Supplementary Information for

An Autonomous Lab for Data-Driven Homogeneous Catalysis

Table of Contents

1. Chemicals

All chemicals were used as received. Fluorophosphite ligand L1 (12-(tert-butyl)-6-fluoro-2,4,8,10-tetrakis(2-phenylpropan-2-yl)-12H-dibenzo[d,g][1,3,2]dioxaphosphocine) was provided by Eastman. Synthesis details and spectra of ligand L1 are reported in prior work[1,2]. Tris(1-naphthyl) phosphine (L2), 97%, Triphenylphosphine (L3), 99%, and Tris(2,4-ditertbutylphenyl) phosphite (L15), 98%, were purchased from Thermo Scientific Chemicals. Toluene (Anhydrous 99.8%), tridecane analytical standard, aldehyde standard (isobutyraldehyde and n-butyraldehyde), and ligands 4-14, and 16 (**Scheme 1** in the main text) were purchased from MilliporeSigma with purities of: Tris(2,4,6-trimethoxyphenyl) phosphine (L4), 97%, Tris(4-methoxyphenyl) phosphine (L5), 95%, Tris(2,4,6-trimethylphenyl) phosphine (L6), 97%, Tris(3,5-dimethylphenyl) phosphine (L7), 96%, Tris(4-methylphenyl) phosphine (L8), 98%, Tris(2-methylphenyl) phosphine (L9), 97%, Tris(3,5-bistrifluoromethylphenyl) phosphine (L10), 97%, Tris(4-trifluoromethylphenyl) phosphine (L11), 97%, Tris(2,3,4,5,6-pentafluorophenyl) phosphine (L12), 97%, Tris(4-fluorophenyl) phosphine (L13), 98%, Triphenyl phosphite (L14), 97%, and Tris(2-furyl) phosphine (L16), 99%. Acetylacetonato dicarbonyl rhodium(I) ($Rh(CO)_2(acac)$), 97%, was purchased from Alfa Aesar. CO (99.999%), H_2 (99.999%), mixed gas propylene (40% CO, 40% H_2, C_3H_6, 20%), and nitrogen (N_2, 99.999%) were purchased from Airgas.

2. Experimental Setup

Robotic XYZA Axis / Glovebox. The automated robotic chemical workstation (*i.e.*, the core of *Flex-Cat* hardware) Chemspeed robotic system consists of an XYZA gantry arm with swappable tool head built into a customized MBraun glove box (MBraun). The arm can access individual vial locations and storage across the deck with the appropriate tool.

Storage. The deck of the robotic system contains holders for solids, liquids, well plates, and spare vials. The solid containers are PEEK containers with a stainless-steel auger that can contain and dispense up to 50g of free-flowing powders. The vial locations on deck can accommodate 20, 40, and 60 mL standard vials for storage, stock preparation, reaction, or replenishment. The three well plates in the system are used for the MTP reactor block and can each hold 40 (5x8) 2mL GC vials. There is also an 8 space cap rack for 20 mL vial caps and a 50 space cap rack for 2 mL GC vial caps for temporary storage while dispensing solids or sampling.

Grippers / Capping. The container manipulators available on the system include an eccentric vial gripper for transporting 20, 40, and 60 mL vials around the deck, a MTP plate gripper for transporting well plates around the deck, and a screw capping tool that can cap and uncap both the 2 mL GC vials as well as the 20, 40, and 60 mL vials.

Liquid Handling. Liquid handling is accomplished using a 4-channel needle tool with adjustable spacing and depth connected to syringe-based diluter pumps. Each pump has sufficient volume of tubing between the syringe and needle to prevent aspiration of material into the syringe. The four

liquid channels in the system are: 1. 1 mL diluter with small volume GC injection needle. 2. 1 mL diluter with vented needle for precision dosing through septa. 3. 10 mL diluter with vented needle for intermediate dosing through septa. And 4. 25 mL diluter with large bore needle for bulk solvent transfer. Each diluter can aspirate from a solvent reservoir for priming, solvent dispense, and needle rinsing, as well as aspirating through the needle from any location on deck for vial-to-vial transfer. Needle rinsing is accomplished between dispenses by aspirating solvent and dispensing through the needle into the rinse and waste locations in the needle tool dock to remove the air gap and extra volume aspiration while flushing the inside and rinsing the outside of the needle. The 1mL diluter with vented needle was used to create stock solutions of aldehyde for GC calibration to validate the dispense accuracy. The liquid-handling accuracy of needles 2 and 3 was evaluated by measuring dispensed solvent mass in triplicate, as shown in **Fig. S1.**

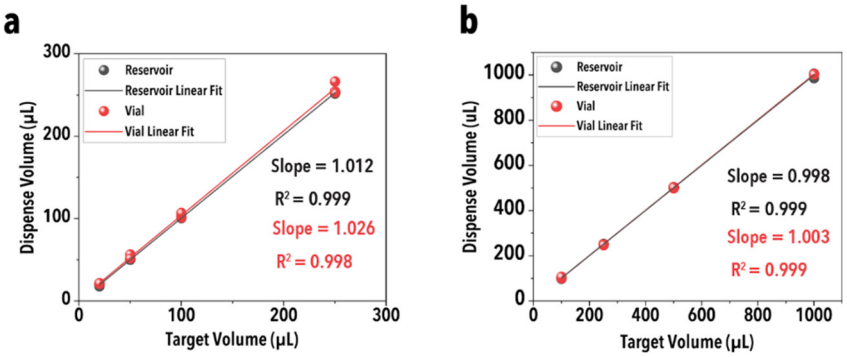

Figure S1| Liquid handling calibration. a, *Liquid dispensing accuracy for both from solvent reservoir and vial-to-vial for needle 2 (1 mL syringe).* ***b,*** *Liquid dispensing accuracy for both from solvent reservoir and vial-to-vial for needle 3 (10 mL syringe).*

Solid Handling. The solid handling tool is the gravitational dispense unit-powder container variant (GDU-Pfd) which consists of a manipulator which can pick up the powder containers from storage, track the mass of the container via balance, and operate the internal auger to release material into the current location. The GDU uses a PID on the auger while tracking the dispensed mass to adjust dispense parameters for materials with differing dispense characteristics. Static charge is mitigated by the presence of de-ionizers placed at several locations around the deck. The GDU can detect when a container is empty and automatically begin using the next container with the same material to be dispensed. The GDU has a dispense tolerance down to 0.1 mg, halts dispense once it crosses the lower boundary (requested dispense - tolerance) and records the dispensed mass in the event log.

Reactors. The deck of the robotic system has three different reactor types: 1. Heater-shaker plate for 20 mL vials. 2. 1mL screening High pressure MTP reactor block. 3. 20 mL scale-up MTP reactor plate. The heater-shaker plate is primarily used for creating stock solutions of the materials in the powder container using the solvent from the diluter pumps, ensuring the stocks are fully dissolved and homogeneous before further utilization. The MTP reactor block enables the addition

of high pressure inert (N_2) and reactive gases (CO, H_2, and propylene mixture) and further elevated temperatures (110C), to run well plates of parallel batches of reactions at 2-20 mL scale. Temperature in the heater-shaker is controlled by a PID and resistive heating element while the MTP uses circulating heat transfer fluid from a Huber Tango Unistat. Reactor temperature for MTP reactions was calibrated by checking the steady-state temperature in the MTP block vs the cryostat set point (**Fig. S2**) The mixing in the heater/shaker plate and MTP reactor block is accomplished by orbital shakers.

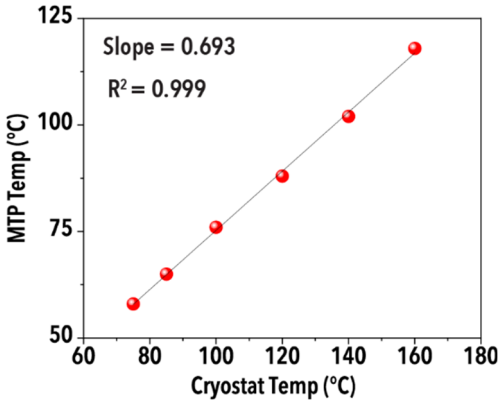

Figure S2| MTP reactor temperature average vs Cryostat Setpoint. *+/- 1°C across multiple wells after two repeated measurements with no temperature change 15 min apart.*

MTP Reactor Well-to-Well Replication. A plate of 32 samples was run in the MTP reactor to investigate the well-to-well variance (**Fig. S3**) at the same experimental condition over the course of the full reaction protocol and analysis. The experimental conditions were Ligand 1, 110°C, 300 psig total, 30 psig propylene, 130 psig CO, 130 psig H_2, 1:1 H_2:CO, 2.5 mM Ligand 1, 0.25mM $Rh(CO)_2(acac)$, 10:1 ligand:Rh at 1 mL scale in toluene. The turnover frequency (TOF) had a coefficient of variation (CV) of 3.75% across the entire plate and the aldehyde regioselectivity (S_N) had a CV of 0.43% indicating good well-to-well agreement through multiple dispense operations as well as heating, mixing, and gas supply.

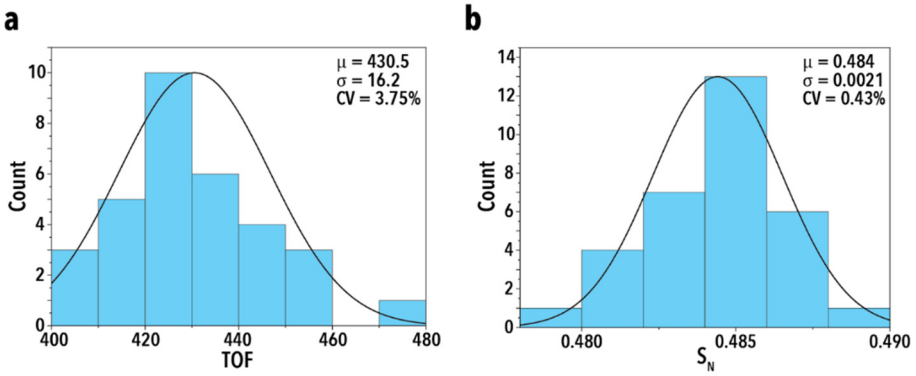

Figure S3 | MTP Reactor well-to-well variation. *32 replicates of Ligand 1, 110°C, 300 psi total, 30 psi propylene, 130 psi CO, 130 psi H_2, 1:1 H_2:CO, 2.5 mM Ligand 1, 0.25mM $Rh(CO)_2(acac)$, 10:1 ligand:Rh at 1 mL scale in toluene.* ***a,*** *TOF histogram, mean, standard deviation, and CV.* ***b,*** *S_N histogram, mean, standard deviation, and CV.*

3. Model/Algorithm Validation

This section details the procedure by which experimental data generated by *Flex-Cat* are leveraged to construct a high-fidelity machine-learning "digital twin" using deep neural networks (DNNs). The objective is to capture the underlying reaction landscape through a virtual reaction simulation. Model performance is assessed using the coefficient of determination (R^2) computed from validation parity plots, which quantifies the fraction of variance in the outputs (aldehyde TOF or regioselectivity) explained by the inputs. All implementation and training were performed in Python 3.9.13 using NumPy, pandas, and TensorFlow. Previous comparisons have shown that a cascade-forward network (CFNN) outperforms a standard feedforward neural network (FFNN), a benefit attributed to its multi-stage feature-extraction pathway, improved propagation and correction of errors across layers, enhanced capacity for high-dimensional nonlinear mapping, and greater resilience to measurement noise[1,3]. For architecture tuning, an architecture search was conducted over network depth (*n* layers) and width (*m* nodes per layer), with the configuration maximizing R^2 selected. **Fig. S4** presents the validation R^2 for aldehyde TOF and regioselectivity (with one-hot encoded ligands L1–L16) as functions of layer count (*x*-axis) and node count (*y*-axis), respectively.

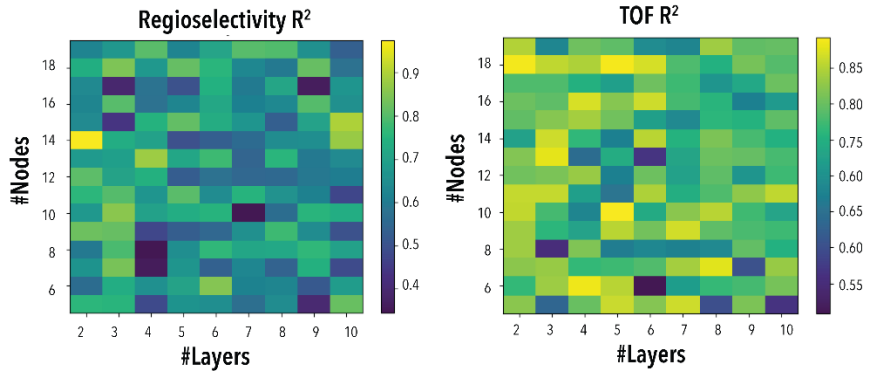

Figure S4| Neural network architecture screening. a, *Heatmap of regioselectivity (R^2) across combinations of hidden layer numbers and nodes per layer.* ***b,*** *Heatmap of TOF (R^2) across combinations of hidden layer numbers and nodes per layer.*

To quantify predictive uncertainty, an ensemble-NN approach was implemented. An optimized range of CFNN architectures was defined through the same screening procedure described (**Fig. 5**). The digital-twin ensemble was constructed by randomly drawing architectures from within this range. Instead of simple averaging, an outer two-layer neural network (stacking) was trained to nonlinearly combine the individual members' predictions, resulting in superior ensemble performance.

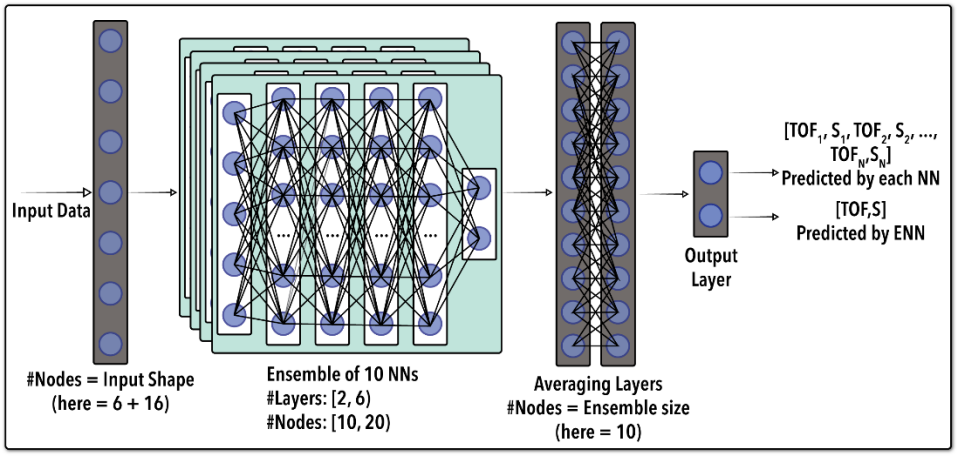

Figure S5| Neural network ensemble architecture. *An ensemble of neural networks with randomly sampled layer numbers and node counts within a defined range. The seven input features include excess CO fraction (y_{CO}), propylene mixture fraction (y_{Mix}), total pressure (P), reactor temperature (T), ligand-to-(ligand + Rh) ratio (X_L), solvent fraction (X_S), and ligand identity. Two final averaging layers integrate the predictions from individual networks to yield the ensemble output.*

Additional hyperparameters, including learning rate and optimizer choice, were fine-tuned to ensure stable convergence (**Fig. S6**). For regression targets, the root–mean–square error loss was minimized using the *RMSprop* optimizer. Hidden layers employed *ReLU* activation, whereas outputs were linear, with an upper-bound constraint (≤ 1.0) imposed to respect the physical limits of fractional selectivity.

