A Journal of the Gesellschaft Deutscher Chemiker

Angewandte International Edition GDCh Chemie

www.angewandte.org

Accepted Article

Title: Revisiting the Paradigm of Reaction Optimization in Flow with a Priori Computational Reaction Intelligence

Authors: Pauline Bianchi and Jean-Christophe M. Monbaliu

To be cited as: *Angew. Chem. Int. Ed.* **2023**, e202311526

Link to VoR: https://doi.org/10.1002/anie.202311526

WILEY VCH

10.1002/anie.202311526

WILEY VCH

RESEARCH ARTICLE

Revisiting the Paradigm of Reaction Optimization in Flow with a Priori Computational Reaction Intelligence

Pauline Bianchi,[a] and Jean-Christophe M. Monbaliu*[a][b]

Dedicated to Professor Klavs F. Jensen on the occasion of his 70th birthday

[a] P. Bianchi, Prof. Dr. J.-C. M. Monbaliu
Center for Integrated Technology and Organic Synthesis (CiTOS),
MolSys Research Unit, University of Liège
B6a, Room3/19, Allée du Six Août 13,
4000 Liège (SartTilman) (Belgium)
E-mail: jc.monbaliu@uliege.be
Homepage: https://www.citos.uliege.be
[b] Prof Dr. J.-C. M. Monbaliu
WEL Research Institute,
Avenue Pasteur, 6,
1300 Wavre (Belgique)

Abstract: The use of micro/meso-fluidic reactors has resulted in both new scenarios for chemistry and new requirements for chemists. Through flow chemistry, large-scale reactions can be performed in drastically reduced reactor sizes and reaction times. This obvious advantage comes with the concomitant challenge of re-designing long-established batch processes to fit these new conditions. The reliance on experimental trial-and-error to perform this translation frequently makes flow chemistry unaffordable, thwarting initial aspirations to revolutionize chemistry. By combining computational chemistry and machine learning, we have developed a model that provides predictive power tailored specifically to flow reactions. We show its applications to translate batch to flow, provide mechanistic insight, contribute reagent descriptors, and to synthesize a library of novel compounds in excellent yields after executing a single set of conditions.

Introduction

Organic chemistry has accompanied and sustained the development of fuels, therapies, and advanced materials throughout human history. Several technological revolutions later, the tools and overall context have changed: while robots and automated systems are being increasingly implemented in labs and on production lines, chemists remain pivotal assets behind innovation. The horizon of their imagination has extended, feeding new ways to access molecular architectures and cover emerging needs. Countless iterations and disruptive advances however still converge to a very pragmatic conclusion: developing new synthetic routes is resource-, time- and waste-intensive. A fundamental shift in how chemists approach their work is necessary.

Flow technology is one of the most impactful paradigm changes in the modern history of organic chemistry, offering superior process control, reduced waste generation, and enhanced safety from R&D to production scale.[1–3] These advantages, which are obtained through the use of micro/mesofluidic reactors (MFRs) can be further enhanced with the integration of artificial intelligence,[4,5] process analytical technology (PAT) tools,[6] and automation.[7,8] This makes flow technology particularly adapted to robust and versatile production systems,[9,10] and yet, breaking away from centuries of macroscopic batch synthesis has remained difficult.[11] There are mainly three aspects that keep making reactions challenging to perform with MFRs: slow kinetics which imposes onerous reactor size requirements, viscous liquids which are difficult to pump, or solid particles which can clog the reactor channel. The last two issues have been addressed by designing MFR configurations with specific technical features. Slow kinetics, on the other hand, represent a particular challenge considering that flow reactions typically require a timeframe below 15 min to be scalable. Traditional chemistry has not been developed to afford high conversions within such a narrow window nor does it indeed have the tools to reach them (Fig. 1A).[11]

Flow reactions are typically accelerated by using process intensification techniques, such as high reaction concentrations and/or increased temperatures.[12] The development of such intensified conditions has, until now, depended on trial-and-error chemistry, occasionally aided by a design-of-experiments approach or machine learning (ML) such as Bayesian Optimization models,[13] to increase the efficiency of the exploration. Predictive models based on mining data from databases such as Reaxys, while successful when applied to batch protocols, have limited use for flow chemistry due to the relatively small number of reactions that have been reported in MFRs.[14] Other types of predictive ML models based on Structure-Reactivity Relationships (SReARs) to relate chemical structure to a desired output (e.g., function or enantiomeric excess) have been developed,[15] yet only in the context of batch reactions. Despite the great insight they provide, the dependence of these models on experimental data represents a greater challenge for a young field such as flow chemistry where very little knowledge exists to even guide initial experiments. If we are to catch up with centuries of accumulated batch chemistry knowledge, the reliance of predictive tools on experimental data needs to be lessened.

While searching for less experiment-dependent ways to predict feasibility in flow, we became interested in the

10.1002/anie.202311526
WILEY VCH

RESEARCH ARTICLE

mathematical relationship described between *kinetics* (***k***), *temperature* (*T*) and the *activation barrier* ($\Delta G^{\ddagger}$) of a reaction afforded by Erying's equation [1] (with *k* = reaction rate constant, κ = transmission coefficient, H = Planck's constant (J s), kB = Boltzmann constant (J K^{-1}), T = temperature (K), R = gas constant (cal K^{-1} mol^{-1}), $\Delta G^{\ddagger}$ = activation barrier (cal mol^{-1})). Considering the rate laws of non-zero order chemical reactions, a kinetic constant (***k***) can then be related to the *rate* and *concentration* of the reaction. In other words, if $\Delta G^{\ddagger}$ can be accessed with high accuracy, Erying's equation becomes an invaluable gateway to predict a reaction time, at any given temperature and concentration.

$$k = \frac{\kappa k_B T}{h} e^{\left(\frac{-\Delta G^{\ddagger}}{RT}\right)} \quad [1]$$

These observations led us to develop a quantum chemistry-guided assistant, a prototype of which was successfully applied to the development on a large scale of an intensified flow synthesis of a contraceptive pharmaceutical.[16] This prototype used DFT to provide an ideal set of reagents, residence times, and temperatures to reach the target conversion, therefore minimizing waste, spending, and exposure to highly active hormonal substances. Despite its effectiveness, the method had two main drawbacks: suboptimal accuracy of the estimated conditions and taxing computational requirements to access high quality $\Delta G^{\ddagger}$ information. Therefore, a more practical solution was sought (Fig. 1B).

