



Contents lists available at ScienceDirect

Chemical Engineering Journal

journal homepage: www.elsevier.com/locate/cej

Rapid, automated determination of reaction models and kinetic parameters

Connor J. Taylor^a, Megan Booth^a, Jamie A. Manson^a, Mark J. Willis^b, Graeme Clemens^c,
 Brian A. Taylor^c, Thomas W. Chamberlain^a, Richard A. Bourne^{a,*}

^a Institute of Process Research and Development, School of Chemistry and School of Chemical and Process Engineering, University of Leeds, Leeds LS2 9JT, UK

^b School of Engineering, University of Newcastle, Newcastle upon Tyne NE1 7RU, UK

^c Chemical Development, Pharmaceutical Technology & Development, AstraZeneca, Macclesfield Campus, SK10 2NA, UK

ARTICLE INFO

Keywords:

Flow chemistry
 Kinetics
 Optimisation
 Reaction model
 Simulation
 Machine learning

ABSTRACT

We herein report a novel kinetic modelling methodology whereby identification of the correct reaction model and kinetic parameters is conducted by an autonomous framework combined with transient flow measurements to enable comprehensive process understanding with minimal user input. An automated flow chemistry platform was employed to initially conduct linear flow-ramp experiments to rapidly map the reaction profile of three processes using transient flow data. Following experimental data acquisition, a computational approach was utilised to discriminate between all possible reaction models as well as identify the correct kinetic parameters for each process. Species that are known to participate in the process (starting materials, intermediates, products) are initially inputted by the user prior to flow ramp experiments, then all possible model candidates are compiled into a model library based on their potential to occur after mass balance assessment. Parallel computational optimisation then evaluates each model by algorithmically altering the kinetic parameters of the model to allow convergence of a simulated kinetic curve to the experimental data provided. Statistical analysis then determines the most likely reaction model based on model simplicity and agreement with experimental data. This automated approach to gaining full process understanding, whereby a small number of data-rich experiments are conducted, and the kinetics are evaluated autonomously, shows significant improvements on current industrial optimisation techniques in terms of labour, time and overall cost. The computational approach herein described can be employed using data from any set of experiments and the code is open-source.

1. Introduction

A major bottleneck in the transition from lab scale chemistry research to process development is the lack of quantitative chemical synthesis information, including critical aspects such as knowing the correct reaction model and precise kinetic parameters[1]. If this kinetic information is available, classical reaction engineering principles can be utilised to shorten process development time and lower overall scale-up costs.[2,3]

Often, when researchers investigate the kinetics of a process, time-series data is obtained under batch conditions due to the ability to collect multiple time points from a single experiment, then chemical intuition is used to postulate a reaction model. Subsequently, more advanced analysis of batch kinetic data is then possible in order to discriminate between possible kinetic models.[4,5] However, kinetic profiling in flow systems can give much more precise and reproducible

reagent addition control, increased temperature and reaction time accuracy, and the ability to explore areas of design space that would be too difficult to consider in a batch setting, i.e. super-ambient conditions above atmospheric boiling points, processes with unstable intermediates or very fast reactions.[6,7]

There are some examples in the literature coupling automated flow chemistry with algorithmic methods, namely Model-Based Design of Experiments (MBoE), for kinetic model discrimination from a defined set of candidate models.[8–12] Some notable recent examples include the confirmation of the kinetic models of a Pd catalysed C–H activation model by Echtermeyer *et al.*[8] and a benzoic acid esterification by Waldron *et al.*[7] Typically, targeted experiments are sequentially conducted until the confidence in one of the pre-defined reaction models is high enough to infer it is correct - then further experiments are conducted to accurately identify kinetic parameters based on the previously defined model. This can be a powerful tool for kinetic model

* Corresponding author.

E-mail address: R.A.Bourne@leeds.ac.uk (R.A. Bourne).

<https://doi.org/10.1016/j.cej.2020.127017>

Received 2 June 2020; Received in revised form 8 September 2020; Accepted 9 September 2020

Available online 12 September 2020

1385-8947/© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

determination, however, noisy systems may lead to difficulty in asserting confidence in a model due to standalone measurements in otherwise empty chemical space.[13] Furthermore, high-level chemical intuition is necessary in compiling the initial model candidate set for a MBDoE approach to select from, and some models can be overlooked based on their presumed inability to occur.

This work employs the use of pre-defined temperature-sequential automated flow ramps to quickly traverse areas of design space to generate large amounts of time-series data using minimal chemical material. This method is very efficient, using significantly less material than performing steady-state measurements, as time-series data or as part of a MBDoE study, as well as generating appreciably more data-rich kinetic information more quickly. This data is then automatically assessed by a computational method to determine the most likely reaction model and kinetic parameters, which evaluates every possible model based on all mass-balance-allowed chemical transformations.[3] This removes the necessity for high-level chemical intuition when formulating a reaction model, as the fully-comprehensive model identification and verification technique determines statistically the most accurate kinetic model, based on how the data describes the chemistry that is occurring. As part of this machine learning based approach, the kinetic parameters are also identified alongside the model, to quantify not only how the reaction occurs but how fast each of the individual component steps are.

This rapid time-series data generation with an integrated computational approach to kinetic model and parameter determination represents a powerful, automated technique for increasing the level of understanding of the mechanistic aspects of chemical processes. More rigorous coverage of the chemical space whilst using less material, alongside autonomously determining kinetic information without chemical bias or human interaction, can lead to overall savings in material, time and overall cost.

2. Materials and methods

2.1. Computational approach

We adopted a computational approach to reaction model and kinetic parameter determination consisting of two separate, sequential stages: model generation and kinetic fitting. To begin generating the models, all of the known and identifiable species within a process (starting materials, intermediates and products) must be initially inputted into the system as a ‘mass matrix’, describing the number of each atom in every species. A Mixed-Integer Linear Programming (MILP) optimisation algorithm then finds every possible reaction between these chemicals based solely on mass balance.[3] If the mass of all reactants in a chemical reaction equals the mass of the products, the reaction is considered as feasible regardless of whether chemical intuition believes it to be possible or not. As the computer program does not have in-built synthetic knowledge, it remains impartial and unbiased when preparing feasible reaction models, meaning model determination is achieved through purely statistical methods. This leads to consideration of every possible reaction model, including models that may have otherwise been dismissed by a trained chemist due to their prior impressions of the validity of particular reaction pathways.

2.1.1. Model generation

A library of plausible reaction models is then generated, where 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 reactions (δ). For example, if the MILP optimisation identifies 5 possible reactions ($\delta = 5$), then each of the 5 reactions themselves can be a standalone model, plus every combination of 2 reactions make a model, plus every combination of 3 reactions etc. The summation of all combinations then gives η equal to 31, where each of the participating species in these models are defined

as first-order. This example is shown in a model representation in Fig. 1, where this could be any fitting reaction where: SM = starting material, Int1 = intermediate 1, Int2 = intermediate 2, P = product and Imp = impurity. A specific chemical example showing material balances is shown in the ESI (Section 1.2.3).