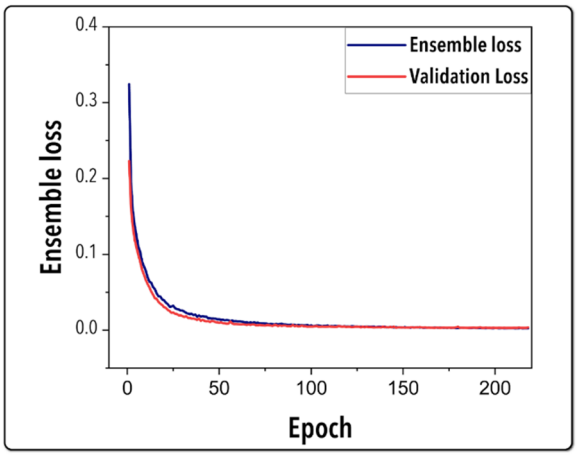

Figure S6 | Training (ensemble) and validation loss. *The model's training and validation loss across epochs. Both curves drop rapidly in early training and plateau near zero by ~200 epochs; their close overlap indicates stable convergence with minimal overfitting.*

The data-driven reaction digital twin was modeled by partitioning the experimental LHS data into 80% training, 10% validation, and 10% test sets. Early stopping based on validation loss was applied to prevent overfitting, and final model accuracy was evaluated on the held-out test set. **Fig. S7** presents parity plots comparing ensemble predictions with experimental measurements, demonstrating that the digital twin reproduces *Flex-Cat* outcomes with high fidelity. Model performance, expressed as R^2, reached 0.72 (TOF, test set) and 0.85 (S_N, test set), and 0.87 (TOF, full dataset) and 0.95 (S_N, full dataset).

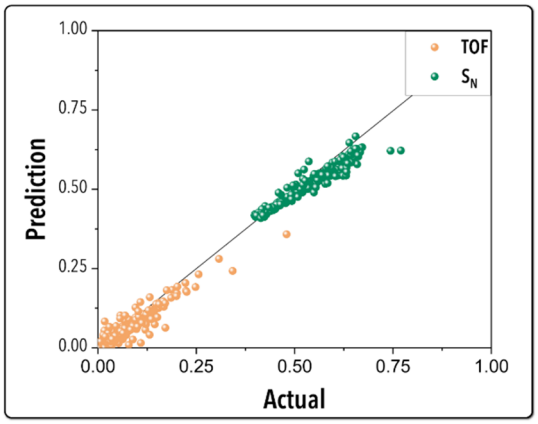

Figure S7* | *Parity plots for S_N and TOF. *Predicted values by the ML model vs the actual values are depicted for both selectivity and normalized TOF. R^2 values reached 0.72 for TOF (test set) and 0.85 for S_N (test set), and 0.87 for TOF (full dataset) and 0.95 for S_N (full dataset).*

Initialization. An initial LHS of 10 plates × 16 ligands × 2 ligand and solvent fractions (ligand / (ligand + Rh), solvent fraction) which is 10 rounds of 32 wells in a reaction plate is generated, holding plate-based reaction conditions (T, P, p_{CO}, p_{H_2}) constant within each plate. These 320 experiments are used to train the two Gaussian-Process (GP) surrogate models for regioselectivity (S_N) and turnover frequency (TOF). After model initialization and prior to the Bayesian optimization cycles, batch-size screening was conducted with the trained digital twin to determine the optimal number of experiments per MTP. Three batch sizes (10, 20, and 40; **Fig. 8**) were evaluated by calculating the average Pareto-front area. Screening revealed that 10 reactions per plate achieved the highest coverage, with an average area of 0.244.

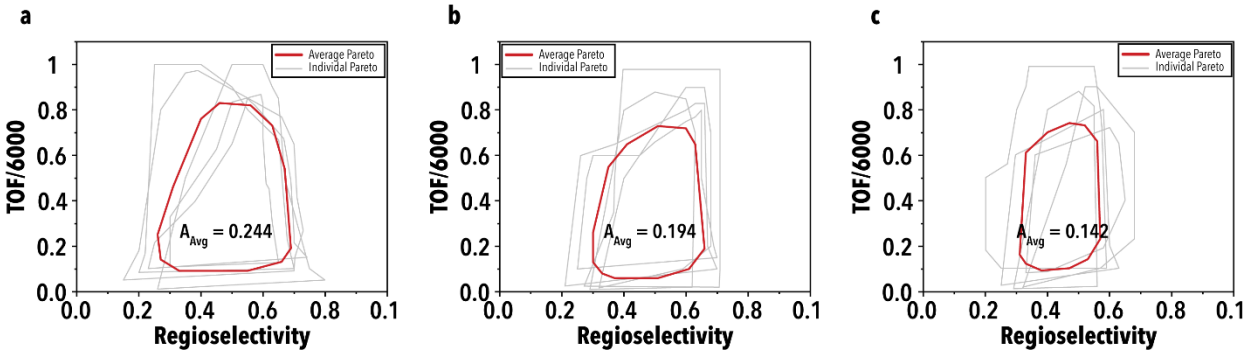

Figure S8* | *Experiment batch size optimization. *Fine-tuning of the BO batch size to identify the optimal number of reaction vials per MTP.* ***a,*** *10 experiments per plate yielded the largest Pareto-front area (0.244).* ***b,*** *20 experiments per plate resulted in an average area of 0.194.* ***c,*** *40 experiments per plate gave the smallest area of 0.142.*

Algorithm Benchmarking and Comparative Analysis. To rigorously evaluate the performance of our bespoke optimization framework, we performed an *in-silico* benchmarking of the custom-developed machine learning (ML) pipeline (using the data-driven reaction digital twin) against three alternative strategies with the same total budget, including **(i)** random sampling, **(ii)** Latin

hypercube sampling (LHS), and **(iii)** linear objective optimization (without *ReLU* bonus). Each method was repeated for five independent runs, and the average results are presented in **Figs. S9-S11**.

Random Sampling. As a baseline, we first considered uniform random sampling under a total experimental budget of 688 conditions. Specifically, 43 random combinations were generated for the continuous variables, and each condition was expanded across all 16 ligands, resulting in 688 total experiments. For each condition, the ensemble neural network (ENN) model trained on the experimental data (*i.e.*, data-driven digital twin of the tested oxo reaction) was queried to predict S_N and TOF.

The red curve in **Fig. S9** demonstrates that random sampling fails to effectively improve the objective function

$$Z = 0.5S + 0.5\mathrm{TOF} + 1.5\ \mathrm{ReLU}(S - 0.5).$$

After approximately 200 *in silico* runs, the objective value reaches a plateau, indicating limited exploration efficiency and poor convergence toward high-performing regions of the design space.

LHS. We next evaluated LHS as a structured design of experiment alternative to random sampling. In this case, 43 LHS levels were generated for the continuous variables and again expanded across all 16 ligands to maintain the same experimental budget. As shown in **Fig. S9**, similar to random sampling, LHS fails to substantially improve the objective value *Z*, indicating that uniform space-filling designs alone are insufficient for identifying high-performing regions in this complex, mixed-variable space.

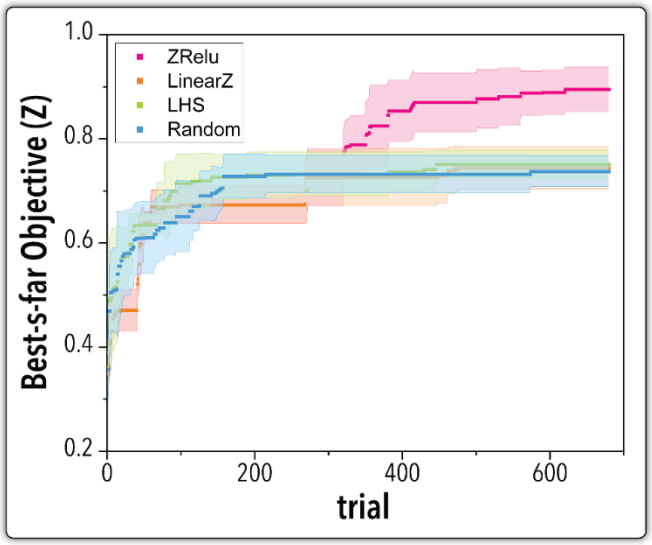

Figure S9| In-silico benchmarking. *Comparison of maximum obtained acquisition function score for 4 sampling methods tested in silico (random sampling, LHS, BO with linear weights (Z), and BO with a selectivity boost, ZReLU). Experimental conditions were determined using the associated acquisition function but plotted using the ZReLU acquisition function to maintain the same scale on output.*

Linear Objective Optimization (Without ReLU Bonus). To further probe the influence of the objective formulation, we implemented a third strategy using a simplified objective function without the *ReLU* bonus term. Specifically, 10 LHS levels were generated for the plate variables (x_1-x_4) and 20 LHS levels for well concentrations (x_5, x_6). These were expanded across all 16 ligands, yielding 320 initialization experiments.

A surrogate model was trained using this dataset, and optimization was conducted using the linear objective:

$$Z_{\text{linear}} = 0.5S + 0.5\text{TOF}.$$

For fair comparison across methods, the final performance metric reported on the vertical axis in **Fig. S9** is the global objective *Z* (including the *ReLU* bonus term), even for campaigns optimized using the linear formulation. Importantly, both *in-silico* linear objective and the *ReLU*-enhanced BO campaigns were initialized using identical LHS data to ensure consistency in surrogate training and eliminate initialization bias.

As shown in **Fig. S9**, the BO campaign under the linear objective produces gradual improvements in *Z*. However, it does not reach the optimum identified by the bespoke BO which directly optimizes the full objective including the *ReLU* bonus term.

Budget Considerations. The experimental campaign for our method used a total budget of 680 experiments. Because 680 is not divisible by 16 (the number of ligands), the random and full LHS baselines required eight additional virtual well plates to reach 688 conditions. This minor difference does not affect the comparative conclusions but is noted for completeness.

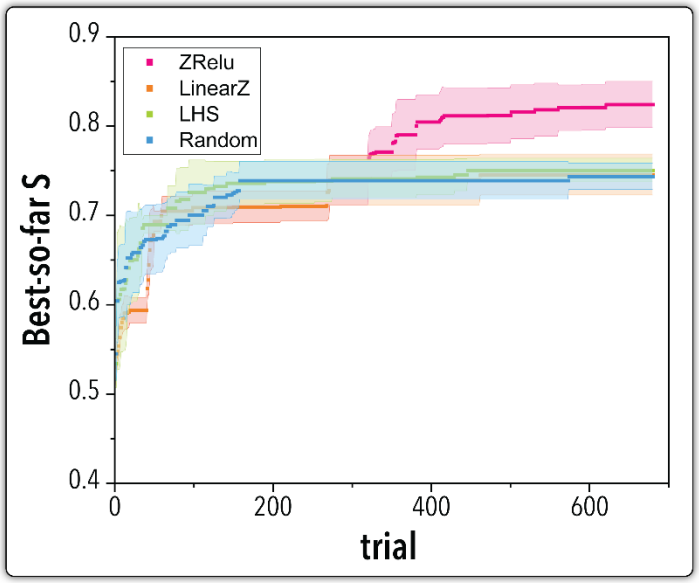

Figure S10| Digital benchmarking. *Comparison of maximum obtained aldehyde regioselectivity for 4 sampling methods tested in silico (random sampling, LHS, BO with linear weights (LinearZ), and BO with a selectivity boost, ZReLU).*

Regioselectivity and TOF Evolution. Since Z is a composite mathematical objective, we additionally report the evolution of S_N and TOF individually for each *in silico* tested sampling method (**Fig. S10** and **S11**). The bespoke hierarchical BO strategy with the *ReLU*-enhanced objective achieves the highest aldehyde regioselectivity within the same experimental budget.

Although the linear objective produces slightly higher TOF values in some cases, the corresponding gain (~10%) is substantially smaller than the improvement in regioselectivity (~20%) achieved by the *ReLU*-based method. From an industrial perspective, enhanced regioselectivity is often more valuable than marginal improvements in TOF, as higher selectivity directly translates to more pure products, reduced separation costs, and lower waste generation.

Overall, these results demonstrate that incorporating a threshold-based bonus term into the BO objective meaningfully reshapes the search landscape, guiding the algorithm toward chemically and industrially relevant optima that are not accessible through uniform sampling or purely linear objective formulations. Among all tested sampling/optimization methods, the hierarchical BO using the *ReLU*-enhanced objective achieved the widest accessible range of S_N values while maintaining competitive TOF performance. A broader benchmark against additional batch acquisition strategies, including qEI/qNEI, Thompson sampling, qUCB, and hypervolume-based multi-objective policies, is outside the scope of the present experimental study but is enabled by the released modular codebase.

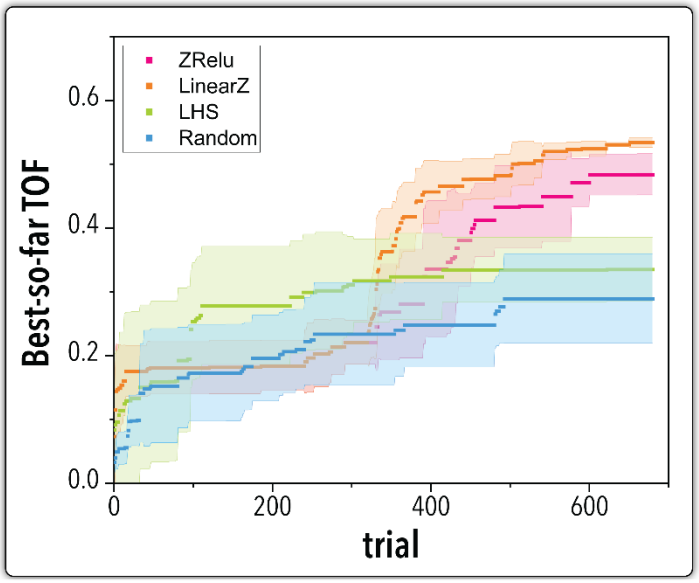

Figure S11| Digital benchmarking. *Comparison of maximum obtained normalized TOF for 4 sampling methods tested in silico (random sampling, LHS, BO with linear weights (LinearZ), and BO with a selectivity boost, ZReLU).*

4. Bayesian Optimization

Plate-Based Reaction Candidate Generation. The Bayesian optimization workflow is shown in **Fig. S12.** Random Sampling of reaction plate-based parameters (128 arrays of [CO, H_2, P, T]): 128

distinct combinations of syngas composition, total pressure, and temperature are drawn. Random Sampling of Local Conditions (512 arrays of [ligand / (ligand + Rh), solvent fraction] + ligand): For each plate-based reaction condition, 512 ligand identities (one-hot encoded) and reagent-concentration variations are sampled, yielding 65,536 virtual experiments.

Plate-Based UCB Scoring. For each plate-based reaction condition candidate, the two GP models predict $Z = 0.5 \cdot S_N + 0.5 \cdot \text{TOF} + 1.5 \cdot ReLU(S_N - 0.5)$ across its 512 local variations. The mean Z plus $\beta \cdot$(standard deviation) is computed to give an upper-confidence-bound (UCB) score for each plate-based reaction combination.