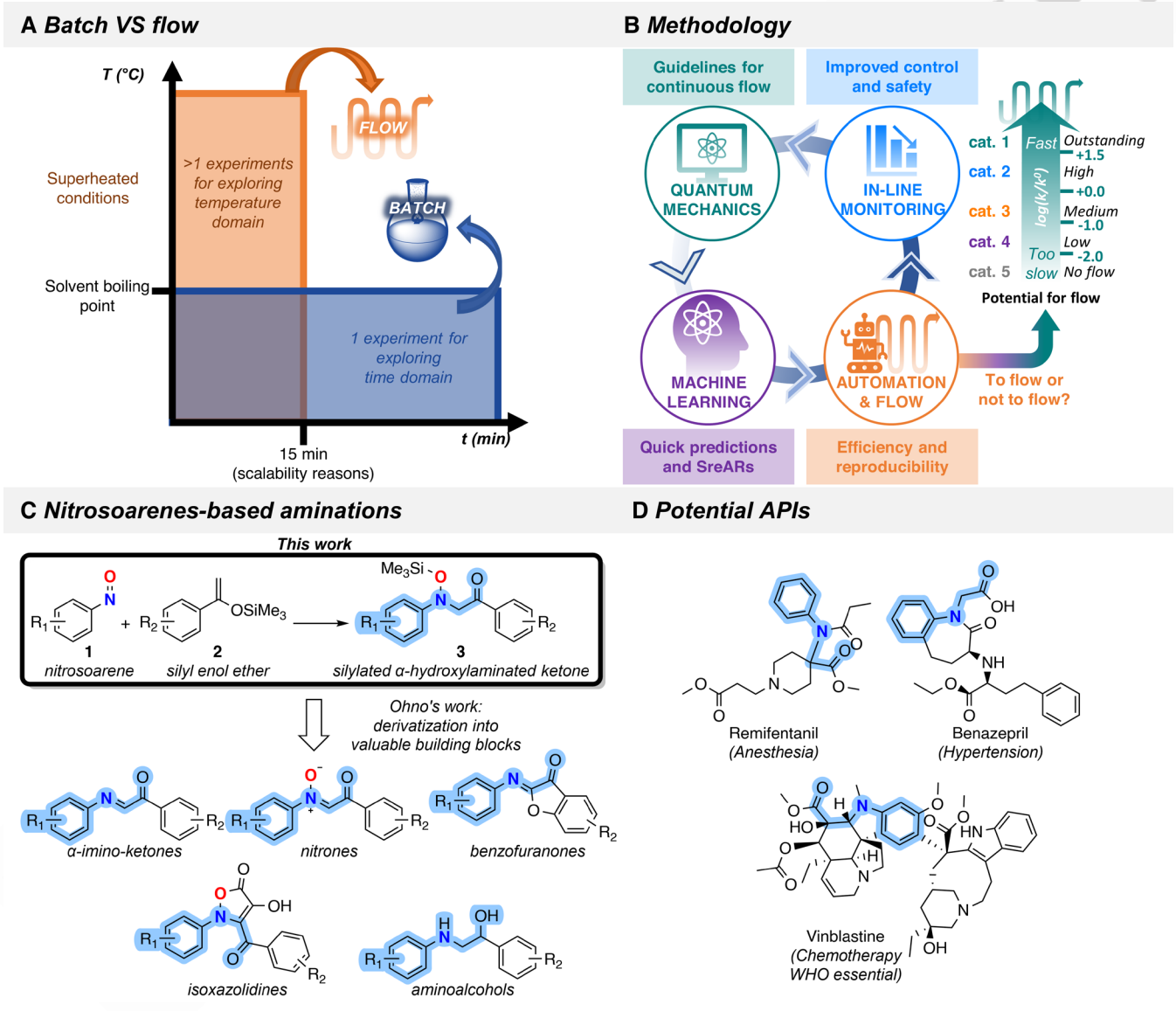


Figure 1. Overview of the interdisciplinary approach for revisiting the paradigm of reaction optimization. (**A**) One of the main challenges in translating from batch to flow conditions: batch chemistry was developed relying on extended reaction times whereas flow requires process intensification. (**B**) Our rationale for the adoption of quantum chemistry and machine learning to access rational and intelligent flow processes. (**C**) Case study: electrophilic amination between nitrosoarenes **1** and silyl enol ethers **2** toward hydroxylaminated ketones **3**. Compounds **3** are valuable aminated building blocks. (**D**) Representative examples of APIs potentially synthesizable through electrophilic aminations with nitrosoarenes.

The use of a relative logarithm (**log(*k*/*k*$_0$)**), comparing a reaction of interest (***k***) to the kinetics of a reference reaction (***k*$_0$**), was proposed as a first step to increase the accuracy of our estimates. Aside from providing data normalization and variance stabilization, **log(*k*/*k*$_0$)** can provide immediate intuitive insight about how fast or slow a reaction is, compared to the model reaction. Reactions involving nitrosoarene electrophiles **1** (Fig. 1C) were selected to test this approach. The choice was based on their potent nitrogen-transfer capacities which unfortunately remain underexploited due to their capricious reactivity.[17–19] Developing a protocol for the application of **1** would provide a precious tool for the preparation of pharmaceutically relevant aminated derivatives (Fig. 1D).

Results and Discussion

A Roadmap to Flow

The reaction between nitrosobenzene (**1a**) and unsubstituted silyl enol ether (**2a**) results in the highly selective formation of the corresponding N-addition product (**3a**) and was chosen as the model reaction with kinetic constant ***k*$_0$**.[20] The analogous reaction between 2-nitrosotoluene (**1b**) and **2a** was used to define a first relative constant **log(*k*$_{1b}$/*k*$_0$)**. Extensive computational and experimental probing of these two reactions was performed to show the robust accuracy of this metric (Supporting Information, see Section 7.1.).

Current records for precise $\Delta G^{\ddagger}$ determination feature a mean absolute error (MAE) between 0.5 and 1 kcal mol^{-1} which translates to a ~5-fold change in the predicted reaction time.[21–23] This means that an estimated residence time of 5 min could take place anywhere between 1 to 25 min, and therefore that the margin of error exceeds the narrow time window for optimal flow conditions (i.e., >15 min). Twenty computational methods to determine $\Delta G^{\ddagger}$ were assessed, from which a combination of Minnesota functionals[24] with corrective factors for concentration and quasi-harmonic oscillators,[25] provided values that were within 0.3 kcal mol^{-1} of experimental $\Delta G^{\ddagger}$ values, reducing the uncertainty for predicted time to ~1.5-fold (Supporting Information, see Section 5.3.). The method was translated into a DIY LabVIEW™ software (SnapPy) to automatically extract corrected $\Delta G^{\ddagger}$ and kinetic data from output files generated by a Quantum Mechanics software.