2.1.2. Kinetic fitting and statistical analysis

Kinetic fitting is then performed on every model within the library using a gradient-based local minimisation algorithm.[14] Using this in conjunction with ordinary differential equation (ODE) solvers allows an assessment of how the concentration of a species changes over time when given particular rate constants, or k values. Therefore, it becomes possible to compare a particular model’s simulated ODEs to experimental work to identify the correct k values for that dataset. The algorithm optimises for the k values within the reaction model that it is provided, by minimising the sum of squared error (SSE) difference between the experimental data points and the ODE solver - as shown in Fig. 2. Therefore, the objective function that is minimised for each model evaluation is simply this SSE value. Ideally, data from multiple experiments with different starting concentrations can be supplied to minimise correlations in these fitted rate constants. This use of optimisation algorithms and ODE solvers allows for the rapid convergence of the simulated ODEs with the provided data points, giving an indirect route to identifying the k values for the reaction. Thus, this methodology removes the necessity for data point linearization via mathematical transformations that are seen in more conventional kinetic analysis techniques to determine rate constants.

During this fitting stage, there may be many reaction models that are unsuitable and provide a poor fit to the experimental data. In order to select the most likely model from the library, Corrected Akaike’s Information Criterion (AIC_c) is used as a statistical measure to identify the models that balance model simplicity and convergence to the data.[15] AIC_c , mathematically shown in eqn. (1), favours a small SSE value whilst penalising extra reaction model terms, where: N_{Data} = number of data

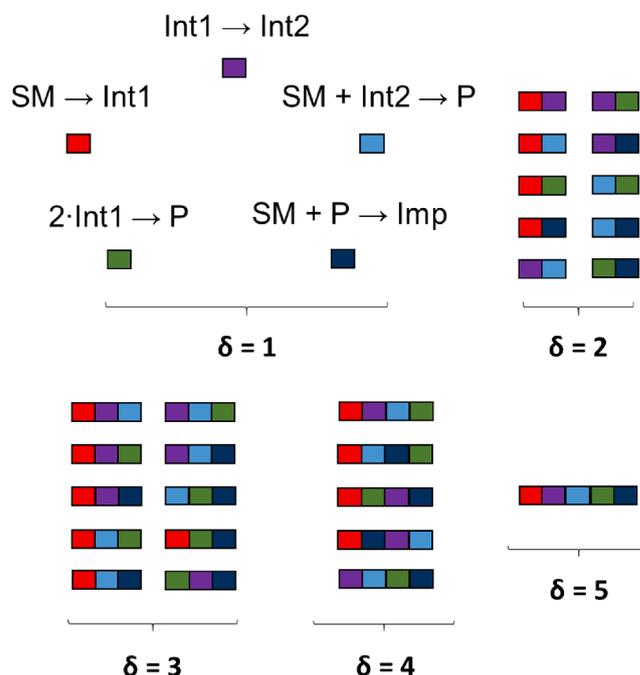


Fig. 1. A visual representation of all possible reaction models when given 5 potential sample reactions, each shown as a different coloured block. When $\delta = 1$, each reaction is in itself a model, and when $\delta > 1$, each reaction behaves as a model fragment. These fragments when combined in different ways provide full and unique reaction models, each of which are to be considered for kinetic fitting.

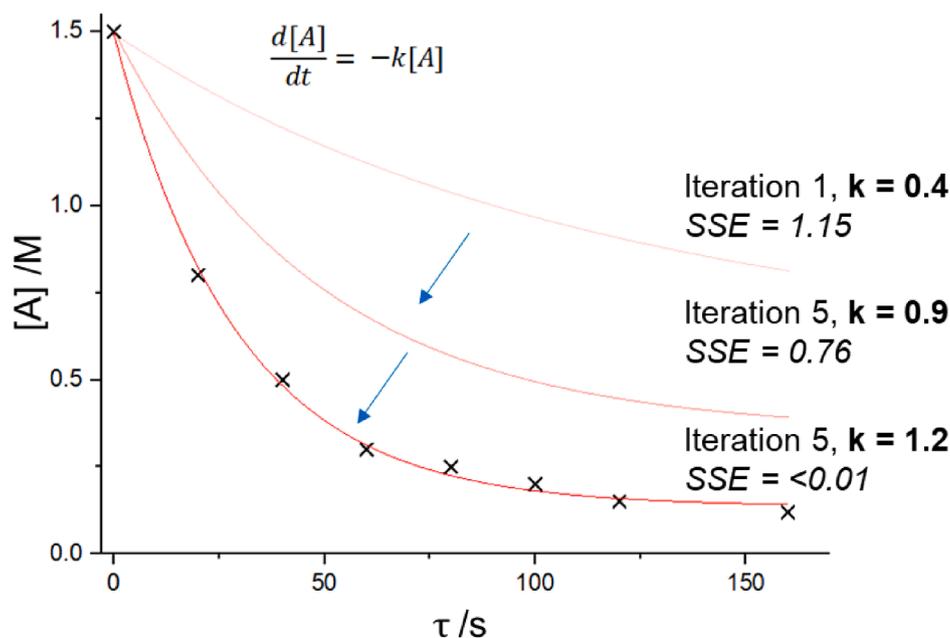


Fig. 2. A sample representation showing an optimisation algorithm's fit of an ODE (—) to experimental data (x) for a simple reaction model. This model's rate is described by differential equations and therefore altering the inputted rate constant affects the overall convergence of the curve(s), thereby minimising the outputted SSE value.

points sampled, $SSE = \text{sum of squared error}$, $\delta = \text{number of model terms}$. This ensures that there are no unnecessary terms in the model that provide little value to the fit of the ODEs to the data, which would cause an overfitting of potential models to the data. It is also possible to add unknown species to the system, which can determine if any changes in mass balance due to analysis is significant enough to suggest that there may be other degradation pathways or side-reactions. If the AIC_C value provides similar evaluations for multiple models, it would then be possible to perform further reactions to distinguish between them. An overview of the computational approach is shown in Fig. 3.

$$AIC_C = N_{Data} \cdot \ln\left(\frac{SSE}{N_{Data}}\right) + 2\delta + \frac{2\delta(\delta + 1)}{N_{Data} - \delta - 2} + N_{Data} \cdot \ln(2\pi) + N_{Data} \quad (1)$$

2.2. Data acquisition

2.2.1. Flow chemistry kinetics

Kinetic data acquisition methods for most chemists can typically be categorized as either steady-state sampling in flow, or more classically, sampling several time points from one batch experiment. Although

kinetic data is easily obtained under batch conditions as you can sample multiple times from the same reaction, the use of specialist and expensive continuous measurement systems is often required for accurate, valid kinetic profiling - however, this is very dependent on the chemistry and timescale of the reaction, and is particularly true of fast/sensitive reactions. This is because discrete analytical sampling alone in a batch system with short reaction times or small volumes may not provide data of sufficient quantity for the kinetic analysis of reaction progress, and is typically used effectively only as a check that an in situ technique remains accurate throughout the reaction. [5]

Acquiring kinetic data in a micro-fluidic system is advantageous in several ways. Continuous flow experiments have many benefits, such as enhanced mass transfer due to the small operating scale, precise temperature and reagent addition control, real-time reaction sampling, improved safety and the ability to automate reactions. [6] However, using traditional steady-state sampling methods may be unappealing when obtaining large amounts of data because of the material usage cost (approximately 1.5–3 reactor volumes are required per measurement) and the related, long wait times that are necessary to get the reactor to steady-state. [16] This may be problematic when material is expensive, as the material is purged to waste for each discrete steady-state sampling point, as well as when flow reactors with larger volumes are utilised, meaning that these relative material usage costs are translated to a higher magnitude of material consumption.