Plate-based Surrogate Model and Acquisition. A third GP is trained with inputs of 128 plate-based parameter sets and outputs of the calculated UCB scores. An Expected Improvement (EI) acquisition function then proposes the most promising plate-based reaction condition by balancing exploration (uncertainty) and exploitation (high predicted Z).

Local Candidate Selection under Fixed Plate-based Reaction Conditions. With the new plate-based reaction condition set fixed, a constrained local BO is performed to select ligand identity (one-hot encoded) and reagent concentrations that maximize Z. Candidate diversity is enforced by checking for repeats; if encountered, switching to alternative acquisition functions (*e.g.*, UCB with random α or Noisy EI) to encourage exploration.

Experiment Execution and Model Update. The selected batch of experiments was executed using the MTP reactor, with newly acquired Z-measurements appended to the dataset and all three GP models subsequently retrained. To balance exploration and exploitation, every third EI-driven iteration was interleaved with a pure exploitation step, maximizing Z to more aggressively refine the known optimum. The closed-loop cycle continued until the experimental budget was exhausted.

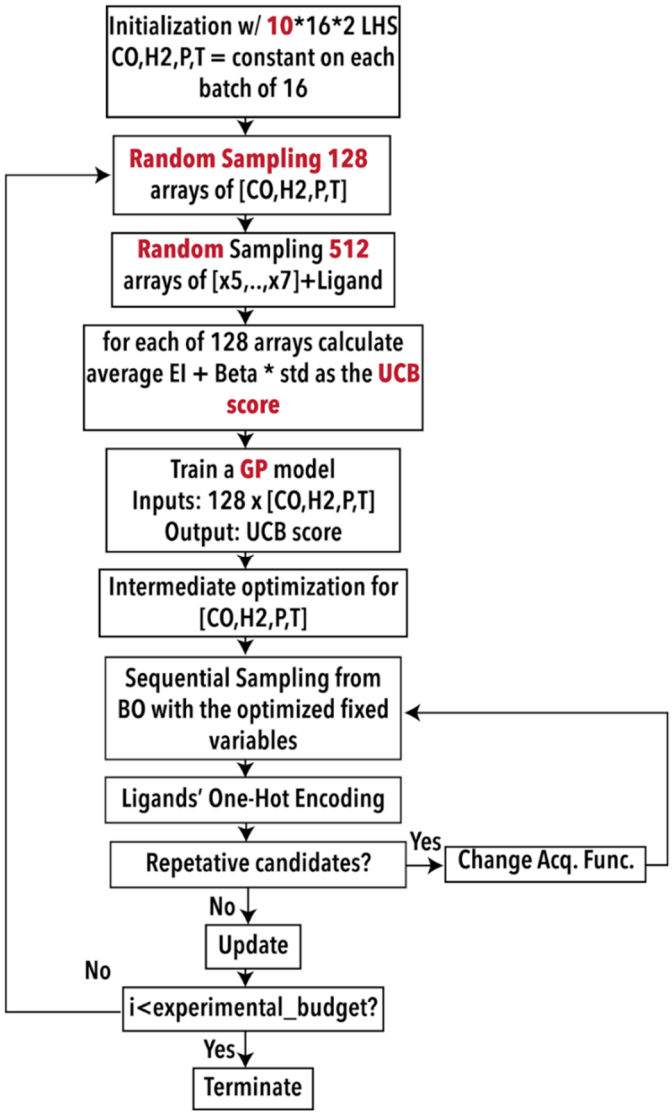

Figure S12* | *Flex-cat's BO workflow. *The hierarchical, two-stage BO workflow efficiently navigates the mixed discrete-continuous parameter space.*

5. Regioselectivity and Flexibility Evolution

Cycle-wise evolution of the minimum and maximum regioselectivity, flexibility scores, and TOF is shown for all ligands in the library (**Figs. S13-S44**). Each ligand trajectory begins with the initialization phase (10 LHS cycles), followed by subsequent optimization cycles in which the ligand is actively selected. The frequency of selection varies across ligands, reflecting the belief model's assessment of whether further improvement in the objective function (Z) or flexibility score is possible. During optimization, each cycle tries to push regioselectivity toward either the linear or branched direction, and these shifts inevitably alter the flexibility score if there is a change in the minimum or maximum regioselectivity. For ligands that consistently exhibit extreme

regioselectivity on either the linear or branched side (ligands 2, 6, 9, 12, and 16), the flexibility score is negative, and its evolution depends on the direction of the regioselectivity shift. In contrast, ligands that span both sides of the threshold (minimum regioselectivity below 0.5, l/b = 1, and maximum regioselectivity above 0.5) exhibit positive flexibility, which increases with any broadening of their regioselectivity range. A notable exception is Ligand 13: during initialization, increases in maximum regioselectivity do not alter flexibility because the minimum regioselectivity remains fixed at 0.5, yielding a score of zero. Only after the first linear EPLT cycle, when the minimum regioselectivity drops below 0.5, does the flexibility score rise sharply. The same cycle-wise evolution was calculated for TOF, showing the minimum and maximum values in each cycle. In certain cycles, a decrease in TOF is observed, which likely reflects the trade-off between optimizing selectivity and optimizing TOF.

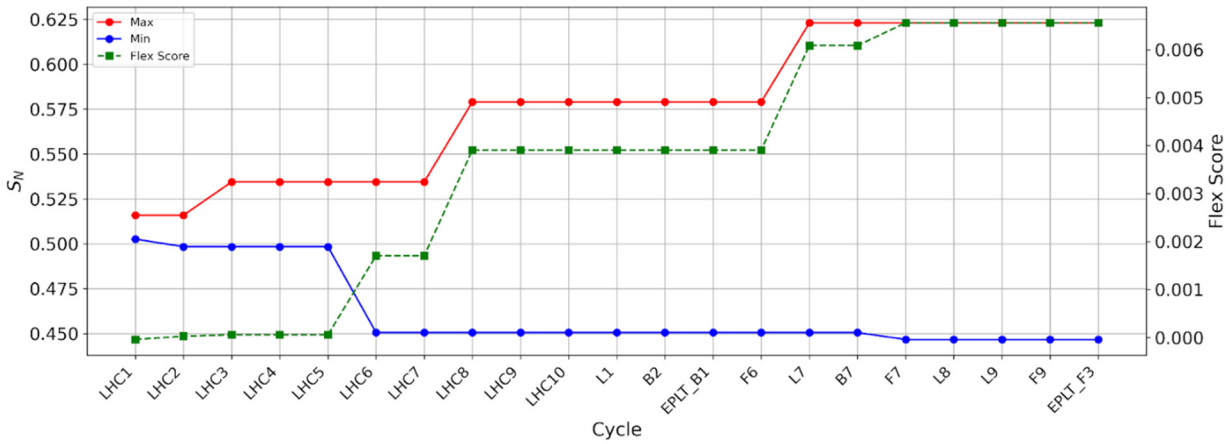

Figure S13| Regioselectivity and flexibility evolution for L1. *Cycle-wise evolution of the minimum and maximum regioselectivity and flexibility score for ligand 1. EPLT cycles are annotated by campaign and order, with EPLT_B1 denoting the first exploitation cycle in the branched optimization campaign.*

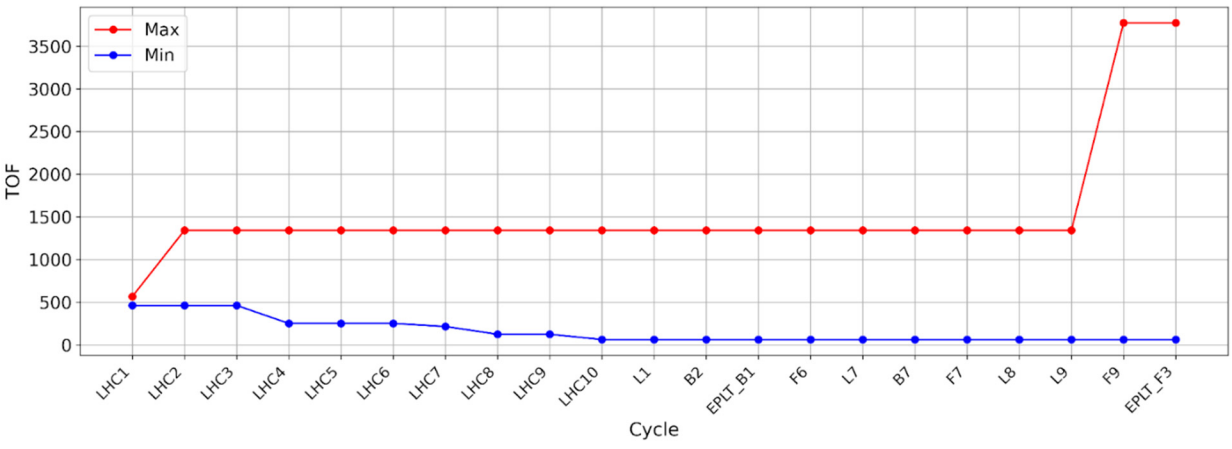

Figure S14| TOF evolution for L1. *Cycle-wise evolution of the minimum and maximum TOF for ligand 1.*

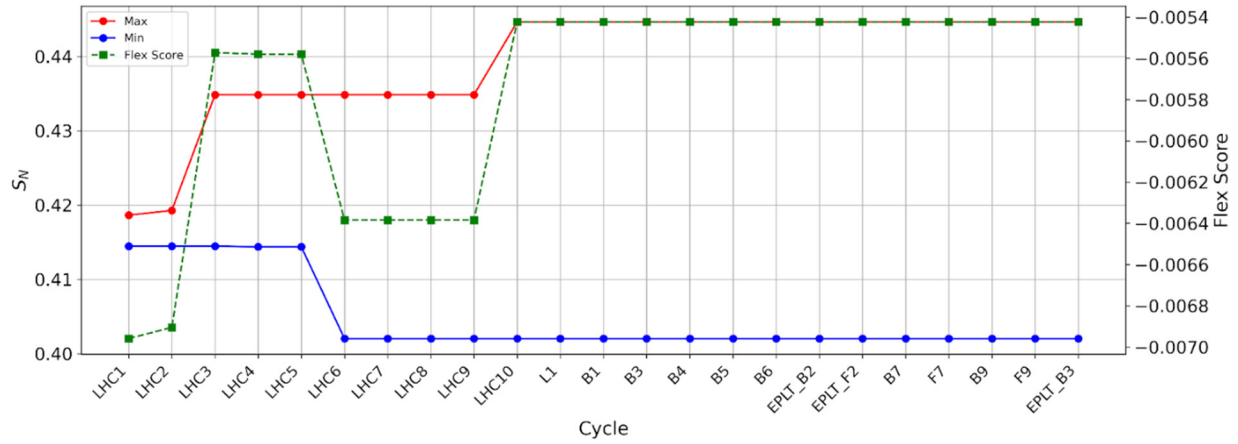

Figure S15| Regioselectivity and flexibility evolution for L2. *Cycle-wise evolution of the minimum and maximum regioselectivity and flexibility score for ligand 2.*

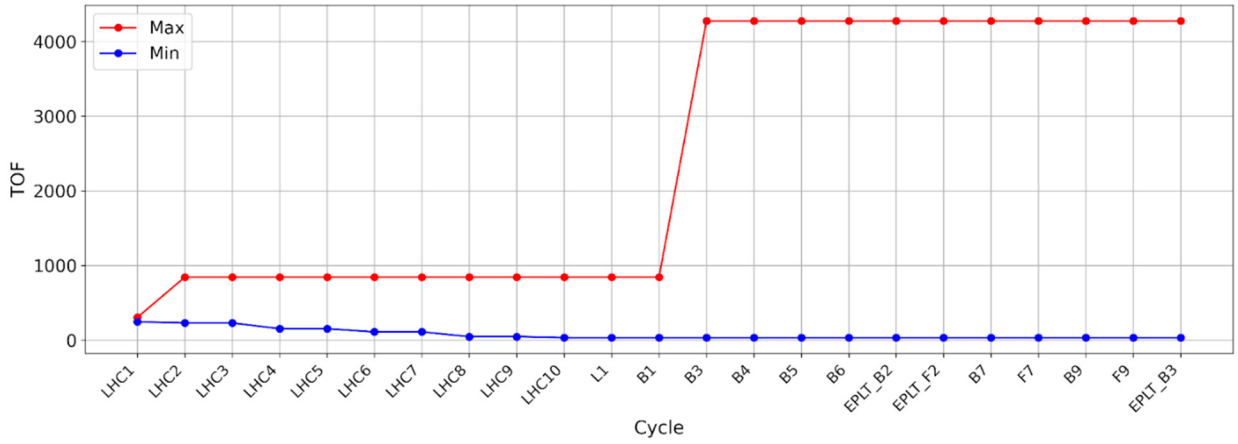

Figure S16| TOF evolution for L2. *Cycle-wise evolution of the minimum and maximum TOF for ligand 2.*

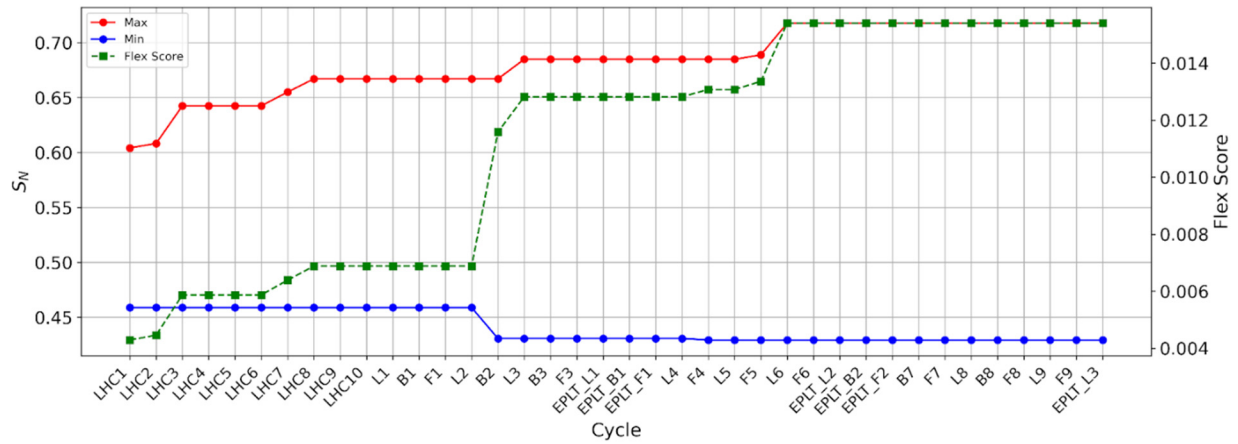

Figure S17| Regioselectivity and flexibility evolution for L3. *Cycle-wise evolution of the minimum and maximum regioselectivity and flexibility score for ligand 3.*

