Reaction Chemistry & Engineering

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d5re00082c

Reaction kinetics for the synthesis of an anti-cancer drug (adavosertib) precursor†

Matthew Blair, [a] Mazaher M. Chalchooghi, [b] Robert J. Cox [b] and Dimitrios I. Gerogiorgis *[a]

Received 18th February 2025,
Accepted 14th April 2025

DOI: 10.1039/d5re00082c

rsc.li/reaction-engineering

The development of kinetic models which can accurately describe drug synthesis reactions is an important part of process design in the pharmaceutical industry. Correctly identifying these models can be difficult, since the reaction pathways used to manufacture new pharmaceutical compounds are often extremely complex. Consequently, many kinetic modelling and parameter estimation tools have been developed in recent years to allow drug manufacturers to test and compare a variety of reaction models before selecting the one which provides the best predictions. The present paper employs a multistart parameter estimation code (in MATLAB®) to parameterise a range of kinetic models describing the synthesis of a key intermediate required for the production of a new anti-cancer drug, Adavosertib (AZD1775). Furthermore, the Akaike and Bayesian Information Criteria are used to rank these models based upon their complexity and fidelity to reflect real-world experimentation.

1. Introduction

Since 2005, public spending on anti-cancer drugs has steadily increased across Europe (Fig. 1a) and the United States; leading to a total global spending of 150 billion USD in 2020.[1,2] Whilst some of this spending can be attributed to rising incidence rates of various cancer types around the world[2–6] (with 18.1 million new cases and 10 million deaths reported in 2020 alone[7]), it is clear that rising drug prices have contributed to this landscape (Fig. 1b). Consequently, pharmaceutical companies must find a way to reduce the price of their medications if treatments are to be affordable in the future.[1,2]

To accomplish this goal, drug manufacturers must first find a way to intensify their manufacturing processes. Moreover, they must do so whilst minimising experimentation so as to not simply inflate costs elsewhere in their pipeline. Therefore, many authors[8–16] propose that fit-for-purpose process models should be developed, to accurately describe the different unit different operations involved in pharmaceutical processes. Once these models are correctly parameterised, they can provide an efficient way to visualise and optimise processes. This has already been demonstrated by Jolliffe *et al.*,[15] Diab *et al.*,[8] and Cuthbertson *et al.*[17] for ibuprofen, diphenhydramine and amoxicillin, respectively.

Numerous studies have been published in recent years on the importance of developing and parameterising kinetic models for chemical synthesis[11,18–36] and crystallisation processes;[37,38] with the former receiving particular attention (Table 1). For example, Schenk *et al.*[39] demonstrated how parameter estimation tools can be used to compare different kinetic models available to describe drug synthesis processes, using the production of an asymmetrical urea compound required to manufacture various active pharmaceutical ingredients (APIs) as a case study. Earlier, Grom *et al.*[22] had shown that computational modelling techniques can be used to study the reaction mechanisms underpinning Lorcaserin synthesis (a complicated reaction network, consisting of 27 reaction steps and 15 chemical species), leading to the robust development of a temperature-dependent kinetic model with 29 parameters. Published studies have been conducted to show how kinetic modelling can be used to comparatively assess the benefits of batch *vs.* continuous manufacturing processes.[12,40] The paper of Kraus *et al.*[41] considered the eco-friendly production of carbamazepine from urea and iminostilbene. Because of these efforts, various established software packages are now available to conduct this type of studies, and they are summarised in Table 2.

[a] Institute for Materials and Processes (IMP), School of Engineering, University of Edinburgh, The King's Buildings, Edinburgh, EH9 3FB, UK.
E-mail: D.Gerogiorgis@ed.ac.uk; Tel: +44 (0)131 6517072
[b] Chemical Development, Pharmaceutical Technology and Development, Operations, AstraZeneca, Macclesfield, SK10 2NA, UK

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d5re00082c

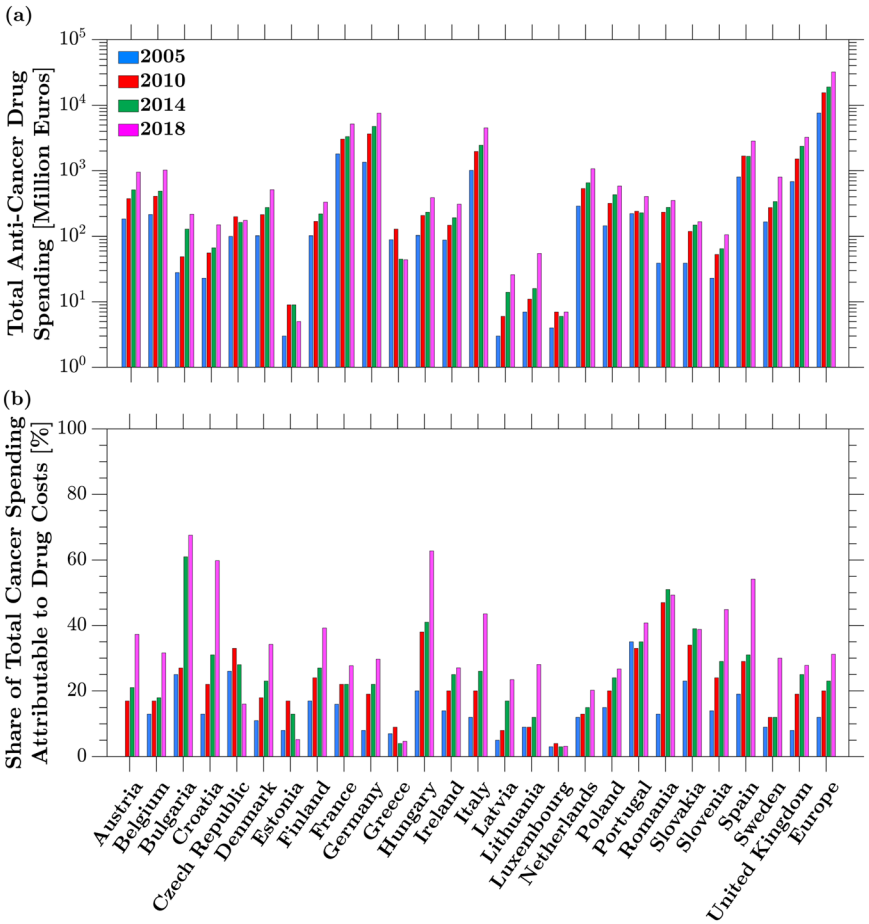

Fig. 1 Anti-cancer spending in Europe, 2005–2018 (data: ref. 2 and 3). (a) Anti-cancer drug expenditure; (b) drug cost fraction.

Several studies have also been conducted to show how computational modelling can simplify conceptualising and ranking the kinetic models developed for these reactions.[53] Tsu *et al.*[54] recently used integer linear programming (ILP) to identify stoichiometric coefficients for synthetic chemical reactions, whilst August *et al.*[55] used it to identify promising chemical reaction networks for biological systems. Meanwhile, Taylor *et al.*[23,56] and Willis *et al.*[57] have shown that Mixed Integer Linear Programming (MILP) can be used to automatically generate and test different rate laws and reaction networks for systems where the reagents, products and intermediates are already known. Ultimately, the grand vision is allowing all feasible unimolecular and bimolecular reactions between different chemical species to be determined automatically,[53,54,57] before leveraging metrics such as the Akaike[23,56,58] and Bayesian information criterions[59] to identify models which have the simplest structure, yet can accurately predict experimental observations.[23,60]

A promising cancer therapy which can particularly benefit from accelerated process R&D is Adavosertib (AZD1775) – an experimental oral medication which may inhibit tyrosine kinase WEE1 activity during cell signalling, growth and division[61] – since it has shown clinical efficacy as a monotherapy against a range of cancers. Most notably, it is potent against non-small cell lung cancer (NSCLC) and pancreatic cancers,[62] which together account for over 50% of all cancer-related deaths in Europe today.[4] This compound has also shown promise as a combinatorial therapy when treating NSCLC, ovarian cancer and leukemia using Sotorasib,[63] chemotherapy[62] and Cytarabine[64] as partner therapies.

Considering this, in the present paper we propose a range of kinetic models which can be used to describe the synthesis of an important precursor required for adavosertib production: AZD1775 hydroxymethylsulfanyl (HMS). Neither a reaction mechanism nor any kinetic model have ever been previously proposed for the synthesis of this compound, but understanding its synthesis and production is critical for the cost-effective manufacture of Adavosertib. To conduct this kinetic study, we have used an original parameter estimation code written in MATLAB® to parameterise 64 candidate kinetic models, before ranking them based on their complexity and fidelity to reflect lab-scale experimentation. Remarkably, each of the models hereby postulated and considered have been developed by invoking knowledge of similar chemical reactions.

Table 1 Summary of published reaction modelling and kinetic parameter estimation studies conducted for different APIs and drug precursors

API/precursor	Condition treated	Study outcomes	Software	Ref.
Carbamazepine	Epilepsy	Kinetic model & Arrhenius rate law	MATLAB®	41
Unspecified	Unspecified	Kinetic model & isothermal rate constants	KIPET	39
Lorcaserin	Obesity	Kinetic model & Arrhenius rate law	—	22
Lomustine	Brain tumours, Hodgkin's lymphoma	Kinetic models, isothermal rate constants	MATLAB®	18
Osimertinib intermediate	Non-small cell lung cancer	Kinetic model & Arrhenius rate law	Dynochem®	19
Carfilozomib intermediate	Myeloma	Kinetic model & Arrhenius rate law	Dynochem®	20
Merestinib intermediate	Biliary tract & non-small cell lung cancer	Kinetic model & isothermal rate constants	—	21
Paracetamol	Moderate pain, fevers	Kinetic model & isothermal rate constants	MATLAB®	23
Metoprolol	High blood pressure	Kinetic model & isothermal rate constants	MATLAB®	23
Unspecified	Unspecified	Kinetic model & isothermal rate constants	KIPET	24
Ibuprofen	Moderate pain, fevers, inflammation	Kinetic model & isothermal rate constants	MATLAB®	25
Diphenhydramine	Hay fever, common cold, short-term insomnia	Kinetic model & isothermal rate constants	MATLAB®	8
Nevirapine	HIV	Arrhenius rate law	MATLAB®	11
Aziridines (building block)	Cancer therapies (mitomycin, azinomycin)	Arrhenius rate law	gPROMS®	26
Pyrroles (building block)	Cancer therapies (sunitinib), pain relief (ketorolac), heart disease (atorvastatin)	Kinetic models & Arrhenius rate law	COMSOL Multiphysics®	27
Abemaciclib	Advanced/metastatic breast cancers	Arrhenius rate law	Dynochem®	28
Thiazolidine intermediate	Diabetes	Kinetic model & Arrhenius rate law	—	29
Glitazone intermediate	Diabetes	Kinetic model & Arrhenius rate law	—	30
Tryptophol	Insomnia	Kinetic model & Arrhenius rate law	COMSOL Multiphysics®	31
Dolutegravir intermediate	HIV	Kinetic model & Arrhenius rate law	COMSOL Multiphysics®	32

2. Experimental

The experimental data required for kinetic modelling is obtained once AZD1775 HMS is synthesised from a feedstock of pyrimidine (Pyr) and bromopyridine (PyBr) in the presence of potassium acetate (KOAc), using copper(I) iodide (Cu(I)I) and racemic ± *trans*-*N*,*N*′-dimethylcyclohexane-1,2-diamine (CyDMEDA) acting as a catalyst and ligand, respectively (Fig. 2).

