SOLVE chemistry

A fully automated kinetics workflow: From data collection to microkinetic modelling

SOLVE Chemistry, Scale Space, 58 Wood Lane, W12 7RZ, United Kingdom

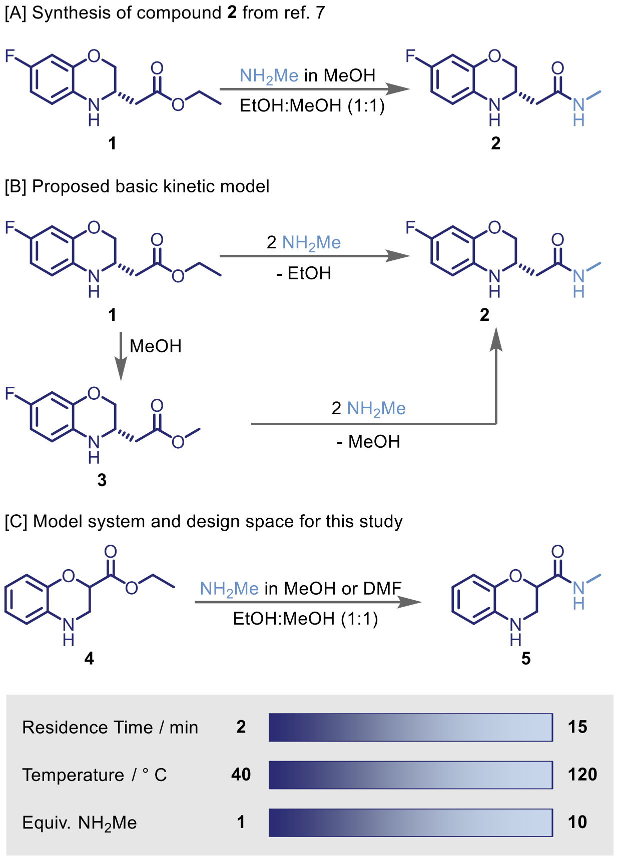
Abstract We report a fully automated kinetics workflow integrating transient flow data collection with AI-assisted microkinetic modelling. Using a model substrate analogous to one studied during AstraZeneca's development of Balcinrenone, we collected 224 data points across a broad design space in just 4 experimental runs and 17 hours of automated reactor time. Our proprietary AI agent iteratively proposed, fitted, and chemically validated 10 mechanistic models in under an hour. The best-performing model reproduced key mechanistic features identified in the prior literature while revealing additional mechanistic detail, namely reversible transesterification and rationalising the dual role of methylamine - delivering a 58% reduction in model error relative to the literature benchmark

Introduction Pharmaceutical process development relies on efficient and robust data collection methods. The rise of automation in synthetic organic chemistry is a powerful tool that enables this kind of data to be collected at an accelerated rate. Reliable data of this sort underpins kinetic analysis and developing mechanistic understanding of reactions that is vital to process development campaigns.

Continuous flow synthesis coupled with on-line analytics is an increasingly popular way to collect high quality kinetic data in an automated fashion.[1,2] Within this field, transient flow regimes can be used to create dynamic reaction environments by varying reaction inputs during the course of a reaction.[3] This lends itself exceptionally well to data-rich experimentation, with each analytical result corresponding to a unique set of reaction conditions. This technique allows concentrations, temperatures and reaction times to be varied in a single experimental run – something that is infeasible on automated batch platforms. At SOLVE Chemistry, we have developed automated transient flow reactors capable of collecting high-quality data in a high-throughput manner.

Judicious choice in reaction inputs means the resulting outputs are suitable not only for rapid reaction optimisation, but also global kinetic interpretation.[3,4] The wider use of machine learning in chemistry has reduced the barrier to using this data for microkinetic modelling, maximising the insights from data collection. The biggest impacts are led by better optimisation software and algorithms, as well as significant improvements in Large Language Models for chemistry.[5,6]

We sought to explore if we could fully automate the entire kinetic analysis workflow – from data collection to machine-learning assisted microkinetic modelling. When considering a model system to study for this, we were inspired by recent work from García-Morales, Cox and co-workers, who published details about the development of Balcinrenone.[7] En route to the final compound, the conversion of ethyl ester **1** to methyl amide intermediate **2** (Figure 1A) was studied. Four reactions of varied temperatures and concentrations enabled the development of a basic kinetic model, Figure 1B. Kinetic analysis in this work was based on product concentration data.