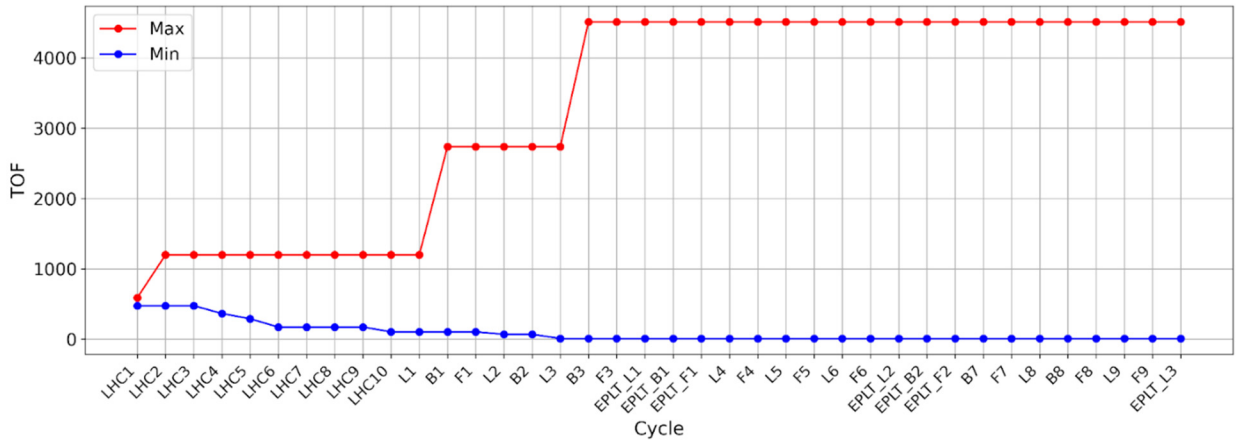

Figure S18| TOF evolution for L3. *Cycle-wise evolution of the minimum and maximum TOF for ligand 3.*

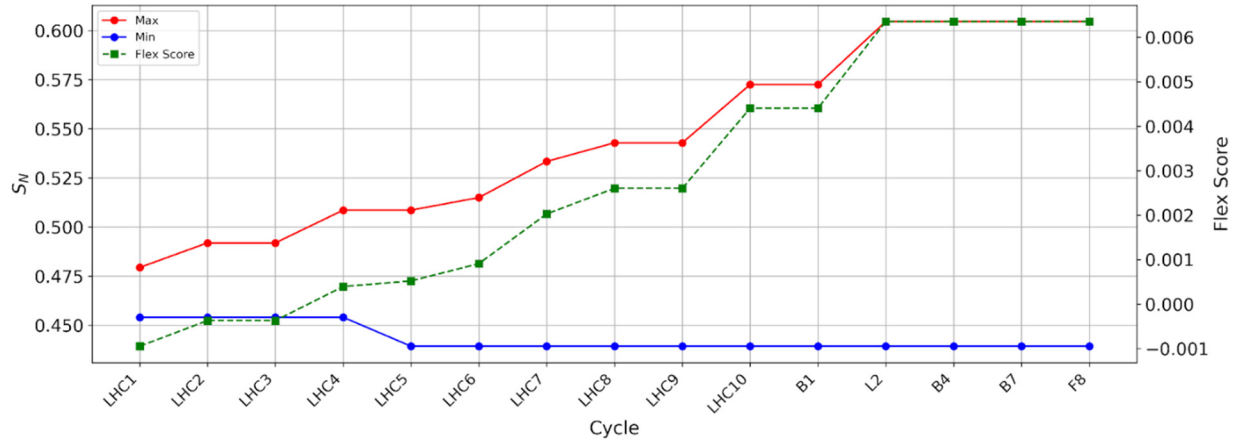

Figure S19| Regioselectivity and flexibility evolution for L4. *Cycle-wise evolution of the minimum and maximum regioselectivity and flexibility score for ligand 4.*

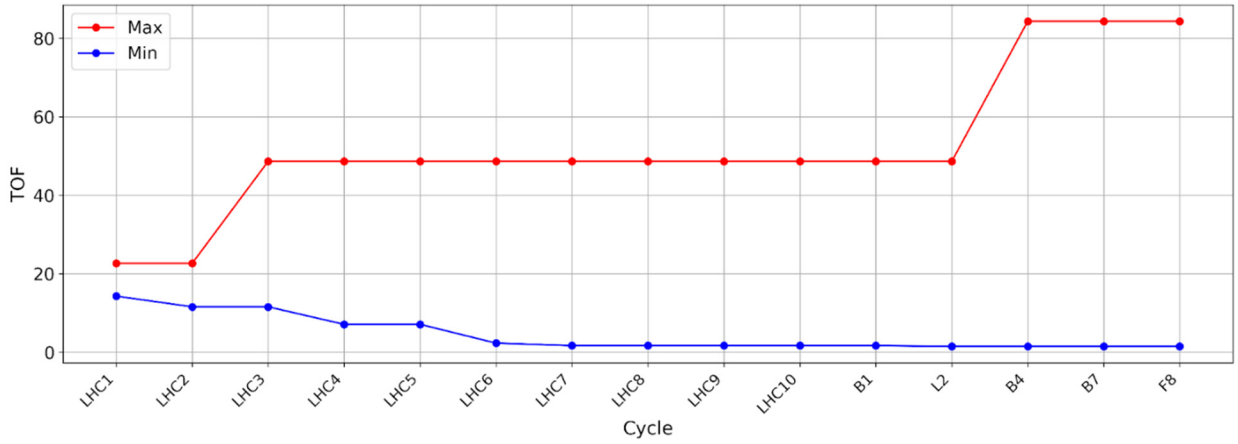

Figure S20| TOF evolution for L4. *Cycle-wise evolution of the minimum and maximum TOF for ligand 4.*

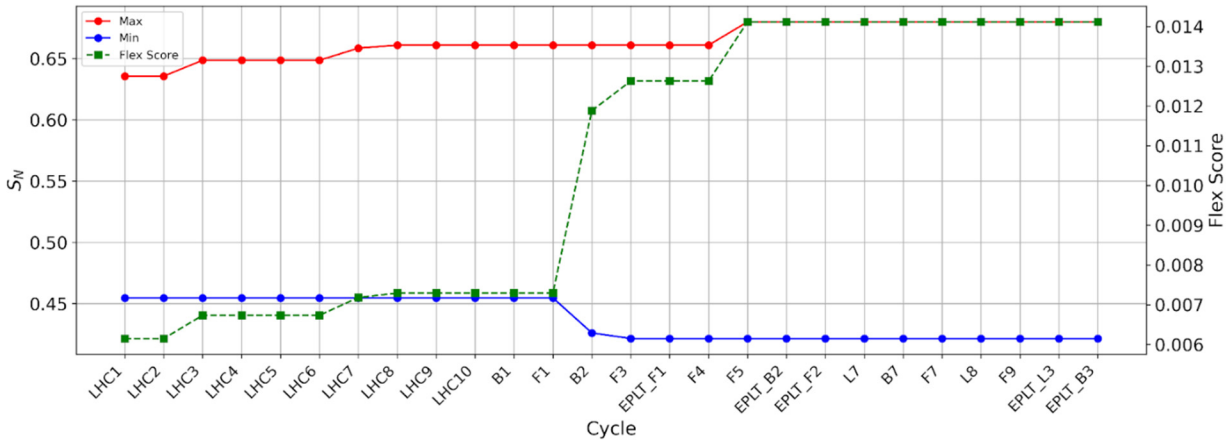

Figure S21| Regioselectivity and flexibility evolution for L5. *Cycle-wise evolution of the minimum and maximum regioselectivity and flexibility score for ligand 5*

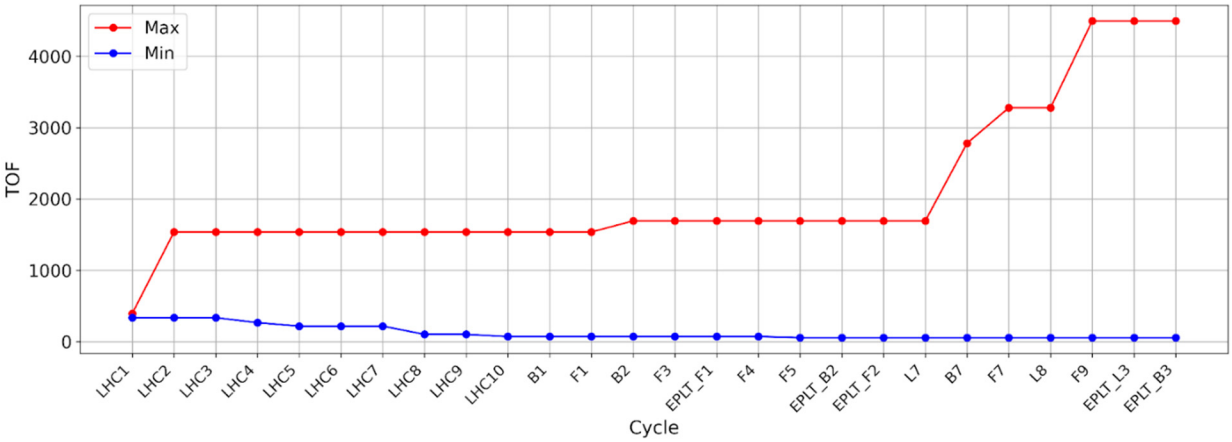

Figure S22| TOF evolution for L5. *Cycle-wise evolution of the minimum and maximum TOF for ligand 5.*

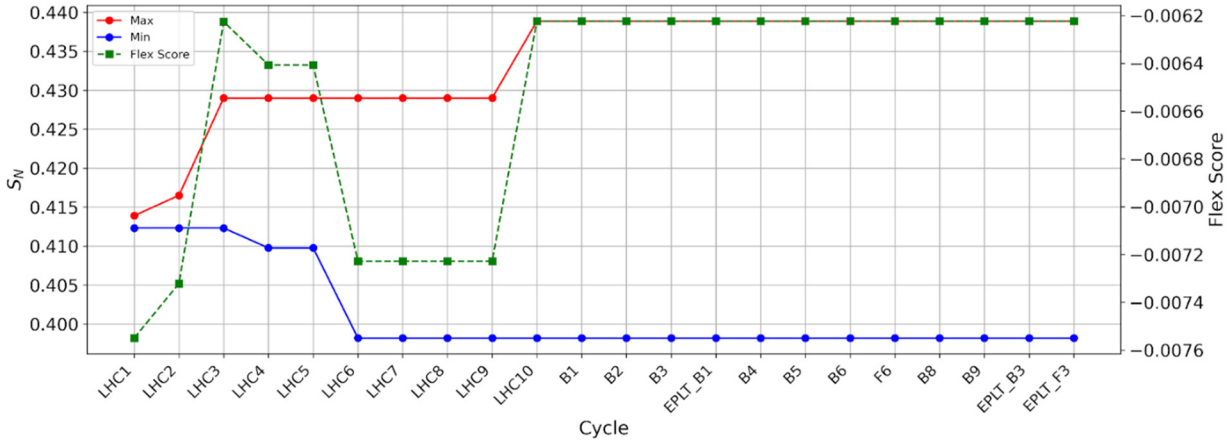

Figure S23| Regioselectivity and flexibility evolution for L6. *Cycle-wise evolution of the minimum and maximum regioselectivity and flexibility score for ligand 6.*

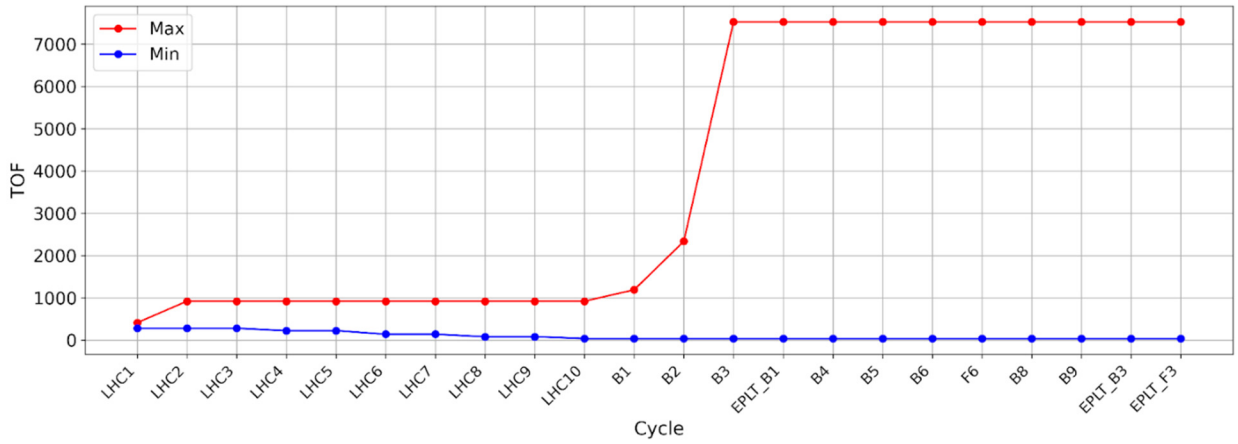

Figure S24| TOF evolution for L6. *Cycle-wise evolution of the minimum and maximum TOF for ligand 6.*

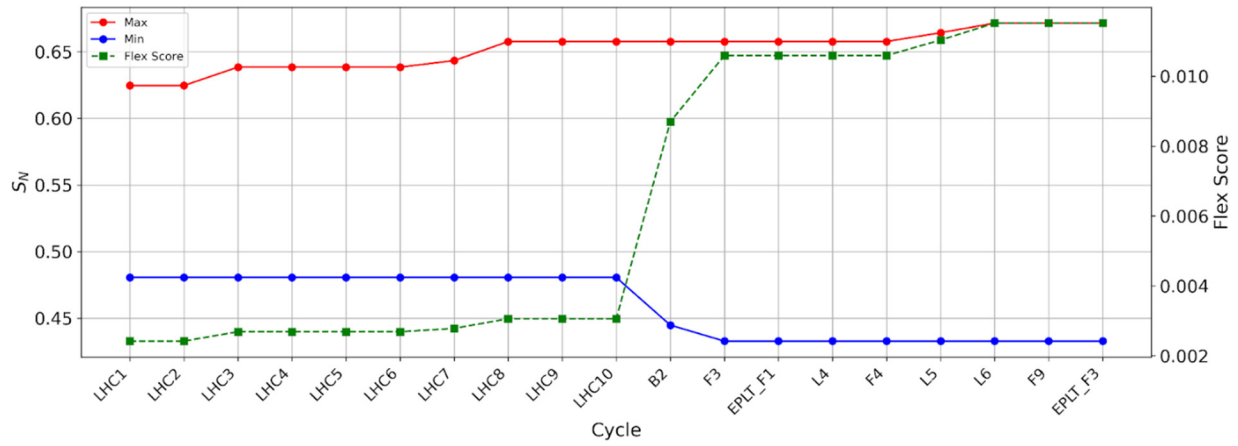

Figure S25| Regioselectivity and flexibility evolution for L7. *Cycle-wise evolution of the minimum and maximum regioselectivity and flexibility score for ligand 7.*