The chemical synthesis protocol involves first weighing appropriate amounts of solids (*i.e.* Pyr, PyBr, KOAc, Cu(I)I and CyDMEDA) into vials, before purging them with nitrogen, and adding the reaction solvent (degassed MeCN) to these reagents under inert conditions only. Consequently, an Amigo Chem® workstation (Fig. 3) has been used to carry out the synthesis reaction, over a range of reaction temperatures (338.15, 348.15, and 355.15 K) and initial reagent concentrations (*cf.* ESI,† Table S1). High Performance Liquid Chromatography (HPLC) has been used to determine the concentration-time profile of each reagent during these reactions (hereafter, referred to as experiments a–n), measuring the concentration of species Pyr, PyBr, PyI and HMS over the course of the reaction. For the final experimental run (experiment 'n'), Pyr concentration was not recorded.

Key properties of these compounds are provided in Table 3.

3. Reaction mechanism & kinetic model development

To develop a reaction mechanism which can accurately describe AZD1775 HMS synthesis, inspiration has been taken from a similar reaction already studied by Strieter *et al.*[67] (specifically, the copper-catalysed *N*-arylation of amides), since HMS synthesis occurs *via* Ullmann-type coupling.[68,69] Herein, we have adapted the reaction mechanism presented by these authors to allow the consideration of several characteristics unique to HMS synthesis. These include:

1. The production of an aryl iodide compound (PyI), due to the presence of a side reaction involving PyBr (Fig. 4).

2. The production of acetic acid (HOAc), which can hinder the progress of the reaction (Fig. 4).

3. The presence of potassium acetate (KOAc), which may influence reaction rates (Fig. 4).

Table 2 Software packages for kinetic parameter estimation (CT = concentration-time data; AT = absorbance-time data; V = various data types)

	Software elements	Required packages	Input data	Open source
Dynochem[42]	• ODE solvers (*e.g.*, Rosenbrock) • Local-optimisation algorithms (*e.g.*, Levenberg–Marquardt) and search tools (*e.g.*, multiple start-point)	• Dynochem®	V	✗
MATLAB[43]	• ODE solvers (*e.g.*, Rosenbrock, Runge–Kutta, variable step variable order solvers) • Various optimisation algorithms (*e.g.*, Levenberg–Marquardt, trust-region-reflective, simulated annealing, Nelder–Mead, genetic algorithm, particle swarm, surrogate polynomial optimisation, pattern search) and search tools (*e.g.*, multiple start-point)	• MATLAB® • Global optimisation toolbox	V	✗
gPROMS[44,45]	• ODE solvers (*e.g.*, DAEBDF, or DASOLV variable step variable order backward differentiation formulae, SRADAU variable step Runge–Kutta) • Local-optimisation algorithms (*e.g.*, maximum likelihood) and search tools (*e.g.*, multiple start-point)	• gPROMS® Process or Formulated Products	V	✗
SciPy[46]	• ODE solvers (*e.g.*, Runge–Kutta, variable step variable order solvers) • Local-optimisation algorithms (*e.g.*, Nelder–Mead, trust region reflective, Newton-CG, sequential least squares programming)	• Python • SciPy • NumPy	V	✓
KIPET[24,47,48]	• ODE solvers (*e.g.*, orthogonal collocation on finite elements) • Local-optimisation algorithms (*e.g.*, maximum likelihood) and search tools (*e.g.*, multiple start-point)	• Python • Pyomo • SciPy • NumPy • KIPET	CT AT	✓
GEKKO[49]	• ODE solvers (*e.g.*, orthogonal collocation on finite elements) • Local optimisation algorithms and Hyperopt search tools (*e.g.*, grid search, random search, tree-structured parzen estimator, adaptive tree-structured parzen estimator)	• Python • GEKKO • NumPy	V	✓
GDOC[33,35,36,50]	• ODE solvers (*e.g.*, CVODES) • Local-optimisation algorithms (*e.g.*, sequential quadratic programming *via* NPSOL, *etc.*) and search tools (*e.g.*, branch and bound algorithm coupled with convex relaxation considerations)	• Fortran, C, or C++ • NPSOL • Any ANSI compliant Fortran, C, or C++ compiler	CT	✗
COMSOL[51,52]	• ODE solvers and optimisation modules (*e.g.*, optimisation module, LiveLink™ for MATLAB) • Local-optimisation algorithms (*e.g.*, Levenberg–Marquardt, bound optimisation BY quadratic approximation)	• COMSOL Multiphysics®	V	✗

Using this approach, we have defined three fundamental regimes under which different kinetic models can be developed to describe HMS synthesis (Fig. 4). These are:

- Case 1: AZD1775 HMS synthesis disregarding PyI production.
- Case 2: AZD1775 HMS synthesis considering irreversible PyI production.
- Case 3: AZD1775 HMS synthesis considering reversible PyI production.

Each of these cases can be described using systems of differential algebraic equations (DAEs), as shown in Table 4, where R_1 and R_2 denote HMS synthesis from PyBr and PyI, respectively, whilst R_3 and R_4 denote the production and reversible consumption of PyI, respectively (Fig. 4). To summarise their differences, we note that in Case 1 the production of PyI from PyBr is completely ignored, allowing us to treat PyBr and PyI as a single pseudo-aryl halide (PyHal). Conversely, in cases 2 and 3, we explicitly consider

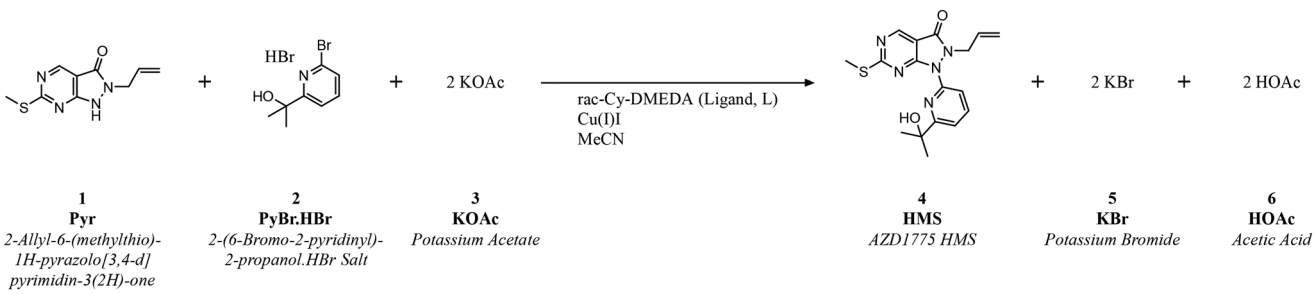

Fig. 2 The overall reaction scheme for AZD1775 hydroxymethylsulfanyl (HMS) batch synthesis.

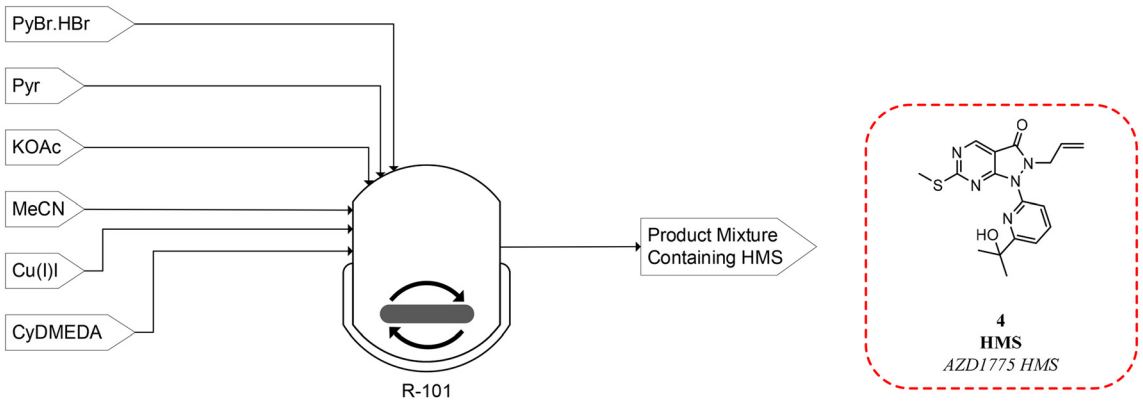

Fig. 3 Schematic representation of a magnetically stirred Amigo Chem® batch reactor used for AZD1775 HMS synthesis.

Table 3 Properties of compounds used in AZD1775 HMS synthesis

Species	Role	#CAS	Chemical formula	M [g mol^{-1}]	T_{bp} [K]	T_{mp} [K]
Pyr	Reagent	955368-90-8	$C_9H_{10}N_4OS$	222.27	671.05[a]	—
PyBr·HBr	Reagent	—	$C_8H_{11}Br_2NO$	296.99	—	—
PyBr	Reagent	638218-78-7	$C_8H_{10}BrNO$	216.08	546.35[a]	339.36[a]
HBr	Reagent	10035-10-6	HBr	80.91	206.15[b]	186.15[b]
KOAc	Reagent	127-08-2	$C_2H_3KO_2$	98.14	—	565.15[b]
PyI	Intermediate	—	$C_8H_{10}INO$	263.08	—	—
KBr	By-product	7758-02-3	KBr	119.00	1708.15[b]	1003.15[b]
HOAc	By-product	64-19-7	$C_2H_4O_2$	60.05	391.15[b]	289.85[b]
HMS	Product	955369-56-9	$C_{17}H_{19}N_5O_2S$	357.43	838.75[a]	—
Cu(I)I	Catalyst	7681-65-4	CuI	190.45	1563.15[a]	878.15[a]
CyDMEDA	Ligand	67579-81-1	$C_8H_{18}N_2$	142.25	459.95[a]	283.38[a]
MeCN	Solvent	75-05-8	C_2H_3N	41.05	354.82[b]	228.15[b]

[a] Ref. 65. [b] Ref. 66.

rate laws for PyI production *via* irreversible (Case 2) and reversible (Case 3) halide substitution reactions. Thereby, we allow for PyBr and PyI to both react with Pyr independently in order to produce HMS.

Within each case different rate laws describe:

1. The mechanism by which HMS synthesis occurs (from PyBr, PyI or the pseudo-aryl halide PyHal).
2. The influence of HOAc on reaction rate.
3. The impact of KOAc concentration on reaction rate.

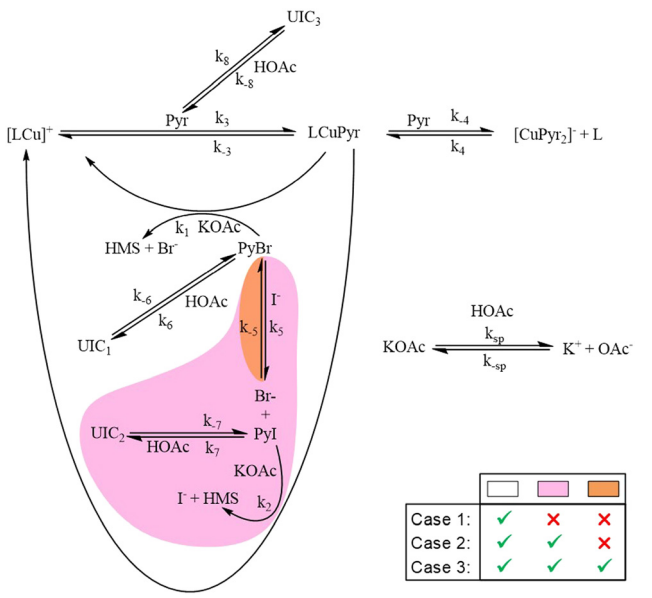

Fig. 4 Proposed reaction mechanism for AZD1775 HMS synthesis: UIC$_i$ denotes unidentified compounds due to inhibitory HOAc action.