Therefore, to collect concentration–time data for chemical processes with a high data density laying on the curvature of the kinetic plot, we opted to use linear flow ramp gradients with on-line HPLC analysis. Steady-state is initially achieved within the system, then sampling begins during the transition to the next steady-state condition, utilizing the transitory information that would otherwise be lost as waste prior to the next measurement. Pump flow rates are incrementally lowered, meaning that as time passes, portions of the reaction mixture spend a progressively longer time in the reactor. This flow rate manipulation enables treatment of each sequential fluid element as a “pseudo batch” reaction that passes through the reactor in a time that is unique for that reaction medium. [17] This is shown in Fig. 4. This method allows the capture of accurate, reproducible data by capturing information from transient flow, as well as saving time and material by not performing

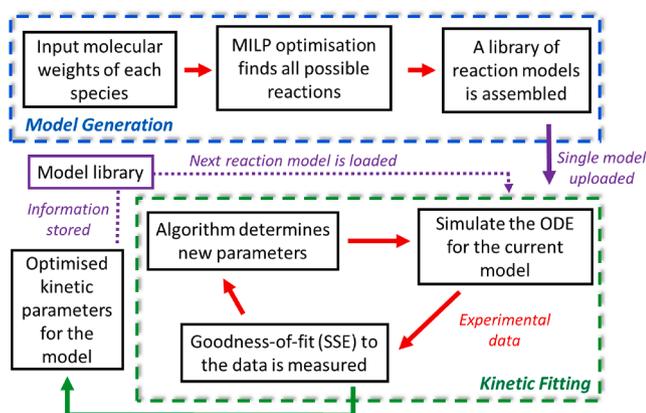


Fig. 3. A flow chart representing the key steps of the computational approach to reaction model and kinetic parameter determination.

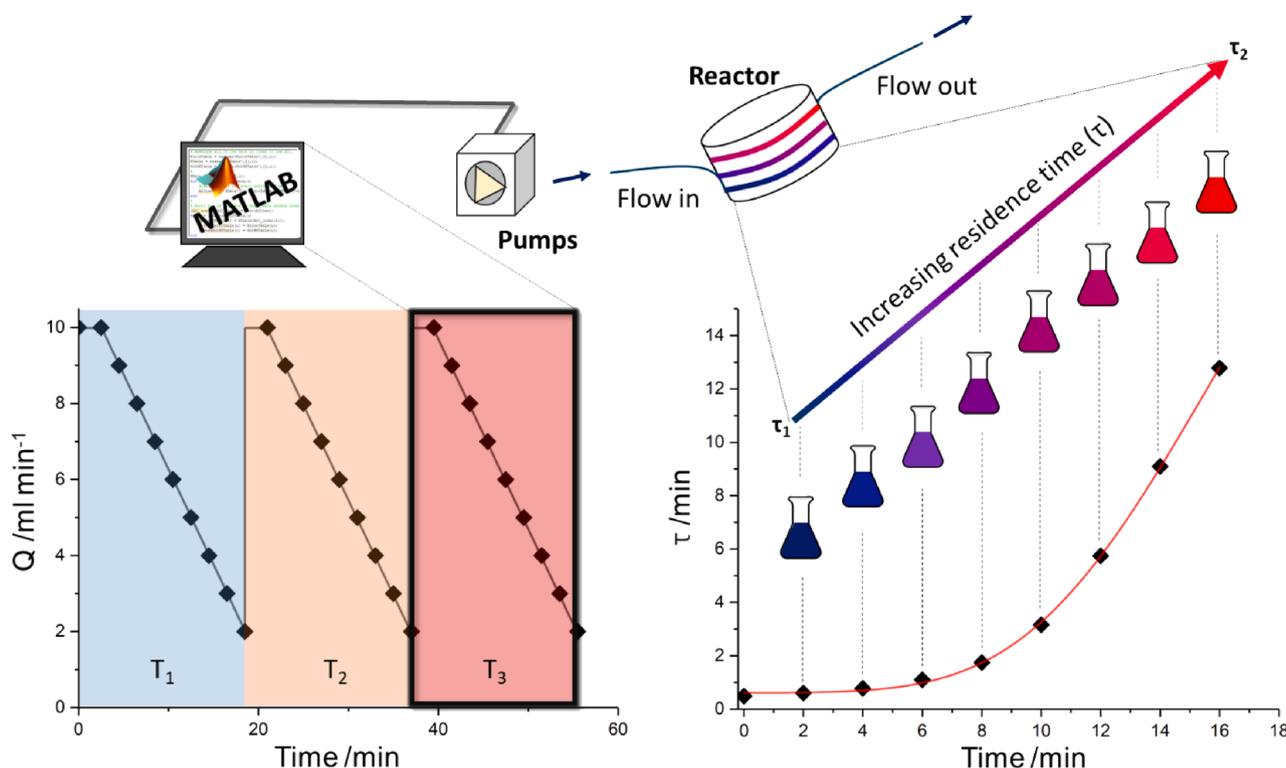


Fig. 4. A representation of how linear gradient flow ramps can be utilised to sample with a high data density on the initial curvature of the kinetic plot. Where: \blacklozenge = data point, T_n = experiment temperature, Q = total flow rate, Time = time the reaction has been running, τ = residence time that the reaction mixture experiences.

steady-state flow measurements - comparisons of reaction material usage between these techniques is calculated in the ESI (Section 1.4.5).

For each chemical process, at least two flow ramp experiments at different temperatures were conducted, with HPLC sampling every 2–3 min. When conducting steady-state experiments, the residence time can typically be calculated as a function of simply the flow rate and reactor volume. However, in a flow-rate-manipulated experiment, the flow rate is of course not constant and is related to the deceleration of individual pump flow rates over time. Therefore, the residence time was calculated for every sampling point in each experiment using eqn. (2), using: τ (residence time), α (deceleration of pump flow rate), μ_0 (initial flow rate), t (experiment time) and L (reactor volume), following derivations highlighted by Hone[16]:

$$\tau = \frac{\alpha \cdot t - \mu_0 + \sqrt{(\mu_0 - \alpha \cdot t)^2 + 2 \cdot L \cdot \alpha}}{\alpha} \quad (2)$$

Several authors have generated kinetic data similarly by using transient flow measurements. Mozharov *et al.* first employed a step change in the flow rate to investigate a Knoevenagel condensation reaction.[18] Jensen and Bourne then independently reported an improved, more precise methodology by ramping the flow rate to sample the transient flow.[16,17] Temperature ramping experiments have also been reported, as well as more recently, efficient temperature and flow ramping experiments simultaneously.[19]

Using this flow ramp methodology, it is not only possible to automate the generation of kinetic data, but also to obtain large amounts of data for very fast reactions which would be infeasible by batch methods. The use of flow ramps for the reactions herein described allowed discrete sampling of over 40 time points with less than one minute residence time in some instances, allowing kinetic insights into processes that would be too difficult to monitor by other means. Therefore, depending on the speed of the reaction, *a priori* estimations of reaction completion for particular residence times may need to be conducted. It is also important to highlight that this computational approach to determining kinetic

information is not limited to fast reactions, and is equally applicable to time-series data for all speeds of reaction, obtained by any means (batch sampling, in-line analysis etc).