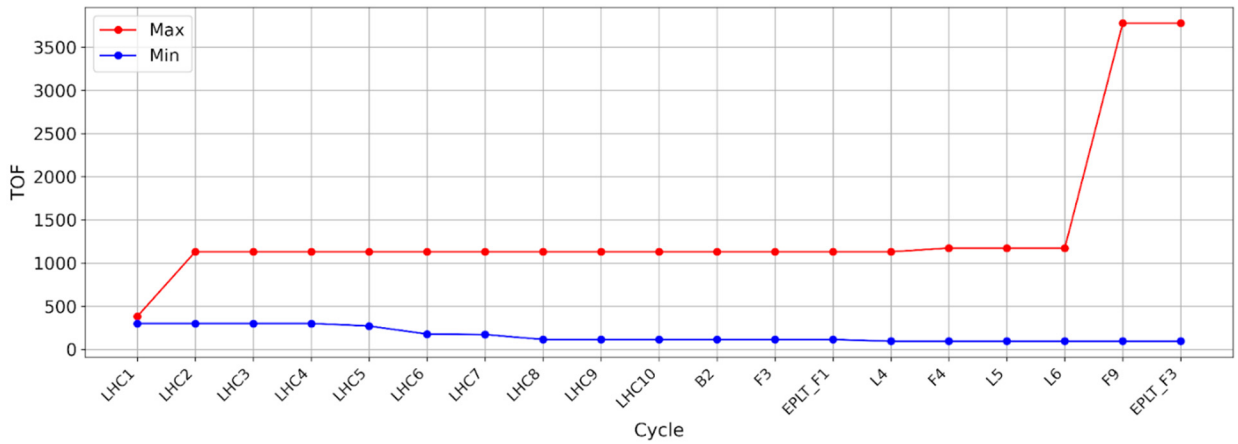

Figure S26| TOF evolution for L7. *Cycle-wise evolution of the minimum and maximum TOF for ligand 7.*

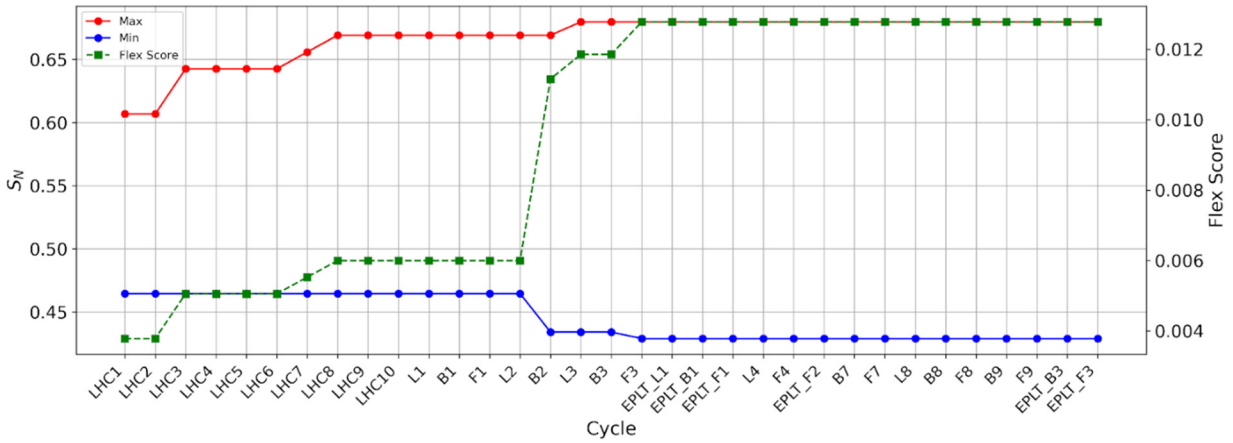

Figure S27| Regioselectivity and flexibility evolution for L8. *Cycle-wise evolution of the minimum and maximum regioselectivity and flexibility score for ligand 8.*

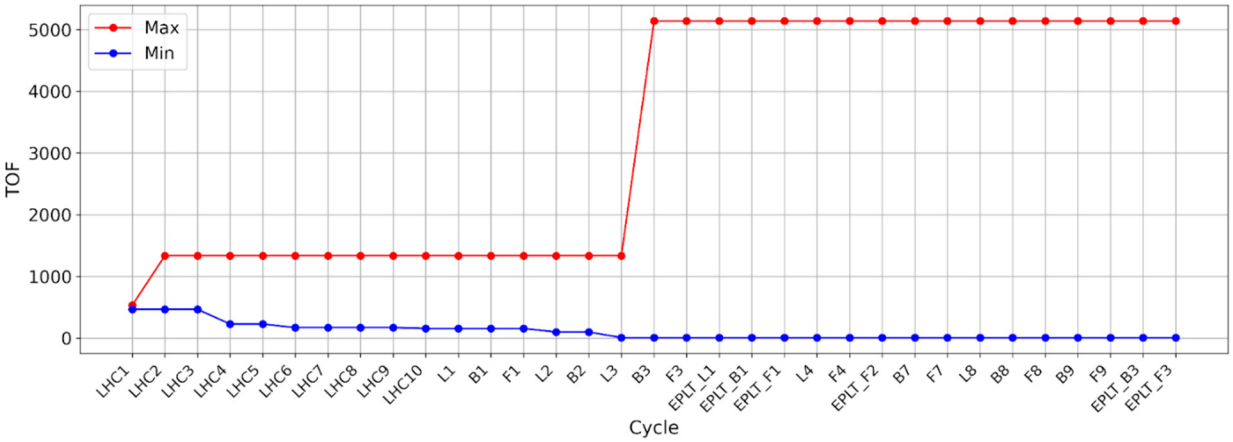

Figure S28| TOF evolution for L8. *Cycle-wise evolution of the minimum and maximum TOF for ligand 8.*

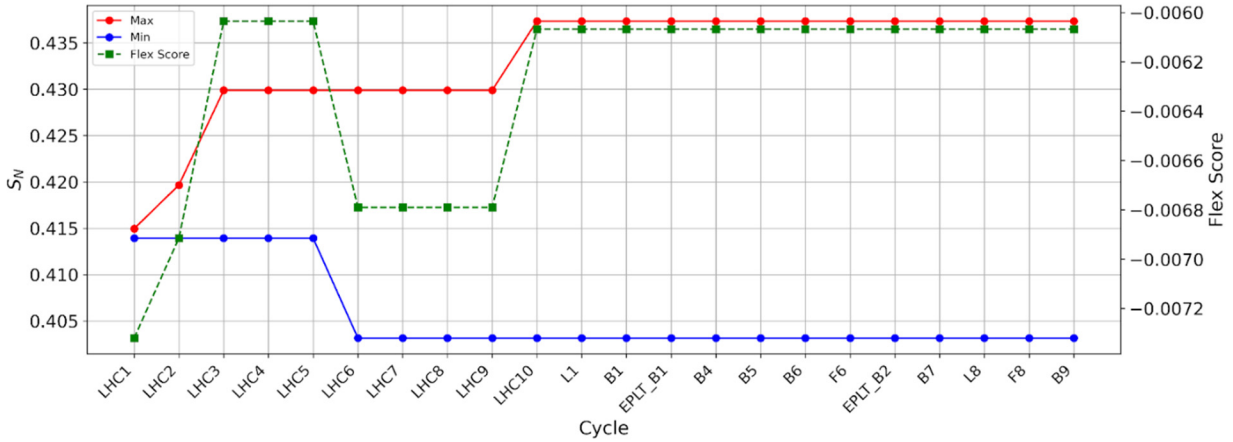

Figure S29| Regioselectivity and flexibility evolution for L9. *Cycle-wise evolution of the minimum and maximum regioselectivity and flexibility score for ligand 9.*

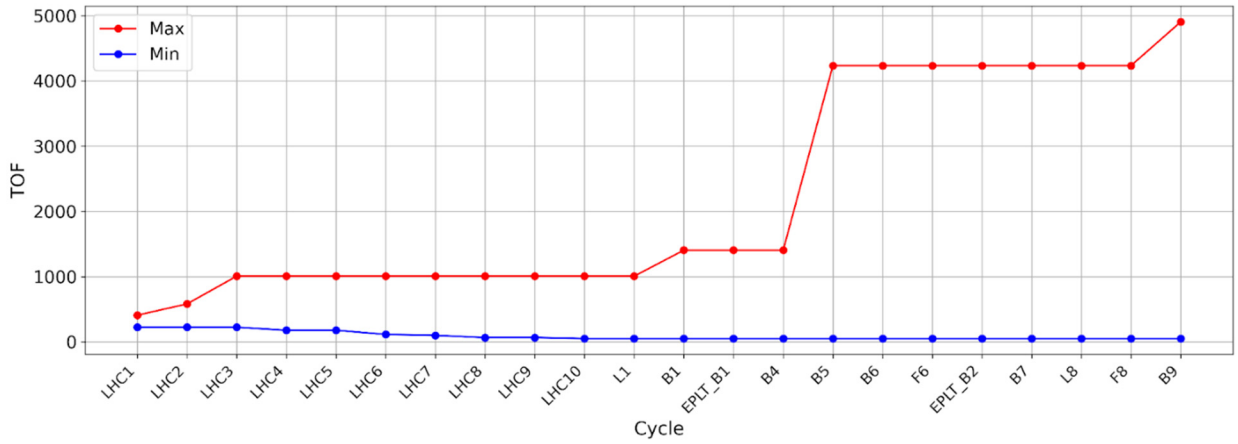

Figure S30| TOF evolution for L9. *Cycle-wise evolution of the minimum and maximum TOF for ligand 9.*

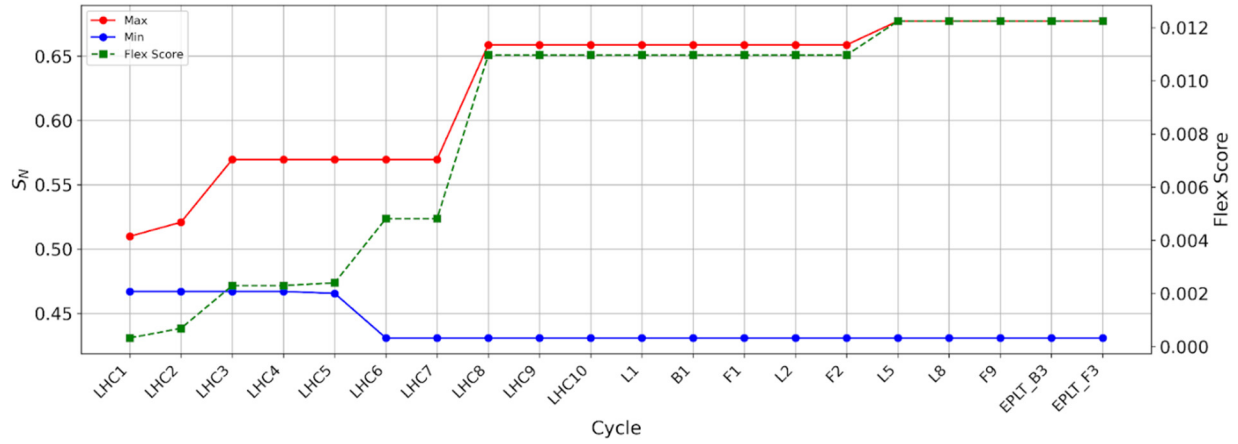

Figure S31| Regioselectivity and flexibility evolution for L10. *Cycle-wise evolution of the minimum and maximum regioselectivity and flexibility score for ligand 10.*

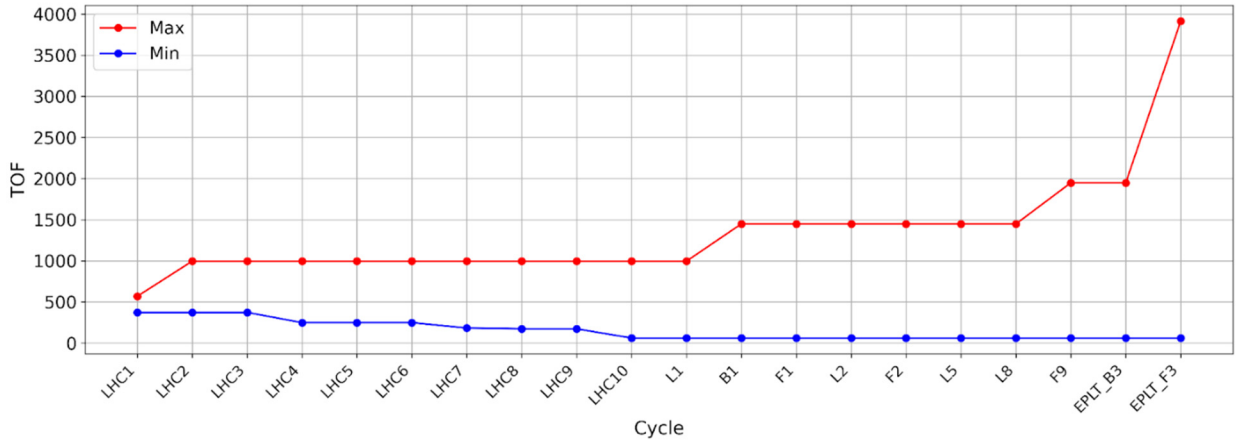

Figure S32| TOF evolution for L10. *Cycle-wise evolution of the minimum and maximum TOF for ligand 10.*

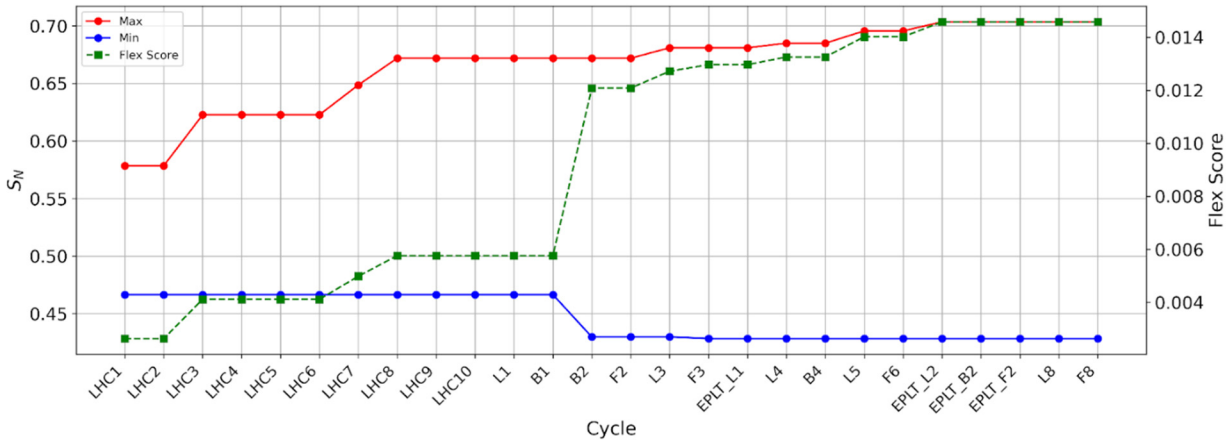

Figure S33| Regioselectivity and flexibility evolution for L11. *Cycle-wise evolution of the minimum and maximum regioselectivity and flexibility score for ligand 11.*

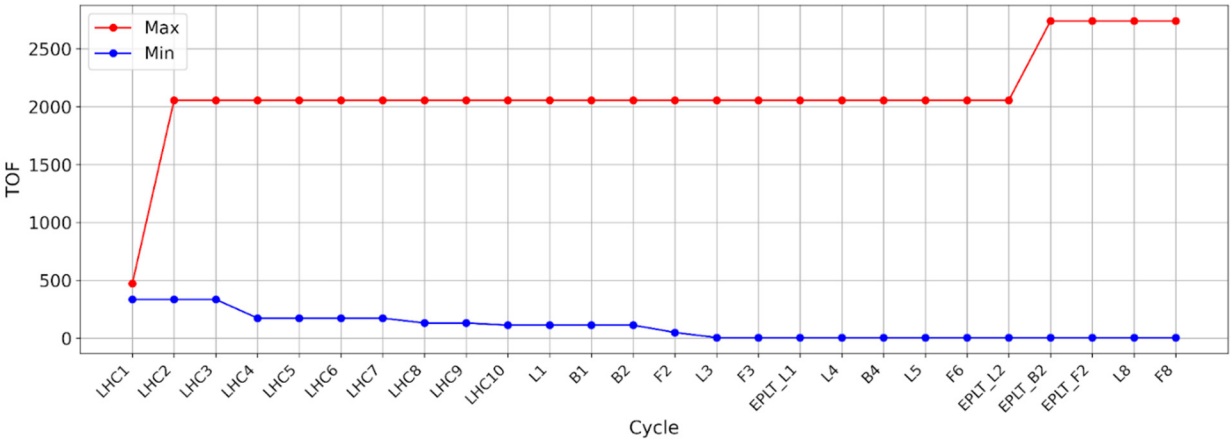

Figure S34| TOF evolution for L11. *Cycle-wise evolution of the minimum and maximum TOF for ligand 11.*

