nature chemical engineering

Article

https://doi.org/10.1038/s44286-024-00033-5

Autonomous reaction Pareto-front mapping with a self-driving catalysis laboratory

Received: 29 July 2023

Accepted: 25 January 2024

Published online: 27 February 2024

Check for updates

J. A. Bennett[1], N. Orouji[1], M. Khan[1], S. Sadeghi[1], J. Rodgers[2] & M. Abolhasani[1]

Ligands play a crucial role in enabling challenging chemical transformations with transition metal-mediated homogeneous catalysts. Despite their undisputed role in homogeneous catalysis, discovery and development of ligands have proven to be a challenging and resource-intensive undertaking. Here, in response, we present a self-driving catalysis laboratory, Fast-Cat, for autonomous and resource-efficient parameter space navigation and Pareto-front mapping of high-temperature, high-pressure, gas–liquid reactions. Fast-Cat enables autonomous ligand benchmarking and multi-objective catalyst performance evaluation with minimal human intervention. Specifically, we utilize Fast-Cat to perform rapid Pareto-front identification of the hydroformylation reaction between syngas (CO and H_2) and olefin (1-octene) in the presence of rhodium and various classes of phosphorus-based ligands. By reactor benchmarking, we demonstrate Fast-Cat's knowledge scalability, essential to fine/specialty chemical industries. We report the details of the modular flow chemistry platform of Fast-Cat and its autonomous experiment-selection strategy for the rapid generation of optimized experimental conditions and in-house data required for supplying machine-learning approaches to reaction and ligand investigations.

Transition metal-based homogeneous catalysts and their associated complexing ligands play a major role in many widespread chemical processes such as hydrogenations, carbonylations, oxidations and various coupling reactions[1]. These reactions are present in bulk chemical syntheses up to millions of tons per year as well as more specialized fine chemical and pharmaceutical syntheses with diverse and challenging transformations[2–4]. Finding new catalysts to perform organic synthesis steps more efficiently can lead to dramatic savings in energy and chemical requirements for both the reaction as well as downstream purification steps.

Ligand discovery in homogeneous catalysis remains a multifaceted problem that requires both molecular structure and reaction condition optimization. The most optimal ligand may result in poor or no catalytic activity at all if used in a homogeneous transition metal-catalyzed reaction with incorrect process conditions, and no two ligands will have identical optimum conditions. This variability in ligand performance (ligand and reaction condition-dependent yield and regioselectivity) discourages classical fixed-condition or design of experiment reaction-screening strategies that may not truly capture a candidate ligand's true potential. To address this complex problem, a different catalyst/ligand system benchmarking strategy is required to efficiently explore a given ligand's available reaction space and ensure the true extent of the ligand's capabilities is used when comparing performance (Extended Data Fig. 1). Manual one-parameter-at-a-time experimentation is generally limited by labor, precursor availability and time, which can make finding optima with multiple competing output metrics challenging. Automation can help by speeding up sampling for high-throughput experimentation and incorporating design of experiments methods for evenly covering the reaction space, but a large amount of experimental data are still required to obtain a holistic

[1]Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC, USA. [2]Eastman Chemical Company, Kingsport, TN, USA. e-mail: abolhasani@ncsu.edu

view of the reaction system. Autonomous systems seek to use closed-loop feedback with dynamically evolving understanding of the reaction system to select new experimental conditions more efficiently, with the goal of minimizing the experimental cost of exploring the limits of a new reaction system.

Rhodium (Rh)-catalyzed hydroformylation (Fig. 1) of olefins to produce linear and branched aldehydes is a prominent homogeneous reaction in contemporary industrial processes, with broad reaching applications ranging from commodity chemicals to specialty pharmaceuticals and fragrances[1–4]. A substantial amount of work has been performed in academia and industry with the aim of developing specialty ligands to help direct the Rh center into producing the desired aldehyde product with high selectivity[5–14]. For the most part, ligands have a trade-off between activity and regioselectivity, that is, highly active ligands[15,16] are generally less selective and highly selective ligands[17,18] are generally less active. Common bidentate electron-rich phosphines, such as Xantphos and BISBI, are highly selective towards linear aldehydes with turnover frequencies (TOFs) on the order of 1,000 mol of aldehyde per mol of Rh per hour[14,19,20], while the more active bulky monophosphites approach 10,000 mol of aldehyde per mol of rhodium per hour[15,16] but only have linear-to-branched (l/b) product aldehyde ratios of ~2:1. Various ligands promoting the hydroformylation of internal olefins have been reported[6–8,10,18,21,22], but double bond migration remains a challenge, as well as direct branched selectivity[23] from terminal olefins such as isobutyraldehyde from propylene.

Closed-loop autonomous reaction exploration using self-driving laboratories (SDLs) is a rapidly emerging technology used in accelerated materials and molecular discovery and development[24–44,68–70]. This autonomous experimentation strategy can accelerate multi-objective process optimization to meet specific process requirements by 10–100× compared with conventional manual experimentation strategies[32–34,41,42,45,46]. The core of SDLs lies in automatic generation of experimental data (automation) orchestrated by machine learning (ML) algorithms. In each experimental campaign, the brain of the SDL, that is, the ML algorithm, predicts system response and iteratively acquires an improved set of reaction conditions until the desired objective is attained. Existing SDL software packages, including ChemOS[39] and ARES OS[47], are an open-access, user-friendly starting point for researchers in chemical and material sciences to begin implementing autonomous experimentation. Developing an in-house ML algorithm for a new SDL can allow for fine tuning of model hyper-parameters and architecture for complex reaction spaces, such as ligand-assisted transition metal catalysis. In addition to closed-loop autonomous experimentation using the physical setup of the SDL, the in-house generated experimental data can be utilized to build a digital twin of the SDL hardware using various modeling techniques (for example, Gaussian process regression or deep neural networks, DNNs). Such digital twins can then be employed to query the reaction for a specific condition, visualize the reaction space and investigate the role of each process parameter on the catalytic performance of each catalyst/ligand.

The lack of access to comprehensive experimental data (success and failure conditions) and metadata (detailed reaction and reactor parameters) for transition metal-catalyzed homogeneous catalysis, including hydroformylation reactions, necessitates development of a robust experimental platform to generate big data of high-quality (reproducible with minimum experimental noise) and scalable experimental data to be leveraged with data science tools of SDLs.

Although automated high-throughput experimentation platforms can assist with catalyst screening through parallel batch operation to expand the scope of discreet parameters[37], they can be limited by start-up and shut-down times, slower heat and mass transfer rates and larger amounts of experimental waste. Flow reactors offer a number of benefits in generating large quantities of high-quality experimental data that can be fed into the data science tools of SDLs, facilitating accelerated catalyst discovery and development. One of the key advantages of flow chemistry is the ability to achieve continuous control of reaction conditions[41,42,48–52]. Unlike traditional batch reactors, flow reactors allow for rapid switching between experimental conditions without requiring lengthy downtime or the need for manual cleaning or excessive waste generation. Furthermore, flow reactors provide enhanced heat and mass transport rates compared with conventional batch reactors[48–50,53–56]. This transport rate intensification leads to increased reaction rates and faster thermal equilibration, enabling rapid exploration of reaction conditions and catalyst performance. It should be noted that flow reactors are not without some of their own unique drawbacks, such as difficulty in handling solid species and more complex initial experimental setup. In a flow chemistry platform, in-line reaction characterization or sampling techniques can be readily integrated into the analysis sequence of a given experimental condition. The in-line characterization means that as reactions progress, real-time experimental data can be collected and analyzed. By leveraging closed-loop automated processing of the in-line generated experimental data, the SDL equipped with a flow chemistry platform can continuously update its belief model of the reaction being studied and refine its hypothesis regarding specific catalyst reactivity and the key process parameters that control catalyst performance.

The combination of continuous experimentation, improved transport properties and in-line analysis, as well as reduced chemical consumption and waste generation, enabled by flow reactors, empowers SDLs to rapidly generate and process large volumes of experimental homogeneous catalysis data. These in-house generated experimental data can be efficiently utilized by ML algorithms of SDLs to identify hidden and complex correlations, patterns and trends, ultimately accelerating the catalyst discovery and process development. The ability to rapidly iterate, explore and optimize catalyst performance based on real-time experimental data insights is a major advantage offered by SDLs equipped with flow reactors.

In this Article, we present a fully autonomous self-driving fluidic technology for catalysis acceleration and testing (Fast-Cat). Fast-Cat employs a closed-loop homogeneous catalysis system using a flow chemistry platform to autonomously investigate and expedite the exploration of catalyst/ligand systems. Fast-Cat incorporates modular gas–liquid flow reactors, in-line reaction characterization, advanced process automation and ML techniques to accelerate ligand discovery, catalyst benchmarking and process development for specialty and fine chemicals.

After developing, characterizing and benchmarking the hardware of Fast-Cat versus large-scale conventional batch reactors, we employ it to rapidly map the ligand–regioselectivity–yield relationship for the hydroformylation of olefins. Within a span of only 5 days of continuous autonomous experimentation with no human intervention (45 reactions), Fast-Cat showcases its prowess by swiftly identifying the Pareto front for any desired ligand in the hydroformylation of 1-octene. We report the reaction data and metadata for six full Pareto-front optimization campaigns of selected hydroformylation ligands (Fig. 1), including all tested reactions. Additionally, we highlight an aspect of Fast-Cat, which is the generation of high-quality experimental data used to construct a digital twin of the homogeneous catalytic reaction under investigation. The trained digital twins using the in-house generated experimental data by Fast-Cat serve as virtual representations that aid in understanding the underlying reaction mechanism. Leveraging the information gleaned from the digital twins, we then explore the key process parameters governing product aldehyde yield and regioselectivity.