2.2.2. Reactor modelling

For the kinetic simulations in this study, a plug flow model was adopted. If we consider the general axial dispersion plug flow reactor model[20] in eqn. (3):

$$D_R \frac{d^2 C_i}{dz^2} - u \frac{dC_i}{dz} + r_i = 0 \quad (3)$$

With D_R being the axial dispersion coefficient, C_i the molar concentration of species i , z the length along the reactor, u the superficial velocity and r_i the rate of reaction of species i . Previous studies by Hone [16] and Jensen[10,19,21] investigated dispersion in the microreactor systems similar to the one used in this study, with both finding dispersion only introduces a small deviation from plug flow. Given this, eqn. (3) can be simplified to ignore the second order term, to give eqn. (4)–(6). This PFR design equation can be utilised to model each species in the reaction with respect to residence time.

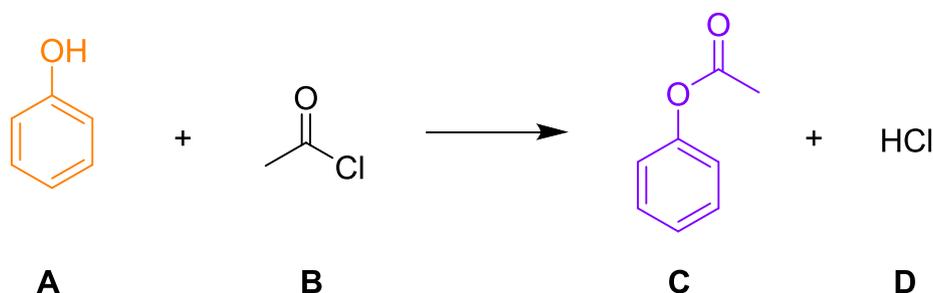
$$u \frac{dC_i}{dz} = r_i \quad (4)$$

$$\tau = \frac{z}{u} \quad (5)$$

$$\frac{dC_i}{d\tau} = r_i \quad (6)$$

3. Results and discussion

The first reaction we explored was a kinetic study to validate the presupposed model of the reaction of phenol with acetyl chloride to form phenyl acetate (Scheme 1), as well as identifying the kinetic parameters for the process. The kinetic data was acquired for this reaction



Scheme 1. The reaction of phenol with acetyl chloride to form phenyl acetate and hydrochloric acid.

by using linear flow ramps at two temperatures, 65 °C and 75 °C (Fig. 5), with the computational approach applied automatically upon completion of the experiments. Specific flow rate manipulations and experimental details can be found in the ESI (Section 1.2).

Based on the four species identified, there were only two reactions that were possible given the mass balance. Therefore, with $\delta = 1$ and 2, there are 3 feasible reaction models that must be evaluated: the pre-supposed forward model (eqn. (7)), the reverse of this reaction (eqn. (8)) and the combination of both (eqn. (9), where an equilibrium occurs). The results from these evaluations are tabulated in Table 1. 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 $1 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$, indicating that the reaction model is in complete disagreement with the experimental data. For the equilibrium model, the error can reach as low as the forward reaction model alone as the optimiser assigns a negligible k value to the reverse reaction. 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 combination of the correct reaction model and kinetic parameters allowed a fit to the experimental data with an average residual of less than $3 \times 10^{-3} \text{ M}$, which corresponds to an average percentage residual error of less than 2%.



Table 1

Evaluation of the feasibility of each reaction model described in Scheme 2. A lower AIC_c evaluation indicates a more likely reaction model.

Reaction Model	k Values / $\times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$		SSE / M	AIC _c Evaluation
	65 °C	75 °C		
Forward	5.15 ± 0.20	10.45 ± 0.42	0.019	-471
Backward	–	–	0.532	-206
Forward + Backward	5.15 ± 0.200	10.45 ± 0.42	0.019	-467



By only inputting the species involved in the reaction, then running two linear flow ramps with the rapid automated computational approach described, the intuitive reaction model was confirmed and the kinetic parameters were identified as: $k_{75 \text{ °C}} = 10.45 \times 10^{-3} \pm 0.42 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$, $E_a = 69.3 \pm 7.8 \text{ kJ mol}^{-1}$. As the flow ramping relies on instantaneous changes in flow characteristics which are not possible, this can lead to some initial non-ideal behaviour in the reactor, sometimes leading to small deviations from the true kinetic curvature of the reaction. This can result in minor effects, observable in some of the case studies herein reported, relating to non-normally distributed residuals which are also reported by other groups.[19,22] However, this effect was shown to have a negligible impact on the ability to extract true kinetic information from these ramped experiments. The equivalent phenyl acetate experiments were also performed using steady-state sampling methods for comparative purposes, to ensure that a transient-flow regime remained accurate in model and parameter

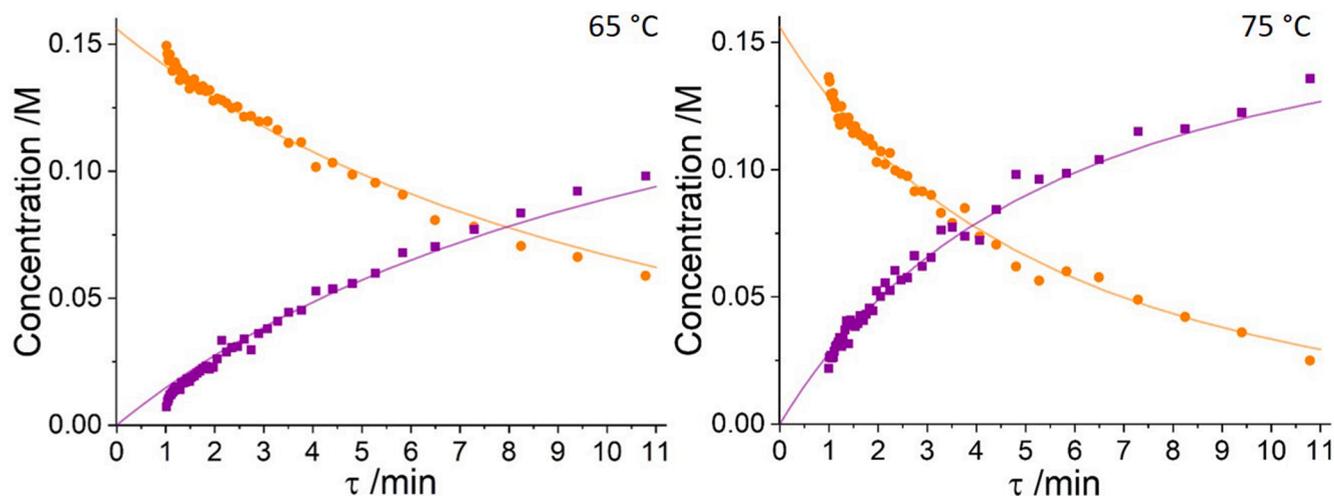


Fig. 5. Kinetic profiles for two flow ramp experiments at 65 °C and 75 °C, where: ● = phenol, ■ = phenyl acetate, — = phenol (ODE), — = phenyl acetate (ODE). See ESI (Section 1.2) for full experimental conditions and raw data.

determination. Results showed agreement in plotted curvature and confirmed the same reaction model to be the most likely, whilst identifying comparable kinetic constants within the error range calculated - see ESI (Section 1.2.5) for more details.