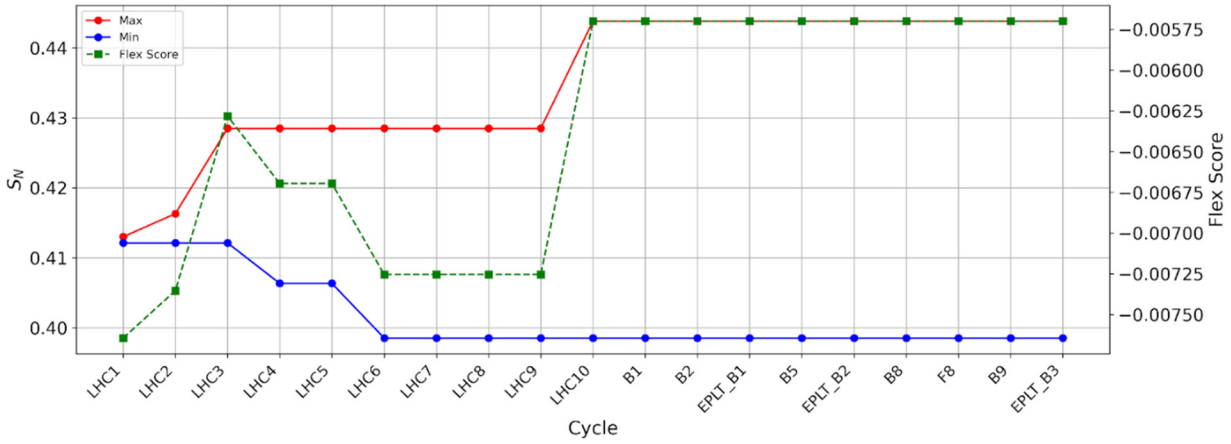

Figure S35| Regioselectivity and flexibility evolution for L12. *Cycle-wise evolution of the minimum and maximum regioselectivity and flexibility score for ligand 12.*

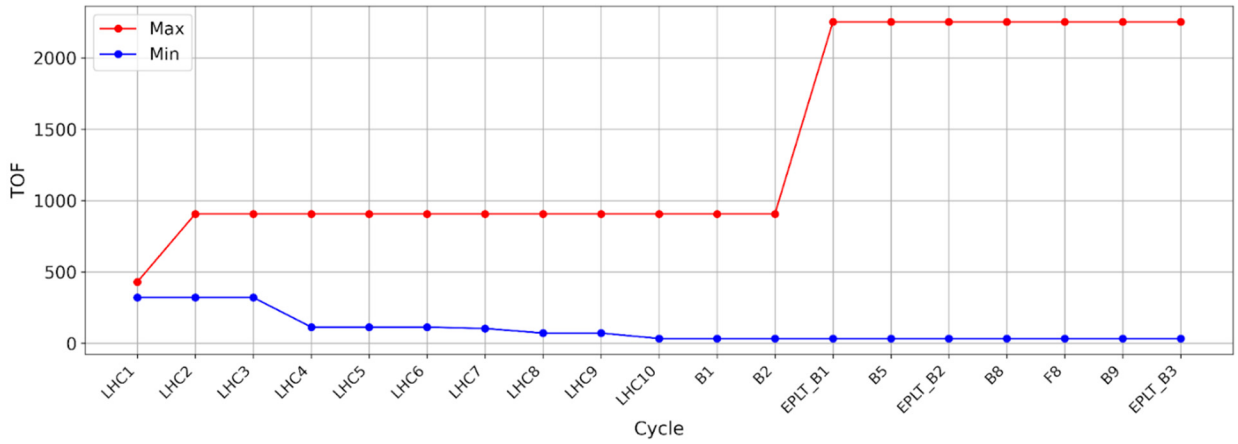

Figure S36| TOF evolution for L12. *Cycle-wise evolution of the minimum and maximum TOF for ligand 12.*

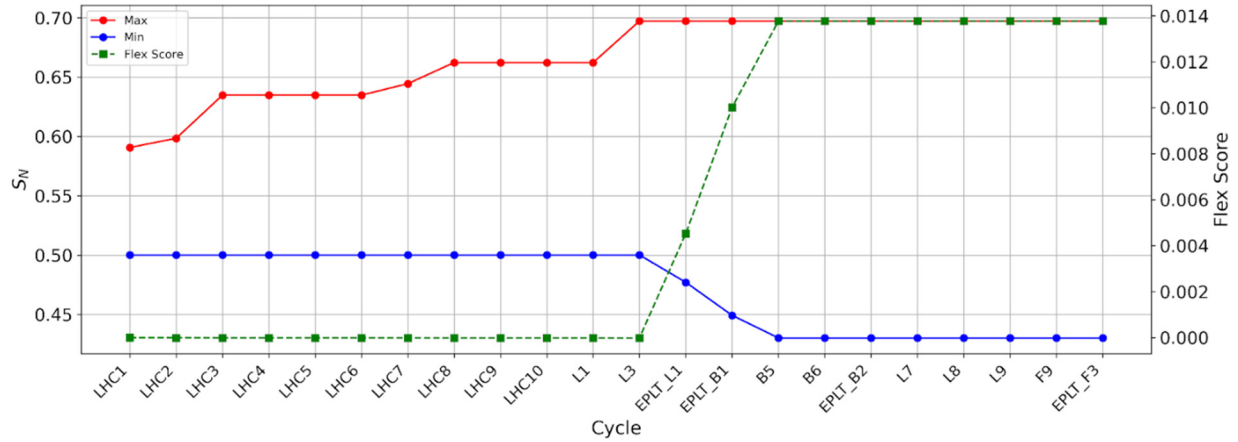

Figure S37| Regioselectivity and flexibility evolution for L13. *Cycle-wise evolution of the minimum and maximum regioselectivity and flexibility score for ligand 13.*

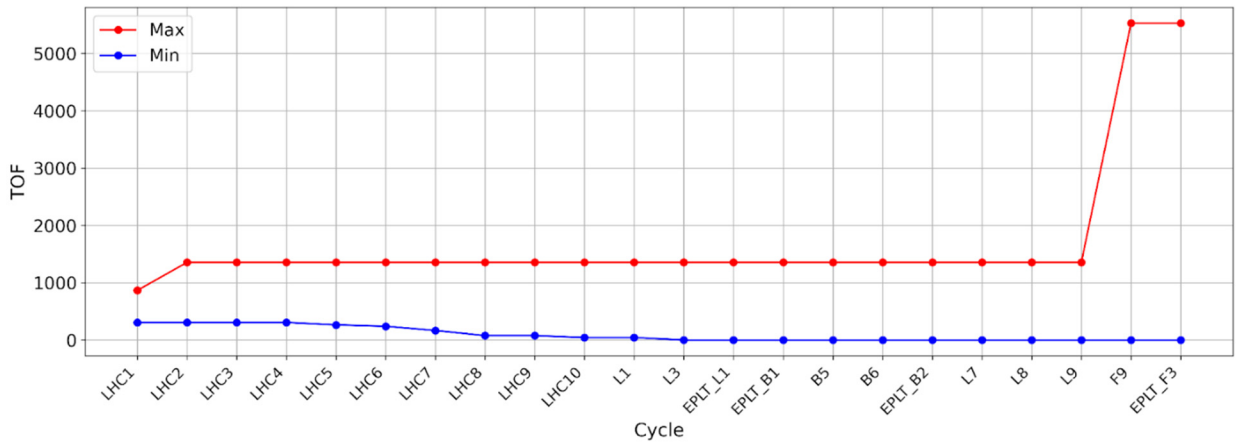

Figure S38| TOF evolution for L13. *Cycle-wise evolution of the minimum and maximum TOF for ligand 13.*

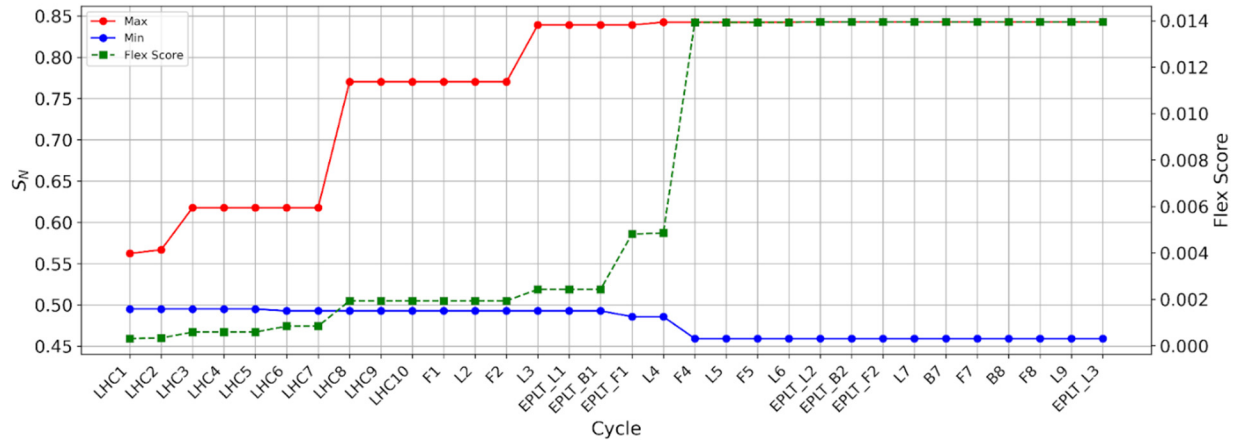

Figure S39| Regioselectivity and flexibility evolution for L14. *Cycle-wise evolution of the minimum and maximum regioselectivity and flexibility score for ligand 14.*

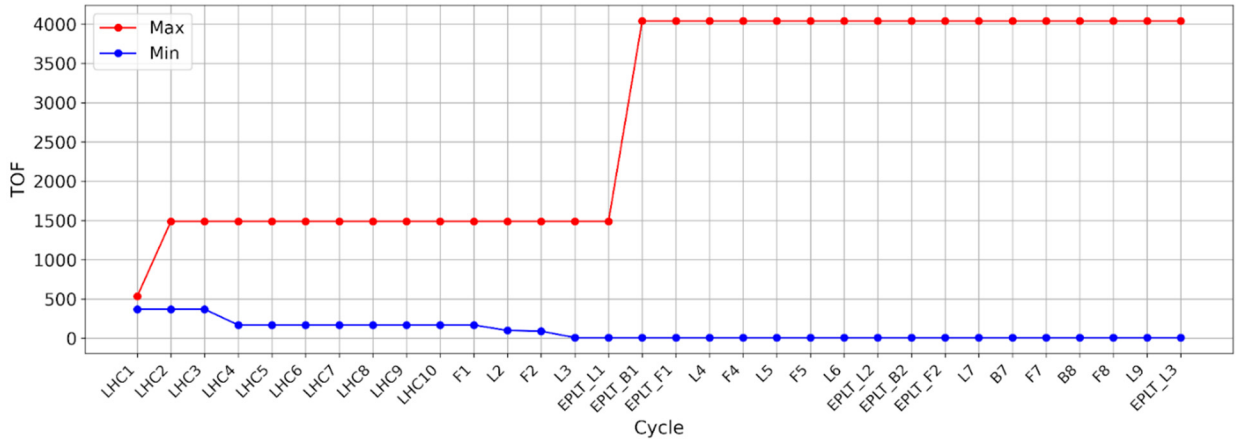

Figure S40| TOF evolution for L14. *Cycle-wise evolution of the minimum and maximum TOF for ligand 14.*

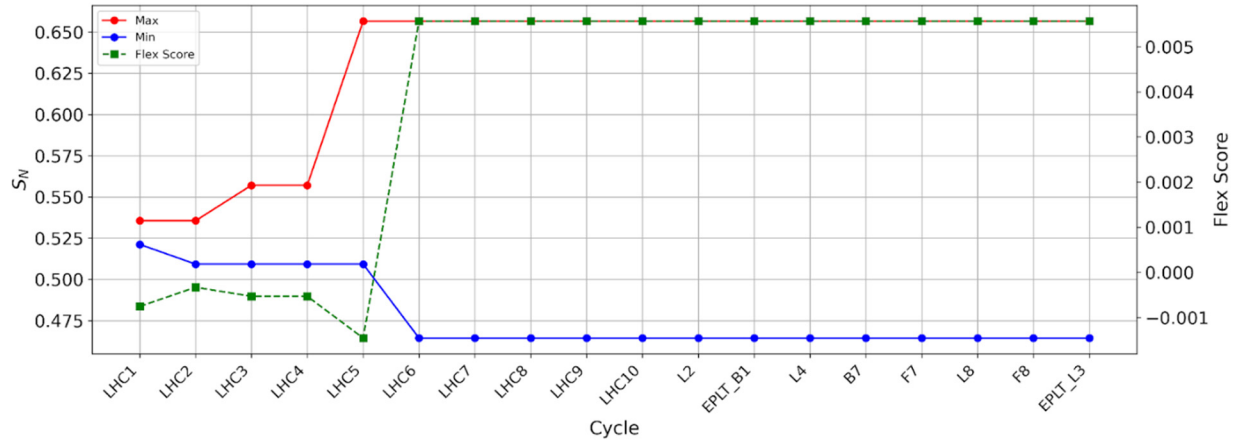

Figure S41| Regioselectivity and flexibility evolution for L15. *Cycle-wise evolution of the minimum and maximum regioselectivity and flexibility score for ligand 15.*

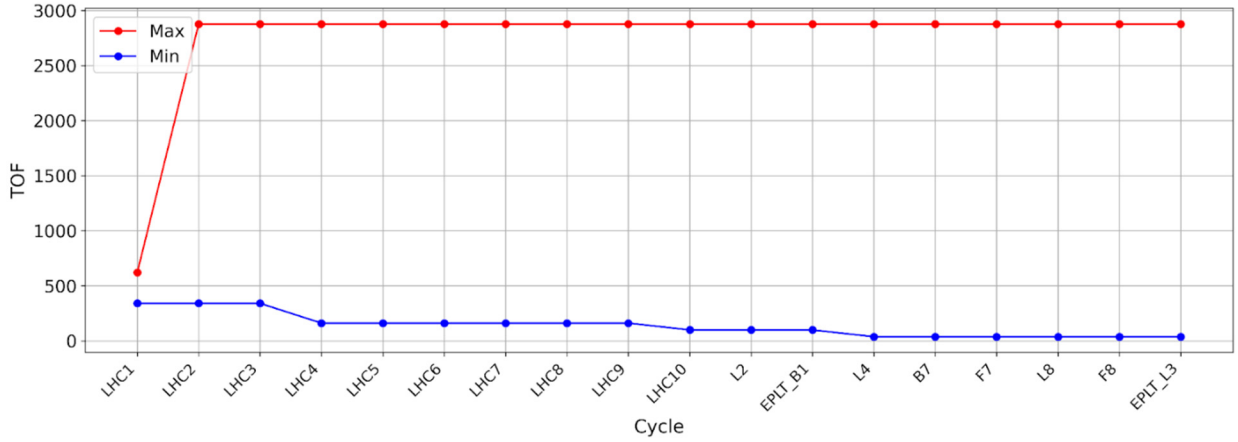

Figure S42| TOF evolution for L15. *Cycle-wise evolution of the minimum and maximum TOF for ligand 15.*

