parameter determination

Connor Jack Taylor

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds

School of Chemical and Process Engineering

December 2020

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapters 2 - 3 are based on work from a jointly authored publication:

- **Connor J. Taylor**, Megan Booth, Jamie A. Manson, Mark J. Willis, Graeme Clemens, Brian A. Taylor, Thomas W. Chamberlain, and Richard A. Bourne. "*Rapid, Automated Determination of Reaction Models and Kinetic Parameters.*" Chemical Engineering Journal (2020): 127017.

The work contained within this publication that is directly attributable to myself is all experimentation, coding of the relevant methodology and preparation of the manuscript. Megan Booth assisted in the setting up and running of some experimentation (notably the phenyl acetate and paracetamol case studies). Jamie A. Manson assisted in the preparation of the manuscript (notably the reactor engineering section). Mark J. Willis assisted in the direction of the coding of the relevant methodology. Graeme Clemens assisted in the setting up of some experimentation (notably the metoprolol case study) and general supervision. Brian A. Taylor assisted in the direction of the approach and general supervision. Thomas W. Chamberlain and Richard A. Bourne both assisted in the direction of the approach, general supervision and preparation of the manuscript.

Acknowledgements

This project could not have been completed without the ongoing help of several individuals who have helped guide the project into the finished thesis and publications. Firstly, I'd like to thank Dr Richard Bourne for his continued support, advice and brilliance in assisting with the project wherever possible, in both running experiments and preparing project directions. Your supervision style of treating a PhD student as an academic partner is well appreciated and has undoubtedly led to an excellent postgraduate experience.

Dr Thomas Chamberlain, thank you also for your ongoing support and contributions to my supervision. Although you're busy, it does not go unnoticed that you are always willing to go the extra mile in proofreading posters/papers and even just transfer report chapters - your sharp wit means that you can sell my science better than anyone can, including myself.

To Brian and Graeme, thank you also for your support in the direction of my project and sourcing opportunities for me to present and develop myself as a researcher, I really appreciate it. Thank you also to Mark Willis, Ian Ashworth, Rob Cox and the rest of the CPUT team at AZ for your invaluable thoughts and discussions.

To Adam, Jamie and Calum, thank you for all of your nonsense talks and debates particularly in the control room that helped keep me sane - even the consistent whinging was funny. More recently, Ollie and Holly have been helping with that also. Ricardo, thank you for all of your advice, as well as your nonsense talking too (but only after you shut off your 'work persona' after 6pm). Brendan, thank you for being a consistent source of hyperactive entertainment, don't let Luke try to calm your ideas! Rosie and Sam, thank you for all of the tea breaks at AZ that also managed to keep us all sane. Mary, thank you for your help in running the lab, and similarly thank you to everyone in the iPRD for your help and support over the years: John, Bao, Nando, Ilias, Claudio, Micaela, Nisha, Alison, Christiane, James, Pia and the ones that have moved on: Alastair, Will, Mike, Becky, Chris Horb, Chris Hone, Alvaro and not forgetting Fanfu!

To Rhiann and my encouraging parents, thank you for always being there and motivating me to be my best and inspiring me to go further.

Abstract

A major bottleneck in the transition from chemistry research at lab scale to process development is a lack of quantitative chemical synthesis information. Critical aspects of this information include knowing the correct reaction model and precise kinetic parameters. If this information is available, classical reaction engineering principles may be utilised to shorten process development times and lower costs.

Identifying the correct reaction model for a particular process, however, can be challenging and time-consuming, particularly for physical-organic chemists and kinetics experts that may be busy with other aspects of process development. The work presented herein describes computational approaches that automatically determine the most likely kinetic model and associated parameters based on the experimental data supplied, without expert chemical intuition.

The concept for these methodologies involves a comprehensive model evaluation tool. The experimental data and the species involved in the process are inputted. Based on mass balance, all mass-balance-allowed transformations between these species are identified. All possible models are then compiled from this list of transformations, featuring unique combinations of these model terms. Every model is then evaluated using ordinary differential equation (ODE) solvers and optimisation algorithms to maximise the convergence of simulated reaction progression with the experimental data, thereby identifying the kinetic parameters. Each model is then statistically evaluated to determine which model is the most likely to be correct.

Using these methodologies allows any chemist to automatically determine a reaction model and kinetic constants for a particular system, by performing all kinetic analysis autonomously. Their most expensive resource, time, can then be focussed on other tasks that cannot be automated.

Table of Contents

List of Schemes

Chapter 1

Chapter 2

List of Figures

Chapter 1

Chapter 4

Chapter 5

List of Tables

List of Abbreviations

A	Pre-exponential factor
Ac	Acetyl
AIC	Akaike's information criterion
AIC_C	Corrected Akaike's information criterion
API	Active pharmaceutical ingredient
Aq	Aqueous
Bn	Benzyl
bp	Boiling point
Bu	Butyl
cm	Centimetre
CSTR	Continuous stirred tank reactor
CRN	Chemical reaction network
DCM	Dichloromethane
E_a	Activation energy
Equiv	Equivalents
Et	Ethyl
g	Gram
GA	Genetic algorithm
GC	Gas chromatography
h	Hour
HPLC	High performance liquid chromatography
Hz	Hertz
ID	Inner diameter
ILP	Integer linear programming
IR	Infrared
k	Rate constant
K	Kelvin
k_{ref}	Reference rate constant

L	Litre
m	Metre
M	Molar
Me	Methyl
MILP	Mixed integer linear programming
min	Minute
mL	Millilitre
MM	Mass matrix
mmol	Millimole
mol	Mole
MS	Mass spectrometry
NMR	Nuclear magnetic resonance
OD	Outer diameter
ODE	Ordinary differential equation
PFR	Plug flow reactor
Ph	Phenyl
R	Ideal gas constant
Re	Reynolds number
S	Stoichiometry
S_NAr	Nucleophilic aromatic substitution
SSE	Sum of squared error
T	Temperature
TFA	Target factor analysis
t_R	Residence time
T_{ref}	Reference temperature
t_{Res}	Residence time
V	Volume

Chapter 1 : Introduction

The main aim of this project is to develop and advance methodologies for kinetic model determination and parameter estimation. The acquisition of the experimental data necessary for these studies is flexible and can be obtained by any means. However, given the numerous advantages of continuous flow chemistry and the availability of a flow chemistry platform, many chemistries were performed using this regime. For this reason, continuous flow chemistry is broadly covered in this introduction, as well as a comprehensive look at methods of obtaining kinetic data in Chapter 3.1.

Current methodologies of obtaining kinetic information are herein broadly covered and critically analysed. However, in-depth discussions about the relevant literature to the kinetic methodologies developed during this project are highlighted in Chapter 2.1. Furthermore, as these kinetic methodologies feature significant use of optimisation algorithms, discussions regarding optimisation problems and common algorithms are also found in this introductory chapter. This coverage depicts the general relevancy of different types of algorithm to particular circumstances, with specific kinetic fitting algorithm discussions featured in Chapter 2.2.2.2.

1.1 Continuous flow chemistry

Continuous flow chemistry is rapidly becoming the optimum way to conduct a wide variety of different chemical reactions. This section introduces the concept of flow chemistry, featuring a discussion of the advantages and disadvantages, as well as when to adopt this methodology. This broad introduction also covers different reports from the literature pertaining to different classes of reaction in flow, as well as how these reactions can be analysed and optimised; either on-line or off-line.

1.1.1 Batch and flow reactions

To discuss flow chemistry, it is first important to understand the differences between this methodology and traditional batch chemistry. Batch reactions are still the ‘normal’ bench practice today after being standardised 200 years ago in the time of Friedrich Wöhler, as there are still advantages to

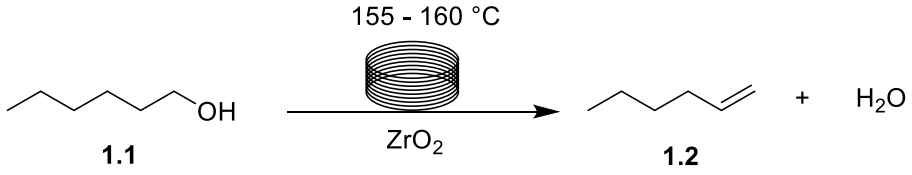
conducting synthesis this way. This batch reaction approach to synthesis, whereby the chemistry is conducted essentially in a vessel, is historically the most fruitful and rewarding approach in terms of discoveries. Suzuki couplings,[1] Sonogashira couplings[2] and the Wittig reaction[3] among hundreds of other ground breaking mechanisms were all discovered via batch reactions. A flow chemistry regime differs in that there is no bulk reaction medium in a vessel. By using pumps, tubing and differing flow reactors, small volumes of reaction medium are constantly flowing and reacting to provide a continuous outlet stream of reacted material. Although this flow methodology has been known for a relatively long time, the historic uptake in synthetic laboratories has been low.[4] However, as flow technology improves and becomes easier to adopt in the laboratory,[5] it is increasingly important to evaluate whether a particular chemical process would be most efficient in batch or flow regimes, in terms of product output or otherwise.

For example when considering exploratory synthesis, in contrast to batch; using flow chemistry, consisting of tubing and microreactors, to attempt to discover new syntheses by exploring discrete options is not typically a good application.[6, 7] This is because the main strengths of flow processes come mainly from already existing reaction pathways where optimising product output is the focus, and to construct a flow system to conduct a purely exploratory reaction is a traditionally inefficient use of time whereas batch can be implemented much faster. In recent years, however, high throughput flow systems have been developed for automated discovery.[8, 9] Unfortunately though, as the equipment for this type of discovery is specialist and can be expensive, it can still be argued that simple exploratory benchtop batch reactions are still much more efficient in terms of time and cost.

There are other general instances where batch is often still favourable for chemical transformations, for example when precipitation drives reaction completion.[7] In batch this isn’t an issue, however when using a flow setup this will frequently result in channel, mixer or pressure regulator clogging - this is also true when using high viscous liquids.[10] As a general rule, precipitation, suspensions and other instances where solids are used can lead to complications in the process - this can be mitigated through reaction dilution but may then suffer from low productivity. The introduction, however, of standardised

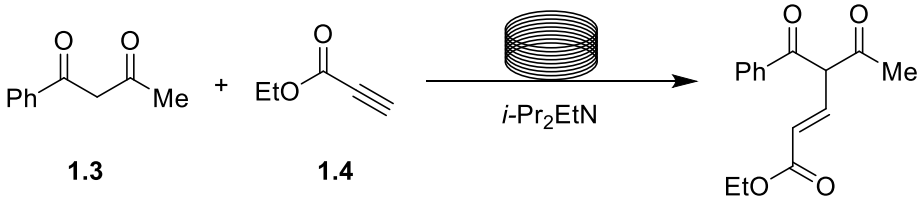
laboratory equipment (such as miniaturised CSTRs: Freactors) for continuous flow can then facilitate such multiphase processes to allow maintained efficiency.[11, 12]

Despite some of these limitations, there are significant benefits to continuous flow processes which have become apparent alongside the tremendous practical advances in recent years; many of which have been utilised in actually transferring batch processes into flow, as more and more batch procedures are superseded.[13-15] Reduction of reaction times[16, 17] and increases in conversion and selectivity may be observed by transferring to flow processes.[18] This may be for a number of reasons, but is likely due to the increased mass transfer that can be expected from flow systems, as well as the ability to increase the temperature of the reaction medium further past the boiling point of the solvent, which is discussed further in Chapter 1.1.2.1. This will lead to an overall increase in both the reaction rate and productivity. Two examples are shown herein where increased conversion was observed. Scheme 1.1 reports the synthesis of hexene from hexanol in flow by Wilson and McCreedy,[19] with a conversion of around 90 %, whereas in conventional batch reactors this reaction conversion does not exceed 60 %. Most notably, there was no degradation of the performance after three days of continuous use, which may be expected from a fixed bed catalyst. Scheme 1.2 shows another direct comparison of conversion in flow and batch by Wiles *et al.*,[20] where the Michael addition is completed in 20 minutes with 100 % conversion in flow, compared to the reaction completion in 24 hours with 89 % conversion in batch.



1.1 **1.2**

Scheme 1.1: The conversion of hexanol, **1.1**, to hexene, **1.2**, in a continuous flow setup.[19]



1.3 **1.4** **1.5**

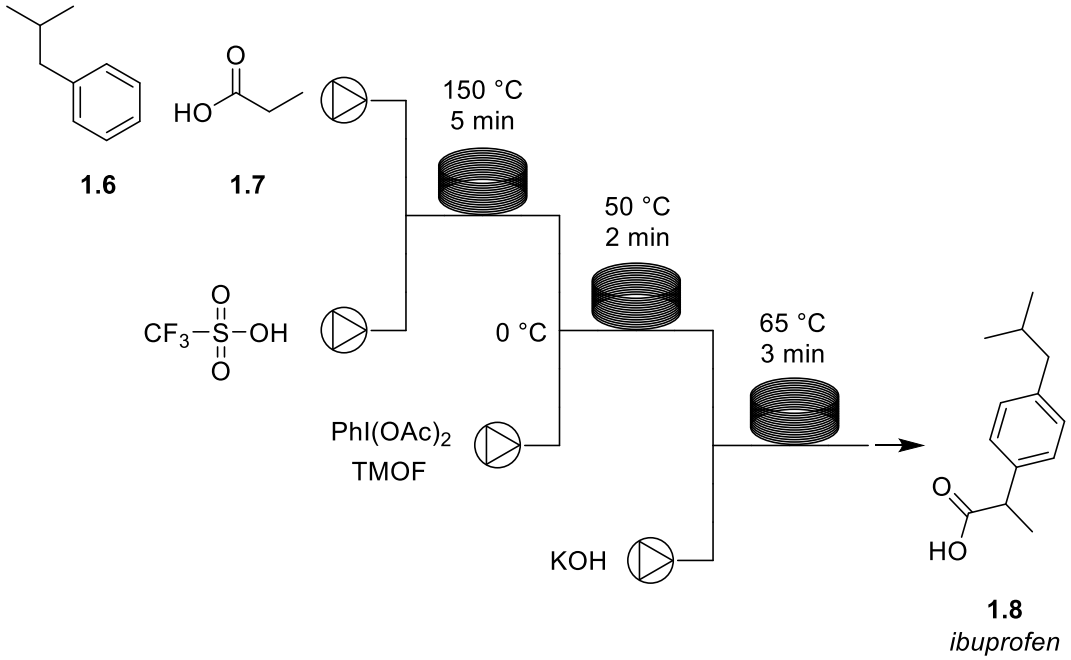
Scheme 1.2: The flow synthesis of the adduct, **1.5**, via a Michael addition between the dicarbonyl, **1.3**, and the alkyne, **1.4**.[20]

A main advantage to using flow chemistry setups to perform reactions arises from the equipment featured within the setups themselves - namely the pumps. Pump flow rates can be set to very precise flows, which can very accurately deliver particular ratios of reagents at given points along the flow setup.[21-23] This ensures that the correct stoichiometry of the reagents is present in the reaction, and these exact stoichiometries can be delivered continuously and with minimal error.[21, 24] This allows for optimisation and kinetics studies with precise reagent additions that would otherwise be troublesome for an experimenter to consistently and dependably deliver. These same flow rates also ensure that the residence time is exact and consistent, meaning sampling points are reliably accurate in sampling the reaction medium after a specific reaction time.[25, 26] This point, however, is highly dependent on the mixing regime present, which is discussed further in Chapter 3.1.

Another noteworthy advantage to flow processes is that a large number of reactions can be executed, and hence a large amount of data can be obtained, using only minimal quantities of a reagent in flow. This is because much less reagent is required for reaction and subsequent on-line analysis in a microreactor compared to bench scale batch setups. This is a particularly desirable attribute of a system if you have an expensive material, or a material that is synthesised by a time consuming process, where you need to be conservative with its use but still obtain lots of experimental data.[24]

One example of a reaction that has been successfully transferred from batch to flow is a multi-step synthesis of ibuprofen, **1.8**, conducted by Bogdan *et al*.[27] This synthesis demonstrates that even flow processes with multiple steps can be achieved, without intermediate purification steps, in a streamlined and efficient manner. This is provided as long as care in the reaction chronology has been undertaken ensuring excess reagents and byproducts from previous steps are compatible with downstream processes, shown in Scheme 1.3. This efficiency is important when creating complex molecules because although the stepwise batch reactor processes are effective, they are also very wasteful, particularly when they involve consecutive work-up procedures. An example of this is in the pharmaceutical industry where, in general, 25 - 100 kg of waste is

produced for every 1 kg of complex molecule.[28] However, to facilitate reaction compatibility, expensive or atom-inefficient transformations may be necessary, such as the use of phenyl-iodine(III) diacetate and TMOF to incur a chemical rearrangement in the reported ibuprofen synthesis below. This may require a careful balance to achieve green metrics in product manufacturing. Further discussion on reaction telescoping is provided in Chapter 1.1.2.6.



Scheme 1.3: The multistep continuous flow synthesis of ibuprofen, **1.8**.

In large scale batch reactor design used in industry the most significant problems to be overcome are the heating and agitation of the system, and it has even been reported that processes cannot be scaled up further due to temperature and mixing gradients.[29] This problem relates to the fact that it is difficult to effectively heat and agitate a large mass completely uniformly, which can lead to different reactivities present at different points within the vessel - this can result in decreased product yield due to an increased rate of side reactions, or more seriously, a runaway reaction when exothermic processes are involved.[30, 31]

By contrast, mesoscale flow systems (channel diameter: 1 - 10 mm) and in particular miniaturised flow reactors (channel diameter: 50 - 1000 µm) are excellent at efficiently heating and mixing reactants.[32] Rapid heating of reaction microchannels through the walls of a preheated reactor ensures the reaction mixture is thoroughly and uniformly heated and mixed, which is attributed to the

large surface area to volume ratios of the microchannels.[18, 33] This heating can be achieved by oil baths, conventional ovens or even microwave heating; all of which achieve the same goal which is precise temperature control, with increased safety.[34-36] The high molecular diffusion across the relatively narrow microchannels allows assumptions of perfect cross-sectional mixing in a 'plug flow' regime[37] (see Chapter 3.1) and mass transfer can further be improved through the use of static mixers[38] or CSTRs.[39]

This increased safety in flow reactors is attributed to only a small amount of reaction material present in the reactor at any one time, enclosed within a tubular vessel. Therefore, although the chemical exposure is the same during reservoir preparation, the experimenter is not exposed to any harmful material during experimentation, as long as all inlets and waste streams are properly contained. The is advantageous especially in processes where there are highly exothermic steps, alongside even the possibility of scale-up of formerly non-scalable reactions.[36, 40] This scalability can be achieved simply by running one flow reactor for a long period of time, or several identical flow reactors in parallel in a process called 'numbering' - any number of flow reactors can be 'numbered up' to achieve the required throughput.[41-43]

1.1.2 Reactions in flow

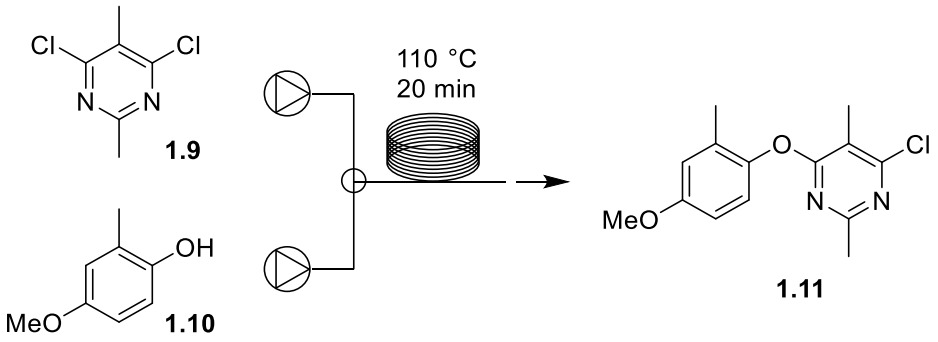
Alongside the aforementioned general advantages that flow processes have over batch, there are several distinct reaction niches which have specific facets whereby translation to flow has distinct benefits. There are many different areas where continuous processes have helped to transform research, as well as industrial processes that may branch from them. This literature review will focus only on the most common types of reaction in flow in which most reactions are likely to encompassed. Also refer to flow chemistry reviews on polymer chemistry[44], the use of extreme temperature[45] and slurry reactors.[46]

1.1.2.1 Temperature-accelerated reactions

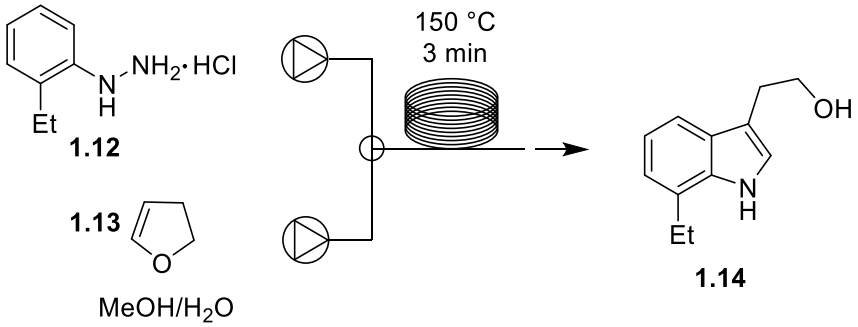
Temperature-accelerated reactions are the most basic and most common use of flow setups, whereby reactants are fed into a heated microreactor to perform a chemical transformation. The use of flow enables access to higher temperatures in a system with excellent safety, which means that reactions can be performed which would otherwise be too slow or unobtainable in other

systems. Temperature-accelerated reactions in a traditional batch setup are often limited by the laboratory equipment and the boiling point of the solvent, and having to switch to different, higher-boiling solvents can lead to complications in both the reaction and the purification.[47, 48] High-pressure batch vessels can be used to facilitate these higher temperature reactions (such as Parr reactors or autoclaves),[49] but additional equipment and safety considerations must be made in these cases.

However, in flow setups there is no need to give significant consideration to changing the solvent as the boiling of the solvent can be suppressed in most cases by the use of a back-pressure regulator.[50] Therefore, when conducting these reactions, where higher temperatures can be accessed in flow setups, the reaction rates are also generally higher.[14] This alongside the aforementioned increased associated safety in using flow setups, can make flow a desirable and convenient option for temperature-accelerated reactions. Scheme 1.4 and Scheme 1.5 show examples of temperature-accelerated reactions that have been conducted in flow, where in both cases the temperature of the reaction was conducted at a higher temperature than the boiling point of the solvent, leading to decreased reaction times. [51, 52] It is also worth noting that unless explicitly stated, a back pressure regulator will have been used in all schemes in this literature review if the temperature of the reaction exceeds the boiling point of the solvent used.



Scheme 1.4: The temperature-accelerated S_NAr reaction of **1.9** with **1.10** to form **1.11**, in THF (boiling point: 66 °C), under flow conditions.[51]



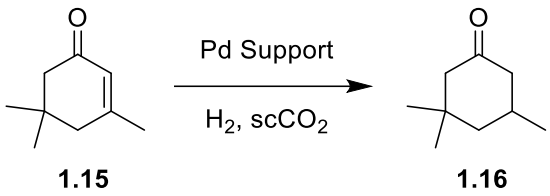
Scheme 1.5*:* The Fischer Indole synthesis of **1.14**, from the hydrazine, **1.12**, and dihydrofuran, **1.13**, in methanol (boiling point: 65 °C) using a flow setup.[52]

1.1.2.2 Supercritical fluid

The use of supercritical fluids (SCFs) as solvents in flow processes have attracted attention in the past 30 years mainly motivated by the prospect of replacing toxic industrial solvents.[53] For example, the environmental interest of using supercritical CO_2 or H_2O ushers in opportunities for greener chemical processes by replacing some of the more harmful solvent systems.[54-56] This is an admirable objective for process chemists and can lead to cleaner and more sustainable reactions.[57]

Whilst the green aspects of using SCFs are attractive, perhaps the greatest quality they possess are their physiochemical “hybrid” properties; an intermediary between liquids and gases whereby properties can be finely tuned by marginal variations in temperature and pressure. As a result, SCFs exhibit gas-like viscosities as well as no surface tension which are objectively advantageous in processes involving interface and surface chemistry.[58-60] Furthermore, the liquid-like densities allow for substantial dissolution of precursors within the SCF and diffusion coefficients are generally at least 100 times greater in SCFs than in liquids. All of these factors contribute towards processes whereby the reaction conversion is significantly improved as a direct effect of using SCFs.[53, 61-63] However, the compatibility of these supercritical solvents must still be tested with the process of interest (as with any solvent), as there may be solubility or reactivity issues.

Supercritical CO_2 ($scCO_2$) is a common solvent in the realm of SCFs because of its easily obtained critical parameters at bench scale; it is also extracted easily with high purity as a byproduct from many processes, as well as being non-toxic and non-flammable.[64] One of the most notable industrial scale applications using $scCO_2$ is the hydrogenation of isophorone, **1.15**, shown in Scheme 1.6 which obtains **1.16** in a sufficiently pure state without any downstream purification.[65] Another common process developed from the use of $scCO_2$, by taking advantage of the non-toxicity of the solvent, is the decaffeination of coffee beans, which has since become one of the most popular decaffeination methods.[66, 67]



Scheme 1.6*:* The hydrogenation of isophorone, **1.15**, to TMCH, **1.16**.[65]

One major problem with using $scCO_2$ in flow processes however is its incompatibility with primary and secondary amines due to carbamate formation, which can precipitate out of solution and therefore block any microchannels in the system.[68] Other supercritical solvents can be used in these cases such as $scNH_3$, although this presents a problem in itself that it cannot be used with aqueous solutions because of ammonium hydroxide formation - nonetheless, supercritical ethane and propane have been used as direct replacements for scCO2 due to similar critical parameters.[69]

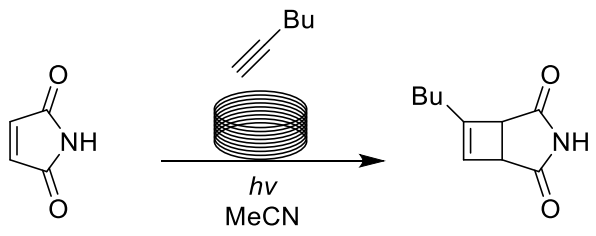
Using continuous reactors for SCFs have the advantage over batch reactors that they do not require depressurisation to add further material to the process or recover products. Also a predominant advantage to using continuous reactors is that reaction parameters such as temperature, flow rate and pressure can all be changed almost independently of each other. In context, this ability allows the properties of the SCF to be altered in real time when optimising reactions which cannot be done easily in batch reactors - this versatility of continuous reactors makes it the favourable choice when working with SCFs.[70]

1.1.2.3 Photochemical reactions

The use of photons at given wavelengths to provide sufficient energy to overcome activation energy barriers is the basis of photochemical reactions; complex molecular structures can be achieved via these unique reaction pathways that are otherwise unobtainable via thermochemical or electrochemical methods.[71, 72] Implementing photochemical reactions is also beneficial as a photon is a “green” reagent which is traceless and thus doesn’t require removal/recycling in the same way many other reagents do.

Photochemistry is not regarded only as a unique and novel pathway to molecules with exotic structures, however - one example of where photochemistry is used extensively in industry is in the Toray process to attain caprolactam, which is used to manufacture Nylon 6.[73] Photochemical reactions have also been found to benefit from continuous flow systems, where large-scale photochemical synthesis is significantly more effective when compared to their corresponding batch approaches.[74] This is due to problems regarding penetration into the bulk reaction medium, which is an important boundary that has to be overcome. Most of the photochemical synthesis occurs within a short radius of the lamp, meaning the reaction conversion is scale dependant. The small tubular diameters in flow systems take advantage of this by exposing the entire reaction medium to the light source as it passes through the reactor.[75]

One example where a photochemical reaction has been implemented into a flow process is reported in the continuous [2 + 2] photocycloaddition producing the cyclic product **1.18**, shown in Scheme 1.7. [74] In this reaction, the conversion is quoted at 83 % and because the reaction conversion is scale independent, the continuous process can be run indefinitely to produce over 500 g per 24 hour period. This process can then be scaled up further, as in any continuous flow process, by numbering up.



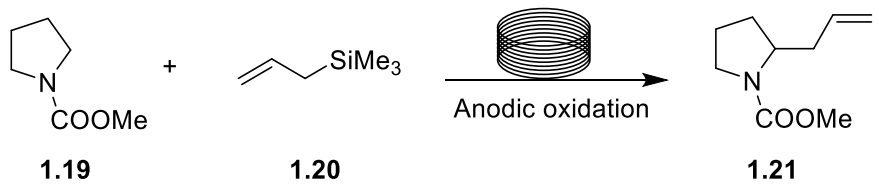
1.17 **1.18**

Scheme 1.7: The [2 + 2] photocycloaddition of malemide, **1.17**, with 1-hexyne to produce the cyclic product **1.18**.[74]

1.1.2.4 Electrochemical reactions

Electrochemistry is another clean and efficient reaction method in the realm of organic synthesis, where the electricity supplied to the reaction system induces the formation of the reactive intermediates from neutral substrates. This method of reaction has significant advantages to conventional batch processes, one of which being that harmful oxidising and reducing agents can be substituted using electrochemistry. By applying precisely controlled current between two electrodes, reactions can be achieved by using milder reagents. This method allows milder reagents as electrons can be added/removed without the need for chemical reducing/oxidising agents which could complicate the reaction, especially in flow systems where downstream processes can be highly affected by excess chemical reagents.[76] Also as the electrons are the “reagent” in these reactions, as well as being widely and readily available, electrochemical reaction pathways become cheaper and less labour-intensive as alternative, costly reagents don’t need to be used.[77]

One example of where electrochemical synthesis in flow has been observed is in the anodic substitution reaction by Horii *et al.*, [78] shown in scheme 1.8, with yields of up to 74%. In flow processes such as this, the laminar flow regime is used as an advantage - two inlets of separate flows join together in the reaction vessel, where their respective microchannel sidewalls are anodic and cathodic opposite to one another. This ensures the dominant oxidation of the substrate, **1.19**, to form a cationic intermediate whilst the nucleophile, **1.20**, diffuses across the tubular diameter to react to afford **1.21**. [79] Laminar flow and other mixing regimes are discussed further in Chapter 3.1.

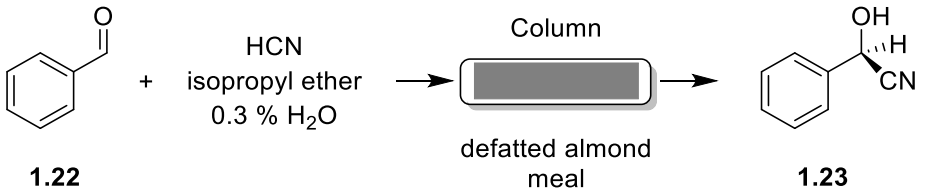


Scheme 1.8: An electrochemical facilitated nucleophilic substitution reaction to afford the product, **1.21**.[78]

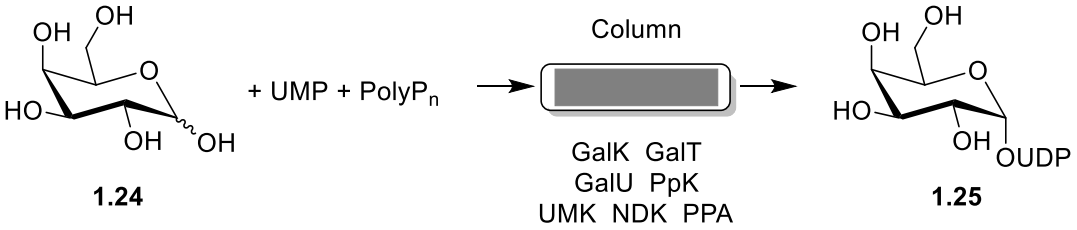
1.1.2.5 Enzymatic reactions

Biocatalytic processes are also an area in which research is continuously progressing in order to develop alternative process routes to the synthesis of fine chemicals. The main issues, however, arise from enzymes only being able to operate efficiently in narrow pH and temperature ranges with many organic substrates having poor solubility in water. However, implementation of enzymes whilst using non-aqueous solvents can dramatically improve their applicability.[80, 81] Within biochemical processes, this niche can be important in achieving reaction optimisation despite given limitations with using biocatalysts.

The use of immobilised active enzymes in the column of a flow system is very attractive and there are many reports in the literature of applied processes.[82, 83] An example of this is the continuous process of chiral cyanohydrin formation, shown in Scheme 1.9, where a column is packed with defatted almond meal.[84] There are also examples where multiple columns with different immobilised enzymes are used as a multistep continuous process in order to reach the desired product - one example is reported where 7 separate columns with differing enzymes achieve the synthesis of UDP-galactose from inexpensive starting materials.[85] This is shown in Scheme 1.10 as a simplified one column synthesis with multiple enzymes. This also shows how advantageous immobilised biocatalysis can be in multistep synthesis, as there are no catalysts or harmful reagents that can affect downstream processes.



Scheme 1.9: An enzyme catalysed continuous cyanohydrin formation forming **1.23**.[84]

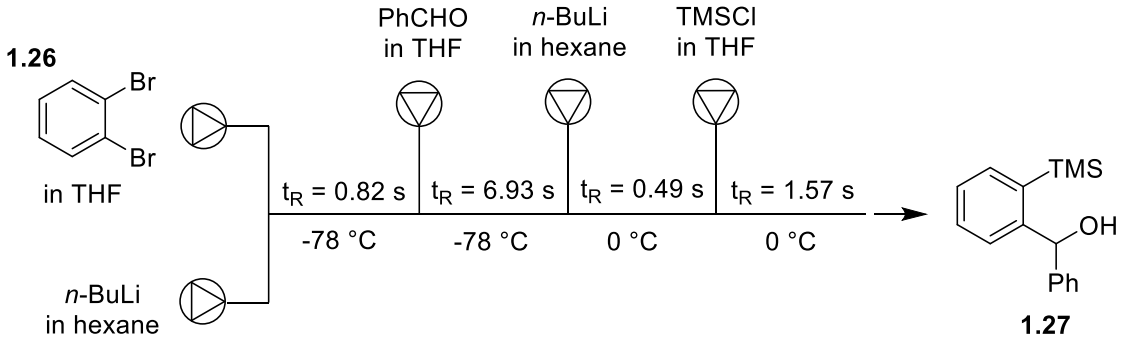


Scheme 1.10: Multiple enzymatic catalysed continuous process to form UDP galactose, **1.25**.[85] Individual enzymes stated.

1.1.2.6 Telescoped reactions

When considering a total synthesis for a pharmaceutical ingredient or otherwise, it would be ideal to limit the number of steps and hence the amount of waste, time and resources. This can be achieved in one way by reaction telescoping.[24] Reaction telescoping is essentially performing multiple chemical transformations to a molecule sequentially without the need for the purification of intermediates or removal of waste, for example the 'one-pot' synthesis of pyridine and quinoline derivatives in batch reported by Kobayashi *et al.*[86]

There are further advantages, however, to reaction telescoping in flow, that are not observed in batch. Using a flow system allows the introduction of reagents in exact stoichiometries at specific points in the reaction cycle simply by addition via inlets further down the line within the system. Pairing these additions with differing fixed-temperature reaction vessels also allow for isolated reaction optimisation for each step along the flow setup in a multi-step synthesis. An example of a multi-step reaction where the success of the synthesis is attributed to the effective residence time and temperature control of each step is shown in Scheme 1.11, in the synthesis of **1.27**.[87]



Scheme 1.11: The multi-step flow synthesis of **1.27**, from coupling o-dibromobenzene, **1.26**, with varying electrophiles via sequential halogen-lithium exchange reactions.[87]

The main drawback of reaction telescoping is that downstream transformations have to be compatible with side products and waste from upstream reactions, as any incompatibility leads to a lesser conversion of any intermediates. This serves as a hindrance to the other transformations down the line, hence leading to poor conversion for the overall process. Therefore when considering reaction telescoping, great care must be taken into account for reagent compatibility and reaction chronology.

A further advantage to using a flow system, however, is the opportunity of in-line purification which can solve reagent compatibility issues - this gives

greater accessibility to a variety of different reactions that can all be accomplished in a single process. This in-line purification is made possible by an in-line liquid-liquid separator, such as the separator reported by Zaiput Flow Technologies, that allows immiscible liquids to be separated into two streams, one organic and one aqueous, via filtration using a hydrophobic membrane.[88] This separator is shown in Figure 1.1. An example where a membrane separation has been used, shown in Scheme 1.12, is in an ibuprofen multi-step synthesis where Snead and Jamison improved on previous work aforementioned by Bogdan *et al*.[27, 89] This in-line separation avoided the need for triflic acid and increased the yield to 98 % by incorporating this aspect within the process.

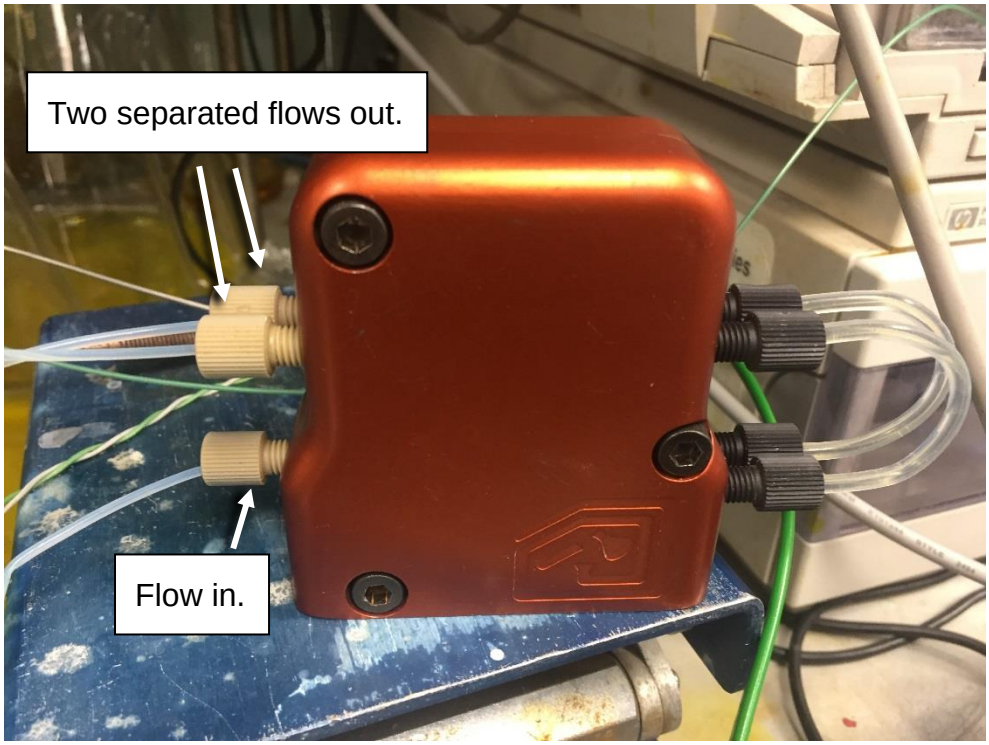
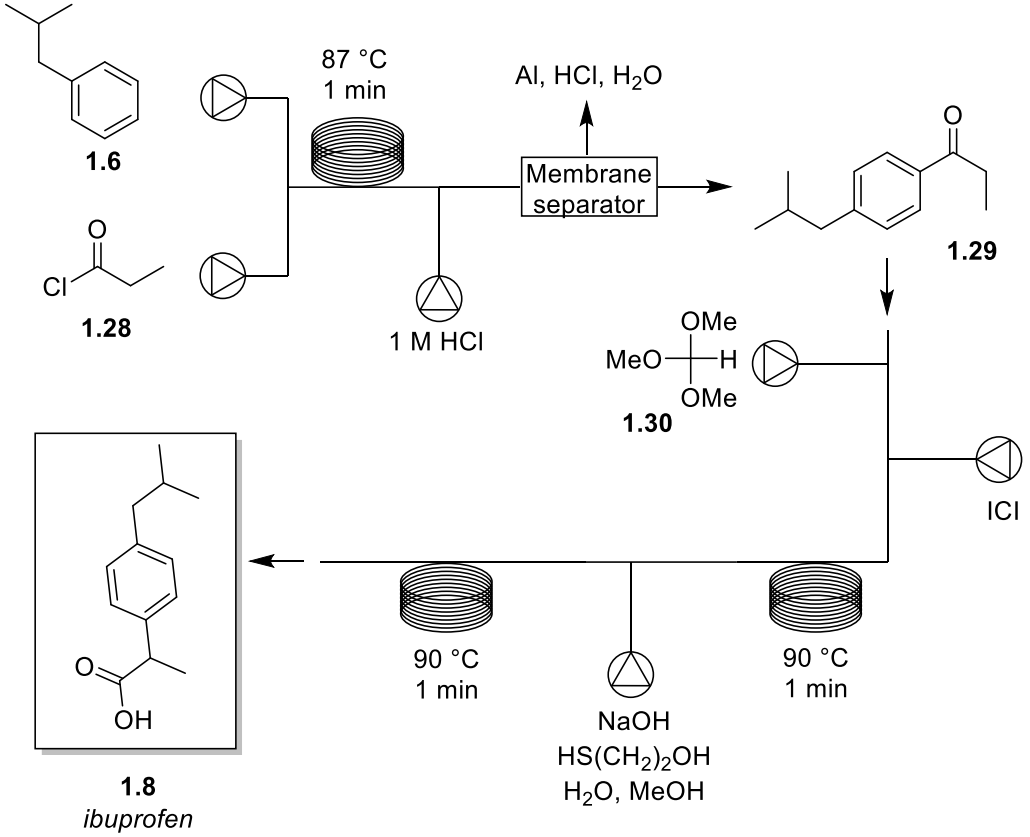


Figure 1.1: An annotated picture of an in-line liquid-liquid separator, developed by Zaiput Flow Technologies.



Scheme 1.12: A multi-step flow synthesis of ibuprofen, **1.8**, involving an in-line purification step.[89]

There are also several other methods of in-line purification that can be considered, alongside or instead of liquid-liquid separation. Gas-liquid separation[90], micro-distillation[91], solid-supported scavenging[50] and gravity-based separation[92] have all been reported in the literature, but have not be covered in detail as these methods are not well established.

1.1.3 Reaction analysis

Reaction analysis in benchtop batch processes are relatively time-consuming and labour-intensive, as samples must often be removed by hand from the reaction mixture to determine reagent conversion, product composition etc. Therefore the elimination of these practices and adoption of their respective flow processes has many potential cost benefits, particularly in fine chemical and pharmaceutical production.[93]

On the other hand, for these same methods of analysis, the sampling process can be fully automated by the use of sampling valves. A sample valve

is a device attached to the line of a flow system, where the reaction mixture passes through undisturbed as if it was an extension of the tubing; however there is a ‘switching’ mechanism whereby the flow is diverged from its path into a separate flow system, instantaneously, then the original flow is restored milliseconds later. This diverted flow then leads to an analytical system (HPLC, mass spectrometer etc.) and has much thinner tubing with a constant mobile phase flowing, ready to carry the injected sample to the analytical equipment. This sampling is termed ‘on-line’ as the flow is rerouted from the flow along the reactor. ‘In-line’ therefore refers to sampling without the need to divert reaction material and is typically non-destructive. In all cases, great care must be taken in calibrating the machine to chemistries of interest, as quantitative analysis can often only be performed by calibrating the response of the equipment to known standard concentrations of each chemical species.

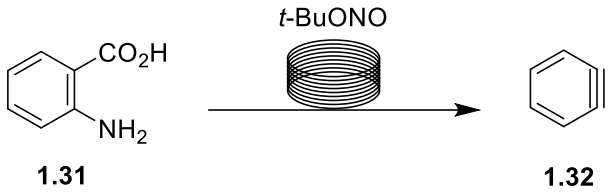
This section will report on reaction analysis in flow by the use of well-established and common techniques, such as mass spectrometry and high-performance liquid chromatography (HPLC), as well as fast developing and high interest areas such as NMR and IR. In each case, the system is assumed to be at ‘steady-state’ in which the flow output is consistent - this is explained in detail in Chapter 3.1. Also refer to specific papers covering on/in-line analysis by UV[94, 95], Raman[96], fluorescence[97], GC[98, 99], XAFS[100] and NIR[101-103].

1.1.3.1 Mass spectrometry

The use of mass spectrometry in reaction analysis is a common technique and when coupled with flow systems, on-line mass analysis can be employed to qualitatively and quantitatively monitor reactions for the identification of products and intermediates, as well as analysis of relative composition.[104-106] All of this can be performed in real-time due to the short method times, and is ideal in the context of reaction monitoring as lots of information can be gained from the mass spectrometer with little need for data manipulation.[107]

An example where mass spectrometry has been used for on-line analysis in order to monitor reactive intermediates in flow is in the formation of benzyne, **1.32**, shown in Scheme 1.13.[108] This experiment was conducted to show how mass spectrometers are ideally suited to on-line analysis applications, although

they are rarely used in this way as traditional mass spectrometry instruments are too large and bulky to be conveniently coupled to flow systems.



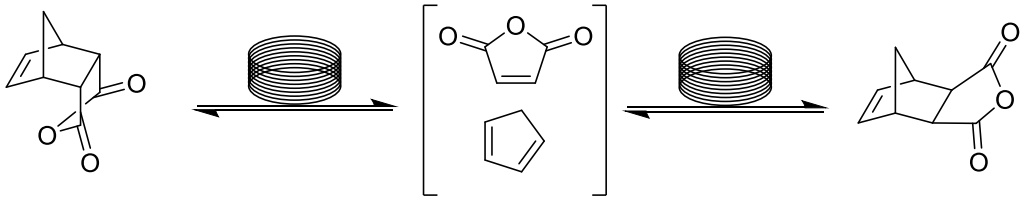
1.31 **1.32**

Scheme 1.13: A reaction scheme showing the formation of benzyne, **1.32**, which was followed by analysis via mass spectrometry.[108]

1.1.3.2 HPLC

HPLC can also give very good qualitative and quantitative analysis of a reaction mixture - however, the advantage that HPLC has over mass spectrometry is that it is typically easier to quantify the individual components of a mixture.[109] This is a particularly useful characteristic for modelling the kinetics of a reaction as the relative abundancies of products from a reaction can be measured precisely by their peak areas in the chromatograph at different times,[110] as well the ability to differentiate molecules with the same mass and even different isomers of the same mass.

An example where the kinetics of a reaction were studied using the coupling of a flow system with on-line HPLC, is in the thermal isomerization of the endo molecule, **1.33a**, to its exo counterpart, **1.33b**, shown in Scheme 1.14.[111] This example shows how combining on-line HPLC with an automated flow system with varying pump flow rates and temperature, can lead to an understanding of the factors influencing yield and purity of a process, with the ability to provide useful data to be utilised for generating response surfaces, shown in Figure 1.2.



1.33a **1.33b**

Scheme 1.14: A reaction scheme showing the thermal isomerization of **1.33a** to **1.33b**, which was followed by analysis via HPLC.[111]

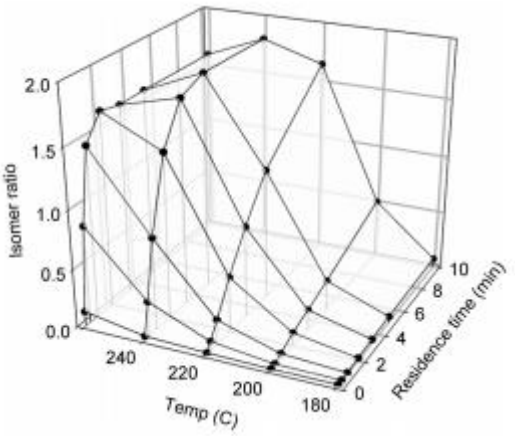
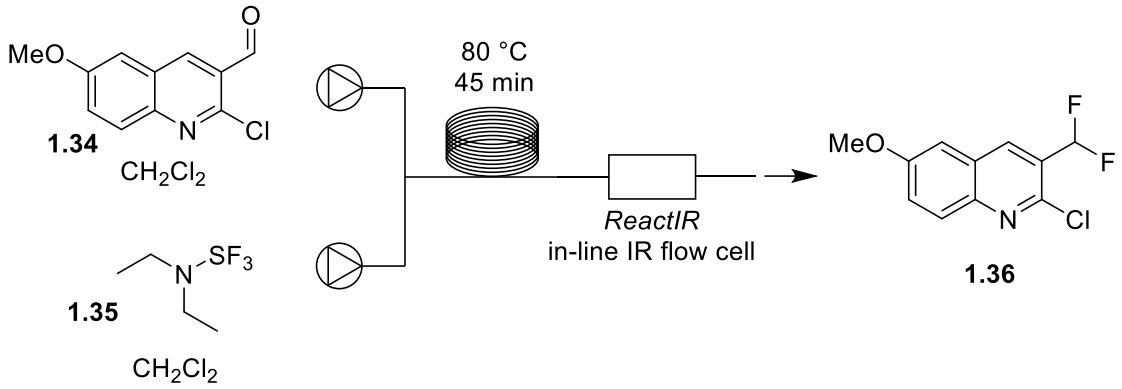


Figure 1.2: The response surface showing temperature and residence time effects on the thermal isomerisation of **1.33a**.[111] Reproduced with permission.

1.1.3.3 IR

The use of in-line IR analysis in flow systems is different as it is not quantitative in the same respect as mass spectrometry or HPLC. IR rather detects the relative abundancies of particular species by monitoring individual peak signals. The advantage to in-line IR is that key, identifiable peaks from functional groups in an IR spectrum (O-H bonds, C=O bonds, etc.) can be tracked over time as disappearance or appearance of peaks can serve as an indicator of reaction completion.[112-114] IR analysis is fundamentally non-destructive, which in the context of a flow system, means that in-line IR analysis has been developed and integrated into systems which do not require samples to be removed from the flow: a truly continuous, non-interrupted stream but still continuously analysed by IR.[115-117]

An example of where in-line IR analysis has been used to monitor the completion of a reaction is in the fluorination reaction to yield **1.36** shown in Scheme 1.15.[112] The in-line probe brand used was the ReactIR. In this reaction, the region where C-F bonds are expected to absorb was monitored during the reaction, in order to establish a reaction completion trend based on the intensity of the peak produced - this is shown in Figure 1.3.



Scheme 1.15: A reaction scheme showing a fluorination reaction yielding **1.36**, which was followed by analysis via in-line IR.[112]

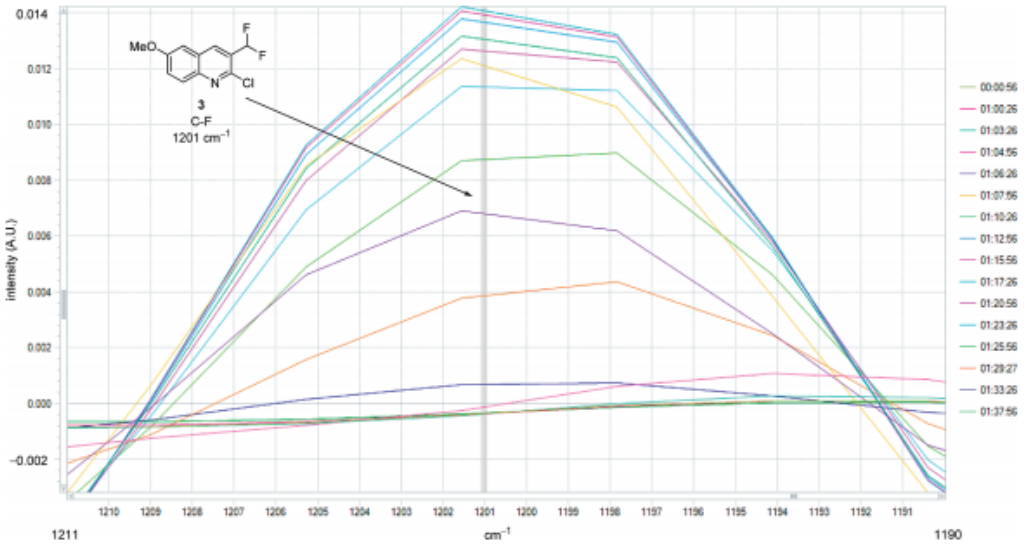


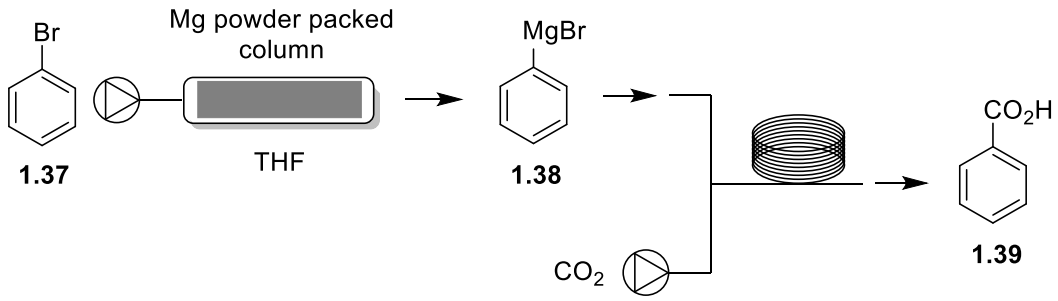
Figure 1.3: An IR absorption showing the detailed rise of the C-F bond peak over time, indicating the extent of reaction completion.[112] Reproduced with permission.

1.1.3.4 NMR

The use of on-line NMR is also a good quantitative option for flow system analytics. The potential to incorporate high resolution NMR systems into a flow system to show both structural and conformational changes in chemical transformations is a desirable potential analytical technique. At current, generally only small, low-field systems have been developed to be conveniently coupled with flow systems. However, this technique thus far has led to the increased value of using NMR spectroscopy as a non-invasive method as more process development applications are available.[118] Although there are some reports of using high resolution on-line NMR[119, 120], most cases generally utilise low field NMR systems. The low resolution analytical technique is unlike traditional NMR analysis in a high resolution system (> 300 MHz), as low-field systems tend to

analyse the relaxation times and relaxation weighted signals to identify product.[121]

The main downside to this technique, is that as the resolution is often too low for accurate interpretation of NMR spectra because of small differences in chemical shifts, therefore it is difficult to use for multi-component reaction analysis.[122] However, there are examples where this technique has been useful. In a report by Goldbach *et al.*,[123] low field on-line NMR was used to monitor the reagent and product concentrations of the Grignard reaction shown in Scheme 1.16. This reaction conversion was monitored by peak areas in the aromatic region over time, shown in Figure 1.4, where over 90% conversion was reported by the examination of peak appearance and disappearance.



Scheme 1.16: The Grignard reaction that was tracked by on-line NMR.[123]

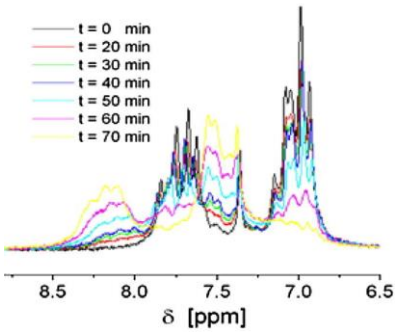


Figure 1. 4*:* The on-line NMR analysis in the aromatic region of the Grignard reaction shown in Scheme 1.16 at different time intervals.[123] Reproduced with permission.

1.1.3.5 Summary

There is no best method of flow reaction analysis, as each individual case would benefit from an independent examination of the necessary results to be generated. This is because different reactions and outcomes would be more suited to different analytical systems. A table summarising these main analytical methods, as well as other methods, with their respective sensitivities, acquisition speeds and limitations are shown in Table 1.1.[124]

Table 1.1: A table summarising reaction analysis techniques in flow systems, adapted from Houben and Lapkin.[124]

Technique	Type of information	Sensitivity /mol%	Typical speed of acquisition /s	Limitations
Mid-IR	Chemical identity, concentration	~10^{-1}	~1	Steps have to be taken to tolerate water within the system
Near-IR	Chemical identity, concentration	~10^{-1}	~1	Less information than Mid-IR
Raman	Chemical identity, concentration, crystal structure	~10^{-1}	~1-100	Fluorescence masking Raman signal
UV-vis	Chemical identity, concentration	~10^{-4}	<1	Limited number of species
NMR	Molecular structure, concentration	~10^{-3}	~10	Flow method is currently limited in sensitivity and resolution
GC	Concentration	~10^{-6}	10-1500	Typically slow, cannot identify unknown compounds
HPLC	Concentration	~10^{-6}	200-1500	Long method times, must be combined with mass spec. for proof of molecular identity
Mass spec.	Chemical identity, concentration	~10^{-8}	~5-20	Requires chemoinformatics expertise

1.1.4 Continuous flow chemistry summary

Flow chemistry is a powerful methodology that can be employed to obtain enhanced reaction control over similar batch processes: increased heat and mass transfer, accurate residence times and precise continuous addition of reagents. These advantages, alongside automated sampling procedures, mean that this regime is an ideal methodology for generating time-series data for the purposes of this project. As kinetic studies benefit from this additional control, as well as further hardware-manipulation techniques discussed in Chapter 3.1 that are not possible in batch, continuous flow will be used when possible. However, not all of the identified niches of flow will be exploited for this project, such as SCF usage, electrochemistry etc. and have been covered as a holistic overview of the field.

Flow chemistry is not the answer to every case study and it's likely that there will be instances where chemistry is better suited to batch conditions, however. Each chemistry must be evaluated to see if the study would more appropriate for a flow or a batch regime. For totally homogenous systems, or for very fast reactions, it is likely that a flow system would be used for the generation of kinetic data. For experimental case studies that require sampling of a suspension in a reaction medium, it is likely that a batch system would be more appropriate. In any case, the availability of an automated flow system is very useful (Chapter 3.2) and unlocks powerful reaction control that can be difficult to obtain in batch, but batch systems must be employed if the chemistry requires it.

1.2 Kinetic methodologies

The kinetic methodologies applied in this project are computational, which utilise ordinary differential equations (ODEs) to describe kinetic models, as well as optimisation algorithms to maximise the convergence of these ODE curves to the experimental data - this is described in detail in Chapter 2.2.2. This differs from traditional analytical techniques used more generally in chemistry, as well as specific applied methods in determining reaction order, described herein. These methods, although useful, are difficult to scale as they require user input and at times qualitative analysis, which are not suitable for automatic kinetic information determination. Furthermore, these methods may struggle with

complex reaction systems or systems with loss of mass balance. For these reasons, these methods are covered to describe common kinetic analysis techniques, but are not utilised in this project.

1.2.1 Conventional methodologies

There are several conventional, but considered outdated, methodologies to determining kinetic information that are still employed. The most basic analytical solutions for simple reactions can still be powerful in identifying reaction order and rate constants.[125, 126] Consider the first order reaction of species A reacting to form species B, shown in eqn. 1.1. The change in the concentration of A as time progresses can be described as a differential equation, eqn. 1.2, which can then be integrated to give the integrated rate equation, eqn. 1.3. This integrated rate equation can then be used to determine the concentration of A at any reaction time, and a plot of ln[A] vs. time allows the determination of the rate constant, as shown in Figure 1.5. This method also allows qualitative confirmation of the reaction order, as the experimental data points should fit to a linear line.

$$A \xrightarrow{k} B \qquad \text{eqn. 1.1}$$

$$\frac{d[A]}{dt} = -k[A] \qquad \text{eqn. 1.2}$$

$$\ln[A]_t = -kt + \ln[A]_0 \qquad \text{eqn. 1.3}$$

Where:

- A = starting material
- B = product
- k = rate constant
- $[A]$ = concentration of A
- t = time
- $[A]_t$ = concentration of A at time t
- $[A]_0$ = initial concentration of A at t = 0

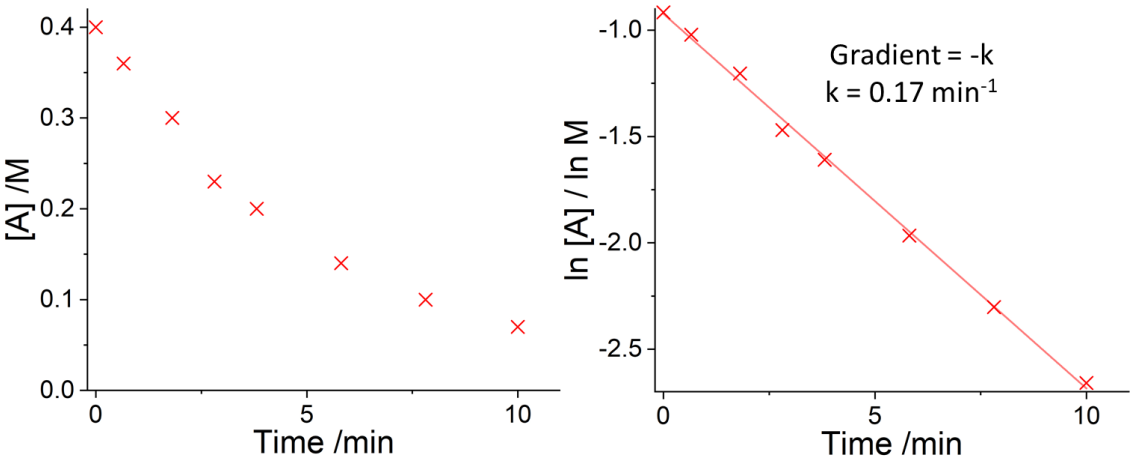


Figure 1.5: A graphical representation of generated time-series data for a first order reaction shown in eqn. 1.1. The data is log-transformed to give a linear fit where the gradient is equal to the rate constant multiplied by -1.

If the plot of ln[A] vs. time does not produce a linear fit, then it is likely that the reaction order is not 1 with respect to A, and may be zero order, second order or even another non-integer order. To deduce if the reaction is zero order, the reaction can still be viewed as eqn. 1.1, but the rate of change of A must be described differently as the zero order reaction is not dependant on the concentration of A at any time. Therefore, this differential equation can be described as eqn. 1.4, which can then be integrated to give the integrated rate equation, eqn. 1.5. A simple plot of [A] vs. time deduces if the reaction is zero order, as a linear fit indicates this, where the rate constant can also be identified, as shown in Figure 1.6.

$$\frac{d[A]}{dt} = -k \qquad \text{eqn. 1.4}$$

$$[A]_t = -kt + [A]_0 \qquad \text{eqn. 1.5}$$

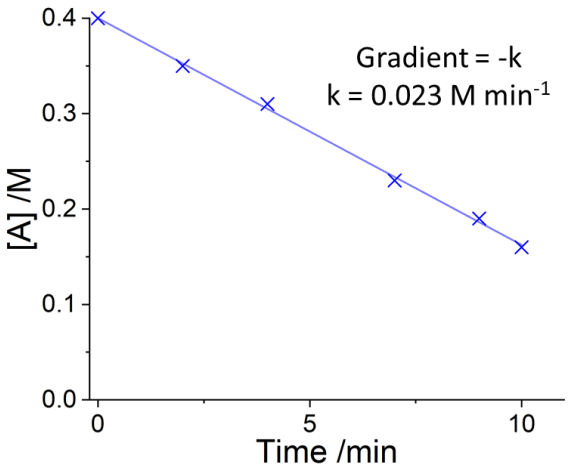


Figure 1.6: A graphical representation of generated time-series data for a zero order reaction shown in eqn. 1.1. This data does not need to be transformed as the linear fit is equal to the rate constant multiplied by -1.

If the reaction is second order with respect to A, the rate of change of A must be described as in eqn. 1.6, which when integrated gives the integrated rate equation shown in eqn. 1.7. A plot of 1/[A] vs. time then allows the chemist to deduce that the reaction is second order, if a linear fit is obtained. The gradient of this linear fit can then be ascribed as the rate constant, as shown in Figure 1.7.

$$\frac{d[A]}{dt} = -k[A]^2 \qquad \text{eqn. 1.6}$$

$$\frac{1}{[A]_t} = kt + \frac{1}{[A]_0} \qquad \text{eqn. 1.7}$$

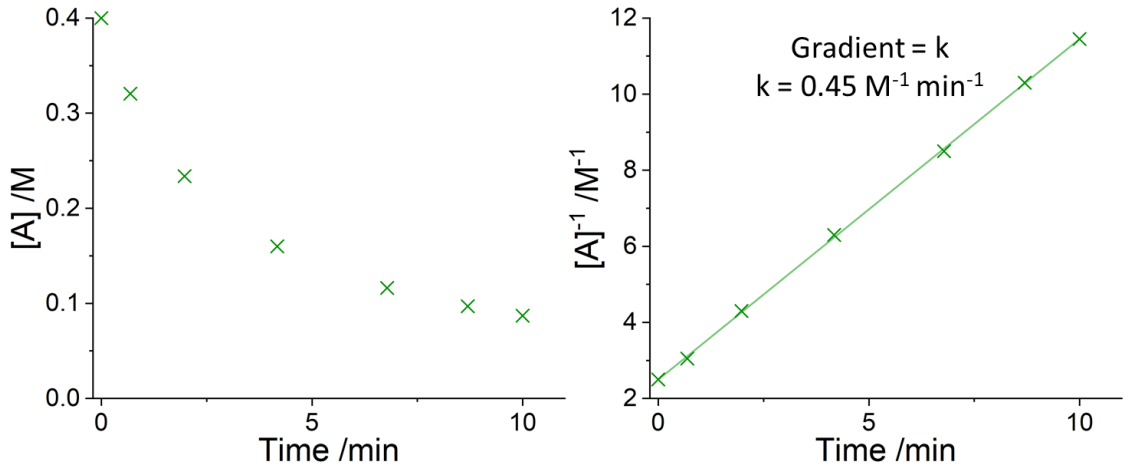


Figure 1.7: A graphical representation of generated time-series data for a second order reaction shown in eqn. 1.1. This data is plotted as the reciprocal of the concentration vs. time and the gradient of the linear fit to this data is the rate constant.