Following this, we investigated the reaction of 4-aminophenol with acetic anhydride forming paracetamol, followed by the consecutive reaction to form diacetamide, as shown in Scheme 2. 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 diacetamide, at 30/60 °C and 160/180 °C respectively. Each of these ramps were conducted with differing reactor sizes and hence residence times, as well as starting concentrations of acetic anhydride. The reactor volumes for the lower temperature reactions were 0.25 mL and 0.5 mL, whilst the higher temperature reactions were carried out in a 3.5 mL reactor. 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. The kinetic profiles for these experiments are shown in Fig. 6.

63 potential models were identified from the 5 principal reactions possible by mass balance. Of those 63 models, the reaction model shown in Scheme 2 was determined to be the most likely representation of the system. The approach also determined the kinetic parameters of step one: $k_{60\text{ }^\circ\text{C}} = 6.45 \pm 0.26 \text{ M}^{-1} \text{ s}^{-1}$, $E_a = 3.2 \pm 1.2 \text{ kJ mol}^{-1}$ and step two: $k_{180\text{ }^\circ\text{C}} = 4.27 \times 10^{-2} \pm 0.17 \times 10^{-2} \text{ M}^{-1} \text{ s}^{-1}$, $E_a = 97.9 \pm 6.5 \text{ kJ mol}^{-1}$. This allows 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. Using this efficient methodology, we achieved effective sampling rates of more than one HPLC sample per second, with reaction times of less than 10 s - this would be very difficult to replicate in a batch system. Furthermore, calculations were performed using the kinetic parameters obtained for the formation of diacetamide, showing that to reach the same conversion as shown in Fig. 6 (180 °C experiment) and hence gain the same process understanding, would take approximately 70 days in a batch vessel at constant reflux.

The identified model alongside the determined kinetic parameters fits to the experimental data very accurately, with an average residual of less than $1 \times 10^{-4} \text{ M}$, corresponding to an average percentage residual error of less than 0.5%. It is also possible to view this reaction in terms of a more complex model to obtain a better fit to the experimental data; a full data table in the ESI (Section 1.3.3) shows that some models describe the data more accurately. This is likely to only be true within the bounds of the explored design space, as extrapolation using over-fitted (or under-fitted) models will tend to lead to a loss of accuracy in the predicted values[23] - this is why model ranking incorporates model simplicity as well as the fit to the data.

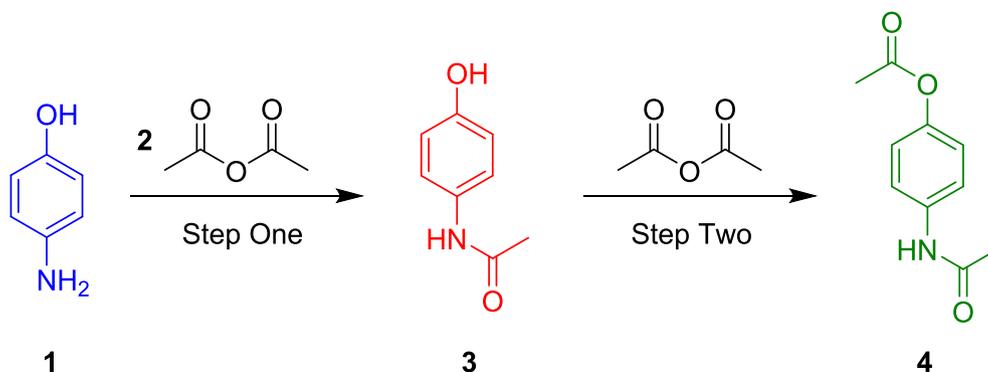
Finally, we assessed the methodology on one of the final reaction steps required in the synthesis of the current pharmaceutical product,

metoprolol, shown in Scheme 3. Metoprolol, **8**, 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.[24] The reaction of interest is shown in Scheme 3, depicted as the presumed reaction model with a consecutive reaction to form the tertiary amine product, **9**.

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 the ESI (Section 1.4).

The kinetic profiles from the flow ramp experiments are shown in Fig. 7. The reaction pathway, shown in Scheme 3, was determined to be the most likely out of 63 potential models, and the kinetic parameters for the formation of metoprolol were determined to be: $k_{190\text{ }^\circ\text{C}} = 9.95 \times 10^{-3} \pm 0.07 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$, $E_a = 60.8 \pm 2.0 \text{ kJ mol}^{-1}$ and for the formation of **9**: $k_{190\text{ }^\circ\text{C}} = 1.67 \times 10^{-3} \pm 0.40 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$, $E_a = 72.4 \pm 1.9 \text{ kJ mol}^{-1}$. This kinetic information provides an excellent fit to the experimental data, with the average residual of less than $2 \times 10^{-3} \text{ M}$, corresponding to an average percentage residual error of less than 0.4%.

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, [25] 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 ESI (Section 1.4.5) for full details. Other optimisation strategies, such as MBDoE,[26] may also lead to improvements in material consumption as singular experiments can be run in flow rather than full reaction profiles, but this comparison was not within the scope of this study. As these optimisation techniques may deliver different model formats (empirical or physical), the main assumption in these comparisons is that the optimised experimental conditions are found regardless. 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.



Scheme 2. The reaction of 4-aminophenol, **1**, with acetic anhydride, **2**, to form paracetamol, **3**, in step one, followed by a further reaction with acetic anhydride to form diacetamide, **4**, in step two.