Table 4 System of DAEs corresponding to each case developed

Case 1	Case 2	Case 3
$\frac{d[Pyr]}{dt} = -R_1$	$\frac{d[Pyr]}{dt} = -R_1 - R_2$	$\frac{d[Pyr]}{dt} = -R_1 - R_2$
$\frac{d[PyrBr]}{dt} = -R_1$	$\frac{d[PyBr]}{dt} = -R_1 - R_3$	$\frac{d[PyBr]}{dt} = -R_1 - R_3 + R_4$
$\frac{d[KOAc]}{dt} = -2R_1$	$\frac{d[PyI]}{dt} = R_3 - R_2$	$\frac{d[PyI]}{dt} = R_3 - R_2 - R_4$
$\frac{d[HOAc]}{dt} = 2R_1$	$\frac{d[KOAc]}{dt} = -2R_1 - 2R_2$	$\frac{d[KOAc]}{dt} = -2R_1 - 2R_2$
$\frac{d[KBr]}{dt} = 2R_1$	$\frac{d[HOAc]}{dt} = 2R_1 + 2R_2$	$\frac{d[HOAc]}{dt} = 2R_1 + 2R_2$
$\frac{d[HMS]}{dt} = R_1$	$\frac{d[KBr]}{dt} = 2R_1 + 2R_2$	$\frac{d[KBr]}{dt} = 2R_1 + 2R_2$
	$\frac{d[HMS]}{dt} = R_1 + R_2$	$\frac{d[HMS]}{dt} = R_1 + R_2$

4. PyI production.

Consequently, candidate rate laws for each of these aspects have been covered in section 3.1 for each case.

3.1. Candidate rate laws for cases 1–3

Full kinetic models developed for each case have been provided in the ESI† (Case 1: Table S2, Case 2: Table S3, and Case 3: Table S4). The individual rate laws used in each of these kinetic models are discussed in the following sections.

3.1.1. AZD1775 HMS synthesis disregarding PyI production (Case 1). The production of HMS occurs *via* Ullmann-type coupling as per the study of Strieter *et al.*[67] Thus, assuming PyBr and PyI can be treated as a single compound (PyHal), HMS synthesis may be described using the rate law presented in eqn (1) and (2).

$$R_1 = k_1[\mathrm{LCuPyr}][\mathrm{PyHal}] \quad (1)$$

$$[\mathrm{LCuPyr}] = \frac{K_3K_4[\mathrm{CuI}]_0[\mathrm{L}][\mathrm{Pyr}]}{K_3[\mathrm{Pyr}]^2 + K_4[\mathrm{L}] + K_3K_4[\mathrm{Pyr}][\mathrm{L}]} \quad (2)$$

Moreover, we postulate that potassium acetate (KOAc) may impact reaction rate. Thus, an alternative rate law for HMS synthesis is defined, to account for this possibility, as shown in eqn (3).

$$R_1 = k_1[\mathrm{LCuPyr}][\mathrm{PyHal}](K_{sp}[\mathrm{KOAc}])^{\alpha} \quad \text{where: } \alpha \in [0, 1] \quad (3)$$

HMS may also be produced *via* a two-step Ullmann-type coupling reaction (Fig. 5). Hence, it is also possible that the reaction proceeds with 2nd order dependence on the ligated Pyrimidine compound (LCuPyr);[68,69] leading to the development of a second candidate rate law, as shown in eqn (4).

$$R_1 = k_1[\mathrm{LCuPyr}]^2[\mathrm{PyHal}](K_{sp}[\mathrm{KOAc}])^{\alpha} \quad \text{where: } \alpha \in [0, 1] \quad (4)$$

Beyond these considerations, it is also important to consider the possibility that acetic acid (HOAc) impacts the rate of reaction. Hence, reversible processes are defined, which temporarily take reactants off-cycle from the main reaction to model this effect (eqn (5) and (6)).

$$[\mathrm{PyHal}] = \frac{[\mathrm{PyHal}]_{\mathrm{total}}}{1 + [\mathrm{HOAc}]K_6} \quad (5)$$

$$[\mathrm{Pyr}] = \frac{[\mathrm{Pyr}]_{\mathrm{total}}}{1 + [\mathrm{HOAc}]K_8} \quad (6)$$

The nature of the compounds produced by these side processes (eqn (5) and (6)) is unknown. Thus, we have simply represented them as UIC_1 and UIC_3 in this study (Fig. 4).

Considering each of these options, it is clear that either eqn (3) or (4) can be used to describe HMS synthesis, by setting α to 0 or 1. Moreover, eqn (5) and (6) can be used to augment eqn (3) and (4) to account for the effects of HOAc. Consequently, a total of 16 kinetic models can be built for Case 1 – full details of which can be found in the ESI† (Table S2). These models are named 1–16.

3.1.1.1. Calculating ligand concentration. To calculate the free ligand concentration required for each model, according to eqn (2), we define its concentration in terms of observable species only. To do this, we note that the concentration of the catalyst and free-ligand may be defined in terms of the Cu-bound and ligated species as per eqn (7) and (8): when combined and rearranged, these yield eqn (9).

$$[\mathrm{L}]_{\mathrm{total}} = [\mathrm{L}] + [\mathrm{LCu}]^+ + [\mathrm{LCuPyr}] \quad (7)$$

$$[\mathrm{CuI}]_{\mathrm{total}} = [\mathrm{CuI}]_0 = [\mathrm{LCu}]^+ + [\mathrm{LCuPyr}] + [\mathrm{CuPyr_2}]^- \quad (8)$$

$$[\mathrm{L}] = [\mathrm{L}]_{\mathrm{total}} - [\mathrm{CuI}]_0 + [\mathrm{CuPyr_2}]^- \quad (9)$$

Accordingly, we use Fig. 4 to define eqn (10), before combining it with eqn (2) and (9) to produce eqn (11) (after extensive algebraic rearrangement).

$$K_4 = \frac{[\mathrm{LCuPyr}][\mathrm{Pyr}]}{[\mathrm{CuPyr_2}]^-[\mathrm{L}]} \quad (10)$$

Eqn (11) is a simple quadratic equation, solvable at each timestep for the reaction.

$$(K_3K_4[\mathrm{Pyr}] + K_4)[\mathrm{L}]^2 + (K_3[\mathrm{Pyr}]^2 + K_3K_4[\mathrm{CuI}]_0[\mathrm{Pyr}] + K_4[\mathrm{CuI}]_0 - K_4[\mathrm{L}]_{\mathrm{total}} - K_3K_4[\mathrm{Pyr}][\mathrm{L}]_{\mathrm{total}})[\mathrm{L}] - K_3[\mathrm{Pyr}]^2[\mathrm{L}]_{\mathrm{total}} = 0 \quad (11)$$

3.1.2. AZD1775 HMS synthesis considering irreversible PyI production (Case 2). For Case 2, HMS synthesis is modelled assuming Ullmann-type coupling, just as in Case 1. Herein, however, we also consider PyI formation, assuming that it is produced irreversibly from PyBr. Consequently, HMS production for Case 2 is defined using eqn (12) and (13) to describe HMS synthesis from PyBr (eqn (12): one-step coupling, eqn (13): two-step coupling), whilst eqn (14) and (15) are used to describe its production from PyI (eqn (14): one-step coupling, eqn (15): two-step coupling).

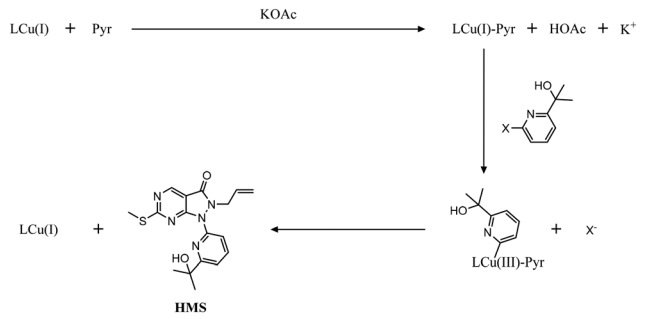

Fig. 5 Ullmann-type coupling[68,69] hypothesis for AZD1775 HMS formation.

$$R_1 = k_1[\text{LCuPyr}][\text{PyBr}](K_{sp}[\text{KOAc}])^{\alpha} \quad \text{where: } \alpha \in [0, 1] \quad (12)$$

$$R_1 = k_1[\text{LCuPyr}]^2[\text{PyBr}](K_{sp}[\text{KOAc}])^{\alpha} \quad \text{where: } \alpha \in [0, 1] \quad (13)$$

$$R_2 = k_2[\text{LCuPyr}][\text{PyI}](K_{sp}[\text{KOAc}])^{\alpha} \quad \text{where: } \alpha \in [0, 1] \quad (14)$$

$$R_2 = k_2[\text{LCuPyr}]^2[\text{PyI}](K_{sp}[\text{KOAc}])^{\alpha} \quad \text{where: } \alpha \in [0, 1] \quad (15)$$

Following these definitions, expressions (as per eqn (5) and (6) of Case 1) are defined to describe the inhibitory effect of HOAc on PyBr, PyI and Pyr independently (eqn (16)–(18)).

$$[\text{PyBr}] = \frac{[\text{PyBr}]_{\text{total}}}{1 + [\text{HOAc}]K_6} \quad (16)$$

$$[\text{PyI}] = \frac{[\text{PyI}]_{\text{total}}}{1 + [\text{HOAc}]K_7} \quad (17)$$

$$[\text{Pyr}] = \frac{[\text{Pyr}]_{\text{total}}}{1 + [\text{HOAc}]K_8} \quad (18)$$

A variety of rate laws are tested for the production of PyI from PyBr, all of which are reminiscent of those in other areas of chemistry where halogenic substitution takes place.[70–79] For example, it is noted that some studies[71–73] have found that halogenic substitution reactions involving pyridines follow second-order rate laws (which are first-order in both the substrate and solvent). Moreover, a recent study by Kundu *et al.*[74] explored reaction mechanisms underpinning ligand-assisted substitutions involving pyridines and their derivatives, concluding that they proceed *via* third-order reactions (first-order in catalyst and second-order in aryl halide substrate).[74] Other authors[70,71] have suggested that such reactions proceed with rate laws of non-integer order (*i.e.*, between 1 and 2). Consequently, two variable-order rate laws are hereby tested to allow each of these possibilities to be examined simultaneously, limiting the reaction order with respect to PyBr to between 1 and 2 (eqn (19) and (20)). This is a critical problem simplification, removing the need to explicitly quantify acetonitrile and copper(I) iodide concentrations, whilst defining β as an extra kinetic parameter to be estimated.

$$R_3 = k_5'[\text{PyBr}]^{\beta} \quad \text{where: } k_5' = k_5[\text{MeCN}] \approx k_5[\text{CuI}] \quad (19)$$

$$R_3 = k_5[\text{CuI}]_0[\text{PyBr}]^{\beta} \quad (20)$$

This approach results in a total of 32 models for Case 2 (named models 17–48), with details in the ESI† (Table S3).

3.1.3. AZD1775 HMS synthesis considering reversible PyI production (Case 3). For Case 3, we retain eqn (2) from Case 1 and eqn (12)–(18) from Case 2 to describe the production of HMS from PyBr and PyI. However, we replace our treatment of PyI to consider its reversible production from PyBr. Specifically, this is postulated to occur *via* a reversible copper-catalysed Finkelstein reaction (eqn (21) and (22)).[75–79]

$$R_3 = k_5[\text{CuI}]_0[\text{PyBr}][\text{I}^-] \quad \text{where: } [\text{I}^-] \approx [\text{I}^-]_0 + [\text{CuI}]_0 - [\text{PyI}] \quad (21)$$

$$R_4 = k_{-5}[\text{KBr}][\text{PyI}] \quad (22)$$

This leads to the development of 16 kinetic models for Case 3, resulting in a total of 64 models across all three cases. Consequently, the kinetic models arising from Case 3 are named models 49–64. A full description of each model developed as part of Case 3 can be found in the ESI† (Table S4).