If the reaction is more complex and features bimolecular reactions of different species, such as the reaction shown in eqn. 1.8, then there are two possible scenarios that may be considered to determine the rate constant and confirm the reaction order. The first scenario is where the initial concentrations of species A and species B are the same. In this case, the rate of change of A can be written as eqn. 1.9. However, at any time, [A] = [B], meaning that the scenario mimics the second order reaction illustrated in eqn. 1.6, and kinetic information can be identified by using the corresponding aforementioned methodology.

$$A + B \xrightarrow{k} C \qquad \text{eqn. 1.8}$$

$$\frac{d[A]}{dt} = -k[A][B] = -k[A]^2 \qquad \text{eqn. 1.9}$$

The second possible scenario is where the initial concentrations of A and B are not equal. In this case, it is useful to define a term, x, where this relates to the concentration of each of the species that have reacted at a particular time, t. The expression of the rate law then becomes eqn. 1.10, which can be rearranged to eqn. 1.11:

$$-\frac{dx}{dt} = -k([A]_0 - x)([B]_0 - x) \qquad \text{eqn. 1.10}$$

$$\frac{dx}{([A]_0 - x)([B]_0 - x)} = k\,dt \qquad \text{eqn. 1.11}$$

If eqn. 1.11 is then integrated between time zero, and t, the time of interest:

$$\int_0^x \frac{dx}{([A]_0 - x)([B]_0 - x)} = k \int_0^t dt \qquad \text{eqn. 1.12}$$

Then integrating by using the method of partial fractions, eqn. 1.12 becomes:

$$\int_0^x \frac{dx}{([A]_0 - x)([B]_0 - x)} = \frac{1}{[B]_0 - [A]_0}\left(\ln\frac{[A]_0}{[A]_0 - x} - \ln\frac{[B]_0}{[B]_0 - x}\right) \qquad \text{eqn. 1.13}$$

Substituting [A] for $[A]_0 - x$ and [B] for $[B]_0 - x$, eqn. 1.13 simplifies to:

$$\int_0^x \frac{dx}{([A]_0 - x)([B]_0 - x)} = \frac{1}{[B]_0 - [A]_0}\left(\ln\frac{[B][A]_0}{[A][B]_0}\right) \qquad \text{eqn. 1.14}$$

Therefore resulting in the overall integrated rate equation:

$$\frac{1}{[B]_0 - [A]_0}\left(\ln\frac{[B][A]_0}{[A][B]_0}\right) = kt \qquad \text{eqn. 1.15}$$

Which can be rearranged to obtain:

$$\ln\frac{[B][A]_0}{[A][B]_0} = k([B]_0 - [A]_0)t \qquad \text{eqn. 1.16}$$

This obtained integrated rate equation, shown as eqn. 1.16, can then be used to determine the rate constant. A plot of $\ln\frac{[B][A]_0}{[A][B]_0}$ vs. time confirms a first order dependence on both species in the bimolecular reaction if a linear fit is

observed, as shown in Figure 1.8, where the gradient can be used to calculate the rate constant. If a bimolecular reaction is known to occur but this graphing indicates curvature, then it's possible that one or more of the species may have a reaction order not equal to one. Under these circumstances, other techniques must be used to confirm reaction order, such as computational approaches to fitting differential equations, which is further described in Chapter 2.2.2.

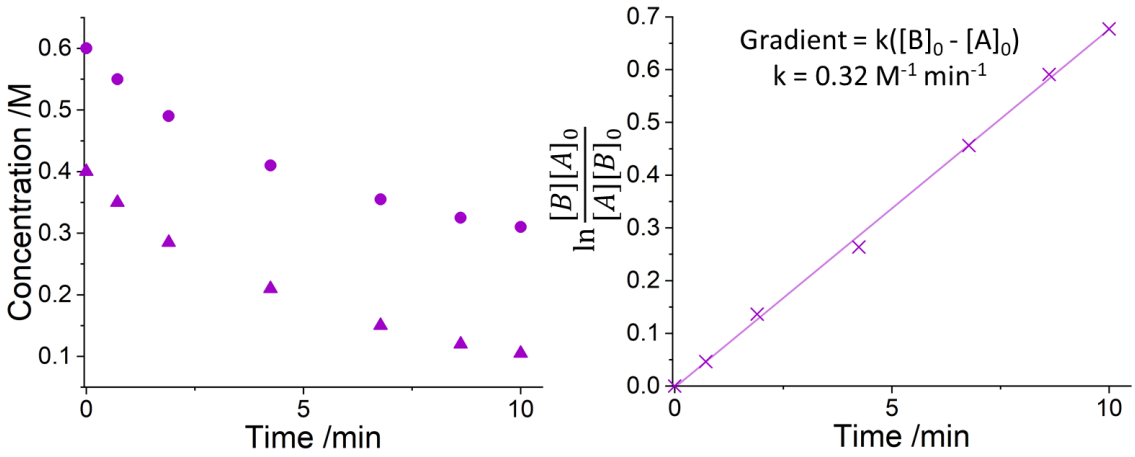


Figure 1.8: A graphical representation of generated time-series data for a second order reaction shown in eqn. 1.8, where: [A] = ▲, [B] = ●. This data is plotted as a log-transformed concentration fraction vs. time. The gradient of the linear fit to this data is the rate constant multiplied by the initial concentration of B minus the initial concentration of A.

For kinetic studies involving either biological (enzymatic) or chemical catalysts, the reaction system can still be described by elementary reaction steps and their respective differential equations. If we look at one of the simplest catalytic systems, where an enzyme, E, binds with a substrate, S. This is a reversible binding step, where the forward reaction step is k_1 and the backward reaction step is k_{-1}, in which the enzyme-substrate complex, ES, is formed. This complex can then react to form the product, P, as well as regenerating the enzyme catalyst, E. This reaction is shown in eqn. 1.17, and as before, the differential equations with respect to each of these species are described in eqn. 1.18 - 1.21.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P \qquad \text{eqn. 1.17}$$

$$\frac{d[S]}{dt} = -k_1[E][S] + k_{-1}[ES] \qquad \text{eqn. 1.18}$$

$$\frac{d[E]}{dt} = -k_1[E][S] + k_{-1}[ES] + k_2[ES] \quad \text{eqn. 1.19}$$

$$\frac{d[P]}{dt} = k_2[ES] \quad \text{eqn. 1.20}$$

$$\frac{d[ES]}{dt} = k_1[E][S] - k_{-1}[ES] - k_2[ES] \quad \text{eqn. 1.21}$$

However, although this system can be described in this way, it is more commonly described as approximated models, as simple assumptions can derive a model known as Michaelis-Menten kinetics.[127, 128] This provides a more simplistic method of catalytic modelling and prediction, by relating the catalyst concentration to the rate enhancement of a reaction in a different way. Michaelis-Menten kinetics employ the steady-state approximation with respect to the enzyme-substrate complex, ES. This means that it is assumed that this complex will rapidly approach a state where it is generated at the same rate that is consumed. This means that the overall rate of change, as described in eqn. 1.22, is zero - this allows a new relationship to be derived between the enzyme, substrate and complex, as shown in eqn. 1.23.

$$\frac{d[ES]}{dt} = k_1[E][S] - (k_{-1} + k_2)[ES] = 0 \quad \text{eqn. 1.22}$$

$$k_1[E][S] = (k_{-1} + k_2)[ES] \quad \text{eqn. 1.23}$$

In order to describe the system as an overall equation relating to the rate of the reaction, we must be able to determine the rate of product formation, shown in eqn. 1.20. However, as it is unknown what the free enzyme concentration is at any point after the reaction has progressed, it is more appropriate to describe the free enzyme, E, as the total concentration of all enzyme, E_T, minus the complex concentration: $[E] = [E_T] - [ES]$. This substitution into the steady-state approximation derived equation, eqn. 1.23, allows the complex concentration to be described by other concentrations:

$$k_1([E_T] - [ES])[S] = (k_{-1} + k_2)[ES] \quad \text{eqn. 1.24}$$

Followed by expansion:

$$k_1[E_T][S] - k_1[ES][S] = (k_{-1} + k_2)[ES] \quad \text{eqn. 1.25}$$

Then rearrangement and factorisation:

$$k_1[E_T][S] = (k_{-1} + k_2)[ES] + k_1[ES][S] \quad \text{eqn. 1.26}$$

$$k_1[E_T][S] = [ES]\big((k_{-1} + k_2) + k_1[S]\big) \quad \text{eqn. 1.27}$$

Then a final rearrangement allows the overall relationship to be described as:

$$[ES] = \frac{k_1[E_T][S]}{(k_{-1} + k_2) + k_1[S]} = \frac{[E_T][S]}{\left(\frac{k_{-1} + k_2}{k_1}\right) + [S]} \quad \text{eqn. 1.28}$$

Therefore, substitution of this relationship into the product formation step in eqn. 1.20, where *v* is the rate of the reaction, relates the overall rate with only the total enzyme concentration, substrate concentration and kinetic rate constants, shown as eqn. 1.29:

$$v = k_2[ES] \; = \frac{k_2[E_T][S]}{\left(\frac{k_{-1} + k_2}{k_1}\right) + [S]} \quad \text{eqn. 1.29}$$

To now arrive at the familiar form of the Michaelis-Menten equation, final substitutions are made for V_{max} and K_M, shown in eqn. 130, producing the overall equation shown in eqn. 1.31:

$$V_{max} = k_2[E_T];\; K_M = \frac{k_{-1} + k_2}{k_1} \quad \text{eqn. 1.30}$$

$$v = \frac{V_{max}[S]}{K_M + [S]} \quad \text{eqn. 1.31}$$

These terms used in the Michaelis-Menten equation, V_{max} and K_M, are different indicators to what are typically found in other areas of kinetics. V_{max} relates to the maximum velocity of the reaction, which is independent of substrate concentration. K_M is a ratio measure of the breakdown of the enzyme-substrate complex with relation to the its corresponding formation. Different enzymes have different K_M values and can be affected by a range of conditions, such as pH and temperature. This equation, and other variations when the reaction model differs (such as enzyme catalysed bimolecular reactions), are commonly used when studying catalytic kinetics to predict how a system will behave upon changing concentrations.

This model is also useful as the apparent reaction order of the system can be easily described. When the concentration of substrate, [S], is very low in comparison to K_M, this means that the substrate concentration is deemed to be

negligible, leading to a direct proportionality between the rate and [S]. This means the reaction appropriates first order kinetics, as shown in eqn. 1.32. When [S] is very large in comparison to K_M, this means that K_M is deemed to be negligible, meaning that the equation cancels any occurrence of [S], leading to overall zero order kinetics. This is shown in eqn. 1.33.

$$v = \frac{V_{max}[S]}{K_M} \qquad \text{eqn. 1.32}$$

$$v = V_{max} \qquad \text{eqn. 1.33}$$

V_{max} and K_M can be determined experimentally, by measuring the initial rates of reaction at varying substrate concentrations. After these rates are determined, the graph of reaction rate vs. substrate concentration can be plotted, where V_{max} is the reaction rate that the curve tends to. K_M is found via plotting the relationship of reaction rate to substrate concentration, indicated by a curve, whereby K_M is equal to the substrate concentration (x axis) when the reaction rate is half of V_{max} (y axis). The experimental determination of these constants is shown in Figure 1.9.

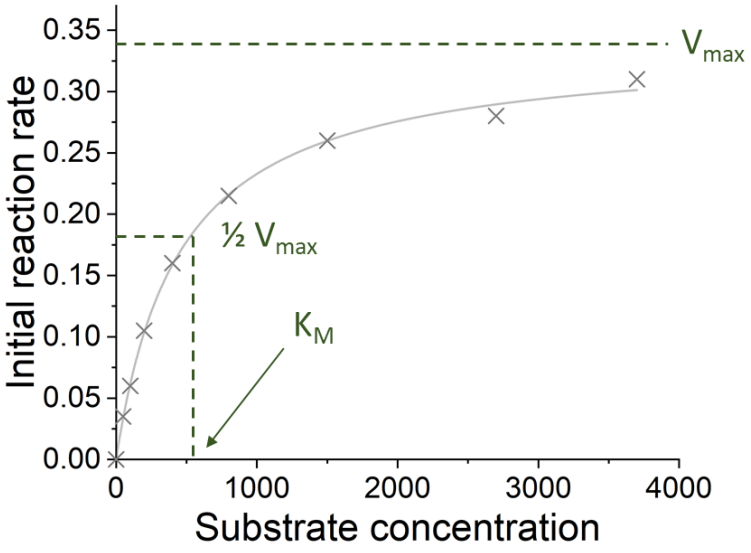


Figure 1.9: A representation of Michaelis-Menten kinetics, where the initial reaction rate of individual experiments with differing substrate concentrations, x, is plotted. This plot is then used to determine the kinetic parameters: V_{max} and K_M.

Further development in these Michaelis-Menten kinetic plots then linearised and simplified this technique, leading to a simpler interpretation of the kinetics, as well as easier adoption in determining kinetic parameters. This form of the Michaelis-Menten equation, known as the Lineweaver-Burk equation shown in eqn. 1.34, is simply the double-reciprocal form of eqn. 1.31.[129] This

allows less complex plotting methods to determine V_{max} and K_M, as shown in Figure 1.10.

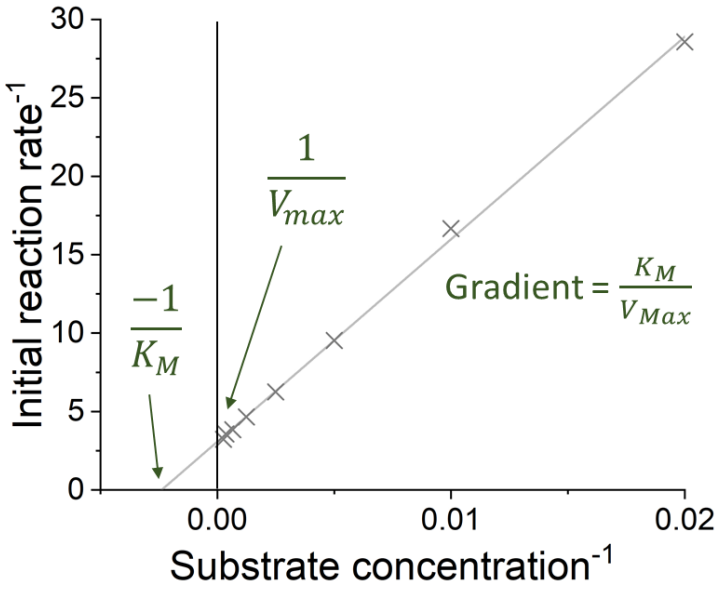


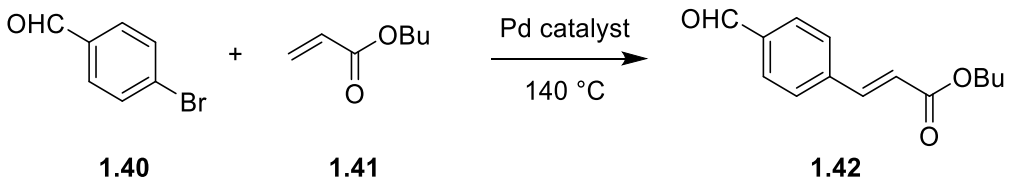
Figure 1.10: A representation of the Lineweaver-Burk equation, where the inverse of the initial reaction rate of individual experiments with differing substrate concentrations, x, is plotted. This plot is then used to determine the kinetic parameters: V_{max} and K_M.

1.2.2 Visual kinetic analysis

In order to identify catalytic reaction behaviour and the order of the species in these processes, different methodologies with the theme of visual, qualitative analysis were developed to allow many scientists to easily extract kinetic information from their reactions. Although the root of these analyses are still highly mathematical, visual kinetic analyses depend on the subjective analysis of the scientist. Therefore, precise kinetic parameters cannot be elucidated; however, the plots required are simple to construct and easy to interpret, allowing easy determination of the order of the catalyst/reagents or of product inhibition/catalyst deactivation.[130]

Reaction progress kinetic analysis (RPKA) employs a checklist to visually determine this kinetic information efficiently with very few reactions.[131] These reactions, which feature in situ measurements, are combined with mathematical manipulations to construct graphical rate equations where kinetic information can be identified. This methodology comprises three different analyses for three sets of experiments. These experiments identify: product inhibition/catalyst deactivation, the order in the catalyst and the order in any of the other components of the reaction.

For a product inhibition/catalyst deactivation probing study, curves of reaction rate vs. substrate concentration are compared for reactions that are started at different points - one with higher concentrations (reaction 1) and one with lower (reaction 2). This means that although the initial starting material concentrations are different, both reactions at some point will have the same concentration of all of the starting materials as the other reaction. This is true as long as both experiments have the same difference in the concentration of reactants, referred to as ‘same excess’ experiments. Consequently, at these points in the reaction where there are the same concentrations of starting materials, the reaction with the greater initial concentrations (reaction 1) will have a greater concentration of product, as well as a greater number of catalytic turnovers. This means that there are two sets of experimental data, where the reaction profiles should overlay to indicate the lack of product inhibition and catalyst deactivation. A case study where there was shown to be no indication of inhibition/deactivation is shown in Scheme 1.17 and Figure 1.11.[132]



1.40 **1.41** **1.42**

Scheme 1.17: A Heck coupling of an aryl bromide, **1.40**, with **1.41** to form the adduct **1.42**. This case study showed no indication of product inhibition or catalyst deactivation.

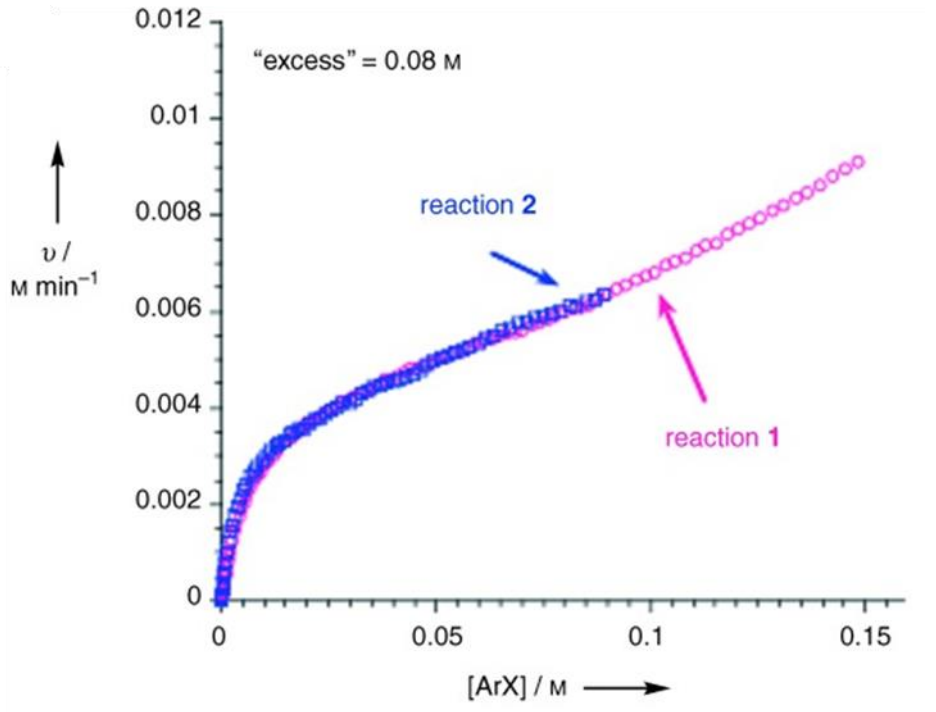
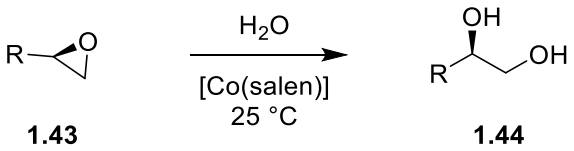


Figure 1.11: An overlay of two experimental datasets with the same 'excess' of reactant concentration, for the reaction shown in Scheme 1.17. This overlap indicates no product inhibition or catalyst deactivation. Reproduced with permission.

If these curves don't overlay, it can be inferred that product inhibition or catalyst deactivation is occurring. An example of this is shown in Scheme 1.18 and Figure 1.12.[133] To discern between these two possibilities, a third experiment is required where there is more product added. This experiment should feature the same initial concentrations as the experiment initiated at lower concentrations (reaction 2), but the reaction should also feature the same product concentration as generated by the other reaction until this point. This results in two reactions that are identical in stoichiometric composition, but vary in the number of turnovers completed by the catalyst. Therefore, overlay of these two reaction profiles indicate product inhibition, whilst no overlay indicates catalyst deactivation.



1.43 **1.44**

Scheme 1.18: The epoxide opening of **1.43** to form **1.44** using a catalytic system. This case study indicates some product inhibition or catalyst deactivation.

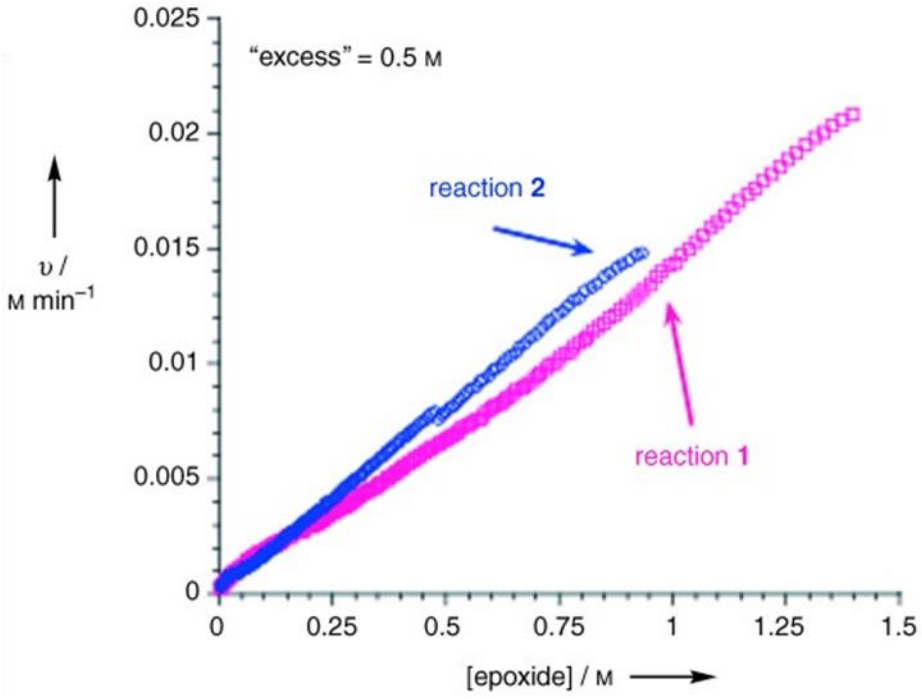
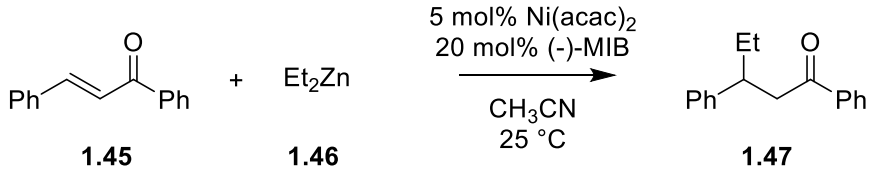


Figure 1.12: As there is no overlay in the two experimental datasets with the same 'excess' of reactant concentration, in the reaction system shown in Scheme 1.18, this indicates some product inhibition or catalyst deactivation. A further experiment is necessary to determine what is occurring. Reproduced with permission.

To then use RPKA to elucidate the order in catalyst, the datasets of different reactions run with different catalyst concentrations are plotted as rate/$[cat]^{\gamma}$ vs. substrate concentration. The value of γ is changed until all of the curves overlay - this value is then the correct order in the catalyst. Similarly for determining the order in a given reactant "B", the datasets of 'different excess' reactions are obtained i.e. where all concentrations are constant apart from the specific substrate of interest, in this case reactant B. The profiles are plotted as rate/$[B]^{\beta}$ vs. [A], and the value of β is altered until the curves overlay. This value of β then indicates the correct order for reactant B. A reported example of an alkylation reaction, shown in Scheme 1.19, was analysed in this manner to determine the order with respect to Et_2Zn, **1.46**, by performing 'different excess' experiments, as shown in Figure 1.13a. This data was then plotted using this methodology in Figure 1.13b to confirm the reaction order of this species to be 1.[131]



1.45 **1.46** **1.47**

Scheme 1.19: A nickel-catalysed alkylation reaction of chalcone, **1.45**, with Et_2Zn, **1.46**, to form the adduct **1.47**.

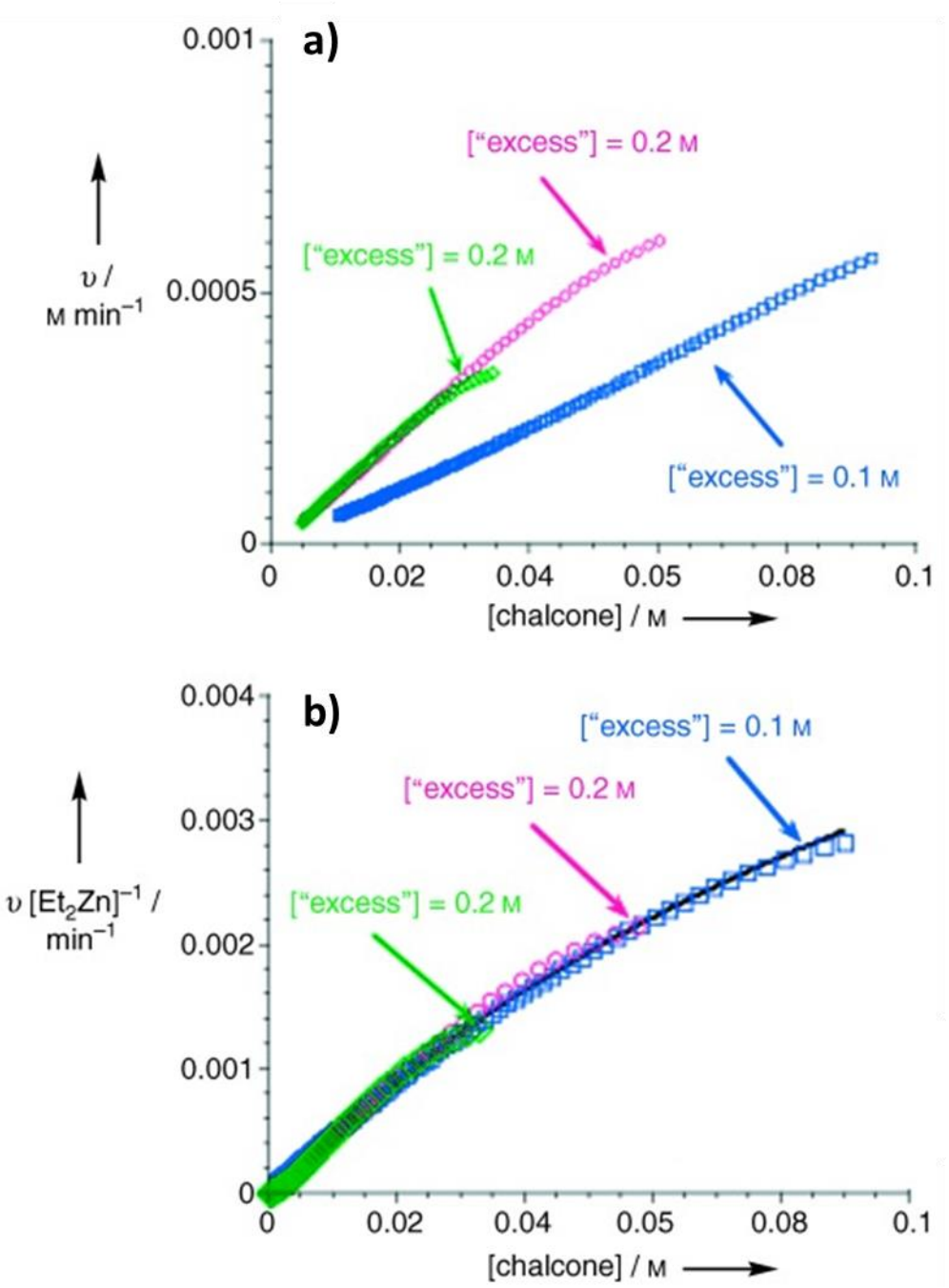


Figure 1.13: Graphical rate equations for the alkylation reaction shown in Scheme 1.19. **a)** Standard graphical rate equation. **b)** Using the RPKA methodology to observe any overlap in the 'different excess' reaction curves. This overlap indicates that the reaction order of Et_2Zn is 1. Reproduced with permission.

Another form of visual kinetic analysis, Variable Time Normalisation Analysis (VTNA) has also been reported to determine this same information by utilising exclusively concentration-time reaction profiles.[134, 135] This technique may be preferential to RPKA as this data is typically more readily available from

the use of almost any reaction monitoring technique, when compared to the data handling necessary to extract rate information.

To identify product inhibition or catalyst deactivation, the reaction profiles of two or more reactions are examined, where the reactions have different starting concentrations - these datasets are both plotted together. The profile of the reaction with the lower initial concentration of starting materials is then shifted on the time axis, until the first data point overlaps with the second (higher initial concentrations of starting materials) reaction profile. If there is an overlay of the two concentration profiles at this point, this suggests that there is no product inhibition or catalyst deactivation present in the system. This procedure is highlighted in Figure 1.14.[130] However, a lack of overlay indicates that one of these scenarios is present and a third experiment must be conducted to determine the cause of this lack of overlay. This third experiment must contain added product, in the same way as previously mentioned for the RPKA methodology, where an overlay of this new curve with existing data indicates product inhibition, and the converse outcome indicates catalyst deactivation.

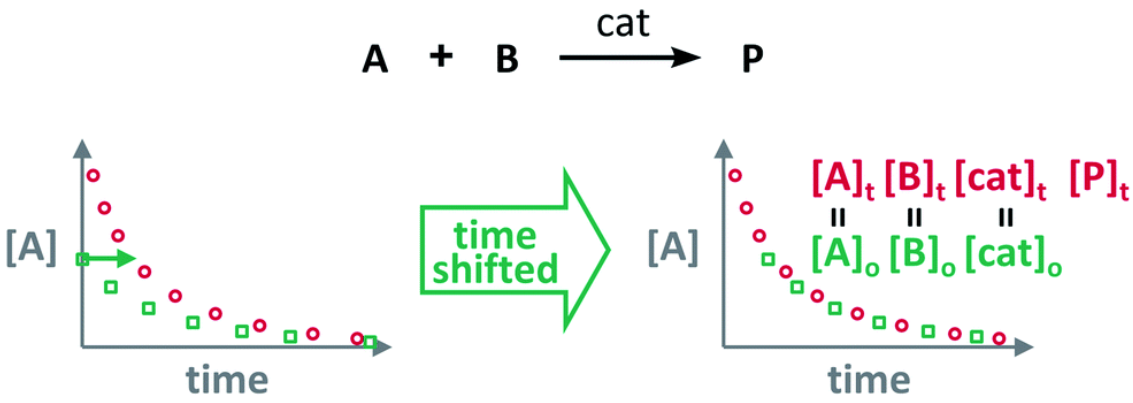


Figure 1.14: A time shift of the profiles of two reactions to observe overlaying plots. This allows the comparison of two profiles with the same starting material concentrations, but different product concentrations, therefore identifying signs of product inhibition or catalyst deactivation. Reproduced with permission.

To then elucidate the order in any of the species present, two reactions or more are run with different initial concentrations of this species, S, but with the same initial concentrations of all other species. The time scale of these reactions are then replaced with $\Sigma[S]^{\gamma}\Delta t$ - this expression is shown in eqn. 1.34.[136, 137] A plot of concentration of S vs. $\Sigma[S]^{\gamma}\Delta t$ is then graphed, where the value of γ that produces an overlay of the reaction profiles is the order in species S, in a similar

manner to RPKA plotting methods. This plotting method can be used for identifying both the order in catalysts and reactants in a chemical system.

$$\sum_{i=1}^{n}\left(\frac{[S]_i + [S]_{i-1}}{2}\right)^{\gamma}(t_i - t_{i-1}) \qquad \text{eqn. 1.34}$$

Where:

- n = number of experiments
- i = current experiment
- γ = order of reaction
- t = time of measurement

These visual kinetic analysis methods may be used in different circumstances for the simple determination of reactant/catalyst orders, as well as to identify any occurrences of product inhibition or catalyst deactivation. Where reaction rate information is available, RPKA may be used, and where concentration-time data is available, VTNA can be used. There are many applications reported by academic and industrial research groups in a wide range of catalytic reactions, using both RPKA[138-140] and VTNA.[141-143] Figure 1.15 shows a summary of the analysis of these techniques, with the information that each technique provides given the particular experimental protocol.[130]

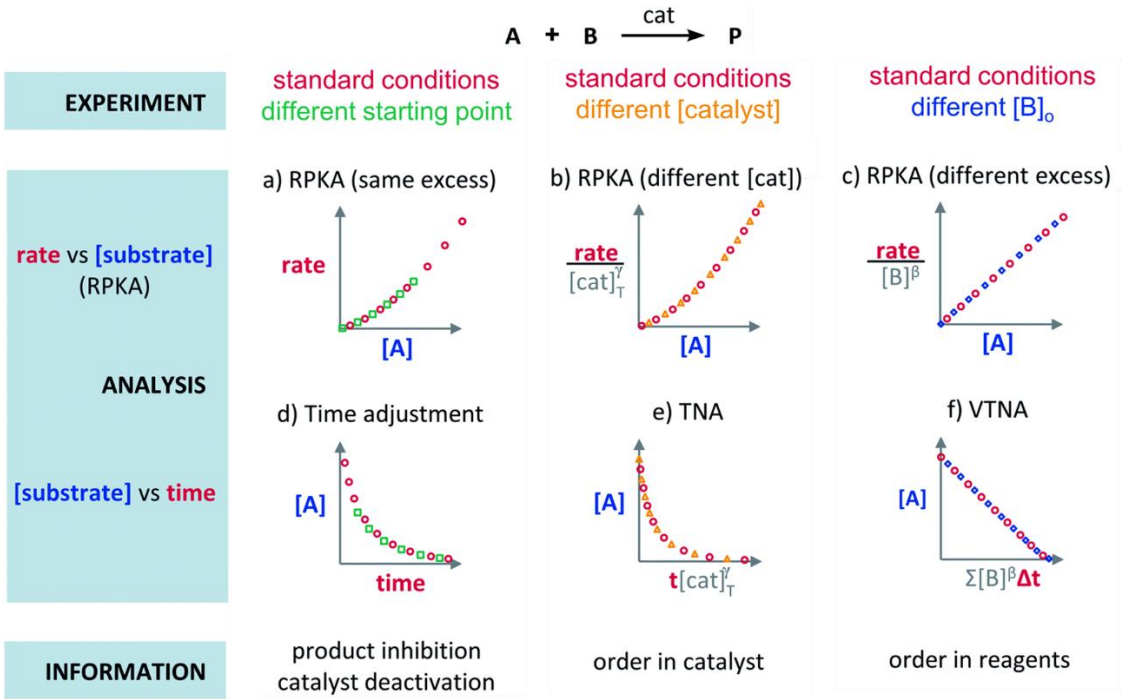


Figure 1.15: A summary of the use of RPKA and VTNA. **a)** catalyst deactivation or product inhibition, **b)** catalyst order, **c)** reagent order, **d)** catalyst deactivation or product inhibition, **e)** catalyst order, **f)** reagent order. Reproduced with permission.

1.2.3 Kinetic methodologies summary

Herein reported are several methodologies to determine kinetic information, all of which have become convention to varying degrees and adopted by many experts and non-experts alike. It is because of the large adoption of these techniques that is why these particular methods have been covered in this introduction, although the usage of them within the context of this project will be minimal. There are many other more complex methodologies peripheral to this project that, although are very powerful in particular circumstances, are not widely adopted by research groups. For this reason, the concepts of these methodologies are not introduced, but for further reading there are reports to be found on: model-based design of experiments (mbDoE),[144-146] hybrid modelling[147] and soft modelling,[148-150] as well as other more niche techniques that can be found in the literature.[151-153]

These introduced methodologies are unlikely to be used in the context of this project, as many require qualitative assessments or data manipulation by a user that is difficult to automate. Furthermore, the analysis of complex reaction models may become more and more difficult with increasing model terms, meaning that precise kinetic parameters may also be more difficult to determine. These factors produce tough challenges in the context of the main aims of the project (described further in Chapter 1.4), as the identification of kinetic models and parameters cannot be scaled effectively or automated. Therefore, other procedures are utilised in answering the challenge of creating a fully autonomous kinetic model and parameter determination methodology, which are introduced in detail in Chapter 2.2.

1.3 Optimisation

When searching for the optimum solution(s) to a particular optimisation problem, the best approach in gaining the most amount of information possible would be to simply measure the response of the optimisation function at every combination of input parameters at infinitesimally small variations. This would give a complete and perfect insight into the impact of the input parameters on

the response, meaning that it would be easy for a user to even qualitatively identify the optimum. However, as function evaluations can be expensive (computationally and otherwise), performing this resource- and time-intensive optimisation is not feasible.[154, 155] Approximations of this approach can be performed in some circumstances by applying grid search methodology,[156] but for many applications, the number of function evaluations necessary to explore such a large number of parameter combinations means that this approach is not viable. Due to the fixed-space exploration of these approaches, parameter estimations may also be imprecise.

For both efficient and accurate determination of optimum parameter inputs for a given function, it is often appropriate to use an optimisation algorithm. These algorithms heavily vary in their approaches to identifying optimised outputs, but a major criteria that they attempt to fulfil is to obtain this information in an efficient manner - specifically saving computational time and memory. The choice of algorithm can then depend on the nature of the problem, specifically the type of optimisation (local or global) as well as the nature of the optimisation problem. An algorithm can still generally achieve optimality in many circumstances, however, specific algorithms have been found to perform more efficiently in different situations. [157]

Many algorithms can be broadly described as either a local or a global optimiser. Local optimisers are fast, can handle large-scale problems and are widely applicable to many scenarios. For this reason, they are employed extensively in a variety of circumstances. However, in local optimisation, the compromise is to accept that the identified optimum may only be locally optimal, which does not guarantee a result that is globally optimal, i.e. a better function evaluation than all other feasible points.[158] These optimisers also require a starting point, which is a critical difference between local and global optimisers, as this can affect the objective value of the local solution that is found.

In global optimisation, by contrast, the true global optimum of the system is found. The compromise, however, is often efficiency. Global optimisation algorithms are commonly used where the computational cost of the function evaluations is not critical. This is because the number of function evaluations is traditionally very high, which can also lead to very large computational times.[158] Therefore, the choice to use a global optimisation algorithm may be made in

circumstances where the value of certifying the true global optimum outweighs the computational cost to arrive at this identification. Careful considerations must therefore be made in selecting the optimal algorithm based on what is desired in each scenario.

The nature of the optimisation problem itself can also be a large factor in the choice of algorithm used, as some algorithms can identify optima more efficiently based on the type of problem present.[159] Some common optimisation problems are highlighted in Figure 1.16. Specific classes of algorithms can then be utilised in different scenarios. Some classes of these algorithms include: linear programming (LP),[160] mixed integer linear programming (MILP),[161] nonlinear programming (NLP),[162] convex programming (CP),[163] quadratic programming (QP)[164] and more.[158] These algorithm classes can then also be further separated into discrete and continuous algorithms, then further still in many ways.

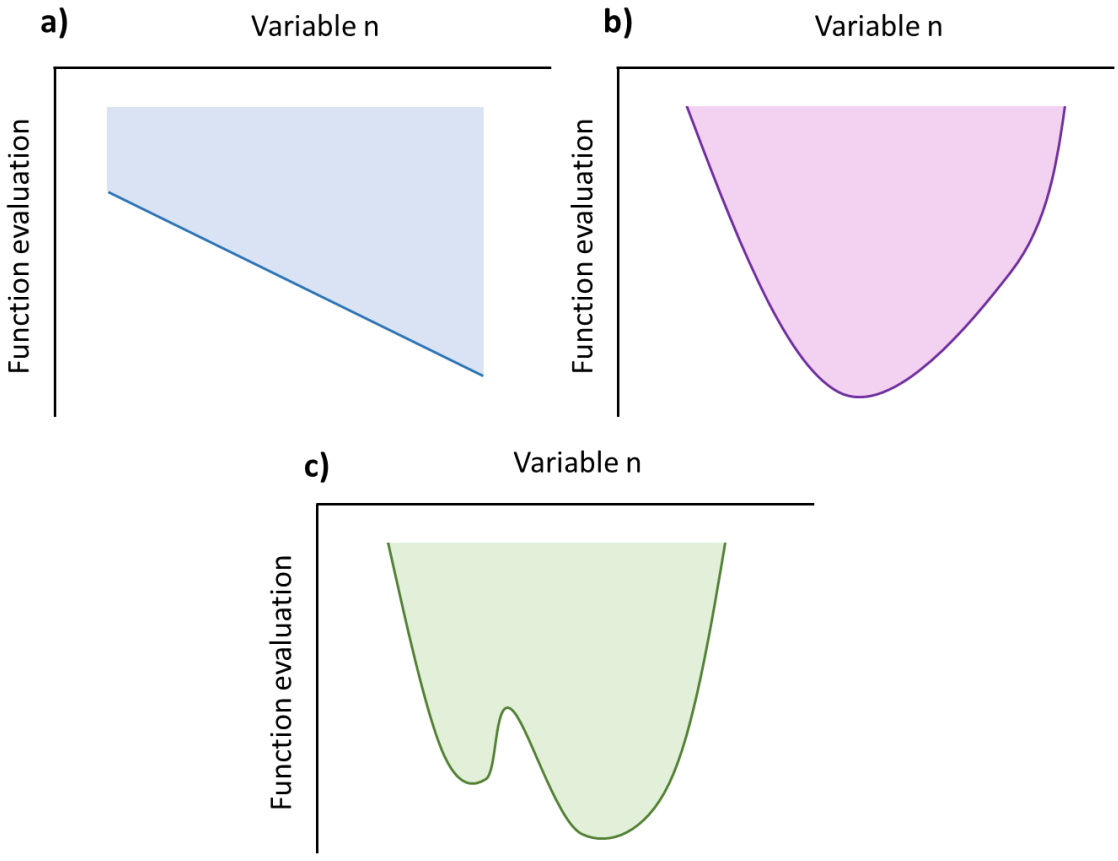


Figure 1.16: Some different classes of optimisation problem, where variable n is an input variable and the function evaluation is a measure of how the function is minimised with the changing variable. **a)** Linear optimisation problem. **b)** Convex optimisation problem. **c)** Non-linear optimisation problem.

As it is beyond the scope of this project to compare and discuss classes of optimisation algorithms, classifications are made only on their local/global nature. Commonly utilised local and global algorithms are qualitatively reported, with reference to further reading for mathematical proofs. As this project also involves heavy use of specific optimisation algorithms (in a black box manner), those utilised are also introduced qualitatively.

1.3.1 Local algorithms

Many local optimisation algorithms typically iterate towards an optimum by using gradients in the response or approximating the local response surface around the measurements. One of the most intuitive and simplest form of gradient-based algorithm is the steepest descent algorithm, that iterates by performing measurements along the trajectory of 'steepest-descent' towards the minimum.[165] An example steepest-descent representation is shown in Figure 1.17. Shown are two variables on the x and y axes respectively: X_1 and X_2, where the blue contoured area represents a minimum in the response and the red area represents a maximum in the response. A steepest-descent algorithm initially takes a guess of the inputs to measure the response. The algorithm then runs exploratory measurements in each direction from the initial guess and calculates the change in the response for each direction. The most favourable direction in n-dimensional space is identified, and measurements continue along this trajectory towards the minimum, until a decrease in the response is observed. The most favourable direction is then identified again, until the measurements can no longer observe a favourable change in the output. This terminates the algorithm as it has converged on a local minimum.[166]

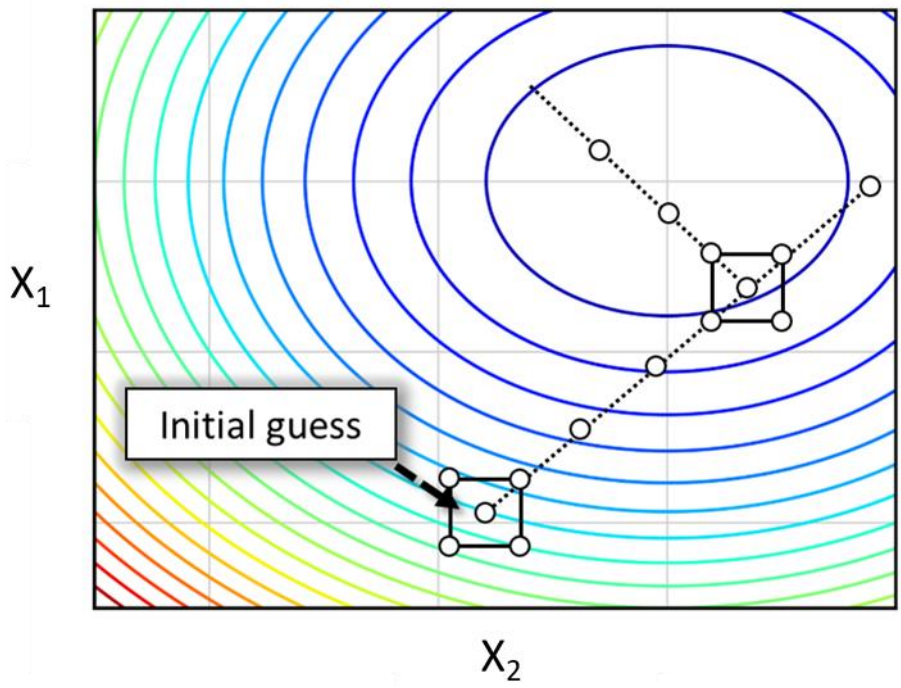


Figure 1.17: A representation of a steepest-descent algorithm minimising a 2-dimensional contoured parameter space, where **o** indicates a measurement. Where red areas are function maxima and blue areas are function minima.

Another intuitive gradient-based optimisation algorithm is the simplex algorithm.[167] This method uses convex polyhedra formed of n + 1 vertices (where n is the number of variables) - an individual polyhedron is referred to as a simplex. The algorithm begins by conducting either user-defined or random measurements of the response function at particular inputs, as shown in Figure 1.18 in 2-dimensional space, where each vertex of the initial simplex represents a function evaluation measurement. The worst performing vertex is then replaced upon each iteration of the algorithm via a reflection, resulting in a new simplex that explores a new area of parameter space. This approach locates areas with a more optimal response and hence successive simplex iteration converge on a local optima at point.

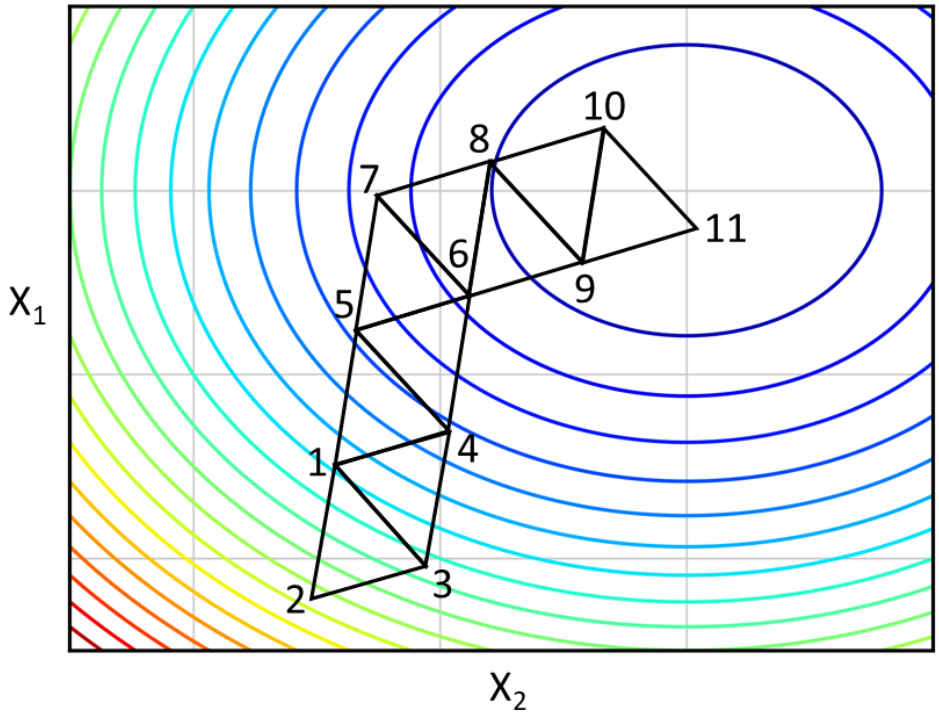


Figure 1.18: A representation of a simplex optimisation for a 2-dimensional contoured parameter space, where the numbered vertices of each polyhedron indicate a measurement. Where red areas are function maxima and blue areas are function minima.

As shown in Figure 1.18, the initial simplex, comprising of the measurements 1 - 3, is evaluated via the response function at each vertex. The worst vertex, measurement 2, is then replaced via reflection to evaluate the response function at measurement 4. Similarly, measurement 3 is then replaced by measurement 5 in a further simplex iteration. The optimisation typically stops when a better response function evaluation cannot be found, indicating that a local optimum has been identified - this is illustrated by measurement 11. Further modifications have also been adapted from this methodology, notably the Nelder-Mead simplex algorithm,[168] that allows further geometric transformations as well as just reflections. These transformations are highlighted in Figure 1.19. Because of the efficient and intuitive nature of these simplex algorithms, particularly the adapted Nelder-Mead variation, their use in applications in the literature are vast: the original Nelder-Mead report has over 30,000 citations.

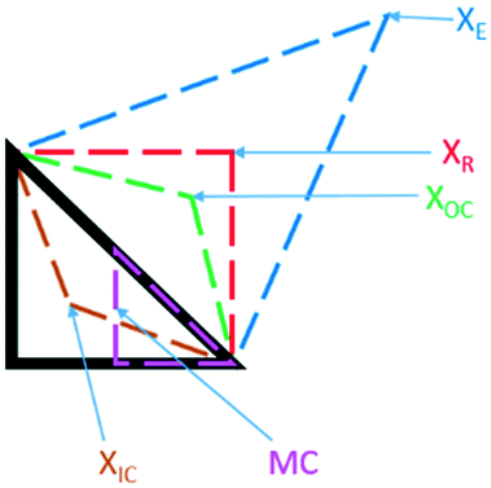


Figure 1.19: The different geometric transformations of the Nelder-Mead simplex: inside contraction (X_{IC}), multiple contraction (MC), outside contraction (X_{OC}), reflection (X_R) and expansion (X_E).

Another notable form of local optimisation methodology is the interior-point method. The interior-point method is a class of algorithm that is very effective at solving both nonlinear and convex problems, by approximating local regions around measurements as a 'trust region'.[157] An initial trust region, t_0, is identified around the starting point, shown as x_0 in Figure X.Y, where the trust region has an approximated objective function to minimise. The gradient of the trust region determines both how the region is approximated, as either linear or quadratic, and also the steps necessary in minimising the approximated objective function. This optimisation is shown as a 1-dimensional minimisation in Figure 1.20, where the x-axis is the magnitude of the variable and the y-axis is the measure of the approximated objective function.

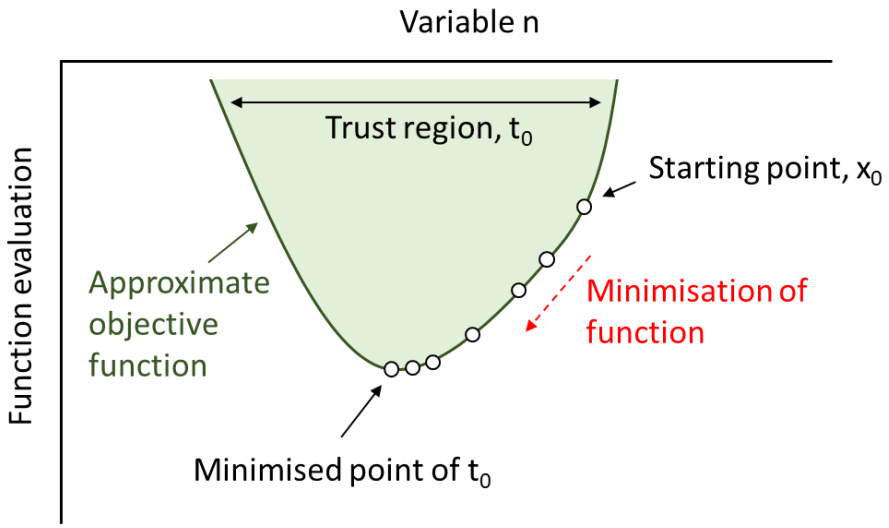


Figure 1.20: A 1-dimensional optimisation using an interior-point method. The approximate objective function is minimised for the starting trust region, t_0.

When this initial trust region objective function is minimised, a new trust region around this minimum is then approximated, shown as t_1 in Figure 1.21. This new approximated objective function is then also minimised. The algorithm continues to iterate via trust regions along what can be described holistically as the 'central path' towards the optimum for the true objective function, where no further improvements can be made. This central path can be viewed as an averaged direction based on the actual, algorithm path taken. In this example, the is objective function was minimised in the fourth trust region, t_3, finding the optimum x_{best}. More detailed algorithmic and mathematical discussions of this optimisation technique can be found in some of the more relatively recent reports.[169, 170]

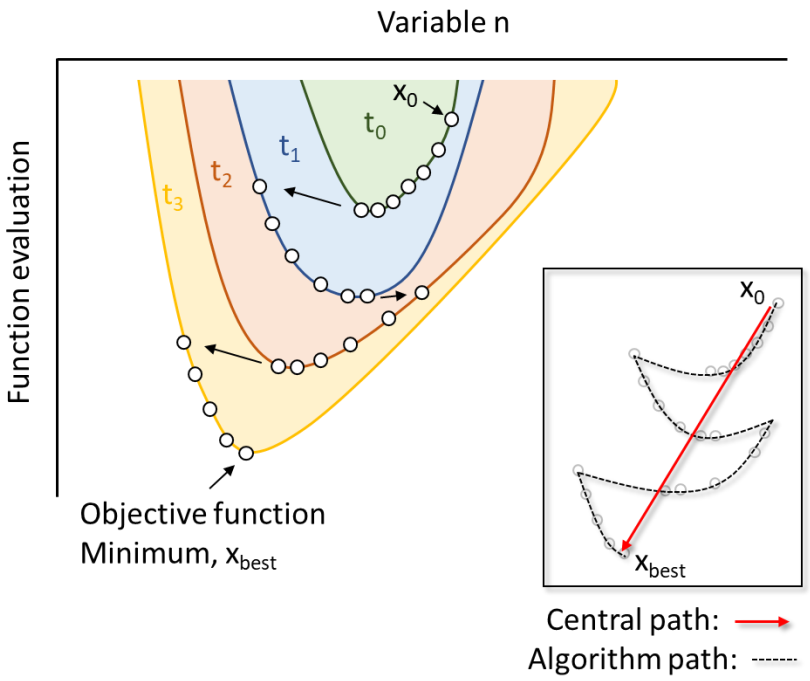


Figure 1.21: A 1-dimensional optimisation using an interior-point method. The approximate objective function is minimised for the sequential trust regions until the overall function is minimised. The path that the algorithm can take can be described as the algorithm path or the central path towards the optimum.

1.3.2 Global

There is much variation in the methodologies applied by global optimisation algorithms, but one of the most common techniques relates to what can broadly be described as a genetic algorithm.[171] Genetics algorithms perform computations to search for a global optimum, utilising techniques inspired by

Darwinian evolutionary biology. Although there are many forms of genetic algorithms, the methodology of a simple genetic algorithm (SGA) is qualitatively introduced herein. The six stages of SGA can be categorised as: initialisation, evaluation, selection, crossover, mutation and replacement. After initialisation, the other stages of SGA are repeatedly iterated until the optimum has been found, which leads to termination of the algorithm. The optimum is typically assumed to be found if the evaluation stage does not produce a better function evaluation (fitness value) for the given objective function (fitness function).[158] The workflow of the algorithm is depicted in Figure 1.22.

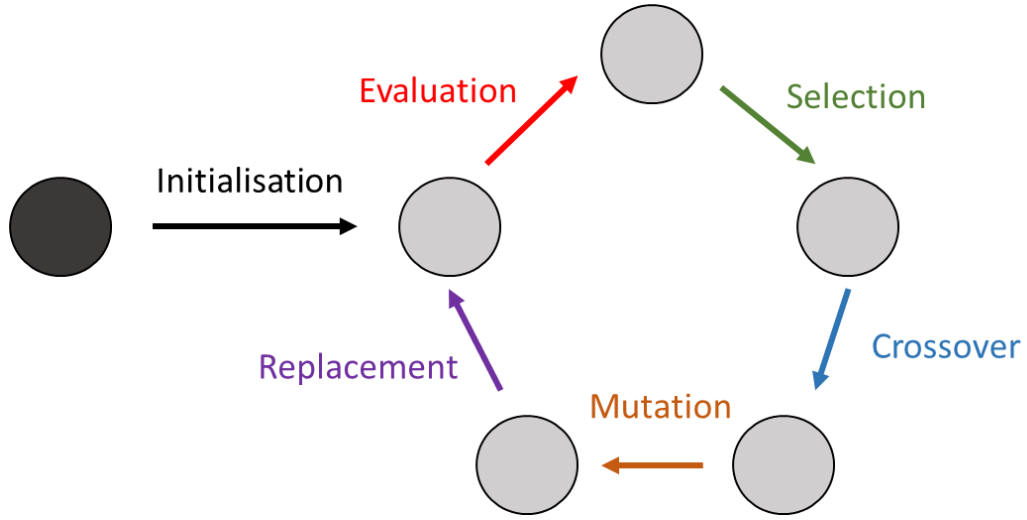


Figure 1.22: A depiction of the flow of a simple genetic algorithm.

The selection stage aims to find the best parents from the current population to generate the next population, where a population is a collection of all of the sets of inputs for the given variables, and a parent is one potential set of inputs from this population. The parents are often selected simply based on their fitness value. In the crossover stage, each individual variable input value (chromosome) from each parent has a probability of 0.5 to exist in the resulting children set. Each of the children in the following population then have another given probability of a random chromosome mutating in the mutation stage. This stage prevents a genetic algorithm from becoming stuck in a local minima. After this cycle has occurred, this represents one iteration of the algorithm which then successively occurs until the optimum has been reached. This SGA methodology is highlighted in the adapted simple example shown in Figure 1.23, whereby a fitness function (the sum of all inputs) is maximised by the input of integer chromosomes.[158] The ability of genetic algorithms to obtain an optimised result

for a given function is very desirable, even with a very large number of input variables. Despite the high computational expenditure to this methodology requires, it's reliability has resulted in high adoption in the general literature for many applications.

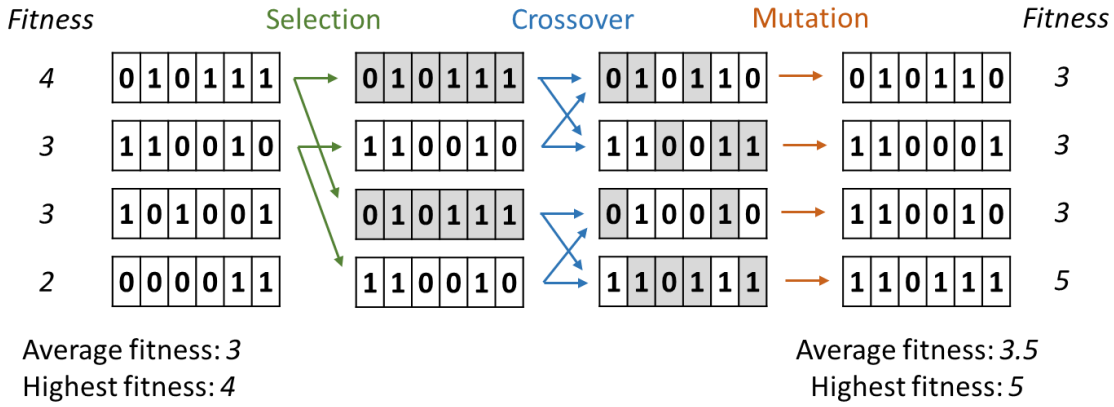


Figure 1.23: A depiction of one iteration of a simple genetic algorithm. Figure adapted from works reported by Boyd & Vandenberghe.[158]

Bayesian optimisation is another category of global optimisation methods that generally proceed in a more efficient manner than genetic algorithms, by utilising approximated (surrogate) models to optimise expensive-to-evaluate objective functions.[166] Random evaluated data points are initially conducted, then a surrogate model is built to fit these data points as well as a corresponding acquisition function. Bayesian optimisation methods obtain their efficiency from their use of these acquisition functions, which balance exploration of the parameter space and exploitation of already known high-performing regions of parameter space.[172] This acquisition function is maximised after each iteration of the algorithm to determine the optimum measurement point to evaluate next, in order to explore parameter space but also attempt to gain a better evaluation.

Iterations of a Bayesian optimisation for the minimisation of an arbitrary function, f(x), with respect to the 1-dimensional optimisation of variable x is shown in Figure 1.24.[166] Two initial measurements are inputted, shown as red dots in the first iteration, (i). The surrogate model is fitted in blue, where the blue area indicates the uncertainty in the model. The acquisition function is shown in red, with a vertical red line indicating the next parameter measurement to be evaluated - this evaluation is shown specifically in (i) as another red dot. From the second iteration, (ii), there is now enough data to fit 95 % confidence intervals to the surrogate model, indicating the regions of highest uncertainty. The

maximum of the acquisition function again determines the next parameter measurement. As aforementioned, this parameter measurement is made based on the compromise of exploring known regions of high performance, and regions where these confidence intervals are the largest. As further measurements are made in the iterations thereafter, (iii) - (viii), the confidence intervals shrink around areas richer with measurements and the surrogate model gains confidence in its predictions. Therefore, as the algorithm gains confidence in the model, this means that the algorithm also gains confidence that it has achieved the global optimum for the particular function.

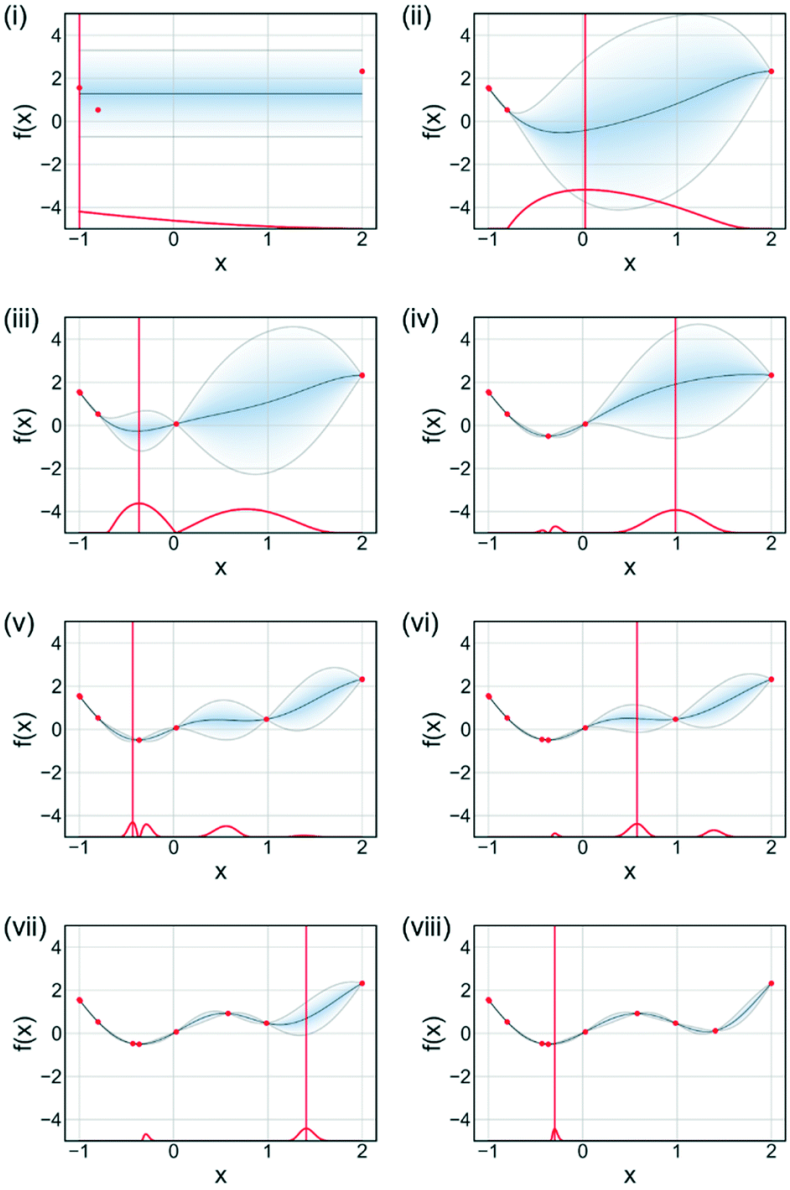


Figure 1.24: The Bayesian optimisation of f(x), where (i) - (viii) represent the sequential iterations of the minimisation. The acquisition function is shown in red, measurements are shown as red dots and the current estimated surrogate model is shown in blue with its corresponding 95 % confidence interval. The maximum of the acquisition function indicates the next measurement to be taken.