The research acceleration framework of Fast-Cat enables rapid and comprehensive studies of high-temperature/pressure, metal- and ligand-mediated gas–liquid reactions, ubiquitous in specialty/fine chemical industries. Specifically, we demonstrate that self-driving catalysis laboatories can dramatically enhance the speed and

https://doi.org/10.1038/s44286-024-00033-5

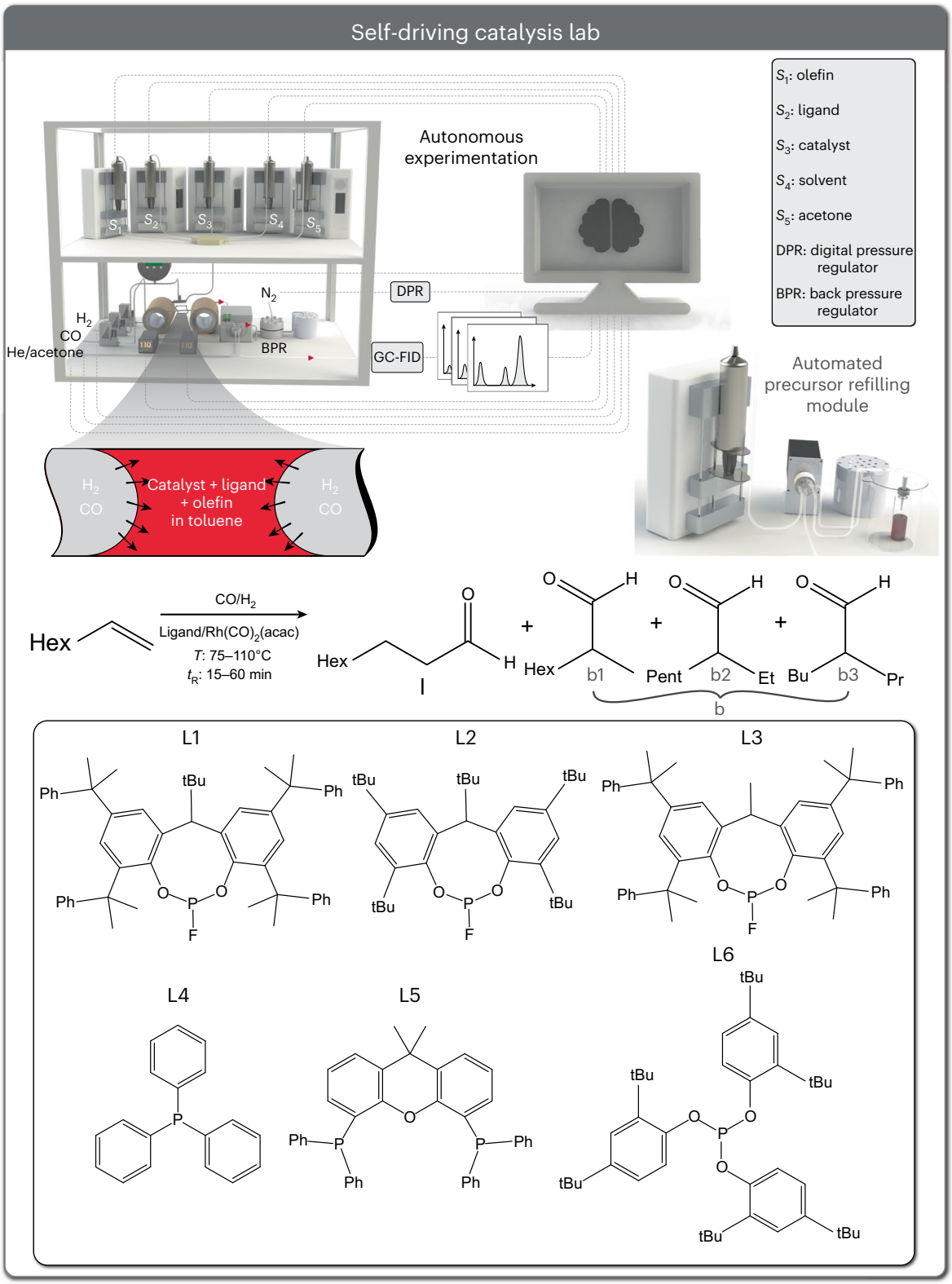


Fig. 1 | Schematic of the developed self-driving catalysis lab technology. Homogeneous catalyst/ligand development is accelerated using autonomous flow reactors with a general reaction scheme for the hydroformylation of 1-octene to all aldehyde products. Product characterization is performed using gas chromatography equipped with a flame ionization detector (GC-FID). Autonomously mapped ligands by Fast-Cat. L1, 12-(*tert*-butyl)-6-fluoro-2,4,8,10-tetrakis(2-phenylpropan-2-yl)-12*H*-dibenzo[d,g][1,3,2] dioxaphosphocine; L2, 2,4,8,10,12-penta-*tert*-butyl-6-fluoro-12*H*-dibenzo[d,g][1,3,2]dioxaphosphocine; L3, 6-fluoro-12-methyl-2,4,8,10-tetrakis(2-phenylpropan-2-yl)-12*H*-dibenzo[d,g][1,3,2]dioxaphosphocine; L4, triphenylphosphine; L5, Xantphos, L6, tris(2,4-di-*tert*-butylphenyl) phosphite.

experimental efficiency of reaction space exploration for similar reaction systems by allowing for robust and comprehensive evaluation of the performance capabilities of catalyst species. Furthermore, Fast-Cat unveils the role of ligand structure on the important reaction parameter controlling the product regioselectivity and yield. Fast-Cat closes the knowledge scalability gap of SDLs by demonstrating the scalability of the machine-discovered knowledge from miniaturized flow to batch reactors. The insights gained from Fast-Cat's investigations offer avenues for exploring new ligands, optimizing catalyst performance and advancing the realm of homogeneous hydroformylation.

Results

Fast-Cat's hardware

The automated flow chemistry platform of Fast-Cat (Fig. 1, see 'Experimental setup' section in Supplementary Information) comprises five modules, including precursor loading/formulation, precursor refilling, flow reactor, reaction sampling and in-line characterization using gas chromatography (GC).

The hydroformylation of 1-octene requires a liquid phase containing the olefin, Rh catalyst, ligand and solvent as well as a gas phase to supply the syngas. The phosphorus-based ligands used in this work are detailed in Fig. 1. These ligands were selected to represent a diverse set of phosphine and phosphite ligands for hydroformylation of olefins. Ligand selection was driven by structural diversity and prior literature knowledge. L1[57] was selected as a ligand candidate with a relatively large flexibility range of regioselectivity (that is, being able to produce primarily linear and primarily branched products at unique reaction conditions). L2 and L3 were selected as closely related fluorophosphite candidates with steric structure modifications. L4, L5 and L6 were selected to expand the scope of Fast-Cat classified ligands into common mono- and bis-phosphines as well as phosphite candidates.

In an industrial setting, gas–liquid reactions are commonly performed in a gas-fed continuous-stirred tank reactor (CSTR)[58,59]. However for small-scale research applications, segmented gas–liquid flow format (Fig. 1) is a great alternative with benefits to process safety through reduced liquid and gas holdup as well as improved heat and gas–liquid mass transfer rates compared with a batch (autoclave) reactor[48–50,53]. The primary reason for utilizing a gas–liquid segmented flow format versus other reactor configurations (for example, CSTR or batch) for Fast-Cat's hardware was minimization of the chemical consumption and waste generation. While the utilized segmented flow format by Fast-Cat is miniaturized in comparison with CSTRs or batch reactors, it should be noted that it is possible to map conditions from a miniaturized segmented flow reactor and achieve equivalent results in a batch reactor as well as extracting fundamental knowledge and trends from the in-flow obtained experimental data. In the gas–liquid segmented flow configuration, the liquid and gas streams are mixed to the desired composition independently and then brought together at a fluidic T-junction, causing the liquid phase to be broken up into a train of uniform slugs by the reactive gas phase (Fig. 1). The no-slip boundary conditions between the liquid slugs and the inside of the reactor tube walls sets up an axisymmetric recirculation pattern within each liquid slug[60,61], lowering the diffusive length scale and improving the transport of the syngas from the gas–liquid interface at either end of the liquid slug.

Fast-Cat's ML brain

The digital component of Fast-Cat includes automated data processing of the in-flow hydroformylation reactions and automated experiment selection. The ML sequence of Fast-Cat begins by reading the in-house database of previously evaluated experimental conditions. The reaction data consist of seven input parameters (X_i, i = 1:7) corresponding to reaction conditions (CO flowrate, H_2 flowrate, total reaction pressure, reactor temperature, dilution, ligand-to-Rh ratio and olefin-to-Rh ratio) and two objective values (total aldehyde yield and fraction of linear aldehyde versus all aldehydes, S_N). The belief model (DNN) of the current reaction system is then automatically created with the supplied experimental data generated by Fast-Cat. The DNN model used by Fast-Cat is an ensemble of randomized cascade-feed forward networks. Each sub model of the ensemble supplies an individual output (model uncertainty) and contributes to an aggregated ensemble output (average model prediction value). The output of the surrogate ensemble model (average model value and uncertainty) becomes the starting point for a Bayesian optimization framework of Fast-Cat.

The primary acquisition function used in this work is batched noisy expected hypervolume improvement (qNEHVI[62], Fig. 2), which is a multi-objective policy relative to the classical Bayesian single-objective expected improvement. Classical expected improvement ranks predicted belief model outputs by the expectation of the predicted objective value being better than the best observed objective value. Applying a fixed weighting to the target objectives before sampling can bias the reaction space search. Exploring the Pareto boundary of the objective space and then selecting a satisfactory objective tradeoff is more expensive in terms of sampling but much more likely to result in the global optimum in the explored reaction space. qNEHVI achieves these criteria by ranking future experimental conditions based on their potential in improving on the best-seen data in any objective (that is, pushing the Pareto boundary outwards, increasing the contained hypervolume). The cycle starts with building a model of the system on a small amount of initial data. The observed experimental conditions are then partitioned into the Pareto and dominated set, and the current hypervolume is calculated. Samples are then taken from the model and added to the partitioning. The difference in the Pareto front containing the predicted value and the Pareto front of the previously observed experimental data is the difference in the total hypervolumes enclosed by each Pareto front compared with a fixed reference point (Fig. 2). This Pareto mapping strategy explores reaction conditions that improve on the reaction yield or regioselectivity or a combination of the two for a hydroformylation of terminal olefins without bias. During the optimization, the observed Pareto front should converge on the ground truth (that is, true objective limits of the reaction system within the available experimental space). Once an autonomous Pareto-front campaign is completed, the aggregated data consisting of pairs of reaction conditions and outcomes are compiled in a file for further meta-analysis and building of the digital twin of the reaction system.