To be applicable to a library of compounds, the time required to predict **log(*k*/*k*$_0$)** needed to be reduced from hours to a few minutes. The need for a method to determine **log(*k*/*k*$_0$)** without having to compute $\Delta G^{\ddagger}$ was therefore evident. Instead, easily computed reagent properties (e.g., electronic, steric, or vibrational) to find **log(*k*/*k*$_0$)** through SReARs were explored.[15,26] Fourteen different reagent properties for each of 25 nitrosoarenes **1** (Fig. 2A) and 25 silyl enol ethers **2** (Fig. 2A) were computed and analyzed to identify feature intercorrelations. The size of the data set and its inherent features were selected based on previous works.[27–29] This resulted in the selection of two properties to be used as features for a ML training set, specifically *the nitrogen charge* and the *B*$_5$ *Sterimol* steric parameter for nitrosoarenes **1**, and the *C=C stretching* and the *B*$_1$ *Sterimol* steric parameter for silyl enol ethers **2** (Fig. 2B). The $\Delta G^{\ddagger}$ values for the reactions between these same 25 nitrosoarenes **1** (Fig. 2A) with **2a** and for the 25 silyl enol ethers **2** with **1a** were computed to calculate **log(*k*/*k*$_0$)**. We then searched for a ML model to relate **log(*k*/*k*$_0$)** to the selected features. Model generalization was assessed with k-fold cross-validation (k = 5). A total of four ML models were evaluated, namely, Least Absolute Shrinkage and Selection Operator (LASSO), Ridge, K-Nearest Neighbors (KNN) and Random Forest (RF) regressions (Supporting Information, see Sections 6.1. and 6.2.). Among the four, the Ridge regression afforded the clearest correlation for both the nitrosoarene **1** (R^2_{train} = 0.93) and the silyl enol ether **2** training sets (R^2_{train} = 0.93) (Fig. 2C).

Having trained our system to determine **log(*k*/*k*$_0$)** using only *in silico* reagent properties, an experimental test procedure was performed. The trained ML-model was used to determine optimal temperatures to experimentally access $\Delta G^{\ddagger}$ for the reactions between five novel nitrosoarenes (**1c-g**) with **2a**, and between five silyl enol ethers (**2b-f**) with **1a** (Fig. 2C). This was done by classifying reactions in 5 categories according to their computationally derived **log(*k*/*k*$_0$)** value. Each of the 5 categories suggested a temperature and residence time range at which relevant kinetic information could be obtained (Fig. 2C, table). Category 5 reactions, for which no reactivity was predicted using MFRs, were also recorded as negative data.[30,31] This experimental guidance is particularly relevant to design kinetics experiments in MFRs where, contrary to batch procedures, time is defined by the MFR setup, meaning kinetics cannot be performed simply by monitoring the reaction over time, as it is typically done for batch kinetics.

The experiments were performed using an automated flow setup inspired by Jensen *et al*.[32] Our setup incorporated real-time monitoring of temperature, pressure and conversion in four compact units: (1) a dosing unit including feed solutions and syringe pumps; (2) a reactor unit combining a proportional−integral−derivative (PID) temperature controller and a stainless steel coil reactor; (3) a downstream analytical unit featuring an in-line IR spectrometer and (4) a DIY control unit relying on LabVIEW™ (enFLOW) to coordinate all auxiliaries and execute instructions from an Excel sheet.

The data from the test set was plotted into the Ridge regressions which provided a clear correlation for the reactions varying the nitrosoarene (**1c-g**) (R^2_{test} = 0.96) and the reactions varying the silyl enol ethers (**2b-f**) (R^2_{test} = 0.96, Fig 2B). The $\Delta G^{\ddagger}$ values extracted from our ML model afforded a MAE of ~0.2 kcal mol^{-1} and a root-mean-square error (RMSE) of ~0.3 kcal mol^{-1} in comparison to the experimental values (Supporting Information, see Section 7.2.).

Extension to Mayr's Database

This ML model was used to contribute to the reaction prediction scale developed by Mayr which predicts a reaction rate using experimentally derived reactivity parameters for the electrophile (*E*) and the nucleophile (s_N and *N*) according to equation [2].[33–35]

$$\log(k_{20\,^{\circ}C}) = s_N(E + N) \qquad [2]$$

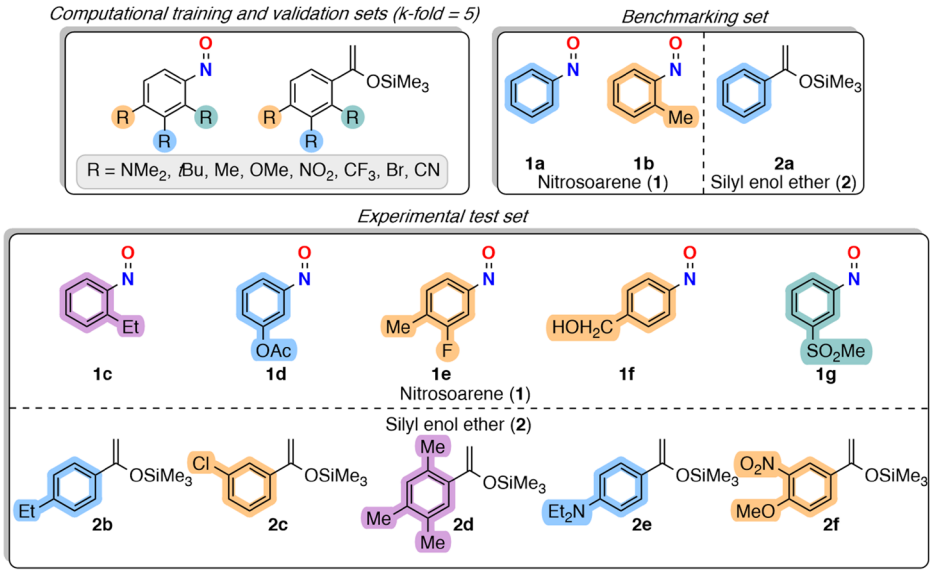
To add nitrosoarene electrophiles **1** to the database, the reactivity parameters (s_N and *N*) associated with our reference nucleophile (**2a**) needed to be obtained in acetonitrile. This was performed by determining the kinetics of the reactions between silyl enol ether **2a** and four azodicarboxylates (**4**, Fig. 3A) for which the reactivity parameter *E* was reported.[36] The