[A] Synthesis of compound **2** from ref. 7

[B] Proposed basic kinetic model

[C] Model system and design space for this study

Figure 1. [A] Reaction of interest from ref. 7. [B] Basic kinetic model developed in ref. 7. [C] Reaction and design space studied in this investigation.

The experimental data fit a mechanistic proposal with two pathways from **1** to **2**. The first is direct, irreversible aminolysis with methylamine. The second proceeds via irreversible transesterification with methanol to access methyl ester **3**, which then undergoes aminolysis to furnish **4**. The experimental results indicated that the esterification pathway dominates and the aminolysis from the methyl ester is 3.6 times faster than from the ethyl ester. The authors were able to use this kinetic model for a full factorial simulation and additional experimentation refined the kinetic model further, enabling the development of a telescoped process with improved sustainability and throughput.

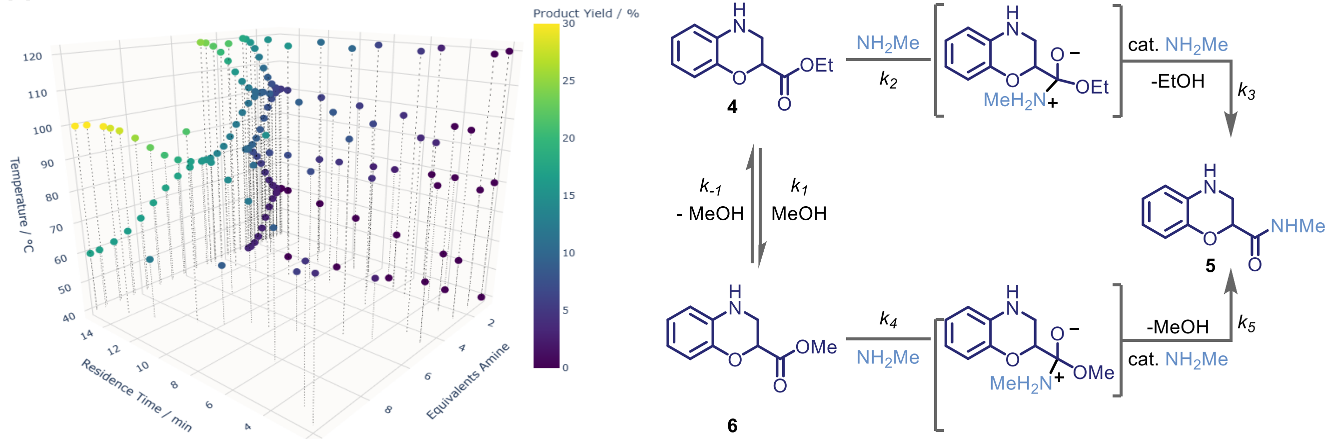
To benchmark our approach against this work, we set out to investigate whether our automated transient flow reactors could collect data in minimal experimental time and hypothecate and fit improved interpretable mechanistic models. We selected analogous ethyl ester **4** as our model substrate and considered the design space highlighted in Figure 1C. Working under a flow regime enabled us to work at higher temperatures than the previous study, in turn facilitating short reaction times.

Results and Discussion Our campaign required only 4 experimental runs and 17 hours of automated reactor time in total – achieved in 4 days. Across the design space a total of 224 data points were collected, each corresponding to a unique set of reaction conditions, Figure 2A. As with the previous study, only product concentration was monitored during the reactions. Each data point was then fed to our proprietary AI agent for mechanistic kinetic model discovery. The agent iteratively proposed kinetic models, and then the model parameters are fitted to the data. After each fit, the agent analysed the model and the data, and underwent a chemical plausibility check. Based on the best-fitting models, the agent refined or proposed alternative mechanisms.

A total of 10 models were automatically probed by our modelling tool in under an hour; the mechanistic output from the best-performing model is shown in Figure 2B, along with the fitted reaction activation energies and Arrhenius factors calculated as part of the workflow, Figure 2C.

Our tool identified the same transesterification with methanol from the solvent as identified by García-Morales, Cox and co-workers, resulting in two pathways to the product. The best performing model discovered by our tool differs from the kinetic model proposed in the previous work by identifying the transesterification as a reversible process. This was only proposed in the previous work after additional experiments and further refinement of the basic model. Although the model in the original work proposes a second-order dependency in methylamine from either ester, our best performing model identified a tetrahedral intermediate in these pathways, which collapses via proton transfer, facilitated by another equivalent methylamine and expulsion of ethanol or methanol, in turn regenerating methylamine. Our best-performing model splits the second-order effect of methylamine into separate stoichiometric and catalytic steps which provides both better mechanistic understanding and a quantitatively better model fit.