3.2. Kinetic rate law temperature dependence

To capture the temperature dependence of each reaction step, Arrhenius and van't Hoff relationships are embedded within each of the kinetic models proposed (eqn (23)–(25)).

$$k_n = k_{n,\text{ref}} e^{\frac{-E_{a,n}}{R}\left(\frac{1}{T} - \frac{1}{T_{\text{ref}}}\right)} \quad (23)$$

$$K_n = K_{n,\text{ref}} e^{\frac{-\Delta H_n^0}{R}\left(\frac{1}{T} - \frac{1}{T_{\text{ref}}}\right)} \quad (24)$$

$$K_n = \frac{k_n}{k_{-n}} \quad (25)$$

The temperature dependence of the solubility of KOAc within HOAc[80] has also been estimated using the van't Hoff equation (eqn (26) and (27))[81] (Fig. 6).

$$K_{sp} = x_{\text{KOAc}}{}^2 \quad (26)$$

$$\ln(x_{\text{KOAc}}) = -\frac{\Delta H_d^0}{RT} + \frac{\Delta S_d^0}{R} \quad (27)$$

4. Parameter estimation & model discrimination

Parameter estimation and model discrimination methods are hereby used in tandem to identify the highest-fidelity kinetic model for HMS synthesis from the 64 models developed in section 3. Consequently, the computational procedure is

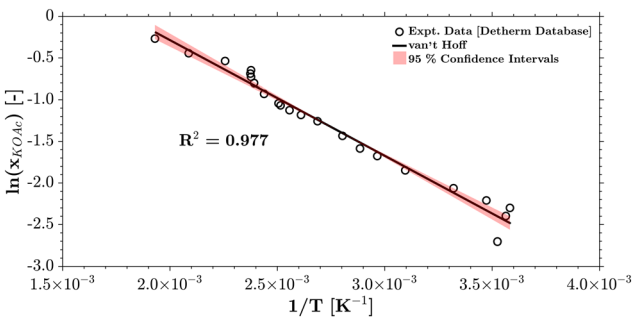

Fig. 6 Temperature-dependent KOAc solubility in HOAc (data: ref. 80).

summarised in Fig. 7, and sections 4.1 and 4.2 are provided below to offer detailed explanations of each of the blocks in the software architecture illustrated in this diagram.

Each case (and model) has been studied independently using multistart nonlinear programming (NLP). Hence, parameter estimation efforts are conducted for each model individually, and model discrimination methods are subsequently employed to analyse the results from each of these studies.

4.1. Kinetic parameter estimation methodology

The parameter estimation problem is formulated as a multistart constrained optimisation problem for each model, using the objective function and constraints of eqn (28)–(30).

$$\min f(\boldsymbol{\theta}) \tag{28}$$

$$\text{s.t.} \begin{cases} f(\boldsymbol{\theta}) = \text{WLS} \\ \boldsymbol{\theta} = \left[k_{n,\text{ref}},\, E_{\text{a},n},\, \Delta H_n^0,\, K_{n,\text{ref}},\, \Delta H_\text{d}^0,\, \Delta S_\text{d}^0,\, \beta\right] \\ \boldsymbol{\theta}_\text{lb} \le \boldsymbol{\theta} \le \boldsymbol{\theta}_\text{ub} \\ \boldsymbol{\theta} \in \mathbb{R} \end{cases} \tag{29}$$

$$\text{WLS} = \frac{1}{2} \sum_{i=1}^{N_\text{species}} \sum_{j=1}^{N_\text{data}} W_i \left(C_{i,j}^\text{expt} - C_{i,j}^\text{model}\right)^2 \tag{30}$$

Using this approach, parameter estimations are initialised from 1000 different start-points for each model, using 1000 different random guess-vectors containing different parameter values within the specified bounds (Table 5). To avoid divergence, parameter estimations resulting in non-numeric, infinite or imaginary objective function values must be rejected, forcing the optimisation algorithm to modify the search vector space.

A detailed discussion of the code structure is provided in section 4.1.1 below. The kinetic parameters and reaction orders required by each model are estimated using an in-house parameter estimation code written in MATLAB® (Fig. 7), employing its *fminsearch* command (which implements the Nelder–Mead algorithm) at the centre of a while loop, so as to iteratively minimise the sum of weighted least squares (eqn (30)) for each model: this occurs by setting weights (W_i) to the reciprocal of the square uncertainties of experimental measurements, whilst also placing user-specified bounds on each of the parameters to be estimated (Table 5).

To enable this approach, it has been stipulated that kinetic rate constants within the kinetic models tested can take values between 0–6 × 10^{11}, since liquid-phase bimolecular reactions do not exceed rates of 6 × 10^{11} M^{-1} min^{-1}.[82–86]

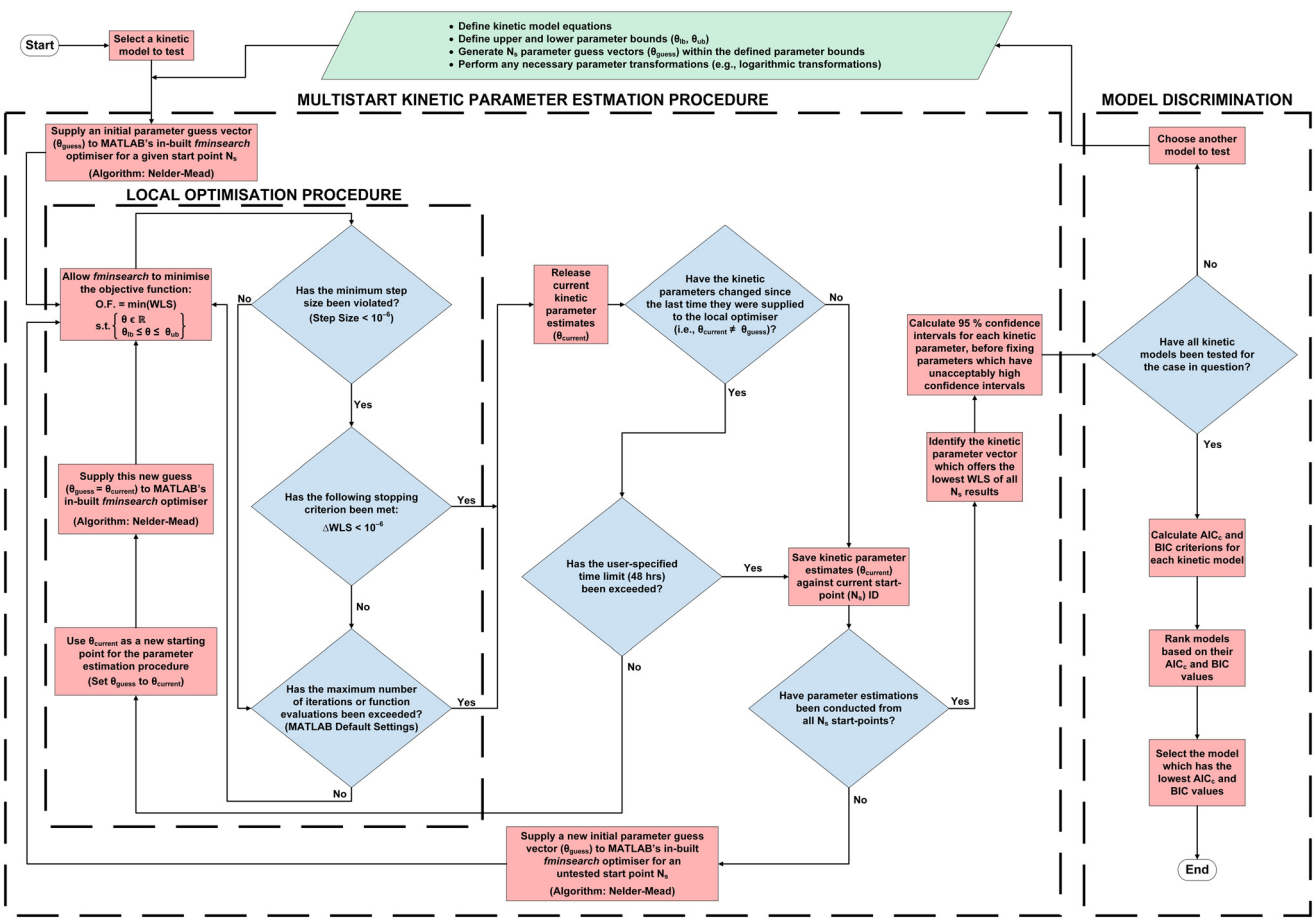

Fig. 7 The computational procedure for kinetic parameter estimation and reaction model discrimination, implemented in MATLAB®.

Table 5 Parameter bounds

	$k_{n,\mathrm{ref}}$ [various]	$K_{n,\mathrm{ref}}$ [various]	$E_{\mathrm{a},n}$ [kJ mol^{-1}]	ΔH_n^0 [kJ mol^{-1}]	ΔH_d^0 [kJ mol^{-1}]	ΔS_d^0 [kJ mol^{-1} K^{-1}]	β [—]
Lower bound ($\boldsymbol{\theta}_\mathrm{lb}$)	10^{-10}	0.01	1	−1000	10.702	0.0184	1
Upper bound ($\boldsymbol{\theta}_\mathrm{ub}$)	6×10^{11}	100	150	1000	12.365	0.0230	2

Likewise, it has been specified that activation energies ($E_{\mathrm{a},n}$) take values between 1–150 kJ mol^{-1}, to align with data from classical organic synthesis reactions.[87] Moreover, it has been acknowledged that standard reaction enthalpies (ΔH_n^0) for organic reactions are within ±1000 kJ mol^{-1}, whilst their equilibrium constants (K_n) have values of 0.01–100, whenever equilibrium concentrations are measurable.[88] Reaction orders, β, must be bounded between 1 and 2 (for the reasons discussed in section 3), and KOAc dissolution properties (ΔH_d^0 and ΔS_d^0) are set within limits equal to their 95% confidence intervals (as per Fig. 6).

Given the range of each of these bounds, the search space for each problem is reduced by natural logarithm transforms of parameters $k_{\mathrm{ref},n}$, $E_{\mathrm{a},n}$ and K_ref, as per recent studies.[89,90] This approach is not possible for ΔH_n^0 parameters, as it would result in imaginary solutions with no physical meaning (it is also not required for β parameters due to range brevity).

4.1.1. Code structure. Parameter estimation has been carried out for each kinetic model independently (*i.e.*, models 1–64) using original MATLAB® code based on the next concepts:

1) *Local optimisation*: MATLAB®'s *fminsearch* command (implementing the Nelder–Mead algorithm) is used to minimise the objective function of each model (eqn (28)–(30)) by varying the value of their model parameters ($\boldsymbol{\theta}$) within their specified bounds. Moreover, it continues to do so until one of the following criteria is met:

a) The minimum step size (10^{-6}) has been violated, and the objective function cannot be improved upon by more than 10^{-6}.

b) The maximum number of iterations or objective function evaluations is exceeded (MATLAB default settings used).

Once one of these criteria is met, *fminsearch* releases any parameter estimates held ($\boldsymbol{\theta}_\mathrm{current}$). The algorithm then checks whether one of the following criteria has also been satisfied:

c) The kinetic parameters have not changed compared to the values last supplied to the optimiser.

d) The user-specified maximum allowable time for estimations (48 h) has been exceeded.

If one of (c) or (d) has been satisfied, then the current parameter estimates ($\boldsymbol{\theta}_\mathrm{current}$) are saved against their associated start-point identification number (N_s) for later analysis. However, if neither of these criteria have been achieved, then the output parameter estimates ($\boldsymbol{\theta}_\mathrm{current}$) are provided to *fminsearch* again, as a new initial guess. This serves to enhance convergence by repeating step 1 until (a)–(b), plus one of (c) or (d), are satisfied (Fig. 7).

2) *Multistart problem initialisation*: to improve estimations, step 1 is repeated using 1000 different initial guesses ($\boldsymbol{\theta}_\mathrm{guess}$) before selecting the set of parameters which yield the lowest objective function value. Following this, the 95% confidence intervals are calculated for the selected parameter set, and parameters which give unacceptably high confidence intervals (*i.e.*, greater than 15%[23,24,48,56]) should be fixed[39,91] (Fig. 7).