10.1002/anie.202311526

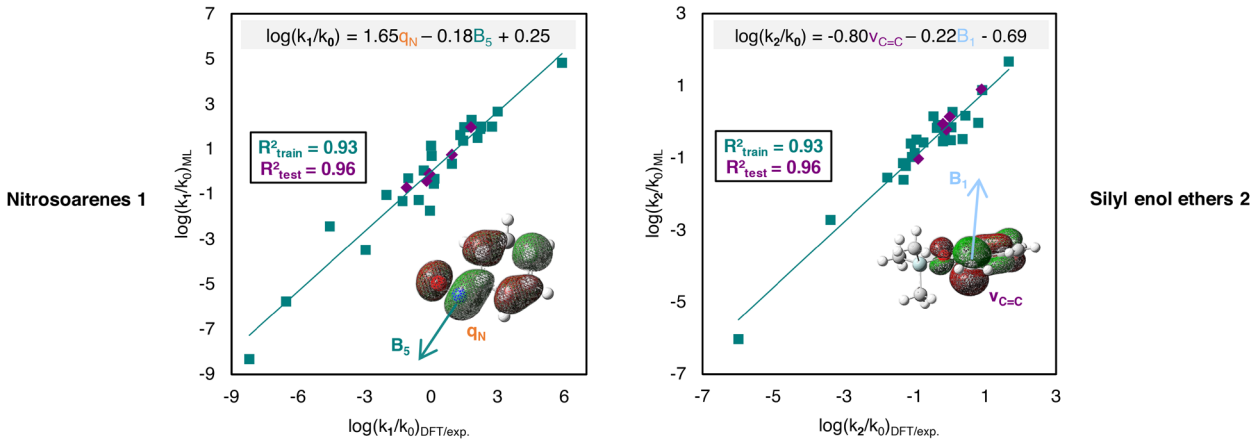
RESEARCH ARTICLE

nucleophilicity parameters for **2a** allowed to calculate the *E* parameter of each nitrosoarene **1a-g** using our kinetics data. The *E* parameters obtained were in the range of -8 to -11, comparable to the values of azodicarboxylates (Supporting Information, see Section 7.3.).

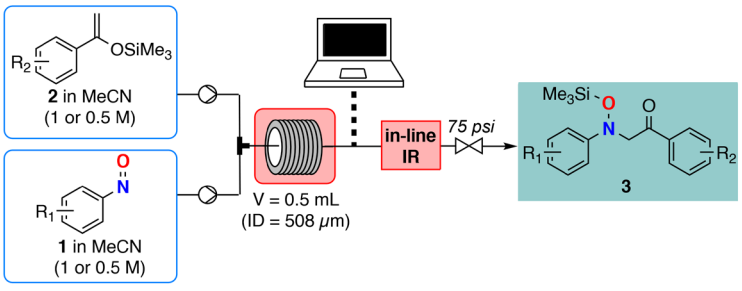
A *Molecular sets*



B *ML predictive model*



C *Experimental testing*



Category	log(*k*/*k*₀)	*t* (min)	T (°C)
1	>1.5	[0.25,1.5]	[30-45]
2	[0.0, 1.5]	[1-6]	[60-90]
3	[0.0, -1.0]	[1-6]	[80-110]
4	[-1.0, -2.0]	[1-6]	[100-130]
5	<-2.0	/	/

Figure 2. Development of the QM/ML predictive model. (**A**) Presentation of the various sets used. Colors represent the category for the kinetics in flow (green: cat. 1, blue: cat. 2, orange: cat. 3; purple: cat. 4). (**B**) Ridge regressions to describe nitrosoarenes **1** (left, drawn orbital: LUMO) or silyl enol ethers **2** (right, drawn orbital: HOMO) for both training and test sets. (**C**) Overview of the experimental testing with the presentation of discriminating categories based on **log(*k*/*k*₀)**. Each category suggested optimum reaction conditions for the kinetics studies.