Fast-Cat hardware validation and benchmarking

An important feature of any SDL is the scalability of the produced knowledge. If the SDL-discovered knowledge (that is, optimized process conditions) is dependent on the physical geometry and characteristics of the SDL hardware, it will pose substantial challenges for scaling up and industrial adoption. Thus, before autonomous Pareto-front mapping of different ligands, the Fast-Cat system was evaluated for stability/reproducibility and knowledge scalability by benchmarking it against standard autoclave batch reactors using a bulky, cyclic fluorophosphite ligand, L1 (ref. 57) (see 'In-flow mixing', 'Batch versus flow benchmarking' and 'Auto-injection replication' sections in Supplementary Information).

An important information for reliable reaction sampling from Fast-Cat's physical platform is its response time needed to reach steady state for each independent input parameter. Fast-Cat's response to changes in total reaction pressure and temperature setpoints was investigated ('Temperature and pressure system response' section in Supplementary Information). The system responses were determined to be approximately a single residence time of the Fast-Cat operation, and the transient flow regime was determined to be approximately three residence times (Extended Data Fig. 2a).

The reproducibility uncertainty of a single reaction condition across repeated automated syringe refills and multiple days of reaction operation was found to be less than 5% (Extended Data Fig. 2b).

https://doi.org/10.1038/s44286-024-00033-5

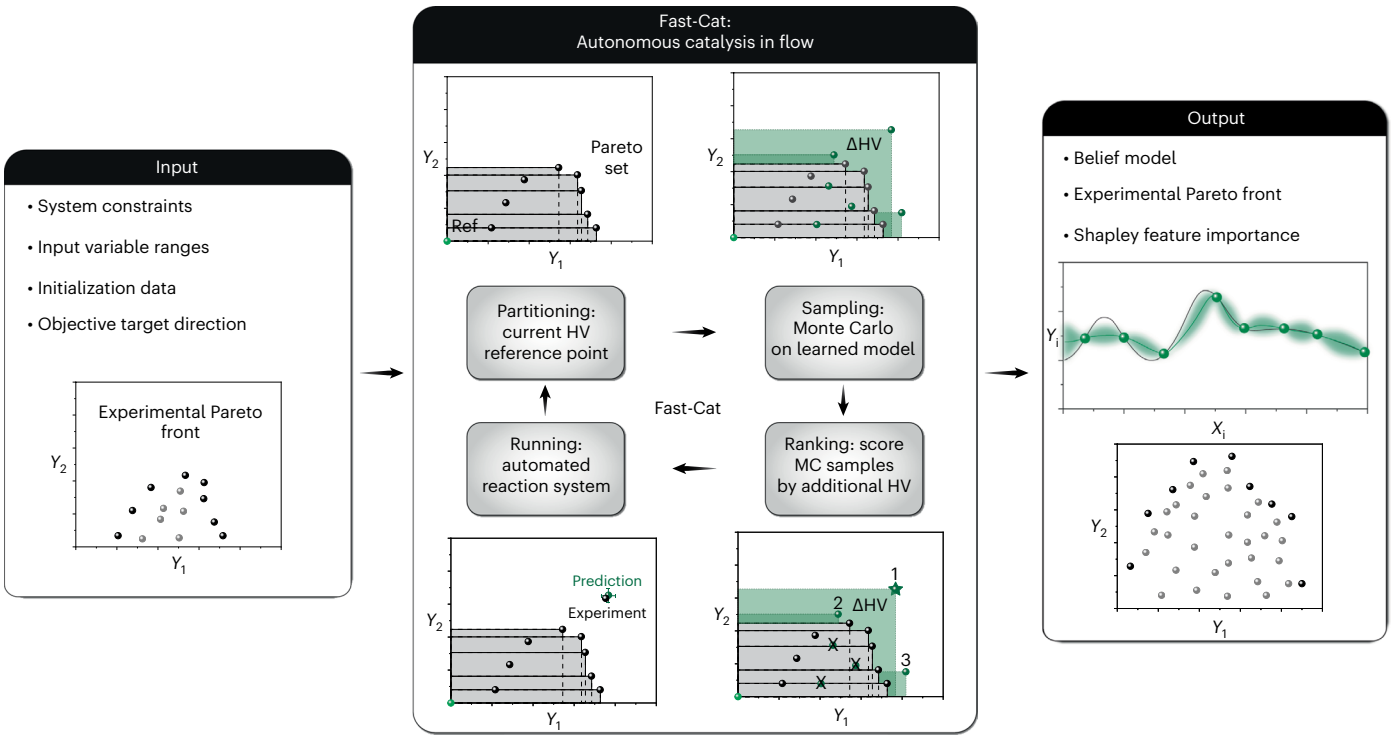


Fig. 2 | Overview of the Fast-Cat's automated experimental-selection workflow using the multi-objective Bayesian optimization technique. Input: prior data, constraints and objectives to optimize. Cycle: determine current hypervolume (HV), Monte Carlo (MC) sample the ML model at new points, rank points on the basis of predicted gained hypervolume (green star at point 1 indicating the prediction with the largest HV improvement followed by points 2 and 3 in order), perform experiment on best prediction and repeat. Output: latest surrogate model (surface plots and feature analysis) and experimental datapoints (Pareto front).

The stability of a single reaction condition over seven consecutive GC injections was investigated (Extended Data Fig. 2c) with a standard deviation of 4% for total aldehyde yield and 2% for the l/b ratio. Next, a reproducibility test was performed automatically to quantify the experimental noise of Fast-Cat by selecting an experimental condition followed by running five randomized experimental conditions and then returning to the initial reaction condition (Extended Data Fig. 2d), while obtaining the same total aldehyde yield and regioselectivity as the initial reaction condition. With Fast-Cat's hardware validated for stability and repeatability, it can be utilized to perform accelerated optimization and complex reaction space exploration with long-term continuous autonomous operation. Full reaction conditions are given in 'Benchmarking and validation conditions' section in Supplementary Information.

Autonomous Pareto-front mapping

Fast-Cat's surrogate model is initialized by four initial experimental conditions. The initial conditions include presumed high and low S_N conditions, according to the hydroformylation literature[57], an intermediate input value and a low residence time condition. Fast-Cat can ab initio rapidly identify the set of Pareto-optimal reaction conditions for both high (predominantly linear aldehyde) and low (predominantly branch aldehyde) l/b aldehyde regioselectivity. New experimental conditions are selected by qNEHVI, a multi-objective optimization algorithm intended to rapidly push the boundaries of the Pareto-front towards high yield and targeted regioselectivity (either high linear or branched aldehyde).

The autonomous Pareto-front mapping campaigns started with L1. Fast-Cat was able to autonomously identify a flexibility range of S_N = 0.37–0.85 (0.8–5.5 l/b) after 60 experimental Pareto-mapping experiments (30 for maximizing S_N: Normal Aldehyde Campaign, and 30 for minimizing S_N: Iso Aldehyde Campaign) within 1 week of continuous experimentations using less than 500 ml reaction solvent and 2 mmol total ligand (Fig. 3). As can be seen by the composite (combined linear and branched optimizations) Pareto front in Fig. 3a, the boundary of the Pareto surface pushes outwards to enclose as much of the reaction outcome space as the given ligand allows, with regions outside the boundary being inaccessible, given the constraints of Fast-Cat's hardware, as the samples converge on the true Pareto front. Once the Pareto front has been created, the data and the model can be investigated to give insights into the reaction system. For L1, the conditions with the highest S_N (that is, high l/b ratio) tend to cluster around a low reaction temperature, low CO:H_2 ratio and high ligand:Rh ratio, while low S_N (that is, low l/b ratio) tends towards the opposite (high reaction temperature, high CO:H_2 ratio and low ligand:Rh ratio). Some process parameters such as total reaction pressure and dilution are less correlated with S_N but highly correlated with the total aldehyde yield.

The proposed mechanism for tunable aldehyde regioselectivity is twofold. The first being the relative abundance of CO versus ligand L1 to impact the Rh complex from a mono-ligated $HRhL(CO)_3$ to a bis-ligated $HRhL_2(CO)_2$, with the increased steric bulk helping to drive linear regioselectivity[63]. The second being the increased reaction temperature also increases the rate of olefin isomerization and that the hydroformylation of the internal olefins produced will only result in branched aldehydes, thus lowering the S_N (refs. 5,8,21,57,63).

It should be noted that Fast-Cat allowed for determination of a broad range of S_N in only 60 autonomously selected experimental conditions while prior high-throughput screenings requiring over 400 experiments and more exotic reaction conditions to identify the same tunable regioselectivity trends for Rh-catalyzed hydroformylation of 1-octene[57]. This result demonstrates that SDL-guided catalysis exploration can dramatically reduce the experimental budget as compared with manual brute-force search or high-throughput screening methods.

https://doi.org/10.1038/s44286-024-00033-5

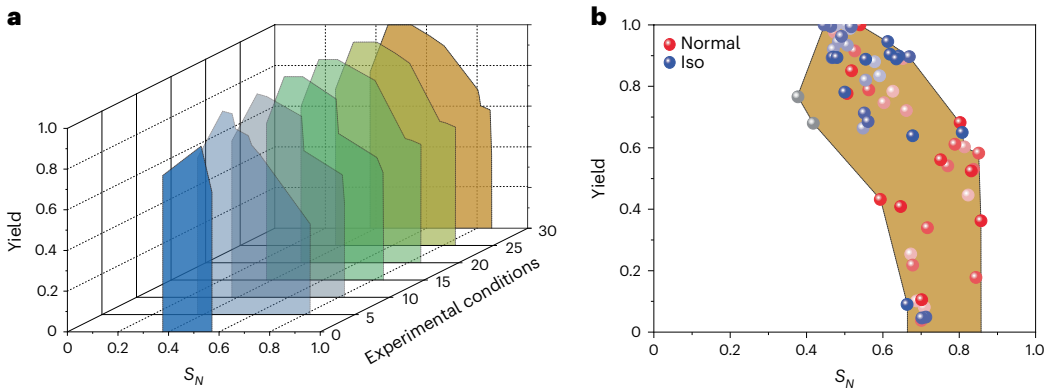


Fig. 3 | Autonomous normal and iso aldehyde Pareto-front mapping of L1 by Fast-Cat. a, Time-series Pareto front from initial data to 30 experimental conditions for both normal and iso aldehyde targets. The colours are only for Pareto-front surface visualization purposes. **b**, Final composite L1 Pareto front obtained autonomously by Fast-Cat. Darker points indicate reactions at a later point in time. Red and blue datapoints are from the S_N maximization (normal aldehyde) and minimization (iso aldehyde) campaigns, respectively.