The parameter estimation code used can be run in series or parallel – depending on the computing facilities available. Running parameter estimations in parallel (splitting jobs across multiple CPU/GPU cores) will invariably provide faster results, as computational cost scales linearly with the number of starting points used.

4.2. Model discrimination

Following parameter estimation, the Akaike Information criterion ($\mathrm{AIC_c}$) is used to rank each model based on its: (i) simplicity, and (ii) fidelity *vs.* experimental observations[23,56] (eqn (31)). Consequently, models with lower $\mathrm{AIC_c}$ values are favoured, since this indicates accurate reproduction of experimental findings whilst avoiding the use of unnecessary terms, thus preventing overfitting and unjustifiable over-parameterisation.[23]

$$\mathrm{AIC_c} = N_\mathrm{obs} \ln\left(\frac{\mathrm{WLS}}{N_\mathrm{obs}}\right) + 2N_\mathrm{param} + \frac{2N_\mathrm{param}\left(N_\mathrm{param}+1\right)}{N_\mathrm{obs}-N_\mathrm{param}-1} \quad (31)$$

Relative Akaike likelihood metrics (eqn (32) and (33)) are simultaneously used to determine the probability that the model with the lowest $\mathrm{AIC_c}$ for a given case is indeed better than all other model candidates available for the same case. Evidence ratios (eqn (34)) and normalized probabilities (eqn (35)) are also calculated for each model, to assess the likelihood that the selected model is better than the next-best model (*i.e.* the probability that model i is more suitable compared to model j).[58,60,92,93]

$$w_{i,\mathrm{AIC_c}} = \frac{\exp\left(-\frac{1}{2}\Delta(\mathrm{AIC_c})_i\right)}{\sum_{j=1}^{N_\mathrm{mdl}} \exp\left(-\frac{1}{2}\Delta(\mathrm{AIC_c})_j\right)} \quad (32)$$

$$\Delta(\mathrm{AIC_c})_i = (\mathrm{AIC_c})_i - (\mathrm{AIC_c})_\mathrm{min} \quad (33)$$

$$(\mathrm{ER})_{\mathrm{AIC_c}} = \frac{w_{i,\mathrm{AIC_c}}}{w_{j,\mathrm{AIC_c}}} \quad (34)$$

$$(\mathrm{NP})_{\mathrm{AIC_c}} = \frac{w_{i,\mathrm{AIC_c}}}{w_{i,\mathrm{AIC_c}} + w_{j,\mathrm{AIC_c}}} \quad (35)$$

The Bayesian Information Criterion (BIC) is also estimated for each model *via* eqn (36) to corroborate this analysis, by calculating normalised probabilities as done for AIC_c. Hence, lower BIC values represent more promising models (higher fidelity), with BIC penalising model complexity more harshly than AIC_c.[59]

$$BIC = N_{obs} \ln\left(\frac{WLS}{N_{obs}}\right) + N_{param} \ln(N_{obs}) \quad (36)$$

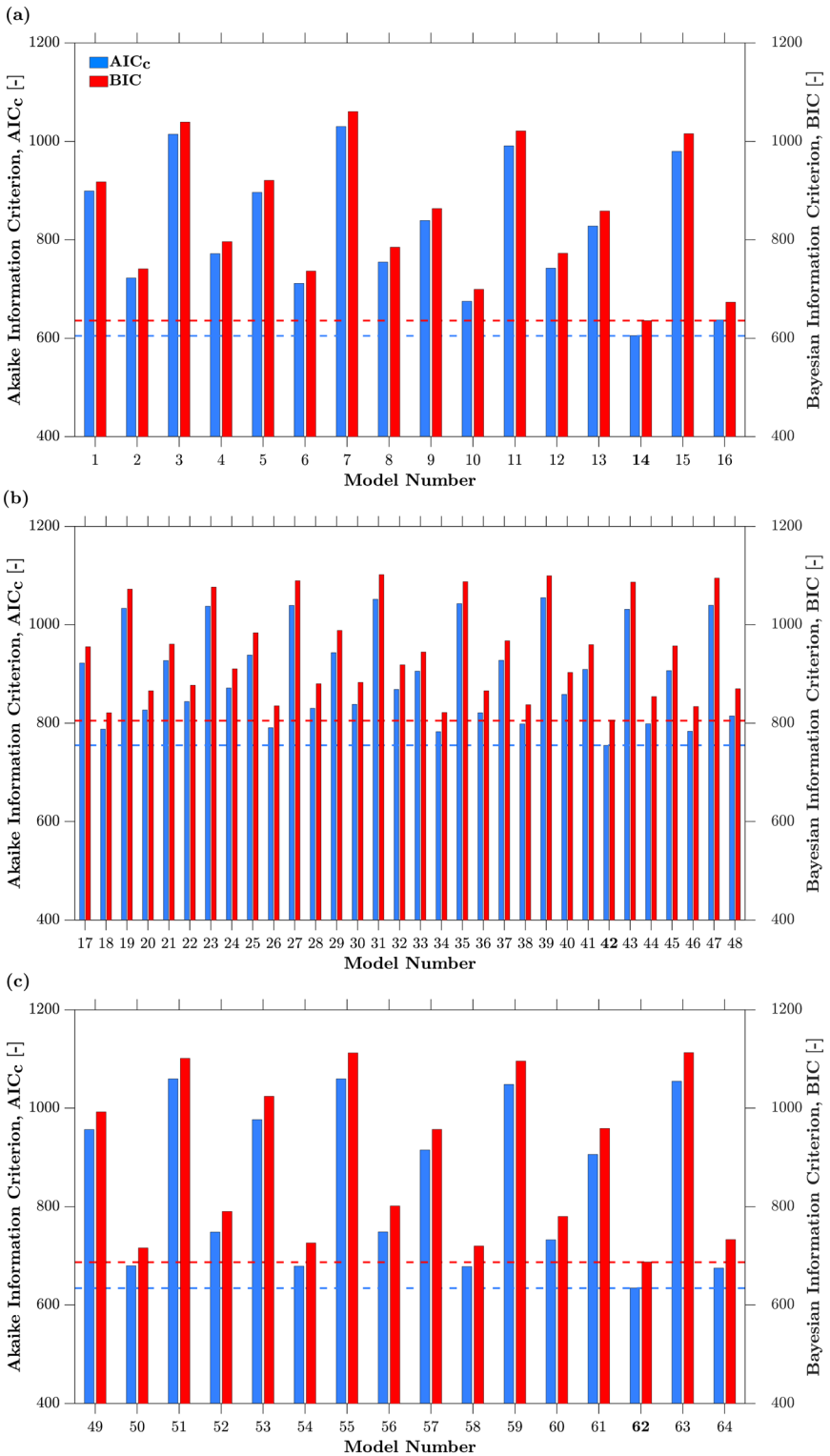

Fig. 8 AIC_c and BIC values for: (a) Case 1 (no Pyl production); (b) Case 2 (irreversible Pyl production); and (c) Case 3 (reversible Pyl production).

5. Results & discussion

Three distinct cases are considered for kinetic models describing HMS synthesis:

- Case 1: AZD1775 HMS synthesis disregarding PyI production.
- Case 2: AZD1775 HMS synthesis considering irreversible PyI production.
- Case 3: AZD1775 HMS synthesis considering reversible PyI production.

Consequently, Fig. 8 provides an overview of AIC_c and BIC values for every single model, ranking all 64 models hereby considered within these three cases.

Models 14, 42 and 62 are determined to be the most promising ones from Cases 1, 2 and 3, respectively. The normalised AIC_c and BIC probabilities both sit at 1.0 for each of these models when compared with the next-best model from their respective cases. Moreover, we observe that among models considering PyI production (Cases 2 and 3), model 62 outperformed model 42 by a significant margin, as it exhibited WLS, AIC_c and BIC values 50.10, 16.03 and 14.66% lower than the latter, respectively. Hence, model 42 is rejected, rendering model 62 as the most reliable *vs.* all other models describing PyI production.

The reaction profiles predicted by models 14 and 62 are compared with experimental data in Fig. 9 and 10, respectively; final kinetic parameters for both these models are provided in Table 6. Both models 14 and 62 provide looser fits for experiment "*l*" than for the rest of the experiments, most likely due to undercalculation of HMS concentrations for this experiment only (possibly caused by non-uniform mixing), since less HMS was produced than Pyr consumed in this experiment (no other species were observed during HPLC measurements for this experiment). The use of Weighted Least Squares (WLS) regression during our original analysis is able to successfully mitigate the effects of this suspected effect, especially since similar percentage uncertainties are computed for each experimental data point measured – leading to larger absolute experimental uncertainties for species with higher concentrations (*e.g.* HMS in experiment "*l*").

From these findings, we observe that both highest-fidelity models share all their main pathway features, with the sole exception of PyI production treatment (*cf.* model structures in Tables S2 and S4 within the ESI†). Thus, we conclude that each of the aspects accounted for by these models plays a key role in the AZD1775 HMS synthesis mechanism. For example, since the production of HMS from aryl halides and Pyr is second-order in LCuPyr concentration for both models, it is likely that a two-step Ullmann coupling (Fig. 5) occurs instead of a single-step mechanism. Conversely, the characteristics ignored by these models (*e.g.* KOAc concentration dependence) are unlikely to have any significant effect on the HMS synthesis reaction.

Furthermore, we can also conclude that PyI is most likely produced *via* a copper-catalysed Finkelstein reaction, since the kinetic models employing this assumption (those in Case 3) often outperform their equivalent counterparts (those in Case 2). However, further experimentation focused on analysing the production of PyI is required to fully confirm this hypothesis. Similarly, because both highest-fidelity models consider inhibitory effects brought about by the presence of HOAc, acetic acid may indeed inhibit the action of Pyr, PyBr and PyI. The exact mechanism by which this inhibition may occur remains unclear (as per section 3, Fig. 4), however. Consequently, future studies should establish the true mechanism by which this HOAc inhibition occurs – a possible explanation is that the presence of HOAc creates a buffer system impacting the deprotonation of pyrimidine, thus slowing down the oxidative addition of aryl halides (Fig. 5).

6. Conclusions

The present paper performs an original parameter estimation and extensive model discrimination to arrive at a novel reaction mechanism (Fig. 4) which successfully captures all known key features of AZD1775 HMS synthesis, with three broad kinetic model classes (Cases 1–3) and 64 individual models developed, tested and comparatively evaluated for the first time. Specifically, the model collection comprises those addressing HMS synthesis by disregarding PyI production entirely (Case 1: models 1–16), as well as those which considered irreversible (Case 2: models 17–48) and reversible (Case 3: models 49–64) PyI production. Thus, a total of 64 candidate kinetic models have been proposed, parameterised and evaluated to describe the copper-catalysed and ligand-assisted HMS synthesis. Candidates 14 and 62 have the highest model fidelity, demonstrating the most promising results by a significant margin and exhibiting the lowest objective function values and the most favourable Akaike and Bayesian (AIC_c and BIC) metrics following their successful parameterisation. Consequently the authors propose that manufacturing facilities and research groups should use model 14 to model HMS synthesis if PyI production can be ignored (as this model has a far simpler structure and a concise parameter set), but employ model 62 if PyI production is to be considered for explicit quantification.