Integer linear programming (ILP) is a class of methodologies that are concerned with variables that can only take the form of integers, and although is not classically a type of global optimisation algorithm, is still a mathematical optimisation that is useful in specific cases and is utilised in this project. An ILP problem is in general very difficult to solve, and many solvers have attempted to simplify the procedure by utilising a much simpler linear programming (LP) approach, then rounding down the values to the nearest integer - this is known as LP relaxation, but may not always be effective in identifying the optimum solution, so other techniques must be used.[158] In ILP optimisations, it is also common to have multiple linear constraints, which define a feasible region that the optimum can be found.

Although ILP algorithms may proceed using different methodologies, the most intuitive introduction would be a 2-dimensional problem with a graphical method. The function to be maximised, f, shown as eqn. 1.35, is bound by 4 linear constraints, C_{1-4}, shown as eqn. 1.36 - 1.39:

$$f = 12x + 7y \qquad \text{eqn. 1.35}$$

$$C_1 = 2x - 3y \leq 6 \qquad \text{eqn. 1.36}$$

$$C_2 = 7x + 4y \leq 28 \qquad \text{eqn. 1.37}$$

$$C_3 = -x + y \leq 2 \qquad \text{eqn. 1.38}$$

$$C_4 = -2x - y \leq 2 \qquad \text{eqn. 1.39}$$

As the problem consists of two variables, x and y, it can be illustrated by a 2-dimensional graphical plane, where each linear constraint is a straight line forming a closed, feasible region - this is indicated in Figure 1.25.[158] Therefore, all feasible solutions to this problem must be an integer in x and y within the bounds of this feasible region. The objective function, f, is represented as a straight line with given inputs of x and y that indicates its current position. This function, however, can be viewed as a dynamic line that moves in parallel with respect to different values of x and y. Therefore, in order to obtain the maximum feasible evaluation of f, the line is moved from its current location in the figure

until it intersects with the allowed two integer-valued inputs within the feasible region for the first time. All integer-valued input combinations within the feasible region are marked with a red dot, where the first intersecting point would be either point 1 (2,3) or point 2 (3,1), indicated as P_1 and P_2 respectively. These points can be evaluated mathematically by the objective function to determinine the true maximum, where P_1 = 43 and P_2 = 45. Although this specific example is 2-dimensional and this intuitive graphical method cannot be easily performed in 3-dimensions and is impossible in further dimensions, many computational ILP algorithms proceed using this generalised concept to identify optimum inputs for a given function.

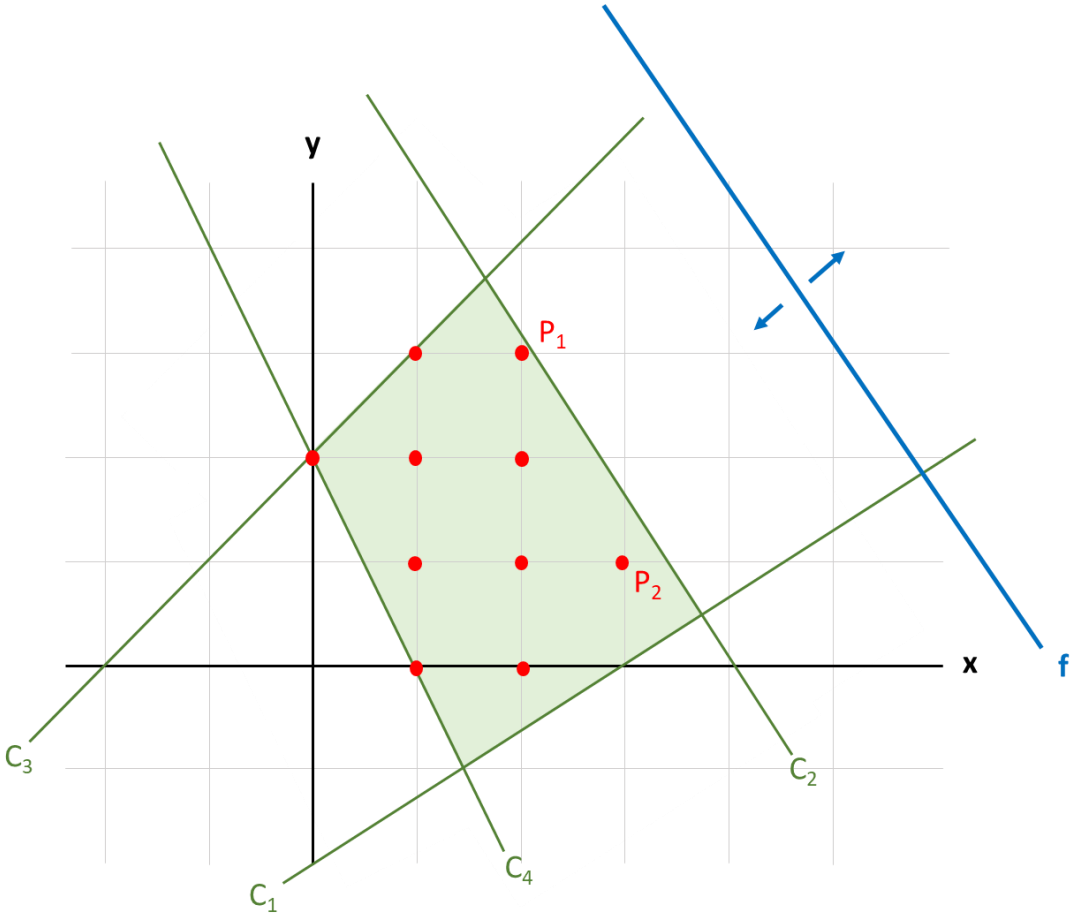


Figure 1.25: A graphical method to solving ILP problems, where C_{1-4} indicate the linear constraints, the green area indicates the feasible region, f shows the objective function, red dots represent the integer values of the inputs and P_1 and P_2 indicate two feasible potential maxima to the problem. Figure adapted from works reported by Boyd & Vandenberghe.[158]

1.3.3 Optimisation summary

There are many methodologies that optimisation algorithms utilise, meaning that there are several classes of algorithm that are available and then

multiple subclasses thereafter. There are specific cases where a particular algorithm class is necessary to solve a problem, but also cases where many algorithms may be suitable; it is then necessary to decide which algorithm to use in this instance based on other factors such as computational efficiency. In this project, these decisions will be explored as optimisation algorithms must be used in multiple unique scenarios. This is because other methodologies such as grid search are too inefficient and imprecise, especially as many input variables (for example, kinetic parameters to be optimised) must be determined in an appropriate length of time and with high precision.

1.4 Research aim

The aim of this project is to develop and advance an automated methodology to kinetic model and parameter determination. The basis of this project arises from a methodology reported by Tsu *et al.*[173] that provides a framework to a technique that achieves this, by evaluating all possible reaction models. This reported framework, however, has limited simulated examples and no real experimental examples that prove its efficiency. This methodology can represent a positive change in the laboratory, where physical-organic chemists can make use of their human resource on other tasks instead of determining a reaction model and kinetic parameters, which can be assigned to a computer. This is the advantage to exploring this automated methodology that this project is predicated on, which leads the specific project goals herein described:

- To build this automated computational approach to kinetic model and parameter determination, based on the framework aforementioned, as well as to test the effectiveness of the approach in simulated/literature case studies. (Chapter 2)
- To prove the effectiveness of the approach in real experimental case studies. This approach will be coupled with automated kinetic profile generation by flow rate manipulation in a continuous flow chemistry experimental setup. (Chapter 3)

Although methodology described is very powerful for many chemical scenarios, there are also instances where this technique does not work. Specifically, this approach is not designed for applications where the reaction

order can be a non-integer, or where particular species may be catalytic in nature. Therefore, further specific project goals are described:

- To further develop this automated approach to kinetic model and parameter determination, to include chemistries that may proceed via a catalytic route or have non-integer orders with respect to one or more species in the reaction model. The effectiveness of this new approach must also be proven in a simulated/literature case study(s). (Chapter 4)
- To prove the effectiveness of this new approach in real experimental case studies to achieve kinetic information. (Chapter 5)

This project will ultimately lead to a tool that any chemist can utilise, by plugging time-series data and chemical species into the automated approach, then observing the output in the form of the most likely kinetic model and associated parameters. Without kinetic expertise, any chemist will be able to use this approach and interpret the output in order to obtain scalable process understanding.

Chapter 2 : Development of the computational approach to kinetic model determination

2.1 Introduction

A chemical system can be seen as a cascading, dynamic reaction scheme, where molecules can orientate themselves in a particular way and with a particular energy to allow a reaction to occur. This reaction of course will follow a particular reaction mechanism. These mechanisms can give a deep understanding of the underlying chemistry that we merely observe as a transformation, and can give real-world benefits as it then becomes clearer and easier to control and understand chemical reactivity.[174, 175]

For the development of a chemical process, however, one of the greater concerns is to mathematically characterise the transformation from the starting materials to products - this can be complex and occur over many reaction steps, and involve many measurable and immeasurable intermediates. This allows quantitative information to be gained regarding the chemical synthesis, allowing

for classical reaction engineering principles to be applied to shorten process development times and lower costs.[173] Therefore, the objective is to develop stoichiometric and kinetic descriptions of the individual transformations, rather than detailed mechanistic insights and rationales.[176] Where stoichiometry refers to the stoichiometric coefficient of a particular species in a reaction, and kinetic descriptions refer to the kinetic parameters involved in the rate equations for a process. For a system with multiple reactions, this is referred to as a chemical reaction network (CRN).

The basis for many stoichiometric and kinetic modelling studies, with the specific aim of CRN determination, is the data-driven work reported by Aris and Mah.[177] This work is an early example of using computation to evaluate aspects of a CRN, specifically how many independent reactions can be shown to account for changes in experimental data, providing that all chemical species are measured. Bonvin and Rippin[178] later reported tar

get factor analysis (TFA), which was shown to derive approximate stoichiometric models for complex systems as well as to test user-inputted stoichiometries for their suitability to experimental (simulated) data. TFA has been used and advanced to verify proposed reaction stoichiometry in a select few reported works.[179, 180]

TFA was then used as the basis for further work reported by Brendel *et al.*[181] and Bhatt *et al.*[182], whereby an incremental identification method was proposed. This is a step-wise method, where the stoichiometry is initially verified using TFA, then other model identification strategies are utilised to identify the structure and parameters of the rate laws within the ODE models. However, TFA initially requires a CRN to be at least postulated by a user before it can be used. This requires a high degree of chemical intuition, but may also be incorrect as there are many possible forms in which the CRN can exist, especially when dealing with a system with a large number of chemical species.

An alternative step-wise method was also reported by Burnham *et al.*[183], in which a global ODE model structure was generated, which is capable of representing an entire set of possible chemical reactions. As this global ODE model structure represents an over-trained model, model terms are sequentially removed using statistical measures until all insignificant terms are removed,

giving the final CRN. As this features several optimisations of models with many kinetic parameters in order to identify insignificant terms, this optimisation can be very slow and also suffer from local optima. This is also the case with the determination of insignificant terms, as the identified optimum model itself may be non-global.

Willis and von Stosch[176] then reported the use of mixed integer linear programming (MILP) to determine the model structure as well as its associated kinetic parameters. This was because previously reported step-wise identification approaches fail to consistently predict the underlying network structure.[184-186] In this work, a global CRN structure was reported, but MILP was used to dismiss all non-mass-balancing reactions from the structure. Decomposition of the global structure then occurs by statistical means, and although the global structure is now much smaller in this work, termination at local optima may still be an issue.

Tsu *et al.*[173] then reported an improvement upon this method where ILP initially identifies all mass-balancing reactions, then builds every possible reaction model. It is this approach that will be focussed on in this chapter, and further description and discussion is provided in Chapter 2.2. Every model is then fitted and statistically evaluated, which avoids the potential pitfall of local optima when determining the correct model. However, this can be computationally demanding as there are many more models to evaluate than in any previously reported work.

Other works in the literature include more algorithmic approaches to CRN identification. Differential evolution[187], genetic programming[188] and multi-objective genetic algorithms[189] have all been used as global optimisers to attempt to determine the CRN. These specific global optimisers, which focus on mimicking real-life Darwinian evolution, are attractive as the CRN is identified with minimal human interaction or expertise. However, this optimisation method can still encounter problems due to local optima and take a very long time to compute due to the number of iterations necessary for convergence upon the optimum.

The main assumption that is universal among these approaches to CRN determination is that all reaction kinetics follow the law of mass action. This

states that the rate of reaction of an elementary reaction is directly proportional to the product of the concentrations of the reactants, raised to the power of their stoichiometric coefficients.[190] For the reaction shown in eqn. 1, based on the law of mass action, the reaction rate can be described by eqn. 2 and therefore eqn. 3. The kinetic rate constant, *k*, therefore determines the speed of the reaction as well as the reactant concentrations.

$$\alpha x_1 + \beta x_2 \overset{k}{\rightarrow} \gamma x_3 \qquad \text{eqn. 2.1}$$

$$rate \propto [x_1]^{\alpha}[x_2]^{\beta} \qquad \text{eqn. 2.2}$$

$$rate = k[x_1]^{\alpha}[x_2]^{\beta} \qquad \text{eqn. 2.3}$$

Where:

- α, β, γ = stoichiometric coefficients
- $x_{1\text{-}3}$ = chemical species
- k = kinetic rate constant
- $rate$ = rate of reaction

The law of mass action is applicable in most cases, however there are particular cases where it is not accurate to describe a process using mass action kinetics, where these approaches are not applicable. Such processes typically occur when concentrations of particular substrates are very low.[191]

2.2 Development of the approach

The computational approach was developed using MATLAB, by loosely following the approach outlined by Tsu *et al.*[173] The approach contains two main, sequential stages: kinetic model generation and kinetic model fitting. Firstly, all mass-balance-allowed reactions are identified from the inputted species (starting materials, intermediates and products). Then, all of the feasible reaction models i.e. all possible combinations of these allowed reactions, are compiled and stored in a model database.

After the model database has been generated for the particular process, the kinetic model fitting stage then evaluates the fit of the ordinary differential equations (ODEs) for that particular model to the experimental data provided. The MATLAB ODE solvers can be used to in such a way to allow an assessment

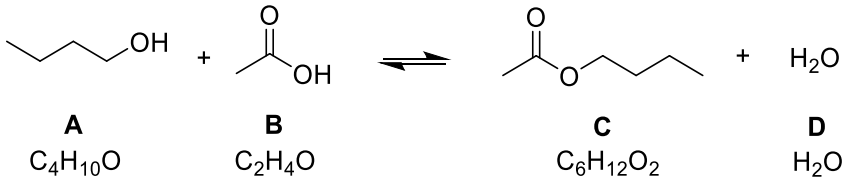
of how the concentration of a species changes over time when given particular k values. Therefore, optimising the k values for a given reaction model, by maximising the convergence of the ODEs with the experimental data, provides an indirect route to their identification. The convergence metric, which is the sum of squared error (SSE) between the simulated ODEs and the data, as well as the k values for each model are then stored.

Statistical analysis is then applied to each of the models. Corrected Akaike's Information Criterion (AIC_C) is used as a measure to identify the models that best balance model simplicity and convergence to the data, and hence, identify the most valid models. When these models have been identified, it is then possible to run Monte Carlo simulations if the error of the experimental data is known. These simulations assume that the error is normally distributed around the data point, and compile both the error in experimental data and error in the kinetic fitting of the ODE to obtain an uncertainty prediction in the identified k values for the process.

2.2.1 Kinetic model generation

In order to generate the kinetic models for evaluation, the first step is to input all of the known species of a chemical process into the system. The species can be identified by any means, typically by experimentation, where all starting materials, intermediates and products are identified. These species are inputted based on their respective number of differing atoms, i.e. number of carbon atoms, number of oxygen atoms etc., which is known as a Mass Matrix (MM).

For example, if you take the known reaction model of A + B ⇌ C + D shown in Scheme 2.1, you could visualise initially inputting these four species into a mass matrix, shown in Figure 2.1.



Scheme 2.1: An example esterification reaction.

A	B	C	D	
4	2	6	0	No. of C atoms
10	4	12	2	No. of H atoms
1	2	2	1	No. of O atoms

Figure 2.1: The Mass Matrix inputted into the approach, assigning mass to each of the species.

It is also possible to describe the reactions (forward and backward) shown in Scheme 2.1 in terms of the stoichiometry of reactants used (S_R) and the products formed (S_P), as a set of matrices in Figure 2.2. When describing these matrices, a 0 represents that the species does not participate, a 1 represents that the species does participate with a kinetic order of 1, and 2 represents that the species participates with a reaction order of 2. For example, in the forward reaction, both species A and B participate as a reactant with reaction order 1, meaning that they are represented by '1', whereas C and D do not participate as reactants, so are represented by '0'. These stoichiometric matrices are then combined to form an overall stoichiometry, S, where $S = S_P - S_R$. This overall stoichiometry for an unknown model would otherwise be identified using an Integer Linear Programming (ILP) optimisation, and is necessary to identify all feasible mass-balanced reactions.

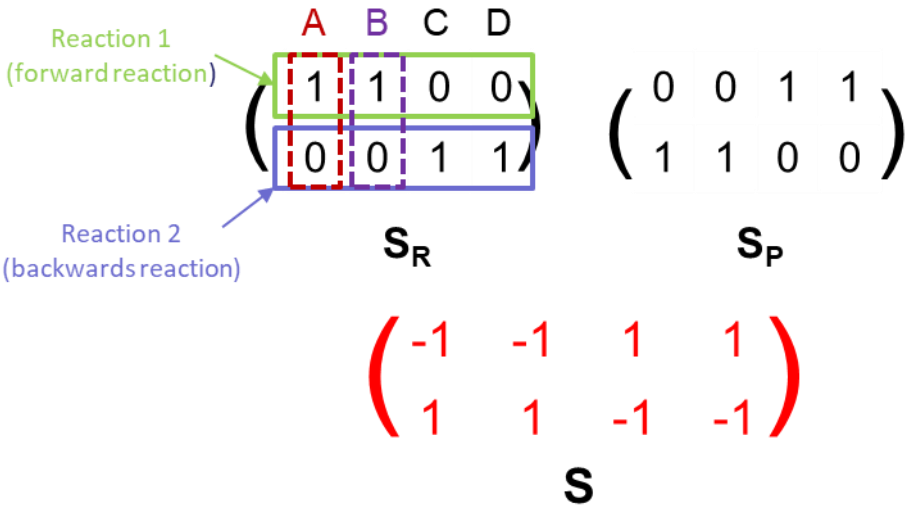


Figure 2.2: The stoichiometric matrices for the reaction model shown in Scheme 2.1, as well as the overall stoichiometry matrix, S.

The ILP optimisation proceeds by finding every feasible stoichiometry that satisfies the objective function of MM·S = 0, where S is the transposed vector of a single reaction with values for each of the species present. Therefore, as there are four species in the esterification process shown in Scheme 2.1, the ILP optimisation will proceed using the known Mass Matrix to identify the values of $X_{1\text{-}4}$ that satisfy the objective function, shown in Figure 2.3. The optimised solution is then the resulting stoichiometric vector which relates to a particular feasible reaction. The upper bound for the number of reactants in a stoichiometric vector is 2, as it is very rare to encounter third order reactions as a single reaction step. When an optimised solution is found, the solution is saved, then successive constraints are added to the ILP optimisation to guarantee that the same combination of binary variables (and hence stoichiometry) are not obtained when the optimisation problem is solved again. The optimisation then ceases when there are no more solutions to be found, and hence all feasible mass-balance-allowed transformations are identified.

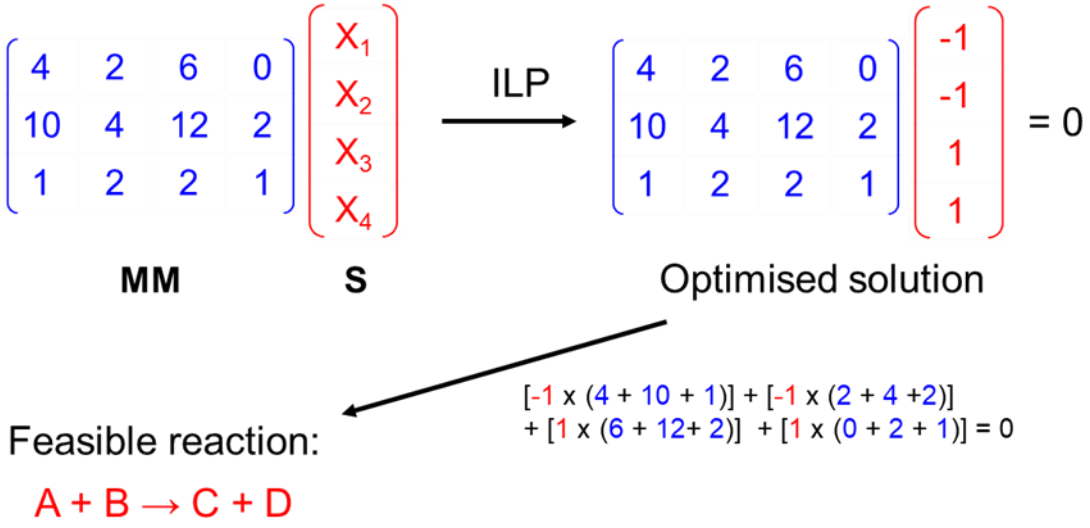


Figure 2.3: The ILP optimisation identifying a feasible solution to the objective function, hence identifying a mass-balance-allowed reaction. Where X_1 = A, X_2 = B, X_3 = C, X_4 = D.

After all feasible reactions are identified, these standalone reactions are themselves potential models. There are also other possibilites for reaction models where there are any combinations of two of the possible reactions, or any combinations of three of the reactions etc. The total number of possible models (η) is equal to the sum of the binomial coefficients (or all combinations)

for every number of reactions in the model up to the total number of possible, identified reactions (δ) - this is shown in eqn 2.4. All of these combinations are then generated and saved to a model database, which will be accessed in the next stage of the approach. For example, if the ILP optimisation identifies 5 possible reactions (δ = 5), the summation of all combinations gives η equal to 31. This example can be shown in a model representation in Figure 2.4, where this could be any appropriate reaction where: SM = starting material, Int1 = intermediate 1, Int2 = intermediate 2, P = product and Imp = impurity. It should also be noted that as all of these forward reactions are mass-balanced, all backward reactions would also be plausible, but are omitted from the representation for concisiveness.

$$\eta = \sum_{i=1,2,3,\ldots,\delta}^{\delta} \frac{\delta!}{i!\,(\delta - i)!} \qquad \text{eqn. 2.4}$$

Where:

- η = total number of models
- δ = number of possible reactions
- i = iterative number of reactions considered for the model

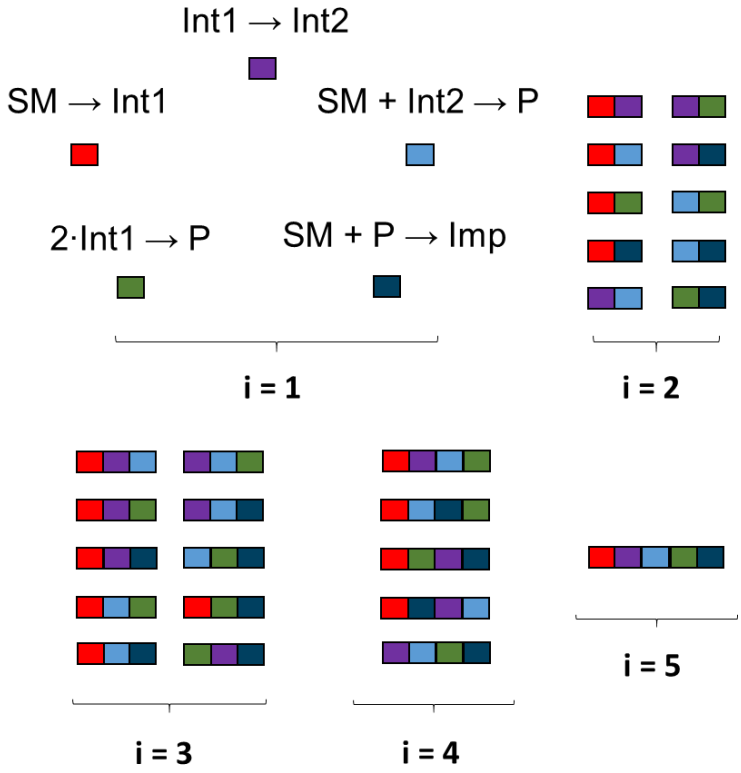


Figure 2.4: A visual representation of all of the possible reaction models when given five mass-balance-allowed sample reactions, each shown as a different coloured block. When i = 1, each reaction is in itself a model, and when i > 1, each reaction behaves as a model fragment. These fragments when combined in different ways provide full and unique reaction models, each of which are to be assessed for their validity with respect to experimental data.

After every possible model has been saved into the model database, each of these models are then automatically transformed into their respective ODEs, which is a form that can be readily used in the kinetic model fitting stage of the approach. The individual stoichiometric reactions that make up each model are compiled into a set of differential equations for each model, where each of the participating species have their own differential equation to show how their concentration changes over time. Figure 2.5 summarises the model generation section of the approach, by using the esterification reaction as an example. This shows how the initial input of the species is first used to identify the feasible mass-balance-allowed reactions, then how these reactions can be ordered into every possible reaction model, then subsequently how these models are transformed into sets of differential equations for use in the kinetic model fitting stage of the approach.

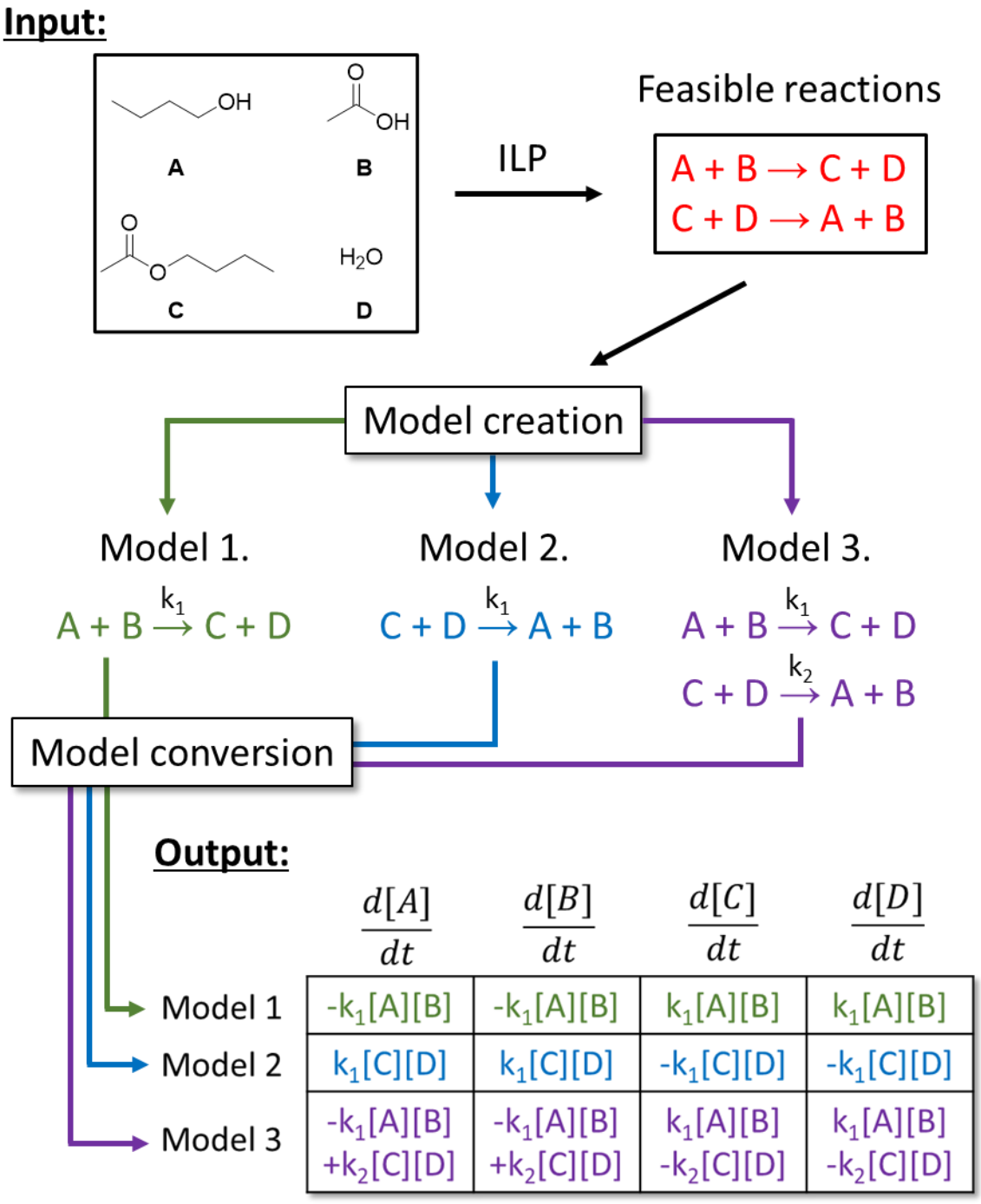


Figure 2.5: A schematic summarising how the model generation stage of the approach progresses. Where the participating species are inputted, and sorted ODE functions that describe all possible mass-balance-allowed reactions are outputted.

2.2.2 Kinetic model fitting

After the model generation stage of the approach is completed and there is a sorted database with all possible models compiled with their respective ODE functions, the next stage of the approach can begin: kinetic model fitting. This

stage proceeds by sequentially loading individual models from the database, then attempting to optimise the kinetic constants for each of the reactions within the model. These optimisations progress based on the simulated ODE results' convergence with experimental data, where more suitable models exhibit a favourable change in ODE convergence based on the kinetic constants provided by the optimiser. Where models do not accurately describe the chemistry taking place in the experimental dataset, the optimiser typically ceases very quickly as there is no way to incur a favourable change in the ODE results as it iterates. After the fitting of a particular model is complete, the kinetic constants (optimised results) as well as the minimised error metric of the ODE convergence (sum of squared error or SSE) are assigned to the model and saved for statistical analysis.

2.2.2.1 The objective function

When a model is initially selected for kinetic fitting, the objective is to use an optimisation algorithm to minimise the SSE between the simulated ODE results and the experimental results. This is shown in Figure 2.6, with a set of time-series data showing a first-order decrease in the concentration of A. The first iteration of the algorithm is shown, where a guess of k = 0.4 is applied to the ODE solver. The 'difference' between the experimental result of data point E_x, and the simulated ODE of data point S_x, is shown as D_x. This value is squared to make the value positive, and summed with all other differences, to give one SSE value: this is shown mathematically in eqn 2.5.

$$SSE = \sum_{x=1,2,3,\ldots} (E_x - S_x)^2 \qquad \text{eqn. 2.5}$$

Where:

- SSE = sum of squared error
- E_x = experimental data point
- S_x = simulated ODE data point

As the algorithm iterates, the aim is to decrease the SSE output by varying the input k values. In this example optimisation shown in Figure 2.6, the optimiser (and hence SSE) reaches a minimum at iteration 11, as the ODE converges very well with the experimental data. This method serves as an indirect route to

identifying kinetic constants, as in this case the k values are the optimised inputs for the optimisation algorithm.

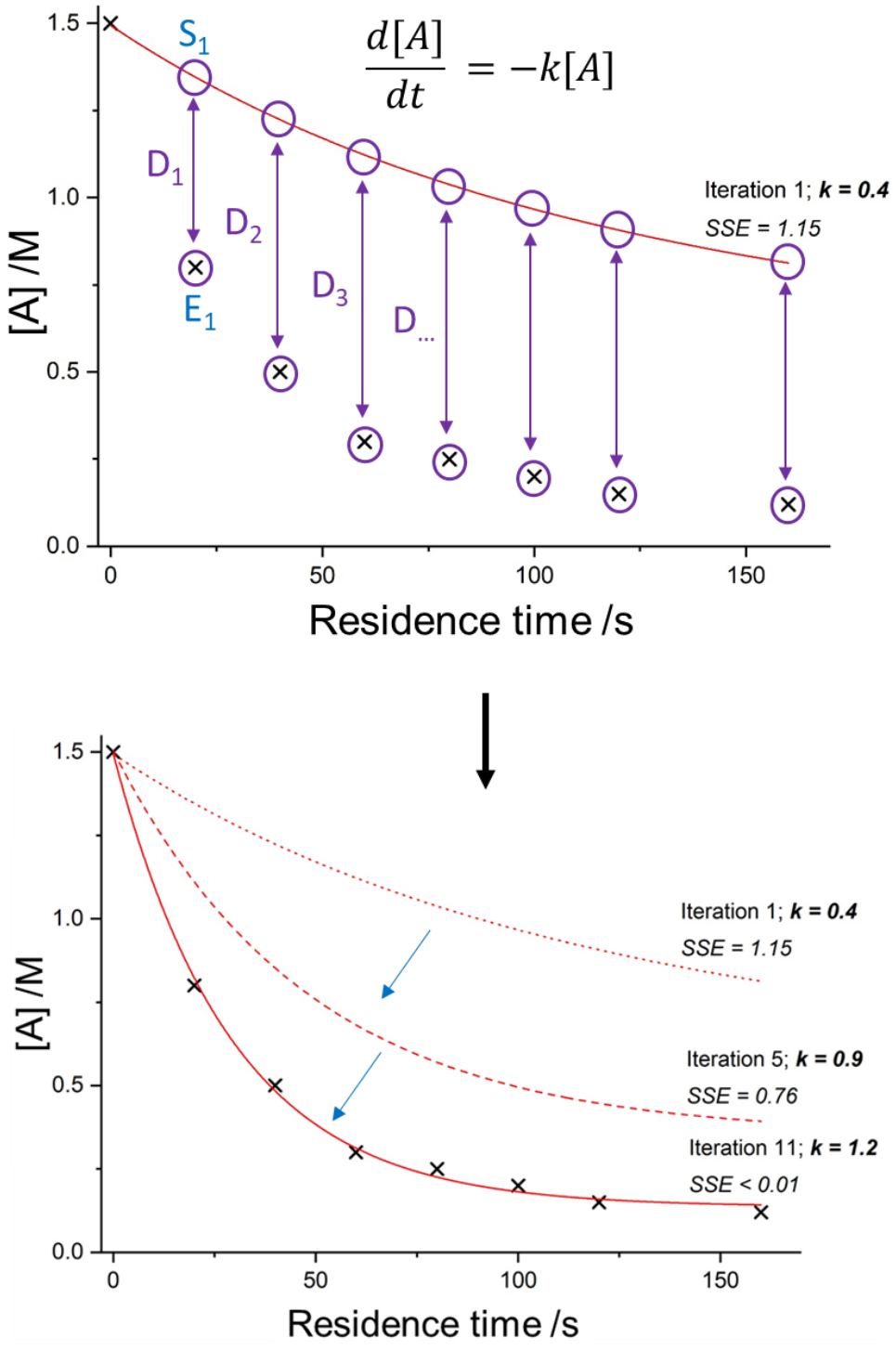


Figure 2.6: The iterations of an optimisation algorithm with respect to the convergence of simulated ODEs with sample data - this occurs via the minimisation of the SSE output.

For each iteration of the algorithm, the inputted k values are used alongside the ODE equations for the current model. The ODE solver simulates how each of the specified initial species' concentrations change over time, given

the reaction rate calculated from the k values, the species concentration at that moment and the reaction model. The ODE solver used for this approach is the MATLAB solver ‘ode15s’. This is a stiff ODE solver, meaning that it can be used to solve ‘stiff’ ODEs. An ODE is stiff when the integration must be completed in unreasonably small increments by typical solvers, that could require millions of evaluations leading to a solver failure. As there are many reaction models that are to be evaluated, it isn’t clear which ODEs would present themselves as stiff problems. Therefore, ode15s was used which can approach curve solutions for both stiff and non-stiff ODEs, as opposed to a non-stiff solver which would be able to solve non-stiff ODEs faster, but fail when stiff ODEs were to be evaluated.

Depending on the nature of the experimental data, different inputs for the optimisation algorithm may also be used. If there is only one data set, or all data sets are at the same temperature, then k values can be optimised. If there are multiple data sets at different temperatures, however, it is often better to optimise directly for the activation energy and reference k value. This is because there are fewer variables for the algorithm to optimise, meaning there is a greater chance of convergence on a global optimum and not a local one, this is discussed further in Chapter 2.2.2.2.

The reference k value is at a particular reference temperature, typically the midpoint of the temperatures explored experimentally. This is to ensure that there are no weightings in the fittings of parameters towards the extremes of the temperatures explored. This reference k value and activation energy value for each of the transformations are optimised, as these inputs generate the k values for each transformation using the re-parameterised Arrhenius equation shown as eqn. 2.6, where the general Arrhenius equation is shown as eqn. 2.7.

$$k = k_{ref} e^{\left[\frac{-E_a}{R}\left(\frac{1}{T}-\frac{1}{T_{ref}}\right)\right]} \qquad \text{eqn. 2.6}$$

$$k = A\, e^{\frac{-E_a}{RT}} \qquad \text{eqn. 2.7}$$

Where:

- k = rate constant
- k_{ref} = reference rate constant at a particular temperature
- E_a = activation energy /kJ mol^{-1}
- A = pre-exponential factor

- T = temperature of reaction /K
- T_{ref} = reference temperature for the reference rate constant /K
- R = ideal gas constant /J mol^{-1} K^{-1}

The re-parameterised Arrhenius equation is used for kinetic fitting as there is only one optimum for the reference k and activation energy values, regardless of the kinetic model. This means that the output of the algorithm can be assumed to be at a discrete, optimised point. This is not the case if optimising for set of values for a pre-exponential factor and activation energy, due to the high correlation between these values.[192, 193] This means that there are several sets of values that can all combine to obtain the same rate constant, which can be avoided when using the re-parameterised equation to fit kinetic parameters to data sets that include multiple temperatures.

The final consideration when compiling the objective function, is the method of measuring the convergence of the simulated ODEs to the experimental data. For all simulated and experimental case studies herein reported, the SSE metric has provided good convergence to time-series data for all species. This could present a problem, however, in cases where particular species occur in only small quantities and SSE does not provide a good fit to the data. This is because the SSE metric reduces the overall error, hence giving a biased fit to species that are present in larger concentrations. This could be prevented by using a different error metric - such as a weighted sum of squared error (WSE) or a relative sum of squared error (RSE). Both of these error metrics serve to normalise the errors in convergence between species present in higher and lower concentrations, and can be used in this approach if the experimental data suggests that it would be appropriate.

2.2.2.2 The optimisation algorithm

When fitting kinetics using this approach, the algorithm used to minimise the error metric described in Chapter 2.2.2.1, and hence maximise the convergence of the simulated ODE with experimental data, is the MATLAB function *fmincon*. fmincon utilises an interior point algorithm, which is described in detail in Chapter 1.3.1. This function has been reported extensively in the literature for the nonlinear optimisation of dynamic models, including considerable uses in the fitting of kinetic parameters to reaction models.[194-196]

fmincon is a local optimisation tool, that leads to fast and accurate convergence on local optima. As the given error metric, SSE, reports a value based on the residuals that is correlated to the magnitude of the concentrations involved, it is not clear based on this value whether true convergence has occurred. However, due to the nature of the problem, it is simple to plot the experimental data and simulated ODE after kinetic fitting, to visually confirm that convergence has occurred. Based on the convergence to the experimental data, it can then be assumed that the optimum found is the global optimum for that model. If non-convergence occurs, then it is either because the given model does not accurately represent the chemistry in the system, or the algorithm became trapped in a local optimum. The likelihood of these scenarios depends on the non-linearity of the data and the number of variables to be optimised.

Generally speaking in these scenarios, it is assumed that when non-convergence occurs, it is because the model does not accurately represent the chemistry. This is because of two reasons. The first is that most kinetic data does not feature sharp inflexions or cliff-edges in which concentration-time data deviates from an otherwise smoothed curvature. It is this curvature in the data that makes local solvers such as fmincon very effective at identifying kinetic parameters, as large aspects of their methodologies involve gradient-based calculations. Therefore, the path that the optimiser takes towards the optimum typically iteratively improves as the optimiser progresses, and with a smoothed parameter space this means that the iterative improvement is unlikely to reach a local optimum. The second reason is that when fitting kinetics, typically the number of variables to be optimised is low, as many models feature chemistry that has fewer than 10 model terms. This means that it is unlikely that fmincon will encounter local optima when optimising and will only find the globally optimum solution.

There is not a discrete reported number of variables to be optimised that means that fmincon will start to encounter local optima issues. Kinetic studies are reported routinely where ~15 or more parameters are optimised using fmincon, both isothermally and across different temperatures, without issues.[197, 198] However, it cannot be assumed *a priori* with certainty that these issues will not occur. Therefore, in cases where there are more than 10 model terms to be optimised, a form of global optimisation called multi-start fmincon (ms-fmincon)

will be used. This methodology employs fmincon, but with multiple starting points from which the optimisation starts. Using ms-fmincon, with 5 randomised starting points, means that from those 5 optimisations the global optimum is very likely to be found when more than 10 parameters are to be optimised. This approach utilises ms-fmincon because it has been proven to be effective in optimising even large scale systems. A recent example of this was reported where 383 parameters were optimised accurately for a large kinetic model.[199]

2.2.3 Statistical analysis

After the kinetic parameters are optimised for each model and the error metric (SSE) recorded, statistical analysis can then be performed to determine the most likely model that describes the experimental data. Akaike's Information Criterion (AIC), eqn. 2.8, provides a relative evaluation of a model for a particular set of data.[200]

$$AIC = n \ln\left(\frac{SSE}{n}\right) + 2k \qquad \text{eqn. 2.8}$$

Where:

- AIC = Akaike's Information Criterion
- n = number of observations
- SSE = sum of squared error
- k = number of model terms

This calculation provides a relative evaluation of balancing the goodness-of-fit of the model to the experimental data, with having as few model terms as possible, where a minimised value is optimal. Therefore, models are more favourable if they only contain terms that have a significant effect on the fit to the data inputted. This means that the optimum model that AIC selects is less likely to be over-trained, as all model terms contribute a significant amount towards the fit of the data.

In the context of this approach, an AIC evaluation can inform the experimenter which reaction model is the most likely to be correct, in relation to other models considered. This model should only contain reactions that significantly contribute towards the convergence of the simulated ODEs with experimental data points. As AIC assumes that the amount of data is sufficiently large in relation to the number of model parameters, it is also beneficial for the

purposes of this approach to use a corrected AIC value to compensate for the reduced sample sizes.

The Corrected Akaike's Information Criterion (AIC_C), eqn. 2.9, both converges to AIC for large sample sizes, but gives a more accurate answer for smaller sample sizes as a greater weighting is placed on the number of model terms.[201] Reported data suggests that an uncorrected AIC value is appropriate only when the ratio of n:k is greater than 40, otherwise an AIC_C evaluation should be used.[202] However, as the model penalisation aspect of AIC_C is favourable to differentiate models with similar SSE values and many of the 'observations' in the context of a kinetic profile can be highly correlated, if the ratio of n:k is less than 40 then the number of observations should be the number of individual experiments rather than individual data points. This is also true where several experiments are used in fitting to contribute towards one SSE value, where each experiment has a vastly different number of measurements within them.

$$AIC_C = AIC + \frac{2k(k+1)}{n-k-2} + n \cdot \ln(2\pi) + n \qquad \text{eqn. 2.9}$$

Where:

- AIC = Akaike's Information Criterion
- AIC_C = Corrected Akaike's Information Criterion
- n = number of observations
- k = number of model terms

2.2.4 Other considerations

There may be some instances where there are intermediates or products formed that cannot be identified. This occurrence can be chemistry-dependant, but it is common that the quantitative analysis of a process can show total mass balance losses as a reaction progresses. This can indicate that either the species calibrations are inaccurate, or there is a real loss of mass/concentration that is unaccounted for. If calibrations are correct with certainty and there is still a loss of mass that is unaccounted for, it can also be beneficial to input an additional species into the system. This will force the approach to identify a new pathway to an unknown species, which can be assigned the total mass loss of the process as the time-series data progresses. Due to the nature of the AIC_C evaluation, it can then be determined statistically if this loss of mass contributes significantly

enough towards the fitting of the data to suggest that there is a chemical pathway occurring that is not described by the user's inputted species. This pathway will be a grouped reaction description with an observed rate constant, k_{obs}. This is because if another pathway is shown to occur, then it is not possible to know how many routes that mass is lost by. Chemical considerations must then be undertaken to identify missing species and the approach must be re-run.

When running the approach, the model generation aspect is very fast, whilst the kinetic fitting stage is significantly slower due to the many optimisations that must be conducted. The kinetic fitting stage is considerably accelerated by utilising parallel computing. This means that each optimisation operation is conducted on a single computer core, meaning that separate tasks can be sent to separate cores within the computer. This means that optimisations are conducted in parallel, as the process speed scales directly with the amount of computer cores available. The time taken for each optimisation is heavily dependent on the model it is fitting and the number of ODE evaluations to be conducted. A typical range for a single optimisation can take from 10 - 120 seconds on a standard Intel i5-2310 processor, however, many optimisations take less than 1 second if the model does not describe the experimental data accurately.

2.2.5 Overview

This computational approach to kinetic model selection and parameter identification was developed in MATLAB by loosely following the methodology reported by Tsu *et al*.[173] The approach first identifies all possible reactions that can happen based on mass balance alone, from the species inputted by the user. All plausible models are then compiled by taking every combination of these reactions, and sequentially optimising the kinetic parameters for each of these models. Each model is then statistically evaluated and ranked, highlighting the most likely model based on the experimental data supplied. The approach outline is summarised in Figure 2.7.

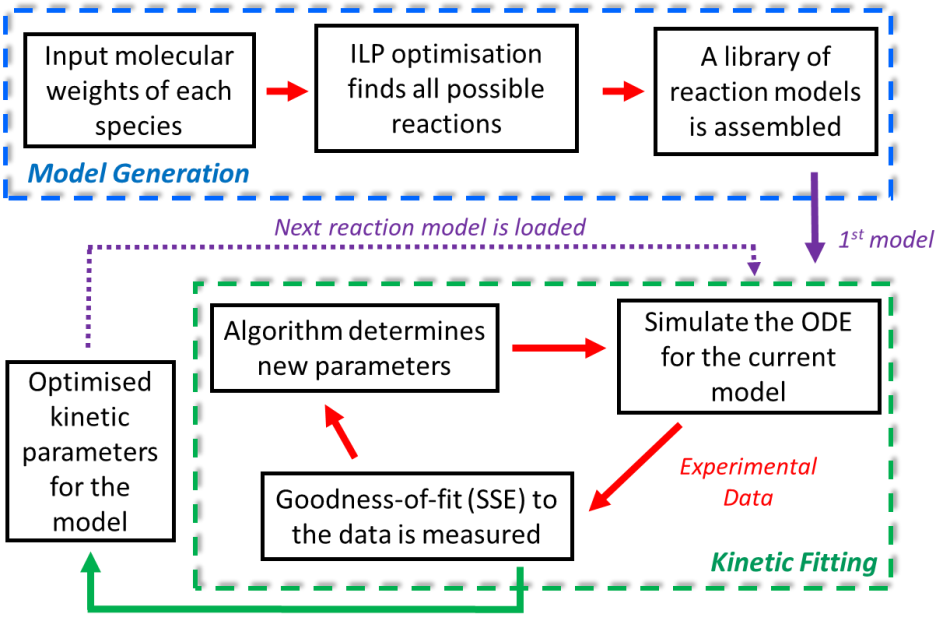


Figure 2.7: An outline of the stages of the computational approach.

This method serves as a comprehensive model evaluator where chemical intuition is removed, thereby removing a chemist's bias and considering a number of possible models that may otherwise not have been considered. It is, however, not the role of this approach to remove all necessity for a physical-organic chemist to supervise the discovery of a kinetic model. This approach is to be used as a complementary tool, whereby all models are evaluated without human effort or interaction, but final considerations are still made by the chemist when the ranked model list is generated. This is because there may be reactions present in the most likely model that describe chemistry that is unlikely or unable to occur, but still give a good fit to the experimental data. It may also be possible that many models give very similar AIC_C evaluations, so many models may be competing.

The benefits to this tool are clear. All models are generated based on logical transformations of the species inputted, and every possibility is comprehensively evaluated without the need for a chemist's or statistician's most important resource: time. Typically a chemist would evaluate a select few models based on chemical intuition, and remove certain chemical possibilities *a priori*, then further refine the model until it is satisfactory. When all possibilities are considered, by an approach that automatically evaluates them, this allows the chemist to work on other projects including more important aspects of laboratory work or theory that cannot be assigned to a machine. However, although

powerful, this tool ultimately is data-driven, and must be used in conjunction with real chemical intuition by an end user to determine that the models identified as most likely are accurate and consistent with the science.

2.3 Simulated verification of the approach

Before any real experimentation takes place, it is important to first verify that the approach works as a means of identifying the correct reaction model and kinetic parameters. Therefore, several simulated case studies were conducted with varying goals to show the adaptability of the approach to different circumstances. These simulated case studies feature reactions from the literature, where the model and kinetic parameters are already identified. Simulated experimental data is generated from this information and used as the basis for verifying the approach.

To simulate the data, first the true model and kinetic parameters reported in a given literature source are inputted into MATLAB. The experiment is then simulated at a particular temperature and set of initial concentrations, by using an ODE solver to evaluate the change in species concentrations over a given experimental timeframe. Individual data points from this timeframe are then extracted, which serve as the experimental data points for the particular simulated case study. To then make the case study more robust, up to 5 % relative error in each measurement is added, to approximate this simulated data more closely to real experimental data. This generation of simulated data is summarised in Figure 2.8.

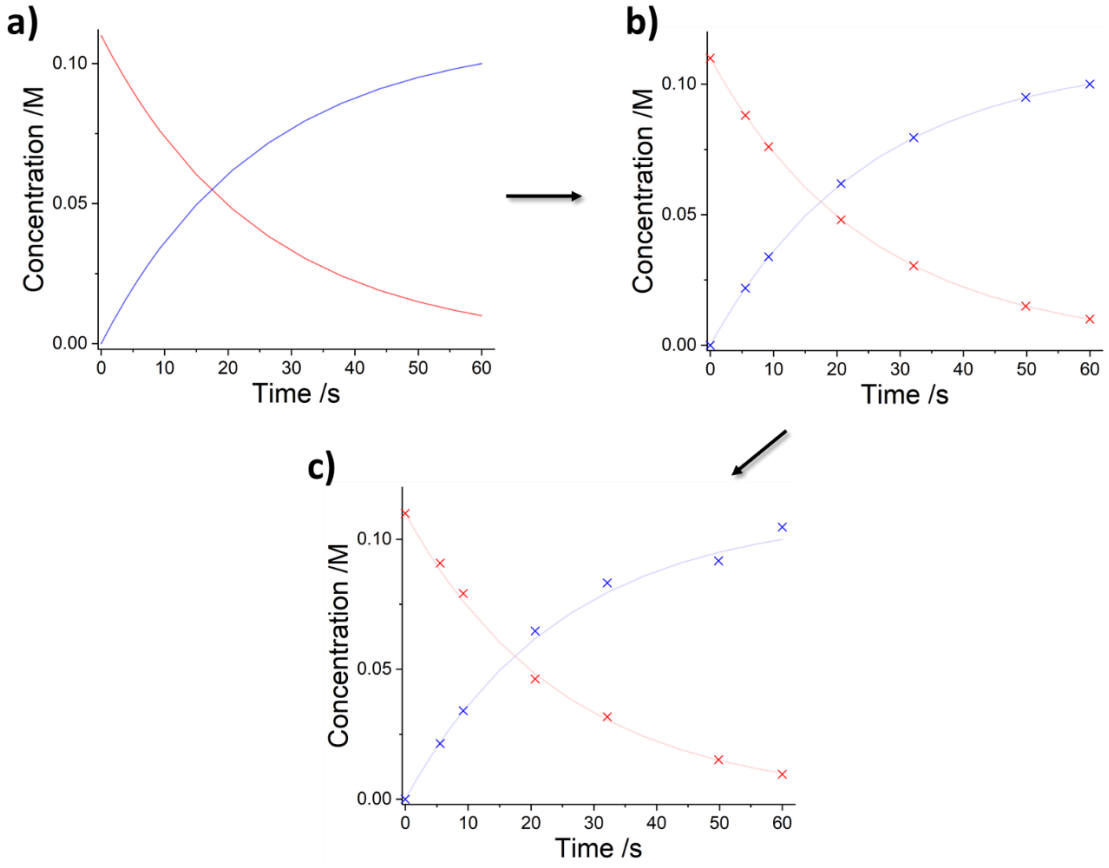
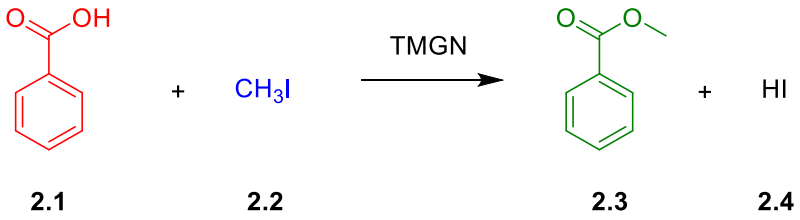


Figure 2.8: An example showing the generation of a simulated data set from the literature. **a)** An ODE is simulated for a particular model with a set of kinetic parameters, in this case A → B, where: — = A, — = B. **b)** Particular time points are taken from this ODE to represent individual measurements, where x = A, x = B. **c)** Up to 5 % relative random error is added to these measurements to more closely resemble real experimental data.

2.3.1 Case study: Benzoic acid alkylation

The first simulated case study was a very basic system, where there are only three species to be considered.[203] The reaction system was benzoic acid, **2.1**, reacting with iodomethane, **2.2**, in the presence of the base, 1,8-bis-(tetramethylguanidino)naphthalene (TMGN), to form the methyl benzoate product, **2.3**, and hydroiodic acid, **2.4**. This reaction is shown in Scheme 2.2. TMGN was present at one equivalent with respect to benzoic acid, and was assumed to deprotonate the benzoic acid so that all of the starting material is in the carboxylate form. There were no rates reported by the authors for this deprotonation, but they indicated that modelling this system as a simple second order reaction under these conditions yielded accurate results.



Scheme 2.2: The reaction of benzoic acid with iodomethane to form methyl benzoate. Modelled as a second order reaction.

2.3.1.1 Data acquisition

The second order reaction model reported by Gholamipour-Shirazi[203] was used to generate the ODEs shown in eqn. 2.10 - 2.13. Three isothermal simulated data sets were generated using these ODEs, with the k value of 0.57 M^{-1} s^{-1}, with the following initial concentrations:

- 0.1 M benzoic acid, **2.1**, 0.08 M iodomethane, **2.2**, 0 M methyl benzoate, **2.3**, 0 M hydroiodic acid, **2.4**.
- 0.1 M benzoic acid, **2.1**, 0.11 M iodomethane, **2.2**, 0 M methyl benzoate, **2.3**, 0 M hydroiodic acid, **2.4**.
- 0.1 M benzoic acid, **2.1**, 0.15 M iodomethane, **2.2**, 0 M methyl benzoate, **2.3**, 0 M hydroiodic acid, **2.4**.

$$\frac{d[\mathbf{2.1}]}{dt} = -k\,[\mathbf{2.1}][\mathbf{2.2}] \qquad \text{eqn. 2.10}$$

$$\frac{d[\mathbf{2.2}]}{dt} = -k\,[\mathbf{2.1}][\mathbf{2.2}] \qquad \text{eqn. 2.11}$$

$$\frac{d[\mathbf{2.3}]}{dt} = k\,[\mathbf{2.1}][\mathbf{2.2}] \qquad \text{eqn. 2.12}$$

$$\frac{d[\mathbf{2.4}]}{dt} = k\,[\mathbf{2.1}][\mathbf{2.2}] \qquad \text{eqn. 2.13}$$

5 % relative error was then added to these simulated data sets, then used as the inputted experimental data for the computational approach. These data sets can be found in Chapter 7.2.1.

2.3.1.2 Results and discussion

Based on the inputs of the molecular weights for this system, there were two mass-balance-allowed reactions, shown in eqn. 2.14 and eqn. 2.15, which are two of the three possible models that can be generated. The third model,

shown in eqn. 2.16, is the combination of these two allowed reactions, resulting in an equilibrium reaction model.

Forward: **2.1** + **2.2** → **2.3** + **2.4** eqn. 2.14

Backward: **2.3** + **2.4** → **2.1** + **2.2** eqn. 2.15

Equilibrium: **2.1** + **2.2** ⇌ **2.3** + **2.4** eqn. 2.16

These three models were then each evaluated based on how well the simulated ODE curves converge to the experimental data inputted. The kinetic parameters, k_x, were optimised for each model, and the SSE was recorded. These results are tabulated in Table 2.1. The concentrations of HI were assumed to not be measured during the kinetic fitting, and were instead inferred.

Table 2.1: A table showing the optimised k values for the identified models in the benzoic acid alkylation case study, alongside each model's SSE and AIC_C.

Reaction	**k_x /M^{-1} s^{-1}**	**SSE /M**	**AIC_C**
Forward	0.5807	0.8101	-6.55
Backward	-	1.0185	-1.74
Equilibrium	0.5817, 0.0009	0.8101	-4.10

The computational approach reveals that the reaction model containing solely the forward reaction gives the minimum overall error and the best (lowest) AIC_C evaluation when assessing the experimental data, indicating that this is the most likely reaction model. For the backward reaction, the minimisation algorithm could not optimise a k value to give a better fit than the initial guess of 1×10^{-3} M^{-1} s^{-1}, indicating that the reaction model is in complete disagreement with the experimental data. For the equilibrium model, the error is as low as the forward reaction model alone as the optimiser assigns a small k value to the reverse reaction, to attempt to fit to the noise of the system. Although this model fits the data equally well, there is the added complexity of a second model term; as this term adds no value in terms of lowering the SSE, it is an unfavourable addition in terms of an AIC_C evaluation which prefers simplistic models, and is therefore considered a less appropriate model than the forward reaction term alone.

This case study has shown that it is possible to use this computational approach to identify the correct model and kinetic parameters for a simple

system, by simply inputting the participating species and their respective changes in time-series data. The k value identified is very similar to the one used for the generation of the simulated data, and varies only because of the 5 % noise added to the system. The fit of the identified model and k value for the data supplied is shown in Figure 2.9, where the data set shown is the middle experiment where the starting concentrations were 0.1 M benzoic acid and 0.11 M iodomethane.

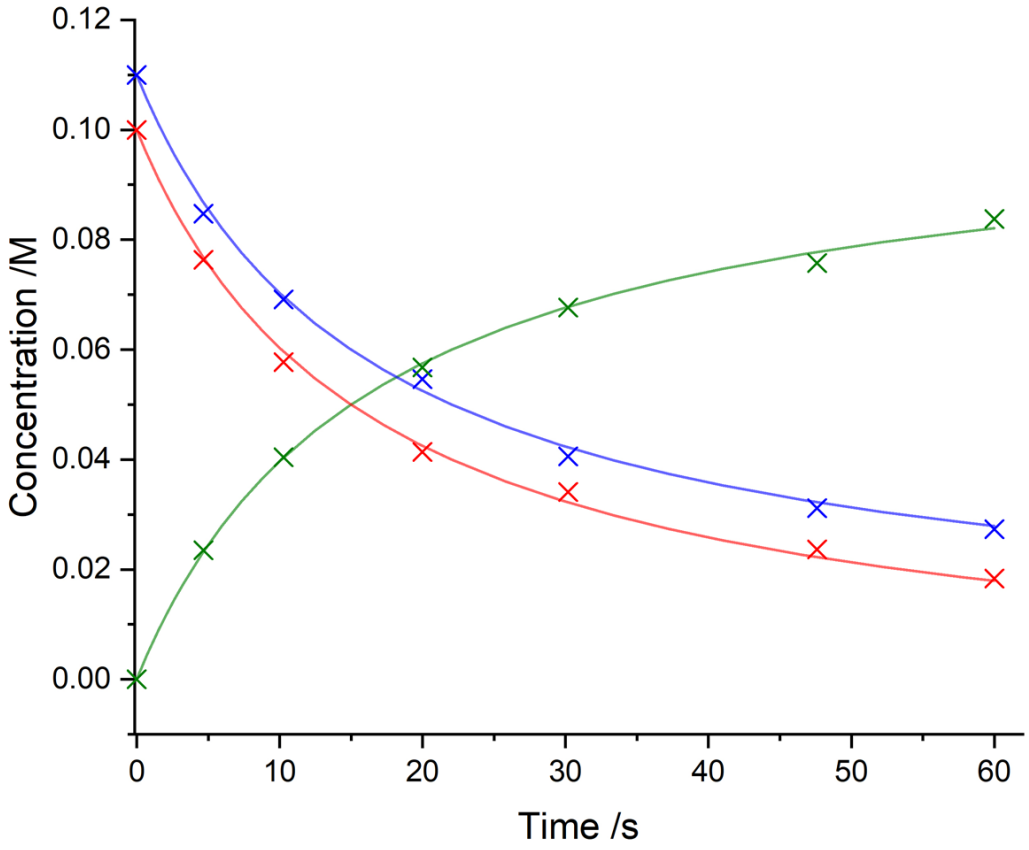
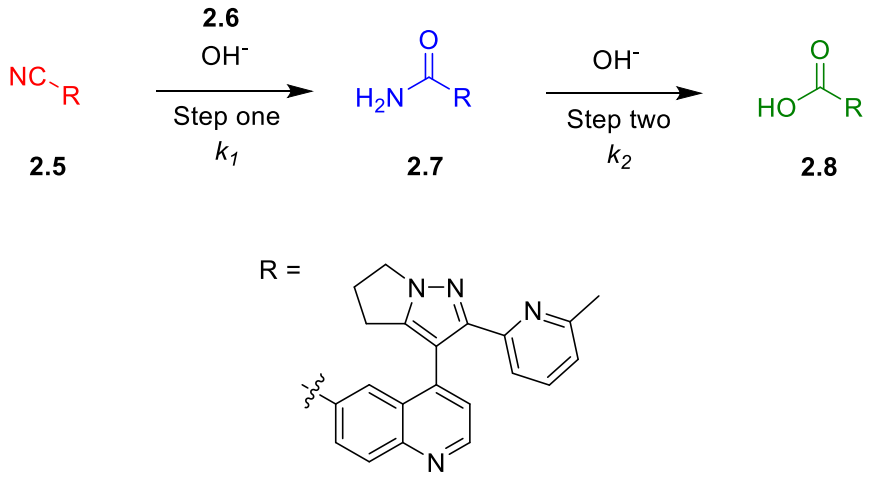


Figure 2.9: The fit of the identified model and kinetic parameters to the generated experimental data with starting concentrations of 0.1 M benzoic acid and 0.11 M iodomethane. Where: x = benzoic acid, x = iodomethane, x = methyl benzoate, — = benzoic acid (ODE), — = iodomethane (ODE), — = methyl benzoate (ODE) .

2.3.2 Case study: Nitrile hydrolysis

This case study features a slightly larger system, where there are 5 species to consider in the hydrolysis of a nitrile.[204] The nitrile of interest, **2.5**, is hydrolysed by hydroxide, **2.6**, to form the amide, **2.7**. This amide is then further hydrolysed to form the carboxylic acid, **2.8**, and ammonia, **2.9**. This reaction is shown in Scheme 2.3. As in many acid/base reactions, there are a lot of fast proton transfer steps, the mass of species **2.6** is therefore added to the system

as water, H_2O, rather than hydroxide, OH^-, in order to preserve mass balance. This allows appropriate reaction models to be generated.