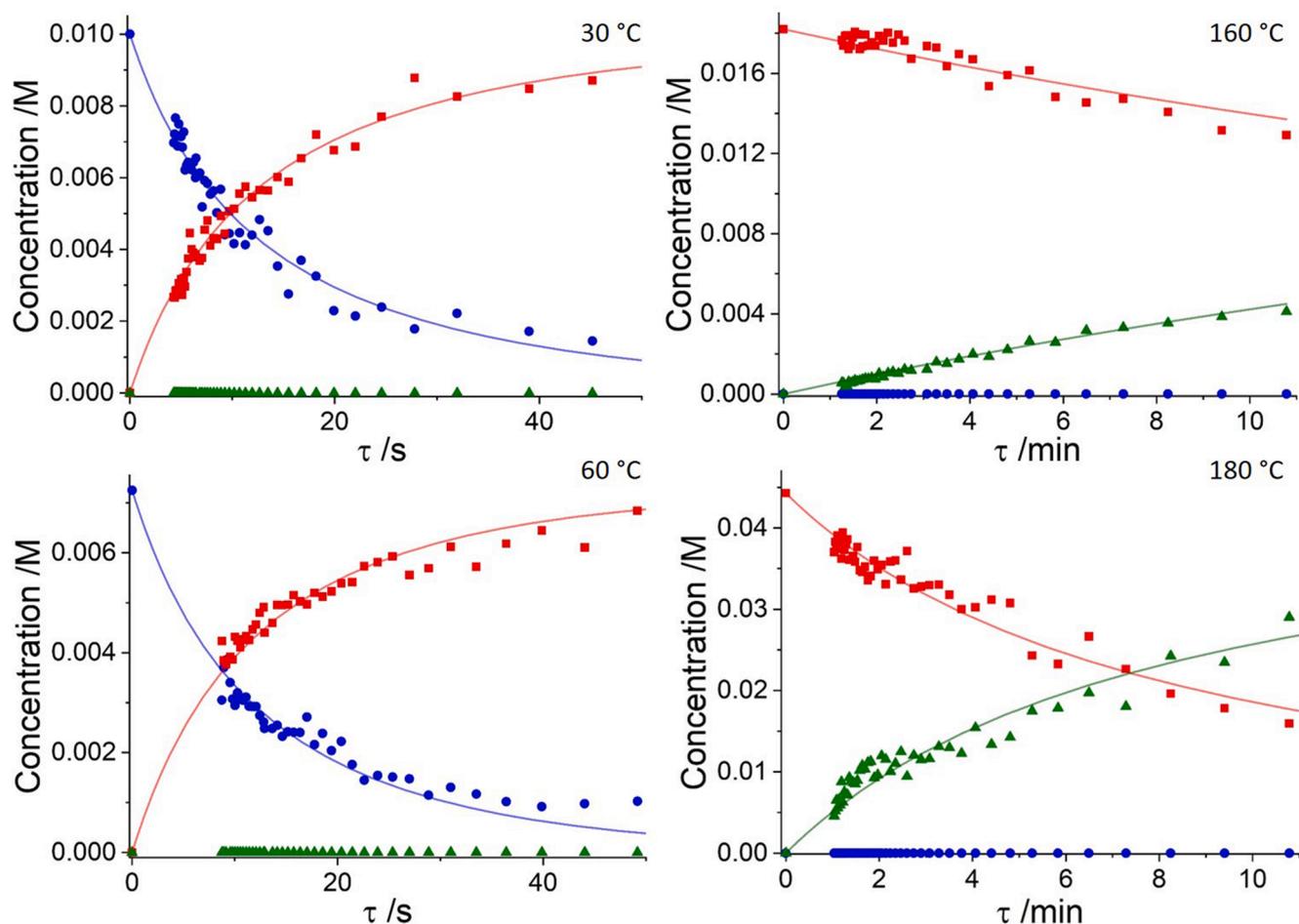
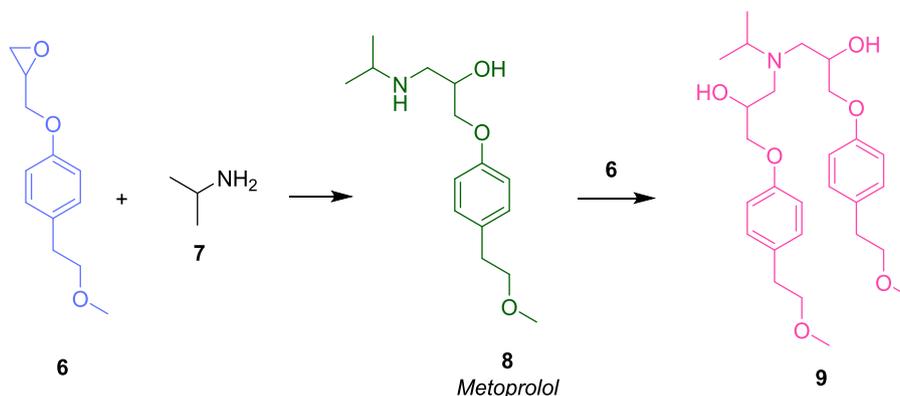


Fig. 6. Kinetic profiles for four flow ramp experiments at 30 °C, 60 °C, 160 °C and 180 °C, where: ● = 4-aminophenol, ■ = paracetamol, ▲ = diacetamate, — = 4-aminophenol (ODE), — = paracetamol (ODE), — = diacetamate (ODE). See ESI (Section 1.3) for full experimental conditions and raw data.



Scheme 3. The reaction of the starting material, 6, with isopropylamine, 7, to form metoprolol, 8, as well as the consecutive reaction to form the bis-substituted product, 9.

4. Conclusion

We have shown in this work that when time-series data for a chemical process is available, scalable process understanding can be achieved with minimal 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 machine learning technique 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 be interpreted by a trained chemist to