The original parameter estimation and model development and discrimination method and code proposed herein does not guarantee uniqueness. Consequently, alternative models may yield comparable results to those achieved here, but they also be far more complex than those presented. Accordingly, future studies should focus on using the kinetic models presented here for the first time, to identify optimal reaction conditions for HMS synthesis, by manipulating process variables such as temperature, reaction time and initial reagent concentrations to maximise product yield whilst minimising side product generation. Furthermore, future parameter estimation studies can incorporate parameter identifiability and estimability analyses[94,95] in the respective workflows. Bootstrapping

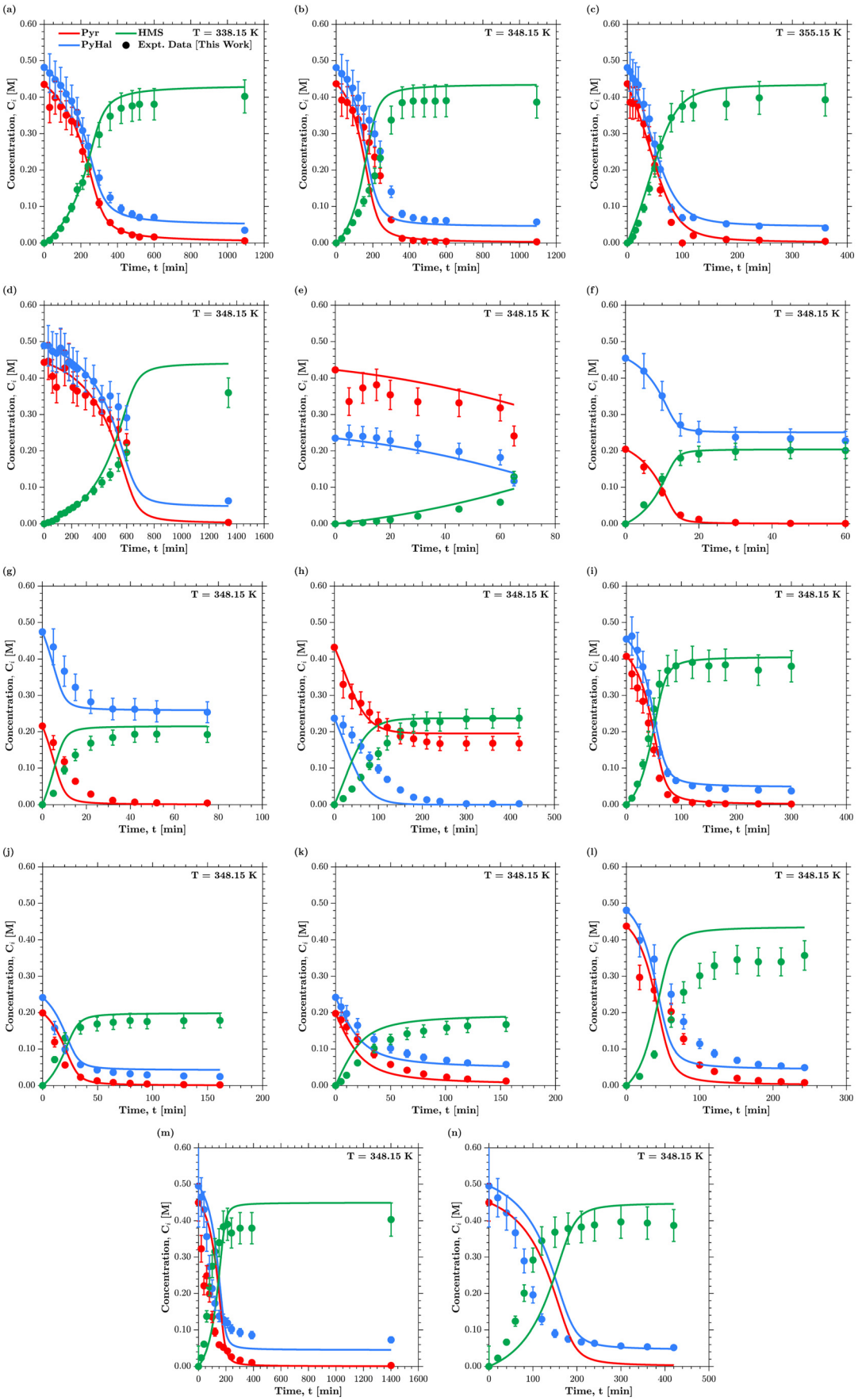

Fig. 9 Kinetic model predictions *vs.* experimental data (model 14).

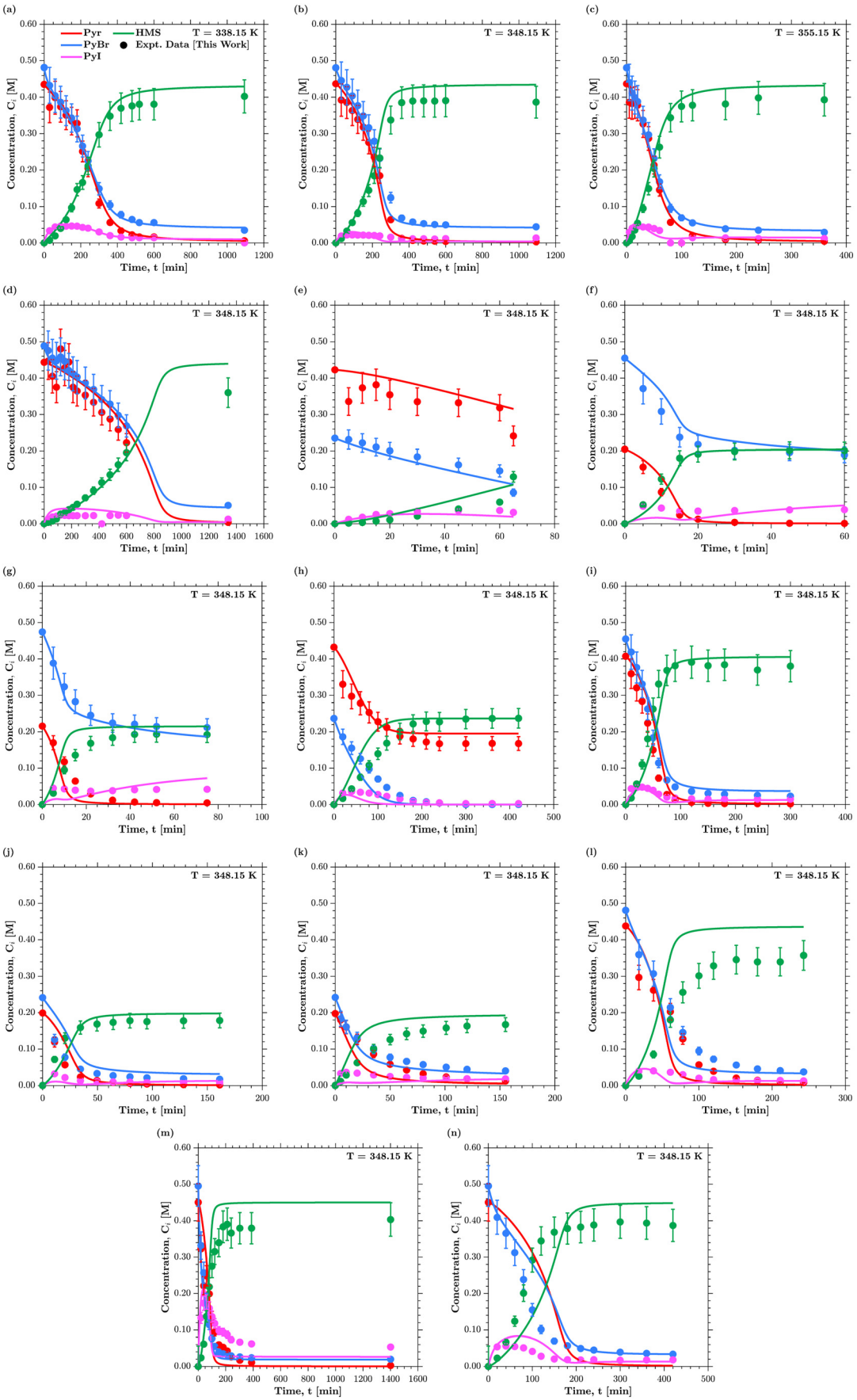

Fig. 10 Kinetic model predictions *vs.* experimental data (model 62).

Table 6 Kinetic parameters (estimated and fixed) for models 14 and 62

	Model 14	Model 62
$k_{1,\mathrm{ref}}$ [M^{-2} min^{-1}]	131.339 (± 1.140%)	852.151 (± 2.164%)
$k_{2,\mathrm{ref}}$ [M^{-2} min^{-1}]	—	12 273.399 (fixed)
$k_{5,\mathrm{ref}}$ [M^{-2} min^{-1}]	—	1.482 (± 0.014%)
$K_{3,\mathrm{ref}}$ [M]	99.993 (± 0.011%)	8.004 (± 0.086%)
$K_{4,\mathrm{ref}}$ [—]	0.166 (± 0.148%)	0.052 (± 0.063%)
$K_{5,\mathrm{ref}}$ [M^{-1}]	—	9.284 (fixed)
$K_{6,\mathrm{ref}}$ [M]	4.112 (± 10.168%)	0.549 (± 2.408%)
$K_{7,\mathrm{ref}}$ [M]	—	6.977 (fixed)
$K_{8,\mathrm{ref}}$ [M]	8.057 (± 6.120%)	2.296 (± 5.839%)
$E_{a,1}$ [kJ mol^{-1}]	146.134 (± 7.539%)	31.272 (± 9.268%)
$E_{a,2}$ [kJ mol^{-1}]	—	1.045 (fixed)
$E_{a,5}$ [kJ mol^{-1}]	—	60.497 (± 3.047%)
ΔH_3^0 [kJ mol^{-1}]	200.464 (± 1.030%)	−309.601 (± 1.146%)
ΔH_4^0 [kJ mol^{-1}]	0.825 (± 1.045%)	9.908 (± 2.465%)
ΔH_5^0 [kJ mol^{-1}]	—	253.532 (fixed)
ΔH_6^0 [kJ mol^{-1}]	404.302 (± 6.580%)	−515.590 (± 3.211%)
ΔH_7^0 [kJ mol^{-1}]	—	−584.127 (± 10.448%)
ΔH_8^0 [kJ mol^{-1}]	102.235 (fixed)	−217.411 (± 8.344%)

methods can also be incorporated into future workflows, especially for codes which may study similarly many kinetic model scenarios under limited experimental data availability.