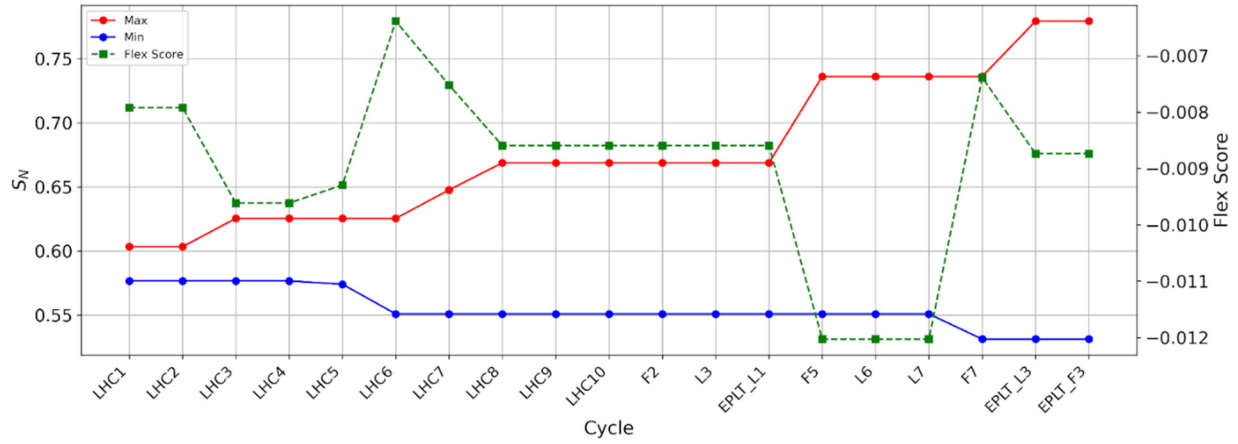

Figure S43| Regioselectivity and flexibility evolution for L16. *Cycle-wise evolution of the minimum and maximum regioselectivity and flexibility score for ligand 16.*

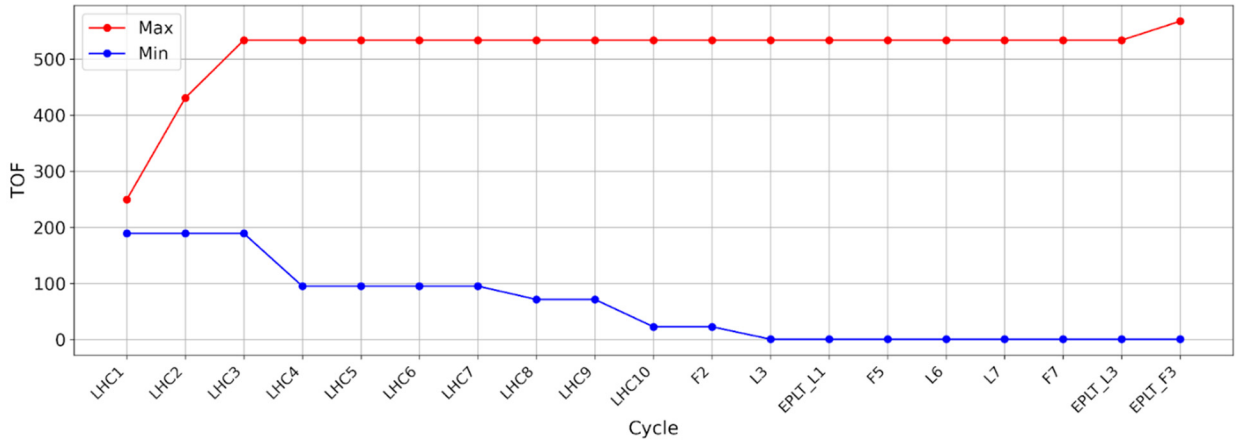

Figure S44| TOF evolution for L16. *Cycle-wise evolution of the minimum and maximum TOF for ligand 16.*

6. Reaction Parameters Evolution

The selection trends for reaction parameters (T, P, y_{CO}, y_{Mix}) across BO cycles are shown in **Fig. S45**. All parameter values are normalized for comparison. The temperature ranges from 75°C to 110 °C, and the pressure ranges from 150 to 300 psig. During the initialization phase (LHS cycles), no specific trend is expected since LHS sampling aims to evenly explore the entire parameter space. In contrast, distinct patterns emerge during optimization. In linear cycles, sampling initially covers diverse regions of the parameter space but ultimately converges toward the lower range of values. In branched cycles, the model rapidly identifies that higher parameter values favor branched regioselectivity, leading to consistent sampling in the upper range. Flexibility cycles exhibit a different behavior: the model alternates between high and low parameter values as it seeks to push each ligand toward both linear and branched regioselectivity.

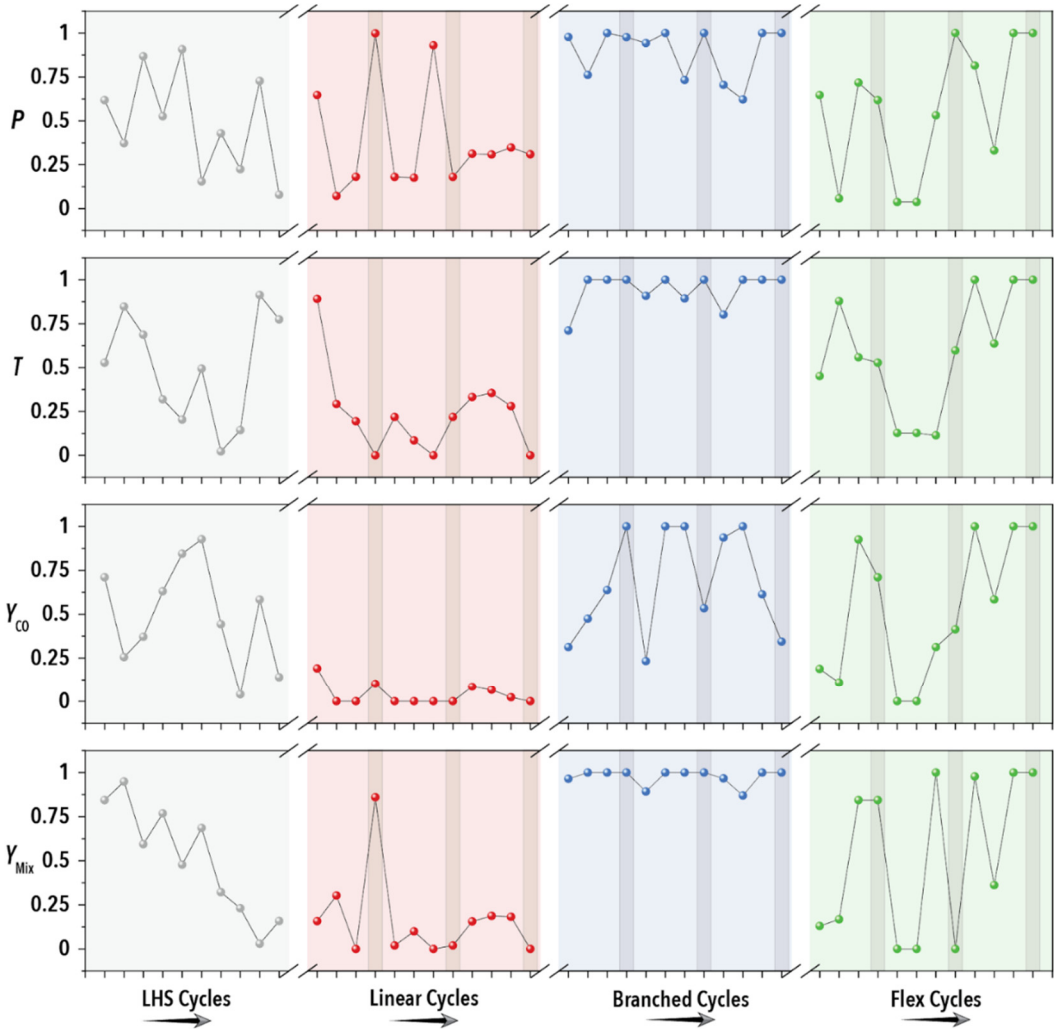

Figure S45| Cycle-wise global reaction parameters evolution. *Normalized values of selected global reaction parameters (temperature, pressure, gas composition) across optimization cycles. Linear campaigns primarily sample near the lower bounds of the parameter space, while branched campaigns favor upper-bound conditions. Flex cycles cover a wider span, reflecting adaptive exploration of the continuous parameter space. Each optimization campaign highlights EPLT cycles, which occur after every three EI-based optimization cycles.*

7. Reaction Parameters: Pareto-front

The effects of different reaction parameters (total pressure, partial pressures, H_2:CO ratio, Rh concentration, L:Rh ratio, and temperature) on the composite Pareto-front are depicted in **Fig. S46-47**. Plotting each parameter on the TOF *vs*. S_N pareto-front as a color map allows for rapid identification of the correlations between input parameters and the output space. For example, total pressure and propylene partial pressure are correlated and show similar influence with high values trending to high TOF and low S_N and low values trending towards lower TOF and higher S_N. Colormaps for p_{CO} and p_{H2} and H_2:CO show the effective elevated hydrogen directing towards linear aldehydes and elevated CO directing towards branched product with intermediate H_2:CO having the strongest TOF with excess CO beginning to interfere with high activity. Rh and L:Rh plots show the interplay of catalyst loading and ligand to rhodium ratio on reaction performance. High L:Rh ratio tends towards linear aldehyde regioselectivity albeit with significant variation in absolute regioselectivity between ligand species while catalyst loading tends to achieve higher TOF with lower absolute rhodium concentrations. Increasing temperature increases the TOF and causes regioselectivity to trend toward the branched product.

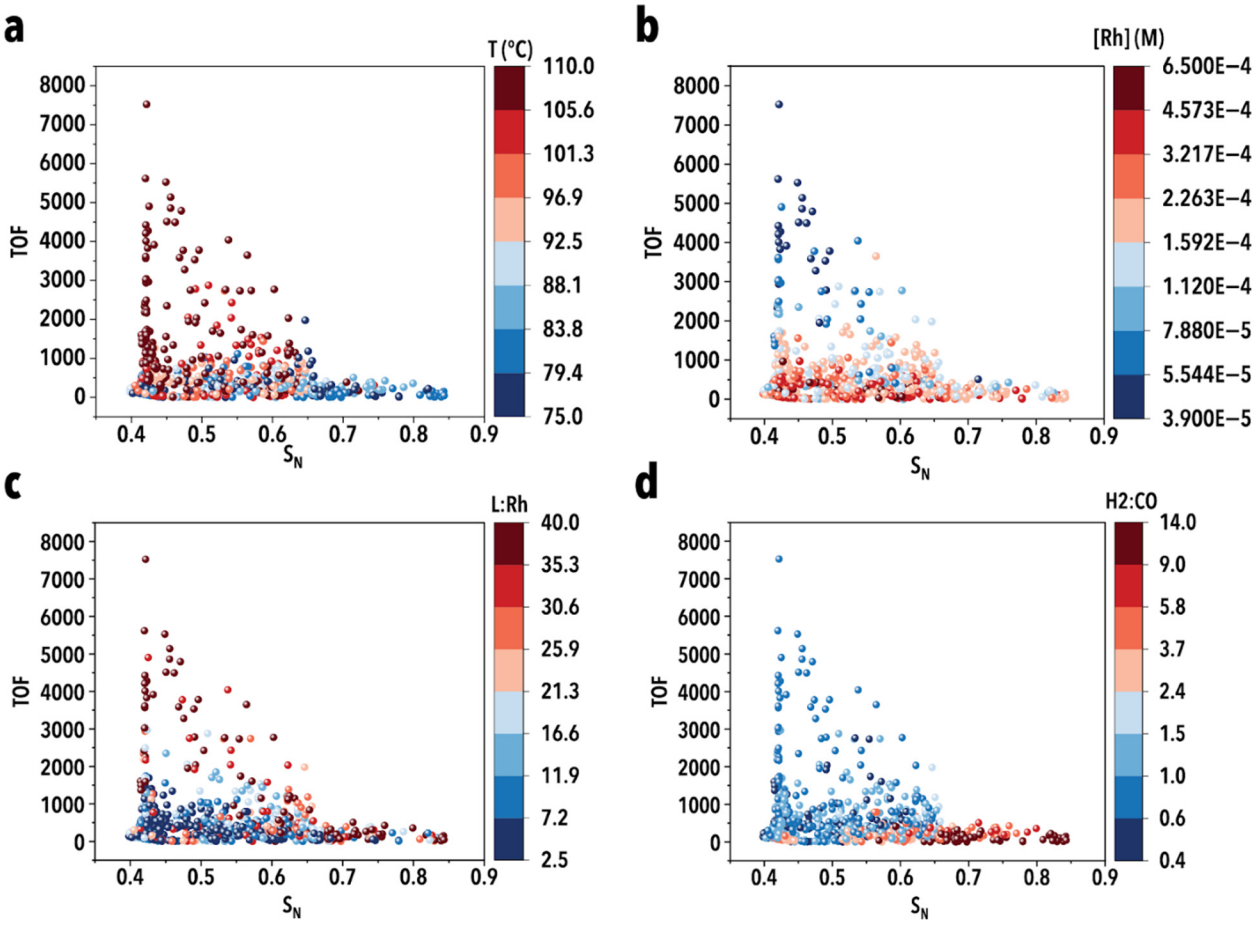

Figure S46| Pareto front color-coded by reaction parameters. a-d, *Final composite Pareto front combining all BO campaigns and initialization data, with data points color-coded by temperature, Rh concentration, ligand-to-Rh ratio, and H_2-to-CO ratio.*