10.1002/anie.202311526

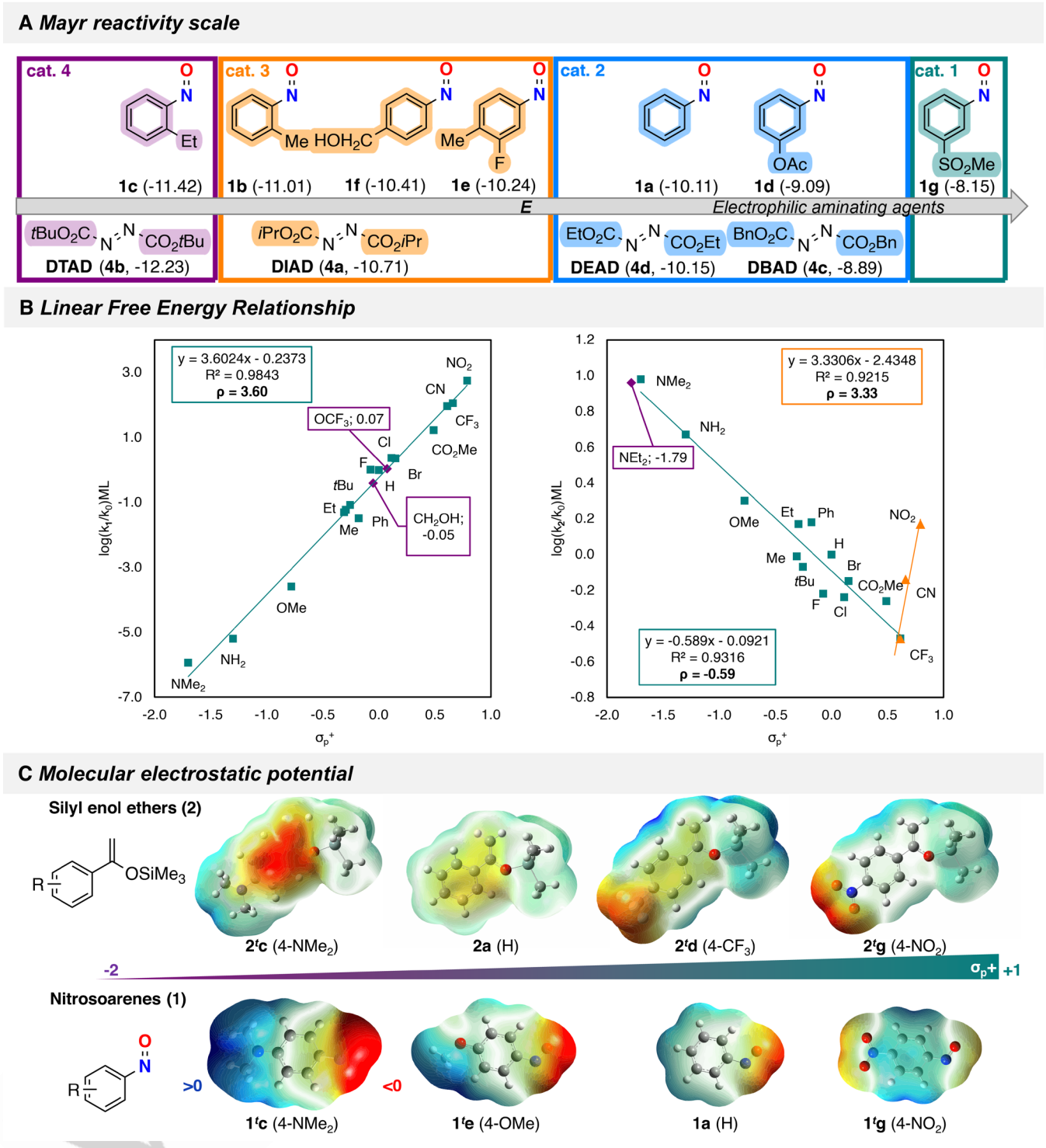


Figure 3. Applications of this work in Physical Organic Chemistry. (**A**) Electrophilicity Mayr indexes determined for nitrosoarenes (**1**), as compared to azodicarboxylates (**4**). (**B**) Linear free-energy relationships for nitrosoarenes (**1**, left) and silyl enol ethers (**2**, right) based on the electrophilic substituent constants σ_p^+ from Brown.[40][Purple values highlight predicted σ_p^+ from unreported substituents. (**C**) Molecular electrostatic potential (MEP) surfaces of various nitrosoarenes (**1**) and silyl enol ethers (**2**) ranked based on the electrophilic substituent constants σ_p^+ from Brown[40] (in red: negative potential, in blue: positive potential).

According to Mayr's categories, their reactions with silyl enol ethers such as **2a** are correctly classified as occurring under thermal activation,[37,38] except for **1g** which reacts spontaneously (Fig. 3A). Aside from contributing the experimentally obtained *E* parameters for a new family of electrophiles to Mayr's database, we offer a ML model that could be used in the future to provide estimates for substituted nitrosoarenes of choice.[39]

10.1002/anie.202311526

Application toward LFERs

Using the ML model, Linear Free-Energy Relationships (LFER) analysis of the reaction mechanism could be performed without any additional experiments. The computed **log(k/k_0)** values for 15 nitrosoarenes **1** and silyl enol ethers **2** substituted at the *para* position were plotted against the corresponding electrophilic substituent constant (σ_p^+) reported by Brown (Fig. 3B).[40]

When the molecular probe is borne by nitrosoarene **1**, the rate of the electrophilic amination is accelerated with increasingly powerful electron withdrawing groups (EWG). This reaction acceleration, indicated by a positive reaction constant (ρ), reflects the incoming electron flow at the transition state. A ρ value of +3.6 highlights the high sensitivity of the reactivity of nitrosoarenes to electronic modulation. On the other hand, when the molecular probe is located on silyl enol ether **2**, two trends are observed. Initially, EWGs moderately deter the reaction due to the decreasing electron density of the nucleophile (ρ = -0.59). However, when strong EWGs are used (*i.e.* 4-CF_3, 4-CN, 4-NO_2), the trend inverts and a ρ value of +3.33 is obtained. Such a drastic change in behavior reflects the ambivalent nature of both nitrosoarenes **1** and silyl enol ethers **2**.[17,41] The incorporation of strong EWGs in **2** results in an inverse electronic demand by which **2** behaves as the electrophile and **1** as the nucleophile. To further confirm this finding, the molecular electrostatic potential (MEP) has been computed, clearly showing a change in the electrophilic and nucleophilic centers (Fig. 3C). The broader dispersion of the scatter plot of **2** emphasizes a higher sensitivity towards steric hindrance, rather than the electronic factors described by Brown's electrophilic constants. This observation agrees with the main descriptor identified for the ML model, $v_{c=c}$, which gathers both electronic and steric information.[42] The ML model can also be used to predict unreported σ_p^+, as illustrated for compounds **1f,j** and **2e** (Fig. 3B).

Predicting optimum reaction conditions

A particularly powerful application of our ML model is the capacity to predict the time needed to reach a target conversion at a given concentration and temperature for a combination of reagents **1**,**2**. We have expanded the use of **log(k/k_0)**-based categories developed for the testing of the ML model (Fig. 2C, table) to rank reactions based on their efficiency in a MFR. We propose to use the **log(k/k_0)** value to sort reactions in five categories based on the predicted residence time. Firstly, because categories depend on reaction conditions, a temperature and concentration need to be defined before a reaction can be sorted. From there, reactions can be considered as follows: *Category 1* (residence time <1 min) for reactions with *outstanding* potential in MFR; *category 2* (residence time of 1-5 min) indicates a *high* potential in MFR; *category 3* reactions (residence time of 5-15 min) are considered to have *medium* potential; and *category 4* (residence time of 15 min) have a *low* potential to be applied in MFR. *Category 5* reactions require more than 15 min under the chosen conditions and are considered *unsuitable for flow*. Considering the properties of our system, 150 °C and 0.50 M (or 0.25 when required for solubility) have been chosen as reference conditions to classify our reactions (Fig. 4B).

Using the ML model, the **log(k/k_0)** for the reactions of three novel nitrosoarenes **1h-j** with **2a**, and for the reactions of three new silyl enol ethers (**2g-i**) with **1a** were calculated. These values, along with the **log(k/k_0)** of the reagents studied thus far (**1a-g** and **2a-f**), were used to assign a category and to calculate the residence time needed to obtain a target conversion (^{t}conv.) between 85 or 95% at 150 °C (5 bars). These instructions were then transferred to the automated flow platform (Fig. 4A), which successfully synthesized a total of 18 new compounds (Fig. 4C) after testing only one set of conditions per reaction (Supporting Information, see Section 7.4.). Noticeably, these conditions were non-generic, each of them being tailored to fit the reactivity of each different compound.

The predicted conditions led to experimental conversions (^{exp}conv) that matched the targeted conversions with high fidelity (MAE ~3% and RMSE ~4%), with the exception of compounds **3c** and **3m**. For the latter, an azoxy side-product **5c,m** (Fig. 4A) was isolated from the reactor effluent, which indicated a hydride transfer from the aminated product **3c,m** to the dimer **1c,m** as a side-reaction.[43] Increasing the excess of silyl enol ether to 3 equivalents reinstated the expected conversion.