Nomenclature & acronyms

Latin letters

$C_{i,j}^{\mathrm{expt}}$	Experimental concentration of species i at datapoint j [M]
$C_{i,j}^{\mathrm{model}}$	Modelled concentration of species i at datapoint j [M]
$E_{a,n}$	Activation energy associated with forward reaction n [kJ mol^{-1}]
$f(\boldsymbol{\theta})$	Temperature dependent fitting objective function [—]
ΔH_d^0	Standard enthalpy of dissolution for KOAc in HOAc [kJ mol^{-1}]
ΔH_n^0	Standard reaction enthalpy with reaction n [kJ mol^{-1}]
$[i]$	Molar concentration of component i [M]
$[i]_0$	Initial molar concentration of component i [M]
k_n	Kinetic rate constant associated with forward reaction n [various]
$k_{n,\mathrm{ref}}$	Pre-exponential reference constant for forward reaction n [various]
k_{-n}	Kinetic rate constant associated with reverse reaction n [various]
K_n	Equilibrium rate constant associated with reaction n [various]
K_{sp}	Solubility product [M^2]
M	Molecular weight [g mol^{-1}]
N_{data}	Number of experimental data points [—]
N_{mdl}	Number of models tested [—]
N_{obs}	Number of experimental observations (*i.e.*, time-points) [—]
N_{param}	Number of parameters [—]
N_s	Number of start-points [—]
N_{species}	Number of species [—]
R	Universal gas constant [J mol^{-1} K^{-1}]
R_j	Reaction rate j [mol L^{-1} min^{-1}]
$\mathbb{R}$	Real numbers [—]
ΔS_d^0	Standard entropy of dissolution for KOAc in HOAc [kJ mol^{-1} K^{-1}]
t	Reaction time [min]
T	Reaction temperature [K]
T_{bp}	Boiling point temperature [K]
T_{mp}	Melting point temperature [K]
T_{ref}	Reference temperature [K]
$w_{i,\mathrm{AIC_c}}$	Normalised relative Akaike likelihoods [—]
W_i	Objective function weight associated with measurement i [M^{-2}]
x_{KOAc}	Solubility of KOAc in HOAc at a given temperature [mol mol^{-1}]

Greek letters

α	Binary decision variable [—]
β	Unknown rate order [—]
$\boldsymbol{\theta}$	Parameter vector [various]
$\boldsymbol{\theta}_{\mathrm{current}}$	Parameter vector outputted at interim points during estimations [various]
$\boldsymbol{\theta}_{\mathrm{guess}}$	Initial parameter vector guess for a given start-point [various]
$\boldsymbol{\theta}_{\mathrm{lb}}$	Parameter vector lower bounds [various]
$\boldsymbol{\theta}_{\mathrm{ub}}$	Parameter vector upper bounds [various]

Acronyms

$\mathrm{AIC_c}$	Corrected Akaike's information criterion [—]
$(\mathrm{AIC_c})_i$	Corrected Akaike information criterion of model i [—]
$(\mathrm{AIC_c})_{\mathrm{min}}$	Minimum corrected Akaike information criterion obtained [—]
$\Delta(\mathrm{AIC_c})_i$	Corrected Akaike differences of model i [—]
BIC	Bayesian information criterion [—]
$(\mathrm{ER})_{\mathrm{AIC_c}}$	Corrected Akaike evidence ratios [—]
$(\mathrm{NP})_{\mathrm{AIC_c}}$	Corrected Akaike normalized probabilities [—]
WLS	Weighted least squares [—]

Data availability

A legal confidentiality agreement (2020) is in effect and has been signed between the University of Edinburgh and AstraZeneca plc, governing the publication of results (under the auspices of the Engineering & Phys. Sciences Research Council, EPSRC). Tabulated, literature reference and digitised reaction (concentration) data are all provided and suffice for the reproduction of all Figs./results presented in the article. Specific requests for original (raw) experimental data and/or the software (MATLAB®) code by Mr M. Blair can be considered in writing, to ensure the above is not violated.

Conflicts of interest

The authors report no conflict of interest whatsoever.

Acknowledgements

This study has employed computational resources provided by the University of Edinburgh's Compute and Data Facility (ECDF) (**https://www.ecdf.ed.ac.uk/**). M. Blair acknowledges the financial support provided by the Engineering and Physical Sciences Research Council (EPSRC) and AstraZeneca UK *via* an Industrial CASE (iCASE) PhD Fellowship. Prof. D. I. Gerogiorgis also acknowledges financial support from UKRI/EPSRC (RAPID: ReAl-time Process Modeling and Diagnostics: Powering Digital Factories, EP/V028618/1) as well as a recent Royal Society Short Industrial Fellowship (2020-22), and an ongoing Royal Society International Exchanges Programme grant (IES\R2\232014, 2023-25). All authors acknowledge the critical leadership and support of Dr. Martin Jones, Principal Scientist (Process Design, AstraZeneca U.K.), for the inception and critical support of this research project.

References

1 V. Prasad, K. De Jesus and S. Mailankody, The high price of anticancer drugs: origins, implications, barriers, solutions, *Nat. Rev. Clin. Oncol.*, 2017, **14**(6), 381.
2 B. Jönsson, T. Hofmarcher, P. Lindgren and N. Wilking, The cost and burden of cancer in the European Union 1995–2014, *Eur. J. Cancer*, 2016, **66**, 162.
3 T. Hofmarcher, P. Lindgren, N. Wilking and B. Jönsson, The cost of cancer in Europe 2018, *Eur. J. Cancer*, 2020, **129**, 41.
4 T. Dyba, G. Randi and F. Bray, *et al.*, The European cancer burden in 2020: Incidence and mortality estimates for 40 countries and 25 major cancers, *Eur. J. Cancer*, 2021, **157**, 308.
5 Cancer Research UK, Cancer Statistics for the UK, [cited 2023 23rd March]; Available from: **https://www.cancerresearchuk.org/health-professional/cancer-statistics-for-the-uk**.
6 R. L. Siegel, K. D. Miller, H. E. Fuchs and A. Jemal, Cancer Statistics, 2021, *Ca-Cancer J. Clin.*, 2021, **71**(1), 7.
7 Cancer Research UK, Worlwide cancer statistics, [cited 2023 25th May]; Available from: **https://www.cancerresearchuk.org/health-professional/cancer-statistics/worldwide-cancer#heading-Zero**.
8 S. Diab and D. I. Gerogiorgis, Process Modeling, Simulation, and Technoeconomic Evaluation of Separation Solvents for the Continuous Pharmaceutical Manufacturing (CPM) of Diphenhydramine, *Org. Process Res. Dev.*, 2017, **21**(7), 924.
9 S. Diab and D. I. Gerogiorgis, Process modelling, simulation and technoeconomic evaluation of crystallisation antisolvents for the continuous pharmaceutical manufacturing of rufinamide, *Comput. Chem. Eng.*, 2018, **111**, 102.
10 S. Diab and D. I. Gerogiorgis, Process modelling, simulation and technoeconomic optimisation for continuous pharmaceutical manufacturing of (S)-warfarin, in *Computer Aided Chemical Engineering*, ed. A. Friedl, *et al.*, Elsevier, 2018, p. 1643.
11 S. Diab, D. T. McQuade, B. F. Gupton and D. I. Gerogiorgis, Process Design and Optimization for the Continuous Manufacturing of Nevirapine, an Active Pharmaceutical Ingredient for HIV Treatment, *Org. Process Res. Dev.*, 2019, **23**(3), 320.
12 S. Diab, N. Mytis, A. G. Boudouvis and D. I. Gerogiorgis, Process modelling, design and technoeconomic Liquid–Liquid Extraction (LLE) optimisation for comparative evaluation of batch vs. continuous pharmaceutical manufacturing of atropine, *Comput. Chem. Eng.*, 2019, **124**, 28.
13 O. L. Watson, S. Jonuzaj and J. McGinty, *et al.*, Computer Aided Design of Solvent Blends for Hybrid Cooling and Antisolvent Crystallization of Active Pharmaceutical Ingredients, *Org. Process Res. Dev.*, 2021, **25**(5), 1123.
14 H. G. Jolliffe and D. I. Gerogiorgis, Plantwide design and economic evaluation of two Continuous Pharmaceutical Manufacturing (CPM) cases: Ibuprofen and artemisinin, *Comput. Chem. Eng.*, 2016, **91**, 269.
15 H. G. Jolliffe and D. I. Gerogiorgis, Technoeconomic Optimization of a Conceptual Flowsheet for Continuous Separation of an Analgaesic Active Pharmaceutical Ingredient (API), *Ind. Eng. Chem. Res.*, 2017, **56**(15), 4357.
16 M. A. McDonald, G. D. Marshall and A. S. Bommarius, *et al.*, Crystallization Kinetics of Cephalexin Monohydrate in the Presence of Cephalexin Precursors, *Cryst. Growth Des.*, 2019, **19**(9), 5065.
17 A. B. Cuthbertson, A. D. Rodman, S. Diab and D. I. Gerogiorgis, Dynamic Modelling and Optimisation of the Batch Enzymatic Synthesis of Amoxicillin, *Processes*, 2019, **7**(6), 318.
18 S. Diab, M. Raiyat and D. I. Gerogiorgis, Flow synthesis kinetics for lomustine, an anti-cancer active pharmaceutical ingredient, *React. Chem. Eng.*, 2021, **6**(10), 1819.
19 C. A. Hone, A. Boyd and A. O'Kearney-McMullan, *et al.*, Definitive screening designs for multistep kinetic models in flow, *React. Chem. Eng.*, 2019, **4**(9), 1565.
20 E. Içten, A. J. Maloney and M. G. Beaver, *et al.*, A Virtual Plant for Integrated Continuous Manufacturing of a Carfilzomib Drug Substance Intermediate, Part 1: CDI-Promoted Amide Bond Formation, *Org. Process Res. Dev.*, 2020, **24**(10), 1861.
21 K. P. Cole, B. M. Campbell and M. B. Forst, *et al.*, An Automated Intermittent Flow Approach to Continuous Suzuki Coupling, *Org. Process Res. Dev.*, 2016, **20**(4), 820.
22 M. Grom, G. Stavber, P. Drnovšek and B. Likozar, Modelling chemical kinetics of a complex reaction network of active pharmaceutical ingredient (API) synthesis with process optimization for benzazepine heterocyclic compound, *Chem. Eng. J.*, 2016, **283**, 703.
23 C. J. Taylor, M. Booth and J. A. Manson, *et al.*, Rapid, automated determination of reaction models and kinetic parameters, *Chem. Eng. J.*, 2021, **413**, 127017.
24 C. Schenk, M. Short and J. S. Rodriguez, *et al.*, Introducing KIPET: A novel open-source software package for kinetic parameter estimation from experimental datasets including spectra, *Comput. Chem. Eng.*, 2020, **134**, 106716.