Scheme 2.3: The reaction of a nitrile with hydroxide to form the corresponding amide, which is susceptible to further hydrolysis to form the carboxylic acid. Modelled as sequential second order reactions.

2.3.2.1 Data acquisition

The sequential second order reaction model reported by Niemeier *et al.*[204] was used to generate the ODEs shown in eqn. 2.17 - 2.21. The reported kinetic constants for the two steps were:

- Step one: $k_{75\,°c}$ = 9.27 x 10^{-3} M^{-1} s^{-1}, E_a = 87.1 kJ mol^{-1}
- Step two: $k_{75\,°c}$ = 3.63 x 10^{-5} M^{-1} s^{-1}, E_a = 74.5 kJ mol^{-1}

Four simulated data sets were generated at four separate temperatures, by using eqn. 2.6 to calculate k values, with the following initial concentrations:

- 0.8 M nitrile, **2.5**, 1.8 M hydroxide, **2.6**, 0 M amide, **2.7**, 0 M carboxylic acid, **2.8**, 0 M ammonia, **2.9**, 60 °C.
- 0.8 M nitrile, **2.5**, 1.8 M hydroxide, **2.6**, 0 M amide, **2.7**, 0 M carboxylic acid, **2.8**, 0 M ammonia, **2.9**, 70 °C.
- 1 M nitrile, **2.5**, 2 M hydroxide, **2.6**, 0 M amide, **2.7**, 0 M carboxylic acid, **2.8**, 0 M ammonia, **2.9**, 80 °C.
- 1 M nitrile, **2.5**, 2 M hydroxide, **2.6**, 0 M amide, **2.7**, 0 M carboxylic acid, **2.8**, 0 M ammonia, **2.9**, 90 °C.

$$\frac{d[\mathbf{2.5}]}{dt} = -k_1[\mathbf{2.5}][\mathbf{2.6}] \qquad \text{eqn. 2.17}$$

$$\frac{d[\mathbf{2.6}]}{dt} = -k_1[\mathbf{2.5}][\mathbf{2.6}] - k_2[\mathbf{2.7}][\mathbf{2.6}] \quad \text{eqn. 2.18}$$

$$\frac{d[\mathbf{2.7}]}{dt} = k_1[\mathbf{2.5}][\mathbf{2.6}] - k_2[\mathbf{2.7}][\mathbf{2.6}] \quad \text{eqn. 2.19}$$

$$\frac{d[\mathbf{2.8}]}{dt} = k_2[\mathbf{2.7}][\mathbf{2.6}] \quad \text{eqn. 2.20}$$

$$\frac{d[\mathbf{2.9}]}{dt} = k_2[\mathbf{2.7}][\mathbf{2.6}] \quad \text{eqn. 2.21}$$

5 % relative error was then added to these simulated data sets, then used as the inputted experimental data for the computational approach. These data sets can be found in Chapter 7.2.2.

2.3.2.2 Results and discussion

Based on the inputs of the molecular weights for this system, there were four mass-balance-allowed reactions, shown in eqn. 2.22 - 2.25. All possible models were then compiled, subject to eqn. 2.4, resulting in 15 models.

$$\mathbf{2.5} + \mathbf{2.6} \rightarrow \mathbf{2.7} \quad \text{eqn. 2.22}$$

$$\mathbf{2.7} \rightarrow \mathbf{2.5} + \mathbf{2.6} \quad \text{eqn. 2.23}$$

$$\mathbf{2.6} + \mathbf{2.7} \rightarrow \mathbf{2.8} + \mathbf{2.9} \quad \text{eqn. 2.24}$$

$$\mathbf{2.8} + \mathbf{2.9} \rightarrow \mathbf{2.6} + \mathbf{2.7} \quad \text{eqn. 2.25}$$

All 15 models were evaluated sequentially, then ranked based on their AIC_C. It was found that the highest ranked model was the correct model, with kinetic parameters optimised to be very close to the real values. As artificial error was added to the results, this optimisation is assumed to be correct as the relative error in the kinetic parameters is negligible. The top three performing models are shown in Table 2.2, where the reference k values at 75 °C are shown, as well as the activation energy, SSE and AIC_C.

Table 2.2: A table showing the top three ranked models for the nitrile hydrolysis case study. The kinetic parameters for each reaction within the model is shown, as well as the SSE and AIC_C evaluation.

Model rank	Model	$k_{x,75\,°C}$ /M^{-1} s^{-1} or s^{-1}	E_a /kJ mol^{-1}	SSE /M	AIC_C
1	2.5 + 2.6 → 2.7	9.23 x 10^{-3}	86.7	0.0179	-162.3

	2.6 + 2.7 → 2.8 + 2.9	4.49 x 10^{-5}	71.6		
2	2.5 + 2.6 → 2.7 2.6 + 2.7 → 2.8 + 2.9 2.8 + 2.9 → 2.6 + 2.7	9.28 x 10^{-3} 4.30 x 10^{-5} 1.94 x 10^{-4}	86.7 75.5 26.3	0.0179	-156.7
3	2.5 + 2.6 → 2.7 2.7 → 2.5 + 2.6 2.6 + 2.7 → 2.8 + 2.9	9.32 x 10^{-3} 3.63 x 10^{-5} 4.20 x 10^{-5}	86.5 37.3 74.8	0.0180	-156.6

This case study has shown that it is possible to identify the correct model and kinetic parameters for a multistep chemical system using the computational approach. Interestingly, this study has shown again that the overall error is very similar in the top-ranked competing models, meaning that the deciding factor then becomes the number of model terms, due to the AIC_C evaluation. The data inputted was at different temperatures and the kinetic parameter optimisation proceeded smoothly when optimising reference k values and activation energies directly. An example kinetic plot showing the fit of the identified model to the experimental data is shown in Figure 2.10 - this is the fourth experiment where the starting concentrations were 1 M nitrile and 2 M hydroxide at 90 °C.

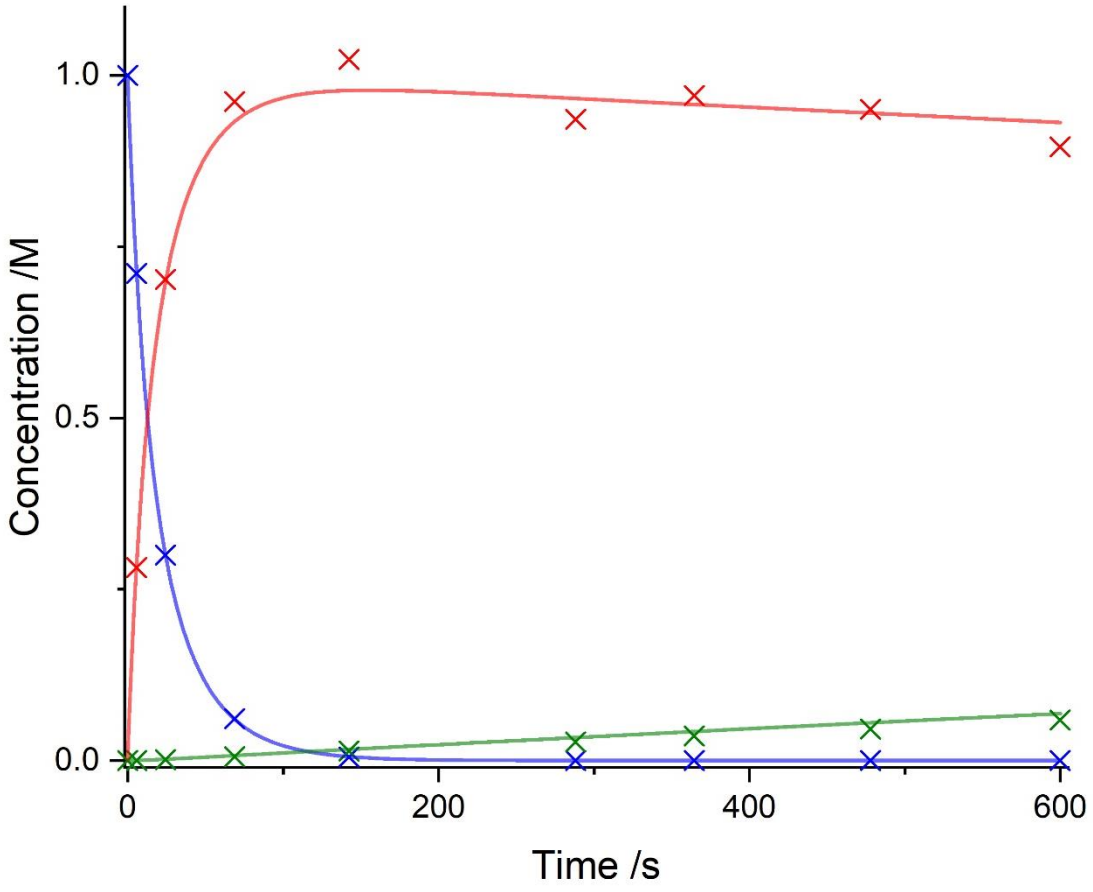
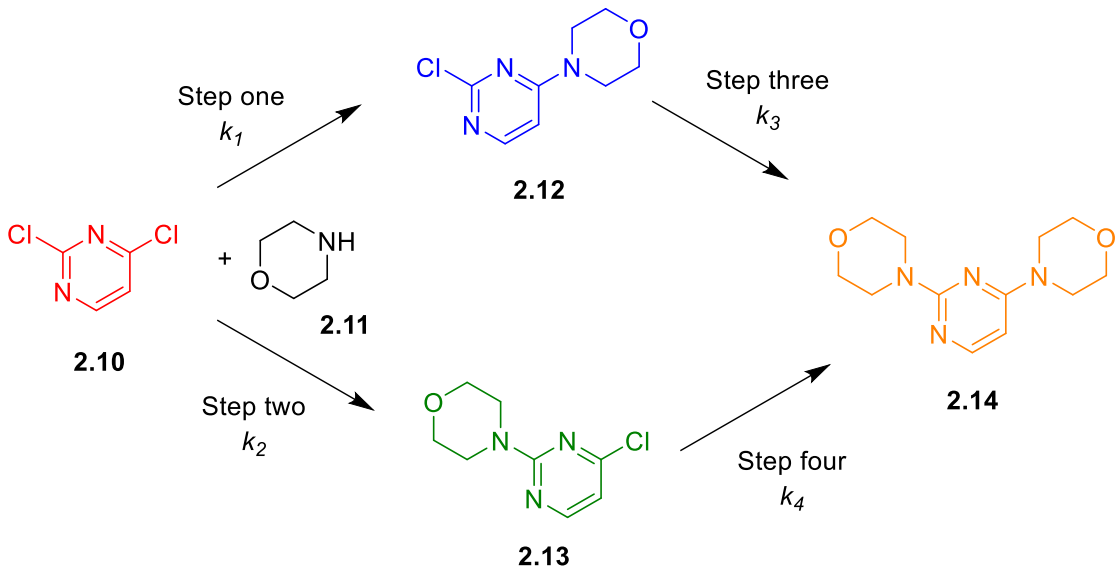


Figure 2.10: The fit of the identified model and kinetic parameters to the generated experimental data with starting concentrations of 1 M nitrile and 2 M hydroxide at 90 °C. Where: x = nitrile, x = amide, x = carboxylic acid, — = nitrile (ODE), — = amide (ODE), — = carboxylic acid (ODE) .

2.3.3 Case study: S_NAr kinetics

The final simulated case study involves a larger scale multistep system, where there are 6 species in a S_NAr chemical setting. 2,4-dichloropyrimidine (starting material, SM), **2.10**, reacts with morpholine, **2.11**, to form either the 4-substituted product, **2.12**, or the 2-substituted product, **2.13**. These products can then further react with another equivalent of morpholine to produce the bis-substituted product, **2.14**. Each of these reactions form hydrochloric acid, **2.15**, as a byproduct that is neutralised by an excess of base. This system is shown in Scheme 2.4.



Scheme 2.4: The reaction of 2,4-dichloropyrimidine, **2.10**, with morpholine, **2.11**, to form the 4-substituted product, **2.12**, and the 2-substituted product, **2.13**, and the subsequent bis-substituted product, **2.14**.

2.3.3.1 Data acquisition

The multistep S_NAr reaction model reported by Reizman and Jensen consists of 4 second-order reactions. These reactions were used to generate the ODEs shown in eqn. 2.26 - 2.31. The reported kinetic constants for these three steps were:

- Step one: log(A) = 3.4 M^{-1} s^{-1}, E_a = 27.0 kJ mol^{-1}
- Step two: log(A) = 3.5 M^{-1} s^{-1}, E_a = 32.1 kJ mol^{-1}
- Step three: log(A) = 4.9 M^{-1} s^{-1}, E_a = 60.0 kJ mol^{-1}
- Step four: log(A) = 3.0 M^{-1} s^{-1}, E_a = 45.0 kJ mol^{-1}

Three simulated data sets were generated at 90 °C, by using eqn. 2.7 to calculate k values, with the following initial concentrations:

- 1 M SM, **2.10**, 2.2 M morpholine, **2.11**, 0 M 4-substituted product, **2.12**, 0 M 2-substituted product, **2.13**, 0 M bis-substituted product, **2.14**, and 0 M hydrochloric acid, **2.15**.
- 1.2 M SM, **2.10**, 2.8 M morpholine, **2.11**, 0 M 4-substituted product, **2.12**, 0 M 2-substituted product, **2.13**, 0 M bis-substituted product, **2.14**, and 0 M hydrochloric acid, **2.15**.
- 0.9 M SM, **2.10**, 3.2 M morpholine, **2.11**, 0 M 4-substituted product, **2.12**, 0 M 2-substituted product, **2.13**, 0 M bis-substituted product, **2.14**, and 0 M hydrochloric acid, **2.15**.

$$\frac{d[\mathbf{2.10}]}{dt} = -k_1[\mathbf{2.10}][\mathbf{2.11}] - k_2[\mathbf{2.10}][\mathbf{2.11}] \quad \text{eqn. 2.26}$$

$$\frac{d[\mathbf{2.11}]}{dt} = -k_1[\mathbf{2.10}][\mathbf{2.11}] - k_2[\mathbf{2.10}][\mathbf{2.11}] - k_3[\mathbf{2.12}][\mathbf{2.11}] - k_4[\mathbf{2.13}][\mathbf{2.11}] \quad \text{eqn. 2.27}$$

$$\frac{d[\mathbf{2.12}]}{dt} = k_1[\mathbf{2.10}][\mathbf{2.11}] - k_3[\mathbf{2.12}][\mathbf{2.11}] \quad \text{eqn. 2.28}$$

$$\frac{d[\mathbf{2.13}]}{dt} = k_2[\mathbf{2.10}][\mathbf{2.11}] - k_4[\mathbf{2.13}][\mathbf{2.11}] \quad \text{eqn. 2.29}$$

$$\frac{d[\mathbf{2.14}]}{dt} = k_3[\mathbf{2.12}][\mathbf{2.11}] + k_4[\mathbf{2.13}][\mathbf{2.11}] \quad \text{eqn. 2.30}$$

$$\frac{d[\mathbf{2.15}]}{dt} = k_1[\mathbf{2.10}][\mathbf{2.11}] + k_2[\mathbf{2.10}][\mathbf{2.11}] + k_3[\mathbf{2.12}][\mathbf{2.11}] + k_4[\mathbf{2.13}][\mathbf{2.11}] \quad \text{eqn. 2.31}$$

2 % relative error was then added to these simulated data sets, then used as the inputted experimental data for the computational approach. These data sets can be found in Chapter 7.2.3.

2.3.3.2 Results and discussion

From these 6 species, there are 16 mass-balance-allowed reactions that can be identified, shown in eqn. 2.32 - 2.47. All possible reaction models were generated, subject to eqn. 2.4, which resulted in 65535 unique models to be evaluated by the approach.

$$\mathbf{2.13} \rightarrow \mathbf{2.12} \quad \text{eqn. 2.32}$$

$$\mathbf{2.12} \rightarrow \mathbf{2.13} \quad \text{eqn. 2.33}$$

$$\mathbf{2.10} + \mathbf{2.11} \rightarrow \mathbf{2.12} + \mathbf{2.15} \quad \text{eqn. 2.34}$$

$$\mathbf{2.12} + \mathbf{2.15} \rightarrow \mathbf{2.10} + \mathbf{2.11} \quad \text{eqn. 2.35}$$

$$\mathbf{2.10} + \mathbf{2.14} \rightarrow \mathbf{2.12} + \mathbf{2.12} \quad \text{eqn. 2.36}$$

$$\mathbf{2.10} + \mathbf{2.11} \rightarrow \mathbf{2.13} + \mathbf{2.15} \quad \text{eqn. 2.37}$$

$$\mathbf{2.11} + \mathbf{2.12} \rightarrow \mathbf{2.14} + \mathbf{2.15} \quad \text{eqn. 2.38}$$

2.12 + **2.12** → **2.10** + **2.14** eqn. 2.39

2.14 + **2.15** → **2.11** + **2.13** eqn. 2.40

2.13 + **2.15** → **2.10** + **2.11** eqn. 2.41

2.11 + **2.13** → **2.14** + **2.15** eqn. 2.42

2.10 + **2.14** → **2.12** + **2.13** eqn. 2.43

2.10 + **2.14** → **2.13** + **2.13** eqn. 2.44

2.14 + **2.15** → **2.11** + **2.12** eqn. 2.45

2.12 + **2.13** → **2.10** + **2.14** eqn. 2.46

2.13 + **2.13** → **2.10** + **2.14** eqn. 2.47

All 65535 models were evaluated by the computational approach, which incurred a computation time of 52 hours (see Chapter 7.6.1 for details on computer specifications). Each of these models were ranked based on their AIC_C, and it was found that the highest ranked model was also the correct model, with kinetic parameters optimised to values that were very close to the generated values. As in the nitrile hydrolysis case study, Chapter 2.3.2, this optimisation is assumed to be correct as the error in the fitted parameters is negligible, considering that artificial error was added to the dataset. The top three performing models are shown in Table 2.3, where the optimised k values are shown, as well as their corresponding SSE and AIC_C values.

Table 2.3: A table showing the top three ranked models for the S_NAr case study. The kinetic parameters for each reaction within the model is shown, as well as the SSE and AIC_C evaluation.

Model rank	Model	k_x /M^{-1} s^{-1}	SSE /M	AIC_C
1	2.10 + 2.11 → 2.12 + 2.15 2.10 + 2.11 → 2.13 + 2.15 2.11+ 2.12 → 2.14 + 2.15 2.11 + 2.13 → 2.14 + 2.15	0.32945 0.07686 0.00018 0.00037	0.0024	-211.2
2	2.10 + 2.11 → 2.12 + 2.15 2.10 + 2.11 → 2.13 + 2.15 2.11+ 2.12 → 2.14 + 2.15	0.32948 0.07679 0.00018	0.0024	-208.6

	2.11 + 2.13 → 2.14 + 2.15 2.13 + 2.15 → 2.10 + 2.11	0.00033 0.00004		
3	2.10 + 2.11 → 2.12 + 2.15 2.10 + 2.11 → 2.13 + 2.15 2.11+ 2.12 → 2.14 + 2.15 2.11 + 2.13 → 2.14 + 2.15 2.10 + 2.14 → 2.13 + 2.13	0.32949 0.07672 0.00018 0.00036 0.15890	0.0024	-208.5

This final simulated case study has shown that it is still possible to identify the correct reaction model and kinetic parameters for a more complicated multistep chemical system, even when there are several thousand models to consider. Many of these models are very similar and the top-ranked competing models contain the four model terms that are appropriate, as well as typically one other model. These models are top-ranking because the additional terms are optimised by the algorithms to either have a negligible effect on the output, or to minimise the error in the kinetic fitting to a certain extent. It is then the role of the AIC_C evaluation to determine whether these additional model terms are producing a significant effect in order for their consideration in the correct reaction model. In this case, the SSE value was unchanged among the top-ranked competing models, meaning that the deciding factor then became the number of model terms. An example kinetic plot showing the fit of the identified model to the experimental data is shown in Figure 2.11 - this is the first experiment where the starting concentrations were 1 M 2,4-dichloropyrimidine and 2.2 M morpholine at 90 °C.

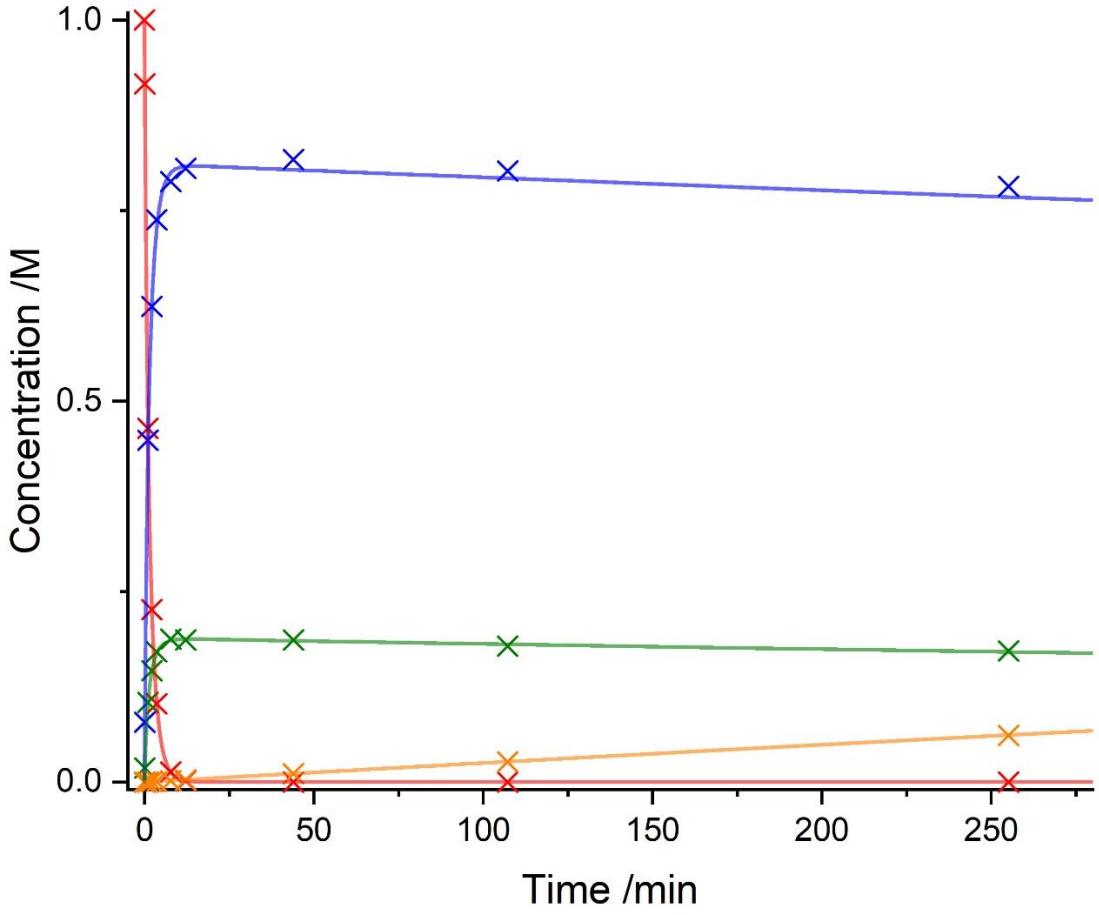


Figure 2.11: The fit of the identified model and kinetic parameters to the generated experimental data with starting concentrations of 1 M 2,4-dichloropyrimidine and 2.2 M morpholine at 90 °C. Where: x = 2,4-dichloropyrimidine, x = 4-substituted product, x = 2-substituted product, x = bis-substituted product, — = 2,4-dichloropyrimidine (ODE), — = 4-substituted product (ODE), — = 2-substituted product (ODE), — = bis-substituted product. Graph is only shown to 260 minutes to show curvature of the initial data points.

2.4 Conclusion

It has been shown that a computational approach to kinetic modelling, loosely based on work by Tsu *et al.*, can be programmed in MATLAB. This approach takes chemical species information, as well as experimental datasets, to identify both the correct reaction model and kinetic parameters for the chemical process by utilising optimisation algorithms and statistical analysis. This approach has been proven to work successfully for three simulated case studies of varying optimisation difficulty, with different sets of experimental data inputs with varying artificial errors. The success shown from these studies inspires confidence that the approach can be implemented with real experimental data, and be used as an effective tool in process development to automatically identify the correct model and kinetic parameters of various

chemical systems. The next step is to prove the efficiency of this approach alongside real experimentation.

Chapter 3 : Experimental applications of the computational approach to kinetic model and parameter determination

3.1 Introduction

With the development of the computational approach completed and verified with multiple simulated case studies, the viability of the approach on real experimental data must also be tested. This experimental application can utilise data from any source to determine the most likely reaction model and kinetic parameters, but for reasons that shall be discussed further, all experimentation for this chapter was conducted using an automated continuous flow reactor platform.

There is a long history of conducting and analysing chemical and enzymatic kinetics using flow techniques, with the first reported case published around 100 years ago by Hartridge *et al.* as a means to study very fast reactions.[205] During this time, many different processes, analytical methods and experimental setups have been reported, here citing only a few.[206-209] A widely regarded essential requirement for kinetic experiments, however, is for the reaction system to operate in a turbulent flow regime.[210] Turbulent flow leads to fast, continuous mixing and can be predicted by calculating the Reynolds number, Re, in eqn. 3.1.

$$Re = \frac{\nu d \rho}{\eta} \qquad \text{eqn. 3.1}$$

Where:

- Re = Reynolds number
- ν = average flow velocity
- d = tube diameter
- ρ = fluid density
- η = fluid viscosity

Using the unitless Reynolds number, the flow regime can be predicted to be turbulent[210] when Re exceeds 2000, or laminar[211] when Re is less than 2000. These regimes are illustrated in Figure 3.1a and 3.1b respectively. Under laminar flow conditions, there is a parabolic velocity profile - meaning that the flow velocity at the centre of the tube is double the average velocity due to friction on the walls of the tubing known as dispersion. This dispersion was originally thought to blur the time axis leading to a substantial distortion of the observed kinetics, and hence meaningful kinetic experiments under these conditions might be impossible.[212]

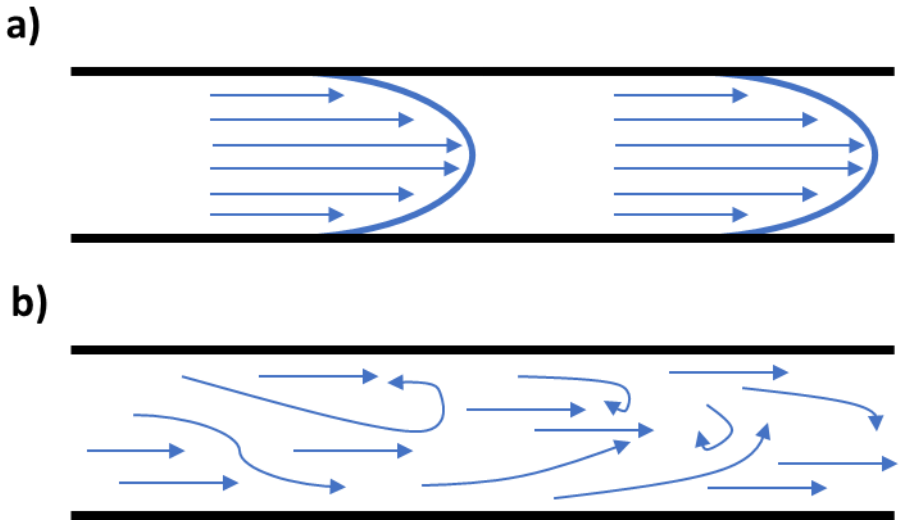


Figure 3.1: A diagram to show the direction and flows within a flow regime, where: **a)** laminar flow, **b)** turbulent flow.

Interestingly, these assumptions remained untested until a number of reported kinetics works from the Douglas group showed that accurate kinetic experiments could be conducted under laminar flow conditions.[213-215] The Reynolds numbers for the experiments conducted ranged between 2.8 and 8.5, implying laminar flow. However, the kinetics monitored in these systems agreed very well with results obtained conventionally by stopped-flow methods. Further work suggests that molecular diffusion has a very significant role, and under certain circumstances it can be a good approximation to neglect the effects of laminar flow i.e. for particular flow velocities with given internal diameters.[212, 216] With this approximation, the kinetics can then be analysed as if the tubular flow was homogeneous, which can lead to accurate kinetic studies even in laminar flow regimes.

In many continuous flow settings for micro- or meso-flow volumes, it is typically simpler to adopt an idealised plug flow reactor model. This model states that each infinitely thin section of flow, known as a plug, travels in the axial direction of the reactor and is perfectly mixed in the radial direction only, where a uniform distribution of the reactor concentrations occurs. This means that the residence time of the plug is a direct function of the length of the tubular reactor and the velocity of the fluid. This is shown in Figure 3.2. This model is commonly employed to simplify tubular mixing and remains an accurate approximation for kinetic measurements.[37] All kinetic experiments by other authors covered in this introduction employ this model, and further experimentation as part of this project also utilises this model - this is discussed further in Chapter 3.2.

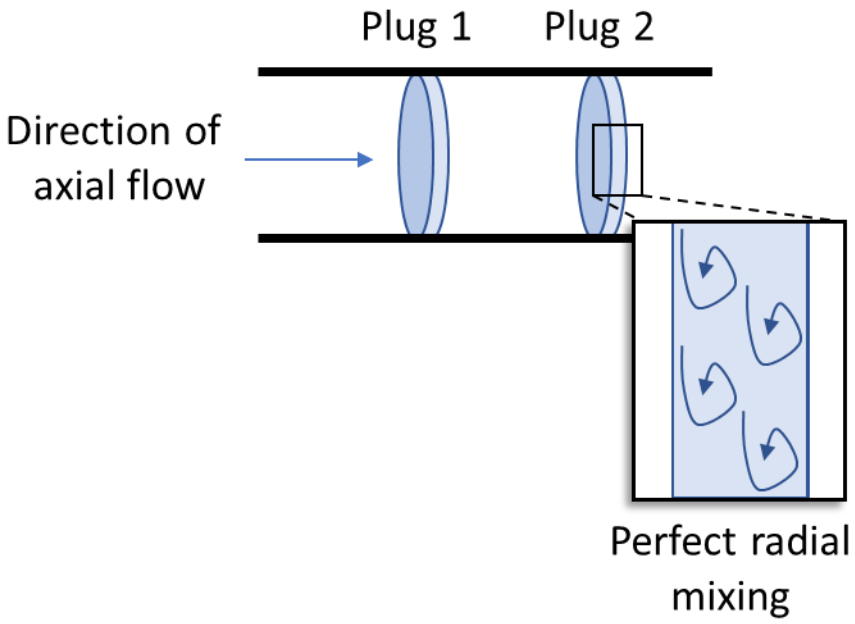


Figure 3.2: A plug flow reactor model, where there is perfect mixing in the radial direction but no forward or backward mixing in the axial direction. Plugs 1 and 2 are examples of the infinitely short plugs existing within this reactor model.

Many of the advantages of continuous flow chemistry, as previously stated in Chapter 1.1, are also particularly attractive for running kinetic experiments. Increased heat and mass transfer[217, 218] ensures that the reaction is well controlled, meaning that it can be asserted with confidence that the experimental conditions applied to it are truly experienced by the reaction medium. Precise reagent control arises out of utilising flow chemistry, as specific flow rates allow accurate addition of chemicals at various points within the reactor system[22, 23, 35] - this is very important when deducing the order of the species

within a system, as reagent stoichiometry is present in exact quantities. Coupling of flow reactor platforms with quantitative analysis, such as UV, HPLC or GC, also unlocks automation capabilities,[6, 219] where multiple experiments can be queued for the system to run autonomously. The increased safety,[220] the efficiency of operating at reaction temperatures above the boiling point of the solvent[221, 222] and the ability to run and analyse extremely fast reactions,[223, 224] means that flow experiments can be run that would otherwise be very difficult or impossible to run in a batch system. For these reasons, it is more desirable in particular circumstances to run kinetic experiments in flow, rather than more traditional batch kinetic experiments, whereby sampling occurs throughout the progression of a single reaction. These advantages will be exploited throughout these experimental acquisitions of data.

The main disadvantage to kinetic modelling in flow systems, however, is the necessity of the system to reach ‘steady-state’. Steady-state simply refers to the state in which the reactor is in, where the responses from the system are a direct consequence of the conditions applied to it. This is an important distinction because as the reaction parameters are changed, a certain amount of time is then required for those reaction parameters to be applied to the system to achieve a consistent output.[225] This can be very wasteful, as a system typically needs approximately 1.5 - 3 reactor volumes of reagent in the acclimation to steady-state, prior to each measurement.[226] This wastes a lot of precious reaction material, as well as time, as the time taken to reach steady-state is in direct relation to the residence time required i.e. a 15 minute residence time results in up to a 45 minute wait for steady-state.

Kinetic experiments in flow have been performed in this way for a long time, as the advantageous properties of continuous flow chemistry were worth the wastage from the steady-state measurements. However, in recent years deviations from these conventional techniques have become more popular, that take advantage of the transitionary period between two steady-state measurements. This is because during this period, transient reaction information is available but is otherwise lost because of traditional sampling methods. However, if the pump flow rates are manipulated to structure this transient data in a way that can be translated to regular time-series data, an entire reaction profile can be mapped very quickly and efficiently. This then negates the need

for multiple steady-state measurements and rectifies this flow chemistry disadvantage, as kinetic profiles can be obtained with minimal material.

The first transient flow method reported by Mozharov *et al.*[96] is an instantaneous step change in the flow rates, where the reaction initially takes place at a low flow rate, F_1, and then the flow rate is increased by an order or magnitude to F_2. This high flow rate then pushes out the transient data profile of the reaction medium, whilst the liquid is monitored by an in-line analysis technique sensitive enough to detect a change in product concentration, for example Raman spectroscopy as reported by the authors. Using this in-line technique, many measurements are taken in short time intervals to generate a reaction profile, from which kinetic information can be derived. This is possible as the magnitude of F_2, the experimental times of the analytical measurements (t_n) and the dimensions of the flow path in the system are known, therefore the reaction profile along the microreactor capillary can be recreated and plotted. This concept is depicted in Figure 3.3, with the corresponding equation shown in eqn. 3.2 to convert the species concentration in experimental time to the species concentration in residence time. This conversion then maps the kinetic profile of the chemical process, between the calculated residence times.

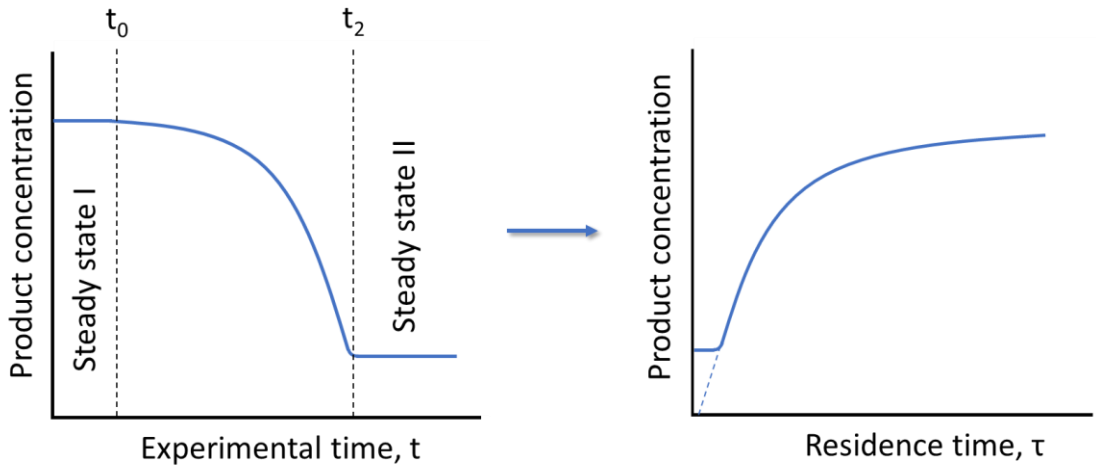


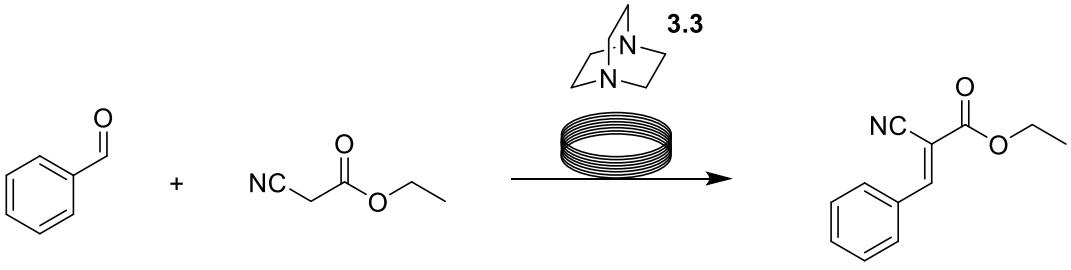
Figure 3.3: A depiction of how a kinetic experiment can be run utilising a step change in flow rates between two steady-states (I and II), allowing time-series data to be plotted from transient data. Figure adapted from Mozharov *et al.*[96]

$$\tau = t_2 \frac{F_2}{F_1} - t \frac{F_2 - F_1}{F_1} \qquad \text{eqn. 3.2}$$

Where:

- τ = residence time of the reaction medium
- t = experimental time experienced by the reactor
- t_2 = experimental time taken to reach steady-state
- F_1 = initial flow rate
- F_2 = higher flow rate that the pumps are instantaneously changed to

This concept was also used experimentally as a means to test the accuracy of the technique in relation to conventional steady-state measurements. The base catalysed Knoevenagel condensation between benzaldehyde, **3.1**, and ethyl cyanoacetate, **3.2**, alongside the base, **3.3**, to form the corresponding adduct, **3.4**, is shown in Scheme 3.1. The corresponding kinetic profiles were generated by conventional, sequential steady-state measurements in Figure 3.4, and by the described step-change methodology in Figure 3.5, where the reaction proceeded at a low flow rate before a sudden step change to a higher flow rate. It is shown that the two sets of time-series data are very comparable, however, the step-change methodology has many advantages. This technique can determine a kinetic profile from only one experiment, with a much lower consumption of reagents and significantly reduced experimental time.



3.1 **3.2** **3.4**

Scheme 3.1: The base catalysed Knoevenagel condensation between benzaldehyde and ethyl cyanoacetate to form **3.4**.

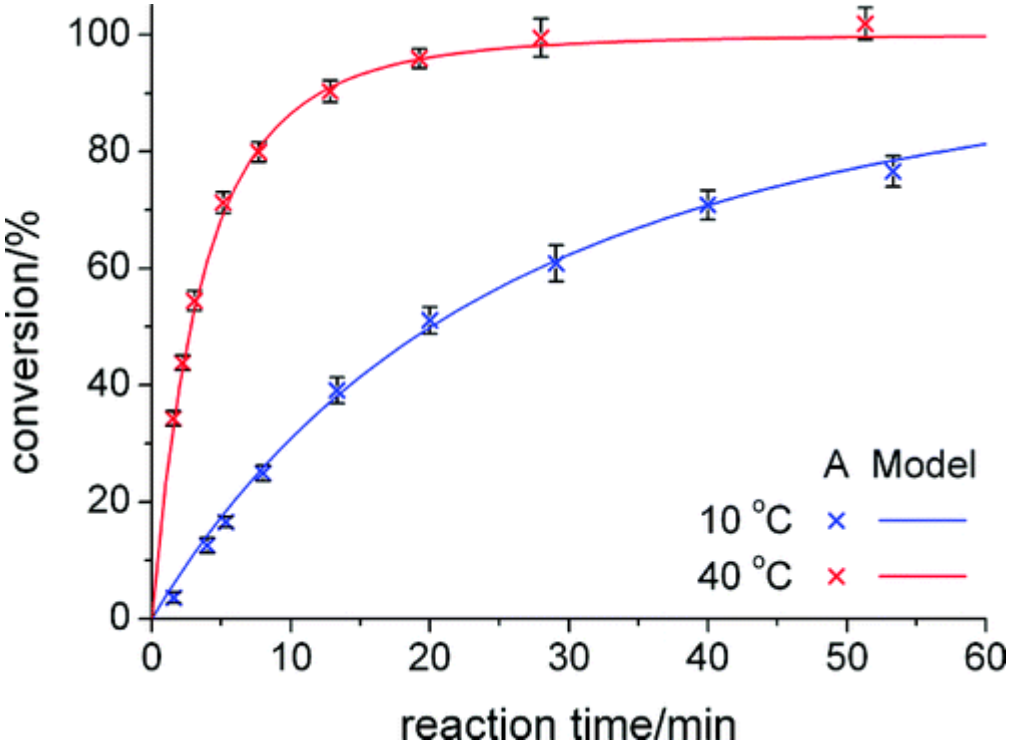


Figure 3.4: A kinetic profile generated from the Knoevenagel condensation shown in Scheme 3.1, where kinetic information was obtained by measuring conversation at given residence times at steady-state, at two temperatures. Steady-state markers, x, are shown as part of the 'A Model'. Reproduced with permission.

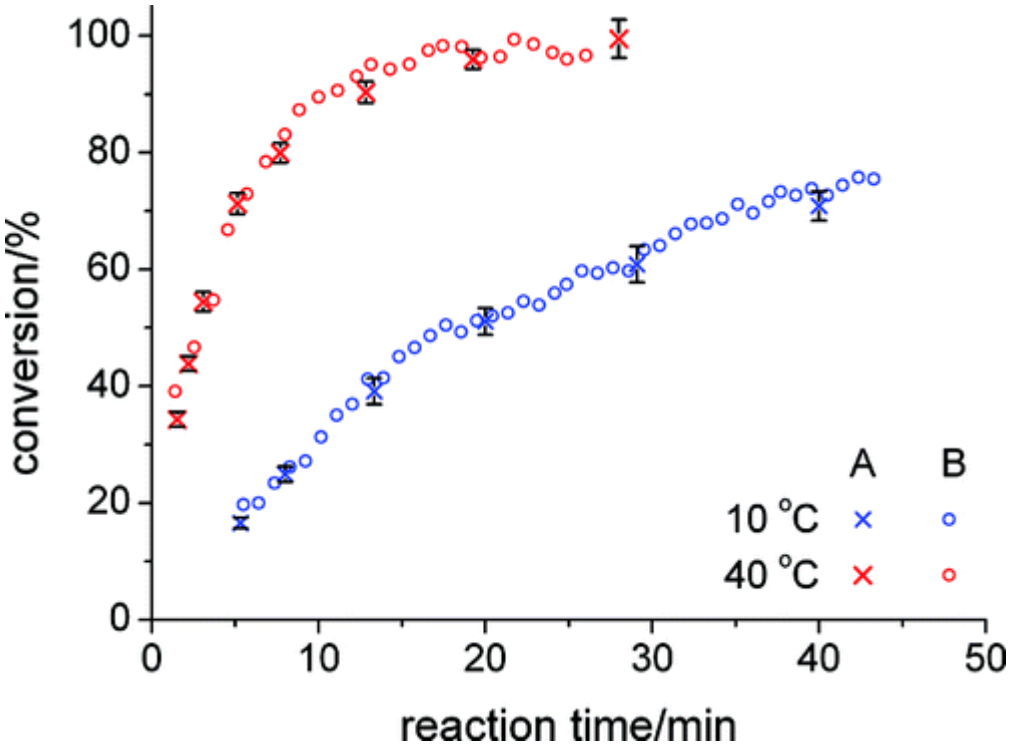


Figure 3.5: A kinetic profile generated from the Knoevenagel condensation shown in Scheme 3.1, where kinetic information was obtained using a flow rate step-change, at two temperatures. Steady-state markers, x, are shown as part of the 'A Model', where step-change markers, o, are shown as part of the 'B Model'. Reproduced with permission.

The main critique of this technique, as stated in the original publication, is that the step increase in flow rate is never perfect as the system always needs time to speed up to the higher flow rate, therefore the exact function F(τ) during this transitional period is uncertain.[96] This non-ideality is caused by several experimental factors, such as non-rigidity of the tubing walls and the syringe, preventing an immediate change in both the flow rates and the pressure profile throughout the system.[227]

Moore and Jensen[227] then reported a new concept that involves a controlled ramp instead of a step change, which leads to less uncertainty in the determination of the residence times, leading to greater accuracy in the time-series data. This therefore also leads to greater accuracy in the kinetic parameters obtained from the experiment. This report introduced the concept of "pseudo-batch" reactors, referring to each fluid element passing through the flow reactor in a time that is unique, which can be thought of as many successive pseudo-batch reactions. This concept is shown in Figure 3.6.

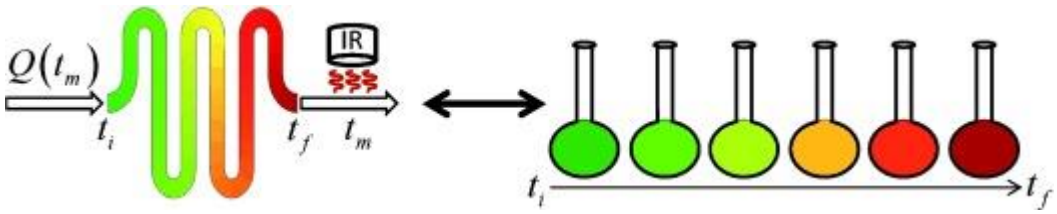
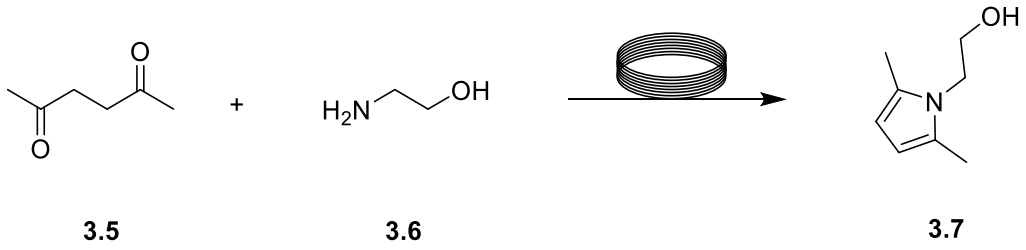


Figure 3.6: A depiction of how a continuous flow reactor may be described as a series of sequential pseudo-batch reactors, where the colour represents the extent of conversion from low (green) to high (red). Q represents the total flow rate, t_i represents the initial time of each pseudo-batch reactor entering the reactor, t_f represents the final time, t_m is the time at which the concentration is actually measured by the IR probe. Reproduced with permission.

This controlled ramp method results in a more predictable and accurate residence time profile when compared to the step change method as the residence time uncertainty decreases, as well as a greater sampling rate with a data density 10-fold higher than previously reported in Mozharov *et al.*'s technique. This was shown experimentally in Scheme 3.2, in the Paal-Knorr reaction of 2,5-hexanedione, **3.5**, and ethanolamine, **3.6**, generating **3.7** whilst constantly monitored by an in-line IR probe. This greater data density also reduces the error within the kinetic profiles generated, whilst still agreeing with

kinetic profiles generated purely from steady-state measurements, as shown in Figure 3.7, where these datasets are shown to be very precise and reproducible.



3.5 **3.6** **3.7**

Scheme 3.2: The Paal-Knorr reaction of 2,5-hexanedione, **3.5**, and ethanolamine, **3.6**, to yield **3.7**, to show how the kinetics of a process can be observed by using a controlled ramp technique.

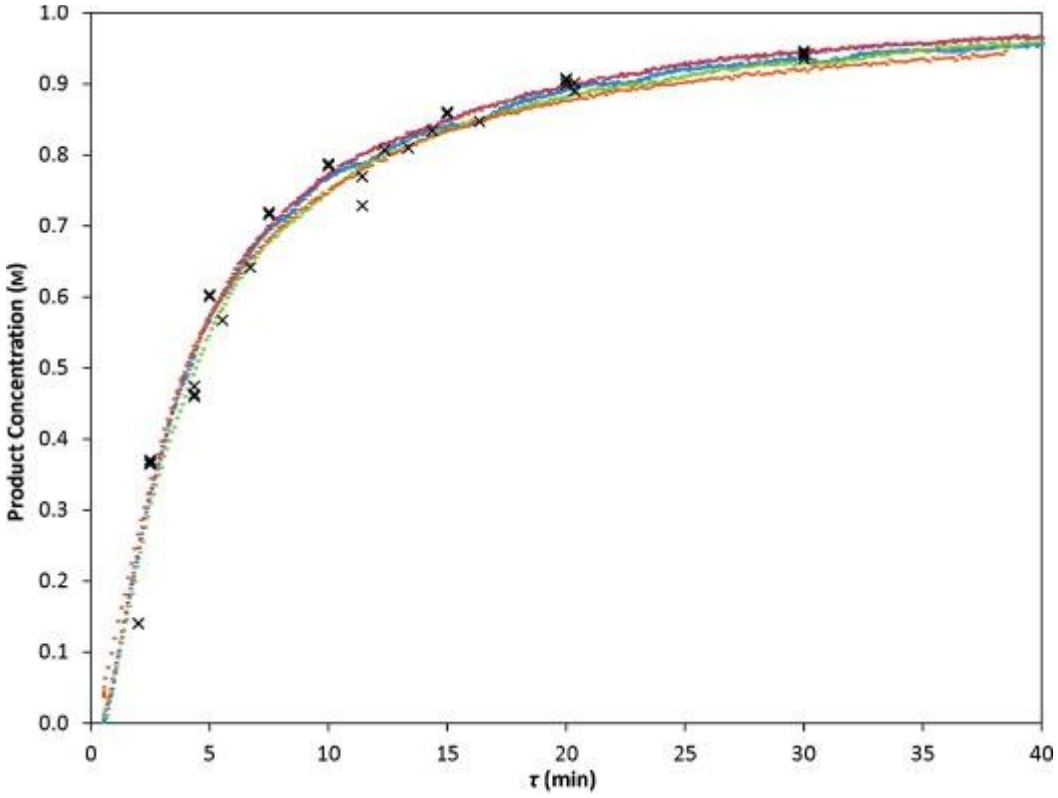


Figure 3.7: A combination of the kinetic models for the reaction yielding **3.7** in Scheme 3.2, where the differing colours represent different values of S, where S is a corrective residence time multiplier to show the reproducibility of the controlled flow ramp methodology. S = 1/4 (blue), S = 1/3 (red), S = 1/2 (green), S= 2/3 (orange), steady-state experiments = x. Reproduced with permission.

Controlled ramps have now been reported many times because of the experimental advantages that they possess.[225, 228] They feature all of the advantages of conventional flow chemistry, as reported earlier in this introduction, without the disadvantages that arise from typical steady-state sampling. This allows users to generate accurate, data-rich kinetic information from a process whilst using minimal material and time. This methodology will

feature heavily in the data acquisition portions of this chapter, as we opted to use controlled ramps for all of our experimentation to exploit these advantages.

More recently, further efficiency has also been realised by the combination of temperature gradients[229] with controlled flow ramps.[146, 230-232] This allows for the fitting of kinetic parameters, including activation energies, to a non-isothermal dataset obtained from one flow experiment. This methodology therefore results in even less material and time consumption as all kinetic parameters can be determined from one flow experiment. Although this relatively new technique has only been reported for very simple systems, it represents an advancement in performing exceptionally efficient kinetic experiments using continuous flow.

3.2 Experimental setup

All flow experiments were conducted using a tubular reaction vessel built in-house, consisting of a 1/16” OD (1/32” ID) stainless steel tubing coiled around a cylindrical aluminium heated block. Reagents were pumped using JASCO PU980 dual piston HPLC pumps and flow streams were mixed using Swagelok SS-100-3 tee-pieces. Sampling was conducted by using a VICI Valco EUDA-CI4W.5 sample loop with a 0.5 µL aliquot volume. This results in small aliquots of reaction mixture automatically transferring to the HPLC for on-line analysis.

The reaction system was maintained under a fixed back pressure using an Upchurch Scientific 1000 PSI back pressure regulator. Quantitative analysis was performed using an Agilent 1100 series HPLC instrument fitted with a Sigma Ascentis Express C18 reverse phase column (5cm x 4.6mm, 2.7 µm). In all experiments biphenyl was added to one reservoir as an internal standard. This experimental setup is shown in Figure 3.8.

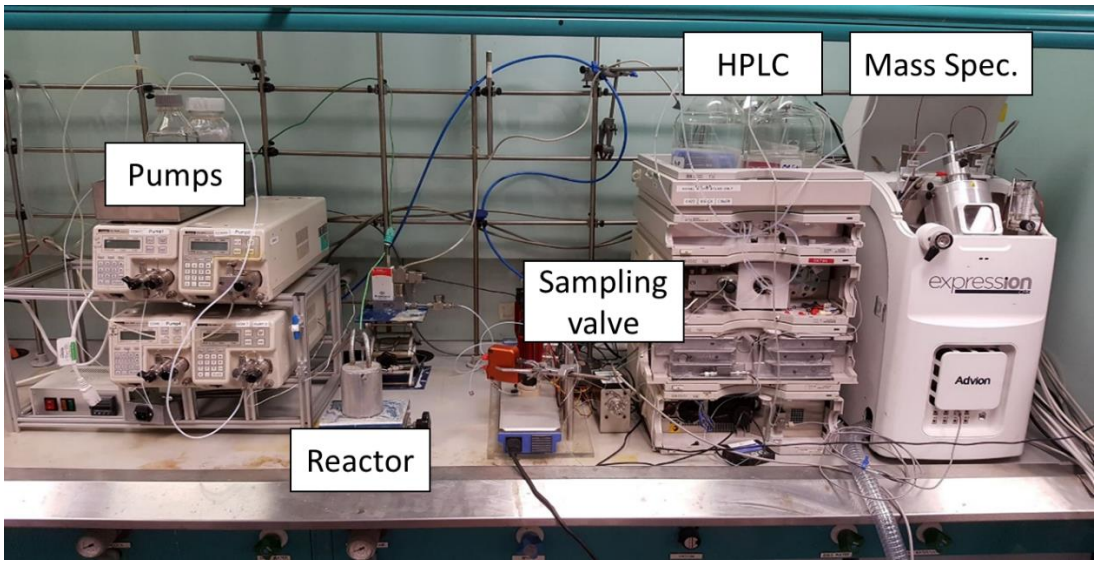


Figure 3.8: A photograph of the automated continuous flow reactor used for this work.

Previous work using this experimental setup has shown that transient flow experiments show an error of less than 4 % in the observed rate constants due to dispersion, whilst steady-state experiments show an absolute error of 0.5 % and a relative error of 0.24 %.[233, 234] Therefore, when considering the kinetic simulations in these studies, a plug flow model was adopted. If we consider the general axial dispersion plug flow reactor model[235] in eqn. 3.3:

$$D_R \frac{d^2C_i}{dz^2} - u\frac{dC_i}{dz} + r_i = 0 \qquad \text{eqn. 3.3}$$

Erstwhile studies by Hone[225] on our reactor platform, and similar investigations by Jensen[230, 236, 237] using a comparable system, have found that dispersion only introduces a small deviation from plug flow. Given this, eqn. 3.3 can be simplified to ignore the second order term, to give eqn. 3.4 - 3.6. This plug flow reactor design equation can be utilised to model each species in the reaction with respect to residence time.

$$u\frac{dC_i}{dz} = r_i \qquad \text{eqn. 3.4}$$

$$\tau = \frac{z}{u} \qquad \text{eqn. 3.5}$$

$$\frac{dC_i}{d\tau} = r_i \qquad \text{eqn. 3.6}$$

Where:

- D_R = axial dispersion coefficient

- C_i = molar concentration of species i
- z = the length along the reactor
- u = superficial velocity
- r_i = rate of reaction of species i

For each chemical process, concentration-time data is collected by using linear flow ramp gradients with on-line HPLC analysis. Steady-state is initially achieved within the system, then sampling begins during the controlled ramping to the next steady-state condition, capturing the transitory information. This is shown in Figure 3.9. At least two flow ramp experiments at different temperatures are conducted, with HPLC sampling every 2 - 4 minutes, depending on the length of the HPLC method. The residence time is then calculated for each sampling point using eqn. 3.7 to translate this transient data to a time-series dataset.

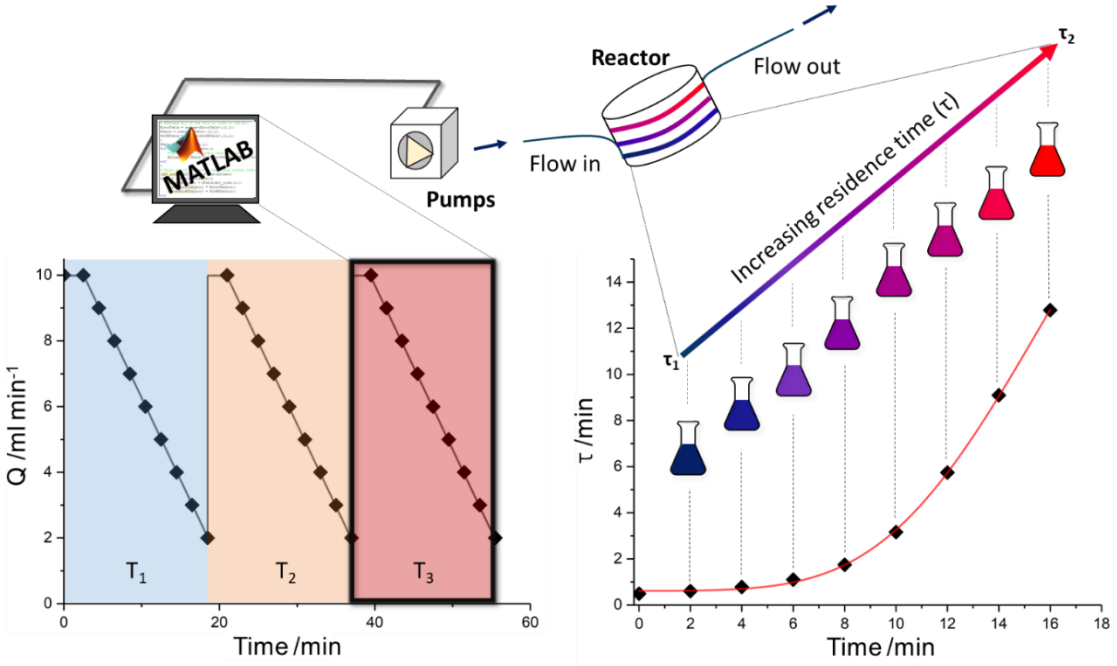


Figure 3.9: A mathematically correct representation of how linear gradient flow ramps can be utilised to sample with a high data density on the initial curvature of the kinetic plot. Where: ♦ = data point, T_n = experiment temperature, Q = total flow rate, Time = time the reaction has been running, τ = residence time that the reaction mixture experiences.

$$\tau = \frac{a \cdot t - \mu_0 + \sqrt{(\mu_0 - a \cdot t)^2 + 2 \cdot L \cdot a}}{a} \qquad \text{eqn. 3.7}$$

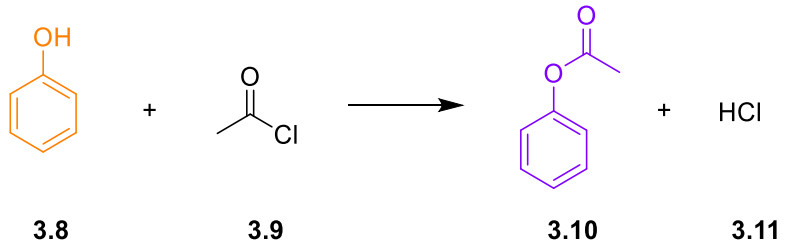
Where:

- τ = residence time
- α = deceleration of flow rate

- μ_0 = initial flow rate
- t = experiment time
- L = reactor volume

3.3 Case study: Phenyl acetate

The first reaction system explored using the computational approach was the presupposed model of the reaction of phenol, **3.8**, with acetyl chloride, **3.9**, to form phenyl acetate, **3.10**, and hydrochloric acid, **3.11**, shown in Scheme 3.3. This first experimental verification of the approach was a final validation of the approach by selecting a simple example with few chemical species. The goal of this study was to confirm the presupposed model, shown in Scheme 3.3, as well as identify the kinetic parameters for this process. The validity of the controlled ramp method would also be determined based on comparisons of the experimental data with steady-state measurements. Full experimental details including the preparation of the feed solutions, experimental setup, flow ramping rates, HPLC analysis and raw data can be found in Chapter 7.3.1.



Scheme 3.3: The reaction of phenol, **3.8**, with acetyl chloride, **3.9**, to form phenyl acetate, **3.10**, and hydrochloric acid, **3.11**.

The kinetic data was obtained using the controlled ramp methodology at two temperatures, 65 °C and 75 °C, and the four participating species were inputted into the computational approach. Based on these four species, only two reactions were calculated to be possible based on mass balance - these are shown as the forward reaction, eqn. 3.8, and the backward reaction, eqn. 3.9, which are two of the three possible models that were generated. The third model, shown as eqn. 3.10, is the combination of these two allowed reactions, resulting in an equilibrium reaction model.

Forward: **3.8** + **3.9** → **3.10** + **3.11** eqn. 3.8

Backward: **3.10** + **3.11** → **3.8** + **3.9** eqn. 3.9

Equilibrium: **3.8** + **3.9** ⇌ **3.10** + **3.11** eqn. 3.10

These three models were then each evaluated automatically by the approach, based on how well the simulated ODE curves converge to the time-series data obtained from the controlled ramps. The k values, SSE error metric and the AIC_C evaluation for each of the models is shown in Table 3.1. This data represents a similar scenario to the simulated benzoic acid esterification case study shown in Chapter 2.3.1, where the identified most likely model has a similar error to another model but with more model terms. The forward reaction model fits both temperature datasets very well, whilst the backward reaction model does not fit to the experimental data at all, as there is no direction in the parameter space from the initial guess that the optimisation algorithm can travel to make a favourable change in the convergence to the data. The equilibrium model fits the data equally as well as the forward model, as it sets the backward rate constant to be negligible, leading to a low SSE. However, as this second term adds no value in further lowering the SSE, it is an unfavourable addition in terms of the AIC_C evaluation which prefers more simplistic models, and is therefore considered a less appropriate model than the forward reaction term alone.

Table 3.1: Evaluation of the feasibility of each reaction model for the phenyl acetate study.

Reaction Model	**k Values / x 10^{-3} M^{-1} s^{-1}**		**SSE /M**	**AIC_C Evaluation**
	65 °C	**75 °C**		
Forward	5.15	10.45	0.019	1.36
Backward	-	-	0.532	8.03
Forward + Backward	5.15 0	10.45 0	0.019	16.36

By only inputting the species involved in the reaction, then running two controlled flow ramps with the automated computational approach described, the intuitive reaction model was confirmed and the kinetic parameters were determined as $k_{75\,°C}$ = 10.45 x 10^{-3} ± 0.42 x 10^{-3} M^{-1} s^{-1}, E_a = 69.3 ± 7.8 kJ mol^{-1}. This combination of the correct reaction model and kinetic parameters allowed a

fit to the experimental data with an average residual of less than 3 x 10^{-3} M, and is shown in Figure 3.10; where the colours of the data and fitted ODEs mimic the colours shown in Scheme 3.3. This study has shown that the transition from simulated to real experimental data does not affect the accuracy of the approach, and that the approach can be applied to experimental data with confidence that the correct reaction model is to be identified.

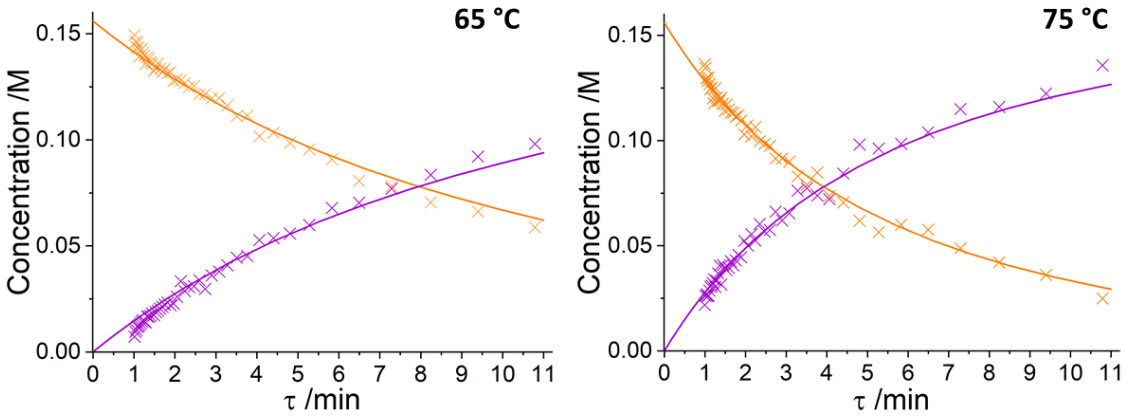


Figure 3.10: Kinetic profiles for two flow ramp experiments at 65 °C and 75 °C, where: x = phenol, x = phenyl acetate, — = phenol (ODE), — = phenyl acetate (ODE).