When high selectivity was obtained, the pure product could be isolated after evaporation of the effluent under reduced pressure. For reactions requiring purification, the isolation of the product was sometimes problematic due to its instability. Nine aminated products were isolated in excellent yields (**3a,d-h**, 86-91%) and seven in moderate isolated yields (**3b,c,i,k,l,o,q**, 60-79%), while products **3m** (23%) and **3r** (42%) were isolated in lower amounts as a result of TMS deprotection or degradation into benzoic acid.[44] Predictions were not impacted by these side-reactions since they occurred after the electrophilic amination (Fig. 4A).

Because substituent effects are additive, our model is not limited to predicting reactions using necessarily either **1a** or **2a**. When considering a reaction between both a new nitrosoarene **1k** and new silyl enol ether **2j** (Fig 4C, **3s**), the relative rate constant **log(k_{1k2j}/k_0)** can be obtained by adding the predicted values for **1k** and of **2j** compared to a same model reaction k_0 (i.e., **log(k_{1k}/k_0)** + **log(k_{2j}/k_0)**). Nitrosoarene **1k** is predicted to be an outstanding electrophile, with a **log(k_{1k}/k_0)** value of +2.0 (*category 1*). Silyl enol ether **2j** is a poorly reactive nucleophile, with a predicted **log(k_{1k}/k_0)** of -1.3 (*category 4*). Their combination predicted a *category 2* reaction with a **log($k_{(1k2j)}/k_0$)** of 0.7, indicating an overall high potential in flow. The predicted conversion (^{t}conv: 95%) was successfully obtained experimentally (^{exp}conv: 96%).

Conclusion

When chemists optimize reactions in MFRs, they typically either feed upon preliminary batch experiments (prior art or trials), trials and errors or guided trials (DoE) in available flow systems. This optimization process is time-, resource-consuming, and produces ample waste. This is particularly concerning when the availability of resources is scarce, when toxic/high activity or unstable compounds are involved. We have developed a different approach. A ML-DFT was constructed for the expedient computation of an *in silico* kinetic metric, **log(k/k_0)**, which predicts with high fidelity experimental data (MAE ~0.2 kcal mol^{-1} and RMSE ~0.3 kcal mol^{-1}). This tool was successfully applied to nitrosoarene electrophilic aminations for which no prior experimental data exists. In addition to drastically accelerating optimization, this tool also offered a powerful gateway to

10.1002/anie.202311526

RESEARCH ARTICLE

incorporate data into known Physical Organic Chemistry databases (Mayr's reactivity scale and LFER). The model here reported is based on the properties of the reagents and therefore is limited to exothermic reactions in which the transition state is closer in energy and structure to the reagents. However, the construction of an analogous model applied to endothermic reactions could be envisaged, as well as models accommodating other factors excluded here such as catalyzed processes or non-homogeneous conditions. Complementary tools to analyze competitive pathways and to suggest corrective actions should also be developed in the future.

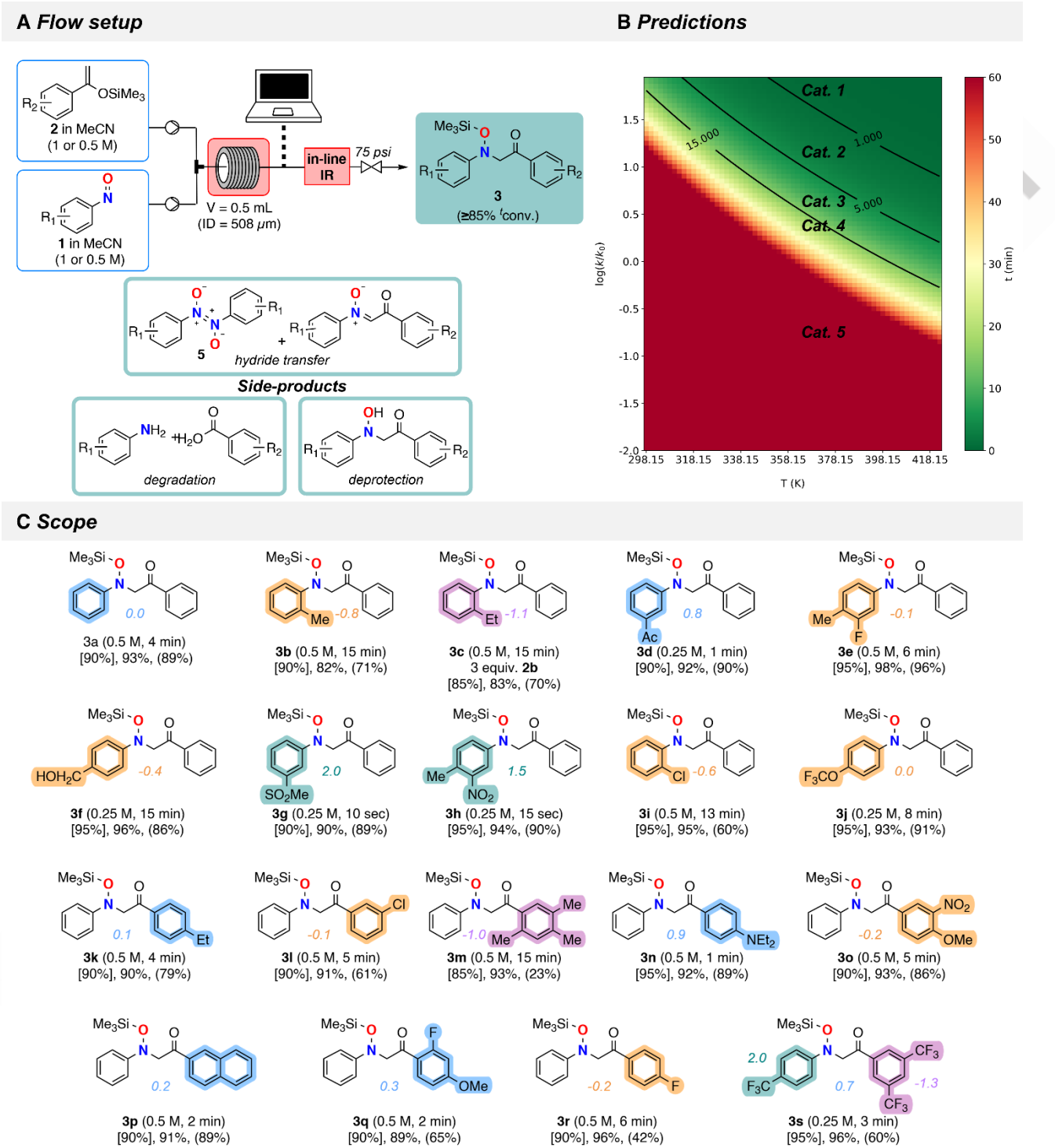


Figure 4. Applications of this work in synthetic organic chemistry. (**A**) Overview of the flow setup for the synthesis of silylated *α*-hydroxylaminated ketones **3** and description of the side-products. (**B**) Illustration of a predictive model to perform the reactions at 0.5 M to obtain 95% conversion. Black lines represent combinations of temperature and **log(*k*/*k*$_0$)** to reach 95% conversion within 1, 5 or 15 min of residence time. (**C**) Products synthesized using conditions predicted by the ML model. Conversions and yields were determined by off-line HPLC analyses using calibration curves at the maximal absorption of each species (DAD detector). Predicted conversions: [^{t}conv.] – Experimental conversions: ^{exp}conv. – Yields: (yield). Values of **log(*k*/*k*$_0$)** are written next to the molecule. Colors represent the category (green: cat. 1 – outstanding, blue: cat. 2 – high, orange: cat. 3 – medium; purple: cat. 4 – low).