Following autonomous mapping of L1 by Fast-Cat, the system was washed, and the next ligand solution was loaded for a new set of autonomous normal and iso aldehyde Pareto-front mapping campaign. Subsequent ligands after L1 (L2–L6) were given an autonomous experimentation budget of 20 experimental conditions per campaign (total 40 experiments) to form their Pareto fronts (Fig. 4a). The bulky fluorophosphite ligand L1 was found to be the ligand with the widest range of regioselectivities, with the closely related L2 and L3 ligands encompassing a slightly more normal-selective region of hydroformylation reaction space. Ligand L4 was a departure from the fluorophosphites in that while the activity and yields were high, the range of regioselectivities available was relatively narrow (-S_N 0.6–0.8, l/b 1.5-4). Similarly for ligand L5, Xantphos, as a bidentate phosphine ligand, the normal regioselectivity was found to be excellent (S_N > 0.98, l/b > 49) while the activity was low. Ligand L6, as a sterically hindered phosphite, showed high yields across most of the autonomously selected reaction conditions, while the regioselectivity remained roughly equal normal and iso product (-S_N 0.4–0.6, l/b 0.66–1.5).

The phosphite ligands tested were the most branched selective of the ligands tested, with L1 being the most branched-selective fluorophosphite followed by L2 and then L3. L1 is the most sterically hindered fluorophosphite with 2-phenylpropan-2-yl substituents and *t*-butyl on the bridging carbon in the phosphite heterocycle. L2 maintained the *t*-butyl bridge but reduced the steric bulk of the outer substituents to *t*-butyl as well, whereas L3 maintained the 2-phenylpropan-2-yl but reduced the bridge to methyl. This trend indicates that bulk at the bridging position may be more important towards branched selective conditions over bulk that is farther from the Rh center. While triphenylphosphine was relatively active towards hydroformylation, the simple structure and lack of steric bulk does not appear to allow much regioselectivity flexibility beyond changing 1-octene isomerization. Xantphos, on the other hand, as a bis-phosphine is driven to only the bis-ligated form of the Rh complex[19,63–65] by virtue of intra- versus inter-molecular complexing of the second phosphorus verusu monophosphine species. This weighting of the bis-ligated Rh center drives Xantphos to the highest S_N of the tested ligands by Fast-Cat, and the low regioselectivity flexibility indicates the difficulty of displacing one of the phosphines once Xantphos is complexed at all. The full table of autonomously tested reaction conditions by Fast-Cat for ligands L1–L6 and reaction outcomes are listed in sections 'Model Parameters and Normalization' and 'qNEHVI Ligand Screening Data' in Supplementary Information.

Next, to demonstrate the knowledge scalability of Fast-Cat, the ligands and hydroformylation reaction conditions with high and low regioselectivity values and yields above 50% were selected to be tested in a batch reactor (10-ml autoclave). The knowledge scalability results shown in Fig. 4b illustrate that the knowledge generated can be directly transferred from a flow to a batch reactor and retain consistent reaction results. This result closes the gap of knowledge transfer and scalability between lab-scale SDL-driven discovery and scaled-up production relevant to the industry. Details on the reaction conditions of the knowledge scalability experiments can be found in 'Knowledge scalability' section in Supplementary Information.

The wealth of reaction data captured across the various ligand Pareto-front mapping experiments performed by Fast-Cat allows for digitized reaction models to be created for each tested ligand. In the next section, we discuss the digital hydroformylation reaction models built from in-house generated experimental data to determine the magnitude and direction of the response of the system (aldehyde yield and regioselectivity) to the various model input features. Additional visualization of the L1 Pareto front for important input parameters is in 'Visualization' section in Supplementary Information.

Digital twin

The design of the hydroformylation reaction digital twin involves fine tuning the hyper-parameters of the DNN to optimize its performance. The prediction accuracy of the digital twin is quantified by comparing the trained ML model prediction on the input reaction conditions which have not been previously used for ML model training and validation (that is, test set), versus the actual values obtained from Fast-Cat. Using the test set that was not utilized for ML model training ensures that the prediction accuracy is not influenced by overfitting. To remove bias from the ML model and create uncertainty for the predicted data, we adapted an ensemble ML modeling strategy for the digital twin[66]. Full model tuning information is supplied in 'Digital twin architecture' section in Supplementary Information. Next, the fine-tuned digital twins built from the in-house generated experimental data were employed to autonomously study the impact of each reaction parameter on the aldehyde yield and regioselectivity as well as visualizing the reaction space.

Shapley values provide valuable insights into the impact of different input variables on the outcomes of a predictive model. Shapley analysis is a method used to understand the contribution or importance of model features (that is, hydroformylation reaction variables) on the model's outputs (that is, aldehyde yield and regioselectivity). Shapley additive explanations[67] values provide a metric for discerning the features that Fast-Cat considers to have the most substantial influence on the outputs of the ML model (digital twin). Utilizing the

https://doi.org/10.1038/s44286-024-00033-5

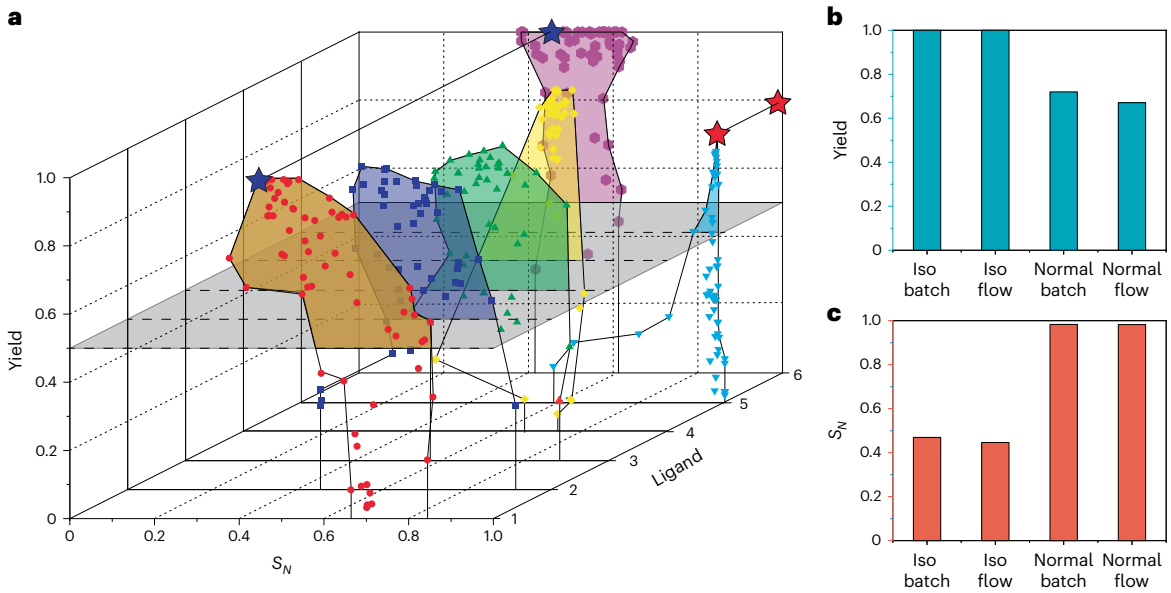


Fig. 4 | Autonomous Pareto-front mapping of different tested ligands by Fast-Cat. a, Final composite Pareto fronts for all six ligands autonomously mapped by Fast-Cat. Red and blue stars indicate condition with highest normal and iso aldehyde regioselectivity, respectively, above 50% yield across all ligands. The colours are only for Pareto-front surface visualization purposes of different ligands. **b**,**c**, Comparison between autonomously identified normal and iso aldehyde reaction conditions by Fast-Cat between a 10 ml autoclave batch reactor and Fast-Cat's flow reactor. The iso-selective condition was identified with L1, and the normal-selective condition was identified with L5.

trained digital twin model for each ligand in this study, we performed Shapley analysis to investigate the importance of process parameters on the yield and regioselectivity of the hydroformylation reaction. Figures 5a,b shows Shapley analysis of L1's digital twin. Shapley additive explanations values of L1 suggest that the total reaction pressure had the largest effect on the hydroformylation yield, with an increase in pressure leading to an increase in the reaction yield. Increasing the reaction pressure enhances the concentration of syngas in the liquid phase and the reaction time, thereby increasing the overall reaction yield. However, reaction temperature demonstrated the greatest influence on the aldehyde regioselectivity tuning (inverse relationship) in the presence of L1. The inverse relationship between the reaction temperature and normal aldehyde regioselectivity in the presence of L1 can be attributed to the dominant effect of olefin isomerization, the sensitivity of catalyst pre-activation to temperature and the preferential formation of branched aldehydes at higher temperatures due to the destabilization and conversion of specific ligand-catalyst species[57]. Interestingly, as the olefin-to-Rh ratio increases, it slows down the full conversion of olefin at a constant catalyst TOF, but it favors the formation of linear product. Increased formation of the linear versus branched aldehyde can be attributed to the enhanced hydroformylation over isomerization of the terminal olefin as the syngas-to-Rh ratio increases, while any isomerization leads to primarily branch product. Moreover, a higher ligand-to-Rh ratio increases both aldehyde regioselectivity and yield. According to a mono versus bis-ligation switching mechanism of L1 (refs. 23,63), higher L1 concentration will promote bis-ligated L-Rh-L species formation with higher steric hindrance and, thus, favor the formation of linear aldehyde. Additionally, L1 is an activating ligand, that is, the TOF of the catalyst increases at higher ligand-to-Rh ratios, therefore, increasing its concentration will lead to higher aldehyde yield.

