Transient Flow-Assisted Kinetic Modelling and Reaction Network Identification for Pyrazole Synthesis

Fernando Vega-Ramon,a Linden Schrecker,b Miguel Angel de Carvalho Servia,c King Kuok (Mimi) Hii,b Dongda Zhang a,\*

a Department of Chemical Engineering, the University of Manchester, Engineering Building, Manchester M13 9PL, United Kingdom

b Department of Chemistry, Imperial College London, 82, Wood Lane, London W12 0BZ, United Kingdom

c Department of Chemical Engineering, Imperial College London, Exhibition Road, London SW7 2AZ, United Kingdom

\* Corresponding Author. Email: dongda.zhang@manchester.ac.uk

Abstract

Due to their structural versatility and wide range of biological activities, pyrazoles serve as important motifs in a variety of organic chemicals such as pharmaceuticals and pesticides. An established and economically favourable route for the synthesis of pyrazoles is the Knorr reaction, which involves the condensation of di-carbonyl and hydrazine reactants. Despite being researched extensively, previous kinetic studies on the Knorr reaction have been mainly qualitative and based on empirical observations. Because of this, the development of a comprehensive kinetic model capable of predicting product distribution and regioselectivity under different operating conditions remains a challenge.

In this study, a microkinetic analysis framework was combined with transient PFR experimentation techniques to perform kinetic modelling and reaction network identification. On the basis of literature observations and gathered spectroscopic and time-series kinetic data, two plausible reaction mechanisms were proposed for the condensation of symmetrical 1,3-diketones with phenyl hydrazine. Sparse-regression approaches were then adopted to identify kinetically negligible pathways within the proposed reaction networks, reducing the complexity of their corresponding microkinetic models while retaining their original fitting accuracy. The robustness of the reduced models was further confirmed through uncertainty analysis; the results indicate that most kinetic parameters are identifiable and statistically significant, suggesting the advantage of high throughput data generation via transient experiments.

Finally, a model-based design of experiments (MbDoE) methodology was adopted to identify the optimal set of experimental conditions for discriminating among the two proposed models. This work therefore provides a new avenue for the systematic analysis of complex organic reactions assisted by transient flow experimentation.

**Keywords**: Transient flow reactor, pyrazoles, kinetic modelling, model discrimination.

* 1. Introduction

Owing to the wide availability of 1,3-diketone and hydrazine precursors, the Knorr reaction is a popular pathway for the synthesis of substituted pyrazoles. Previous literature studies (Sloop et al., 2008; Norris et al., 2005) have qualitatively demonstrated the influence of reaction parameters (such as feed ratios, reactant substituent groups, or pH) on the reaction rate and product regioselectivity. The construction of a detailed kinetic model capable of describing such effects is therefore a vital requirement for process optimisation, reaction network identification, and scale-up considerations.

A common challenge in the development of comprehensive kinetic models for complex chemical systems is the need for reliable reaction data over a diverse range of conditions. An emerging technique for the efficient generation of time-series data is transient flow experimentation, whereby a flow reactor is operated under unsteady-state conditions to emulate a collection of several batch experiments. In the present work, this high throughput experimentation tool was combined with kinetic model construction and MbDoE methodologies to elucidate potential reaction mechanisms for the addition of phenyl hydrazine to substituted diketones, namely acetyl acetone and heptane-3,5-dione.

* 1. Experimental Methodology
     1. Transient Flow Setup

Ethanolic reagent solutions were kept in bottles; inlet tubing lines connected these to two Gilson 305 HPLC pumps controlled via a master/slave system. The pump outlets fed into a Valco T-piece stainless steel mixer submerged in a silicon oil bath. The mixer outlet fed into a 5.10 m (4.13 mL) stainless steel tubular reactor, which was also submerged in the oil bath. The operating temperature was maintained around the desired set point (70 °C) by means of PID control. The reactor effluent was then passed through a custom cooling system consisting of a Peltier thermoelectric cooler connected to an aluminum block heat sink. In-line IR spectroscopic analysis was then performed using a Mettler Toledo ReactIRTM 15 equipped with a micro flow cell. A back pressure regulator was set to 8.1 bar to ensure all solvents remain in solution for the tested temperatures. A more detailed account of the experimental setup can be found on previous work (Schrecker et al., 2023).