10.1002/anie.202311526

The computation power required to develop this type of model is available to many chemistry labs, facilitating its widespread adoption. To assist any person wishing to explore this approach, we have included a step-by-step tutorial in our supporting information (see Supporting Information, Scheme S1). Indeed, the growing accessibility of quantum mechanics and machine learning makes it increasingly hard to justify a guesstimate approach to synthesis. This is particularly crucial for flow syntheses where the expense of re-writing centuries of batch protocols threatens to outweigh the potential benefits of flow. It is in great deal the lack of rational ways of performing this translation that has deferred the transformation that flow technology was once hailed to bring to the chemical world.

Aside from the pragmatic purpose of reaction optimization, the ease these tools provide to advance physical organic chemistry descriptors is an invitation to increase the depth of analysis of chemical developments. The DFT methods we identified also go beyond the application shown here, being sufficiently accurate to immediately provide an intensified temperature protocol that would suggest if and how a known batch reaction can be applied in flow. It is meant to serve, in a way, as a *chemical translator* from batch to flow. We expect that expediting the development of flow chemistry will have further repercussions in industrial reshoring efforts, by providing safer and greener methods in accordance with modern regulations.

If more than one process needs to be evaluated, e.g., when constructing chemical libraries, the ML model becomes a door towards rational synthesis development. Because the ML model provides reactivity predictions as well as mechanistic insight solely through computational data, it reduces the need for experimentation. This is showcased by the successful use of categories based purely on quantum mechanics and machine learning to predict optimal conditions to perform the testing. Although test sets are an important part of ML methods, the strength of our approach lies in the construction of a model with predictive power based purely on computational data. In other words, reactions and conditions for which no previous knowledge exists, can be preliminarily modeled without experimentation. The ML model becomes therefore a scout for the vast and new chemical space that flow chemistry provides.

Acknowledgements

We thank Prof. Raphaël Robiette (Université catholique de Louvain, Belgium) for the stimulating discussions on computational chemistry; Prof. Kristof Van Hecke (Ghent University, Belgium) for the X-ray diffraction on samples **3a** and **1g**; Laurent Collard (Université catholique de Louvain, Belgium) for the HRMS analyses on compounds **1g, 1h**, **1i**, **2e**, **2f**, **2h**, and **3a-s**; Michaël Schmitz (CiTOS, University of Liège, Belgium) and Dr. Martin Cattoen (CiTOS, University of Liège, Belgium) for elements of the automated system design and the flow platform. Dr. Diana Silva (CiTOS, University of Liège, Belgium), Dr. Yi-Hsuan Tsai (CiTOS, University of Liège, Belgium) are acknowledged for helpful discussions related to the preparation of the manuscript and for the proofreading of the manuscript. We also thank Sébastien Schoenmaeckers (ISLV, University of Liège, Belgium) for the proofreading of the manuscript.

This work is supported by the Walloon Region as part of the funding for the FRFS-WEL-T strategic axis. The authors acknowledge the WEL Research Institute (grant WEL-T-CR-2023 A – 05, "Smart Flow Systems"), the University of Liège ("Crédit d'opportunité stratégique du Conseil universitaire de la Recherche et de la Valorisation" under Grant No. OPP_CURV_22-44) and the F.R.S.-FNRS (Incentive grant for scientific research MIS F453020F, J.C.M.M.; PhD Fellowship ASP 1.A.054.21F, P.B.) for funding. Computational resources were provided by the "Consortium des Équipements de Calcul Intensif" (CÉCI), funded by the "Fonds de la Recherche Scientifique de Belgique" (F.R.S.-FNRS) under Grant No. 2.5020.11a and by the Walloon Region.

Keywords: • Amination • Automation • Continuous Flow • Density Functional Calculations • Machine Learning

[1] A. Adamo, R. L. Beingessner, M. Behnam, J. Chen, T. F. Jamison, K. F. Jensen, J. C. M. Monbaliu, A. S. Myerson, E. M. Revalor, D. R. Snead, T. Stelzer, N. Weeranoppanant, S. Y. Wong, P. Zhang, *Science (80-.).* **2016**, *352*, 61–67.

[2] A. Bonner, A. Loftus, A. C. Padgham, M. Baumann, *Org. Biomol. Chem.* **2021**, *19*, 7737–7753.

[3] T. Noel, L. Capaldo, Z. Wen, *Chem. Sci.* **2023**, *14*, 4230–4247.

[4] M. I. Jeraal, S. Sung, A. A. Lapkin, *Chemistry–Methods* **2021**, *1*, 71–77.

[5] O. J. Kershaw, A. D. Clayton, J. A. Manson, A. Barthelme, J. Pavey, P. Peach, J. Mustakis, R. M. Howard, T. W. Chamberlain, N. J. Warren, R. A. Bourne, *Chem. Eng. J.* **2023**, *451*, 138443.

[6] P. Sagmeister, R. Lebl, I. Castillo, J. Rehrl, J. Kruisz, M. Sipek, M. Horn, S. Sacher, D. Cantillo, J. D. Williams, C. O. Kappe, *Angew. Chemie - Int. Ed.* **2021**, *60*, 8139–8148.

[7] C. P. Breen, A. M. K. Nambiar, T. F. Jamison, K. F. Jensen, *Trends Chem.* **2021**, *3*, 373–386.

[8] S. Chatterjee, M. Guidi, P. H. Seeberger, K. Gilmore, *Nature* **2020**, *579*, 379–384.

[9] A. M. K. Nambiar, C. P. Breen, T. Hart, T. Kulesza, T. F. Jamison, K. F. Jensen, *ACS Cent. Sci.* **2022**, *8*, 825–836.