To better understand the effect of key reaction parameters controlling the total yield and regioselectivity of the hydroformylation reaction in the presence of L1, we utilized the digital twin to visualize the confounding effects as well as Pareto-front heatmaps ('Visualization' section in Supplementary Information) of the top three features of interest (that is, process parameters) identified by Shapley analysis. For the hydroformylation reaction visualization plots based on the digital twin models (Extended Data Fig. 3), three levels of all other process parameters (normalized values 0.25, 0.5 and 0.75) were explored to effectively map the reaction space. The surface plots shown in Extended Data Fig. 3 further confirm the results of Shapley analyses, where temperature was found to inversely correlate with the aldehyde regioselectivity; however, higher aldehyde yields were obtained at higher temperatures (Extended Data Fig. 3a(i),b(i)). Extended Data Fig. 3a(ii,iii) illustrates that an increased olefin-to-Rh ratio corresponds to higher aldehyde yields, as the presence of more reactants facilitates the progression of the hydroformylation reaction. However, despite the results of Shapley analysis (Fig. 5a,b), surface plots of Extended Data Fig. 3a(ii,iii) exhibit low impact for reaction temperature and pressure. This result can be attributed to the fact that altering temperature and pressure can substantially affect the reaction yield only when the remaining parameters are adjusted to values near the optimal conditions. Furthermore, Extended Data Fig. 3a(i) illustrates that increasing the total syngas pressure, which corresponds to increasing the residence time in the flow reactor module of the Fast-Cat platform, enhances the aldehyde yield. Extended Data Fig. 3b(ii,iii) presents the complexity of the effect of hydroformylation reaction parameters on the aldehyde regioselectivity. The overlap of different hydroformylation reaction surface responses indicates that multiple input reaction conditions can lead to the same aldehyde regioselectivity. Although Extended Data Fig. 3b(ii), in contrast to Shapley analysis, suggests that the temperature does not have a large effect on aldehyde regioselectivity tuning in the presence of L1, it has to be noted the presented surface plots correspond to a constant value for process parameters other than the ones shown on each axis. Thus, Extended Data Fig. 3b(ii) implies that if features X_2–X_5 are maintained constant, varying temperature cannot dramatically influence the aldehyde regioselectivity. This result suggests that while temperature serves as a pivotal parameter influencing the hydroformylation reaction, it is imperative to simultaneously tune other process parameters to achieve satisfactory aldehyde regioselectivity. Extended Data Fig. 3b(iii) shows

https://doi.org/10.1038/s44286-024-00033-5

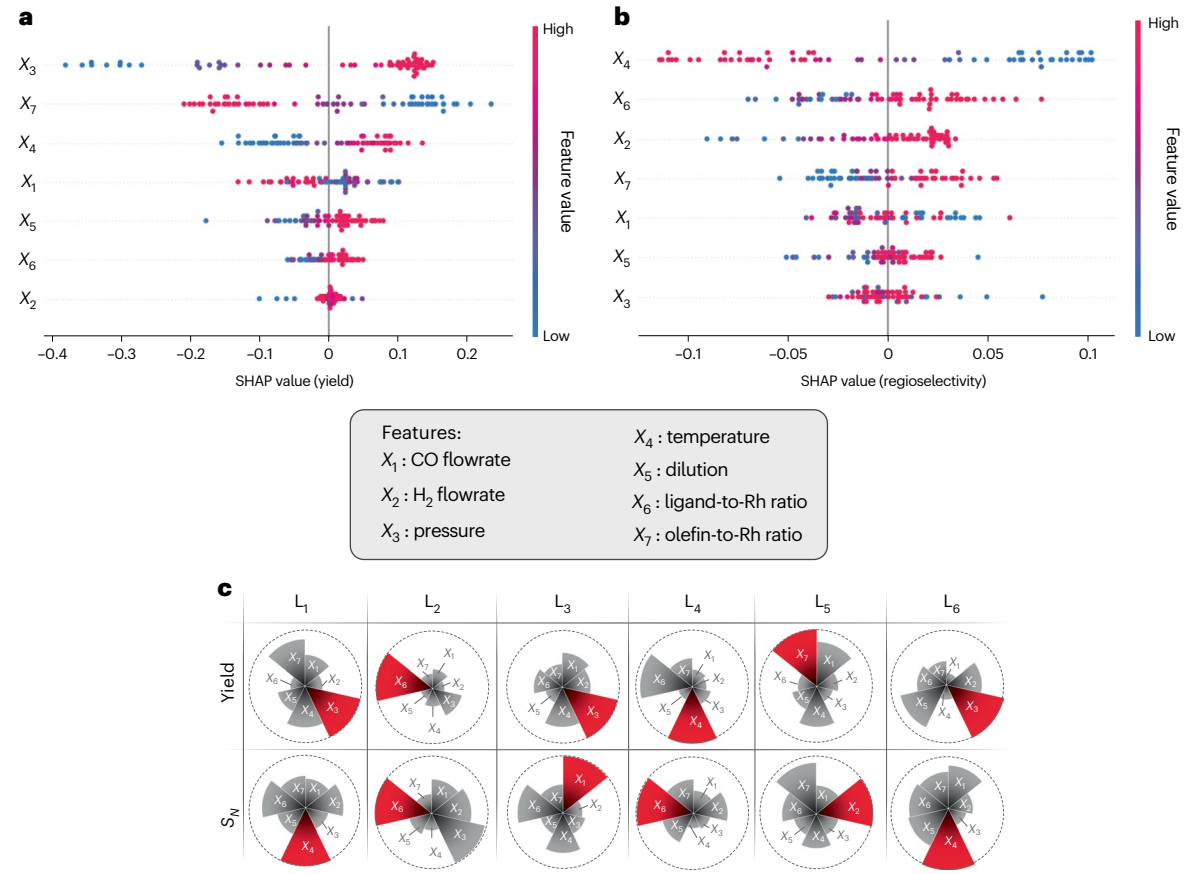


Fig. 5 | ML-assisted assessment of the ligand effect in hydroformylation of 1-octene. a,**b**, Shapley analysis of yield (**a**) and regioselectivity (**b**) using digital twin model of L1 and Fast-Cat process parameters. The Shapley value (SHAP) of a given point (reaction condition) indicates the strength of the influence that a given model input feature has on the output at that specific point. Shapley analysis quantifies the individual contributions of each ML model feature towards the overall yield (**a**) and regioselectivity (**b**) of the product aldehyde in the presence of ligand L1. **c**, Summary of Shapley analysis for the hydroformylation reaction, illustrating variations in the importance of each model feature on aldehyde yield and regioselectivity across ligands L1–L6. These variations can be attributed to inherent structural differences, which potentially influence the binding with Rh and impact the underlying reaction mechanism.

that increasing the L1-to-Rh ratio can tune aldehyde regioselectivity from 0.2 to 0.6.

Similar Shapley analysis on the aldehyde yield and regioselectivities of ligands L2–L6 (Fig. 5c) shows that ligand structure can have a dramatic impact on which process parameters are the most important for being able to control the reaction outcome. The observed variations in the importance of each model feature on the aldehyde yield and regioselectivity across different ligands can be attributed to inherent structural differences, which may give rise to nuanced differences in the Rh coordination environment, affecting the underlying catalytic pathways. For example, the primary contributors to S_N for the fluorophosphites are reaction temperature, ligand:Rh ratio and CO flowrate for L1, L2 and L3, respectively. The relative importance of reaction temperature for L1 can indicate that it has a higher rate of hydroformylation of internal olefins with a temperature increase, both promoting that rate as well as increasing the rate of isomerization, leading to a more pronounced reduction in S_N. L2 and L3, on the other hand, with lower steric bulk, may be more favorable to form bis- complexes with the Rh, therefore resulting in the relative ratios of ligand:Rh and CO flowrate to have greater importance on the overall direction of the aldehyde regioselectivity. Triphenylphosphine has a similar trend with the ligand:Rh ratio being highly impactful as a low steric bulk ligand with lower activity for hydroformylation of internal olefins. Xantphos, on the other hand, is more effected by H_2 flowrate, as the bis-phosphine binding results in bis- complexes with lower influence of the ligand:Rh ratio[63], whereas increased H_2 partial pressure generally promotes the rate of the hydroformylation reaction. Digital twin reaction visualizations and Shapley analysis plots for ligands L2–L6 can be found in 'Digital twin feature analysis' section in Supplementary Information.

Discussion

In summary, this work unveiled Fast-Cat, which is able to rapidly identify the scalable performance Pareto front of a given ligand for transition metal-catalyzed homogeneous reactions using only ~2 mmol of the ligand and a budget of 1 week for continuous autonomous experimentation. Fast-Cat's autonomous capabilities for high-temperature/pressure, metal- and ligand-mediated gas–liquid reactions and data-driven approach improve catalyst discovery and development by enhancing efficiency and speed. Fundamental insights about the role of each ligand on the catalytic reaction can be extracted from the autonomously sampled experimental conditions via digital twin modeling and input Shapley feature analysis. Given the prohibitively large chemical structure space of possible catalysts, it is imperative to have a method of rapidly and comprehensively evaluating the potential capabilities of a given ligand candidate for a targeted catalytic reaction. The Fast-Cat algorithm and workflow proves robust and able to rapidly generate the large amounts of ligand-reaction data required to begin modeling and performing optimization in ligand structure space in search of novel ligand candidates. In addition, the scalability of Fast-Cat-discovered knowledge closes the gap between laboratory-scale discovery and

process development to an industrially relevant reactor. The knowledge scalability of Fast-Cat will facilitate adoption of SDLs for accelerated reaction exploration and process development by fine/specialty chemical industries. As Fast-Cat is modular, future integration with a robotic chemical workstation and other automation tools will serve to expand the library of materials available for reaction space exploration allowing for discrete variable screening and optimization campaigns

Methods

Start-up

The flow reactor is supplied with four separate stock solutions to adjust concentrations of the individual reagent species independently and match the next predicted hydroformylation condition selected by the ML module of Fast-Cat. The stock solutions are prepared such that an equal volumetric flowrate results in 0.5 M 1-octene, 0.25 mM $Rh(CO)_2(acac)$, 2.5 mM ligand and 20 mM tridecane internal standard in toluene as the solvent (list of chemicals can be found in 'Chemicals' section in Supplementary Information). After the reagents are prepared, they are loaded in the automated precursor refill reservoirs ('Automated precursor refilling module' section in Supplementary Information) and syringe pumps using stainless-steel syringes. Next, the flow reactor line is brought up to the target pressure with a 1:1 mixture of carbon monoxide (CO) and hydrogen (H_2). The process control module of Fast-Cat then takes over the operational workflow and the closed-loop catalysis cycles begin.