The equivalent temperature and residence time kinetic profiles were also mapped by performing a series of conventional steady-state experiments, to ensure that a transient-flow regime remained accurate in model and parameter determination. Results showed excellent agreement in plotted curvature and confirmed the same reaction model to be the most likely, whilst identifying very comparable kinetic constants. Figure 3.11 shows a comparison of the two sampling methodologies for this particular reaction. See Chapter 7.3.1.5 for further details.

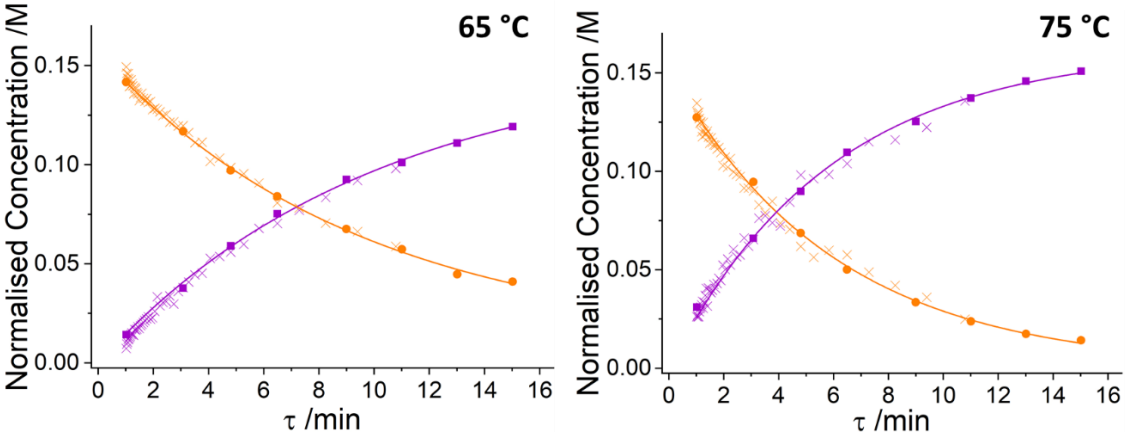
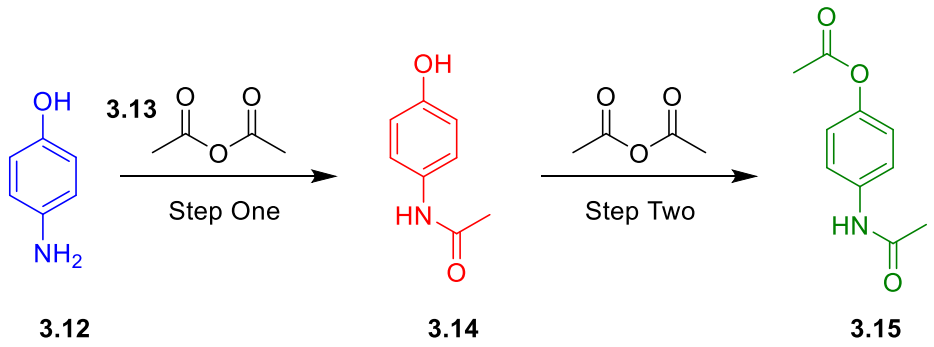


Figure 3.11: Graphs showing the agreement between steady-state and flow ramp measurements for the reaction of phenol with acetyl chloride, where curves are fitted to the steady-state data. Where: ● = phenol (steady-state), x = phenol (flow ramp) ■ = phenyl acetate (steady-state), x = phenyl acetate (flow ramp).

3.4 Case study: Paracetamol

The next reaction system that was explored was the chemical system producing paracetamol, whereby 4-aminophenol, **3.12**, reacts with acetic anhydride, **3.13**, to form paracetamol, **3.14**, and the over-reacted diacetamate impurity, **3.15**, shown in Scheme 3.4 as the intuitive sequential reaction. This reaction system serves as the first real experimental multistep reaction that is studied by this approach in order to determine the correct reaction model and kinetic parameters. However, as there is a large disparity between the reaction kinetics of step one and two, quantitative kinetic analysis of both processes simultaneously, i.e. during a single ramp, was not possible. Therefore, two sets of differing temperature ramps were performed to investigate independently the formation of paracetamol and diacetamate, at 30/60 °C and 160/180 °C respectively. Each of these ramps differ in reactor size and hence residence times, as well as starting concentrations of acetic anhydride. This was performed to illustrate the capability of the approach to handle data from a variety of sources whilst still accurately determining the kinetic properties of a process.



3.12 **3.14** **3.15**

Scheme 3.4: The reaction of 4-aminophenol with acetic anhydride to form paracetamol in step one, followed by a further reaction with acetic anhydride to form diacetamate in step two.

Based on the five species identified, including the acetic acid, **3.16**, formed with each reaction of acetic anhydride, there were six reactions that were calculated to be possible based on mass balance. These reactions are shown as eqn. 3.11 - 3.16:

3.12 + **3.13** → **3.14** + **3.16** eqn. 3.11

3.14 + **3.16** → **3.12** + **3.13** eqn. 3.12

3.14 + **3.13** → **3.15** + **3.16** eqn. 3.13

3.15 + **3.16** → **3.14** + **3.13** eqn. 3.14

3.12 + **3.15** → **3.14** + **3.14** eqn. 3.15

3.14 + **3.14** → **3.12** + **3.15** eqn. 3.16

Interestingly, some of the mass-balance-allowed reactions identified by the ILP optimisation seem very unlikely to happen based on chemical intuition. The reaction of 4-aminophenol with diacetamate shown in eqn. 3.15, for example, suggests that it is possible for this starting material to react with the over-reacted product, to form two equivalents of the desired product. We know from intuition that the ester bond is too strong to break and react without a chemical stimulus, so we may have initial speculations on the feasibility of this reaction. However, it can never be stated with certainty that this reaction cannot happen until its feasibility is studied, and therefore all possible models are still generated to be evaluated. This is a major advantage of this approach, that reactions and models that may otherwise be discarded due to chemical bias are automatically evaluated without a chemist's input. This could lead to unexpected (but statistically and chemically accurate) models that describe time-series data, which can lead to better process efficiency and even discover new reactions.

63 potential models were identified from the five reactions, and the reaction model shown in Scheme 3.4 was determined to be the most likely representation of the system by AIC_C. The approach also determined the kinetic parameters of step one: $k_{60\,°C} = 6.45 \pm 0.26\ M^{-1}\ s^{-1}$, $E_a = 3.2 \pm 1.2\ kJ\ mol^{-1}$ and step two: $k_{180\,°C} = 4.27 \times 10^{-2} \pm 0.17 \times 10^{-2}\ M^{-1}\ s^{-1}$, $E_a = 97.9 \pm 6.5\ kJ\ mol^{-1}$, by fitting k values to each dataset in the kinetic fitting stage of the approach. These values us to assert that step one will likely be very fast at a wide range of temperature ranges, and that step two has a higher sensitivity to changes in temperature when the energy of the system increases. This identified model alongside the identified kinetic parameters fits to the experimental data very accurately, with an average residual of less than 1×10^{-4} M, and is shown in Figure 3.12. The top 5 ranked models are shown in Table 3.2, with their respective SSE error metric and AIC_C evaluation.

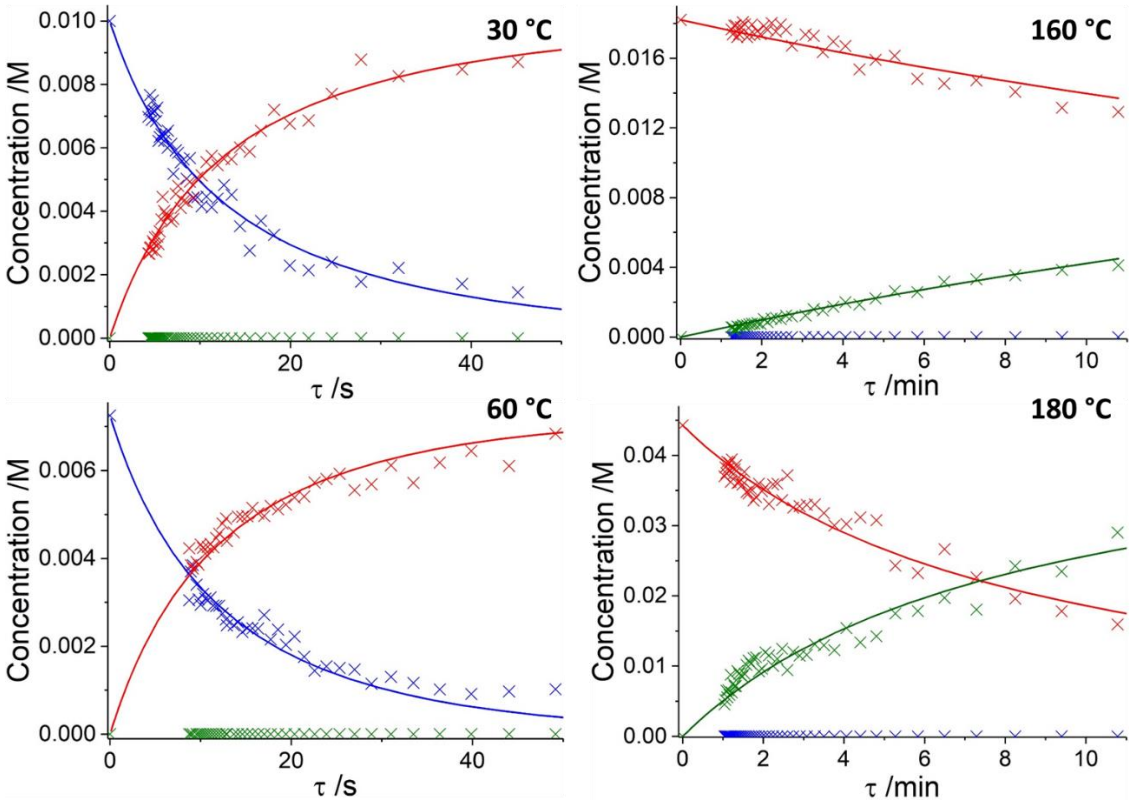


Figure 3.12: Kinetic profiles for four flow ramp experiments at 30 °C, 60 °C, 160 °C and 180 °C, where: x = 4-aminophenol, x = paracetamol, x = diacetamate, — = 4-aminophenol (ODE), — = paracetamol (ODE), — = diacetamate (ODE). See Chapter 7.3.2 for full experimental conditions and raw data.

Table 3.2: Evaluation of the feasibility of each reaction model for the paracetamol study.

Rank	Reaction Model	k values /s^{-1} or /M^{-1} s^{-1}				SSE / x10^3 M	AIC_C Evaluation
		30 °C	60 °C	160 °C	180 °C		
1	3.12 + 3.13 → 3.14 + 3.16 3.14 + 3.13 → 3.15 + 3.16	5.9082 -	6.4471 -	- 0.0129	- 0.0427	0.295	12.92
2	3.14 + 3.16 → 3.12 + 3.13 3.12 + 3.13 → 3.14 + 3.16 3.14 + 3.13 → 3.15 + 3.16	0.1599 5.9509 -	0.9120 6.7519 -	- - 0.0125	- - 0.0427	0.293	36.90
3	3.12 + 3.13 → 3.14 + 3.16 3.12 + 3.15 → 3.14 + 3.14 3.14 + 3.13 → 3.15 + 3.16	5.9067 3.8065 -	6.4348 3.5719 -	- - 0.0129	- - 0.0428	0.295	36.92
4	3.15 + 3.16 → 3.14 + 3.13 3.12 + 3.13 → 3.14 + 3.16 3.14 + 3.13 → 3.15 + 3.16	2.3783 5.9188 -	2.7076 6.4641 -	0.0210 - 0.0130	0.0022 - 0.0430	0.295	36.92

5	3.14 + 3.14 → 3.12 + 3.15 3.12 + 3.13 → 3.14 + 3.16 3.14 + 3.13 → 3.15 + 3.16	0.0799 5.9258 -	0.2270 6.4833 -	- - 0.0127	- - 0.0425	0.295	36.92

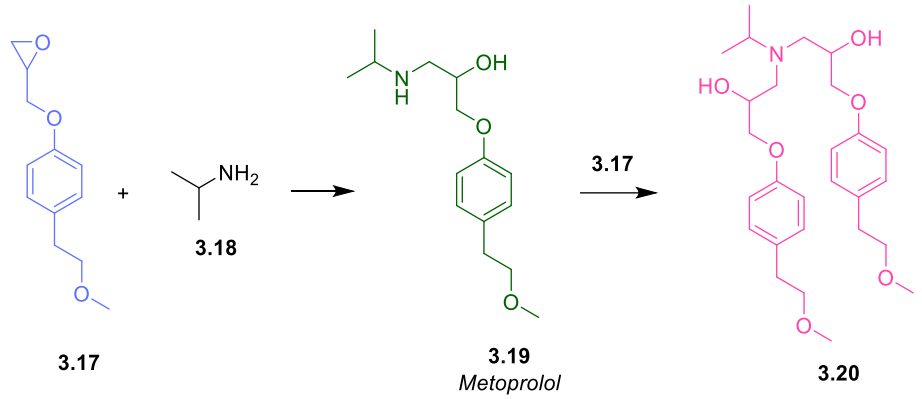
Interestingly, there are many reaction models that were found to have a lower SSE error metric than the identified most likely model. This is again where the AIC_C ranking metric is used to great effect to balance the convergence to experimental data as well as the simplicity of the model. It is in these cases that models that contain transformations discussed previously as very unlikely, such as in the rank 3 model, that the terms are likely being used by the optimisation algorithm to fit experimental noise. For this reason, this approach cannot be completely autonomous and must require evaluation by a chemist when observing the top rated models, to be certain that the quality of the data assures correct model determination - this was discussed in Chapter 2.2.5.

This paracetamol study has shown that the approach can correctly identify the reaction model and kinetic parameters for a multistep reaction using real experimental data from different temperature flow ramps, by inputted only the observed species and time-series data. As a side note - interestingly, contrary to common undergraduate laboratory experiment scripts,[238] the over-reacted diacetamate will not form in any measurable quantities in a conventional undergraduate experiment. Assuming that the reaction is similar in water to acetonitrile, the reaction doesn't need refluxing and will go to full conversion of paracetamol within seconds at room temperature. Furthermore, using the kinetic parameters in this study and the recommended reflux temperature of common laboratory manuals, 100 °C, and a typical 0.2 M reaction mixture of paracetamol and acetic anhydride, it would take approximately 12 years for the reaction to achieve complete conversion to diacetamate.

3.5 Case study: Metoprolol

The final experimental case study that was explored was the chemical system producing metoprolol, whereby the epoxide starting material, **3.17**, reacts with isopropylamine, **3.18**, to form metoprolol, **3.19**, which can further react to form the bis-substituted impurity, **3.20**, shown in Scheme 3.5. Metoprolol is a

cardioselective beta-blocker commonly used for the treatment of hypertension, for which kinetic information would help in the process development stage of manufacture at AstraZeneca. This reaction system serves as the first real experimental multistep reaction that is studied by this approach, where parallel reactions occur in tandem, leading to a more difficult system for which to optimise kinetic parameters and identify the correct reaction model.



Scheme 3.5: The reaction of the epoxide starting material with isopropylamine to form metoprolol, as well as the overreaction to form the bis-substituted product.

Two sets of two-temperature flow ramps were run on parallelised flow reactor platforms, one set in our lab in Leeds (190/210 °C) and one set at AstraZeneca's lab in Macclesfield (130/150 °C). The two reactor platforms differ in equipment specification and reactor volume, and the two experimental sets differ in temperatures and starting concentrations. This parallelisation of experiments on different systems was performed to further confirm the reproducibility of this flow ramp methodology, as corroborating data can be achieved by an operator on separate reactor systems in different locations. The experimental results were then combined and the computational approach was applied. Full details of experimentation can be found in Chapter 7.3.3.1.

Based on the four species identified, there were six reactions that were calculated to be possible based on mass balance. These reactions are shown as eqn. 3.17 - 3.22.

3.17 + **3.18** → **3.19** eqn. 3.17

3.19 → **3.17** + **3.18** eqn. 3.18

3.19 + **3.17** → **3.20** eqn. 3.19

3.20 → **3.19** + **3.17** eqn. 3.20

3.19 + **3.19** → **3.18** + **3.20** eqn. 3.21

3.18 + **3.20** → **3.19** + **3.19** eqn. 3.22

63 reaction models were generated, which correspond to every possible combination of these allowed reactions, and each of them were evaluated sequentially by the approach. The reaction model shown in Scheme 3.5 was identified as the most likely representation of the system by AIC_C, and the kinetic parameters for the formation of metoprolol were found to be: $k_{170\,°C}$ = 0.286 ± 0.012 M^{-1} min^{-1}, E_a = 72.4 ± 2.9 kJ mol^{-1} and for the formation of the bis-substituted product: $k_{170\,°C}$ = 0.019 ± 0.001 M^{-1} min^{-1}, E_a = 75.0 ± 3.0 kJ mol^{-1}. This identified model alongside the kinetic parameters fit to the experimental data very accurately, with an average residual of less than 2 x 10^{-3} M, and is shown in Figure 3.13. The top 5 ranked models are shown in Table 3.3, with their respective SSE error metric and AIC_C evaluation.

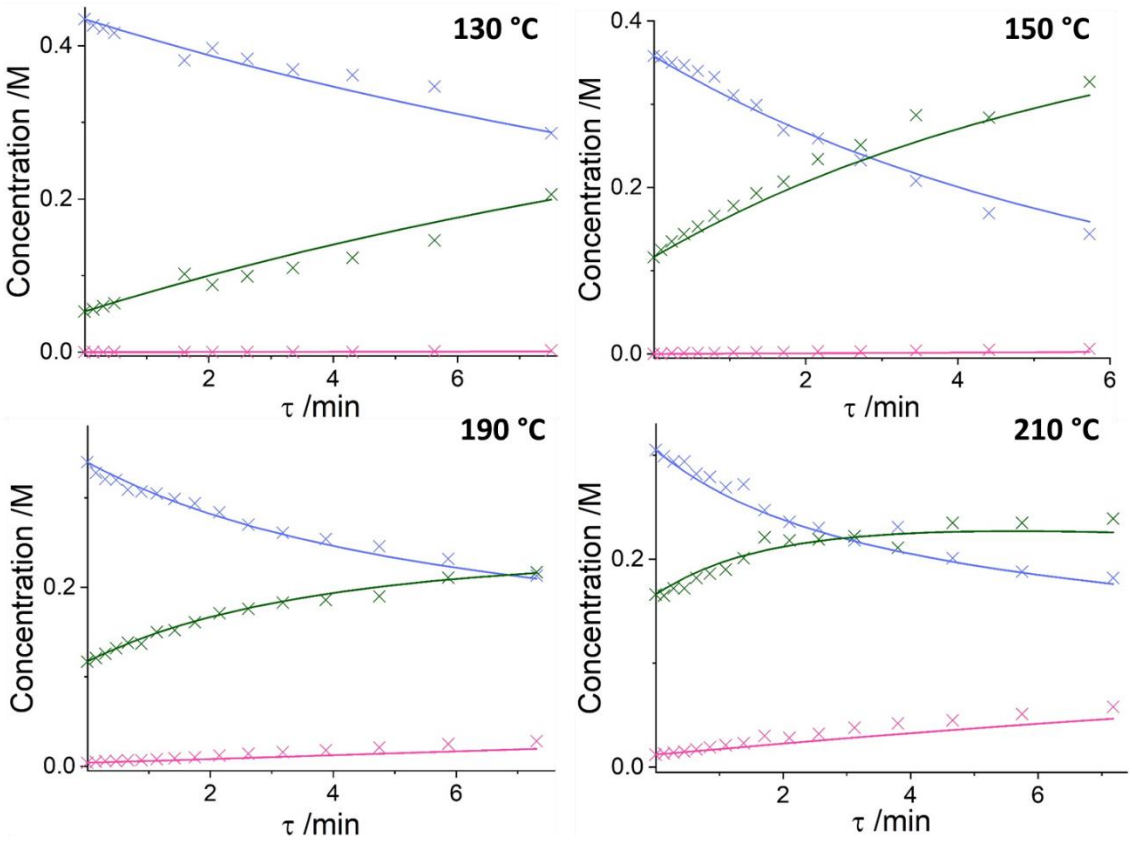


Figure 3.13: Kinetic profiles for the flow ramp experiments at 130 °C, 150 °C, 190 °C and 210 °C, where: x = starting material, x = Metoprolol, x = bis-substituted product, — = starting material (ODE), — = Metoprolol (ODE), — = bis-substituted product (ODE). See Chapter 7.3.3 for full experimental conditions and raw data.

Table 3.3: Evaluation of the feasibility of each reaction model for the metoprolol study.

Rank	Reaction Model	Kinetic parameters		SSE / x10 M	AIC_C Evaluation
		$k_{170\,°C}$ /min^{-1} or M^{-1} min^{-1}	E_a /kJ mol^{-1}		
1	3.17 + 3.18 → 3.19 3.19 + 3.17 → 3.20	0.2857 0.0191	72.42 75.00	0.1212	-311.45
2	3.17 + 3.18 → 3.19 3.19 → 3.18 + 3.20	0.3065 0.0260	75.69 65.52	0.1252	-309.61
3	3.17 + 3.18 → 3.19 3.19 → 3.17 + 3.18 3.17 + 3.19 → 3.20	0.3008 0.0043 0.0188	74.63 53.79 75.57	0.1188	-307.68
4	3.17 + 3.18 → 3.19 3.19 + 3.17 → 3.20 3.20 → 3.19 + 3.17	0.2858 0.0202 0.0158	72.45 82.16 49.20	0.1209	-306.69
5	3.17 + 3.18 → 3.19 3.19 + 3.17 → 3.20 3.18 + 3.20 → 3.19 + 3.19	0.2857 0.0191 0.0003	72.42 75.08 48.98	0.1212	-306.53

The reference k values and activation energies were fitted directly in the kinetic fitting stage of the approach, and the experimental dataset was adjusted to incorporate an artificial-zero time point, which is a common practice when fitting kinetics in order to assert an experimental ‘start time’ when concentrations are known. This can fix any fitting discrepancies that may occur as the first measured time point may not be accurate - it is assumed however, that all time points following this first measurement are correct with respect to the first measurement. This is assumed only for the fitting of the kinetic parameters, as the parameters are likely to be more accurate. In the case of this experiment, small deviations in asserting the time of the initial measurement may occur from thermal expansion of the solvent because of the very high temperatures, or from pump flow rate errors.

It has been shown in this study that it is possible to deconvolute reaction pathways that are happening simultaneously using this computational approach. Differing-temperature experimental datasets were collected from different

experimental systems and collated to show that the correct model could be identified even with data from different sources. Kinetic parameters are also identified from this approach that gives an excellent fit to the experimental data.

This kinetic information can then be used to optimise this process between given limits for temperature, chemical equivalents and reaction time. Using current pricing for the starting material used in this work,[239] other standard industrial optimisation techniques would have been significantly more expensive to implement. When comparing this kinetic approach to other optimisation methods, steady-state kinetic measurements would have cost 24 % more in terms of material consumption, and a screening and full factorial design of experiments (DoE) optimisation would have cost 106 % more - see Chapter 7.3.3.6 for more details. Then of course factoring in the cost of the time of the chemist running the experiments (which hereby would be automated) and the time for interpretation of the data and kinetics (which the approach elucidates), this results in a significant reduction in labour, time and overall cost, which also results in a more comprehensive overview of the possible kinetic models at play.

3.6 Conclusion

It has been shown in this work that when real experimental time-series data for a chemical process is available, total process understanding can be achieved without the need for high-level chemical intuition or human interference. When participating species are known or inferred, complete sets of kinetic information can be obtained via construction of all possible reaction models and identification of their respective kinetic parameters. This was undertaken by coupling an automated flow reactor platform with a computational approach to deduce and evaluate each kinetic model, utilising optimisation algorithms. After post-reaction statistical analysis indicates which models are the most likely to be true based on the experimental data provided, which can be from batch or flow, this information can then be interpreted by a trained chemist to further differentiate reaction pathways based on what should and should not be chemically possible.

This approach has been proven to be powerful in determining the reaction model and kinetic parameters in a variety of experimental circumstances. The approach will be particularly powerful in situations where the reaction model is

not completely understood, for example when there are competing reaction pathways. The approach can be computationally expensive depending on the number of possible reaction models, although all of the optimisations carried out by the approach on the work described was evaluated in less than 5 minutes on a standard 4-core Intel i5-2310 processor. However, the added computational cost is minimal compared to the reduction of time and experimental cost with regards to kinetic evaluation. Furthermore, this approach can be run automatically as bench scientists use their human resource for more challenging tasks that cannot be automated.

This work is the first implementation of the approach on experimental time-series data and has been proven to efficiently interpret kinetic information using minimal amounts of material to generate sufficient experimental data to enable accurate model determination. Using this methodology can considerably outweigh the cost of further experimentation to discriminate speculated kinetic models and can greatly reduce the time and cost barriers to full process understanding.

Chapter 4 : The development of an improved computational approach to kinetic model determination

4.1 Introduction

The aforementioned computational approach to kinetic model and parameter determination, as described in Chapter 2 and Chapter 3, is a powerful tool for many experimental applications. However, there are two major limitations to this tool. The first limitation is in the reaction orders of the particular species that are available - these orders are constrained to two integer orders: 1 and 2. There are instances, however, where reactions can feature zero order reactions, or even have non-integer orders, for which this approach currently cannot facilitate and identify. The second limitation is the inability of the approach to model catalytic reactions and hence determine the reaction order of species within a catalytic reaction. As catalytic reactions are common in research, manufacturing and process development, there are many applications where this approach could ideally be utilised but is not currently applicable.

Further development of the computational approach to rectify these limitations can therefore increase its applicability greatly, which further increases the scope of the tool for chemical applications. As the comprehensive model evaluation aspect of this methodology has been proven to be effective for determining correct reaction models, the skeleton of the approach remains unchanged. This includes the two major steps of the approach, model generation and kinetic parameter fitting, as well as the statistical measurements involved i.e. AIC_C. However, although the basis for the approach remains the same, many structural changes of the programmed tool are necessary for the incorporation of these new applications: non-integer order models and catalytic models. The advancement of the approach to incorporate these changes makes it applicable in almost all chemical processes, which is the final step in answering the overall research aim of the project: to build an automated methodology to kinetic model discrimination and parameter determination.

4.2 Methodology advancement

4.2.1 Capabilities

When mass-balance-allowed reactions are identified by the approach, these reactions can then be compiled into a number of models that feature them, as discussed in Chapter 2.2. However, there is a difference between an allowed reaction based on the mass balance of the reactants and products, and the rate laws that govern the way that these species react. All models are currently compiled in a way that assumes a first order dependence on all species that are featured. Therefore, for the mass-balance-allowed transformation of A to B as shown in eqn. 4.1, the rate law for the change in B is shown in eqn. 4.2:

$$A \xrightarrow{k} B \qquad \text{eqn. 4.1}$$

$$\frac{d[B]}{dt} = k[A] \qquad \text{eqn. 4.2}$$

A fundamental change to this methodology explores the same transformation but using different reactivities, by also investigating different orders of the reacting species. To explore different reaction orders, including non-integer reaction orders, the rate law of the elementary reaction can be described with differing powers of α, shown in eqn. 4.3, where α = 0, 0.5 or 1:

$$\frac{d[B]}{dt} = k[A]^{\alpha} \qquad \text{eqn. 4.3}$$

This allows multiple chemical pathways to be explored that describe the chemistry in different ways, which were not available previously using this methodology. Each of the newly identified reactions are then treated as 'available' reactions, from which reaction models can be constructed in the same way as described in Chapter 2.2. Therefore, for each mass-balance-allowed reaction identified by the approach that is 'unimolecular', there are 3 corresponding reaction orders possible and therefore 3 reactions that are added to the pool of available reactions for kinetic model generation. In the context of this approach, the transformation of a single mass-balance-allowed reaction to every variant of its corresponding rate law is termed 'inflation'.

For 'bimolecular' reactions identified by the approach as mass-balancing, similar chemical descriptions can be written as with the 'unimolecular' reactions. For the reaction of A and B to form C, there are multiple potential combinations of the species orders α and β in the potential rate laws, shown in eqn. 4.4 and 4.5:

$$A + B \overset{k}{\rightarrow} C \qquad \text{eqn. 4.4}$$

$$\frac{d[C]}{dt} = k[A]^{\alpha}[B]^{\beta};\ \alpha + \beta \geq 1; \alpha = \beta \neq 0.5;\ \alpha, \beta \in \{0, 0.5, 1 \qquad \text{eqn. 4.5}$$

These values of α and β can take the form of: 0, 0.5, 1, as before. However, it is extremely rare for two chemical species to react with a zero order dependence on both species. Therefore, in an effort to lower the number of 'available' reactions and hence lower the number of kinetic model evaluations necessary, these particular reactions are seen as unfeasible and are therefore not included in the model generation stage. This is also true of a zero order/0.5 order bimolecular reaction. Therefore, for every 'bimolecular' mass-balance-allowed transformation identified by the approach, there are 5 corresponding reactions that are added to the pool of available reactions for kinetic model generation.

This approach has also been modified to include catalytic reactions in the model generation, if the user selects the option to do so. The user can select to either include catalytic reactions, or exclusively look at catalytic reactions (i.e. no reactions can occur without a catalyst). As before, all mass-balance-allowed

transformations are identified, then for the reaction of A and B to form C, all combinations of integer and non-integer catalytic reactions are generated within the inflation step. At this point, the species that is user-selected to behave catalytically is then incorporated into the rate laws. For each mass-balance-allowed (inflated) transformation, the catalytic species can take the order of: 0.5, 1 or 2, and each of these rate law variations are constructed and added to the pool of available reactions for kinetic model generation. Note also that in these catalytic reactions, the dependence of the orders become empirical to a greater extent, whereby individual species are used and regenerated, as opposed to referring specifically to molecularity (i.e. number of species reacting together at once). When catalytic reactions are included in model generation with the selected catalytic species 'Cat', eqn. 4.4 becomes eqn. 4.6, and the general rate law shown in eqn. 4.5 becomes eqn. 4.7, where γ = order of catalyst:

$$A + B + Cat \xrightarrow{k} C + Cat \qquad \text{eqn. 4.6}$$

$$\frac{d[C]}{dt} = k[A]^{\alpha}[B]^{\beta}[Cat]^{\gamma}; \alpha + \beta \geq 1;\ \alpha = \beta \neq 0.5;\ \alpha, \beta \in \{0, 0.5, 1\};\ \gamma \in \{0.5,1,2\} \qquad \text{eqn. 4.7}$$

All combinations of catalytic and non-catalytic reactions are generated. For each mass-balance-allowed 'bimolecular' reaction, where catalytic variants are also generated, there are 15 possible corresponding reactions that are added to the pool of available reactions for kinetic model generation. This inflation step is summarised qualitatively in Figure 4.1, showing how individual mass-balance-allowed reactions can generate multiple transformations that are dependent on different rate laws.

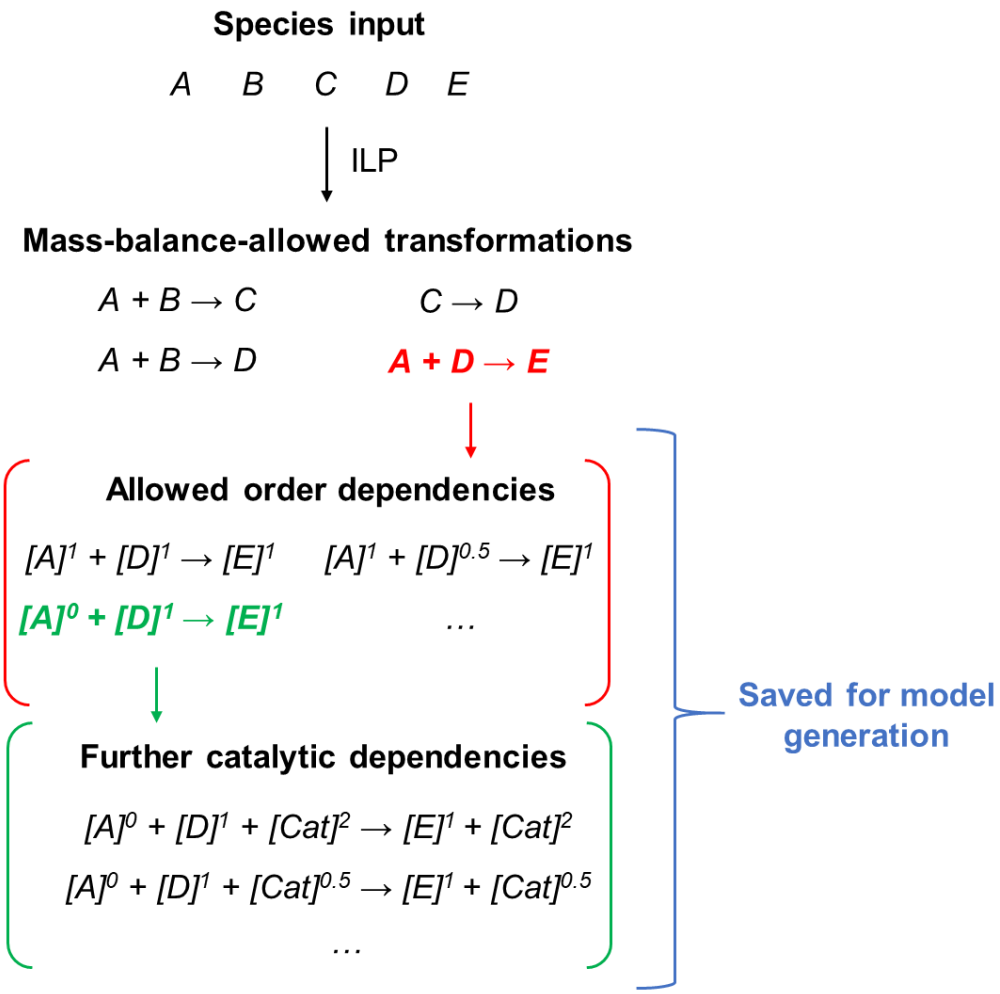


Figure 4.1: An illustration of how the 'inflation' step takes all mass-balance-allowed transformations sequentially, then deduces all allowed integer and non-integer orders and generates corresponding rate laws for these transformations. If catalytic reactions are to be explored also, rate laws for these catalytic dependencies are also generated. These transformations are all then saved, from which full reaction models can be constructed.

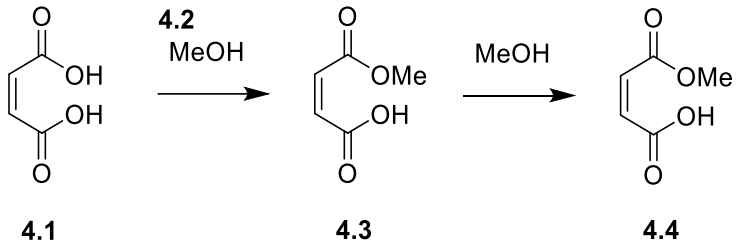
These new capabilities incorporated into the overall computational approach allow kinetic model determination for more complex chemical systems, for catalytic/non-catalytic processes with integer or non-integer orders. This approach can then be used as before, whereby the chemical species and experimental datasets are uploaded and the program runs an extensive kinetic model evaluation loop, followed by statistical analysis to determine the most likely model. However, the major disadvantage to evaluating all possible models generated after the newly incorporated inflation step is the huge computational cost. As the pool of available reactions is expanded greatly, from simple mass-balance-allowed reactions to all rate laws possibly governing these reactions, the number of possible models to be generated from these reactions grows

exponentially. It is therefore not feasible, with current technology, to evaluate every possible model. However, many of these possible models are also unfeasible due to their inability to occur anyway. Upon generation of a reaction model, the fitness of the model can be determined by employing simple and appropriate logical and chemical constraints, or rules. These rules ensure that generated models with an infeasibility to occur are identified and discarded, without the need for further evaluation and consumption of computational resources.

4.2.2 Rules

In order to evaluate only the models that are feasible, *a priori* logical and chemical rules must be applied to the available models to discard large numbers of them, as many models may feature redundant or impossible model terms within them. To postulate such rules requires logical, yet simplistic, rationalisations of how reactions occur - rationalisations that seem obvious upon realisation but are seldom taught or reported in the literature.

To show the importance of these rules when running this approach, Scheme 4.1 shows a model reaction of maleic acid, **4.1**, reacting with methanol, **4.2**, to form the mono-product, **4.3**, and the di-product, **4.4**, which can be used to show the number of models to be evaluated.[240] In this reaction, the maleic acid is self-catalytic, so this species behaves as both a starting material and a catalyst. Using these 5 species, there are 6 possible mass-balance-allowed reactions. From these 6 reactions, after inflation steps to look at all integer and non-integer species dependencies and all catalytic and non-catalytic reactions, the total number of models to evaluate is >1,854,900,872. A specific number cannot be quoted, as just compiling this number of reaction models takes over 110 GB of memory and so an accurate estimate cannot be achieved with the equipment available. The evaluation of this number of models would likely take several years to complete, even when using high-performance computing clusters. It is therefore imperative that these rules drastically reduce the number of models to evaluate, to enable this methodology as a practical process development application.



Scheme 4.1: The self-catalysed reaction of maleic acid, **4.1**, and methanol, **4.2**, to form the mono-product, **4.3**, and the di-product, **4.4**.[240]

Rule #1: One variant of each reaction

There are many variations of the same elementary transformations to be considered when building models for evaluation. Examples of these variations could be:

- $A + B \rightarrow C$
- $A + B^{0} \rightarrow C$
- $A^{0.5} + B \rightarrow C$
- etc.

It is very unlikely, however, that the same reactants will undergo the same chemical transformation to products with many different rate laws. It would not be expected that an elementary step would proceed in numerous ways, in both an overall second order and an overall first order manner to result in the same product, for example. Therefore, when compiling reaction models for evaluation, only one variant of each elementary reaction can be present in a single model. This also reduces the maximum number of model terms in a single model to the total number of elementary reactions, which in this example is 6. When this rule is enforced during model generation for the model chemical system, the total number of models for evaluation is: 8,156,735.

Rule #2: Iterative model evaluations

Although it cannot be known what the ‘average’ number of model terms in a ‘general’ model is, the assumption can be made that the user is investigating a single synthetic step in a chemical process. With a single synthetic step, there may be consecutive reactions and impurities formed, but it would undoubtedly be regarded as unlikely if any model was proposed with 30 model terms, for example. In accepting this postulation, it must therefore also be true that there is a continuum of likelihood whereby a model will contain a certain number of model

terms, especially when evaluated by AIC_C. Therefore, there must also exist a limit of model terms that once reached, it can be regarded that the corresponding model is too unlikely to occur to be treated as a serious model candidate. Without a vast literature study, it is not possible to know what this number should be, and this prospective study extends past the scope of this project - this discussion is highlighted to simply introduce the concept of the existence of this upper limit of model terms.

Although this upper limit is unknown, it is possible to start at a prospective number of model terms, then the user can decide if this number should be increased based on the responses observed. From chemical intuition based on the findings of AIC_C values thus far in the project, much of the data can be fitted effectively with ≤ 4 model terms and adding further model terms *in general* leads to diminishing model evaluations. This is, of course, a biased generalisation based on the chemical systems that have been studied within the project. However, using an initial upper limit of 4 model terms in the model generation stage may be generally appropriate for single-step chemical systems - if this limit does not give a satisfactory model that fits the supplied experimental data, it can be iteratively increased to 5 model terms, then 6, etc. As there are many models that would require computationally expensive evaluations with over 4 model terms, this rule allows the user to determine if further evaluations with larger models are necessary, which could otherwise potentially waste computational resources and more importantly, time. When this rule is enforced during model generation in the model chemical system, as well as the previous Rule #1, the total number of models for evaluation is: 668,735.

Rule #3: Catalytic reactions considered only

In order to reduce the number of available reactions from which to build models, and hence reduce the number of unnecessary model evaluations, it is assumed from the user's input of a catalytic species that all mass-balance-allowed reactions present will proceed only with a catalyst. This is assumed due to the available lower energy pathways in the presence of the catalyst. This can then be changed by user input if a suitable model is not found, where the model generation is re-run to incorporate reactions that both do and do not proceed using the specified catalyst. When this rule is enforced during model generation,

as well as each previous rule, the total number of models for evaluation in the model chemical system is 173,711.

Rule #4: There must be at least one reaction from starting materials to products

As the starting concentrations are initially inputted into the system, it can be deduced which of the species are ‘starting materials’. In some instances, there may be species that are considered by the user to be the products of the reaction at the first time point, but in the context of this rule are also considered to be a starting material. For clarity, this definition of ‘starting material’ refers only to species that have mass in the initial concentrations. The rule states that there must be at least one term in the model that describes a decrease in the concentration of one of the identified starting materials. This logic-based rule ensures with certainty that the models that are evaluated feature terms that exhibit a change in concentrations. Although seemingly obvious, this rule is necessary to remove these commonly occurring models that would only lead to wasted computational resources.

For example, if there are four species in a chemical system (A, B, C, D) and their initial concentrations are 1 M, 0 M, 0 M and 0 M respectively, the only starting material identified is A. Many models that are feasible according to mass balance may not be feasible according to this rule, as they may not exhibit a change in concentration. Table 4.1 shows some examples of allowed models based on these starting concentrations that will progress onto the kinetic fitting stage of the approach, as well as disallowed models that will be discarded. As shown, entries 3 and 4 will both feature a change in concentrations, whereas entries 1 and 2 will not, and are hence discarded. When this rule is enforced, as well as each previous rule, the total number of models for evaluation in the model chemical system is 121,836.

Table 4.1: A table showing examples of allowed and disallowed models according to Rule #4.

Entry	1	2	3	4
Model	$B \rightarrow C$ $B + B \rightarrow D$	$B + B \rightarrow D$ $D \rightarrow A$	$A \rightarrow D$ $B \rightarrow C$	$A \rightarrow B + B$ $B \rightarrow C$
Allowed?	✗	✗	✓	✓

Rule #5: New 'starting materials' available from products

As an extension of Rule #4, relating to identifying 'starting materials' from initial concentrations, further starting materials can also be identified from the products of these earlier reactions. For example, in entry 4 in Table 4.1, B is not initially identified as a starting material. However, as A is a starting material forming B as a product, there is certain to be some concentration of B as time progresses, meaning that B can also be viewed as a starting material at some point. Therefore, as starting materials in entry 4 are A and B, both reactions in this model are allowed because they both feature a change in concentration from a starting material to another species.

Under Rule #5, models that feature model terms incurring a decrease in concentration from these newly identified starting materials *only* can be allowed. Therefore, entry 4 will be a model that is progressed onto the model evaluation stage. However, in entry 3 in Table 4.1, it can be noted that the initially identified starting material (A) forms D as a product, meaning that the starting materials in this model are A and D. The second model term features a reaction of B to form C, but this logic-based rule highlights that there is no possible scenario where B will be formed in order to further react. If this model were to progress onto model evaluation, it would certainly be an over-trained model and would rank lowly according to an AIC_C evaluation, as the second model term will not provide a benefit to the model over just having the first model term alone. Therefore, under Rule #5, entry 3 would be disallowed and not progress onto the model evaluation stage, thereby saving computational resources on this redundant model. When all rules are enforced, including Rule #5, the total number of evaluations for the model system is 98,725.

Final considerations for model generation

Other considerations were also made when compiling models in the model generation stage, specifically referring to the allowed orders of particular species. For catalytic species, the allowed orders were determined to be: 0.5, 1 and 2. This is because the order in a catalytic species of a system being greater than >2 or =1.5 are too rare to warrant serious consideration and would incur a significantly greater computational expense to evaluate all of the generated models.

Much thought was exercised as to whether a zero-order dependence on a catalyst is possible, as this discussion is missing from the literature and general teaching. It can be postulated that since the presence of a catalyst allows a more favourable lower energy pathway, there must therefore be some dependence on the concentration of the catalyst for this lower energy pathway to progress. This rationalisation infers that it is not possible for a zero-order dependence of catalyst, but there may always be selected cases where this can be empirically true. In any case, if a zero-order dependence of catalyst is possible, it is too unlikely to warrant serious consideration as this would generate many more models to be evaluated.

For non-catalytic systems, this approach attempts to build reaction models from base-level elementary reactions (where the both the stoichiometry and order of participating species are 1, for each occurrence of the species in the reaction scheme). However, there are two notable exceptions: 0 and 0.5 orders, which are inherently more empirical in nature. A zero-order dependence on a particular species in this approach will still have a stoichiometry of 1, whereby one molecule of this species is consumed in the reaction, but the reaction rate does not depend upon its concentration. A 0.5 order dependence will have both a stoichiometry of 0.5 and an order of 0.5. There are many other integer and non-integer empirical dependencies of varying stoichiometry and order that are possible in complex chemical reaction networks (CRNs). However, these dependencies are very rare and not common enough to warrant serious consideration for this computational approach as this would generate many more models to proceed into model evaluation, resulting in a much higher computational expense. Therefore, reactions are only considered for model generation with species that have orders in the rate laws of: 0, 0.5 or 1.

4.2.3 Overview

This updated computational approach to kinetic model selection and parameter identification was developed in MATLAB by adjusting the previously developed tool reported in Chapter 2 and Chapter 3. Firstly, the user inputs experimental data and the participating species into the system. The approach identifies all possible reactions that can happen based on mass balance, then generates all conceivable rate laws subject to basic chemical intuition in the inflation step (and subject to the user defining a catalytic species). Every

combination of these rate laws are then compiled into different reaction models, followed by logical rule enforcements to remove vast numbers of redundant or wasteful models. All remaining models are sequentially loaded, then kinetic parameters are optimised to increase convergence of ordinary differential equation (ODE) curves to the experimental data. Finally, each model is then statistically evaluated and ranked, highlighting the most likely model based on the experimental data supplied. The approach overview is highlighted in Figure 4.2.

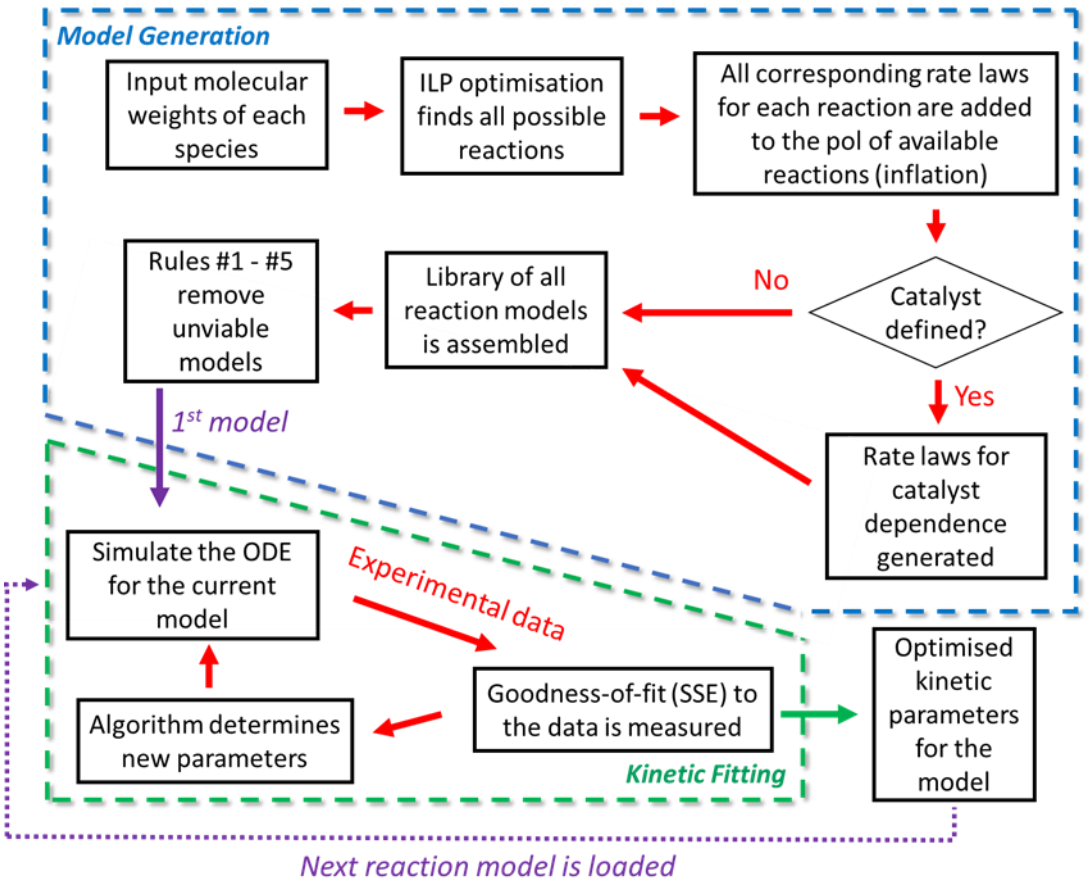


Figure 4.2: An overview of the updated computational approach to kinetic model and parameter determination.

This updated methodology represents a significantly improved, automatic tool for process development, by comprehensively evaluating every possible model without the need for chemical intuition, or more importantly: chemists' time. This approach has been upgraded to allow both catalytic reactions and reactions with non-integer orders to be identified, therefore vastly expanding its scope. However, it is still important to note that although this tool is powerful, it is ultimately data-driven and subsequently must only be used as a complementary tool. Therefore, after all model evaluations, the tool must still be

used in conjunction with real chemical intuition by an end user to determine that the models identified as most likely are accurate and consistent with the science.

4.3 Simulated verifications of the approach

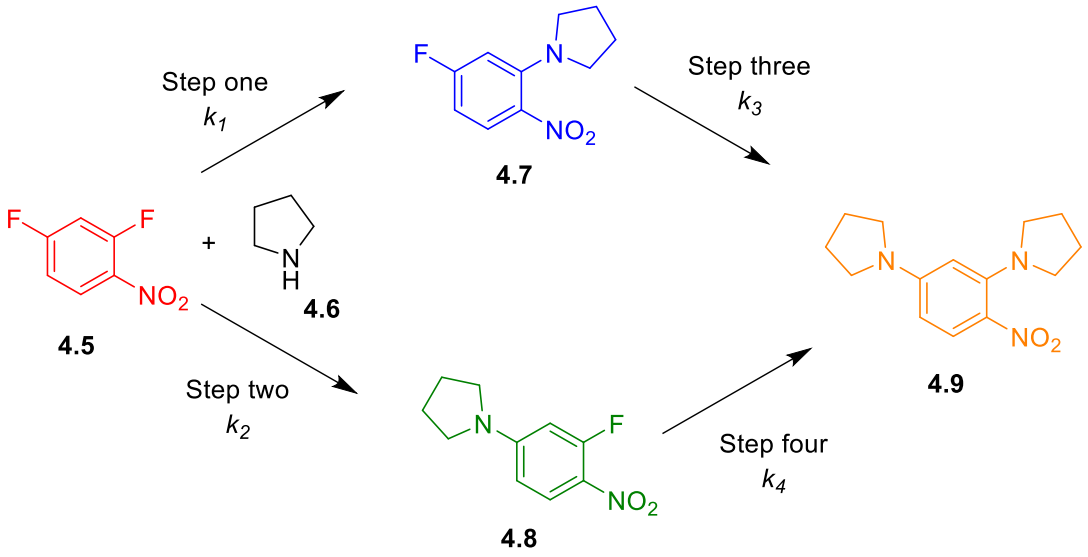
In the same manner as the previous methodology, it is first important to verify that the approach identifies the correct models and kinetic parameters before conducting any real experimentation. Several simulated case studies were conducted that feature different true kinetic models, showing how this methodology has been adapted to identify catalytic species' orders as well as non-integer and zero order species dependencies in chemical systems. These case studies feature reactions from the literature, whereby simulated experimental data is generated from known models and kinetic parameters.

To simulate the experimental data, as shown in Chapter 2.3, the true model and kinetic parameters reported in the literature source are inputted into MATLAB. The experiments are then simulated at particular temperatures and initial sets of concentrations, by using an ODE solver to evaluate the change in species concentrations over a given experimental timeframe. Individual data points are then extracted from this timeframe, which then represent experimental data points for the simulated case study. Up to ± 2 % relative error is then added to each measurement, to approximate this simulated data more closely to real experimental data.

4.3.1 Case study: S_NAr kinetics

The first simulated case study was the S_NAr reaction described in Scheme 4.2, where 2,4-difluoronitrobenzene (starting material, SM), **4.5**, reacts with pyrrolidine, **4.6**, to form either the ortho-substituted product, **4.7**, or the para-substituted product, **4.8** - both of which can then react further with pyrrolidine to form the bis-substituted product, **4.9**, and form hydrofluoric acid, **4.10**, as a byproduct in each step.[225] This case study features 'simple' elementary reactions in the true model as there are no non-integer or zero order dependencies. However, this scenario is still important to study, as the new approach must still be able to correctly identify model structures and kinetic parameters of systems that feature only these elementary reactions. In this case, both the old approach and the new approach were run in order to determine if

they both arrive at the same identified model, as they should. This should confirm that even in a multicomponent process, the new approach will still identify the correct elementary reactions instead of substituting non-integer or zero order rate laws.



Scheme 4.2: The reaction of 2,4-difluoronitrobenzene, **4.5**, with pyrrolidine, **4.6**, to form the ortho-substituted product, **4.7**, and the para-substituted product, **4.8**. Consecutive reactions then occur to form the bis-substituted product, **4.9**. In each reaction, hydrofluoric acid, **4.10**, is formed.[225]

4.3.1.1 Data acquisition

The multistep S_NAr reaction model reported by Hone *et al.* consists of 4 second-order reactions. These reactions were used to generate the ODEs shown in eqn. 4.1 - 4.6. The reported kinetic constants for these three steps were:

- Step one: $k_{90\,°C}$ = 0.579 M^{-1} s^{-1}, E_a = 33.3 kJ mol^{-1}
- Step two: $k_{90\,°C}$ = 0.027 M^{-1} s^{-1}, E_a = 35.3 kJ mol^{-1}
- Step three: $k_{90\,°C}$ = 0.009 M^{-1} s^{-1}, E_a = 38.9 kJ mol^{-1}
- Step four: $k_{90\,°C}$ = 0.016 M^{-1} s^{-1}, E_a = 44.8 kJ mol^{-1}

Five simulated data sets were generated at 30 °C, by using eqn. 2.7 to calculate k values, with the following initial concentrations:

- 1 M SM, **4.5**, 1.1 M pyrrolidine, **4.6**, 0 M ortho-substituted product, **4.7**, 0 M para-substituted product, **4.8**, 0 M bis-substituted product, **4.9** and 0 M hydrofluoric acid, **4.10**.

- 0.8 M SM, **4.5**, 1.5 M pyrrolidine, **4.6**, 0 M ortho-substituted product, **4.7**, 0 M para-substituted product, **4.8**, 0 M bis-substituted product, **4.9** and 0 M hydrofluoric acid, **4.10**.
- 1 M SM, **4.5**, 2 M pyrrolidine, **4.6**, 0 M ortho-substituted product, **4.7**, 0 M para-substituted product, **4.8**, 0 M bis-substituted product, **4.9** and 0 M hydrofluoric acid, **4.10**.
- 1 M SM, **4.5**, 2.5 M pyrrolidine, **4.6**, 0 M ortho-substituted product, **4.7**, 0 M para-substituted product, **4.8**, 0 M bis-substituted product, **4.9** and 0 M hydrofluoric acid, **4.10**.
- 1.2 M SM, **4.5**, 3 M pyrrolidine, **4.6**, 0 M ortho-substituted product, **4.7**, 0 M para-substituted product, **4.8**, 0 M bis-substituted product, **4.9** and 0 M hydrofluoric acid, **4.10**.

$$\frac{d[\mathbf{4.5}]}{dt} = -k_1[\mathbf{4.5}][\mathbf{4.6}] - k_2[\mathbf{4.5}][\mathbf{4.6}] \quad \text{eqn. 4.1}$$

$$\frac{d[\mathbf{4.6}]}{dt} = -k_1[\mathbf{4.5}][\mathbf{4.6}] - k_2[\mathbf{4.5}][\mathbf{4.6}] - k_3[\mathbf{4.7}][\mathbf{4.6}] - k_4[\mathbf{4.8}][\mathbf{4.6}] \quad \text{eqn. 4.2}$$

$$\frac{d[\mathbf{4.7}]}{dt} = k_1[\mathbf{4.5}][\mathbf{4.6}] - k_3[\mathbf{4.7}][\mathbf{4.6}] \quad \text{eqn. 4.3}$$

$$\frac{d[\mathbf{4.8}]}{dt} = k_2[\mathbf{4.5}][\mathbf{4.6}] - k_4[\mathbf{4.8}][\mathbf{4.6}] \quad \text{eqn. 4.4}$$

$$\frac{d[\mathbf{4.9}]}{dt} = k_3[\mathbf{4.7}][\mathbf{4.6}] + k_4[\mathbf{4.8}][\mathbf{4.6}] \quad \text{eqn. 4.5}$$

$$\frac{d[\mathbf{4.10}]}{dt} = k_1[\mathbf{4.5}][\mathbf{4.6}] + k_2[\mathbf{4.5}][\mathbf{4.6}] + k_3[\mathbf{4.7}][\mathbf{4.6}] + k_4[\mathbf{4.8}][\mathbf{4.6}] \quad \text{eqn. 4.6}$$

2 % relative error was then added to each data point in these simulated data sets, then used as the inputted experimental data for both the old and new computational approach. These data sets can be found in Chapter 7.4.1.

4.3.1.2 Results and discussion

As the initial ILP optimisation proceeds in the same manner for both the old and new approaches, the mass-balance-allowed reactions identified in each case are the same. It is only the subsequent rate laws and models that differ in

the model generation steps between the old and new approaches. From the 6 species inputted, there are 16 mass-balance-allowed reactions that were identified, shown in eqn. 4.7 - 4.22:

4.8 → **4.7**	eqn. 4.7
4.7 → **4.8**	eqn. 4.8
4.5 + **4.6** → **4.7** + **4.10**	eqn. 4.9
4.7 + **4.10** → **4.5** + **4.6**	eqn. 4.10
4.5 + **4.9** → **4.7** + **4.7**	eqn. 4.11
4.5 + **4.6** → **4.8** + **4.10**	eqn. 4.12
4.6 + **4.7** → **4.9** + **4.10**	eqn. 4.13
4.7 + **4.7** → **4.5** + **4.9**	eqn. 4.14
4.9 + **4.10** → **4.6** + **4.8**	eqn. 4.15
4.8 + **4.10** → **4.5** + **4.6**	eqn. 4.16
4.6 + **4.8** → **4.9** + **4.10**	eqn. 4.17
4.5 + **4.9** → **4.7** + **4.8**	eqn. 4.18
4.5 + **4.9** → **4.8** + **4.8**	eqn. 4.19
4.9 + **4.10** → **4.6** + **4.7**	eqn. 4.20
4.7 + **4.8** → **4.5** + **4.9**	eqn. 4.21
4.8 + **4.8** → **4.5** + **4.9**	eqn. 4.22

Old approach

When running the old approach, every combination of the 16 identified mass-balance-allowed transformations was constructed, resulting in 65535 unique models to be evaluated. All of these resulting models were evaluated by the approach, which incurred a computation time of 13 hours (see Chapter 7.6.2 for details on computer specifications).

Each of these models were ranked based on their AIC_C evaluation and it was found that the highest ranked model was also the correct model, with kinetic parameters optimised to values that were very close to the literature values used to generate the data. As with previous simulations described in Chapter 2.3, the

optimisation is assumed to be correct as the error in the fitted parameters is negligible, considering that artificial error was added to the data set. The top five performing models are shown in Table 4.2, where the optimised k values are shown, as well as their corresponding SSE and AIC_C values.

Table 4.2: A table showing the top five ranked models for the S_NAr case study as identified by the old approach. The kinetic parameters for each reaction within the model are shown, as well as the SSE and AIC_C evaluation.

Model rank	Model	k_x /M^{-1} min^{-1}	SSE /M	AIC_C
1	4.5 + 4.6 → 4.7 + 4.10 4.5 + 4.6 → 4.8 + 4.10 4.7 + 4.6 → 4.9 + 4.10 4.8 + 4.6 → 4.9 + 4.10	0.0656 0.0026 0.0007 0.0008	0.0040	-306.27
2	4.5 + 4.6 → 4.7 + 4.10 4.5 + 4.6 → 4.8 + 4.10 4.7 + 4.6 → 4.9 + 4.10 4.8 + 4.6 → 4.9 + 4.10 4.5 + 4.9 → 4.7 + 4.8 4.9 + 4.10 → 4.6 + 4.8	0.0657 0.0026 0.0007 0.0010 0.0065 <0.0001	0.0039	-301.25
3	4.5 + 4.6 → 4.7 + 4.10 4.5 + 4.6 → 4.8 + 4.10 4.7 + 4.6 → 4.9 + 4.10 4.8 + 4.8 → 4.9 + 4.5	0.0657 0.0027 0.0007 0.0174	0.0044	-300.97
4	4.5 + 4.6 → 4.7 + 4.10 4.5 + 4.6 → 4.8 + 4.10 4.7 + 4.6 → 4.9 + 4.10 4.8 + 4.6 → 4.9 + 4.10 4.5 + 4.9 → 4.7 + 4.8	0.0663 0.0028 0.0007 0.0010 0.0026	0.0042	-300.53
5	4.5 + 4.6 → 4.7 + 4.10 4.5 + 4.6 → 4.8 + 4.10 4.7 + 4.6 → 4.9 + 4.10 4.8 + 4.6 → 4.9 + 4.10	0.0655 0.0026 0.0007 0.0009	0.0041	-299.46

	4.5 + 4.9 → 4.7 + 4.7 4.5 + 4.9 → 4.8 + 4.8	0.0143 0.0065		

New approach

When running the new approach, all allowed combinations of the mass-balance-allowed reactions and corresponding non-integer and zero order rate laws were compiled, subject to the rules described earlier in Chapter 4.2.2. This model generation resulted in 141,505 unique models to be evaluated by the approach, which incurred a 28 hour computation time (see Chapter 7.6.2 for details on computer specifications).

Each of these models were ranked based on their AIC_C, and it was found that the highest ranked model was also the correct model, with kinetic parameters optimised to values that were very close to the literature values that were used to generate the data. As highlighted in the old approach section, the optimisation is assumed to be correct as the error in the fitted parameters is negligible, considering that artificial error was added to the dataset. The top five performing models from this new approach are shown in Table 4.3, where the optimised k values are shown, as well as their corresponding SSE and AIC_C values.

Table 4.3: A table showing the top five ranked models for the S_NAr case study as identified by the new approach. The kinetic parameters for each reaction within the model are shown, as well as the SSE and AIC_C evaluation. α denotes a variable molar dependence within the kinetic parameter units, depending on the rate law.

Model rank	**Model**	**k_x /M^{α} min^{-1}**	**SSE /M**	**AIC_C**
1	4.5 + 4.6 → 4.7 + 4.10 4.5 + 4.6 → 4.8 + 4.10 4.7 + 4.6 → 4.9 + 4.10 4.8 + 4.6 → 4.9 + 4.10	0.0656 0.0026 0.0007 0.0008	0.0040	-306.27
2	4.5 + 4.6 → 4.7 + 4.10 4.6 + 4.7 → 4.9 + 4.10 4.8^0 → 4.7 4.7 → 4.8	0.0680 0.0007 0.0501 0.0394	0.0042	-303.33

3	4.5 + 4.6 → 4.7 + 4.10 4.5 + 4.6 → 4.8 + 4.10 4.7 + 4.6 → 4.9 + 4.10 4.8 + 4.8 → 4.9 + 4.5	0.0657 0.0027 0.0007 0.0174	0.0044	-300.97
4	4.5 + 4.6 → 4.7 + 4.10 4.5 + 4.6 → 4.8 + 4.10 4.7 + 4.6 → 4.9 + 4.10 $4.8 + 4.10^{0.5} \rightarrow 4.5 + 4.6$	0.0640 0.0024 0.0007 0.0008	0.0061	-285.21
5	4.5 + 4.6 → 4.7 + 4.10 4.7 + 4.6 → 4.9 + 4.10 4.7 → 4.8 4.8 → 4.7	0.0670 0.0007 0.0041 0.0997	0.0065	-282.76

Discussion

This initial simulated case study has shown that it is still possible to identify the correct reaction model and kinetic parameters with this new approach for a 'simple' reaction system featuring a first-order dependencies on all species. The most-likely identified model, and corresponding SSE/AIC_C values, were the same in this instance for both the old and new approach which is to be expected if the new approach is functioning correctly. As shown in the old approach, there are models that have a greater convergence to the experimental data (and hence a lower SSE), but the statistical analysis correctly assigns these models an unfavourable AIC_C based on their extra (insignificant) model terms. However, as the initial limit of 4 terms in a model is enforced in the new approach, it can be determined by the user that compiling and evaluating the 5-term-models is not required based on the marginal possible SSE benefits. These small benefits are very unlikely to warrant extra terms based on AIC_C evaluation, as the convergence to experimental data is already excellent, meaning that the model identified is very likely to be correct. An example kinetic plot showing the fit of the identified model to the experimental data is shown in Figure 4.3 - this is the fifth simulated experiment where the starting concentrations were 1.2 M 2,4-difluoronitrobenzene and 3 M pyrrolidine at 30 °C.