Each successive iteration of the automated modelling workflow produced better fitting models , Figure 2D. A simple model (of the form A+B → C) showed the greatest sum of squared errors (SSE) (point A, error = 1.49×10^{-5}). The best performing model (point C, error = 0.56×10^{-5}) showcased a better fit than the model most similar to the basic model proposed in the literature (point B, error = 1.33×10^{-5}) with a reduction of almost 58% in relative error. The major difference mechanistically being the reversibility of the transesterification step and the separation of the stoichiometric and catalytic effects of methylamine in the amidation step.



[A] Reaction Data

[B] SOLVE-AI Proposed Microkinetic Model

[C] Calculated Kinetic Parameters

Entry	Step	Ea (kJ/mol)	A
1	k_1	18.4	0.3
2	k_{-1}	0.0013	0.064
3	k_2	18.8	4.65
4	k_3	19.0	0.06
5	k_4	14.0	0.25
6	k_5	46.1	13.3

[D] Model Performance Evolution

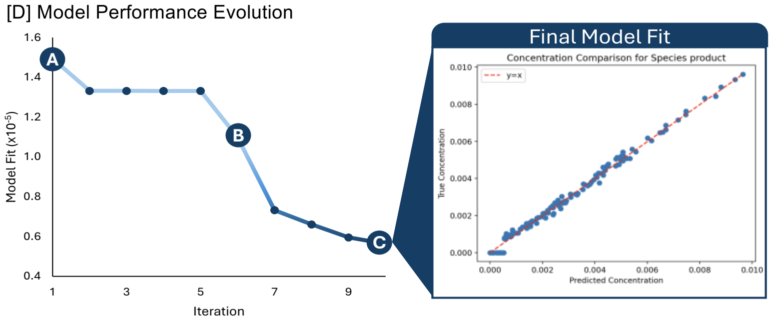


Figure 2. [A] Reaction output data collected experimentally. [B] Proposed mechanistic pathway. [C] Calculated kinetic parameters. [D] Model fit evaluation across iterations as SSE. Inset is the model performance validation, comparing the model predictions against the true data to assess kinetic model fit to the data.

Conclusion

Our automated workflow delivers mechanistic kinetic understanding with remarkable speed: from chemistry entering the lab to fully fitted microkinetic models in just 4 days, requiring only 17 hours of unattended reactor time, 4 experimental runs, and yielding 224 data points across the design space. This represents a significant acceleration over traditional manual approaches, reducing time to process development decisions and manual effort required from scientists.

The quality of insight generated also surpasses that of previous manual approaches. Features that emerged only after further experimental effort in the original work, notably the reversibility of the transesterification step, were identified automatically in our first modelling campaign. Beyond reproducing the literature mechanism, our AI agent uncovered deeper mechanistic detail: by resolving the observed second-order methylamine dependence into discrete stoichiometric and catalytic steps, we obtained both a more chemically interpretable model and a quantitatively superior fit, with nearly 58% lower error than the closest literature-equivalent model.

References

1 B. Zhang, A. Mathoor and T. Junkers, *Angew. Chem. Int. Ed.*, DOI:10.1002/anie.202308838.

2 B. J. Reizman and K. F. Jensen, *Org. Process Res. Dev.*, 2012, **16**, 1770–1782.

3 L. Schrecker, J. Dickhaut, C. Holtze, P. Staehle, M. Vranceanu, A. Wieja, K. Hellgardt and K. K. Hii, *React. Chem. Eng.*, 2024, **9**, 1077–1086.

4 C. J. Taylor, M. Booth, J. A. Manson, M. J. Willis, G. Clemens, B. A. Taylor, T. W. Chamberlain and R. A. Bourne, *Chemical Engineering Journal*, 2021, **413**, 127017.

5 J. Burés and I. Larrosa, *Nature*, 2023, **613**, 689–695.

6 M. Á. de Carvalho Servia, I. O. Sandoval, K. K. (Mimi) Hii, K. Hellgardt, D. Zhang and E. Antonio del Rio Chanona, *Digital Discovery*, 2024, **3**, 954–968.

7 C. García-Morales, D. Dave, Z. Neouchy, H. Pointon, M. J. Foulkes, A. Page, T. O. Ronson and R. J. Cox, *Org. Process Res. Dev.*, 2024, **28**, 4455–4466.