25 H. G. Jolliffe and D. I. Gerogiorgis, Process modelling and simulation for continuous pharmaceutical manufacturing of ibuprofen, *Chem. Eng. Res. Des.*, 2015, **97**, 175.
26 A. Echtermeyer, Y. Amar, J. Zakrzewski and A. Lapkin, Self-optimisation and model-based design of experiments for developing a C–H activation flow process, *Beilstein J. Org. Chem.*, 2017, **13**, 150.
27 K. C. Aroh and K. F. Jensen, Efficient kinetic experiments in continuous flow microreactors, *React. Chem. Eng.*, 2018, **3**(1), 94.
28 M. O. Frederick, M. A. Pietz and D. P. Kjell, *et al.*, Development of a Leuckart–Wallach Reaction in Flow for the Synthesis of Abemaciclib, *Org. Process Res. Dev.*, 2017, **21**(9), 1447.
29 R. R. de Oliveira Silva, P. V. Cuesta Calvo and M. Fernandes da Silva, *et al.*, Flow Synthesis of a Thiazolidine Drug Intermediate in Capillary Microreactors, *Chem. Eng. Technol.*, 2019, **42**(2), 465.
30 D. D. S. Pinheiro, R. R. D. O. Silva and P. V. C. Calvo, *et al.*, Microreactor Technology as a Tool for the Synthesis of a Glitazone Drug Intermediate, *Chem. Eng. Technol.*, 2018, **41**(9), 1800.
31 Y. R. Dubhashe, V. M. Sawant and V. G. Gaikar, Process Intensification of Continuous Flow Synthesis of Tryptophol, *Ind. Eng. Chem. Res.*, 2018, **57**(8), 2787.
32 C. T. Armstrong, C. Q. Pritchard and D. W. Cook, *et al.*, Continuous flow synthesis of a pharmaceutical intermediate: a computational fluid dynamics approach, *React. Chem. Eng.*, 2019, **4**(3), 634.
33 A. B. Singer, Global Dynamic Optimization, *PhD Thesis*, Department of Chemical Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, 2004.
34 A. B. Singer and P. I. Barton, Global solution of optimization problems with parameter-embedded linear dynamic systems, *J. Optim. Theory Appl.*, 2004, **121**(3), 613.
35 A. B. Singer and P. I. Barton, Global Optimization with Nonlinear Ordinary Differential Equations, *J. Glob. Optim.*, 2006, **34**(2), 159.
36 A. B. Singer, J. W. Taylor, P. I. Barton and W. H. Green, Global Dynamic Optimization for Parameter Estimation in Chemical Kinetics, *J. Phys. Chem.*, 2006, **110**, 971.
37 M. Trampuž, D. Teslić and B. Likozar, Crystallization of fesoterodine fumarate active pharmaceutical ingredient: Modelling of thermodynamic equilibrium, nucleation, growth, agglomeration and dissolution kinetics and temperature cycling, *Chem. Eng. Sci.*, 2019, **201**, 97.
38 A. Pohar and B. Likozar, Dissolution, Nucleation, Crystal Growth, Crystal Aggregation, and Particle Breakage of Amlodipine Salts: Modeling Crystallization Kinetics and Thermodynamic Equilibrium, Scale-up, and Optimization, *Ind. Eng. Chem. Res.*, 2014, **53**(26), 10762.
39 C. Schenk, L. T. Biegler, L. Han and J. Mustakis, Kinetic Parameter Estimation from Spectroscopic Data for a Multi-Stage Solid–Liquid Pharmaceutical Process, *Org. Process Res. Dev.*, 2021, **25**(3), 373.
40 Q. Su, Z. K. Nagy and C. D. Rielly, Pharmaceutical crystallisation processes from batch to continuous operation using MSMPR stages: Modelling, design, and control, *Chem. Eng. Process.: Process Intesif.*, 2015, **89**, 41.
41 H. F. Kraus, D. Acevedo and W. Wu, *et al.*, Kinetic modelling of an environmentally friendly carbamazepine synthesis via urea and iminostilbene in batch and continuous processes, *React. Chem. Eng.*, 2023, **8**(2), 402.
42 Scale-up Systems Ltd., *Dynochem*, Scale-up Systems Ltd.
43 The MathWorks Inc., *MATLAB*, MathWorks.
44 Process Systems Enterprise, *gPROMS Process*, PSE.
45 Process Systems Enterprise, *gPROMS Formulated Products*, PSE.
46 P. Virtanen, R. Gommers and T. E. Oliphant, *et al.*, SciPy 1.0: fundamental algorithms for scientific computing in Python, *Nat. Methods*, 2020, **17**(3), 261.
47 M. Short, L. T. Biegler, S. García-Muñoz and W. Chen, Estimating variances and kinetic parameters from spectra across multiple datasets using KIPET, *Chemom. Intell. Lab. Syst.*, 2020, **203**, 104012.
48 M. Short, C. Schenk and D. Thierry, *et al.*, KIPET – An Open-Source Kinetic Parameter Estimation Toolkit, *Comput.-Aided Chem. Eng.*, 2019, **47**, 299.
49 L. D. R. Beal, D. Hill, R. A. Martin and J. D. Hedengren, GEKKO Optimization Suite, *Processes*, 2018, **6**(8), 106.
50 A. B. Singer and P. I. Barton, Bounding the Solutions of Parameter Dependent Nonlinear Ordinary Differential Equations, *SIAM J. Sci. Comput.*, 2006, **27**(6), 2167.
51 COMSOL, *COMSOL Multiphysics*, COMSOL.
52 COMSOL, *Chemical Reaction Engineering Module*, COMSOL.
53 S. C. Burnham, D. P. Searson, M. J. Willis and A. R. Wright, Inference of chemical reaction networks, *Chem. Eng. Sci.*, 2008, **63**(4), 862.
54 J. Tsu, V. H. G. Díaz and M. J. Willis, Computational approaches to kinetic model selection, *Comput. Chem. Eng.*, 2019, **121**, 618.
55 E. August and A. Papachristodoulou, Efficient, sparse biological network determination, *BMC Syst. Biol.*, 2009, **3**(25), 1752.
56 C. J. Taylor, H. Seki and F. M. Dannheim, *et al.*, An automated computational approach to kinetic model discrimination and parameter estimation, *React. Chem. Eng.*, 2021, **6**(8), 1404.
57 M. J. Willis and M. V. Stosch, Inference of chemical reaction networks using mixed integer linear programming, *Comput. Chem. Eng.*, 2016, **90**, 31.
58 H. Akaike, A new look at the statistical model identification, *IEEE Trans. Autom. Control*, 1974, **19**(6), 716.
59 N. N. Narisetty, Chapter 4 - Bayesian model selection for high-dimensional data, in *Handbook of Statistics*, ed. A. S. R. Srinivasa Rao and C. R. Rao, Elsevier, 2020, p. 207.
60 K. P. Burnham, *Model selection and multimodel inference: a practical information-theoretic approach*, ed. D. R. Anderson and K. P. Burnham, Springer, New York, 2nd edn, 2002.
61 N. Takebe, A. R. Naqash and G. O'Sullivan Coyne, *et al.*, Safety, Antitumor Activity, and Biomarker Analysis in a Phase I Trial of the Once-daily Wee1 Inhibitor Adavosertib (AZD1775) in Patients with Advanced Solid Tumors, *Clin. Cancer Res.*, 2021, **27**(14), 3834.

62 A. A. Mortlock, D. M. Wilson and J. G. Kettle, *et al.*, 5.02 - Selective Kinase Inhibitors in Cancer, in *Comprehensive Medicinal Chemistry III*, 2017, p. 39.
63 G. Yamamoto, K. Tanaka and R. Kamata, *et al.*, WEE1 confers resistance to KRASG12C inhibitors in non-small cell lung cancer, *Cancer Lett.*, 2025, **611**, 217414.
64 G. O. Rodríguez-Vázquez, A. O. Diaz-Quiñones and N. Chorna, *et al.*, Synergistic interactions of cytarabine-adavosertib in leukemic cell lines proliferation and metabolomic endpoints, *Biomed. Pharmacother.*, 2023, **166**, 115352.
65 Royal Society of Chemistry, *ChemSpider*, 2023, [cited 2023 10/03/2023]; Available from: **https://www.chemspider.com/**.
66 S. Kim, J. Chen and T. Cheng, *et al.*, PubChem 2023 update, *Nucleic Acids Res.*, 2023, **51**(D1), D1373.
67 E. R. Strieter, B. Bhayana and S. L. Buchwald, Mechanistic Studies on the Copper-Catalyzed N-Arylation of Amides, *J. Am. Chem. Soc.*, 2009, **131**(1), 78.
68 X. Ribas and I. Güell, Cu(I)/Cu(III) catalytic cycle involved in Ullmann-type cross-coupling reactions, *Pure Appl. Chem.*, 2014, **86**(3), 345.
69 C. Sambiagio, S. P. Marsden, A. J. Blacker and P. C. McGowan, Copper catalysed Ullmann type chemistry: from mechanistic aspects to modern development, *Chem. Soc. Rev.*, 2014, **43**(10), 3525.
70 J. F. Bunnett, Mechanism and reactivity in aromatic nucleophilic substitution reactions, *Q. Rev., Chem. Soc.*, 1958, **12**(1), 1.
71 A. J. Kirby and A. G. Varvoglis, The reactivity of phosphate esters: reactions of monoesters with nucleophiles. Nucleophilicity independent of basicity in a bimolecular substitution reaction, *J. Chem. Soc. B*, 1968, 135.
72 T. Cohen and I. Cristea, Kinetics and mechanism of the copper(I)-induced homogeneous Ullmann coupling of o-bromonitrobenzene, *J. Am. Chem. Soc.*, 1976, **98**(3), 748.
73 H. K. Oh, M. H. Ku, H. W. Lee and I. Lee, Nucleophilic Substitution Reactions of Aryl Dithioacetates with Pyridines in Acetonitrile, *J. Org. Chem.*, 2002, **67**(11), 3874.
74 A. Kundu, M. Inoue and H. Nagae, *et al.*, Direct ortho-C–H Aminoalkylation of 2-Substituted Pyridine Derivatives Catalyzed by Yttrium Complexes with N,N′-Diarylethylenediamido Ligands, *J. Am. Chem. Soc.*, 2018, **140**(23), 7332.
75 G. Evano, A. Nitelet, P. Thilmany and D. F. Dewez, Metal-Mediated Halogen Exchange in Aryl and Vinyl Halides: A Review, *Front. Chem.*, 2018, **6**, 114.
76 X. Jin and R. P. Davies, Copper-catalysed aromatic-Finkelstein reactions with amine-based ligand systems, *Catal. Sci. Technol.*, 2017, **7**(10), 2110.
77 A. Klapars and S. L. Buchwald, Copper-Catalyzed Halogen Exchange in Aryl Halides: An Aromatic Finkelstein Reaction, *J. Am. Chem. Soc.*, 2002, **124**(50), 14844.
78 R. D. Pace and Y. Regmi, The Finkelstein Reaction: Quantitative Reaction Kinetics of an SN2 Reaction Using Nonaqueous Conductivity, *J. Chem. Educ.*, 2006, **83**(9), 1344.
79 D. S. Surry and S. L. Buchwald, Diamine ligands in copper-catalyzed reactions, *Chem. Sci.*, 2010, **1**(1), 13.
80 DECHEMA, *DETHERM*, [accessed 05/10/2022]; Available from: **https://detherm.cds.dechema.de/**.
81 M. Tao, Z. Wang and J. Gong, *et al.*, Determination of the Solubility, Dissolution Enthalpy, and Entropy of Pioglitazone Hydrochloride (Form II) in Different Pure Solvents, *Ind. Eng. Chem. Res.*, 2013, **52**(8), 3036.
82 J. Keizer, Diffusion effects on rapid bimolecular chemical reactions, *Chem. Rev.*, 1987, **87**(1), 167.
83 A. I. Awaji, B. Köksoy and M. Durmuş, *et al.*, Novel Hexadeca-Substituted Metal Free and Zinc(II) Phthalocyanines; Design, Synthesis and Photophysicochemical Properties, *Molecules*, 2019, **24**(1), 77.
84 G. B. Dutt and N. Periasamy, Electron-transfer distance in intermolecular diffusion-limited reactions, *J. Chem. Soc., Faraday Trans.*, 1991, **87**(24), 3815.
85 T. R. Waite, General Theory of Bimolecular Reaction Rates in Solids and Liquids, *J. Chem. Phys.*, 1958, **28**(1), 103.
86 F. C. Collins and G. E. Kimball, Diffusion-controlled reaction rates, *J. Colloid Sci.*, 1949, **4**(4), 425.
87 R. R. Shaydullin, A. S. Galushko and V. V. Ilyushenkova, *et al.*, Are activation barriers of 50–70 kcal mol−1 accessible for transformations in organic synthesis in solution?, *Chem. Sci.*, 2025, **16**(12), 5289.
88 R. J. Ouellette and J. D. Rawn, 3 - Introduction to Organic Reaction Mechanisms, in *Organic Chemistry*, ed. R. J. Ouellette and J. D. Rawn, Academic Press, 2018, 2nd edn, p. 51.
89 D. M. Espie, *The use of nonlinear parameter estimation for dynamic chemical reactor modelling*, University of London, 1987.
90 D. J. Pritchard and D. W. Bacon, Statistical assessment of chemical kinetic models, *Chem. Eng. Sci.*, 1975, **30**(5), 567.
91 K. A. P. McLean and K. B. McAuley, Mathematical modelling of chemical processes—obtaining the best model predictions and parameter estimates using identifiability and estimability procedures, *Can. J. Chem. Eng.*, 2012, **90**(2), 351.
92 H. T. Banks and M. L. Joyner, AIC under the framework of least squares estimation, *Appl. Math. Lett.*, 2017, **74**, 33.
93 P. Kletting and G. Glatting, Model selection for time-activity curves: The corrected Akaike information criterion and the F-test, *Z. Med. Phys.*, 2009, **19**(3), 200.
94 A. D. Rodman and D. I. Gerogiorgis, Parameter estimation and sensitivity analysis for dynamic modelling and simulation of beer fermentation, *Comput. Chem. Eng.*, 2020, **136**, 106665.
95 D. W. Pilarski and D. I. Gerogiorgis, Systematic Parameter Estimation And Dynamic Simulation of Cold Contact Fermentation for Alcohol-Free Beer Production, *Processes*, 2022, **10**(11), 2400.