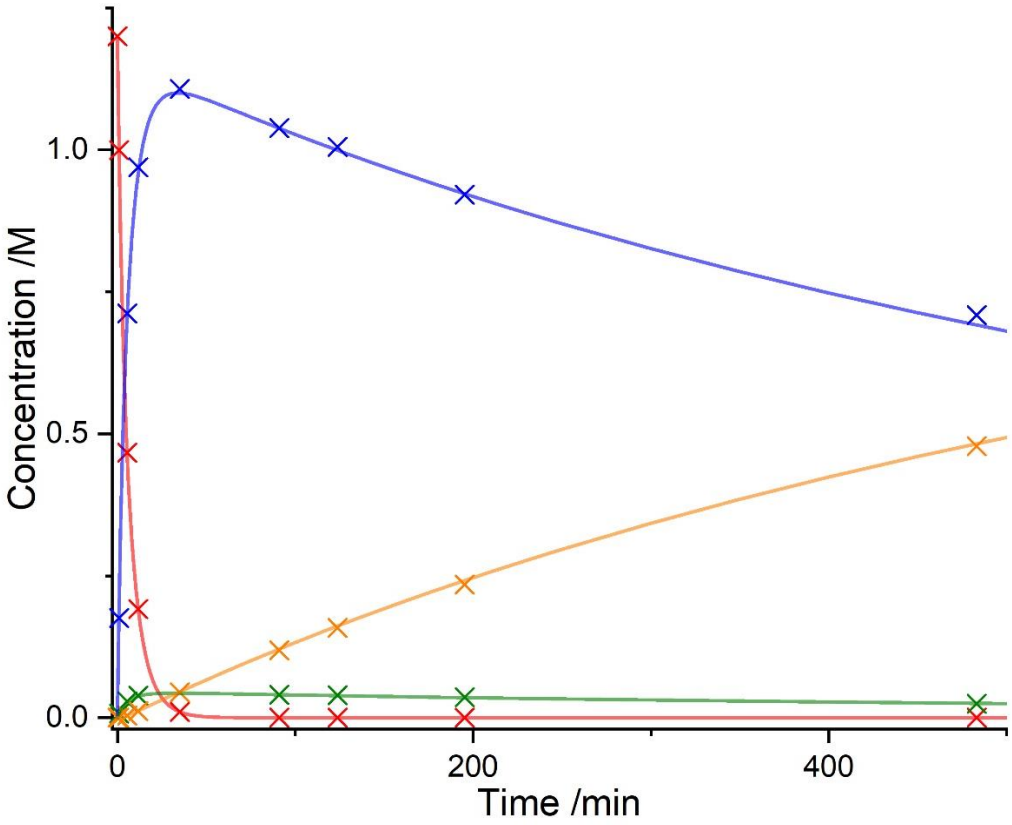


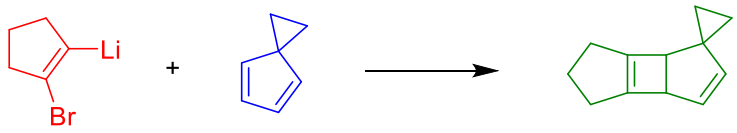
Figure 4.3: The fit of the identified model and kinetic parameters to the generated experimental data with starting concentrations of 1.2 M 2,4-difluoronitrobenzene and 3 M pyrrolidine at 30 °C. Where: x = 2,4-difluoronitrobenzene, x = ortho-substituted product, x = para-substituted product, x = bis-substituted product, — = 2,4-difluoronitrobenzene (ODE), — = ortho-substituted product (ODE), — = para-substituted product (ODE), — = bis-substituted product. Plot only shown to 500 minutes to the show curvature of the initial data points.

It has been shown that the new approach can identify the correct model and kinetic parameters for a multistep process with first-order dependencies on all species, with the same accuracy as expected from the old approach. Therefore, it is now possible to confidently move on to simulated case studies where rate laws are more complex, featuring catalytic reactions and orders that are non-integer or zero.

4.3.2 Case study: Pentyne kinetics

This simulated case study features a chemical system that reacts with a zero-order dependence on one of the species.[241] The starting material, **4.11**, reacts with the spirodiene, **4.12**, to form the product, **4.13**, and LiBr, **4.14**, as shown in Scheme 4.3. The true rate law, however, shows a first-order dependence on the starting material, and a zero-order dependence on the spirodiene. Interestingly, this reaction order has been used as evidence for the existence of a pentyne species, as this order suggests an initial reaction of the starting material to form

a much more reactive species, leading to the zero-order dependence on the spirodiene.



4.11 **4.12** **4.13**

Scheme 4.3: The reaction of the starting material, **4.11**, with the spirodiene, **4.12**, to form the product, **4.13**, and LiBr, **4.14**.[241]

4.3.2.1 Data acquisition

The overall first-order reaction model reported by Gilbert *et al.*[241] was used to generate the ODEs shown in eqn. 4.23 - 4.26. Three isothermal simulated data sets were generated using these ODEs with the k value of 0.00846 min^{-1} at -78 °C, with the following initial concentrations:

- 1 M 2-bromocyclopentenyl lithium (starting material), **4.11**, 0.8 M spirodiene, **4.12**, 0 M product, **4.13**, 0 M LiBr, **4.14**.
- 1.2 M 2-bromocyclopentenyl lithium (starting material), **4.11**, 1.5 M spirodiene, **4.12**, 0 M product, **4.13**, 0 M LiBr, **4.14**.
- 1 M 2-bromocyclopentenyl lithium (starting material), **4.11**, 1 M spirodiene, **4.12**, 0 M product, **4.13**, 0 M LiBr, **4.14**.

$$\frac{d[\mathbf{4.11}]}{dt} = -k_1[\mathbf{4.11}] \qquad \text{eqn. 4.23}$$

$$\frac{d[\mathbf{4.12}]}{dt} = -k_1[\mathbf{4.11}] \qquad \text{eqn. 4.24}$$

$$\frac{d[\mathbf{4.13}]}{dt} = k_1[\mathbf{4.11}] \qquad \text{eqn. 4.25}$$

$$\frac{d[\mathbf{4.14}]}{dt} = k_1[\mathbf{4.11}] \qquad \text{eqn. 4.26}$$

2 % relative error was then added to these simulated data sets, then used as the inputted experimental data for the computational approach. These data sets can be found in Chapter 7.4.2.

4.3.2.2 Results and discussion

From these 3 species, there are 2 mass-balance-allowed reactions that can be identified, shown in eqn. 4.27 and eqn. 4.28. From these two reactions, the corresponding rate laws were generated and all models were compiled, subject to the rules defined in Chapter 4.2.2 - this resulted in 30 unique models to be evaluated by the approach.

4.11 + 4.12 → 4.13 + 4.14 eqn. 4.27

4.13 + 4.14 → 4.11 + 4.12 eqn. 4.28

All 30 models were evaluated by the computational approach, incurring a computation time of around <1 minute (see Chapter 7.6.1 for details on computer specifications). Each of these models were ranked based on their AIC_C, and it was found that the highest ranked model was also the correct model, with kinetic parameters optimised to values that were very close to the generated values. The top five performing models are shown in Table 4.4, where the optimised k values are shown, as well as their corresponding SSE and AIC_C values.

Table 4.4: A table showing the top five ranked models for the pentyne case study as identified by the new approach. The kinetic parameters for each reaction within the model are shown, as well as the SSE and AIC_C evaluation. α denotes a variable molar dependence within the kinetic parameter units, depending on the rate law.

Model rank	Model	k_x /M^{α} min^{-1}	SSE /M	AIC_C
1	$4.11 + 4.12^0 \rightarrow 4.13 + 4.14$	0.0085	0.0046	-134.903
2	$4.11 + 4.12^0 \rightarrow 4.13 + 4.14$ $4.13 + 4.14 \rightarrow 4.11 + 4.12$	0.0085 <0.0001	0.0046	-132.515
3	$4.11 + 4.12^0 \rightarrow 4.13 + 4.14$ $4.13 + 4.14^{0.5} \rightarrow 4.11 + 4.12$	0.0085 <0.0001	0.0046	-132.515
4	$4.11 + 4.12^0 \rightarrow 4.13 + 4.14$ $4.13^0 + 4.14 \rightarrow 4.11 + 4.12$	0.0085 <0.0001	0.0046	-132.514
5	$4.11 + 4.12^0 \rightarrow 4.13 + 4.14$ $4.13 + 4.14^0 \rightarrow 4.11 + 4.12$	0.0085 <0.0001	0.0046	-132.514

Interestingly, all five of these top-ranked models all have the same SSE values and ranks 2 - 5 all have very similar AIC_C evaluations. It is clear from these models that the primary reason that they score highly is due to the

presence of the true model term: 4.11 + 4.12^0 → 4.13 + 4.14. One extra model term is present that is rendered negligible, but as the algorithm is forced to fit it, it is fitted with a value <0.0001. In the old approach, these negligible reactions could be found but not repeated, whereas in the new approach, every variation of the rate law of a particular negligible reaction can be repeated and score highly according to AIC_C.

Nonetheless, the highest ranked model according to the approach is the true kinetic model and the corresponding k value was also identified correctly. This case study shows that it is possible for the approach to identify models that feature reactions that do not have typical species order dependencies of 1, therefore also that it can be used for real experimentation featuring zero or non-integer order model terms. An example kinetic plot showing the fit of the identified model to the simulated experimental data is shown in Figure 4.4 - this is the first experiment where the starting concentrations were 1 M starting material and 0.8 M spirodiene at -78 °C.

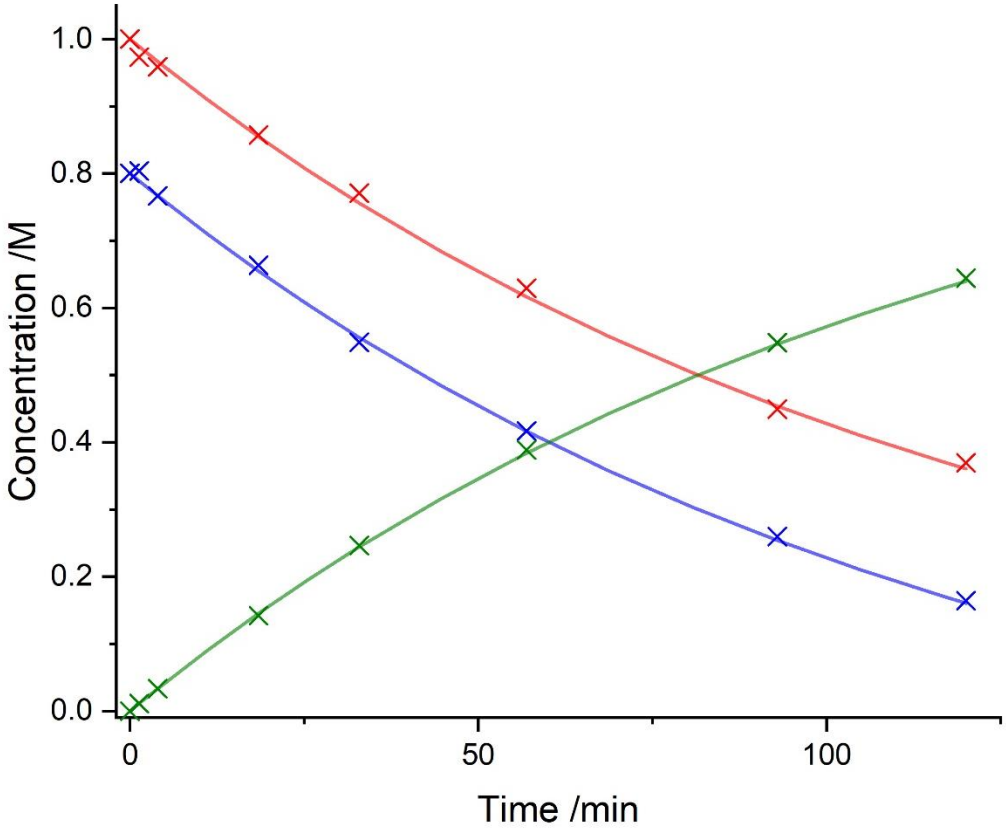
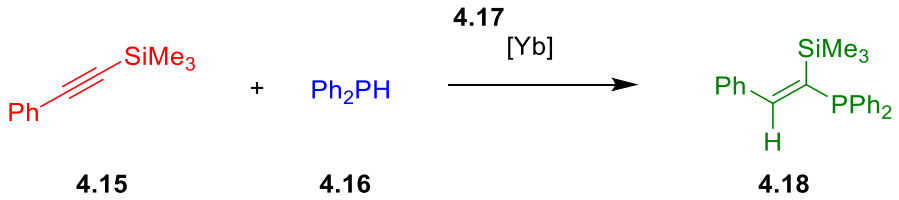


Figure 4.4: The fit of the identified model and kinetic parameters to the generated experimental data with starting concentrations of 1 M starting material and 0.8 M spirodiene at -78 °C. Where: x = starting material, x = spirodiene, x = product, — = starting material (ODE), — = spirodiene (ODE), — = product (ODE).

4.3.3 Case study: Ytterbium catalysis

The final simulated case study involves a ytterbium catalysed reaction with 'non-normal' empirical rate laws.[242] 1-Phenyl-2-trimethylsilylacetylene (starting material), **4.15**, reacts with diphenylphosphine, **4.16**, in the presence of a ytterbium catalyst ([Yb(η^2-Ph$_2$CNPh)(hmpa)$_3$]), **4.17**, to form the product, **4.18**, as shown in Scheme 4.4.



Scheme 4.4: The reaction of the starting material, **4.15**, with diphenylphosphine, **4.16**, in the presence of the ytterbium catalyst, **4.17**, to form the product, **4.18**, where [Yb]: Yb(η^2-Ph$_2$CNPh)(hmpa)$_3$.[242]

4.3.3.1 Data acquisition

The complex order dependencies for the reaction model reported by Takaki *et al.*[242] consists of 1 reaction that was used to generate the ODEs shown in eqn. 4.29 - 4.32. In the true empirical rate law, the order in the starting material is 1, the order in diphenylphosphine is 0 and the order of the ytterbium catalyst is 2. Four isothermal simulated data sets were generated at 25 °C using these ODEs, with the k value of 9.324 M^2 h^{-1}, with the following initial concentrations:

- 2 M starting material, **4.15**, 2 M diphenylphosphine, **4.16**, 0.2 M ytterbium catalyst, **4.17**, 0 M product, **4.18**.
- 1.5 M starting material, **4.15**, 2 M diphenylphosphine, **4.16**, 0.1 M ytterbium catalyst, **4.17**, 0 M product, **4.18**.
- 1.4 M starting material, **4.15**, 1.8 M diphenylphosphine, **4.16**, 0.3 M ytterbium catalyst, **4.17**, 0 M product, **4.18**.
- 2.2 M starting material, **4.15**, 1.6 M diphenylphosphine, **4.16**, 0.25 M ytterbium catalyst, **4.17**, 0 M product, **4.18**.

$$\frac{d[\mathbf{4.15}]}{dt} = -k\,[\mathbf{4.15}][\mathbf{4.17}]^2 \qquad \text{eqn. 4.29}$$

$$\frac{d[\mathbf{4.16}]}{dt} = -k\,[\mathbf{4.15}][\mathbf{4.17}]^2 \qquad \text{eqn. 4.30}$$

$$\frac{d[\mathbf{4.17}]}{dt} = -2k\,[\mathbf{4.15}][\mathbf{4.17}]^2 + 2k\,[\mathbf{4.15}][\mathbf{4.17}]^2 \qquad \text{eqn. 4.31}$$

$$\frac{d[\mathbf{4.18}]}{dt} = k\,[\mathbf{4.15}][\mathbf{4.17}]^2 \qquad \text{eqn. 4.32}$$

2 % relative error was then added to these simulated data sets, then used as the inputted experimental data for the computational approach. These data sets can be found in Chapter 7.2.1.

4.3.3.2 Results and discussion

Based on the inputs of the molecular weights for this system, there are two mass-balance-allowed reactions. However, as it is known that this is a catalytic system, these two reactions must be facilitated by the user-defined ytterbium catalyst, **4.17**, shown in eqn. 4.33 and 4.34. It is then the role of the approach to identify all combinations of allowed reactions and rate laws to determine which is the most likely model.

$$\mathbf{4.15 + 4.16} \xrightarrow{\mathbf{4.17}} \mathbf{4.18} \qquad \text{eqn. 4.33}$$

$$\mathbf{4.15 + 4.16} \xrightarrow{\mathbf{4.17}} \mathbf{4.18} \qquad \text{eqn. 4.34}$$

After the model generation stage, 105 models were identified and evaluated by the computational approach, incurring a time of around 1 minute (see Chapter 7.6.2 for details on computer specifications). Each of these models were ranked based on their AIC_C, and it was found that the highest ranked model was also the correct model, with the k value optimised to a value that was very close to the value used for data generation. The top five performing models are shown in Table 4.5, where the optimised k values are shown, as well as their corresponding SSE and AIC_C values.

Table 4.5: A table showing the top five ranked models for the ytterbium catalysis case study as identified by the new approach. The kinetic parameters for each reaction within the model are shown, as well as the SSE and AIC_C evaluation. α denotes a variable molar dependence within the kinetic parameter units, depending on the rate law.

Model rank	Model	k_x /M$^{\alpha}$ h^{-1}	SSE /M	AIC_C
1	$4.15 + 4.16^0 \xrightarrow{4.17^2} 4.18$	9.3191	9.28 x 10^{-4}	-241.38

2	$4.15 + 4.16^0 \xrightarrow{4.17^2} 4.18$ $4.18 \xrightarrow{4.17^{0.5}} 4.15 + 4.16$	9.3191 <0.0001	9.28×10^{-4}	-239.10
3	$4.15 + 4.16^0 \xrightarrow{4.17^2} 4.18$ $4.18 \xrightarrow{4.17^1} 4.15 + 4.16$	9.3191 <0.0001	9.28×10^{-4}	-239.10
4	$4.15 + 4.16^0 \xrightarrow{4.17^2} 4.18$ $4.18 \xrightarrow{4.17^2} 4.15 + 4.16$	9.3191 <0.0001	9.28×10^{-4}	-239.10
5	$4.15 + 4.16 \xrightarrow{4.17^2} 4.18$	8.0784	1.1549	-13.35

As in the previous pentyne case study, four of these top-ranked models all have the same SSE values and ranks 2 - 4 all have very similar AIC_C evaluations. The reasoning for this appears to be the same as previously discussed, where the primary reason that they score highly is due to the presence of the true model term: $4.15 + 4.16^0 \xrightarrow{4.17^2} 4.18$. One extra model term is present in ranks 2 - 4 that the algorithm is forced to fit although it does not provide any significance to the model, hence it is fitted with a value <0.0001.

Nonetheless, the highest ranked model according to the approach is the true kinetic model and the corresponding k value was also identified correctly. This case study shows not only that it is possible to identify models that feature catalytic species, but also that the orders of the catalytic and non-catalytic species can be identified. As a ‘typical’ catalytic example is more likely to be first-order with respect to all species, the ability that the approach has shown to identify non-normal order dependencies in the rate law for multiple species inspires confidence that the approach will be successful in its use in real experimentation. An example kinetic plot showing the fit of the identified model to the simulated experimental data is shown in Figure 4.5 - this is the second experiment where the starting concentrations were 1.5 M starting material, 2 M diphenylphosphine and 0.1 M ytterbium catalyst at 22 °C.

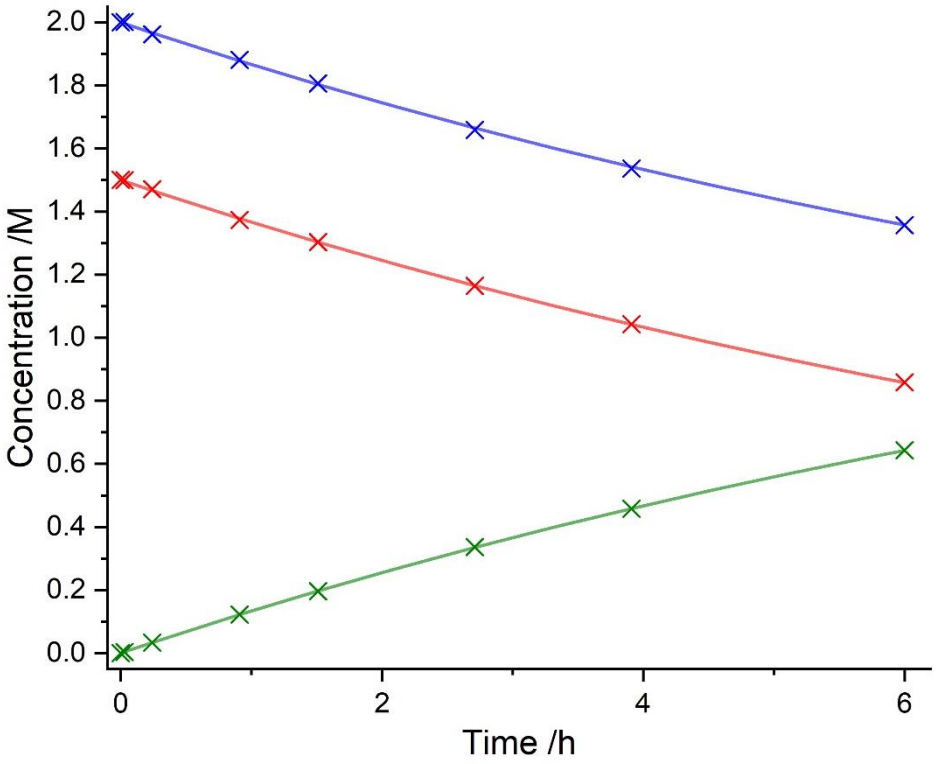


Figure 4.5: The fit of the identified model and kinetic parameters to the generated experimental data with starting concentrations of 1.5 M starting material, 2 M diphenylphosphine and 0.1 M ytterbium catalyst at 22 °C. Where: x = starting material, x = diphenylphosphine, x = product, — = starting material (ODE), — = diphenylphosphine (ODE), — = product (ODE).

4.4 Conclusion

It has been shown that the previously built computational approach to kinetic modelling can be adapted to catalytic systems and rate laws with non-normal order dependencies. This new approach has been proven to work successfully for three simulated case studies featuring different modelling criteria, with sets of experimental data inputs with artificial errors. The success shown from these studies inspires confidence that the approach can be implemented with real experimental data.

The scope of the approach has increased substantially following the improvements in this work and can now be used for many more chemical applications. However, due to the comprehensive nature of the approach, the computational expense has also increased as there are many more reactions that must be considered when building reaction models. This computational stress, as before, can be reduced by replacing existing computer hardware with more computer logical cores - this is because the code is highly parallelised and

the speed of the approach is directly proportional to the number of computer cores available. Although the approach is very powerful, this computational expense may still be cumbersome for desktop computers depending on the number of species involved in a particular chemical system. However, as time passes, processing technology consistently improves and there will undoubtedly be a time where millions of models can be evaluated at a desk in a matter of hours via comprehensive methodologies such as this.

This comprehensive approach to kinetic model and parameter determination has been shown to be successful in simulated case studies - the next step is to prove the efficiency of this methodology with real experimental case studies. The aim of these case studies is to show that this approach can correctly identify models of various-complexity chemical processes, thereby realising the main aim of this project in developing an automated kinetic analysis tool with a wide scope for process development.

Chapter 5 : Experimental applications of the new approach

5.1 Introduction

With the development of the new computational approach complete and verified with multiple simulated case studies, the viability of the approach for real experimental data must also be tested. Real experimental case studies are to be conducted to measure the ability of this new and final computational approach.

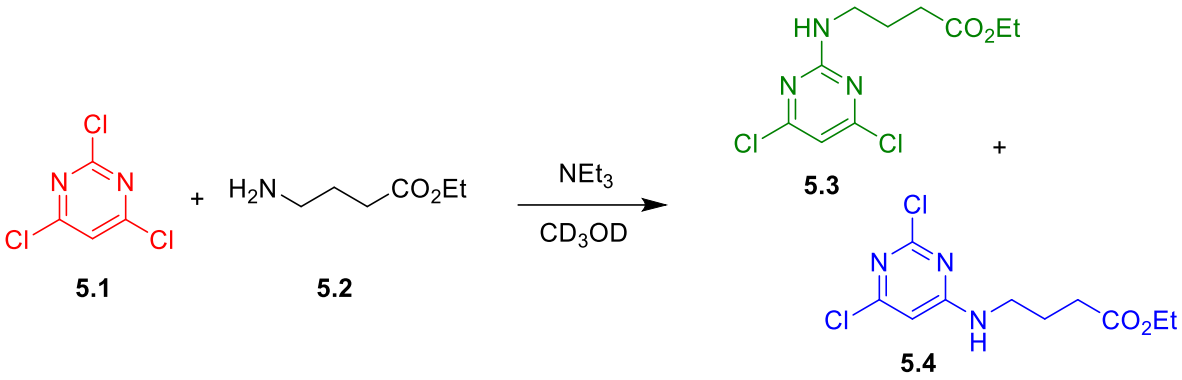
As the scope of this new approach has been broadened to include catalytic reactions and species order dependencies of zero or 0.5, it is important to study chemical systems that may exhibit these properties to prove that the new approach can identify these models. It is also important that the approach can continue to identify more ‘simple’ models that have species order dependencies of 1, as well as utilise kinetic data from any means. This is because the kinetic motifs often cannot be determined *a priori* and the beneficial feature of this methodology is that models are identified and ranked according to statistical measures. This removes bias and therefore considers all rate laws for every possible model, meaning that the approach does not know initially if a

model is supposed to contain a zero order term or not, for example, and will only determine this after the comprehensive model evaluations.

Experimental case studies from any reaction process (batch, flow, etc.) and analytical technique (HPLC, NMR, etc.) will be used in these final experimental verifications. These case studies will prove that the approach can identify models including 'normal', zero or 0.5 species order dependencies for both catalytic and non-catalytic processes. The research project aim will thereby be realised, by ultimately developing a powerful, automated methodology to be used in process development for the modelling of real chemical systems.

5.2 Case study: S_NAr kinetics

The first reaction system explored using the new computational approach was the presupposed model of the reaction of 2,4,6-trichloropyrimidine (SM), **5.1**, with ethyl 4-aminobutanoate, **5.2**, to form the major 2-substituted S_NAr product, **5.3**, and the minor 4-substituted S_NAr product, **5.4**, as shown in Scheme 5.1. As the ethyl 4-aminobutanoate is present in its hydrochloride salt, the hydrochloric acid present (and generated from subsequent S_NAr reactions as **5.5**) is neutralised by an excess of triethylamine.



Scheme 5.1: The reaction of 2,4,6-trichloropyrimidine (SM), **5.1**, with ethyl 4-aminobutanoate, **5.2**, to form the major 2-substituted S_NAr product, **5.3**, and the minor 4-substituted S_NAr product, **5.4**.

As the desired major product from this reaction can be further reacted to synthesise bioactive derivatives of pharmaceuticals,[243, 244] this initial transformation is of interest in the process development of these molecules. A collaborator in the Spring group at Cambridge University, Hikaru Seki, has a particular interest in these bioactive molecules for his research project and

wanted to maximise the synthetic yield of the 2-substituted S_NAr product, **5.3**, by obtaining kinetic understanding of the system. Therefore, all practical experimental work in this case study was conducted by Hikaru Seki at Cambridge University, then the results were forwarded to Leeds for kinetic analysis.

Four kinetic experiments were conducted, at -25 °C, 0 °C, 25 °C and 50 °C, with three replicate experiments at 25 °C. Each experiment was conducted in batch in an NMR tube within a 500 MHz NMR spectrometer, which allowed continuous NMR acquisition at a rate of approximately one sample per 80 seconds. This continuous reaction monitoring generates a large data density for the subsequent kinetic analysis using the new computational approach, by monitoring the proton peaks of the individual species, shown in Figure 5.1. Full experimental details can be found in Chapter 7.5.1.

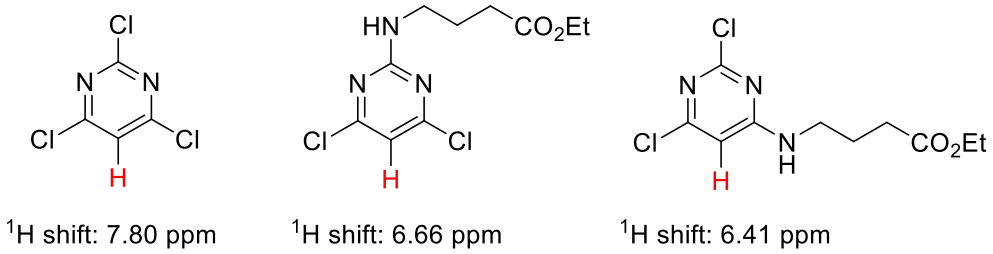


Figure 5.1: The distinctive proton peaks, shown in red, that are monitored via NMR as time progresses in the S_NAr case study.

Based on the five species identified, there were six reactions that were calculated to be possible based on mass balance - these reactions are shown as eqn. 5.1 - 5.6. From these reactions, all corresponding rate laws were generated and all models were compiled, subject to the rules defined in Chapter 4.2.2. This resulted in 3320 unique models to be evaluated by the approach.

5.3 → 5.4	eqn. 5.1
5.4 → 5.3	eqn. 5.2
5.1 + 5.2 → 5.3 + 5.5	eqn. 5.3
5.3 + 5.5 → 5.1 + 5.2	eqn. 5.4
5.1 + 5.2 → 5.4 + 5.5	eqn. 5.5
5.4 + 5.5 → 5.1 + 5.2	eqn. 5.6

All 3320 models were evaluated by the computational approach, incurring a computation time of around 9 hours (see Chapter 7.6.1 for details on computer

specifications). Each of these models were then ranked based on their AIC_C evaluation, and it was found that the most likely representation of the system is the model shown in Scheme 5.1 with 'typical' first-order species dependencies in the rate laws. The approach also determined the kinetic parameters of the transformation to the major product: $k_{25\,°C}$ = 0.499 ± 0.006 $M^{-1}\,min^{-1}$, E_a = 44.19 ± 0.57 kJ mol^{-1} and to the minor product: $k_{25\,°C}$ = 0.384 ± 0.009 $M^{-1}\,min^{-1}$, E_a = 36.57 ± 0.88 kJ mol^{-1}. This model and the corresponding kinetic parameters allowed a fit to the experimental data with an average residual of less than 2.6 x 10^{-4} M and is shown in Figure 5.2. The top 5 ranked models are shown in Table 5.1, with their respective SSE error metric and AIC_C evaluation.

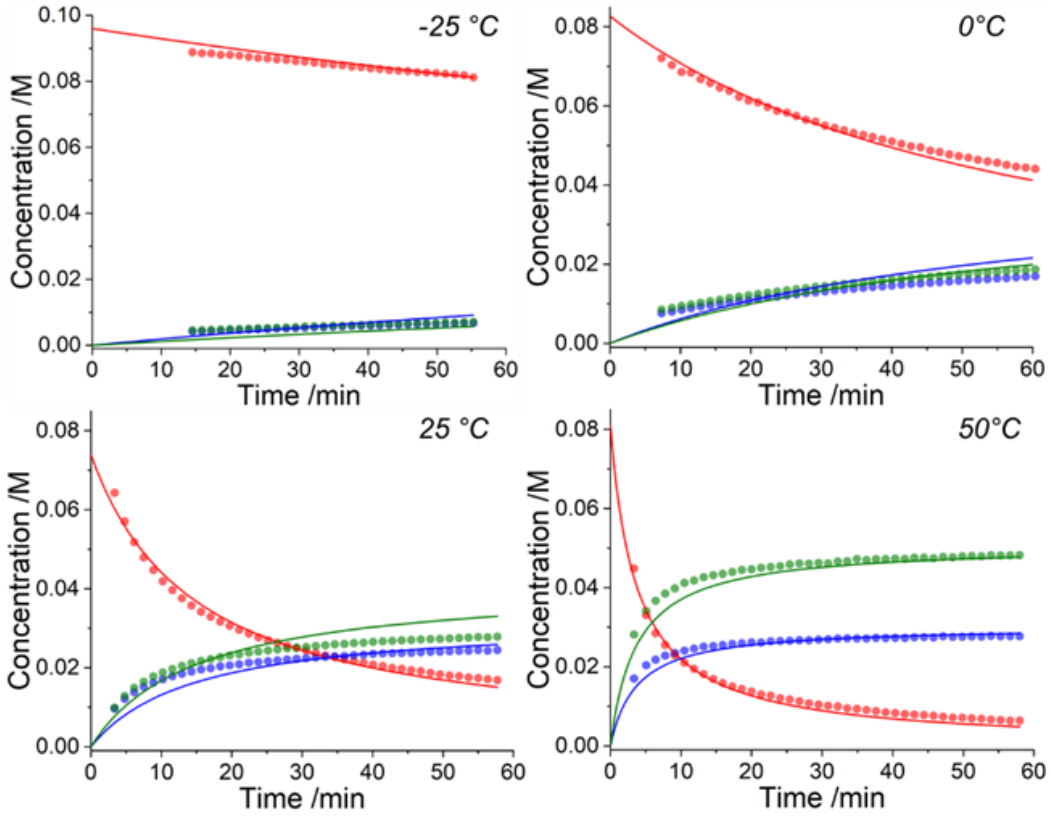


Figure 5.2: Kinetic profiles for four kinetic experiments at -25 °C, 0 °C, 25 °C and 50 °C, where: ● = starting material, ● = 2-substituted product, ● = 4-substituted product, — = starting material (ODE), — = 2-substituted product (ODE), — = 4-substituted product (ODE). See Chapter 7.5.1 for full experimental conditions and raw data.

Table 5.1: Evaluation of the feasibility of each reaction model for the S_NAr case study, where α is variable for each model depending on the overall model order.

Rank	Reaction Model	Kinetic parameters		SSE /10^3 M	AIC_C
		$k_{25\,°C}$ /M^{α} min^{-1}	E_a /kJ mol^{-1}		
1	5.1 + 5.2 → 5.3 + 5.5	0.499	44.19	1.614	-51.91

	$5.1 + 5.2 \rightarrow 5.4 + 5.5$	0.384	36.57		
2	$5.1 + 5.2 \rightarrow 5.3 + 5.5$ $5.1 + 5.2^{0.5} \rightarrow 5.4 + 5.5$	0.407 0.073	40.61 31.01	3.723	-48.57
3	$5.1 + 5.2 \rightarrow 5.3 + 5.5$ $5.1^{0} + 5.2 \rightarrow 5.4 + 5.5$	0.466 0.017	42.25 27.56	3.777	-48.51
4	$5.1 + 5.2 \rightarrow 5.3 + 5.5$ $5.1 + 5.2^{0} \rightarrow 5.4 + 5.5$	0.466 0.017	42.25 27.56	3.777	-47.29
5	$5.1 + 5.2^{0.5} \rightarrow 5.3 + 5.5$ $5.1 + 5.2 \rightarrow 5.4 + 5.5$	0.092 0.298	38.06 32.27	5.127	-46.64

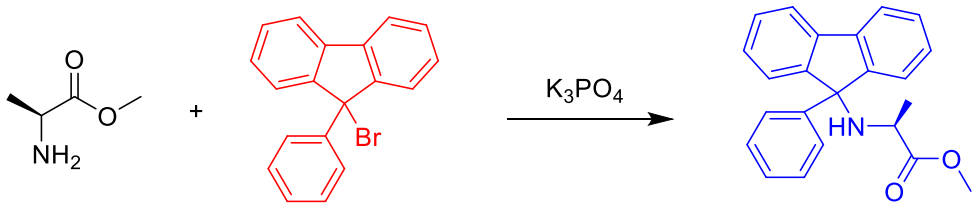
It is clear from the kinetic plots that the fit of the ODEs to the experimental data do not result in normally-distributed residuals. This could be for one of many reasons, but is most likely due to either errors in producing the correct concentration of amine, **5.2**, or NMR integrations errors, or evaporation of solvent (CD_3OD) into the headspace at higher temperatures. This must be considered by a trained chemist upon completion of the computational approach and this is an example showing that this methodology must be used in conjunction with chemical expertise to confirm the approach output. In this case, however, the residuals are only a small consideration and it can be observed that the model still fits the experimental data well, so the identified model is still valid.

This first experimental case study has shown that the new approach can correctly identify the reaction model and kinetic parameters, even when considering a multitude of differing rate laws. In this case, it was found that each the rate laws featured first-order dependencies on each of the species, which is common in chemical processes. This kinetic information can then be used to optimise this process for the highest yields of the 2-substituted (major) product, **5.3**, for our collaborator's research project which can then be used as a building block for bioactive materials.

5.3 Case study: PfBr

The next case study that was explored was a chemical system featuring a protection of an amino acid for further functionalisation, where alanine methyl

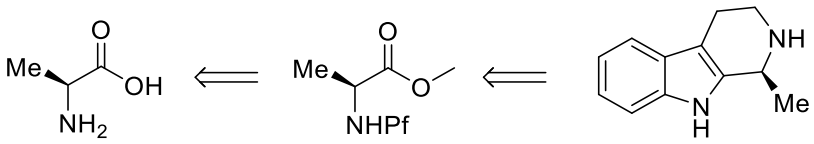
ester (Al-Me), **5.6**, reacts with 9-bromo-9-phenylfluorene (PfBr), **5.7**, to form the protected amino acid (Pf-Al-Me), **5.8**, and hydrobromic acid, **5.9**, as a side-product, as shown in Scheme 5.2. The alanine methyl ester is present in its hydrochloride salt form, so this hydrochloric acid and the formation of hydrobromic acid are both neutralised by an excess of suspended potassium phosphate.



5.6 **5.7** **5.8**

Scheme 5.2: The reaction of alanine methyl ester (Al-Me), **5.6**, with PfBr, **5.7**, to form the protected amino acid, **5.8**. Hydrobromic acid, **5.9**, is also generated as a side product.

PfBr is the reagent to introduce the 9-phenylfluorene (Pf) protecting group in a synthesis. Pf is a pharmaceutically relevant and bulky protecting group, that can be introduced as a more acid stable alternative to the more commonly used trityl protecting group.[245, 246] The chemical system of interest is the protection step in the total synthesis of (S)-eleagnine from L-alanine, as shown in Scheme 5.3 as a retrosynthetic methodology.[247] Understanding this transformation, by performing kinetic analysis, would therefore accelerate process development when scaling up this process.

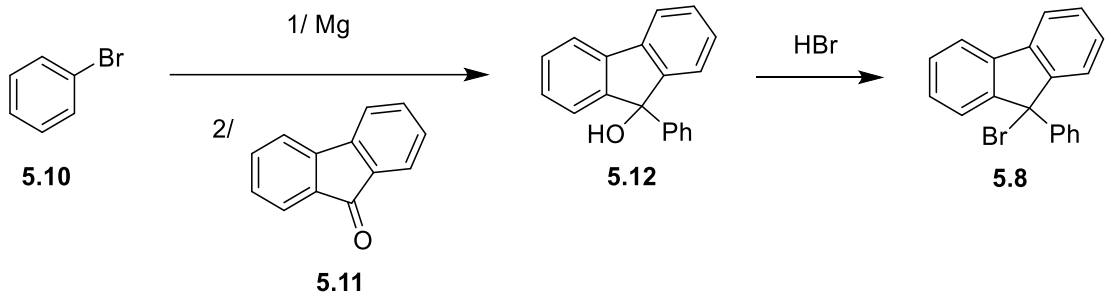


L-alanine **5.8** *(S)-eleagnine*

Scheme 5.3: The retrosynthetic methodology of the transformation of L-alanine to (S)-eleagnine via the Pf-protected alanine methyl ester (Pf-Al-Me).[247]

Due to the high cost of PfBr as a starting material (£137 for 5 g),[248] PfBr was synthesised from cheaper materials as it was unknown how much would be necessary for kinetic analysis. Bromobenzene, **5.10**, was initially treated with magnesium, then reacted with fluorenone, **5.11**, to make the intermediate PfOH, **5.12**. PfOH was then isolated and reacted with hydrobromic acid to form the desired PfBr product, as shown in Scheme 5.4.[249] The most expensive material

in the synthesis of this PfBr product is fluorenone, costing £76 for 500 g, resulting in a much cheaper method to obtaining this desired material.[250] This reaction was carried out on a 40 g scale, and full experimental details can be found in Chapter 7.5.2.1. As there was an absence of an OH peak from the IR spectrum of PfBr, the PfBr material was analysed by HPLC and the purity was determined to be >99 %. With pure PfBr as a starting material, the kinetic case study was then possible to begin.



5.10 **5.11** **5.12** **5.8**

Scheme 5.4: The synthetic route from cheap starting materials: bromobenzene, **5.10**, and fluorene, **5.11**, to the intermediate PfOH, **5.12**, and finally the desired PfBr material, **5.8**.

Three experiments were conducted in a three-neck round bottomed flask, at 30 °C, 35 °C and 40 °C. Samples (0.5 mL) were extracted manually from the bulk reaction medium with a syringe, followed by syringing approximately 0.3 mL of water and shaking the syringe - this dissolves the suspended potassium phosphate and allows injection into the sample loop for analysis via HPLC. The batch setup for experimentation is shown in Figure 5.3.

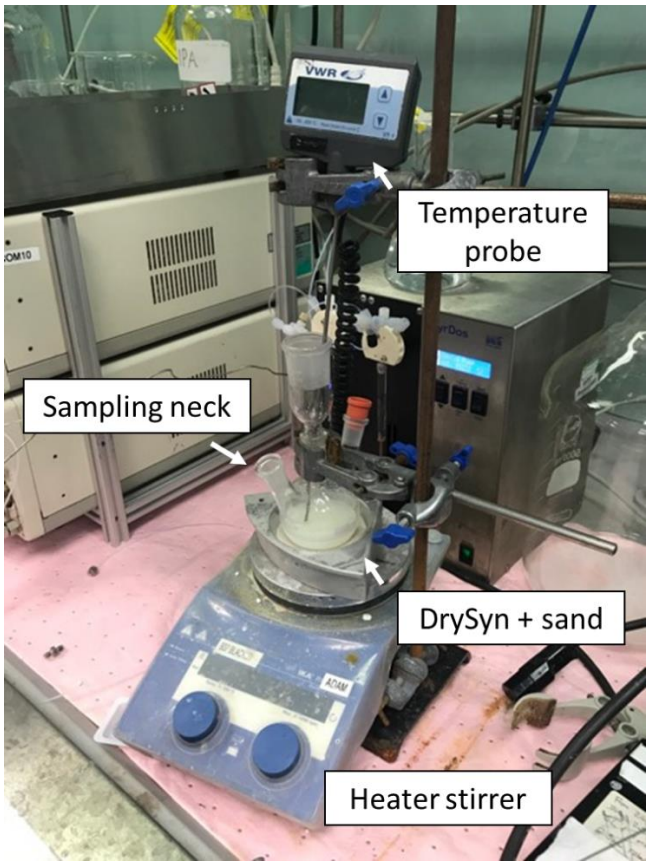


Figure 5.3: The batch setup for experimentation in the PfBr case study, where the temperature probe is submerged in the reaction medium that is heated via a heater stirrer.

After all experiments were run, the experimental data and the identified species were inputted into the new approach. Based on these four species, there were two feasible reactions calculated - these reactions are shown as eqn. 5.7 and eqn. 5.8. From these reactions, all corresponding rate laws were generated and all models were compiled, subject to the rules defined in Chapter 4.2.2. This resulted in 30 unique models to be evaluated by the approach.

$$\mathbf{5.6 + 5.7 \rightarrow 5.8 + 5.9} \qquad \text{eqn. 5.7}$$

$$\mathbf{5.8 + 5.9 \rightarrow 5.6 + 5.7} \qquad \text{eqn. 5.8}$$

All 30 models were evaluated by the new computational approach, incurring a computation time of less than 2 minutes (see Chapter 7.6.1 for details on computer specifications). Each of these models were then ranked based on their AIC_C evaluation, and it was found that the most likely representation of the system is the model shown in Scheme 5.2, but with a zero-order dependence on the alanine methyl ester in the rate law. This makes sense chemically as the bulky aromatic rings would stabilise the cation formed if the reaction were to proceed via a traditional S_N1 mechanism; this means that the rate-determining-

step is likely to be the loss of the bromide ion, followed by a fast reaction with the alanine methyl ester. The approach also determined the kinetic parameters of this transformation as $k_{35\,°C}$ = 1.06 x 10^{-2} ± 0.01 x 10^{-2} min^{-1}, E_a = 62.91 ± 0.23 kJ mol^{-1}. This model and the corresponding kinetic parameters allowed a fit to the experimental data with an average residual of 2.6 x 10^{-4} M and is shown in Figure 5.4 in a combined plot. The top 5 ranked models are shown in Table 5.2, with their respective SSE error metrics and AIC_C evaluations.

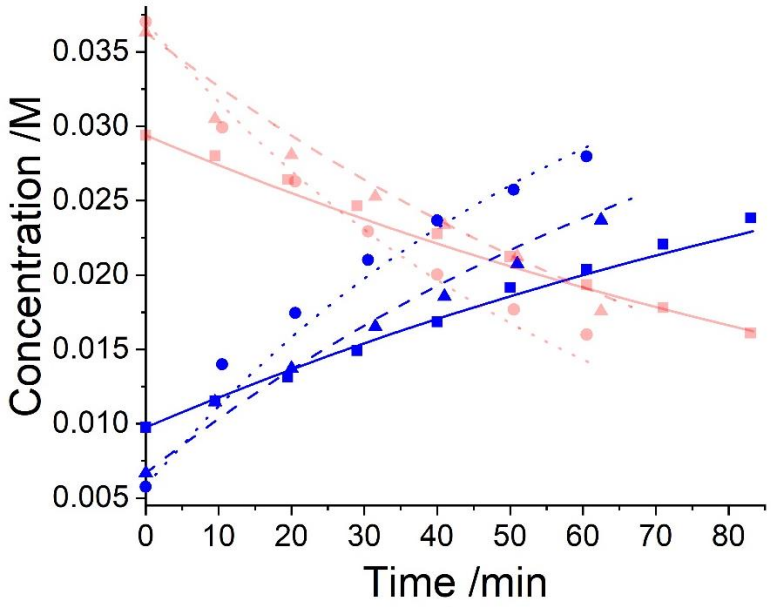


Figure 5.4: Kinetic profiles for three kinetic experiments at 30 °C, 35 °C and 40 °C, where red plots indicate PfBr concentrations and blue plots indicate Pf-Al-Me concentrations. At 30 °C: ■ = experimental data, — = ODE. At 35 °C: ▲ = experimental data, - - - = ODE. At 40 °C: ● = experimental data, ····· = ODE. See Chapter 7.5.2 for full experimental conditions and raw data.

Table 5.2: Evaluation of the feasibility of each reaction model for the PfBr case study, where α is variable for each model depending on the overall model order.

Rank	Reaction Model	Kinetic parameters		SSE /10^{-5} M	AIC_C
		$k_{35\,°C}$ /M^{α} min^{-1}	E_a /kJ mol^{-1}		
1	$5.6^0 + 5.7 \rightarrow 5.8 + 5.9$	0.0106	62.91	3.51	-238.15
2	$5.6^0 + 5.7 \rightarrow 5.8 + 5.9$ $5.8 + 5.9^{0.5} \rightarrow 5.6 + 5.7$	0.0115 0.0144	67.09 70.00	2.76	-238.05
3	$5.6^0 + 5.7 \rightarrow 5.8 + 5.9$ $5.8 + 5.9^0 \rightarrow 5.6 + 5.7$	0.0129 0.0034	65.27 70.00	2.76	-238.04
4	$5.6^0 + 5.7 \rightarrow 5.8 + 5.9$	0.0128	65.71	2.81	-237.64

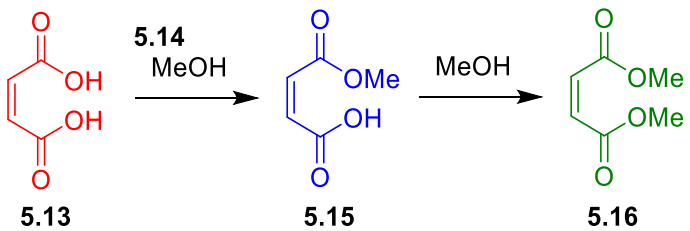
	$5.8^0 + 5.9 \rightarrow 5.6 + 5.7$	0.0057	70.00		
5	$5.6^0 + 5.7 \rightarrow 5.8 + 5.9$ $5.8^{0.5} + 5.9 \rightarrow 5.6 + 5.7$	0.0112 0.0167	66.06 70.00	2.95	-236.54

Interestingly, the top 5 ranked models all feature the main 'important' model term of $5.6^0 + 5.7 \rightarrow 5.8 + 5.9$, with models 2 - 5 also featuring some variant of the reversible reaction. This is a similar outcome to the simulated pentyne kinetics case study in Chapter 4.3.2. In this experimental case, however, the reversible reaction should be very unlikely as the hydrobromic acid is removed by the excess potassium phosphate present. It can be concluded therefore that these extra model terms are just fitting to the experimental noise, especially as the magnitude of the SSE metrics are so low.

It has been shown in this case study that the new computational approach can correctly identify the reaction model and kinetic parameters of a system where there is a zero-order dependency on one of the species. This is the first instance of an identified model featuring non-first-order rate laws in a real chemical process and further proves the broadened applicability of the new approach as a kinetic analysis tool. This kinetic information can then be used to shorten process development times and lower costs if (S)-eleagnine was synthesised using this route.

5.4 Case study: Maleic acid

The final experimental case study that was explored was the chemical system involving maleic acid, **5.13**, reacting with methanol, **5.14**, to form the monomethylated maleic acid ester (mono-product), **5.15**, and the dimethylated maleic acid ester (di-product), **5.16**, shown in Scheme 5.5. Each reaction also liberates a molecule of water, **5.17**.



Scheme 5.5: The reaction of maleic acid, **5.13**, and methanol, **5.14**, to form the mono-product, **5.15**, and the di-product, **5.16**.

Our collaborators at AstraZeneca were interested in this reaction following the contamination of a batch of an API maleate salt with the corresponding monomethyl maleate salt, as this contamination could be mitigated if the impurity formation was well understood. Kinetic experiments were run and the findings were published, as it was found that the reaction forming the mono-product was autocatalytic.[240] In this case, the reaction was found to be catalytic with an order of 0.5, meaning that the overall order with respect to the maleic acid was 1.5 in forming the mono-product. The consecutive reaction forming the di-product is also catalytic with respect to the maleic acid with a 0.5 order dependence. Although this information is now known and reported, it took several months for physical-organic chemists at AstraZeneca to decipher this reaction model as this model is not immediately intuitive. Therefore, this case study was run to retrospectively show the efficiency of this approach in elucidating non-intuitive reaction models, and how this methodology can serve as a viable substitute for many hours spent on a project by experts that could be working on other aspects of process development.

All experimental data was obtained from our collaborators at AstraZeneca and was collected using batch experimentation and ^{1}H NMR sampling. Five experiments were conducted at differing starting concentrations and temperatures in the range of 40 - 60 °C. This experimental data was inputted into the computational approach as well as the identified species. Based on these five species, there were six feasible reactions identified - these reactions are shown as eqn. 5.9 - 5.14:

5.13 + 5.14 → 5.15 + 5.17 eqn. 5.9

5.15 + 5.17 → 5.13 + 5.14 eqn. 5.10

5.15 + 5.14 → 5.16 + 5.17 eqn. 5.11

5.16 + 5.17 → 5.15 + 5.14 eqn. 5.12

5.15 + 5.15 → 5.13 + 5.16 eqn. 5.13

5.13 + 5.16 → 5.15 + 5.15 eqn. 5.14

As the experiments were ran with methanol as a solvent, this exhibits an effective methanol concentration of ~24 M, which is far greater than the maleic

acid concentration, that had a maximum concentration of 0.8 M. As the methanol is in such a high concentration, its concentration remains effectively unchanged throughout the reaction, meaning that the rate law would approximate pseudo-first-order kinetics. Therefore, after all rate laws were compiled using these mass-balance-allowed reactions, all rate laws pertaining to non-zero-order with respect to methanol concentration were removed. As the methanol concentration would be effectively unchanged, it would be very difficult to differentiate between models with differing order dependencies on the methanol - this would likely lead to many models tied for 'most likely' with exactly equivalent SSE values.

After the subsequent rate laws were compiled, the models were generated, subject to the rules defined in Chapter 4.2.2 and with the maleic acid defined as a catalyst. This resulted in 5086 unique models to be evaluated by the approach, which incurred a computation time around 17 hours (see Chapter 7.6.1 for details on computer specifications). Each of these models were then ranked based on their AIC_C evaluation, and it was found that the most likely representation of the system is the reaction model shown in Scheme 5.5, but each step is catalysed by maleic acid with a species order dependence of 0.5 - these are the same findings that our collaborators made during their kinetic analysis. The approach also determined the kinetic parameters of this transformation to be: $k_{mono\ 50\ °C} = 3.85 \times 10^{-3} \pm 0.01 \times 10^{-3}\ M^{-0.5}\ min^{-1}$, $E_a = 72.61 \pm 0.12$ kJ mol^{-1} and $k_{di\ 50\ °C} = 4.66 \times 10^{-4} \pm 0.01 \times 10^{-4}\ M^{-0.5}\ min^{-1}$, $E_a = 69.74 \pm 0.10$ kJ mol^{-1}. These parameters were again very similar to those obtained by our collaborators, although our collaborators reported the kinetic parameters as Arrhenius constants and activation energies. This model and the corresponding kinetic parameters allowed a fit to the experimental data with an average residual of 1.1×10^{-3} M and two experimental fittings at 50 °C (with initial concentrations of 0.8 M and 0.4 M) are shown in Figure 5.5. The top 5 ranked models are shown in Table 5.3, with their respective SSE error metrics and AIC_C evaluations.

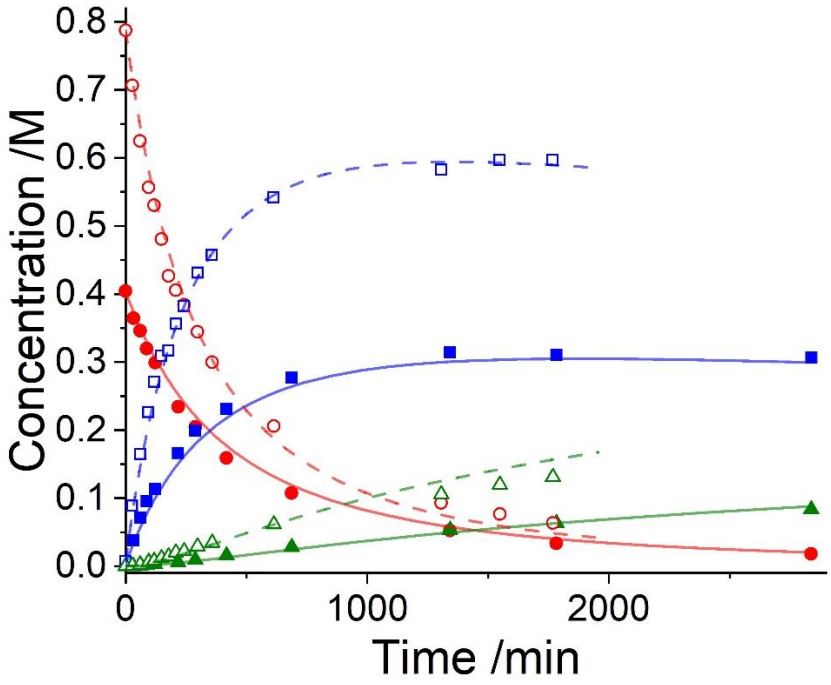


Figure 5.5: Kinetic profiles for two kinetic experiments at 50 °C, with the initial concentration of maleic acid at 0.4 M and 0.8 M. At 0.4 M: ● = maleic acid, — = maleic acid (ODE), ■ = mono-product, — = mono-product (ODE), ▲ = di-product, — = di-product (ODE). At 0.8 M: ○ = maleic acid, - - - = maleic acid (ODE), □ = mono-product, - - - = mono-product (ODE), △ = di-product, - - - = di-product (ODE). See Chapter 7.5.3 for full experimental conditions and raw data.

Table 5.3: Evaluation of the feasibility of each reaction model for the maleic acid case study, where α is variable for each model depending on the overall model order.

Rank	Reaction Model	Kinetic parameters		SSE /10^{-2} M	AIC_C
		$k_{50\,°C}$ /10^{-3} M^{α} min^{-1}	E_a /kJ mol^{-1}		
1	$5.13^{1.5} + 5.14^{0} \rightarrow 5.15 + 5.13^{0.5} + 5.17$ $5.13^{0.5} + 5.15 + 5.14^{0} \rightarrow 5.16 + 5.13^{0.5} + 5.17$	3.85 0.47	72.61 69.74	0.28	-49.68
2	$5.13^{1.5} + 5.14^{0} \rightarrow 5.15 + 5.13^{0.5} + 5.17$ $5.13 + 5.15^{2} \rightarrow 5.13^{2} + 5.16$	4.08 1.92	78.32 79.77	0.33	-49.08
3	$5.13^{1.5} + 5.14^{0} \rightarrow 5.15 + 5.13^{0.5} + 5.17$ $5.13 + 5.15 + 5.14^{0} \rightarrow 5.16 + 5.13 + 5.17$	3.91 1.03	74.19 79.95	0.33	-49.02

4	$5.13^{1.5} + 5.14^{0} \rightarrow 5.15 + 5.13^{0.5} + 5.17$ $5.13^{0.5} + 5.15^{2} \rightarrow 5.13^{1.5} + 5.16$	3.97 0.76	76.52 73.06	0.42	-48.06
5	$5.13^{2} + 5.14^{0} \rightarrow 5.15 + 5.13 + 5.17$ $5.13 + 5.15 + 5.14^{0} \rightarrow 5.16 + 5.13 + 5.17$	5.60 0.78	79.98 79.36	0.53	-47.15

As in previous case studies, the top ranked models all feature the main 'important' model term that describes the convergence to experimental data in the best way, with models 1 - 4 featuring this term: $5.13^{1.5} + 5.14^{0} \rightarrow 5.15 + 5.13^{0.5} + 5.17$. This is because the SSE metric is a measure of the total error in the fit to the data and as this model term features the transformation of 5.13 to 5.15, which are the species present in the highest concentrations, this model term therefore has the highest impact on minimising the SSE.

It has been shown in this case study that the new computational approach can correctly identify the reaction model and kinetic parameters of a real chemical system where there are both catalytic species and species with non-integer order dependencies. This is the first instance of an identified model featuring a catalytic species and non-integer-order rate laws using real experimental data. As the model was already identified by our collaborators, this methodology has corroborated their findings and shown that process understanding can be accelerated using this approach. This approach automatically identified the correct reaction model in ~17 hours, whereas our collaborators using more traditional 'trial and error' approaches arrived at the correct model in months. Retrospectively, this would have allowed faster implementation of measures to reduce the amount of monomethyl maleate salt impurity in their API maleate salt, thereby proceeding with process development and cutting costs.

5.5 Conclusion

It has been shown that using this new computational approach, total process understanding can be automatically achieved without the need for high-level chemical intuition. The widened scope of this new approach has been proven to be effective in three experimental case studies, each with differing rate laws and kinetic structural motifs. The approach has been proven to be effective and applicable for ‘normal’ kinetic models where each species has a first-order dependence, for zero-order and non-integer rate laws and for catalytic species of variable integer- and non-integer-order.

It has also been shown from these case studies that using this methodology can greatly accelerate process development and reduce the workload of kinetic analysis for the physical-organic chemists. After experimentation, all kinetic analysis can be automated whilst experts focus on other aspects of process development. Then, after the computational approach has finished running, they can work in conjunction with the approach to identify which models are the most likely to be both statistically and scientifically correct.

Chapter 6 : Conclusion & future work

The project aim, to develop and advance an automated methodology to kinetic model and parameter determination, has been answered in the thesis chapters herein. The computational approach should be used in conjunction with experts to automate the comprehensive kinetic analysis that is otherwise not feasible using traditional means. The approach takes experimental data and the identified species, then identifies every possible mass-balance-allowed transformation between these species. Each of these reactions are then compiled into every possible unique reaction model, with each feasible corresponding rate law. These models are then evaluated based on their fit to experimental data, by combining the use of ordinary differential equation (ODE) solvers and optimisation algorithms. The models are then statistically ranked based on their convergence to experimental data and the number of model terms in the model, thereby identifying the most likely model and its corresponding kinetic parameters.

The original structure of this comprehensive methodology was originally reported by Tsu *et al.*,[173] which was then replicated, improved and utilised for real experimentation in Chapters 2 - 3. The approach was used in tandem with an automated flow reactor platform, by exploiting flow rate manipulations to obtain full reaction profiles followed by implementation of the approach. The methodology was then further improved, and the scope was widened in Chapter 4, as catalytic systems and differing-/non-integer order species dependencies in the rate laws were also explored. This improved methodology was then proven to be effective in real experimental systems in Chapter 5, as batch experimentation ran at Leeds and by collaborators showed the wide applicability of the approach to different types of reactions. This approach can be used as a tool for process development by anyone, as expert chemical intuition is removed from the kinetic analysis. Therefore, the project aim has been answered in the work described herein, as any chemist can use this approach and interpret the output in order to obtain total process understanding.

Further work on the use and development of this approach for metal-catalysed processes would be beneficial. Metal-catalysed processes are very important in industry, but catalytic cycles can be complex meaning that it is not immediately obvious how to automate a tool to explore these potential models. In some cases it may be sufficient to describe a reaction based on empirical rate laws and this can currently be implemented using the approach developed. However, another research project on the development of a tool to identify these kinetic models and catalytic cycles may be warranted. As a suggestion, this could be performed by inputting many known catalytic cycles into a training set, then using machine learning techniques to build a neural network to identify trends in experimental data to attribute particular catalytic cycles to chemical processes.

Chapter 7 : Appendix

7.1 Chapter 1

No appendix data for Chapter 1.