[10] A. A. Volk, R. W. Epps, D. T. Yonemoto, B. S. Masters, F. N. Castellano, K. G. Reyes, M. Abolhasani, *Nat. Commun.* **2023**, *14*, 1403.

[11] C. Holtze, R. Boehling, *Curr. Opin. Chem. Eng.* **2022**, *36*, 100798.

[12] V. Hessel, D. Kralisch, N. Kockmann, T. Noël, Q. Wang, *ChemSusChem* **2013**, *6*, 746–789.

[13] P. Mueller, A. D. Clayton, J. Manson, S. Riley, O. S. May, N. Govan, S. Notman, S. V. Ley, T. W. Chamberlain, R. A. Bourne, *React. Chem. Eng.* **2022**, *7*, 987–993.

[14] P. P. Plehiers, C. W. Coley, H. Gao, F. H. Vermeire, M. R. Dobbelaere, C. V. Stevens, K. M. Van Geem, W. H. Green, *Front. Chem. Eng.* **2020**, *2*, 5, DOI 10.3389/fceng.2020.00005.

[15] C. B. Santiago, J. Y. Guo, M. S. Sigman, *Chem. Sci.* **2018**, *9*, 2398–2412.

[16] P. Bianchi, A. Dubart, M. Moors, D. Cornut, G. Duhirwe, J. Ampurdanés, J.-C. M. Monbaliu, *React. Chem. Eng.* **2023**, *8*, 1565–1575.

[17] S. Dana, I. Ramakrishna, M. Baidya, *Synth.* **2017**, *49*, 3281–3290.

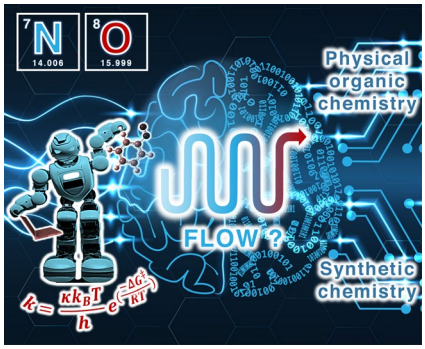
[18] V. E. H. Kassin, R. Morodo, T. Toupy, I. Jacquemin, K. Van Hecke,

R. Robiette, J. C. M. Monbaliu, *Green Chem.* **2021**, *23*, 2336–2351.
[19] P. Bianchi, J.-C. M. Monbaliu, *Org. Chem. Front.* **2022**, *9*, 223–264.
[20] T. Sasaki, Y. Ishibashi, M. Ohno, *Chem. Lett.* **1983**, *12*, 863–866.
[21] K. Jorner, T. Brinck, P. O. Norrby, D. Buttar, *Chem. Sci.* **2021**, *12*, 1163–1175.
[22] K. Spiekermann, L. Pattanaik, W. H. Green, *Sci. Data* **2022**, *9*, 417.
[23] E. H. E. Farrar, M. N. Grayson, *Chem. Sci.* **2022**, *13*, 7594–7603.
[24] N. Mardirossian, M. Head-Gordon, *Mol. Phys.* **2017**, *115*, 2315–2372.
[25] I. Funes-Ardoiz, R. S. Paton, **2018**, DOI 10.5281/ZENODO.1435820.
[26] J. M. Crawford, C. Kingston, F. D. Toste, M. S. Sigman, *Acc. Chem. Res.* **2021**, *54*, 3136–3148.
[27] V. Dhayalan, S. C. Gadekar, Z. Alassad, A. Milo, *Nat. Chem.* **2019**, *11*, 543–551.
[28] S. K. Nistanaki, C. G. Williams, B. Wigman, J. J. Wong, B. C. Haas, S. Popov, J. Werth, M. S. Sigman, K. N. Houk, H. M. Nelson, *Science (80-.).* **2022**, *378*, 1085–1091.
[29] H. S. Peleg, A. Milo, *Angew. Chemie - Int. Ed.* **2023**, *62*, e202219070.
[30] F. Strieth-Kalthoff, F. Sandfort, M. Kühnemund, F. R. Schäfer, H. Kuchen, F. Glorius, *Angew. Chemie - Int. Ed.* **2022**, *134*, e202204647.
[31] M. P. Maloney, C. W. Coley, S. Genheden, N. Carson, P. Helquist, P.-O. Norrby, O. Wiest, *J. Org. Chem.* **2023**, *88*, 5239–5241.
[32] T. Hart, V. L. Schultz, D. Thomas, T. Kulesza, K. F. Jensen, *Org. Process Res. Dev.* **2020**, *24*, 2105–2112.
[33] H. Mayr, T. Bug, M. F. Gotta, N. Hering, B. Irrgang, B. Janker, B. Kempf, R. Loos, A. R. Ofial, G. Remennikov, H. Schimmel, *J. Am. Chem. Soc.* **2001**, *123*, 9500–9512.
[34] H. Mayr, A. R. Ofial, *Pure Appl. Chem.* **2005**, *77*, 1807–1821.
[35] H. Mayr, A. R. Ofial, *J. Phys. Org. Chem.* **2008**, *21*, 584–595.
[36] T. Kanzian, H. Mayr, *Chem. - A Eur. J.* **2010**, *16*, 11670–11677.
[37] R. Lucius, R. Loos, H. Mayr, *Angew. Chemie - Int. Ed.* **2002**, *41*, 91–95.
[38] A. I. Leonov, D. S. Timofeeva, A. R. Ofial, H. Mayr, *Synth.* **2019**, *51*, 1157–1170.
[39] M. Vahl, J. Proppe, *Phys. Chem. Chem. Phys.* **2023**, *25*, 2717–2728.
[40] H. C. Brown, Y. Okamoto, *J. Am. Chem. Soc.* **1958**, *80*, 4979–4987.
[41] V. García-Vázquez, A. Carretero Cerdán, A. Sanz-Marco, E. Gómez-Bengoa, B. Martín-Matute, *Chem. - A Eur. J.* **2022**, *28*, e202201000.
[42] A. Milo, E. N. Bess, M. S. Sigman, *Nature* **2014**, *507*, 210–214.
[43] C. Volpe, S. Meninno, A. Roselli, M. Mancinelli, A. Mazzanti, A. Lattanzi, *Adv. Synth. Catal.* **2020**, *362*, 5457–5466.
[44] M. Ohno, M. Ido, S. Eguchi, *J. Chem. Soc. Chem. Commun.* **1988**, 1530–1531.

10.1002/anie.202311526

RESEARCH ARTICLE

Entry for the Table of Contents



To flow or not to flow ? Assessing feasibility and guiding chemists with *a priori* computational intelligence for reaction optimization under continuous flow conditions.