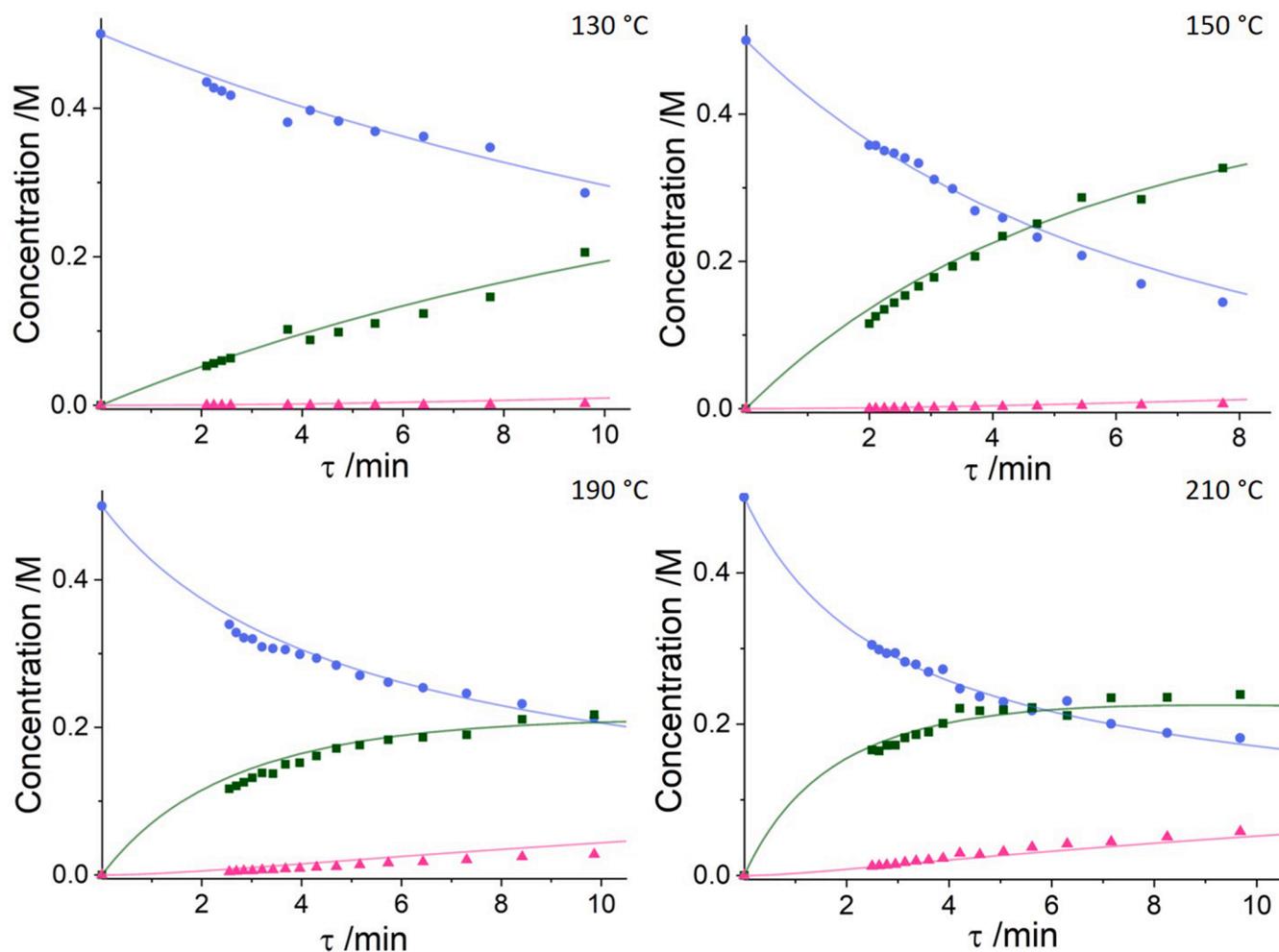


Fig. 7. Kinetic profiles for the flow ramp experiments at 130 °C, 150 °C, 190 °C and 210 °C, where: ● = starting material, ■ = metoprolol, ▲ = bis-substituted product, — = starting material (ODE), — = metoprolol (ODE), — = bis-substituted product (ODE). See ESI (Section 1.4) for full experimental conditions and raw data.

further differentiate reaction pathways based on what should and should not be chemically possible.

This approach will be particularly powerful in situations where the reaction model is not completely understood, for example when there are competing reaction pathways or in cases of catalytic systems where mechanistic insight may be paramount. Further plans for experimental work include expanding the scope of study to homogeneously/heterogeneously catalysed reactions and systems with varying integer and non-integer species orders.

The approach can be computationally expensive depending on the number of possible reaction models, as the number of mass-balance-allowed reactions relates exponentially to the number of models to be evaluated. For example, a system with 5 mass-balance-allowed transformations produces 31 models, whilst a system with 10 reactions relates to 1023 models that must be evaluated. However, as the approach has been parallelised for efficiency, circumstances relating to demanding numbers of model evaluations can be improved by substituting computer hardware for further logical cores. All of the optimisations carried out by the approach on the work described was evaluated in less than 5 min on a standard 4-core Intel i5-2310 processor.

This work is the first implementation of the approach on observed 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.

5. Experimental

5.1. General experimental setup

All flow experiments, excluding two AstraZeneca 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 μ L aliquot volume. 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 (5 cm \times 4.6 mm, 2.7 μ m). This general setup is shown in Fig. 8 and Fig. 9. Full experimental setup details for the University of Leeds and AstraZeneca experiments can be found in the ESI (Section 1).

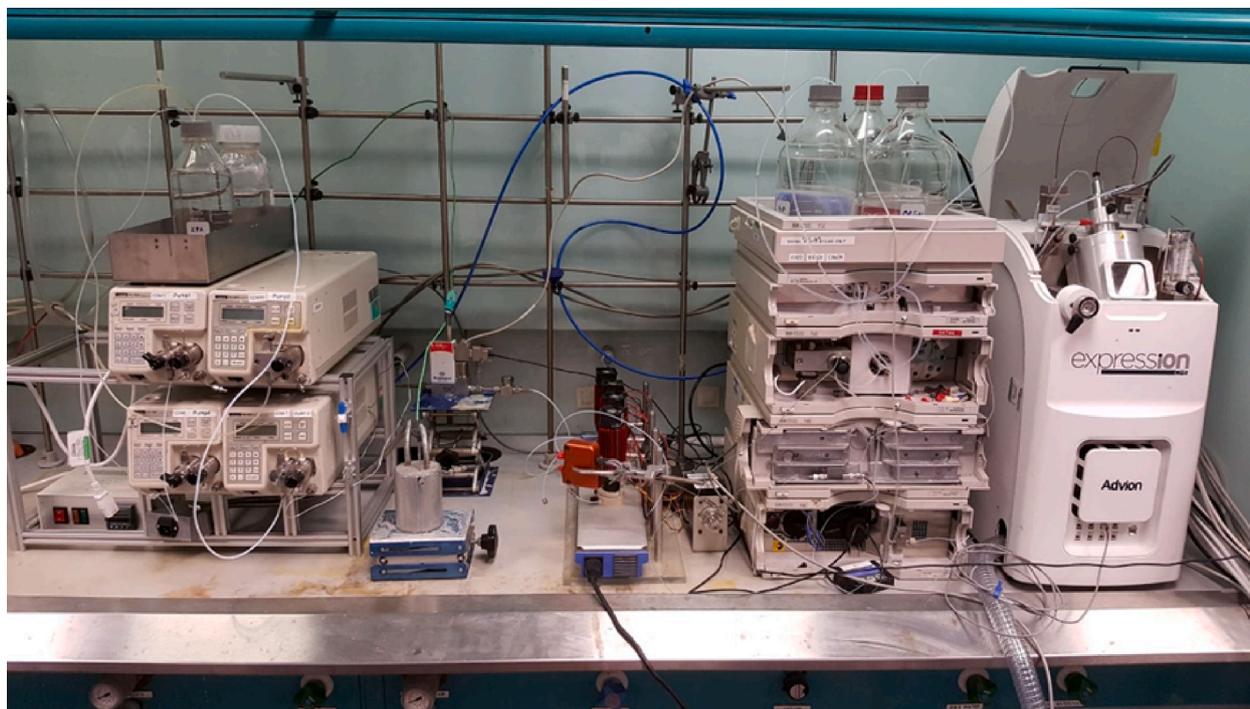


Fig. 8. A photograph of the general experimental setup as used at the University of Leeds.