Operation

The hardware of Fast-Cat automatically adjusts the liquid feed composition by varying the volumetric flowrate of the individual liquid input streams, subject to ratios that maintain stable flow through the mixing junctions. The syngas composition is automatically tuned by varying individual CO and H_2 mass flow controllers to set the flowrate and control for any supply pressure fluctuations. The total volumetric flowrate of the liquid and the gas phases defines the residence time (that is, hydroformylation reaction time) of the reactive liquid slugs within the heated flow reactor section before in-line sampling by GC. The reaction is allowed to equilibrate for four residence times to achieve steady-state operation at the set reaction condition before sampling. The unsampled reactor effluent can be either sent to a waste collection or a sample collection vessel for offline characterization. The automated sampling system switches a sample-loop valve to send the captured sample to the analysis line and trigger the GC method and data acquisition. The sample-loop and GC injection needle are washed between experimental conditions by sending droplets of acetone through the injector line and conducting a wash and GC column bake method. The GC output was recalibrated for the in-line injector system versus off-line sampling ('GC auto-injection calibration' and 'In-flow mixing' sections in Supplementary Information).

Before performing a new reaction condition, Fast-Cat evaluates the current available precursor volume of the syringe pumps and refills them from the precursor refilling module if the new reaction conditions would completely consume the precursors available in the stainless-steel syringes before completing the next hydroformylation reaction.

Pareto screening cycle

Four pre-defined initial experimental conditions are performed to gather the data needed to initialize the surrogate model of the ML brain of Fast-Cat. Once Fast-Cat has enough data to begin training the NN model, the ML algorithm will begin feeding new reaction conditions to be tested to the reaction control module. Upon receiving the results of an experimental condition from automated processing of the GC data file, Fast-Cat updates its ML model and experimental database and predicts a new experiment to be automatically conducted by its hardware. The Pareto-front mapping algorithm initially supplies two conditions, one to perform immediately and one to keep in the queue to begin performing the second reaction immediately after sending the sample of the first reaction into the in-line characterization module (GC). This operation model allows Fast-Cat to begin work on equilibrating the second reaction condition while the product of the first reaction is being analyzed through the GC unit and processed by the ML algorithm. The new predicted experimental condition from the ML brain of Fast-Cat is then placed into the queue to await the next cycle.

Additional details on Fast-Cat's physical operation, benchmarking, validation, computational operation and collected data are presented in Supplementary Information.

Data availability

The authors declare that all data supporting the findings of this study are available within the main text and Supplementary Information.

Code availability

The source code for the Pareto-front mapping and digital twin models have been deposited in the repository 'Fast-Cat' (https://github.com/AbolhasaniLab).

References

1. Franke, R., Selent, D. & Börner, A. Applied hydroformylation. *Chem. Rev.* **112**, 5675–5732 (2012).
2. Whiteker, G. T. & Cobley, C. J. in *Organometallics as Catalysts in the Fine Chemical Industry* 35–46 (Springer, 2012).
3. Botteghi, C., Paganelli, S., Schionato, A. & Marchetti, M. The asymmetric hydroformylation in the synthesis of pharmaceuticals. *Chirality* **3**, 355–369 (1991).
4. Stanley, G. G. in *Kirk-Othmer Encyclopedia of Chemical Technology* 1–19 (Wiley, 2017).
5. Vilches-Herrera, M., Domke, L. & Börner, A. Isomerization–hydroformylation tandem reactions. *ACS Catal.* **4**, 1706–1724 (2014).
6. Klein, H., Jackstell, R., Wiese, K. D., Borgmann, C. & Beller, M. Highly selective catalyst systems for the hydroformylation of internal olefins to linear aldehydes. *Angew. Chem. Int. Ed.* **40**, 3408–3411 (2001).
7. Seayad, A. et al. Internal olefins to linear amines. *Science* **297**, 1676 (2002).
8. Yan, Y., Zhang, X. & Zhang, X. A tetraphosphorus ligand for highly regioselective isomerization–hydroformylation of internal olefins. *J. Am. Chem. Soc.* **128**, 16058–16061 (2006).
9. Breit, B. & Seiche, W. Hydrogen bonding as a construction element for bidentate donor ligands in homogeneous catalysis: regioselective hydroformylation of terminal alkenes. *J. Am. Chem. Soc.* **125**, 6608–6609 (2003).
10. van der Veen, L. A., Kamer, P. C. & van Leeuwen, P. W. Hydroformylation of internal olefins to linear aldehydes with novel rhodium catalysts. *Angew. Chem. Int. Ed.* **38**, 336–338 (1999).
11. Jiao, Y., Torne, M. S., Gracia, J., Niemantsverdriet, J. H. & van Leeuwen, P. W. Ligand effects in rhodium-catalyzed hydroformylation with bisphosphines: steric or electronic? *Catal. Sci. Technol.* **7**, 1404–1414 (2017).
12. Van Leeuwen, P. W. & Claver, C. *Rhodium Catalyzed Hydroformylation* Vol. 22 (Springer, 2002).
13. Zhu, C. et al. Flow chemistry-enabled studies of rhodium-catalyzed hydroformylation reactions. *ChemComm* **54**, 8567–8570 (2018).
14. Raghuvanshi, K. et al. Highly efficient 1-octene hydroformylation at low syngas pressure: from single-droplet screening to continuous flow synthesis. *ACS Catal.* **10**, 7535–7542 (2020).
15. Van Rooy, A., de Bruijn, J. N., Roobeek, K. F., Kamer, P. C. & Van Leeuwen, P. W. Rhodium-catalysed hydroformylation of branched 1-alkenes; bulky phosphite vs. triphenylphosphine as modifying ligand. *J. Organomet. Chem.* **507**, 69–73 (1996).

16. van Rooy, A., Orij, E. N., Kamer, P. C. & van Leeuwen, P. W. Hydroformylation with a rhodium/bulky phosphite modified catalyst. A comparison of the catalyst behavior for oct-1-ene, cyclohexene, and styrene. *Organometallics* **14**, 34–43 (1995).
17. Diebolt, O., Tricas, H., Freixa, Z. & van Leeuwen, P. W. Strong π-acceptor ligands in rhodium-catalyzed hydroformylation of ethene and 1-octene: operando catalysis. *ACS Catal.* **3**, 128–137 (2013).
18. Vogl, C., Paetzold, E., Fischer, C. & Kragl, U. Highly selective hydroformylation of internal and terminal olefins to terminal aldehydes using a rhodium-BIPHEPHOS-catalyst system. *J. Mol. Catal. A* **232**, 41–44 (2005).
19. Kamer, P. C. J., van Leeuwen, P. W. N. M. & Reek, J. N. H. Wide bite angle diphosphines: Xantphos ligands in transition metal complexes and catalysis. *Acc. Chem. Res.* **34**, 895–904 (2001).
20. Ichihara, T., Nakano, K., Katayama, M. & Nozaki, K. Tandem hydroformylation–hydrogenation of 1-decene catalyzed by Rh-bidentate bis(trialkylphosphine)s. *Asian J. Chem.* **3**, 1722–1728 (2008).
21. Fuchs, E., Keller, M. & Breit, B. Phosphabarrelenes as ligands in rhodium-catalyzed hydroformylation of internal alkenes essentially free of alkene isomerization. *Chem. Eur. J.* **12**, 6930–6939 (2006).
22. Kuil, M., Soltner, T., van Leeuwen, P. W. & Reek, J. N. High-precision catalysts: regioselective hydroformylation of internal alkenes by encapsulated rhodium complexes. *J. Am. Chem. Soc.* **128**, 11344–11345 (2006).
23. Besset, T., Norman, D. W. & Reek, J. N. Supramolecular encapsulated rhodium catalysts for branched selective hydroformylation of alkenes at high temperature. *Adv. Synth. Catal.* **355**, 348–352 (2013).
24. Gao, H. et al. Using machine learning to predict suitable conditions for organic reactions. *ACS Cent. Sci.* **4**, 1465–1476 (2018).
25. Zhou, Z., Li, X. & Zare, R. N. Optimizing chemical reactions with deep reinforcement learning. *ACS Cent. Sci.* **3**, 1337–1344 (2017).
26. Epps, R. W. et al. Artificial chemist: an autonomous quantum dot synthesis bot. *Adv. Mater.* **32**, 2001626 (2020).
27. Schweidtmann, A. M. et al. Machine learning meets continuous flow chemistry: automated optimization towards the Pareto front of multiple objectives. *J. Chem. Eng.* **352**, 277–282 (2018).
28. Abolhasani, M. & Kumacheva, E. The rise of self-driving labs in chemical and materials sciences. *Nat. Synth.* **2**, 483–492 (2023).
29. Bennett, J. A. & Abolhasani, M. Autonomous chemical science and engineering enabled by self-driving laboratories. *Curr. Opin. Chem. Eng.* **36**, 100831 (2022).
30. Epps, R. W., Volk, A. A., Ibrahim, M. Y. & Abolhasani, M. Universal self-driving laboratory for accelerated discovery of materials and molecules. *Chem* **7**, 2541–2545 (2021).
31. Volk, A. A. & Abolhasani, M. Autonomous flow reactors for discovery and invention. *Trends Chem.* **3**, 519–522 (2021).
32. Granda, J. M., Donina, L., Dragone, V., Long, D.-L. & Cronin, L. Controlling an organic synthesis robot with machine learning to search for new reactivity. *Nature* **559**, 377–381 (2018).
33. Sans, V., Porwol, L., Dragone, V. & Cronin, L. A self optimizing synthetic organic reactor system using real-time in-line NMR spectroscopy. *Chem. Sci.* **6**, 1258–1264 (2015).
34. Steiner, S. et al. Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* **363**, eaav2211 (2019).
35. Bradford, E., Schweidtmann, A. M. & Lapkin, A. Efficient multiobjective optimization employing Gaussian processes, spectral sampling and a genetic algorithm. *J. Glob. Optim.* **71**, 407–438 (2018).
36. Clayton, A. D. et al. Automated self-optimisation of multi-step reaction and separation processes using machine learning. *J. Chem. Eng.* **384**, 123340 (2020).
37. Christensen, M. et al. Data-science driven autonomous process optimization. *Commun. Chem.* **4**, 112 (2021).
38. MacLeod, B. P. et al. Self-driving laboratory for accelerated discovery of thin-film materials. *Sci. Adv.* **6**, eaaz8867 (2020).
39. Roch, L. M. et al. ChemOS: orchestrating autonomous experimentation. *Sci. Robot.* **3**, eaat5559 (2018).
40. Gómez-Bombarelli, R. et al. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nat. Mater.* **15**, 1120–1127 (2016).
41. Coley, C. W. et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **365**, eaax1566 (2019).
42. Taylor, C. J. et al. Accelerated chemical reaction optimization using multi-task learning. *ACS Cent. Sci.* **9**, 957–968 (2023).
43. Shields, B. J. et al. Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **590**, 89–96 (2021).
44. Ahneman, D. T., Estrada, J. G., Lin, S., Dreher, S. D. & Doyle, A. G. Predicting reaction performance in C–N cross-coupling using machine learning. *Science* **360**, 186–190 (2018).
45. Tabor, D. P. et al. Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat. Rev. Mater.* **3**, 5–20 (2018).
46. Volk, A. A. et al. AlphaFlow: autonomous discovery and optimization of multi-step chemistry using a self-driven fluidic lab guided by reinforcement learning. *Nat. Commun.* **14**, 1403 (2023).
47. Deneault, J. R. et al. Toward autonomous additive manufacturing: Bayesian optimization on a 3D printer. *MRS Bull.* **46**, 566–575 (2021).
48. Gutmann, B., Cantillo, D. & Kappe, C. O. Continuous-flow technology—a tool for the safe manufacturing of active pharmaceutical ingredients. *Angew. Chem. Int. Ed.* **54**, 6688–6728 (2015).
49. Hartman, R. L., McMullen, J. P. & Jensen, K. F. Deciding whether to go with the flow: evaluating the merits of flow reactors for synthesis. *Angew. Chem. Int. Ed.* **50**, 7502–7519 (2011).
50. Movsisyan, M. et al. Taming hazardous chemistry by continuous flow technology. *Chem. Soc. Rev.* **45**, 4892–4928 (2016).
51. Volk, A. A., Campbell, Z. S., Ibrahim, M. Y., Bennett, J. A. & Abolhasani, M. Flow chemistry: a sustainable voyage through the chemical universe en route to smart manufacturing. *Annu. Rev. Chem. Biomol. Eng.* **13**, 45–72 (2022).
52. Nambiar, A. M. K. et al. Bayesian optimization of computer-proposed multistep synthetic routes on an automated robotic flow platform. *ACS Cent. Sci.* **8**, 825–836 (2022).
53. Nieves-Remacha, M. J. & Jensen, K. F. Mass transfer characteristics of ozonolysis in microreactors and advanced-flow reactors. *J. Flow Chem.* **5**, 160–165 (2015).
54. Brzozowski, M., O'Brien, M., Ley, S. V. & Polyzos, A. Flow chemistry: intelligent processing of gas–liquid transformations using a tube-in-tube reactor. *Acc. Chem. Res.* **48**, 349–362 (2015).
55. Hessel, V., Kralisch, D., Kockmann, N., Noël, T. & Wang, Q. Novel process windows for enabling, accelerating, and uplifting flow chemistry. *ChemSusChem* **6**, 746–789 (2013).
56. Plutschack, M. B., Pieber, B. U., Gilmore, K. & Seeberger, P. H. The hitchhiker's guide to flow chemistry||. *Chem. Rev.* **117**, 11796–11893 (2017).
57. Ibrahim, M. Y. S., Bennett, J. A., Mason, D., Rodgers, J. & Abolhasani, M. Flexible homogeneous hydroformylation: on-demand tuning of aldehyde branching with a cyclic fluorophosphite ligand. *J. Catal.* **409**, 105–117 (2022).
58. Joshi, J. B., Pandit, A. B. & Sharma, M. M. Mechanically agitated gas–liquid reactors. *Chem. Eng. Sci.* **37**, 813–844 (1982).
59. Gentric, C., Mignon, D., Bousquet, J. & Tanguy, P. A. Comparison of mixing in two industrial gas–liquid reactors using CFD simulations. *Chem. Eng. Sci.* **60**, 2253–2272 (2005).