7.2 Chapter 2

7.2.1 Generated data set for case study: benzoic acid alkylation

Experiment 1

Time /s	2.1 /M	2.2 /M	2.3 /M
0	0.1	0.8	0
4.75	0.081361	0.060964	0.017126
10.24	0.069511	0.04901	0.031805
20.15	0.054394	0.03336	0.044477
29.5	0.044695	0.027749	0.051252
43.2	0.038039	0.019903	0.062653
60	0.032285	0.013536	0.065719

Experiment 2

Time /s	2.1 /M	2.2 /M	2.3 /M
0	0.1	0.11	0
4.68	0.07631	0.084741	0.023482
10.27	0.057707	0.069104	0.040456
20	0.04138	0.054662	0.056759
30.2	0.034101	0.04059	0.067659
47.6	0.023686	0.031161	0.075772
60	0.018377	0.027305	0.083761

Experiment 3

Time /s	2.1 /M	2.2 /M	2.3 /M
0	1	0.15	0
5	0.069767	0.121755	0.032513
10.5	0.047199	0.098611	0.05277
21.63	0.027758	0.077627	0.072909
30.54	0.019626	0.069013	0.079639
45.78	0.011357	0.059865	0.091724

60	0.006574	0.05691	0.093444

7.2.2 Generated data set for case study: nitrile hydrolysis

Experiment 1, 60 °C

Time /s	**2.5 /M**	**2.7 /M**	**2.8 /M**
0	0.8	0	0
5.235363	0.814005	0.01713	1.00E-06
32.47861	0.67236	0.106334	3.39E-05
67.19836	0.634071	0.199701	0.000124
145.6155	0.463323	0.34173	0.000516
289.9323	0.27438	0.528037	0.001536
368.7175	0.220788	0.551639	0.00212
486.8495	0.162216	0.630518	0.003131
600	0.124203	0.705371	0.004112

Experiment 2, 70 °C

Time /s	**2.5 /M**	**2.7 /M**	**2.8 /M**
0	0.8	0	0
6.133417	0.761239	0.049487	6.83E-06
26.89	0.588801	0.199668	1.15E-04
68.70925	0.426543	0.363734	0.000585
150.1538	0.21105	0.573179	0.00192
297.041	0.079934	0.705184	0.004502
356.4911	0.05351	0.757795	0.005854
486.7485	0.024103	0.789235	0.008408
600	0.013094	0.752509	0.010127

Experiment 3, 80 °C

Time /s	**2.5 /M**	**2.7 /M**	**2.8 /M**

0	1	0	0
5.817383	0.852929	0.15263	4.28E-05
24.71834	0.568389	0.476974	5.89E-04
71.4048	0.218624	0.778816	0.002633
154.3583	0.060192	0.89969	0.006685
303.8829	0.006754	0.944201	0.014173
372.0244	0.002737	0.9556	0.01819
476.5104	0.000608	1.009417	0.023157
600	0.000113	0.9467	0.028902

Experiment 4, 90 °C

Time /s	**2.5 /M**	**2.7 /M**	**2.8 /M**
0	1	0	0
5.910759	0.711158	0.281532	1.72E-04
24.20962	0.30002	0.70196	1.71E-03
68.78985	0.060896	0.961462	0.006075
142.5421	0.005202	1.023471	0.013823
288.1599	5.55E-05	0.935542	0.027497
364.4764	5.21E-06	0.970668	0.036328
478.0807	1.43E-07	0.95015	0.045819
600	1.79E-07	0.89561	0.059331

7.2.3 Generated data set for the case study: S_NAr kinetics

Experiment 1

Time	**2.10 /M**	**2.12 /M**	**2.13 /M**	**2.14 /M**
0	1	0	0	0
0.114724	0.916317	0.078593	0.018182	2.49E-06
1.03049	0.464119	0.448791	0.104213	0.000125
2.141533	0.226584	0.624128	0.146157	0.000376
3.643057	0.102715	0.737636	0.171082	0.000744

7.743801	0.012971	0.788209	0.187458	0.001768
12.16477	0.001476	0.805589	0.18666	0.002965
44.03067	0	0.816998	0.186517	0.010796
107.2065	0	0.801824	0.178389	0.026607
255.1342	0	0.78187	0.172065	0.061291
435.1342	0	0.744149	0.156488	0.102055
600	0	0.702123	0.148841	0.135997

Experiment 2

Time	**2.10 /M**	**2.12 /M**	**2.13 /M**	**2.14 /M**
0	1.2	0	0	0
0.088864	1.073662	0.091451	0.0213	2.94E-06
0.254455	0.913278	0.236597	0.053222	2.17E-05
1.152788	0.407456	0.638051	0.147144	0.000266
3.171611	0.093893	0.900254	0.2064	0.001032
9.39048	0.001601	0.959144	0.228044	0.003623
40.41694	0	0.97195	0.21939	0.016209
195.366	0	0.911133	0.206891	0.075681
311.2192	0	0.891022	0.190707	0.118451
431.2192	0	0.869047	0.184739	0.15486
491.2192	0	0.861757	0.175374	0.176878
600	0	0.839806	0.165531	0.209888

Experiment 3

Time	**2.10 /M**	**2.12 /M**	**2.13 /M**	**2.14 /M**
0	0.9	0	0	0
0.091037	0.797548	0.079495	0.018563	3.13E-06
0.259823	0.653654	0.198581	0.046012	2.19E-05
1.158987	0.239387	0.538734	0.122398	0.000279

4.175397	0.013188	0.712848	0.169522	0.001555
6.793661	0.001174	0.728233	0.166388	0.002729
9.757783	0	0.717645	0.170752	0.003931
44.49079	0	0.719684	0.164371	0.019044
195.6282	0	0.666457	0.14927	0.080022
315.6282	0	0.644589	0.131672	0.126933
495.6282	0	0.600784	0.117343	0.185606
600	0	0.578163	0.107537	0.221849

7.3 Chapter 3

7.3.1 Phenyl acetate

7.3.1.1 Preparation of feed solutions

Three feeds were used for this work, and desired reservoir solutions were prepared by dissolving the desired reagents in acetonitrile under stirring at ambient conditions. The first feed consisted of phenol (13.2 mL, 0.15 mol, 0.5 mol dm^{-3}) and biphenyl (1.15 g, 7.5 mmol, 0.025 mol dm^{-3}) in acetonitrile (300 mL); the second feed contained acetyl chloride (22.5 mL, 0.32 mol, 1.05 mol dm^{-3}); the third feed was a 1:1 mixture of acetonitrile and water for dilution.

7.3.1.2 Flow ramp experiments

Linear gradient flow ramps allowed the generation of a complete reaction profile from a single transient experiment. To obtain transient data, each of the three pumps were initially set at the maximum flow rate to be investigated: 1.75 mL min^{-1}. Steady-state was established in the 3.5 mL reactor, and the flow rate for each pump decreased at a constant rate of 0.0181 mL min^{-1} for 92 minutes. Samples of reactor effluent were injected for HPLC analysis at 2 minute intervals, thus achieving a large data density. The reactor setup schematic is shown in Figure 7.3.1.

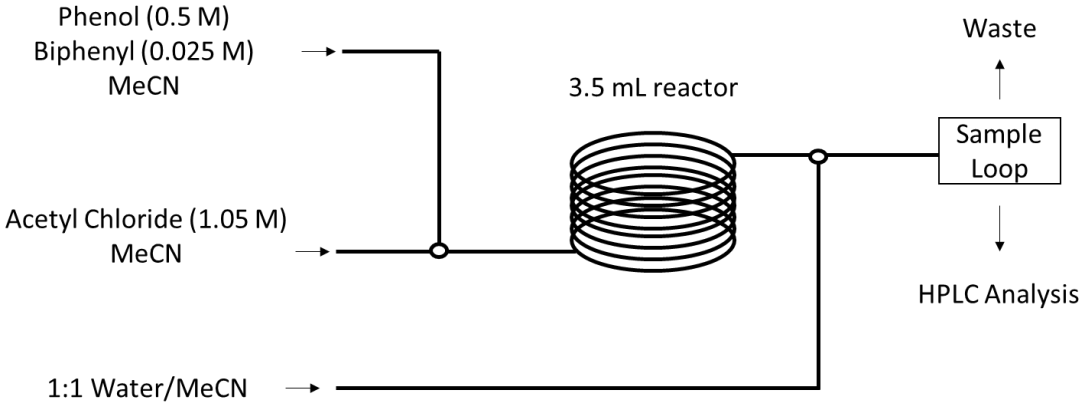


Figure 7.3.1: The flow reactor setup for the phenyl acetate case study experiments.

7.3.1.3 HPLC Analysis

All HPLC analysis was conducted using an Agilent 1100 series HPLC instrument fitted with a Sigma Ascentis Express C18 reverse phase column (5cm x 4.6mm, 2.7 µm). Biphenyl was used as an internal standard. The column temperature was 70 °C and the HPLC method is shown:

Time /min	%A (water, 0.1 % TFA)	%B (acetonitrile, 0.1 % TFA)	Flow rate /mL min^{-1}
0.00	50	50	2
1.50	20	80	2
1.51	50	50	2

An example HPLC chromatogram during a kinetic experiment is shown in Figure 7.3.2:

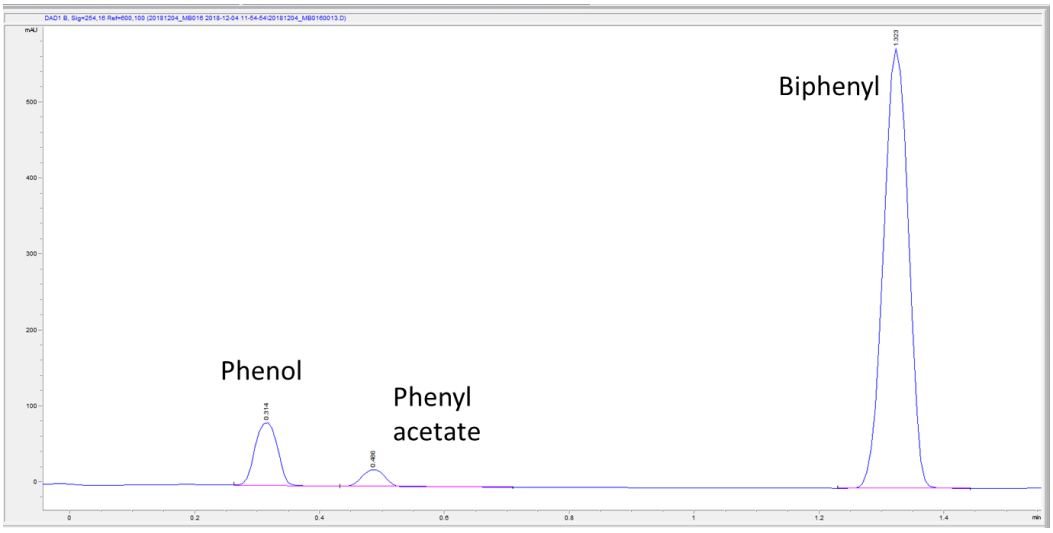


Figure 7.3.2: An example HPLC chromatogram in the phenyl acetate case study at 254 nm.

7.3.1.4 Validation with steady-state measurements

Steady-state measurements were taken at: 1, 3, 5, 7, 9, 11, 13 and 15 minutes, at both 65 °C and 75 °C. Using the identified most likely reaction model from the flow ramps, the k values for the steady-state experiments were found to be: $k_{SS\text{-}65\ °C} = 4.72 \times 10^{-3} \pm 0.01 \times 10^{-3}\ M^{-1}\ s^{-1}$; $k_{SS\text{-}75\ °C} = 9.29 \times 10^{-3} \pm 0.03 \times 10^{-3}\ M^{-1}\ s^{-1}$. These k values give the steady-state kinetic parameters as: $k_{SS\text{-}65\ °C} = 4.72 \times 10^{-3} \pm 0.01 \times 10^{-3}\ M^{-1}\ s^{-1}$, $E_{a\text{-}SS} = 66.3 \pm 0.5\ kJ\ mol^{-1}$ which are very comparable to the flow ramp evaluation of the kinetic parameters: $k_{R\text{-}65\ °C} = 0.0052 \pm 0.0002\ M^{-1}\ s^{-1}$, $E_{a\text{-}R} = 69.3 \pm 7.8\ kJ\ mol^{-1}$. As the steady-state measurements were taken without a dilution pump, the concentrations are different to that of the ramps, hence the steady-state time-series concentration data was normalised to match the concentrations from the ramps to show the agreement in their nature in Figure 3.11.

7.3.1.5 Raw data

Flow ramp experiments

Flow ramp measurements	**65 °C**		**75 °C**	
τ /min	**3.8 /M**	**3.10 /M**	**3.8 /M**	**3.10 /M**
1.02	0.1493	0.0072	0.1346	0.0262
1.04	0.1461	0.0095	0.1297	0.0269
1.06	0.1436	0.0104	0.1284	0.0260
1.08	0.1459	0.0119	0.1300	0.0262
1.11	0.1433	0.0121	0.1268	0.0287
1.14	0.1394	0.0128	0.1244	0.0304
1.16	0.1398	0.0135	0.1251	0.0310
1.19	0.1428	0.0144	0.1200	0.0322
1.22	0.1410	0.0150	0.1176	0.0340
1.25	0.1397	0.0142	0.1249	0.0304
1.29	0.1358	0.0140	0.1204	0.0334
1.32	0.1388	0.0168	0.1188	0.0369

1.36	0.1382	0.0166	0.1190	0.0405
1.40	0.1365	0.0170	0.1204	0.0315
1.44	0.1358	0.0182	0.1174	0.0408
1.49	0.1324	0.0173	0.1142	0.0392
1.53	0.1338	0.0188	0.1170	0.0383
1.58	0.1360	0.0194	0.1148	0.0395
1.64	0.1334	0.0204	0.1134	0.0426
1.70	0.1320	0.0213	0.1132	0.0406
1.76	0.1333	0.0223	0.1113	0.0430
1.82	0.1315	0.0232	0.1121	0.0456
1.89	0.1319	0.0220	0.1095	0.0445
1.97	0.1278	0.0228	0.1029	0.0523
2.05	0.1285	0.0260	0.1072	0.0501
2.14	0.1280	0.0334	0.1021	0.0554
2.24	0.1267	0.0287	0.1065	0.0524
2.35	0.1248	0.0305	0.0996	0.0603
2.47	0.1253	0.0309	0.0984	0.0566
2.60	0.1214	0.0338	0.0974	0.0575
2.74	0.1217	0.0297	0.0914	0.0661
2.90	0.1195	0.0360	0.0914	0.0620
3.08	0.1196	0.0379	0.0900	0.0654
3.28	0.1162	0.0408	0.0830	0.0762
3.51	0.1111	0.0444	0.0790	0.0774
3.77	0.1114	0.0452	0.0848	0.0739
4.06	0.1017	0.0527	0.0737	0.0722
4.41	0.1033	0.0536	0.0705	0.0843
4.81	0.0986	0.0559	0.0618	0.0981
5.28	0.0955	0.0598	0.0564	0.0962
5.83	0.0908	0.0678	0.0599	0.0985
6.49	0.0807	0.0704	0.0576	0.1039

7.29	0.0781	0.0770	0.0488	0.1150
8.24	0.0705	0.0835	0.0421	0.1160
9.40	0.0662	0.0921	0.0360	0.1223
10.79	0.0587	0.0981	0.0249	0.1358

Steady state experiments

Steady-state measurements	**65 °C**		**75 °C**	
τ /min	**3.8 /M**	**3.10 /M**	**3.8 /M**	**3.10 /M**
1.02	0.1771	0.0177	0.1591	0.0387
3.08	0.1461	0.0470	0.1183	0.0824
4.81	0.1214	0.0736	0.0858	0.1123
6.49	0.1050	0.0940	0.0625	0.1370
9.00	0.0844	0.1155	0.0419	0.1567
11.01	0.0717	0.1263	0.0297	0.1715
13.01	0.0559	0.1384	0.0218	0.1822
15.02	0.0512	0.1488	0.0177	0.1886

Normalised steady state experiments

Steady-state measurements (normalised)	**65 °C**		**75 °C**	
τ /min	**3.8 /M**	**3.10 /M**	**3.8 /M**	**3.10 /M**
1.02	0.1417	0.0142	0.1273	0.0310
3.08	0.1169	0.0377	0.0947	0.0660
4.81	0.0971	0.0590	0.0686	0.0898
6.49	0.0840	0.0753	0.0500	0.1096
9.00	0.0675	0.0926	0.0335	0.1253
11.01	0.0573	0.1012	0.0238	0.1372

13.01	0.0447	0.1109	0.0174	0.1458
15.02	0.0409	0.1192	0.0142	0.1509

7.3.2 Paracetamol

7.3.2.1 Preparation of feed solutions

For the work at the temperature of 30 °C, two feeds were used, and desired reservoir solutions were prepared by dissolving the desired reagents in acetonitrile under stirring at ambient conditions. The first feed consisted of 4-aminophenol (0.58 mL, 6 mmol, 0.02 mol dm^{-3}) and biphenyl (1.15 g, 7.5 mmol, 0.025 mol dm^{-3}) in acetonitrile (300 mL); the second feed contained acetic anhydride (0.85 mL, 9 mmol, 0.03 mol dm^{-3}) in acetonitrile (300 mL).

For the work at the temperature of 60 °C, two feeds were used, and desired reservoir solutions were prepared by dissolving the desired reagents in acetonitrile under stirring at ambient conditions. The first feed consisted of 4-aminophenol (0.44 mL, 4.5 mmol, 0.015 mol dm^{-3}) and biphenyl (1.15 g, 7.5 mmol, 0.025 mol dm^{-3}) in acetonitrile (300 mL); the second feed contained acetic anhydride (0.85 mL, 9 mmol, 0.03 mol dm^{-3}) in acetonitrile (300 mL).

For the work at the temperature of 160 °C, two feeds were used, and desired reservoir solutions were prepared by dissolving the desired reagents in acetonitrile under stirring at ambient conditions. The first feed consisted of 4-aminophenol (1.01 mL, 10.5 mmol, 0.035 mol dm^{-3}) and biphenyl (1.15 g, 7.5 mmol, 0.025 mol dm^{-3}) in acetonitrile (300 mL); the second feed contained acetic anhydride (2.98 mL, 31.5 mmol, 0.105 mol dm^{-3}) in acetonitrile (300 mL).

For the work at the temperature of 180 °C, two feeds were used, and desired reservoir solutions were prepared by dissolving the desired reagents in acetonitrile under stirring at ambient conditions. The first feed consisted of 4-aminophenol (2.61 mL, 27 mmol, 0.09 mol dm^{-3}) and biphenyl (1.15 g, 7.5 mmol, 0.025 mol dm^{-3}) in acetonitrile (300 mL); the second feed contained acetic anhydride (5.33 mL, 56.4 mmol, 0.19 mol dm^{-3}) in acetonitrile (300 mL).

7.3.2.2 Flow ramp experiments

For the work at the temperature of 30 °C, a 0.25 mL reactor was used. For the work at the temperature of 60 °C, a 0.5 mL reactor was used. For the

work at the temperatures of 160 °C and 180 °C, a 3.5 mL reactor was used. Linear gradient flow ramps allowed the generation of complete reaction profiles from a single transient experiment. To obtain transient data, each of the two pumps were initially set at the maximum flow rate to be investigated: 1.75 mL min^{-1}. Steady-state was established and the flow rate for each pump decreased at a constant rate of 0.0181 mL min^{-1} for 92 minutes. Samples of reactor effluent were injected for HPLC analysis at 2 minute intervals, thus achieving a large data density. The reactor setups for each of these 4 reactions are shown in Figures 7.3.3 - 7.3.6. For the higher temperature reactions, 160 °C and 180 °C, 4-aminophenol is used and assumed to react instantaneously to form paracetamol, based on the kinetics observed at lower temperatures.

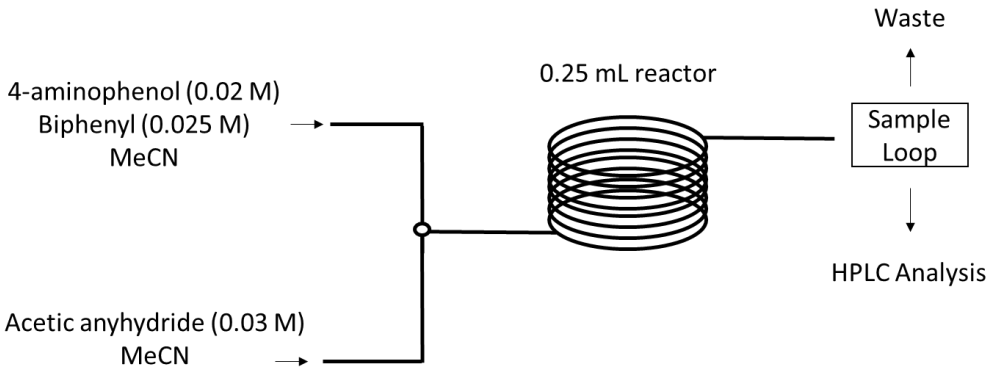


Figure 7.3.3: The reactor setup for the paracetamol flow ramp experiment at 30 °C.

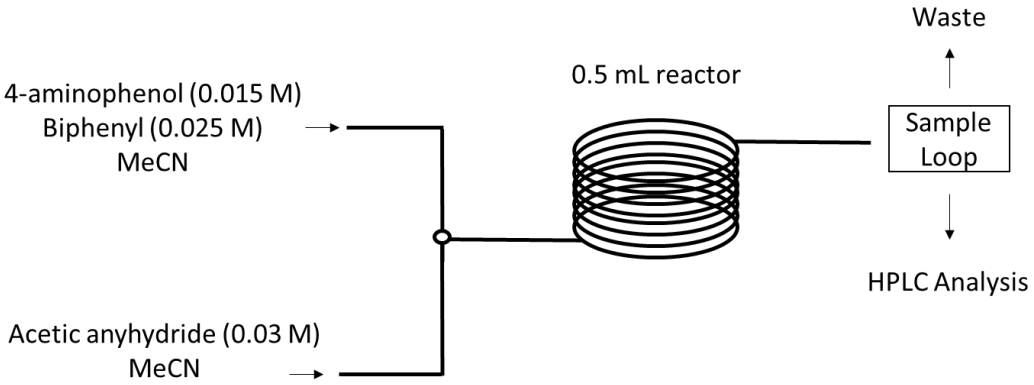


Figure 7.3.4: The reactor setup for the paracetamol flow ramp experiment at 60 °C.

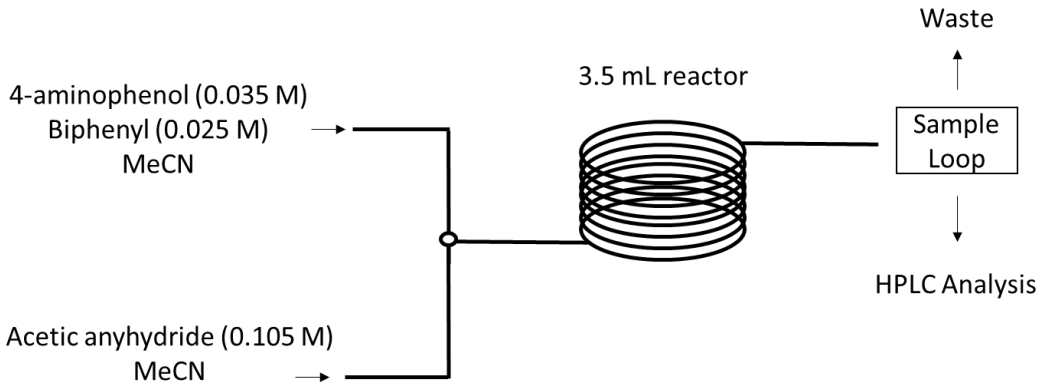


Figure 7.3.5: The reactor setup for the paracetamol flow ramp experiment at 160 °C.

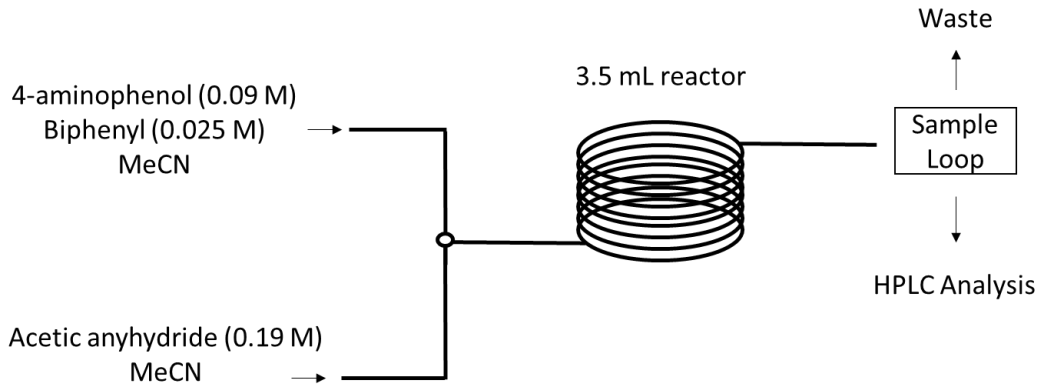


Figure 7.3.6: The reactor setup for the paracetamol flow ramp experiment at 180 °C.

7.3.2.3 HPLC analysis

All HPLC analysis was conducted using an Agilent 1100 series HPLC instrument fitted with a Sigma Ascentis Express C18 reverse phase column (5cm x 4.6mm, 2.7 µm). Biphenyl was used as an internal standard. The column temperature was 70 °C and the HPLC method is shown:

Time /min	%A (water, 0.1 % TFA)	%B (acetonitrile, 0.1 % TFA)	Flow rate /mL min^{-1}
0.00	50	50	2
1.50	20	80	2
1.51	50	50	2

An example HPLC chromatogram showing the separation and analysis of all components at once - these species were each added to the same solution for method development.

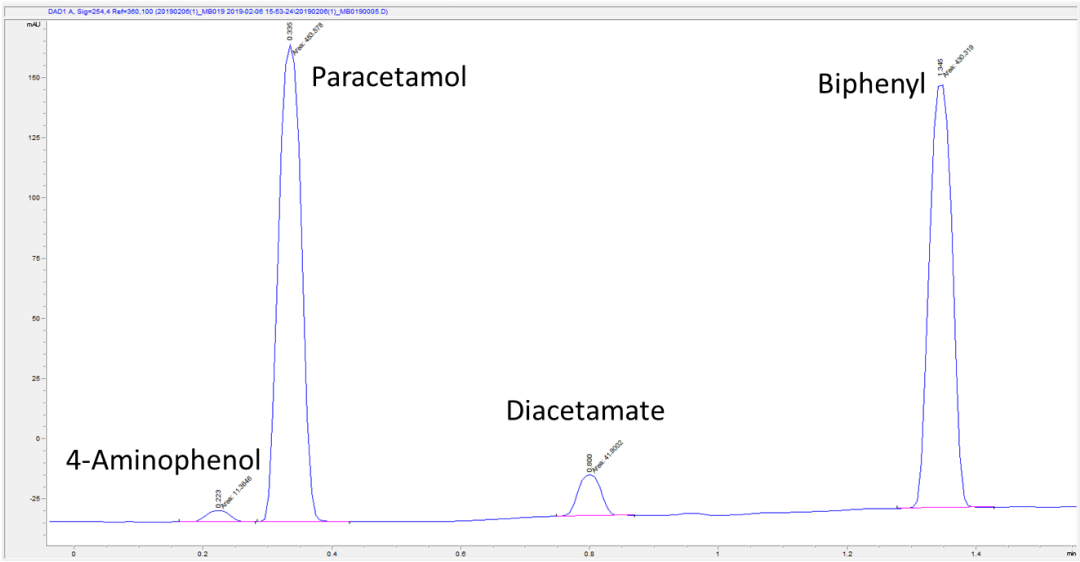


Figure 7.3.7: An example HPLC chromatogram from the paracetamol case study, showing the separation of all reaction components at 254 nm.

7.3.2.4 Raw data

Flow ramp experiment 1, 30 °C

τ /s	3.12 /M	3.14 /M	3.15 /M
4.28	0.00698	0.00267	0
4.37	0.00722	0.00266	0
4.47	0.00766	0.00284	0
4.57	0.00716	0.00287	0
4.67	0.00688	0.00278	0
4.78	0.00749	0.00306	0
4.89	0.00717	0.00299	0
5.01	0.00715	0.00317	0
5.13	0.00685	0.00273	0
5.26	0.00727	0.00320	0
5.40	0.00622	0.00296	0
5.55	0.00635	0.00337	0
5.70	0.00643	0.00374	0
5.86	0.00639	0.00446	0
6.03	0.00623	0.00401	0

6.21	0.00643	0.00391	0
6.40	0.00600	0.00385	0
6.45	0.00654	0.00378	0
6.83	0.00613	0.00368	0
7.06	0.00519	0.00376	0
7.31	0.00592	0.00455	0
7.58	0.00585	0.00481	0
7.87	0.00554	0.00410	0
8.18	0.00564	0.00432	0
8.51	0.00503	0.00429	0
8.88	0.00568	0.00493	0
9.27	0.00441	0.00444	0
9.71	0.00444	0.00506	0
10.19	0.00416	0.00513	0
10.71	0.00446	0.00556	0
11.29	0.00413	0.00575	0
11.94	0.00440	0.00546	0
12.67	0.00483	0.00566	0
13.49	0.00452	0.00564	0
14.43	0.00353	0.00601	0
15.50	0.00276	0.00588	0
16.75	0.00370	0.00654	0
18.20	0.00325	0.00720	0
19.94	0.00229	0.00676	0
22.02	0.00214	0.00686	0
24.59	0.00239	0.00770	0
27.81	0.00178	0.00878	0
31.97	0.00221	0.00826	0
39.00	0.00171	0.00848	0
45.18	0.00145	0.00871	0

Flow ramp experiment 2, 60 °C

τ /s	**3.12 /M**	**3.14 /M**	**3.15 /M**
8.75	0.003048	0.004234	0
8.94	0.003702	0.003846	0
9.13	0.003852	0.003772	0
9.55	0.003402	0.003918	0
9.78	0.003071	0.003860	0
10.02	0.002944	0.004318	0
10.27	0.003195	0.004243	0
10.53	0.003084	0.004111	0
10.80	0.003046	0.004262	0
11.09	0.003112	0.004330	0
11.40	0.002925	0.004253	0
11.72	0.002919	0.004464	0
12.06	0.002921	0.004562	0
12.43	0.002741	0.004796	0
12.81	0.002612	0.004906	0
12.90	0.002480	0.004399	0
13.66	0.002481	0.004597	0
14.12	0.002545	0.004952	0
14.62	0.002325	0.004951	0
15.16	0.002416	0.004964	0
15.73	0.002407	0.005149	0
16.35	0.002401	0.005027	0
17.02	0.002710	0.004966	0
17.75	0.002157	0.005196	0
18.55	0.002382	0.005121	0
19.42	0.002037	0.005228	0

20.37	0.002221	0.005387	0
21.42	0.001760	0.005411	0
22.59	0.001446	0.005728	0
23.89	0.001541	0.005809	0
25.34	0.001507	0.005924	0
26.99	0.001471	0.005553	0
28.86	0.001147	0.005686	0
31.00	0.001304	0.006117	0
33.49	0.001168	0.005719	0
36.41	0.001018	0.006181	0
39.87	0.000916	0.006444	0
44.05	0.000976	0.006104	0
49.18	0.001025	0.006840	0

Flow ramp experiment 3, 160 °C

τ /min	**3.12 /M**	**3.14 /M**	**3.15 /M**
1.25	0	0.01766	0.00056
1.29	0	0.01737	0.00054
1.32	0	0.01786	0.00038
1.36	0	0.01789	0.00042
1.40	0	0.01722	0.00061
1.44	0	0.01738	0.00055
1.49	0	0.01779	0.00059
1.53	0	0.01807	0.00068
1.58	0	0.01794	0.00064
1.64	0	0.01722	0.00067
1.70	0	0.01734	0.00072
1.76	0	0.01792	0.00077

1.82	0	0.01738	0.00076
1.89	0	0.01755	0.00081
1.97	0	0.01738	0.00074
2.05	0	0.01785	0.00102
2.14	0	0.01762	0.00086
2.24	0	0.01801	0.00105
2.35	0	0.01753	0.00111
2.47	0	0.01793	0.00102
2.60	0	0.01763	0.00123
2.74	0	0.01672	0.00118
3.08	0	0.01736	0.00124
3.28	0	0.01728	0.00160
3.51	0	0.01635	0.00153
3.77	0	0.01696	0.00174
4.06	0	0.01670	0.00199
4.41	0	0.01536	0.00187
4.81	0	0.01590	0.00222
5.28	0	0.01614	0.00263
5.83	0	0.01482	0.00258
6.49	0	0.01454	0.00317
7.29	0	0.01474	0.00331
8.24	0	0.01407	0.00355
9.40	0	0.01315	0.00386
10.79	0	0.01292	0.00412

Flow ramp experiment 4, 180 °C

τ /min	**3.12 /M**	**3.14 /M**	**3.15 /M**
1.04	0	0.03703	0.00456

1.06	0	0.03829	0.00519
1.08	0	0.03746	0.00652
1.11	0	0.03910	0.00557
1.14	0	0.03812	0.00656
1.16	0	0.03894	0.00592
1.19	0	0.03622	0.00877
1.22	0	0.03941	0.00625
1.25	0	0.03733	0.00754
1.29	0	0.03798	0.00713
1.32	0	0.03861	0.00717
1.36	0	0.03606	0.00930
1.40	0	0.03615	0.00890
1.44	0	0.03655	0.00870
1.49	0	0.03584	0.00854
1.53	0	0.03765	0.00893
1.58	0	0.03480	0.01017
1.64	0	0.03459	0.01078
1.70	0	0.03522	0.01035
1.76	0	0.03357	0.01114
1.82	0	0.03407	0.01127
1.89	0	0.03598	0.00927
1.97	0	0.03491	0.00963
2.05	0	0.03546	0.01197
2.14	0	0.03307	0.01152
2.24	0	0.03586	0.01004
2.35	0	0.03600	0.01103
2.47	0	0.03364	0.01246
2.60	0	0.03717	0.00943

2.74	0	0.03256	0.01202
2.90	0	0.03278	0.01151
3.08	0	0.03298	0.01160
3.28	0	0.03304	0.01311
3.51	0	0.03179	0.01295
3.77	0	0.03002	0.01227
4.06	0	0.03024	0.01543
4.41	0	0.03117	0.01338
4.81	0	0.03076	0.01426
5.28	0	0.02431	0.01748
5.83	0	0.02326	0.01784
6.49	0	0.02665	0.01973
7.29	0	0.02265	0.01803
8.24	0	0.01960	0.02426
9.40	0	0.01780	0.02347
10.79	0	0.01594	0.02903

7.3.3 Metoprolol

7.3.3.1 AstraZeneca experimental setup

All AstraZeneca experiments were conducted using a tubular reaction vessel (Polar Bear Plus) using PTFE tubing (1 mm ID). Reagents were pumped using JASCO PU-4180 dual piston HPLC pumps and flow streams were mixed using Swagelok SS-100-3 tee-pieces. Sampling was conducted by using a VICI Valco EUDA-CI4W.5 sample loop with a 0.06 µL aliquot volume. The reaction system was maintained under a fixed back pressure using an Upchurch Scientific 150 psi back pressure regulator. Quantitative analysis was performed using an Agilent 1100 series HPLC instrument fitted with an Acquity C18 reverse phase column (3 cm x 4.6 mm, 1.7 µm). This setup is shown in Figure 6.3.8. In all experiments biphenyl was added to one reservoir as an internal standard.

Figure 7.3.8: A photograph of the automated flow system at AstraZeneca.

7.3.3.2 Preparation of feed solutions

For the Leeds experiments, two feeds were used for the flow ramps, and desired reservoir solutions were prepared by dissolving the desired reagents in acetonitrile under stirring at ambient conditions. The first feed consisted of 2-((4-(2-methoxyethyl)phenoxy)methyl)oxirane (SM) (56.76 mL, 0.3 mol, 1 mol dm^{-3}) and biphenyl (1.15 g, 7.5 mmol, 0.025 mol dm^{-3}) in acetonitrile (300 mL); the second feed contained isopropylamine (15.46 mL, 0.18 mol, 0.6 mol dm^{-3}) in acetonitrile (300 mL).

For the AstraZeneca experiments, two feeds were used for the flow ramps, and desired reservoir solutions were prepared by dissolving the desired reagents in acetonitrile under stirring at ambient conditions. The first feed consisted of 2-((4-(2-methoxyethyl)phenoxy)methyl)oxirane (SM) (56.76 mL, 0.3 mol, 1 mol dm^{-3}) and biphenyl (1.15 g, 7.5 mmol, 0.025 mol dm^{-3}) in acetonitrile (300 mL); the second feed contained isopropylamine (51.53 mL, 0.9 mol, 3 mol dm^{-3}) in acetonitrile (300 mL).

7.3.3.3 Flow ramp experiments

For the Leeds experiments, a 3.5 mL reactor was used. Linear gradient flow ramps allowed the generation of complete reaction profiles from a single transient experiment. To obtain transient data, each of the two pumps were initially set at the maximum flow rate to be investigated: 1.75 mL min^{-1}. Steady-state was established and the flow rate for each pump decreased at a constant rate of 0.0181 mL min^{-1} for 92 minutes. Samples of reactor effluent were injected for HPLC analysis at 4 minute intervals, thus achieving a large data density. This reactor setup is shown in Figure 7.3.9.

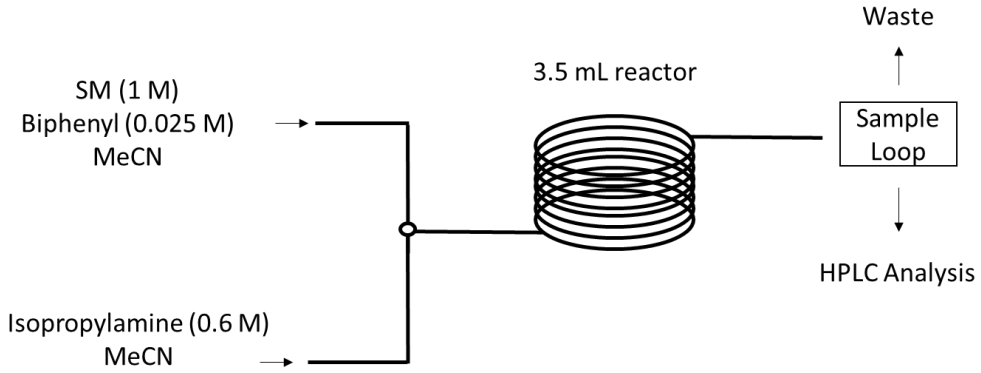


Figure 7.3.9: The reactor setup for the metoprolol flow ramp experiments at 190/210 °C in Leeds.

For the AstraZeneca experiments, a 5 mL reactor was used. Linear gradient flow ramps allowed the generation of complete reaction profiles from a single transient experiment. To obtain transient data, each of the two pumps were initially set at the maximum flow rate to be investigated: 1.25 mL min^{-1}. Steady-state was established and the flow rate for each pump decreased at a constant rate of 0.0125 mL min^{-1} for 96 minutes. Samples of reactor effluent were injected for HPLC analysis at 6 minute intervals, thus achieving a large data density. This reactor setup is shown in Figure 7.3.10.

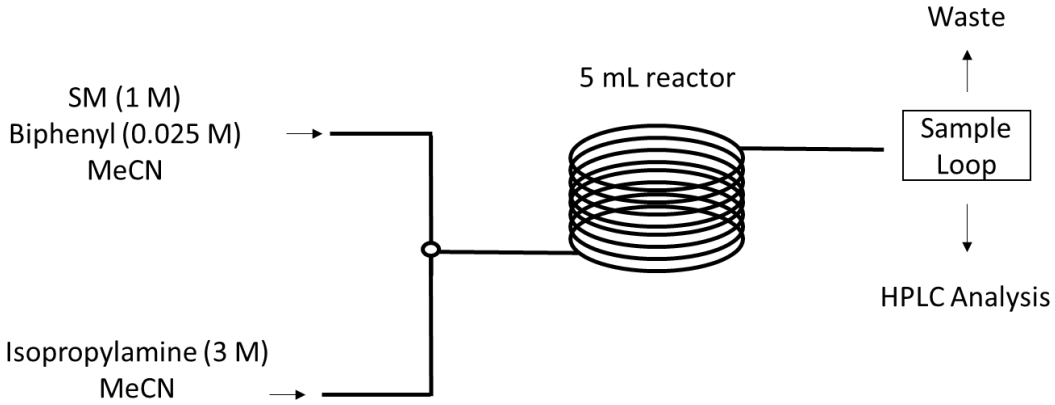


Figure 7.3.10: The reactor setup for the metoprolol flow ramp experiments at 130/150 °C at AstraZeneca.

7.3.3.4 HPLC analysis

All HPLC analysis was conducted using an Agilent 1100 series HPLC instrument fitted with a Sigma Ascentis Express C18 reverse phase column (5cm x 4.6mm, 2.7 µm). Biphenyl was used as an internal standard. The column temperature was 40 °C and the HPLC method is shown:

Time /min	%A (water, 0.1 % TFA)	%B (acetonitrile, 0.1 % TFA)	Flow rate /mL min^{-1}
0.00	77	23	2
3.00	10	90	2
4.00	10	90	2
4.10	77	23	2

An example HPLC chromatogram during a kinetic experiment is shown in Figure 7.3.11:

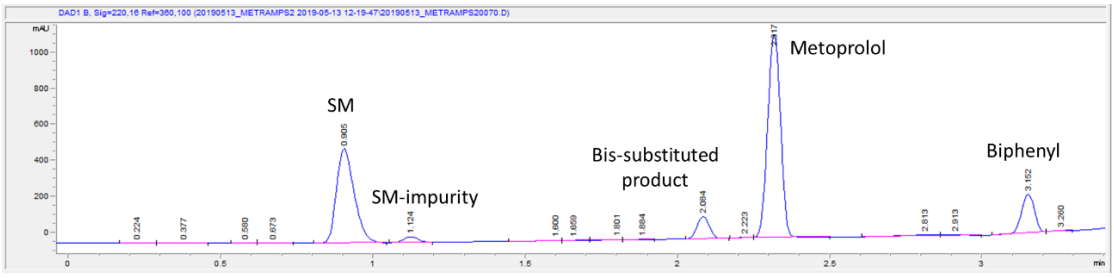


Figure 7.3.11: An example HPLC chromatogram during a kinetic experiment in the metoprolol case study at 220 nm.

7.3.3.5 Cost evaluation

Using the Leeds experimental setup, outlined in Chapter 3.2, and current pricing for the starting material used in the Metoprolol work (£807/25 g),[239] it has been shown that this kinetic approach used less material and was cheaper to implement than other standard industrial optimisation techniques.

As the starting material is the most expensive reaction component by a large margin, this was the only factor taken into consideration - other costs were not considered e.g. solvent, isopropylamine, energy etc. It is assumed that a 1 M solution of starting material is used, and so the cost associated for every 1 mL of this solution used (0.208 g ,1 mmol) is: £6.71. It is also assumed that 2 reactor volumes are consumed prior to all steady-state reactions, and all reactions were

conducted using a 3.5 mL reactor. The addition of skilled labour and time have not been considered, but would incur significantly higher costs for these other techniques, as this technique only requires the initial setup of the reservoirs and reactor because everything else is automated.

Technique	What is required	Solution usage /mL	Cost /£
Flow ramps	4 x linearly decreasing flow ramps, from 1.75 mL min^{-1}, decreasing at a rate of 0.0181 mL min-1 for 92 minutes.	340	£2281
Steady-state kinetics	4 sets of 16 steady-state experiments.	420	£2818
Design of Experiments study	One screening design (19 steady-state experiments) and one optimisation design (81 steady-state experiments) to consider the factors of: temperature, residence time, starting material equivalents and isopropylamine equivalents.	700	£4697

7.3.3.6 Raw data

Flow ramp experiment 1, 130 °C

τ /min	3.17 /M	3.19 /M	3.20 /M
2.10	0.435	0.053	0.000
2.24	0.427	0.056	0.000
2.40	0.423	0.060	0.000
2.58	0.417	0.064	0.000
3.71	0.381	0.102	0.000
4.16	0.397	0.088	0.000
4.72	0.383	0.099	0.000
5.45	0.369	0.110	0.000
6.41	0.362	0.123	0.000
7.73	0.347	0.146	0.001
9.61	0.286	0.206	0.002

Flow ramp experiment 2, 150 °C

τ /min	3.17 /M	3.19 /M	3.20 /M
2.00	0.358	0.116	0.000
2.10	0.357	0.125	0.000
2.24	0.350	0.135	0.001
2.40	0.347	0.144	0.001
2.58	0.340	0.153	0.001
2.80	0.333	0.166	0.001
3.05	0.311	0.178	0.002
3.35	0.299	0.193	0.002
3.71	0.269	0.207	0.002
4.16	0.259	0.234	0.003
4.72	0.233	0.251	0.003
5.45	0.208	0.287	0.004
6.41	0.169	0.284	0.005
7.73	0.144	0.327	0.006

Flow ramp experiment 3, 190 °C

τ /min	3.17 /M	3.19 /M	3.20 /M
2.55	0.340	0.117	0.004
2.69	0.328	0.121	0.005
2.84	0.321	0.126	0.006
3.02	0.320	0.132	0.006
3.21	0.309	0.138	0.007
3.43	0.307	0.137	0.007
3.68	0.305	0.150	0.008
3.97	0.299	0.152	0.009
4.30	0.294	0.161	0.010
4.70	0.284	0.171	0.012
5.17	0.270	0.176	0.014
5.73	0.261	0.183	0.016
6.43	0.254	0.186	0.018
7.30	0.246	0.190	0.021
8.42	0.232	0.211	0.025
9.86	0.213	0.217	0.028

Flow ramp experiment 4, 210 °C

τ /min	3.17 /M	3.19 /M	3.20 /M
2.50	0.305	0.166	0.012

2.63	0.299	0.165	0.013
2.78	0.294	0.172	0.014
2.95	0.294	0.172	0.015
3.14	0.282	0.182	0.017
3.35	0.279	0.186	0.019
3.60	0.269	0.190	0.021
3.88	0.272	0.201	0.023
4.21	0.247	0.221	0.030
4.60	0.236	0.218	0.028
5.06	0.230	0.219	0.032
5.62	0.218	0.222	0.038
6.30	0.231	0.211	0.042
7.16	0.201	0.235	0.045
8.25	0.188	0.235	0.051
9.68	0.182	0.239	0.058

7.4 Chapter 4

7.4.1 Generated data set for the case study: S_NAr kinetics

Experiment 1

Time /s	**4.5 /M**	**4.7 /M**	**4.8 /M**	**4.9 /M**
0	1	0	0	0
1.04232	0.912548	0.068648	0.002878	2.79E-05
5.078096	0.715174	0.266199	0.011023	0.000485
14.75526	0.464384	0.521723	0.020875	0.002315
33.31443	0.269972	0.708549	0.02876	0.006042
56.28294	0.168854	0.80687	0.031878	0.009439
90.17797	0.099139	0.856211	0.03494	0.013887
210.8206	0.032979	0.88963	0.036126	0.024236
332.8275	0.015074	0.929082	0.036808	0.03211
438.2468	0.008534	0.933212	0.036452	0.037487
600	0.004209	0.921486	0.036363	0.043389

Experiment 2

Time /s	4.5 /M	4.7 /M	4.8 /M	4.9 /M
0	0.8	0	0	0
1.302612	0.703325	0.092533	0.003854	6.27E-05
5.712155	0.478466	0.312694	0.012573	0.000883
10.84263	0.329335	0.44882	0.018183	0.002403
20.64908	0.176938	0.596415	0.023908	0.005922
69.41793	0.014924	0.723646	0.029388	0.023796
151.6461	0.000354	0.725866	0.029244	0.051839
300.9398	3.80E-07	0.66851	0.026369	0.099546
515.3318	0	0.621532	0.024249	0.155196
925.3113	0	0.545081	0.019839	0.239656
1200	0	0.506326	0.018372	0.288136

Experiment 3

Time /s	4.5 /M	4.7 /M	4.8 /M	4.9 /M
0	1	0	0	0
0.962483	0.873362	0.112646	0.004695	7.80E-05
10.02158	0.342803	0.632465	0.025508	0.004215
21.06925	0.138132	0.817645	0.033569	0.01147
60.78714	0.008667	0.899281	0.037403	0.036358
101.0498	0.000643	0.895452	0.03658	0.061674
172.7609	7.14E-06	0.851082	0.034174	0.101391
308.2939	0	0.806696	0.030438	0.17491
562.1972	0	0.69294	0.025835	0.270209
904.0341	0	0.598858	0.021781	0.384583
1600	0	0.457229	0.015241	0.527733

Experiment 4

Time /s	4.5 /M	4.7 /M	4.8 /M	4.9 /M

0	1	0	0	0
1.044725	0.854673	0.149213	0.006087	0.000139
5.09551	0.460925	0.515756	0.020548	0.002257
21.32233	0.072138	0.877848	0.035074	0.017175
47.14859	0.005208	0.917717	0.036551	0.041614
69.31167	0.0006	0.887854	0.035809	0.061604
128.0495	2.25E-06	0.86418	0.033593	0.115427
288.6839	0	0.741762	0.02754	0.232766
736.8751	0	0.519088	0.018297	0.455041
1222.483	0	0.381306	0.012248	0.60453
1800	0	0.274267	0.007936	0.705143

Experiment 5

Time /s	**4.5 /M**	**4.7 /M**	**4.8 /M**	**4.9 /M**
0	1.2	0	0	0
0.849478	0.999855	0.175423	0.007178	0.000158
5.548213	0.466871	0.712408	0.028769	0.004137
11.81201	0.191904	0.969199	0.039345	0.012093
35.00272	0.010569	1.10712	0.045047	0.044862
91.1267	1.47E-05	1.038677	0.040673	0.119288
123.7999	6.90E-07	1.005138	0.040148	0.158185
195.3324	0	0.920941	0.03614	0.234604
483.1808	0	0.709117	0.024989	0.478645
632.278	0	0.609582	0.021143	0.562277
900	0	0.502989	0.016122	0.680497

7.4.2 Generated data set for the case study: Pentyne

Experiment 1

Time /min	**4.11 /M**	**4.12 /M**	**4.13 /M**
0	1	0.8	0

1.293971	0.973192	0.803545	0.011085
3.972297	0.958388	0.766506	0.033251
18.43854	0.857109	0.663378	0.141812
32.90479	0.770748	0.54892	0.246498
56.90479	0.629294	0.416498	0.388829
92.90479	0.449379	0.25987	0.548273
120	0.369111	0.164188	0.644284

Experiment 2

Time /min	**4.11 /M**	**4.12 /M**	**4.13 /M**
0	1.2	1.5	0
1.205027	1.199383	1.460822	0.012139
3.721389	1.157799	1.439238	0.037013
18.02628	1.036602	1.347405	0.171608
50.33117	0.77347	1.092225	0.421136
86.33117	0.582851	0.871663	0.614124
140.3312	0.359291	0.678145	0.833511
180	0.259447	0.551321	0.936175

Experiment 3

Time /min	**4.11 /M**	**4.12 /M**	**4.13 /M**
0	1	1	0
1.293971	0.994904	0.974041	0.010778
3.972297	0.975053	0.966892	0.033053
18.43854	0.86422	0.871237	0.145646
32.90479	0.750116	0.752064	0.246904
80.90479	0.508082	0.506177	0.504648
152.9048	0.276204	0.271468	0.72687
240	0.129659	0.132753	0.856012

7.4.3 Generated data set for the case study: Ytterbium catalysis

Experiment 1

Time /h	4.15 /M	4.16 /M	4.18 /M
0	2	2	0
0.054397	1.950736	1.962883	0.040294
0.386129	1.737689	1.725949	0.26888
0.701874	1.545843	1.542334	0.459994
1.901874	0.986091	0.979723	1.020241
3.101874	0.631223	0.628196	1.36402
4.301874	0.40357	0.400764	1.596431
6	0.213539	0.212908	1.784115

Experiment 2

Time /h	4.15 /M	4.16 /M	4.18 /M
0	1.5	2	0
0.029722	1.499822	2.000933	0.004169
0.2413	1.470961	1.962226	0.033316
0.911826	1.373396	1.881085	0.122394
1.511826	1.302576	1.805505	0.196746
2.711826	1.164105	1.659122	0.336104
3.911826	1.042952	1.535555	0.457449
6	0.85898	1.357163	0.642853

Experiment 3

Time /h	4.15 /M	4.16 /M	4.18 /M
0	1.4	1.8	0
0.027487	1.370804	1.774104	0.03184
0.106977	1.284762	1.675632	0.12038
0.249912	1.14023	1.539842	0.264912

0.92138	0.646932	1.043942	0.752256
1.52138	0.389843	0.794653	1.012081
2.12138	0.235788	0.635663	1.164374
3	0.112972	0.511689	1.287393

Experiment 4

Time /h	**4.15 /M**	**4.16 /M**	**4.18 /M**
0	2.2	1.6	0
0.043551	2.153819	1.53798	0.05486
0.241672	1.906782	1.311278	0.288655
0.439793	1.706777	1.105476	0.49592
0.839793	1.351856	0.751683	0.854071
1.239793	1.066839	0.466383	1.129747
1.639793	0.846621	0.246217	1.354339
2	0.682943	0.085826	1.509085

7.5 Chapter 5

7.5.1 S_NAr kinetics

7.5.1.1 Experimental equipment

All experiments were run in an NMR tube using a dedicated 500 MHz NMR spectrometer at The University of Cambridge, as shown in Figure 7.5.1.

Figure 7.5.1: The dedicated NMR spectrometer used for kinetic experiments in the S_NAr kinetics case study.

7.5.1.2 Kinetic experiments

In each experiment, 2,4,6-trichloropyrimidine (700 µL, 200 mM in CD_3OD) was warmed/cooled to the required temperature in the 500 MHz NMR spectrometer. The NMR was then locked onto the solvent, then the probe was tuned and matched to the ^1H nuclei, then shimming ensued. The spectrum of the 200 mM solution of 2,4,6-trichloropyrimidine was acquired.

To this NMR tube, a pre-cooled/warmed solution of CD_3OD containing 200 mM ethyl 4-aminobutyrate hydrochloride and 600 mM Et_3N (700 µL) was added and mixed thoroughly by shaking the NMR tube. This gave the desired reaction concentration of 100 mM pyrimidine, 100 mM amine and 300 mM Et_3N. Spectra were acquired every ~80 seconds. The software was used in each case to obtain absolute integrals for each peak, which were then converted to

concentrations based on the 2,4,6-trichloropyrimidine solution of known concentration.

7.5.1.3 Example NMR spectra

An example ‘stacked’ NMR spectrum is shown in Figure 7.5.2, showing the appearance/disappearance of peaks over time in the 25 °C experiment. Where the peak at 7.80 ppm corresponds to 2,4,6-trichloropyrimidine, 6.66 ppm corresponds to the 2-substituted product and 6.41 corresponds to the 4-substited product.

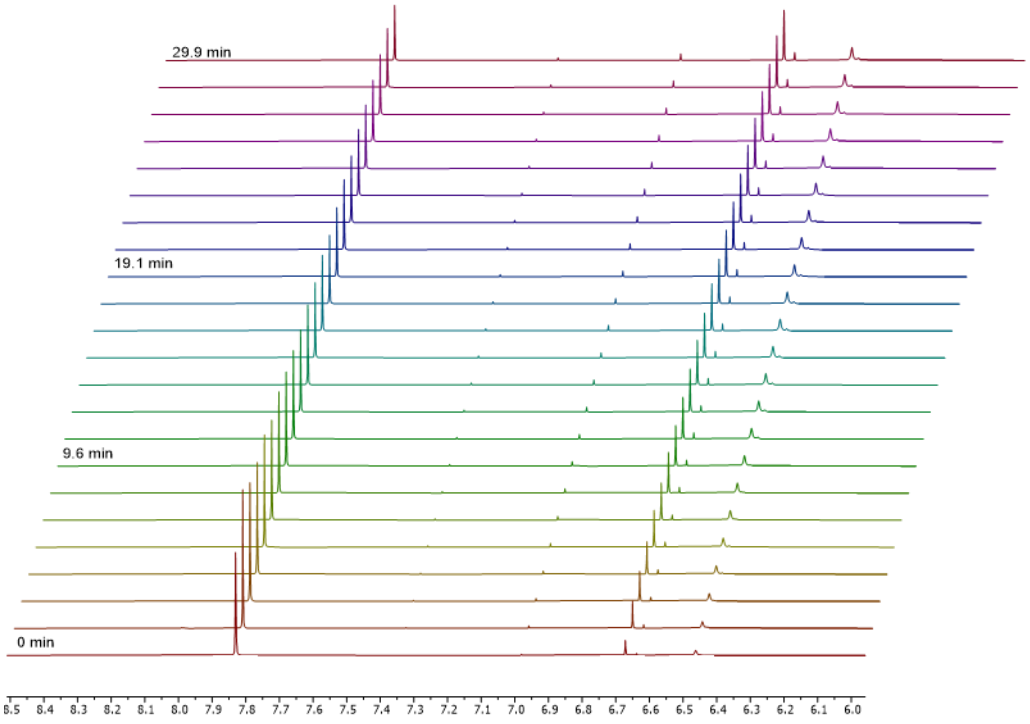


Figure 7.5.2: ‘Stacked’ NMR spectrum showing several NMR spectra over time in the 25 °C experiment.

7.5.1.4 Raw data

Experiment 1, -25 °C

Time /min	5.1 /M	5.3 /M	5.4 /M
14.55	0.088818	0.004156	0.004562
15.91667	0.088538	0.004229	0.004611
17.26667	0.088465	0.004295	0.004727
18.63333	0.088063	0.004362	0.004842
19.98333	0.08805	0.004577	0.004908

21.33333	0.087811	0.004609	0.004962
22.7	0.087412	0.004713	0.005096
24.05	0.087133	0.004827	0.005202
25.41667	0.086922	0.004901	0.005268
26.76667	0.08667	0.004993	0.005345
28.13333	0.086501	0.005049	0.005407
29.48333	0.086142	0.005142	0.005518
30.83333	0.085983	0.005268	0.005704
32.2	0.085639	0.005412	0.005803
33.55	0.085286	0.005468	0.005826
34.91667	0.085108	0.005589	0.00596
36.26667	0.084887	0.005646	0.005947
37.63333	0.084482	0.005781	0.006102
38.98333	0.084308	0.005902	0.006179
40.35	0.083927	0.005914	0.00626
41.7	0.083758	0.006021	0.00639
43.06667	0.083386	0.006157	0.006496
44.41667	0.083205	0.006154	0.006574
45.76667	0.083093	0.006298	0.006627
47.13333	0.08282	0.006377	0.006689
48.48333	0.082663	0.006475	0.006751
49.85	0.082501	0.006492	0.006786
51.2	0.082273	0.00655	0.006939
52.56667	0.08206	0.006658	0.006902
53.91667	0.08186	0.006786	0.007038
55.26667	0.081225	0.006729	0.007105

Experiment 2, 0 °C

Time /min	**5.1 /M**	**5.3 /M**	**5.4 /M**

7.333333	0.072103	0.007607	0.008513
8.8	0.070344	0.008018	0.008965
10.16667	0.068594	0.008452	0.009462
11.51667	0.068423	0.008887	0.009854
12.88333	0.066803	0.009266	0.01028
14.23333	0.065829	0.009635	0.010682
15.6	0.064594	0.010002	0.011061
16.95	0.063823	0.010341	0.011471
18.31667	0.062319	0.010702	0.011804
19.66667	0.061394	0.010984	0.012177
21.01667	0.060906	0.011342	0.012486
22.38333	0.059741	0.011585	0.012806
23.73333	0.058765	0.01182	0.013095
25.1	0.058315	0.012151	0.013445
26.45	0.057463	0.01241	0.013671
27.8	0.056496	0.012685	0.013978
29.16667	0.056066	0.0129	0.014287
30.51667	0.055037	0.013136	0.01449
31.88333	0.054482	0.013382	0.014789
33.23333	0.053852	0.013629	0.015023
34.58333	0.053151	0.013804	0.015225
35.95	0.052684	0.014054	0.01548
37.3	0.052192	0.014261	0.015707
38.66667	0.05156	0.014409	0.01589
40.01667	0.051008	0.014613	0.016133
41.36667	0.050399	0.014842	0.016311
42.73333	0.049812	0.015017	0.016498
44.08333	0.049557	0.015136	0.016666
45.45	0.048743	0.015301	0.016861
46.8	0.048366	0.015482	0.017042

48.16667	0.047763	0.015635	0.01723
49.51667	0.047326	0.015831	0.017433
50.86667	0.046947	0.015973	0.017573
52.23333	0.046384	0.016088	0.017747
53.58333	0.045991	0.016243	0.017903
54.95	0.045672	0.016393	0.018043
56.3	0.045107	0.016496	0.018225
57.66667	0.044674	0.016691	0.018395
59.01667	0.044423	0.016839	0.018509
60.36667	0.044113	0.016972	0.018683
61.73333	0.043619	0.017086	0.01884

Experiment 3, 25 °C

Time /min	**5.1 /M**	**5.3 /M**	**5.4 /M**
3.366667	0.064314	0.009652	0.009883
4.8	0.057052	0.012118	0.012863
6.15	0.051792	0.013813	0.014922
7.5	0.047939	0.015186	0.016476
8.866667	0.044748	0.016215	0.017819
10.21667	0.041922	0.017079	0.018798
11.58333	0.039664	0.017878	0.019678
12.93333	0.037598	0.018491	0.020463
14.28333	0.035777	0.019052	0.021122
15.65	0.034195	0.019558	0.021718
17	0.032928	0.019875	0.022181
18.36667	0.031709	0.020362	0.022676
19.71667	0.030632	0.020622	0.023126
21.06667	0.029623	0.020886	0.023404
22.43333	0.028668	0.02128	0.023818

23.78333	0.027704	0.021452	0.02409
25.15	0.026937	0.021641	0.024389
26.5	0.026303	0.021884	0.024635
27.86667	0.025572	0.022112	0.024925
29.21667	0.024988	0.022265	0.025126
30.56667	0.024358	0.022417	0.025261
31.93333	0.023747	0.022601	0.025533
33.28333	0.023156	0.022788	0.025713
34.63333	0.022678	0.022896	0.025943
36	0.022158	0.022984	0.026042
37.35	0.021731	0.023193	0.026254
38.71667	0.021252	0.023299	0.026356
40.06667	0.020811	0.023542	0.026561
41.43333	0.020451	0.023635	0.026706
42.78333	0.019954	0.023675	0.026824
44.13333	0.01972	0.023743	0.026953
45.5	0.019374	0.023807	0.027033
46.85	0.019129	0.023925	0.027167
48.21667	0.018686	0.024057	0.027249
49.56667	0.018359	0.024079	0.027301
50.93333	0.018011	0.024171	0.027439
52.28333	0.017894	0.024265	0.027549
53.63333	0.017535	0.024319	0.027649
55	0.017377	0.024477	0.027784
56.35	0.01712	0.024351	0.027716
57.71667	0.016878	0.024502	0.027857

Experiment 4, 50 °C

Time /min	**5.1 /M**	**5.3 /M**	**5.4 /M**

3.35	0.044859	0.017049	0.028158
5.066667	0.033058	0.020402	0.034142
6.416667	0.028533	0.021812	0.036713
7.766667	0.025543	0.022842	0.038521
9.133333	0.022916	0.0235	0.039813
10.5	0.021041	0.024107	0.04121
11.85	0.019498	0.024602	0.041829
13.2	0.018156	0.025096	0.042681
14.56667	0.016918	0.025264	0.043145
15.91667	0.015895	0.025503	0.043496
17.28333	0.015031	0.025778	0.043978
18.63333	0.014386	0.025935	0.044504
20	0.013754	0.026166	0.044669
21.35	0.013006	0.026135	0.045027
22.7	0.012731	0.02645	0.04525
24.06667	0.012071	0.026529	0.045439
25.41667	0.011723	0.026593	0.045892
26.78333	0.011273	0.026724	0.046128
28.13333	0.01075	0.026737	0.04605
29.5	0.010443	0.026863	0.046276
30.85	0.010126	0.026964	0.046178
32.21667	0.009967	0.027158	0.046613
33.56667	0.009483	0.0272	0.046695
34.93333	0.009351	0.027183	0.047223
36.28333	0.009066	0.027134	0.047014
37.63333	0.008787	0.027312	0.047229
39	0.008496	0.027291	0.047245
40.35	0.00833	0.027346	0.047363
41.71667	0.00802	0.027313	0.047267
43.06667	0.007943	0.027459	0.047585

44.43333	0.007797	0.027504	0.047612
45.78333	0.007612	0.027442	0.047469
47.13333	0.007363	0.027731	0.047777
48.5	0.007162	0.027549	0.047689
49.85	0.007122	0.027588	0.047921
51.2	0.006956	0.027672	0.048034
52.56667	0.006869	0.027621	0.047948
53.91667	0.006654	0.02781	0.048184
55.28333	0.006511	0.027703	0.048181
56.63333	0.006367	0.027739	0.048193
57.98333	0.006367	0.027751	0.048241

7.5.2 PfBr

7.5.2.1 Synthesis of PfBr material

To synthesise the PfBr material, the synthesis reported by Tian and Menard was approximately followed.[249] To a suspension of magnesium turnings (9.2 g, 382 mmol) in anhydrous THF (5 mL) in a nitrogen atmosphere, was added 0.5 mL of bromobenzene at room temperature to initiate the reaction. The reaction vessel was cooled to 0 °C using an ice bath and a solution of bromobenzene in THF (1.2 M, 270 mL) was added dropwise at 0 °C. The reaction medium was allowed to reach room temperature and stirred vigorously for 2 hours. To this solution was added 9-fluorenone (30.1 g, 186 mmol) in small portions left to stir at room temperature for 1 hour. The reaction was quenched with dropwise addition of hydrochloric acid (1 M, 100 mL) at 0 °C. The solution was extracted with Et_2O (3 x 150 mL), and the combined organic layer was washed with brine, dried over $MgSO_4$, filtered and concentrated under reduced pressure to yield PfOH (42.6 g, 89 %). NMR, IR and mass spectroscopy analysis confirmed the identity of the material.

To prepare PfBr, PfOH (36.1 g, 140 mmol) was dissolved in toluene (150 mL) and aqueous HBr (48 % w/w, 50 mL) was added at room temperature. This suspension was stirred vigorously at room temperature for 48 hours, with the vessel wrapped in foil to reduce light. The mixture was extracted with toluene (3

x 100 mL) then the combined organic layer was washed with brine, dried over $MgSO_4$, filtered and concentrated under reduced pressure to afford the crude product. The product was then recrystallised with hexane to afford PfBr (39 g, 87 %). ^{13}C NMR, IR and HPLC analysis confirmed the identity and purity of the material, as shown in Figure 7.5.3, Figure 7.5.4 and Figure 7.5.5 respectively.