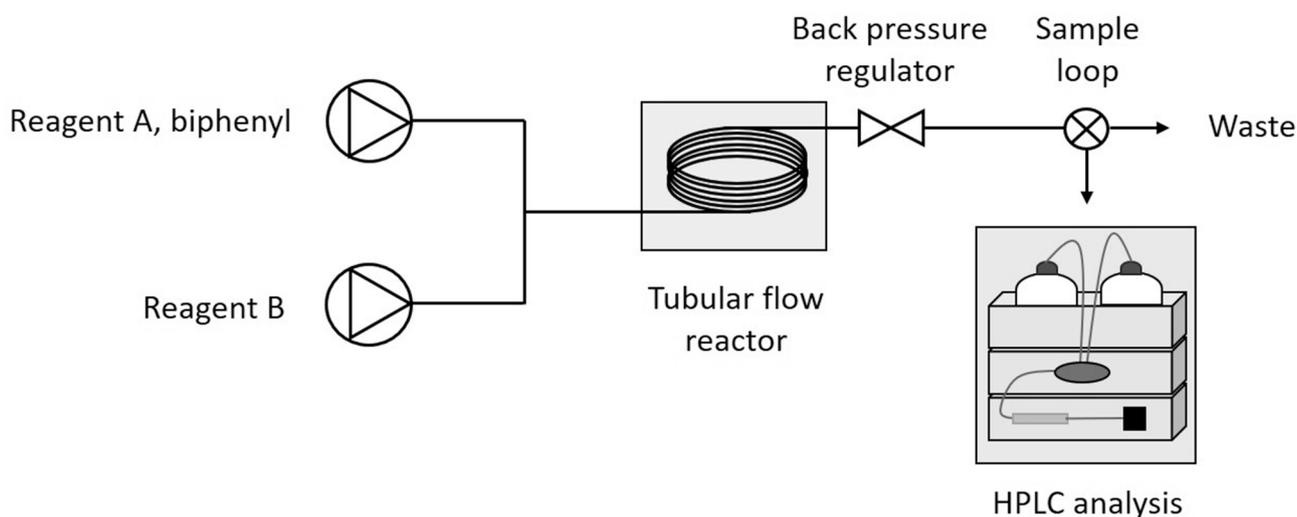


Fig. 9. A generalised schematic of the reactor setup for the case studies reported.

5.2. General reactor setup

A tubular reactor of fixed volume is implemented in all experimentation, where the specific volume is changed based on the case study of interest. The desired reservoir solutions of A and B are prepared by dissolving the desired reagents in a particular solvent under stirring at ambient conditions. In all experiments biphenyl was added to one reservoir as an internal standard. Specific details for each experiment can be found in the [ESI](#).

Linear gradient flow ramps allow 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. Steady state is initially established and the flow rate for each pump decreased at a constant rate for a fixed time. Samples of reactor effluent are injected for HPLC analysis at regular intervals, thus

achieving a large data density. Specific experimental flow regimes can be found in the [ESI](#), as well as all experimental data.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank the School of Chemical and Process Engineering and the School of Chemistry at the University of Leeds for their support, and EPSRC (EP/R032807/1 and EP/N509681/1) and AstraZeneca for their funding and support. RAB was supported by the Royal Academy of

Engineering under the Research Chairs and Senior Research Fellowships scheme (RCSRF1920\9\38).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cej.2020.127017>.

References

- [1] Taylor, C.J. MILP Kinetic Fitting Code. 2019; Available from: <https://github.com/ConnorJTaylor/MILPFitter>.
- [2] J. Tsu, V.H.G. Díaz, M.J. Willis, Computational approaches to kinetic model selection, *Comput. Chem. Eng.* 121 (2019) 618–632.
- [3] M.J. Willis, M. von Stosch, Inference of chemical reaction networks using mixed integer linear programming, *Comput. Chem. Eng.* 90 (2016) 31–43.
- [4] C.-D.-T. Nielsen, J. Burés, Visual kinetic analysis, *Chem. Sci.* 10 (2) (2019) 348–353.
- [5] D.G. Blackmond, Reaction progress kinetic analysis: a powerful methodology for mechanistic studies of complex catalytic reactions, *Angew. Chem. Int. Ed.* 44 (28) (2005) 4302–4320.
- [6] M.B. Plutschack, et al., The hitchhiker's guide to flow chemistry, *Chem. Rev.* 117 (18) (2017) 11796–11893.
- [7] M.J. Pedersen, et al., Optimization of Grignard Addition to Esters: Kinetic and Mechanistic Study of Model Phthalide Using Flow Chemistry, *Industrial Engineering Chemistry Research* 57 (14) (2018) 4859–4866.
- [8] C. Waldron, et al., An autonomous microreactor platform for the rapid identification of kinetic models, *React. Chem. Eng.* (2019).
- [9] A. Echtermeyer, et al., Self-optimisation and model-based design of experiments for developing a C-H activation flow process, *Beilstein J. Org. Chem.* 13 (1) (2017) 150–163.
- [10] J.P. McMullen, K.F. Jensen, Rapid determination of reaction kinetics with an automated microfluidic system, *Org. Process Res. Dev.* 15 (2) (2011) 398–407.
- [11] B.J. Reizman, K.F. Jensen, An automated continuous-flow platform for the estimation of multistep reaction kinetics, *Org. Process Res. Dev.* 16 (11) (2012) 1770–1782.
- [12] M. Cruz Bournazou, et al., Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities, *Biotechnol. Bioeng.* 114 (3) (2017) 610–619.
- [13] G. Franceschini, S. Macchietto, Model-based design of experiments for parameter precision: State of the art, *Chem. Eng. Sci.* 63 (19) (2008) 4846–4872.
- [14] Y. Wang, et al., Parameter inference for discretely observed stochastic kinetic models using stochastic gradient descent, 4 (1) (2010) 99.
- [15] H. Akaike, A new look at the statistical model identification, in: *Selected Papers of Hirotugu Akaike*, Springer, 1974, pp. 215–222.
- [16] C.A. Hone, et al., Rapid multistep kinetic model generation from transient flow data, *React. Chem. Eng.* 2 (2) (2017) 103–108.
- [17] J.S. Moore, K.F. Jensen, “Batch” kinetics in flow: online IR analysis and continuous control, *Angew. Chem. Int. Ed.* 53 (2) (2014) 470–473.
- [18] S. Mozharov, et al., Improved method for kinetic studies in microreactors using flow manipulation and noninvasive Raman spectrometry, *J. Am. Chem. Soc.* 133 (10) (2011) 3601–3608.
- [19] K.C. Aroh, K.F. Jensen, Efficient kinetic experiments in continuous flow microreactors, *React. Chem. Eng.* 3 (1) (2018) 94–101.
- [20] O. Levenspiel, Chemical reaction engineering, *Ind. Eng. Chem. Res.* 38 (11) (1999) 4140–4143.
- [21] K.D. Nagy, et al., Mixing and dispersion in small-scale flow systems, *Org. Process Res. Dev.* 16 (5) (2012) 976–981.
- [22] M. Rubens, J. Van Herck, T. Junkers, Automated polymer synthesis platform for integrated conversion targeting based on inline benchtop NMR, *ACS Macro Lett.* 8 (11) (2019) 1437–1441.
- [23] D.M. Hawkins, The problem of overfitting, *J. Chem. Inf. Comput. Sci.* 44 (1) (2004) 1–12.
- [24] J. Koch-Weser, Metoprolol, *N. Engl. J. Med.* 301 (13) (1979) 698–703.
- [25] Fluorochem. 2019 [cited 2019 27/06/2019]; Available from: <http://www.fluorochem.co.uk/Products/Product?code=303261>.
- [26] M. Quaglio, E.S. Fraga, F. Galvanin, Constrained model-based design of experiments for the identification of approximated models, *IFAC-PapersOnLine* 51 (15) (2018) 515–520.