* + 1. Time-series Kinetic Experiments

Transient ramps were performed by introducing step changes or linear ramps to the volumetric flowrates of the hydrazine and diketone reactants, while simultaneously monitoring the product concentration at the reactor outlet. Namely, three different kinds of transient experiments were performed:

* Residence time ramps, whereby the total volumetric flowrate is decreased while keeping the hydrazine:diketone ratios unchanged. This is equivalent to running a collection of batch experiments with the same initial concentrations but varying reaction times.
* Stoichiometry ramps, whereby the flowrates of reactants are ramped linearly at a fixed cumulative flowrate. This yields a collection of experiments with the same residence time but varying initial concentrations.
* Bivariate ramps, whereby a stoichiometry ramp is performed up to a final diketone:hydrazine ratio, after which a residence time ramp is carried out over a complete reactor volume.
  1. Modelling Methodology
     1. Reaction Mechanism Proposal

Two reaction schemes were constructed on the basis of gathered spectroscopic data and literature observations; these are presented in Fig. 1 below. Equations for the rate of change of concentration of all species were derived by applying the law of mass-action.

A collage of formulas

Description automatically generated with medium confidence

Figure 1: Proposed reaction networks for the reaction between 1,3-diketones and phenyl hydrazine. (a) Model 1 adapted from (Schrecker et al., 2022); (b) Model 2.

Model 1 in Fig. 1(a) corresponds to the reaction network proposed in our previous work (Schrecker et al., 2022). The first step in the mechanism involves the formation of a hydrazone intermediate, which has been isolated and identified in previous literature studies (Sloop et al., 2008). The general consensus in literature is that the hydrazone species undergoes ring-closing (Step 2) to form a hydroxypyrazolidine intermediate, which then aromatises to the final pyrazole product (Step 4). In addition, we also hypothesised an alternative reaction pathway involving the formation (Step 3) and aromatization (Step 5) of a di-addition intermediate, which was detected via mass spectroscopy analysis. The key difference between the two networks concerns the catalysis of the aromatization reactions; whereas in Model 1 these are product- and diketone-catalysed pathways, in Model 2 they are assumed to be hydrazine-catalysed. The rationale behind this modification is that higher hydrazine:diketone ratios favoured product yields on reactant stoichiometry ramp experiments.

* + 1. Sparse Regression-based Parameter Estimation

Kinetic parameter estimation was formulated as a nonlinear, constrained optimisation problem presented in Eq. (2a) – Eq. (2c). Inspired by sparse regression approaches, the objective function to be minimised is a least-squares expression which incorporates an additional sparsity regularization term. The rationale behind this term is that model fitting and model reduction can be accomplished simultaneously by penalizing the number of active reaction pathways: For kinetically significant steps their corresponding constants take some non-zero value and the penalty term approaches unity, whereas for kinetically negligible steps the penalty term approaches 0 as .

|  |  |
| --- | --- |
|  | (2a) |
|  | (2b) |
|  | (2c) |

where andare vectors for the experimentally measured and simulated concentrations of species, respectively. is a weighting matrix used to scale the residuals and is the penalty weight. The number of datapoints is denoted by , while refers to the number of kinetic constants in a proposed reaction network. is a small, positive number. The underscripts and denote lower and upper bounds, respectively.

The differential process constraints Eq. (2b) were discretised into algebraic profiles using Adams method and backward differentiation formulas. The arising NLP was then solved via sequential least squares programming. This procedure was implemented computationally through the S*cipy* environment in *Python* programming language.

* + 1. Kinetic Parameter Uncertainty Analysis

Following parameter estimation, the fidelity of the reduced model structures was further assessed via uncertainty analysis. The first step towards computation of parameter confidence intervals is the linear approximation of the covariance matrix:

|  |  |
| --- | --- |
|  | (3) |

where is the variance-covariance matrix, is the Hessian matrix of the objective function at the optimum (approximated through central differences) and refers to the minimised objective function value.

Joint confidence intervals for the estimated constants were computed from the covariance matrix following the procedure described by Franceschini et al. (2008), while their marginal variances were directly taken from the diagonal elements of the matrix.