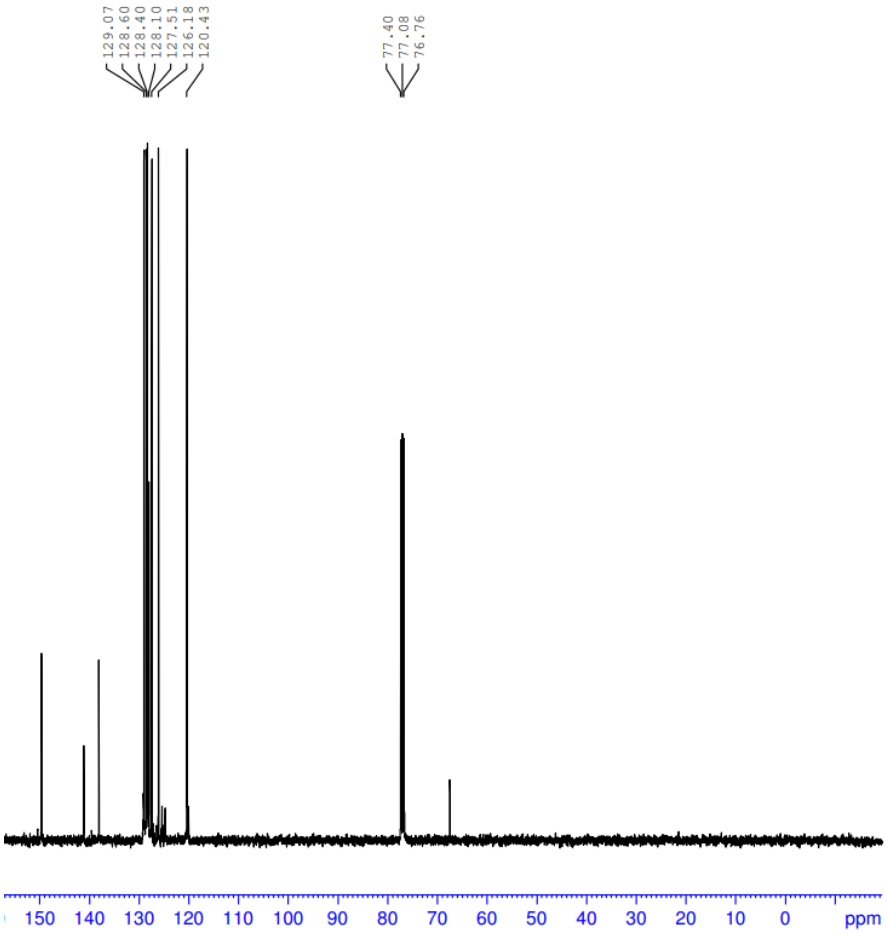


Figure 7.5.3: A ^{13}C NMR confirming the presence of the PfBr material, compared to literature values to identify the material.[249]

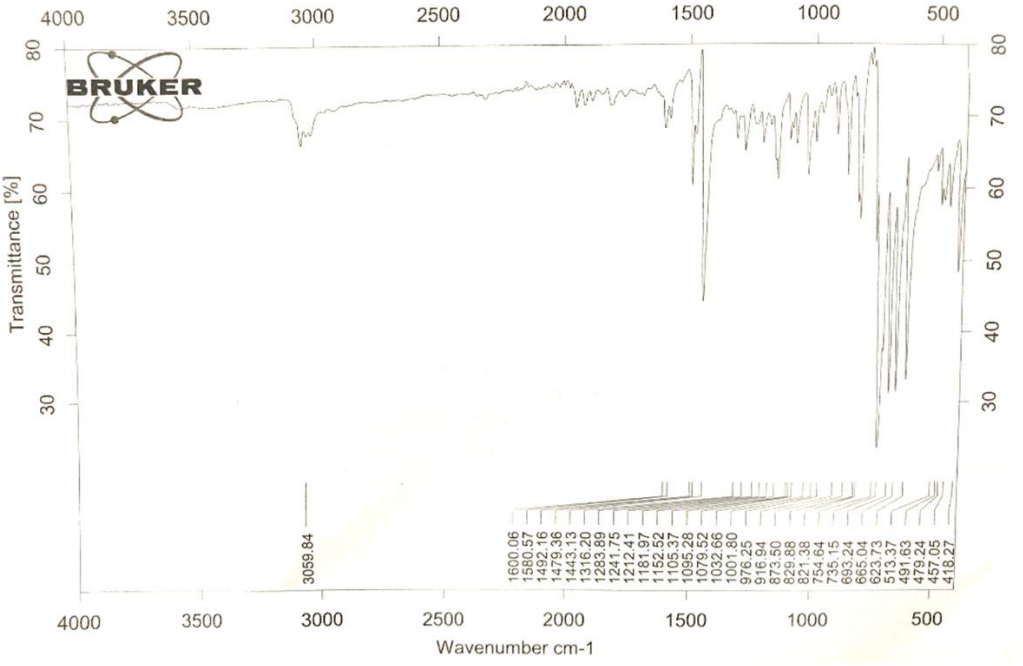


Figure 7.5.4: IR analysis of the PfBr material, showing the absence of a hydroxy peak from any residual PfOH.

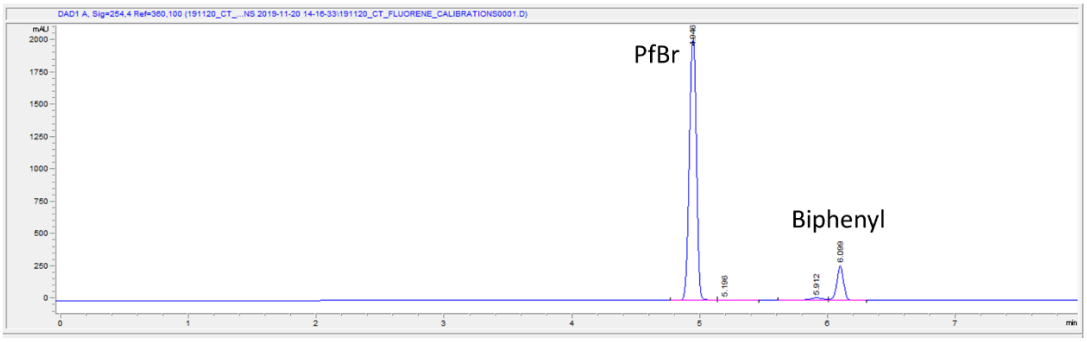


Figure 7.5.5: HPLC analysis of the PfBr material at 254 nm, indicating that a pure product is present.

7.5.2.2 Kinetic experiments

Three kinetic experiments were run at different temperatures: 30 °C, 35 °C and 40 °C. For the first experiment at 30 °C, alanine methyl ester hydrochloride (0.940 g, 0.007 mol), biphenyl (0.195 g, 0.001 mol) and potassium phosphate (3.5 g, 0.017 mol) were added to 50:50 acetonitrile/dichloromethane (37.5 mL) in a vessel. A solution was then prepared of PfBr (0.958 g, 0.003 mol) in 50:50 acetonitrile/dichloromethane (37.5 mL) and added to the reaction vessel. A sample was then immediately taken via HPLC, then the stopwatch was started to track the time of each injection. This first HPLC serves as the artificial

zero time point, then each subsequent HPLC injection followed every ~10 minutes thereafter.

For the second experiment at 35 °C, alanine methyl ester hydrochloride (0.977 g, 0.007 mol), biphenyl (0.199 g, 0.001 mol) and potassium phosphate (3.5 g, 0.017 mol) were added to 50:50 acetonitrile/dichloromethane (37.5 mL) in a vessel. A solution was then prepared of PfBr (1.017 g, 0.003 mol) in 50:50 acetonitrile/dichloromethane (37.5 mL) and added to the reaction vessel. A sample was then immediately taken via HPLC, then the stopwatch was started to track the time of each injection. This first HPLC serves as the artificial zero time point, then each subsequent HPLC injection followed every ~10 minutes thereafter.

For the third experiment at 40 °C, alanine methyl ester hydrochloride (0.841 g, 0.006 mol), biphenyl (0.219 g, 0.001 mol) and potassium phosphate (3.5 g, 0.017 mol) were added to 50:50 acetonitrile/dichloromethane (37.5 mL) in a vessel. A solution was then prepared of PfBr (1.064 g, 0.003 mol) in 50:50 acetonitrile/dichloromethane (37.5 mL) and added to the reaction vessel. A sample was then immediately taken via HPLC, then the stopwatch was started to track the time of each injection. This first HPLC serves as the artificial zero time point, then each subsequent HPLC injection followed every ~10 minutes thereafter.

7.5.2.3 HPLC analysis

All HPLC analysis was conducted using an Agilent 1100 series HPLC instrument fitted with a Sigma Ascentis Express C18 reverse phase column (5cm x 4.6mm, 2.7 µm). Biphenyl was used as an internal standard. The column temperature was 40 °C and the HPLC method is shown:

Time /min	**%A (water, 0.1 % TFA)**	**%B (acetonitrile, 0.1 % TFA)**	**Flow rate /mL min^{-1}**
0.00	80	20	1
7.00	15	85	1
8.00	15	85	1
8.10	80	20	1

An example HPLC chromatogram during a kinetic experiment is shown in Figure 7.5.6:

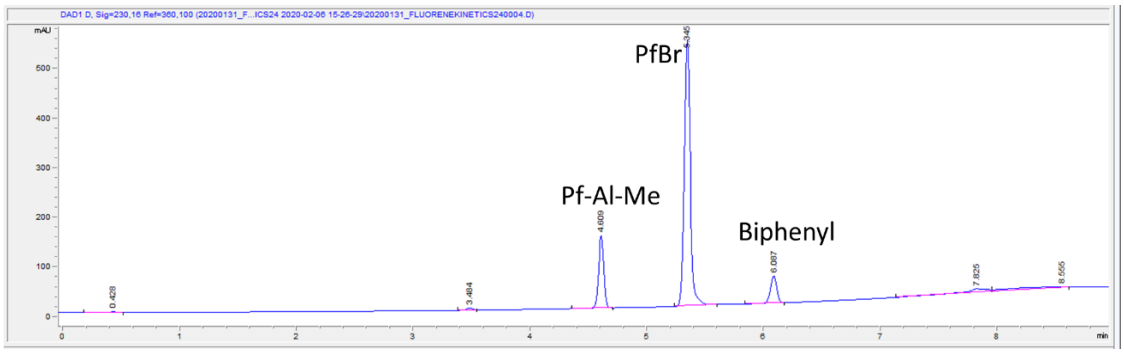


Figure 7.5.6: An example HPLC chromatogram at 230 nm during a kinetic experiment in the PfBr case study.

7.5.2.4 Raw data

Experiment 1, 30 °C

Time /min	5.7 /M	5.8 /M
0	0.029401	0.009756
9.5	0.028029	0.011516
19.5	0.026429	0.013131
29	0.024664	0.014922
40	0.022779	0.016856
50	0.021252	0.019153
60.5	0.019363	0.020372
71	0.017813	0.022065
83	0.01612	0.023823

Experiment 2, 35 °C

Time /min	5.7 /M	5.8 /M
0	0.036314	0.006684
9.5	0.030494	0.011469
20	0.028083	0.01372
31.5	0.025271	0.016536
41	0.023355	0.018562

51	0.021259	0.020741
62.5	0.017572	0.023705

Experiment 3, 40 °C

Time /min	**5.7 /M**	**5.8 /M**
0	0.037032	0.005772
10.5	0.029931	0.013994
20.5	0.026302	0.017446
30.5	0.022947	0.02101
40	0.020038	0.023665
50.5	0.017687	0.025741
60.5	0.016003	0.027979

7.5.3 Maleic acid

7.5.3.1 Example NMR

An example NMR spectrum is shown in Figure 7.5.4, where the peaks of each species are highlighted.

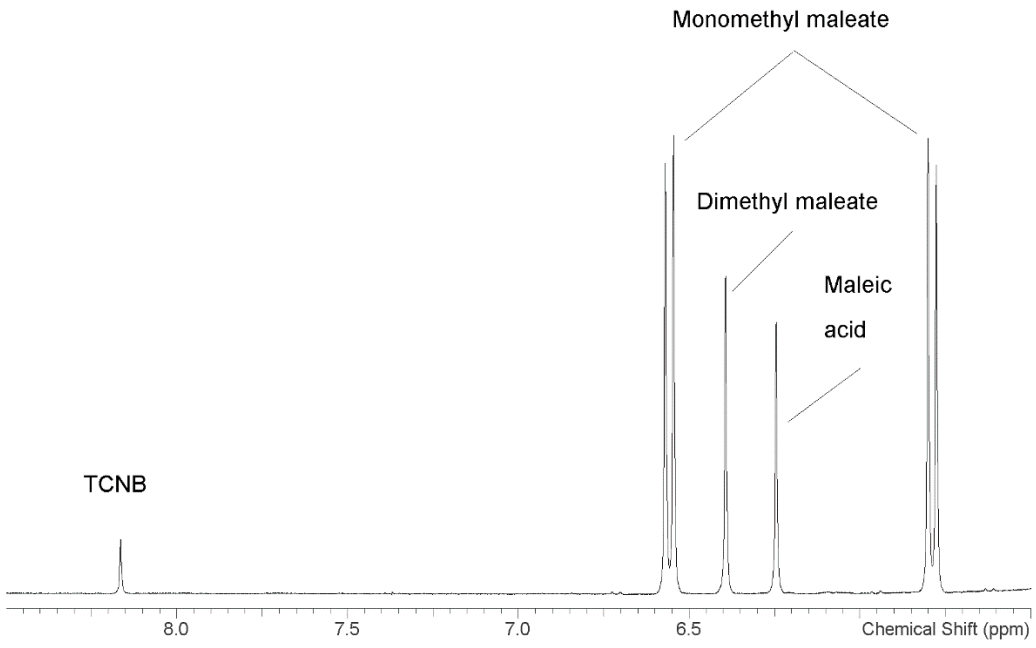


Figure 7.5.7: An example NMR spectrum from the maleic acid case study.

7.5.3.2 Raw data

Experiment 1, 40 °C

Time /min	5.13 /M	5.15 /M	5.16 /M
0	0.761	0	0
120	0.66207	0.097408	0.002283
240	0.554769	0.201665	0.004947
360	0.485518	0.265589	0.009893

Experiment 2, 50 °C

Time /min	5.13 /M	5.15 /M	5.16 /M
0	0.404344	0.003268	0
32.5	0.364451	0.037657	1.86E-05
60.5	0.346181	0.071026	0.000373
87.5	0.319523	0.095447	0.001659
122.5	0.299017	0.113343	0.003225
217.5	0.234143	0.1661	0.006264
288.5	0.204502	0.198537	0.009507
417.5	0.15883	0.230601	0.015846
687.5	0.107378	0.276833	0.028336
1342.5	0.052198	0.313931	0.053689
1782.5	0.033649	0.310389	0.063196
2837.5	0.01799	0.307033	0.083516

Experiment 3, 50 °C

Time /min	5.13 /M	5.15 /M	5.16 /M
0	0.787537	0.007088	0
28	0.706604	0.088537	5.65E-04
60	0.625073	0.164707	0.002563

95	0.556919	0.226165	0.006126
118	0.530477	0.271021	0.007904
148	0.480567	0.308812	0.011678
178	0.426661	0.316753	0.01425
208	0.405376	0.35592	0.020026
242	0.384453	0.382377	0.021991
298	0.344452	0.431289	0.029072
358	0.299951	0.456885	0.034314
613	0.205957	0.541791	0.061416
1306	0.092916	0.582232	0.105274
1549	0.076529	0.596801	0.119691
1768	0.063259	0.597105	0.130915

Experiment 4, 55 °C

Time /min	**5.13 /M**	**5.15 /M**	**5.16 /M**
0	0.687045	0.103644	0.000811
30	0.597449	0.19108	2.97E-03
60	0.531145	0.252908	0.007448
90	0.479052	0.301502	0.010947
168	0.365378	0.402913	0.023209
228	0.31788	0.443385	0.030235
283	0.28467	0.469404	0.037426
348	0.221288	0.520375	0.049837
408	0.200913	0.533579	0.057007

Experiment 4, 60 °C

Time /min	**5.13 /M**	**5.15 /M**	**5.16 /M**
0	0.738693	0.014307	0
60	0.442764	0.296682	1.36E-02

120	0.335085	0.389301	0.029367
180	0.265809	0.447282	0.040662
240	0.224394	0.475143	0.053463
360	0.117468	0.551196	0.084336
480	0.070029	0.575292	0.107679
1320	0.015813	0.538395	0.198792

7.6 Computational setup

7.6.1 Chapter 2 - 3

For all computation in Chapter 2 - 3, MATLAB 2018a was used on a desktop computer with a 4-core Intel i5-2310 processor.

7.6.2 Chapter 4 - 5

For all computation in Chapter 4 - 5, MATLAB 2020a was used on a custom built laptop with an 8-core AMD Ryzen 7 4800H processor.

Chapter 8 : References

1. Miyaura, N., K. Yamada, and A. Suzuki, *A new stereospecific cross-coupling by the palladium-catalyzed reaction of 1-alkenylboranes with 1-alkenyl or 1-alkynyl halides.* Tetrahedron Letters, 1979. **20**(36): p. 3437-3440.
2. Sonogashira, K., Y. Tohda, and N. Hagihara, *A convenient synthesis of acetylenes: catalytic substitutions of acetylenic hydrogen with bromoalkenes, iodoarenes and bromopyridines.* Tetrahedron Letters, 1975. **16**(50): p. 4467-4470.
3. Wittig, G. and W. Haag, *Über Triphenyl-phosphinmethylene als olefinbildende Reagenzien (II. Mitteil. 1).* Chemische Berichte, 1955. **88**(11): p. 1654-1666.
4. Baumann, M., et al., *A Perspective on Continuous Flow Chemistry in the Pharmaceutical Industry.* Organic Process Research & Development, 2020.
5. Noél, T., *Flow into the Chemistry Curriculum.* Chemistry World, 2019.
6. Reizman, B.J. and K.F. Jensen, *Feedback in flow for accelerated reaction development.* Accounts of Chemical Research, 2016. **49**(9): p. 1786-1796.
7. Plutschack, M.B., et al., *The Hitchhiker's Guide to Flow Chemistry.* Chemical Reviews, 2017. **117**(18): p. 11796-11893.

8. Granda, J.M., et al., *Controlling an organic synthesis robot with machine learning to search for new reactivity.* Nature, 2018. **559**(7714): p. 377-381.
9. Peela, N.R., I.C. Lee, and D.G. Vlachos, *Design and fabrication of a high-throughput microreactor and its evaluation for highly exothermic reactions.* Industrial & Engineering Chemistry Research, 2012. **51**(50): p. 16270-16277.
10. Rivera, N.R., et al., *Investigation of a Flow Step Clogging Incident: A Precautionary Note on the Use of THF in Commercial-Scale Continuous Process.* Organic Process Research & Development, 2019. **23**(11): p. 2556-2561.
11. Chapman, M.R., et al., *Simple and versatile laboratory scale CSTR for multiphasic continuous-flow chemistry and long residence times.* Organic Process Research & Development, 2017. **21**(9): p. 1294-1301.
12. Guan, F., et al., *A universal reactor platform for batch and flow: application to homogeneous and heterogeneous hydrogenation.* Reaction Chemistry & Engineering, 2020. **5**(10): p. 1903-1908.
13. Müller, G., et al., *Continuous chemistry in microreaction systems for practical use.* Chimia International Journal for Chemistry, 2006. **60**(9): p. 618-622.
14. Pennemann, H., et al., *Benchmarking of microreactor applications.* Organic Process Research & Development, 2004. **8**(3): p. 422-439.
15. Schwalbe, T., V. Autze, and G. Wille, *Chemical synthesis in microreactors.* Chimia-Zurich, 2002. **56**(11): p. 636-646.
16. Fletcher, P.D., et al., *Micro reactors: principles and applications in organic synthesis.* Tetrahedron, 2002. **58**(24): p. 4735-4757.
17. Haswell, S.J. and P. Watts, *Green chemistry: synthesis in micro reactors.* Green Chemistry, 2003. **5**(2): p. 240-249.
18. Hessel, V. and H. Löwe, *Organic synthesis with microstructured reactors.* Chemical Engineering Technology: Industrial Chemistry-Plant Equipment-Process Engineering-Biotechnology, 2005. **28**(3): p. 267-284.
19. Wilson, N.G. and T. McCreedy, *On-chip catalysis using a lithographically fabricated glass microreactor—the dehydration of alcohols using sulfated zirconia.* Chemical Communications, 2000(9): p. 733-734.
20. Wiles, C., et al., *1, 4-Addition of enolates to α, β-unsaturated ketones within a micro reactor.* Lab on a Chip, 2002. **2**(2): p. 62-64.
21. Wegner, J., S. Ceylan, and A. Kirschning, *Flow chemistry–a key enabling technology for (multistep) organic synthesis.* Advanced Synthesis Catalysis, 2012. **354**(1): p. 17-57.
22. Nagaki, A., et al., *Lithiation of 1, 2-Dichloroethene in Flow Microreactors: Versatile Synthesis of Alkenes and Alkynes by Precise Residence-Time Control.* Angewandte Chemie International Edition, 2012. **51**(13): p. 3245-3248.
23. Mason, B.P., et al., *Greener approaches to organic synthesis using microreactor technology.* Chemical Reviews, 2007. **107**(6): p. 2300-2318.
24. Newman, S.G. and K.F. Jensen, *The role of flow in green chemistry and engineering.* Green Chemistry, 2013. **15**(6): p. 1456-1472.
25. Riva, E., et al., *Reaction of Grignard reagents with carbonyl compounds under continuous flow conditions.* Tetrahedron, 2010. **66**(17): p. 3242-3247.

26. Baumann, M. and I.R. Baxendale, *The synthesis of active pharmaceutical ingredients (APIs) using continuous flow chemistry.* Beilstein Journal of Organic Chemistry, 2015. **11**(1): p. 1194-1219.
27. Bogdan, A.R., et al., *The continuous-flow synthesis of ibuprofen.* Angewandte Chemie International Edition, 2009. **48**(45): p. 8547-8550.
28. Sheldon, R.A., *Catalysis: the key to waste minimization.* Journal of Chemical Technology & Biotechnology: International Research in Process, Environmental & Clean Technology, 1997. **68**(4): p. 381-388.
29. Kummer, A., T. Varga, and L. Nagy, *Semi-batch reactor control with NMPC avoiding thermal runaway.* Computers and Chemical Engineering, 2020. **134**: p. 106694.
30. Kao, C.-S. and K.-H. Hu, *Acrylic reactor runaway and explosion accident analysis.* Journal of Loss Prevention in the Process Industries, 2002. **15**(3): p. 213-222.
31. Fabiano, B., et al., *A perspective on Seveso accident based on cause-consequences analysis by three different methods.* Journal of Loss Prevention in the Process Industries, 2017. **49**: p. 18-35.
32. Malet-Sanz, L. and F. Susanne, *Continuous flow synthesis. A pharma perspective.* Journal of Medicinal Chemistry, 2012. **55**(9): p. 4062-4098.
33. Kunz, U. and T. Turek, *Flow through reactors for organic chemistry: directly electrically heated tubular mini reactors as an enabling technology for organic synthesis.* Beilstein Journal of Organic Chemistry, 2009. **5**(1): p. 70.
34. Kappe, C.O., *Controlled microwave heating in modern organic synthesis.* Angewandte Chemie International Edition, 2004. **43**(46): p. 6250-6284.
35. Yoshida, J.i., A. Nagaki, and T. Yamada, *Flash chemistry: fast chemical synthesis by using microreactors.* Chemistry - A European Journal, 2008. **14**(25): p. 7450-7459.
36. van der Linden, J.J., et al., *Investigation of the Moffatt− Swern oxidation in a continuous flow microreactor system.* Organic Process Research & Development, 2008. **12**(5): p. 911-920.
37. Cutler, A.H., M.J. Antal Jr, and M. Jones Jr, *A critical evaluation of the plug-flow idealization of tubular-flow reactor data.* Industrial & Engineering Chemistry Research, 1988. **27**(4): p. 691-697.
38. Lebl, R., et al., *Scalable continuous flow hydrogenations using Pd/Al2O3-coated rectangular cross-section 3D-printed static mixers.* Catalysis Today, 2020.
39. Jolley, K.E., M.R. Chapman, and A.J. Blacker, *A general and atom-efficient continuous-flow approach to prepare amines, amides and imines via reactive N-chloramines.* Beilstein Journal of Organic Chemistry, 2018. **14**(1): p. 2220-2228.
40. Hessel, V., *Novel process windows–gate to maximizing process intensification via flow chemistry.* Chemical Engineering Technology: Industrial Chemistry-Plant Equipment-Process Engineering-Biotechnology, 2009. **32**(11): p. 1655-1681.
41. Bogdan, A.R. and N.W. Sach, *The Use of Copper Flow Reactor Technology for the Continuous Synthesis of 1, 4-Disubstituted 1, 2, 3-Triazoles.* Advanced Synthesis Catalysis, 2009. **351**(6): p. 849-854.
42. Hornung, C.H., et al., *A microcapillary flow disc reactor for organic synthesis.* Organic Process Research & Development, 2007. **11**(3): p. 399-405.

43. Acke, D.R. and C.V. Stevens, *Study of the Baylis− Hillman reaction in a microreactor environment: first continuous production of Baylis− Hillman adducts.* Organic Process Research & Development, 2006. **10**(3): p. 417-422.
44. McQuade, D.T. and P.H. Seeberger, *Applying flow chemistry: Methods, materials, and multistep synthesis.* J. Org. Chem, 2013. **78**(13): p. 6384-6389.
45. Razzaq, T. and C.O. Kappe, *Continuous Flow Organic Synthesis under High-Temperature/Pressure Conditions.* Chemistry–An Asian Journal, 2010. **5**(6): p. 1274-1289.
46. Hartman, R.L., *Managing solids in microreactors for the upstream continuous processing of fine chemicals.* Organic Process Research & Development, 2012. **16**(5): p. 870-887.
47. Reichardt, C., *Solvent effects in organic chemistry*. 1979: Verlag Chemie.
48. Reichardt, C. and T. Welton, *Solvents and solvent effects in organic chemistry*. 2011: John Wiley & Sons.
49. Singh, D. and P.L. Dhepe, *An efficient catalytic transfer hydrogenation-hydrodeoxygenation of lignin derived monomers: Investigating catalyst properties-activity correlation.* Catalysis Communications, 2021. **149**: p. 106220.
50. Baxendale, I.R., et al., *Multistep synthesis using modular flow reactors: Bestmann–Ohira reagent for the formation of alkynes and triazoles.* Angewandte Chemie, 2009. **121**(22): p. 4077-4081.
51. Alam, M.P., et al., *C O bond formation in a microfluidic reactor: high yield S N Ar substitution of heteroaryl chlorides.* Tetrahedron letters, 2016. **57**(19): p. 2059-2062.
52. Gutmann, B., et al., *On the Fischer Indole Synthesis of 7-Ethyltryptophol□ Mechanistic and Process Intensification Studies under Continuous Flow Conditions.* Organic Process Research & Development, 2013. **17**(2): p. 294-302.
53. Marre, S., Y. Roig, and C. Aymonier, *Supercritical microfluidics: Opportunities in flow-through chemistry and materials science.* The Journal of Supercritical Fluids, 2012. **66**: p. 251-264.
54. Savage, P.E., *Organic chemical reactions in supercritical water.* Chemical reviews, 1999. **99**(2): p. 603-622.
55. Adschiri, T., et al., *Green materials synthesis with supercritical water.* Green Chemistry, 2011. **13**(6): p. 1380-1390.
56. Munshi, P. and S. Bhaduri, *Supercritical CO 2: a twenty-first century solvent for the chemical industry.* Current Science (00113891), 2009. **97**(1).
57. Eckert, C.A., et al., *Sustainable reactions in tunable solvents.* The Journal of Physical Chemistry B, 2004. **108**(47): p. 18108-18118.
58. O'Neil, A. and J.J. Watkins, *Fabrication of device nanostructures using supercritical fluids.* MRS bulletin, 2005. **30**(12): p. 967-975.
59. Romang, A.H. and J.J. Watkins, *Supercritical fluids for the fabrication of semiconductor devices: emerging or missed opportunities?* Chemical reviews, 2010. **110**(1): p. 459.
60. Marre, S., F. Cansell, and C. Aymonier, *Tailor-made surface properties of particles with a hydrophilic or hydrophobic polymer shell mediated by supercritical CO2.* Langmuir, 2008. **24**(1): p. 252.

61. Cansell, F. and C. Aymonier, *Design of functional nanostructured materials using supercritical fluids.* The Journal of Supercritical Fluids, 2009. **47**(3): p. 508-516.
62. Chamberlain, T.W., et al., *Catalytic nanoreactors in continuous flow: hydrogenation inside single-walled carbon nanotubes using supercritical CO 2.* Chemical Communications, 2014. **50**(40): p. 5200-5202.
63. Devetta, L., et al., *Kinetic experiments and modeling of a three-phase catalytic hydrogenation reaction in supercritical CO 2.* Catalysis Today, 1999. **48**(1): p. 337-345.
64. Han, X. and M. Poliakoff, *Continuous reactions in supercritical carbon dioxide: problems, solutions and possible ways forward.* Chemical Society Reviews, 2012. **41**(4): p. 1428-1436.
65. Licence, P., et al., *Chemical reactions in supercritical carbon dioxide: from laboratory to commercial plant.* Green Chemistry, 2003. **5**(2): p. 99-104.
66. Leitner, W., *Supercritical carbon dioxide as a green reaction medium for catalysis.* Accounts of chemical research, 2002. **35**(9): p. 746-756.
67. McHugh, M. and V. Krukonis, *Supercritical fluid extraction: principles and practice*. 2013: Elsevier.
68. Hitzler, M.G., et al., *Selective catalytic hydrogenation of organic compounds in supercritical fluids as a continuous process.* Organic Process Research & Development, 1998. **2**(3): p. 137-146.
69. Cason, J.P. and C.B. Roberts, *Metallic copper nanoparticle synthesis in AOT reverse micelles in compressed propane and supercritical ethane solutions.* The Journal of Physical Chemistry B, 2000. **104**(6): p. 1217-1221.
70. Hyde, J.R., et al., *Continuous catalytic reactions in supercritical fluids.* Applied Catalysis A: General, 2001. **222**(1): p. 119-131.
71. Hoffmann, N., *Photochemical reactions of aromatic compounds and the concept of the photon as a traceless reagent.* Photochemical & Photobiological Sciences, 2012. **11**(11): p. 1613-1641.
72. Hoffmann, N., *Photochemical reactions as key steps in organic synthesis.* Chemical Reviews, 2008. **108**(3): p. 1052.
73. Knowles, J.P., L.D. Elliott, and K.I. Booker-Milburn, *Flow photochemistry: Old light through new windows.* Beilstein journal of organic chemistry, 2012. **8**: p. 2025.
74. Hook, B.D., et al., *A practical flow reactor for continuous organic photochemistry.* The Journal of organic chemistry, 2005. **70**(19): p. 7558-7564.
75. Schuster, E.M. and P. Wipf, *Photochemical flow reactions.* Israel Journal of Chemistry, 2014. **54**(4): p. 361-370.
76. Yoshida, J.-i., et al., *Modern strategies in electroorganic synthesis.* Chemical reviews, 2008. **108**(7): p. 2265-2299.
77. Watts, K., A. Baker, and T. Wirth, *Electrochemical synthesis in microreactors.* Journal of Flow Chemistry, 2014. **4**(1): p. 2-11.
78. Horii, D., T. Fuchigami, and M. Atobe, *A new approach to anodic substitution reaction using parallel laminar flow in a micro-flow reactor.* Journal of the American Chemical Society, 2007. **129**(38): p. 11692-11693.
79. Horii, D., et al., *A novel electrosynthetic system for anodic substitution reactions by using parallel laminar flow in a microflow reactor.* Chemistry-A European Journal, 2008. **14**(33): p. 10382-10387.

80. ZNIDARŠIC-PLAZL, P., *Enzymatic microreactors utilizing non-aqueous media.* Chimica Oggi-Chemistry Today, 2014. **32**: p. 1.
81. Rao, N.N., et al., *Continuous biocatalytic processes.* Organic Process Research & Development, 2009. **13**(3): p. 607-616.
82. Babich, L., et al., *Continuous-Flow Reactor-Based Enzymatic Synthesis of Phosphorylated Compounds on a Large Scale.* Chemistry-A European Journal, 2012. **18**(21): p. 6604-6609.
83. Jas, G. and A. Kirschning, *Continuous flow techniques in organic synthesis.* Chemistry-A European Journal, 2003. **9**(23): p. 5708-5723.
84. Chen, P., et al., *A practical high through-put continuous process for the synthesis of chiral cyanohydrins.* The Journal of organic chemistry, 2002. **67**(23): p. 8251-8253.
85. Liu, Z., et al., *Combined biosynthetic pathway for de novo production of UDP-galactose: catalysis with multiple enzymes immobilized on agarose beads.* ChemBioChem, 2002. **3**(4): p. 348-355.
86. Kobayashi, S., H. Ishitani, and S. Nagayama, *Lanthanide triflate catalyzed imino Diels-Alder reactions; convenient syntheses of pyridine and quinoline derivatives.* Synthesis, 1995. **1995**(09): p. 1195-1202.
87. Usutani, H., et al., *Generation and reactions of o-bromophenyllithium without benzyne formation using a microreactor.* Journal of the American Chemical Society, 2007. **129**(11): p. 3046-3047.
88. Britton, J. and C.L. Raston, *Multi-step continuous-flow synthesis.* Chemical Society Reviews, 2017. **46**(5): p. 1250-1271.
89. Snead, D.R. and T.F. Jamison, *A three-minute synthesis and purification of ibuprofen: Pushing the limits of continuous-flow processing.* Angewandte Chemie International Edition, 2015. **54**(3): p. 983-987.
90. Sahoo, H.R., J.G. Kralj, and K.F. Jensen, *Multistep Continuous-Flow Microchemical Synthesis Involving Multiple Reactions and Separations.* Angewandte Chemie, 2007. **119**(30): p. 5806-5810.
91. Hartman, R.L., et al., *Multistep microchemical synthesis enabled by microfluidic distillation.* Angewandte Chemie International Edition, 2010. **49**(5): p. 899-903.
92. Glöckner, S., et al., *The rapid synthesis of oxazolines and their heterogeneous oxidation to oxazoles under flow conditions.* Organic & biomolecular chemistry, 2015. **13**(1): p. 207-214.
93. Roberge, D.M., et al., *Microreactor technology: a revolution for the fine chemical and pharmaceutical industries?* Chemical engineering & technology, 2005. **28**(3): p. 318-323.
94. Exarchou, V., et al., *LC-UV-solid-phase extraction-NMR-MS combined with a cryogenic flow probe and its application to the identification of compounds present in Greek oregano.* Analytical Chemistry, 2003. **75**(22): p. 6288-6294.
95. Benito-Lopez, F., et al., *Optical fiber-based on-line UV/Vis spectroscopic monitoring of chemical reaction kinetics under high pressure in a capillary microreactor.* Chemical Communications, 2005(22): p. 2857-2859.
96. Mozharov, S., et al., *Improved method for kinetic studies in microreactors using flow manipulation and noninvasive Raman spectrometry.* Journal of the American Chemical Society, 2011. **133**(10): p. 3601-3608.
97. Song, H. and R.F. Ismagilov, *Millisecond kinetics on a microfluidic chip using nanoliters of reagents.* Journal of the American Chemical Society, 2003. **125**(47): p. 14613-14619.

98. Martinuzzi, I., et al., *Reaction mechanism for glycerol dehydration in the gas phase over a solid acid catalyst determined with on-line gas chromatography.* Chemical Engineering Science, 2014. **116**: p. 118-127.
99. Moret, S., et al., *Microwave assisted saponification (MAS) followed by on-line liquid chromatography (LC)–gas chromatography (GC) for high-throughput and high-sensitivity determination of mineral oil in different cereal-based foodstuffs.* Food chemistry, 2016. **196**: p. 50-57.
100. Grunwaldt, J.-D. and B.S. Clausen, *Combining XRD and EXAFS with on-line catalytic studies for in situ characterization of catalysts.* Topics in Catalysis, 2002. **18**(1): p. 37-43.
101. De Beer, T., et al., *In-line and real-time process monitoring of a freeze drying process using Raman and NIR spectroscopy as complementary process analytical technology (PAT) tools.* Journal of pharmaceutical sciences, 2009. **98**(9): p. 3430-3446.
102. Ward, H. and F.E. Sistare, *On-line determination and control of the water content in a continuous conversion reactor using NIR spectroscopy.* Analytica chimica acta, 2007. **595**(1): p. 319-322.
103. Cervera-Padrell, A.E., et al., *Monitoring and control of a continuous Grignard reaction for the synthesis of an active pharmaceutical ingredient intermediate using inline NIR spectroscopy.* Organic Process Research & Development, 2012. **16**(5): p. 901-914.
104. Bristow, T.W., et al., *On-line monitoring of continuous flow chemical synthesis using a portable, small footprint mass spectrometer.* Journal of The American Society for Mass Spectrometry, 2014. **25**(10): p. 1794-1802.
105. Roscioli, K.M., et al., *Real time pharmaceutical reaction monitoring by electrospray ion mobility-mass spectrometry.* International Journal of Mass Spectrometry, 2013. **336**: p. 27-36.
106. Mathieson, J.S., et al., *Continuous parallel ESI-MS analysis of reactions carried out in a bespoke 3D printed device.* Beilstein journal of nanotechnology, 2013. **4**: p. 285.
107. Robert, J. and A. JohnáBlacker, *Online quantitative mass spectrometry for the rapid adaptive optimisation of automated flow reactors.* Reaction Chemistry & Engineering, 2016. **1**(1): p. 96-100.
108. Browne, D.L., et al., *Continuous flow reaction monitoring using an on-line miniature mass spectrometer.* Rapid Communications in Mass Spectrometry, 2012. **26**(17): p. 1999-2010.
109. Zhang, J., et al., *Design and scaling up of microchemical systems: a review.* Annual review of chemical and biomolecular engineering, 2017. **8**: p. 285-305.
110. Rasheed, M. and T. Wirth, *Fast optimisation of a Diels-Alder reaction in a flow microreactor by in-line HPLC analysis.* Chemistry Today, 2011. **29**(3): p. 56.
111. Welch, C.J., et al., *Online analysis of flowing streams using microflow HPLC.* Organic Process Research & Development, 2009. **13**(5): p. 1022-1025.
112. Carter, C.F., et al., *ReactIR flow cell: a new analytical tool for continuous flow chemical processing.* Organic Process Research & Development, 2010. **14**(2): p. 393-404.
113. Malet-Sanz, L., et al., *Preparation of arylsulfonyl chlorides by chlorosulfonylation of in situ generated diazonium salts using a*

continuous flow reactor. Organic & biomolecular chemistry, 2010. **8**(23): p. 5324-5332.
114. Carter, C.F., et al., *The continuous flow synthesis of butane-2, 3-diacetal protected building blocks using microreactors.* Organic & biomolecular chemistry, 2010. **8**(7): p. 1588-1595.
115. Smith, C.J., et al., *A fully automated, multistep flow synthesis of 5-amino-4-cyano-1, 2, 3-triazoles.* Organic & biomolecular chemistry, 2011. **9**(6): p. 1938-1947.
116. Floyd, T.M., M.A. Schmidt, and K.F. Jensen, *Silicon micromixers with infrared detection for studies of liquid-phase reactions.* Industrial & engineering chemistry research, 2005. **44**(8): p. 2351-2358.
117. Carter, C.F., et al., *Synthesis of acetal protected building blocks using flow chemistry with flow IR analysis: preparation of butane-2, 3-diacetal tartrates.* Organic & biomolecular chemistry, 2009. **7**(22): p. 4594-4597.
118. Sans, V., et al., *A self optimizing synthetic organic reactor system using real-time in-line NMR spectroscopy.* Chemical science, 2015. **6**(2): p. 1258-1264.
119. Maiwald, M., et al., *Quantitative high-resolution on-line NMR spectroscopy in reaction and process monitoring.* Journal of Magnetic Resonance, 2004. **166**(2): p. 135-146.
120. Maiwald, M., et al., *Quantitative on-line high-resolution NMR spectroscopy in process engineering applications.* Analytical and bioanalytical chemistry, 2003. **375**(8): p. 1111-1115.
121. Dalitz, F., et al., *Process and reaction monitoring by low-field NMR spectroscopy.* Progress in nuclear magnetic resonance spectroscopy, 2012. **60**: p. 52-70.
122. Silva Elipe, M.V. and R.R. Milburn, *Monitoring chemical reactions by low-field benchtop NMR at 45 MHz: pros and cons.* Magnetic Resonance in Chemistry, 2016. **54**(6): p. 437-443.
123. Goldbach, M., et al., *Preparation of Grignard reagents from magnesium metal under continuous flow conditions and on-line monitoring by NMR spectroscopy.* Tetrahedron Letters, 2016. **57**(1): p. 122-125.
124. Houben, C. and A.A. Lapkin, *Automatic discovery and optimization of chemical processes.* Current Opinion in Chemical Engineering, 2015. **9**: p. 1-7.
125. Cox, B.G., *Modern liquid phase kinetics*. 1994: Oxford University Press.
126. Perrin, C.L., *Linear or nonlinear least-squares analysis of kinetic data?* Journal of Chemical Education, 2017. **94**(6): p. 669-672.
127. Michaelis, L. and M.L. Menten, *Die kinetik der invertinwirkung*. 2007: Universitätsbibliothek Johann Christian Senckenberg.
128. Johnson, K.A., *A century of enzyme kinetic analysis, 1913 to 2013.* FEBS Letters, 2013. **587**(17): p. 2753-2766.
129. Lineweaver, H. and D. Burk, *The determination of enzyme dissociation constants.* Journal of the American Chemical Society, 1934. **56**(3): p. 658-666.
130. Nielsen, C.D.-T. and J. Burés, *Visual kinetic analysis.* Chemical Science, 2019. **10**(2): p. 348-353.
131. Blackmond, D.G., *Reaction progress kinetic analysis: a powerful methodology for mechanistic studies of complex catalytic reactions.* Angewandte Chemie International Edition, 2005. **44**(28): p. 4302-4320.

132. Rosner, T., et al., *Kinetic studies of Heck coupling reactions using palladacycle catalysts: experimental and kinetic modeling of the role of dimer species.* Journal of the American Chemical Society, 2001. **123**(9): p. 1848-1855.
133. Nielsen, L.P., et al., *Mechanistic investigation leads to a synthetic improvement in the hydrolytic kinetic resolution of terminal epoxides.* Journal of the American Chemical Society, 2004. **126**(5): p. 1360-1362.
134. Baxter, R.D., et al., *Mechanistic rationalization of unusual kinetics in Pd-catalyzed C–H olefination.* Journal of the American Chemical Society, 2012. **134**(10): p. 4600-4606.
135. Blackmond, D.G., *Kinetic profiling of catalytic organic reactions as a mechanistic tool.* Journal of the American Chemical Society, 2015. **137**(34): p. 10852-10866.
136. Burés, J., *A simple graphical method to determine the order in catalyst.* Angewandte Chemie International Edition, 2016. **55**(6): p. 2028-2031.
137. Burés, J., *Variable time normalization analysis: General graphical elucidation of reaction orders from concentration profiles.* Angewandte Chemie International Edition, 2016. **55**(52): p. 16084-16087.
138. Mathew, J.S., et al., *Investigations of Pd-catalyzed ArX coupling reactions informed by reaction progress kinetic analysis.* The Journal of Organic Chemistry, 2006. **71**(13): p. 4711-4722.
139. Streuff, J., et al., *Mechanism of the tiIII-catalyzed acyloin-type umpolung: a catalyst-controlled radical reaction.* Journal of the American Chemical Society, 2015. **137**(45): p. 14396-14405.
140. Lin, S. and E.N. Jacobsen, *Thiourea-catalysed ring opening of episulfonium ions with indole derivatives by means of stabilizing non-covalent interactions.* Nature Chemistry, 2012. **4**(10): p. 817.
141. Noda, H., et al., *Catalytic asymmetric synthesis of CF 3-substituted tertiary propargylic alcohols via direct aldol reaction of α-N 3 amide.* Chemical Science, 2017. **8**(4): p. 3260-3269.
142. Lundberg, H., et al., *Mechanistic elucidation of zirconium-catalyzed direct amidation.* Journal of the American Chemical Society, 2017. **139**(6): p. 2286-2295.
143. Vasilenko, V., C.K. Blasius, and L.H. Gade, *One-Pot Sequential Kinetic Profiling of a Highly Reactive Manganese Catalyst for Ketone Hydroboration: Leveraging σ-Bond Metathesis via Alkoxide Exchange Steps.* Journal of the American Chemical Society, 2018. **140**(29): p. 9244-9254.
144. Franceschini, G. and S. Macchietto, *Model-based design of experiments for parameter precision: State of the art.* Chemical Engineering Science, 2008. **63**(19): p. 4846-4872.
145. Galvanin, F., S. Macchietto, and F. Bezzo, *Model-based design of parallel experiments.* Industrial & Engineering Chemistry Research, 2007. **46**(3): p. 871-882.
146. Waldron, C., et al., *Model-based design of transient flow experiments for the identification of kinetic parameters.* Reaction Chemistry & Engineering, 2020. **5**(1): p. 112-123.
147. Von Stosch, M., et al., *Hybrid semi-parametric modeling in process systems engineering: Past, present and future.* Computers and Chemical Engineering, 2014. **60**: p. 86-101.

148. de Juan, A., et al., *Combining hard-and soft-modelling to solve kinetic problems.* Chemometrics Intelligent Laboratory Systems, 2000. **54**(2): p. 123-141.
149. de Carvalho, A.R., et al., *Comparison of PLS and kinetic models for a second-order reaction as monitored using ultraviolet visible and mid-infrared spectroscopy.* Talanta, 2006. **68**(4): p. 1190-1200.
150. Fath, V., et al., *Efficient Kinetic Data Acquisition and Model Prediction: Continuous Flow Microreactors, Inline FT-IR Spectroscopy, and Self-modeling Curve Resolution.* Organic Process Research & Development, 2020.
151. Toch, K., J.W. Thybaut, and G.B. Marin, *A systematic methodology for kinetic modeling of chemical reactions applied to n-hexane hydroisomerization.* AIChE Journal, 2015. **61**(3): p. 880-892.
152. Yang, S.-T., X. Liu, and Y. Zhang, *Metabolic engineering–applications, methods, and challenges*, in *Bioprocessing for Value-Added Products from Renewable Resources*. 2007, Elsevier. p. 73-118.
153. Vasquez, E. and T. Eldredge, *Process modeling for hydrocarbon fuel conversion*, in *Advances in Clean Hydrocarbon Fuel Processing*. 2011, Elsevier. p. 509-545.
154. Paul, S., V. Kurin, and S. Whiteson. *Fast Efficient Hyperparameter Tuning for Policy Gradient Methods*. in *Advances in Neural Information Processing Systems*. 2019.
155. Wistuba, M., N. Schilling, and L. Schmidt-Thieme, *Scalable gaussian process-based transfer surrogates for hyperparameter optimization.* Machine Learning, 2018. **107**(1): p. 43-78.
156. Lerman, P., *Fitting segmented regression models by grid search.* Journal of the Royal Statistical Society: Series C, 1980. **29**(1): p. 77-84.
157. Huang, C.-Y.R., C.-Y. Lai, and K.-T.T. Cheng, *Fundamentals of algorithms*, in *Electronic Design Automation*. 2009, Elsevier. p. 173-234.
158. Boyd, S.P. and L. Vandenberghe, *Convex optimization*. 2004, Cambridge university press. p. 1-16.
159. Vasiljević, D., *Comparison of optimization algorithms*, in *Classical and Evolutionary Algorithms in the Optimization of Optical Systems*. 2002, Springer. p. 83-88.
160. Inuiguchi, M. and J. Ramık, *Possibilistic linear programming: a brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem.* Fuzzy Sets & Systems, 2000. **111**(1): p. 3-28.
161. Bénichou, M., et al., *Experiments in mixed-integer linear programming.* Mathematical Programming, 1971. **1**(1): p. 76-94.
162. Kuhn, H.W. and A.W. Tucker, *Nonlinear programming*, in *Traces and emergence of nonlinear programming*. 2014, Springer. p. 247-258.
163. Bregman, L.M., *The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming.* USSR Computational Mathetmatics and Mathematical Physics, 1967. **7**(3): p. 200-217.
164. Frank, M. and P. Wolfe, *An algorithm for quadratic programming.* Naval Research Logistics Quarterly, 1956. **3**(1-2): p. 95-110.
165. Drummond, L.G. and B.F. Svaiter, *A steepest descent method for vector optimization.* Journal of Computational & Applied Mathematics, 2005. **175**(2): p. 395-414.

166. Clayton, A.D., et al., *Algorithms for the self-optimisation of chemical reactions.* Reaction Chemistry & Engineering, 2019. **4**(9): p. 1545-1554.
167. Spendley, W., G.R. Hext, and F.R. Himsworth, *Sequential application of simplex designs in optimisation and evolutionary operation.* Technometrics, 1962. **4**(4): p. 441-461.
168. Nelder, J.A. and R. Mead, *A simplex method for function minimization.* The Computer Journal, 1965. **7**(4): p. 308-313.
169. Byrd, R.H., J.C. Gilbert, and J. Nocedal, *A trust region method based on interior point techniques for nonlinear programming.* Mathematical Programming, 2000. **89**(1): p. 149-185.
170. Waltz, R.A., et al., *An interior algorithm for nonlinear optimization that combines line search and trust region steps.* Mathematical Programming, 2006. **107**(3): p. 391-408.
171. Whitley, D., *A genetic algorithm tutorial.* Statistics & Computing, 1994. **4**(2): p. 65-85.
172. Brochu, E., V.M. Cora, and N. De Freitas, *A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning.* preprint, arXiv:1012.2599, 2010.
173. Tsu, J., V.H.G. Díaz, and M.J. Willis, *Computational approaches to kinetic model selection.* Computers and Chemical Engineering, 2019. **121**: p. 618-632.
174. de la Mare, P.B.D. *Reaction Mechanism*. Encyclopædia Britannica 2008 [cited 2020 April 13]; Available from: https://www.britannica.com/science/reaction-mechanism.
175. Carey, F.A. and R.J. Sundberg, *Study and Description of Organic Reaction Mechanisms.* Advanced Organic Chemistry: Part A: Structure & Mechanisms, 2000: p. 187-261.
176. Willis, M.J. and M. von Stosch, *Inference of chemical reaction networks using mixed integer linear programming.* Computers and Chemical Engineering, 2016. **90**: p. 31-43.
177. Aris, R. and R. Mah, *Independence of chemical reactions.* Industrial Engineering Chemistry Fundamentals, 1963. **2**(2): p. 90-94.
178. Bonvin, D. and D. Rippin, *Target factor analysis for the identification of stoichiometric models.* Chemical Engineering Science, 1990. **45**(12): p. 3417-3426.
179. Amrhein, M., B. Srinivasan, and D. Bonvin, *Target factor analysis of reaction data: Use of data pre-treatment and reaction-invariant relationships.* Chemical Engineering Science, 1999. **54**(5): p. 579-591.
180. Georgakis, C. and R. Lin. *Stoichiometric Modeling of Complex Pharmaceutical Reactions*. in *Proceedings of the Annual AIChE meeting in Cincinnati, OH, November*. 2005.
181. Brendel, M., D. Bonvin, and W. Marquardt, *Incremental identification of kinetic models for homogeneous reaction systems.* Chemical Engineering Science, 2006. **61**(16): p. 5404-5420.
182. Bhatt, N., et al., *Incremental identification of reaction systems—A comparison between rate-based and extent-based approaches.* Chemical Engineering Science, 2012. **83**: p. 24-38.
183. Burnham, S.C., et al., *Inference of chemical reaction networks.* Chemical Engineering Science, 2008. **63**(4): p. 862-873.

184. Crampin, E.J., S. Schnell, and P.E. McSharry, *Mathematical and computational techniques to deduce complex biochemical reaction mechanisms.* Progress in Biophysics & Molecular Biology, 2004. **86**(1): p. 77-112.
185. Searson, D.P., et al., *Inference of chemical reaction networks using hybrid s-system models.* Chemical Product & Process Modelling, 2007. **2**(1).
186. Srividhya, J., et al., *Reconstructing biochemical pathways from time course data.* Proteomics, 2007. **7**(6): p. 828-838.
187. Searson, D.P., M.J. Willis, and A. Wright, *Reverse Engineering Chemical Reaction Networks from Time Series Data.* Statistical Modelling of Molecular Descriptors in QSAR/QSPR, 2014. **2**(12).
188. Koza, J.R., et al., *Automatic computational discovery of chemical reaction networks using genetic programming*, in *Computational Discovery of Scientific Knowledge*. 2007, Springer. p. 205-227.
189. Hii, C.J., A.R. Wright, and M.J. Willis, *Utilizing a genetic algorithm to elucidate chemical reaction networks: An experimental case study.* International Journal of Chemical Engineering Applications, 2014. **5**(6): p. 516.
190. Chellaboina, V., et al., *Modeling and analysis of mass-action kinetics.* IEEE Control Systems Magazine, 2009. **29**(4): p. 60-78.
191. Sayikli, C. and E.Z. Bagci. *Limitations of using mass action kinetics method in modeling biochemical systems: illustration for a second order reaction*. in *International Conference on Computational Science and Its Applications*. 2011. Springer.
192. Furusjö, E., O. Svensson, and L.-G. Danielsson, *Estimation of kinetic parameters from non-isothermal batch experiments monitored by in situ vibrational spectroscopy.* Chemometrics Intelligent Laboratory Systems, 2003. **66**(1): p. 1-14.
193. Pagano, R.L., et al., *Cure kinetic parameter estimation of thermosetting resins with isothermal data by using particle swarm optimization.* European Polymer Journal, 2008. **44**(8): p. 2678-2686.
194. Zhang, Y.-Y. and J.-H. Liu, *Kinetic study of enantioselective hydrolysis of (R, S)-ketoprofen ethyl ester using immobilized T. laibacchii lipase.* Biochemical Engineering Journal, 2011. **54**(1): p. 40-46.
195. VanBuren, V., L. Cassimeris, and D.J. Odde, *Mechanochemical model of microtubule structure and self-assembly kinetics.* Biophysical Journal, 2005. **89**(5): p. 2911-2926.
196. Gao, D., Y. Xiao, and A. Varma, *Guaiacol hydrodeoxygenation over platinum catalyst: reaction pathways and kinetics.* Industrial Engineering Chemistry Research, 2015. **54**(43): p. 10638-10644.
197. Changi, S., T.M. Brown, and P.E. Savage, *Reaction kinetics and pathways for phytol in high-temperature water.* Chemical Engineering Journal, 2012. **189**: p. 336-345.
198. Zhang, Y., et al., *One-pot Enzymatic Synthesis of Enantiopure 1, 3-oxathiolane by Trichosporon Laibachii Lipase and Kinetic Model.* Organic Process Research & Development, 2020.
199. Villaverde, A.F., et al., *Benchmarking optimization methods for parameter estimation in large kinetic models.* Bioinformatics, 2019. **35**(5): p. 830-838.
200. Akaike, H., *Fitting autoregressive models for prediction.* Annals of the Institute of Statistical Mathematics, 1969. **21**(1): p. 243-247.

201. Hurvich, C.M. and C.-L. Tsai, *Regression and time series model selection in small samples.* Biometrika, 1989. **76**(2): p. 297-307.
202. Burnham, K.P. and D.R. Anderson, *A practical information-theoretic approach*. Model Selection and Multimodel Inference. Vol. 2, 49-96. 2002, New York: Springer.
203. Gholamipour-Shirazi, A. and C. Rolando, *Alkylation of substituted benzoic acids in a continuous flow microfluidic microreactor: kinetics and linear free energy relationships.* Organic Process Research & Development, 2012. **16**(5): p. 811-818.
204. Niemeier, J.K., et al., *Application of kinetic modeling and competitive solvent hydrolysis in the development of a highly selective hydrolysis of a nitrile to an amide.* Organic Process Research & Development, 2014. **18**(3): p. 410-416.
205. Hartridge, H. and F.J.W. Roughton, *A method of measuring the velocity of very rapid chemical reactions.* Proceedings of the Royal Society of London, Series A, Containing Papers of a Mathematical Physical Character, 1923. **104**(726): p. 376-394.
206. Gibson, Q.H., *Rapid mixing: Stopped flow.* Methods in Enzymology, 1969. **16**: p. 187-228.
207. Seong, G.H., J. Heo, and R.M. Crooks, *Measurement of enzyme kinetics using a continuous-flow microfluidic system.* Analytical Chemistry, 2003. **75**(13): p. 3161-3167.
208. Rozich, A.F. and D.J. Castens, *Inhibition kinetics of nitrification in continuous-flow reactors.* Water Pollution Control Federation, 1986: p. 220-226.
209. Christensen, M., et al., *Development of an automated kinetic profiling system with online HPLC for reaction optimization.* Reaction Chemistry & Engineering, 2019. **4**(9): p. 1555-1558.
210. Johnson, K.A., *Rapid quench kinetic analysis of polymerases, adenosinetriphosphatases, and enzyme intermediates.* Contemporary Enzyme Kinetics Mechanism: Reliable Lab Solutions, 2009: p. 407.
211. Probstein, R.F., *Physicochemical hydrodynamics: an introduction*. 2005: John Wiley & Sons.
212. Konermann, L., *Monitoring reaction kinetics in solution by continuous-flow methods: the effects of convection and molecular diffusion under laminar flow conditions.* Journal of Physical Chemistry A, 1999. **103**(36): p. 7210-7216.
213. Konermann, L., et al., *Acid-induced denaturation of myoglobin studied by time-resolved electrospray ionization mass spectrometry.* Biochemistry, 1997. **36**(21): p. 6448-6454.
214. Konermann, L., B. Collings, and D. Douglas, *Cytochrome c folding kinetics studied by time-resolved electrospray ionization mass spectrometry.* Biochemistry, 1997. **36**(18): p. 5554-5559.
215. Zechel, D.L., et al., *Pre-steady state kinetic analysis of an enzymatic reaction monitored by time-resolved electrospray ionization mass spectrometry.* Biochemistry, 1998. **37**(21): p. 7664-7669.
216. Zhou, X., R. Medhekar, and M.D. Toney, *A continuous-flow system for high-precision kinetics using small volumes.* Analytical Chemistry, 2003. **75**(15): p. 3681-3687.
217. Jensen, K.F., B.J. Reizman, and S.G. Newman, *Tools for chemical synthesis in microsystems.* Lab on a Chip, 2014. **14**(17): p. 3206-3212.

218. Hartman, R.L., J.P. McMullen, and K.F. Jensen, *Deciding whether to go with the flow: evaluating the merits of flow reactors for synthesis.* Angewandte Chemie International Edition, 2011. **50**(33): p. 7502-7519.
219. Baxendale, I.R., L. Brocken, and C.J. Mallia, *Flow chemistry approaches directed at improving chemical synthesis.* Green Processing Synthesis, 2013. **2**(3): p. 211-230.
220. Gutmann, B., D. Cantillo, and C.O. Kappe, *Continuous-flow technology—a tool for the safe manufacturing of active pharmaceutical ingredients.* Angewandte Chemie International Edition, 2015. **54**(23): p. 6688-6728.
221. Bedore, M.W., et al., *Aminolysis of epoxides in a microreactor system: a continuous flow approach to β-amino alcohols.* Organic Process Research & Development, 2010. **14**(2): p. 432-440.
222. Zaborenko, N., et al., *Kinetic and scale-up investigations of epoxide aminolysis in microreactors at high temperatures and pressures.* Organic Process Research & Development, 2011. **15**(1): p. 131-139.
223. Monteiro, J.L., et al., *Continuous Flow Homolytic Aromatic Substitution with Electrophilic Radicals: A Fast and Scalable Protocol for Trifluoromethylation.* Chemistry - A European Journal, 2017. **23**(1): p. 176-186.
224. Nagaki, A., D. Ichinari, and J.-i. Yoshida, *Three-component coupling based on flash chemistry. Carbolithiation of benzyne with functionalized aryllithiums followed by reactions with electrophiles.* Journal of the American Chemical Society, 2014. **136**(35): p. 12245-12248.
225. Hone, C.A., et al., *Rapid multistep kinetic model generation from transient flow data.* Reaction Chemistry & Engineering, 2017. **2**(2): p. 103-108.
226. Valera, F.E., et al., *The Flow's the Thing… Or Is It? Assessing the Merits of Homogeneous Reactions in Flask and Flow.* Angewandte Chemie International Edition, 2010. **49**(14): p. 2478-2485.
227. Moore, J.S. and K.F. Jensen, *"Batch" kinetics in flow: online IR analysis and continuous control.* Angewandte Chemie International Edition, 2014. **53**(2): p. 470-473.
228. Haas, C.P., et al., *Automated generation of photochemical reaction data by transient flow experiments coupled with online HPLC analysis.* Reaction Chemistry & Engineering, 2020.
229. Schaber, S.D., et al., *Design, execution, and analysis of time-varying experiments for model discrimination and parameter estimation in microreactors.* Organic Process Research & Development, 2014. **18**(11): p. 1461-1467.
230. Aroh, K.C. and K.F. Jensen, *Efficient kinetic experiments in continuous flow microreactors.* Reaction Chemistry & Engineering, 2018. **3**(1): p. 94-101.
231. Gomez, M.V., et al., *Determination of kinetic parameters within a single nonisothermal on-flow experiment by nanoliter NMR spectroscopy.* Analytical Chemistry, 2015. **87**(20): p. 10547-10555.
232. Waldron, C., et al., *An autonomous microreactor platform for the rapid identification of kinetic models.* Reaction Chemistry & Engineering, 2019. **4**(9): p. 1623-1636.
233. Hone, C.A., et al., *Definitive screening designs for multistep kinetic models in flow.* Reaction Chemistry & Engineering, 2019. **4**(9): p. 1565-1570.

234. Holmes, N., et al., *Online quantitative mass spectrometry for the rapid adaptive optimisation of automated flow reactors.* Reaction Chemistry & Engineering, 2016. **1**(1): p. 96-100.
235. Levenspiel, O., *Chemical reaction engineering.* Industrial & Engineering Chemistry Research, 1999. **38**(11): p. 4140-4143.
236. McMullen, J.P. and K.F. Jensen, *Rapid determination of reaction kinetics with an automated microfluidic system.* Organic Process Research & Development, 2011. **15**(2): p. 398-407.
237. Nagy, K.D., et al., *Mixing and dispersion in small-scale flow systems.* Organic Process Research & Development, 2012. **16**(5): p. 976-981.
238. Afonso, C.A., et al., *Comprehensive Organic Chemistry Experiments for the Laboratory Classroom*. 2016: Royal Society of Chemistry.
239. Fluorochem. 2019 27/06/2019]; Available from: http://www.fluorochem.co.uk/Products/Product?code=303261.
240. Ashworth, I.W., et al., *Where has my acid gone? Understanding the self-catalyzed esterification of maleic acid in methanol during salt formation.* Organic Process Research & Development, 2012. **16**(10): p. 1646-1651.
241. Gilbert, J.C., E.G. McKinley, and D.-R. Hou, *The nature of cyclopentyne from different precursors.* Tetrahedron, 1997. **53**(29): p. 9891-9902.
242. Takaki, K., et al., *Intermolecular hydrophosphination of alkynes and related carbon− carbon multiple bonds catalyzed by organoytterbiums.* The Journal of Organic Chemistry, 2003. **68**(17): p. 6554-6565.
243. Berthelot, D.J.-C., et al., *Heterocyclic amides as modulators of TRPA1*. 2013, Google Patents.
244. Binch, H., et al., *Aminopyridine derivatives having aurora a selective inhibitory action*. 2013, Google Patents.
245. Lood, C.S., et al., *Synthesis of Chiral (Indol-2-yl) methanamines and Insight into the Stereochemistry Protecting Effects of the 9-Phenyl-9-fluorenyl Protecting Group.* European Journal of Organic Chemistry, 2015. **3793**: p. 3805.
246. Poisson, J.-F., A. Orellana, and A.E. Greene, *Stereocontrolled synthesis of (−)-kainic acid from trans-4-hydroxy-L-proline.* The Journal of Organic Chemistry, 2005. **70**(26): p. 10860-10863.
247. Lood, C., M. Nieger, and A. Koskinen, *Enantiospecific gram scale synthesis of (S)-eleagnine.* Tetrahedron, 2015. **71**(30): p. 5019-5024.
248. *9-Bromo-9-phenylfluorene, Sigma Aldrich, Available from: https://www.sigmaaldrich.com/catalog/product/aldrich/368873*. [27/09/2020].
249. Tian, Z. and F. Menard, *Synthesis of Kainoids and C4 Derivatives.* The Journal of Organic Chemistry, 2018. **83**(11): p. 6162-6170.
250. *9-Fluorenone, Sigma Aldrich, Available from: https://www.sigmaaldrich.com/catalog/product/aldrich/f1506*. [27/09/2020].