60. Günther, A., Jhunjhunwala, M., Thalmann, M., Schmidt, M. A. & Jensen, K. F. Micromixing of miscible liquids in segmented gas–liquid flow. *Langmuir* **21**, 1547–1555 (2005).
61. Günther, A., Khan, S. A., Thalmann, M., Trachsel, F. & Jensen, K. F. Transport and reaction in microscale segmented gas–liquid flow. *Lab Chip* **4**, 278–286 (2004).
62. Daulton, S., Balandat, M. & Bakshy, E. Differentiable expected hypervolume improvement for parallel multi-objective Bayesian optimization. *Adv. Neural Inf. Process. Syst.* **33**, 9851–9864 (2020).
63. Jacobs, I., de Bruin, B. & Reek, J. N. Comparison of the full catalytic cycle of hydroformylation mediated by mono-and bis-ligated triphenylphosphine–rhodium complexes by using DFT calculations. *ChemCatChem* **7**, 1708–1718 (2015).
64. Dieleman, C. B., Kamer, P. C., Reek, J. N. & van Leeuwen, P. W. Xantphite: a new family of ligands for catalysis. Applications in the hydroformylation of alkenes. *Helv. Chim. Acta* **84**, 3269–3280 (2001).
65. Zuidema, E. et al. The rate-determining step in the rhodium–Xantphos-catalysed hydroformylation of 1-octene. *Chem. Eur. J.* **14**, 1843–1853 (2008).
66. Shahhosseini, M., Hu, G. & Pham, H. Optimizing ensemble weights and hyperparameters of machine learning models for regression problems. *Mach. Learn. Appl.* **7**, 100251 (2022).
67. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **30**, 106786–6795 (2017).
68. Bateni, F. Smart Dope: a self-driving fluidic lab for accelerated development of doped perovskite quantum dots. *Adv. Energy Mater.* **14**, 2302303 (2024).
69. Sadeghi, S. et al. Autonomous nanomanufacturing of lead-free metal halide perovskite nanocrystals using a self-driving fluidic lab. *Nanoscale* **16**, 580–591 (2024).
70. Koscher, B. A. et al. Autonomous multi-property-driven molecular discovery: from predictions to measurements and back. *Science* **382**, 6677 (2023).

Acknowledgements

J.A.B., N.O., M.K., S.S. and M.A. gratefully acknowledge the financial support from Eastman Chemical Company.

Author contributions

M.A. and J.A.B. conceived the project. J.A.B., N.O., S.S. and M.A. designed the algorithms. J.A.B. programmed the Pareto-mapping algorithm and built the flow platform with M.K. N.O. programmed the digital twin models and performed Shapley analysis under advisement of J.R. and M.A. M.K. conducted the reactor benchmarking experiments. J.A.B. conducted the investigations and data anlayses under advisement of J.R. and M.A. M.A. acquired funding and directed the project. J.A.B., N.O. and M.A. drafted the paper. All authors provided feedback on the paper.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s44286-024-00033-5.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44286-024-00033-5.

Correspondence and requests for materials should be addressed to M. Abolhasani.

Peer review information *Nature Chemical Engineering* thanks Xiaonan Wang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

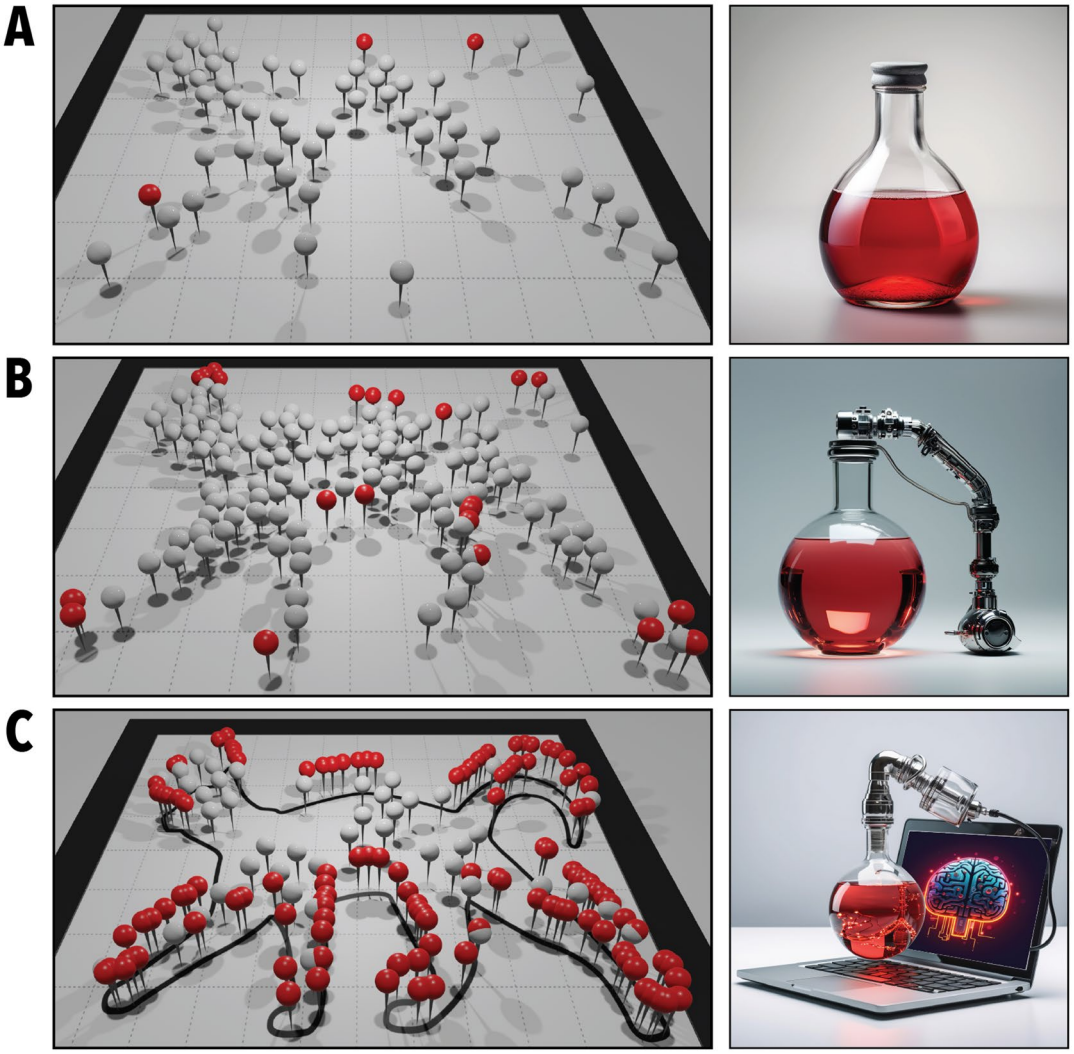
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