* + 1. Model Discrimination via Design of Experiments

In this work, the two proposed reaction networks incorporate different assumptions regarding the product-formation steps. A recurring challenge in kinetic model discrimination is that reaction networks comprising different elementary steps and degrees of detail can yield equally satisfactory fitting of experimental data. An MbDoE methodology was therefore adopted to design a discriminatory transient experiment; this was formulated as an optimisation problem Eq. (4a)-Eq. (4d), whereby the discrepancy between the two candidate models over a residence time ramp was maximised with respect to the feed concentrations:

|  |  |
| --- | --- |
|  | (4a) |
|  | (4b) |
|  | (4c) |
|  | (4d) |

where is a vector of initial conditions (in this case optimisation variables), and is a vector of discrepancies between the product concentration profiles andpredicted by Model 1 and Model 2, respectively.

* 1. Results
     1. Kinetic Model Fitting and Uncertainty Results

The reduced kinetic models exhibited good fitting performance, with an overall mean percentage error of 3.27% and 3.63% for Model 1 and Model 2, respectively. The simulated product profiles for a residence time ramp experiment with the acetyl acetone reactant are presented in Fig. 2 below.

A graph of a function

Description automatically generated with medium confidence

Figure 2: Model fitting results for residence time ramp experiments with acetyl acetone. (a) Model 1; (b) Model 2.

By carefully tuning the sparsity penalty weight, the trade-off between model fitting accuracy and model complexity was controlled. In this fashion, kinetically negligible reaction pathways in the proposed mechanisms were automatically identified and removed from their corresponding model structures. For both models and diketone reactants, the initial nucleophilic attack (Step 1), ring-closure (Step 2) and formation of di-addition intermediate (Step 3) were deemed to be approximately irreversible reactions. Moreover, the computed confidence intervals and correlation matrix for the non-zero kinetic constants indicate that most parameters are statistically identifiable and uncorrelated, as exemplified on Fig. 3 for Model 1 with the acetyl acetone reactant.

A grey circle with a black dot

Description automatically generated

Figure 3: (a) and (b) Joint confidence ellipsoids of some kinetic parameters in Model 1 (acetyl acetone case); (c) Correlation matrix plot for kinetic parameters in Model 1.

Comparison of the estimated constants for the acetyl acetone and heptane-3,5-dione datasets can also provide insight on steric influences on the reaction rate. For example, the formation of the di-addition intermediate was found to be faster for the shorter-chained substrate, whereas the opposite is true for the initial nucleophilic attack.

* + 1. Model Discrimination Results

The optimal initial conditions for a discriminatory residence time ramp experiment were determined following the procedure described in Section 3.4.

A graph of a curve

Description automatically generated

Figure 4: Predicted process trajectories for discriminatory experiment with acetyl acetone reactant, with initial conditions , , , .

From Fig. 4, Model 1 exhibits a more pronounced product formation rate as compared to Model 2 for the given initial concentrations where . This can be rationalised with reference to the key dissimilarities between the two mechanisms: The final aromatization steps are product- and diketone-catalysed for Model 1, as opposed to hydrazine-catalysed pathways in Model 2. Conducting this transient experiment will thus guide the modification and selection of an appropriate model structure for future studies.

* 1. Conclusions

In this work, state-of-the-art transient flow experimentation techniques were coupled with model construction and reduction methodologies to perform kinetics investigations on the Knorr pyrazole reaction. The model fitting accuracy and uncertainty results indicate that the proposed reaction networks are feasible representations of the process. Notably, analysis of the estimated constants in the reduced model structures allow for the identification of kinetically negligible reactions and the evaluation of steric effects on the reaction rate. A preliminary model-based design of experiments was also adopted to exemplify how transient experimentation can facilitate kinetic model discrimination.

References

Franceschini, G., & Macchietto, S. (2008). Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science*, *63*(19), 4846–4872.

Norris, T., Colon-Cruz, R., & Ripin, D. H. B. (2005). New hydroxy-pyrazoline intermediates, subtle regio-selectivity and relative reaction rate variations observed during acid catalyzed and neutral pyrazole cyclization. *Organic & Biomolecular Chemistry*, *3*(10), 1844.

Schrecker, L., Dickhaut, J., Holtze, C., Staehle, P., Vranceanu, M., Hellgardt, K., & Hii, K. K. (Mimi). (2023). Discovery of unexpectedly complex reaction pathways for the Knorr pyrazole synthesis *via* transient flow. *Reaction Chemistry & Engineering*, *8*(1), 41–46.

Sloop, J. C., Lechner, B., Washington, G., Bumgardner, C. L., Loehle, W. D., & Creasy, W. (2008). Pyrazole formation: Examination of kinetics, substituent effects, and mechanistic pathways. International Journal of Chemical Kinetics, 40(7), 370–383.