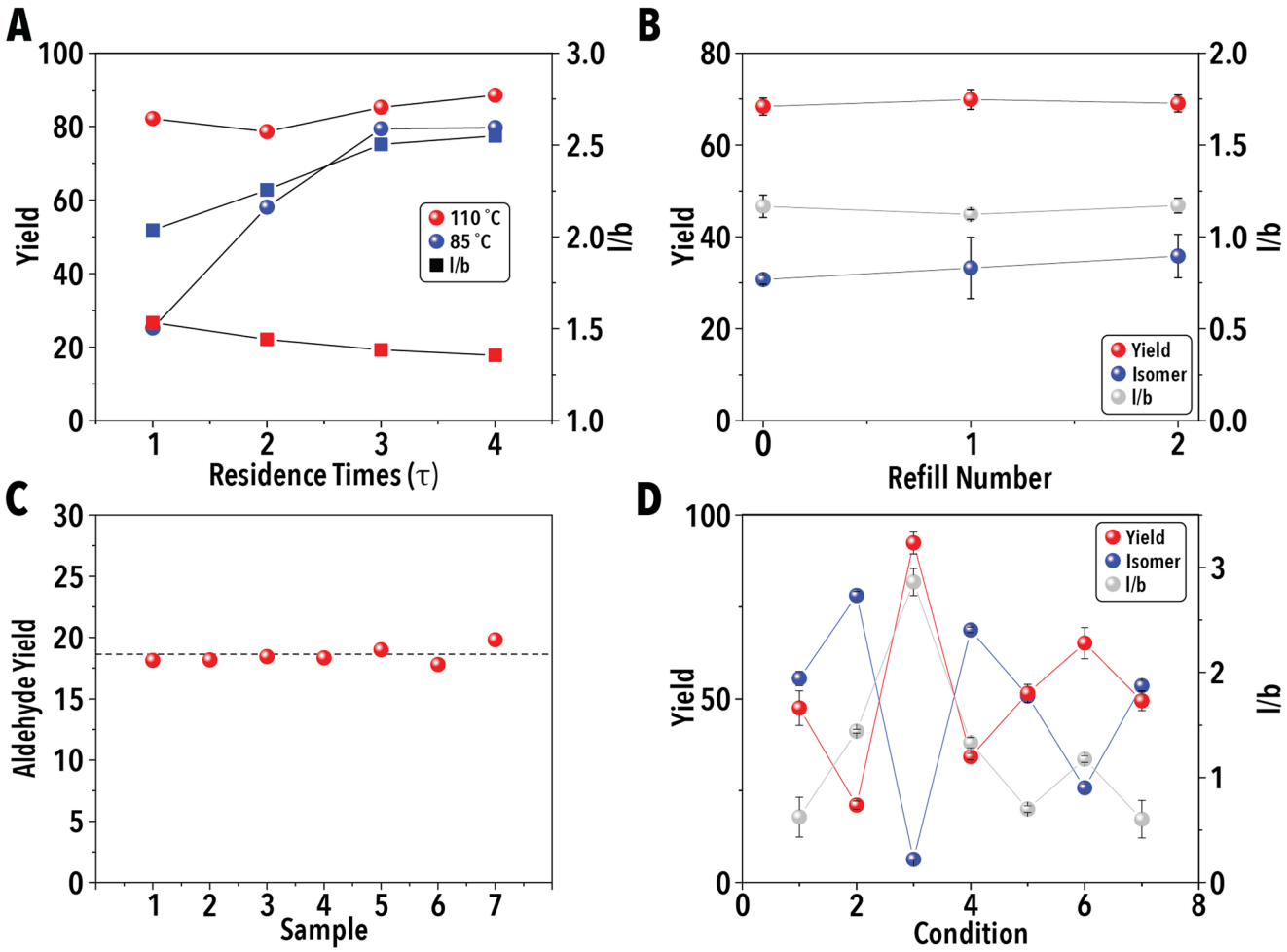
© The Author(s), under exclusive licence to Springer Nature America, Inc. 2024

https://doi.org/10.1038/s44286-024-00033-5



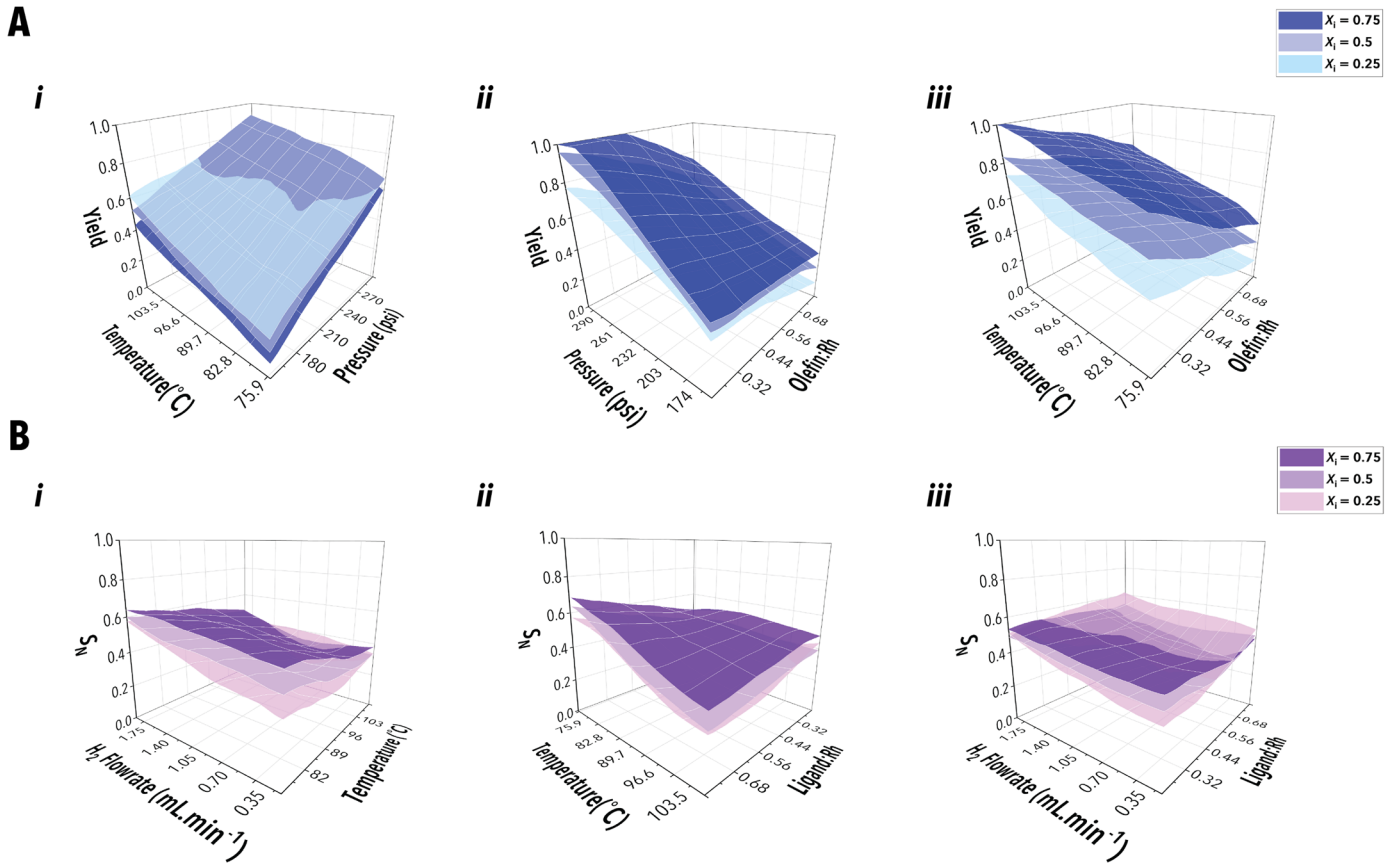
Extended Data Fig. 1 | Different experimental strategies for fundamental and applied studies of homogeneous catalysis. (**a**) Manual vs. (**b**) automated, vs. (**c**) autonomous experimental catalysis. Autonomous experimentation utilizes intelligent experiment-selection to fast-track unveiling the full performance map (red pins) of each catalyst/ligand system with minimum human intervention and experimental cost. The red flask represents a reactor for homogeneous catalysis.

https://doi.org/10.1038/s44286-024-00033-5



Extended Data Fig. 2 | Fast-Cat's hardware validation and benchmarking to quantify the inherent experimental noise. a) Reaction stabilization time for a given experimental condition in multiples of the reactor residence time. **b**) Reaction stability across multiple automatic refills of the syringe pumps from the reagent reservoirs (mean of 3 replicates +/− standard deviation). **c**) Consecutive in-line GC sampling of a single hydroformylation reaction condition. **d**) Fast-Cat's experimental noise via random sampling; starting in-flow hydroformylation reaction with condition 1 followed by 5 randomly selected reaction conditions before returning to condition 1 (labelled as condition 7) (mean of 3 samples at each reaction condition +/− standard deviation). Detailed reaction conditions available in the Supplementary Information section Benchmarking and Validation Conditions.

https://doi.org/10.1038/s44286-024-00033-5



Extended Data Fig. 3 | The hydroformylation reaction space exploration of 1-octene with ligand L1 using the digital twin built by the in-house experimental data of Fast-Cat. a) Surface plots of predicted aldehyde yield with L1 as a function of (i) pressure and temperature, (ii) pressure and olefin to Rh fraction, and (iii) temperature and olefin to Rh fraction. **b**) Surface plots of predicted aldehyde regioselectivity with L1 as a function of (i) temperature and H_2 flowrate, (ii) temperature and ligand to Rh fraction, and (iii) ligand to Rh fraction and H_2 flowrate. For all surface plots shown in panels A) and B), the process parameters other than the two variable parameters used for each surface plot were maintained at a constant value (X_i = 0.25, 0.5 and 0.75) for each surface.