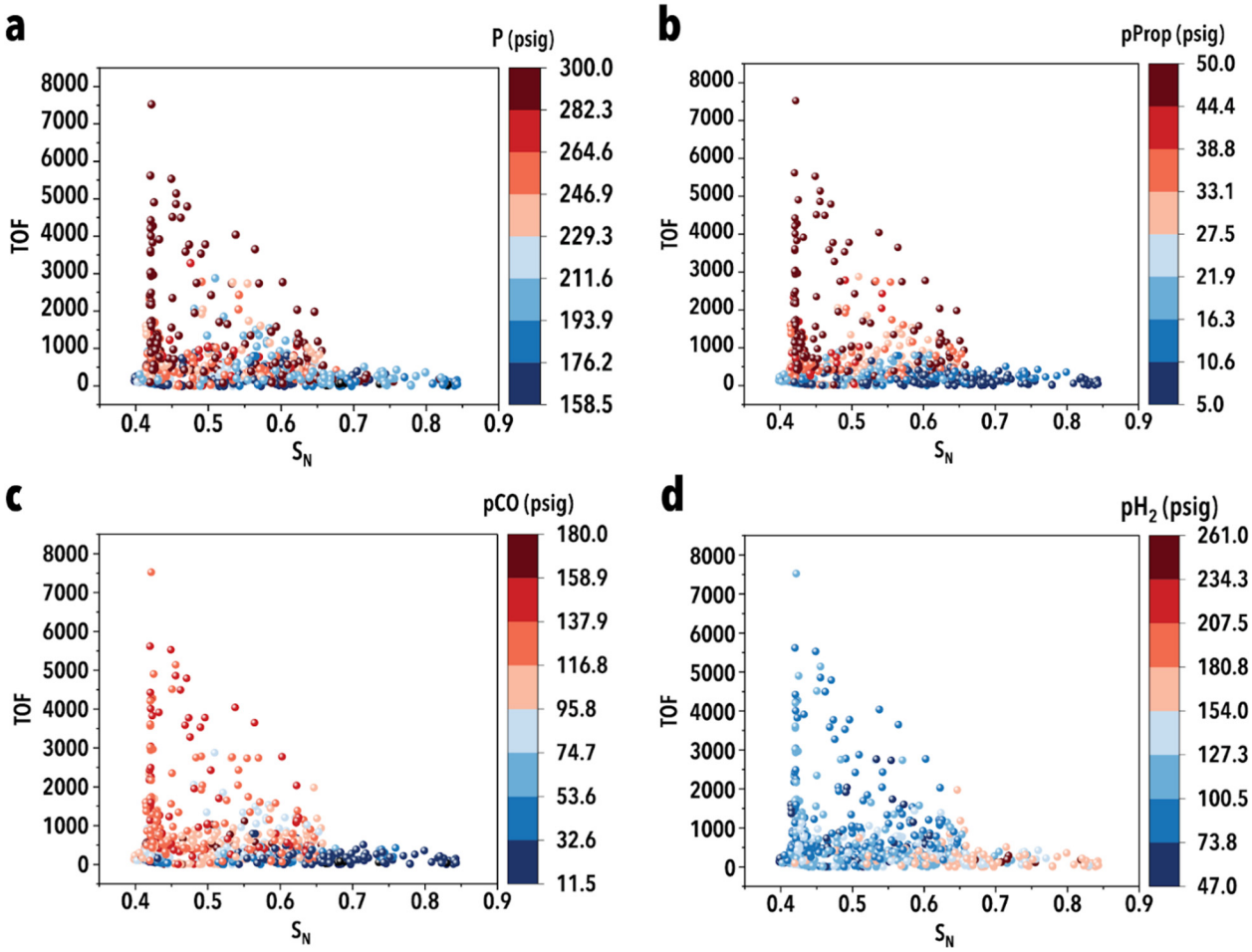

Figure S47| Pareto front color-coded by reaction parameters. a-d, *Final composite Pareto front combining all BO campaigns and initialization data, with data points color-coded by total pressure, propylene partial pressure, CO partial pressure, and H_2 partial pressure.*

8. Ligand Parameterization

To identify a pertinent steric descriptor for the ligand library, we generated scatter plots of S_N *vs.* the DFT calculated descriptors found in the *kraken* database[4] for all 20 conditions tested in the LHS data. Buried volume-based parameters consistently were among the descriptors with the highest trendline R^2 values (**Table S1**). The steric parameter *vbur_ovbur_max_min* (OvBur) was consistently identified among the five descriptors most linearly correlated with S_N within each LHS condition. This parameter quantifies the maximum steric bulk of the octant of a sphere of radius 3.5 Å for the conformation which minimizes this parameter. Ligand L1 was not in the *kraken* database, so we computed the buried volume-based parameters based on the optimized structure using the DFT methods described below. OvBur values used for this analysis are found in **Table S2**.

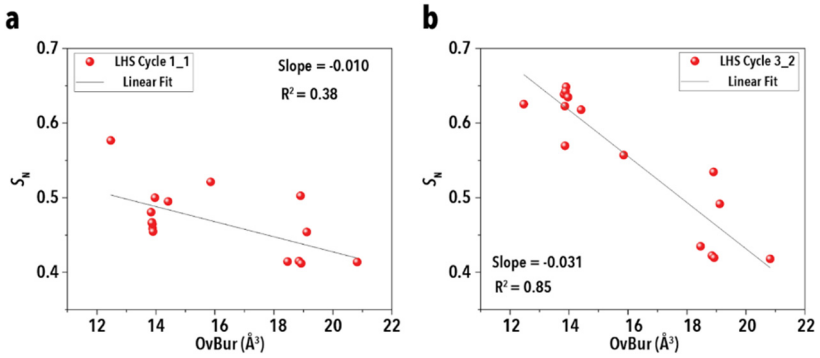

Figure S48| Correlation of regioselectivity with a steric descriptor. *Representative plots of S_N vs. OvBur (octant buried volume, Å³) from the kraken database, shown for selected LHS conditions along with a linear trendline. Panel* ***a*** *shows the LHS cycle with the lowest R^2 (LHS Cycle 1_1)and panel* ***b*** *shows the LHS cycle with the highest R^2 (LHS Cycle 3_2).*

Table S1 | Additional kraken descriptors correlated with S_N. *List of parameters ranked among the top five descriptors across 20 LHS conditions, with counts indicating their frequency of occurrence.*

Descriptor	Count
vbur_ovbur_max_min	20
vbur_qvbur_max_min	18
vbur_ovbur_max_vburminconf	17
vbur_near_vbur_boltzmann_average	17
vbur_vbur_boltzmann_average	13
vbur_near_vbur_vburminconf	7
vbur_near_vbur_min	5
vbur_vbur_vburminconf	1
pyr_alpha_vburminconf	1
vbur_qvbur_max_vburminconf	1

Of the electronic descriptors available in *kraken*, none showed a convincing linear correlation for within a single LHS condition nor across all reaction conditions explored. We instead used calculated Tolman electronic parameters as intuitive electronic descriptors for these ligands (**Table S2**)[5]. All quantum chemical calculations were performed using Gaussian 16[6]. Initial molecular geometries were pre-optimized using the MM2 molecular mechanics force field. Geometry optimizations were carried out in the gas phase using the B3LYP hybrid density functional[6,7] with Grimme's D3 empirical dispersion correction including Becke-Johnson damping (GD3BJ). The 6-

31G(d,p)[8] basis set was employed for all light atoms (C, H, N, O, F, P), while rhodium was described using the Stuttgart RSC 1997 effective core potential and associated valence basis set with a 28-electron core. The integration grid was set to "SuperFine". Frequency calculations were performed to confirm stationary points and obtain thermochemical corrections. Gibbs free energies were calculated at 298.15 K and 1 atm. For systems with multiple conformers, the lowest energy conformer was identified and used for analysis. The electronic properties of the phosphine ligands were evaluated using L-RhH(CO)$_3$ model complexes, where the phosphine and hydride ligands occupy trans axial positions in a trigonal bipyramidal geometry with three equatorial CO ligands. The A_1 symmetric CO stretching frequency of the three equatorial carbonyls was used as an electronic descriptor for each phosphine ligand.

Table S2 | Steric and electronic descriptors for ligands. *List of ligands and associated steric (OvBur) and electronic (υ_{CO}) parameters*

Ligand	**OvBur (Å^3)**	**υ_{CO} (cm^{-1})**
L1	16.86	2134.0
L2	18.46	2114.6
L3	13.88	2120.0
L4	19.11	2093.1
L5	13.90	2117.3
L6	20.82	2109.3
L7	13.84	2117.0
L8	13.88	2118.2
L9	18.85	2116.0
L10	13.86	2132.0
L11	13.86	2125.9
L12	18.92	2127.5
L13	13.96	2121.9
L14	14.41	2130.1
L15	15.85	2124.4
L16	12.47	2127.3

The calculated descriptors are used to color a Pareto front of the tested experimental conditions to investigate the regions available to each ligand. High steric parameters (OvBur) tended towards more branched conditions (S_N = 0.42) with the lower OvBur being more spread out across the linear region. For electronic parameters, high ν_{CO} tended towards linear, intermediate ν_{CO} tended towards branched conditions while excessively low ν_{CO} strongly reduced activity (**Fig. S49**).

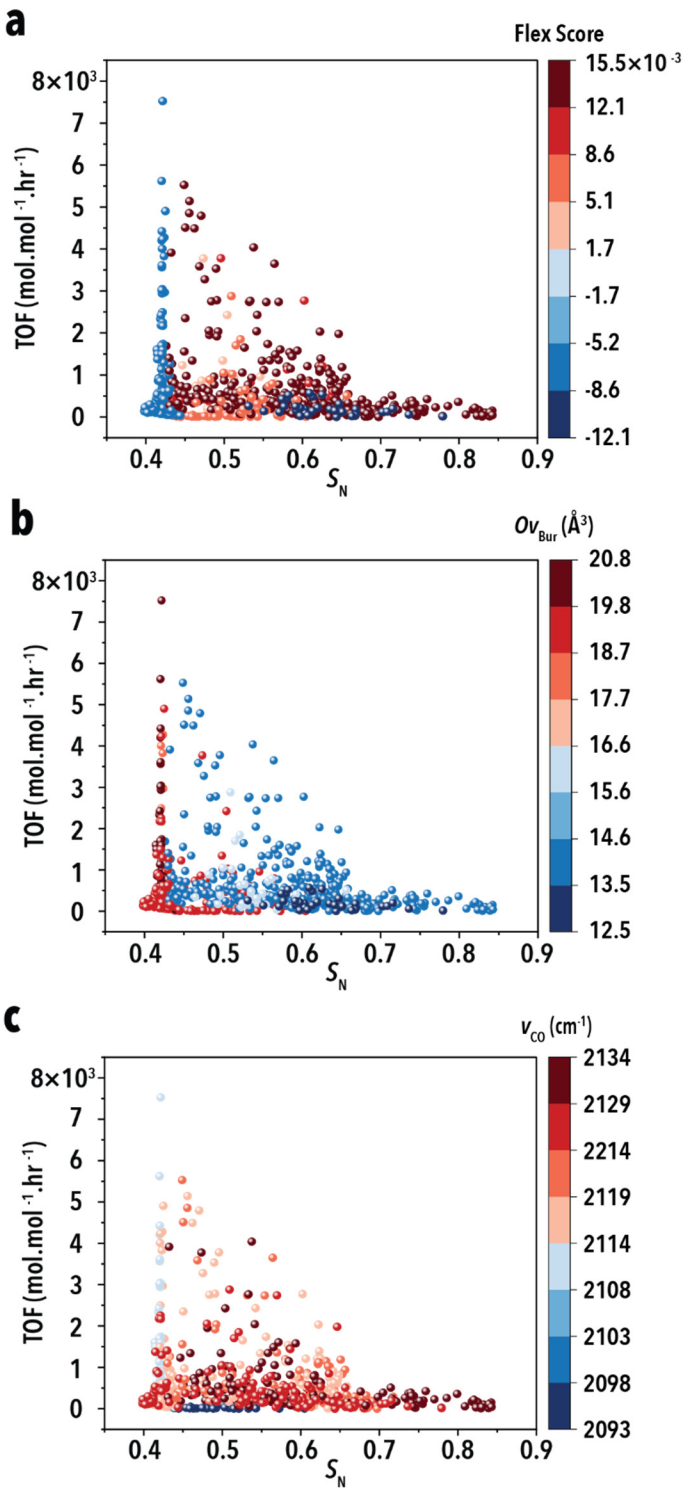

Figure S49| Pareto Colormaps. *Composite Pareto fronts, color-coded by flexibility scores (a), steric (b) and electronic (c) descriptors, illustrating the impact of each descriptor on regioselectivity and TOF.*

9. Champion Condition Prediction

After the BO campaigns, the model was queried on the same reaction conditions identified for scale-up, namely maximum global S_N, minimum global S_N, and Max/Min S_N of the ligand with the highest flex score excluding any condition with TOF under 300 h^{-1} (**Table S3**). Each reaction condition was run for all 16 ligands at 1 mL scale in the MTP reactor block. The S_N parity plot has an R^2=0.87, and the TOF parity plot has an R^2=0.93 (**Fig. S50**).

Table S3 *| Champion conditions. Linear, Max S_N > 300 TOF. Branched, Min S_N > 300 TOF. Flex, Max and Min S_N > 300 TOF for the ligand with the highest flex score. 2h Reaction time, 1mL GC vial with slit septa.*

Cycle	Temperature (°C)	p_{CO} (psi)	p_{H2} (psi)	$p_{Propylene}$ (psi)	[Rh] (mM)	L:Rh
Linear	84.7	26.5	163.9	11.6	0.14	35.7
Branched	97.3	101.7	81.7	16.3	0.11	23.0
Flex_L	75	13.1	176.7	6.5	0.12	39.3
Flex_B	109.4	137.0	111.5	48.9	0.16	2.5

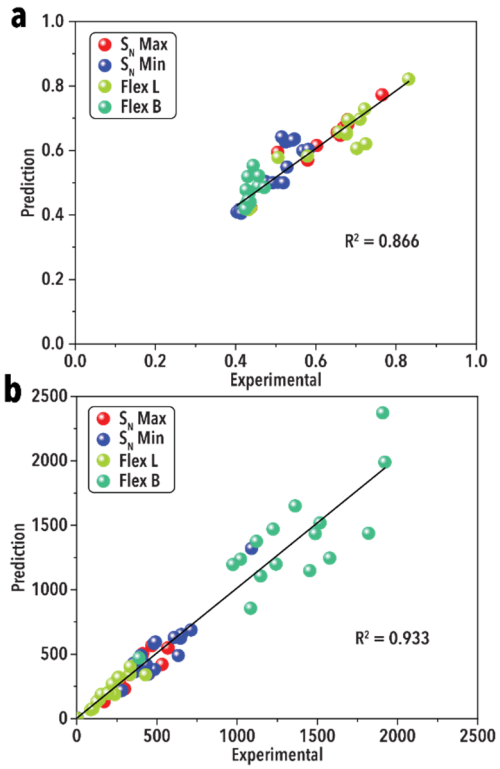

Figure S50| Champion Condition Model Parity. *Parity plots of model prediction vs. experimental result for the 4 scale-up conditions with all 16 ligands per condition. Linear, Max S_N > 300 TOF. Branched, Min S_N > 300 TOF. Flex, Max and Min S_N > 300 TOF for the ligand with the highest flex score, 2h Reaction time, 1mL GC vial with slit septa.*

10. Supplementary References

1 Bennett, J. A. *et al.* Autonomous reaction Pareto-front mapping with a self-driving catalysis laboratory. *Nature Chemical Engineering* **1**, 240-250 (2024). https://doi.org:10.1038/s44286-024-00033-5

2 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. *Journal of Catalysis* **409**, 105-117 (2022). https://doi.org:https://doi.org/10.1016/j.jcat.2022.03.030

3 Orouji, N., Bennett, J. A., Sadeghi, S. & Abolhasani, M. Digital Pareto-front mapping of homogeneous catalytic reactions. *Reaction Chemistry & Engineering* **9**, 787-794 (2024). https://doi.org:10.1039/D3RE00673E

4 Gensch, T. *et al.* A Comprehensive Discovery Platform for Organophosphorus Ligands for Catalysis. *Journal of the American Chemical Society* **144**, 1205-1217 (2022). https://doi.org:10.1021/jacs.1c09718

5 Tolman, C. A. Steric effects of phosphorus ligands in organometallic chemistry and homogeneous catalysis. *Chemical Reviews* **77**, 313-348 (1977). https://doi.org:10.1021/cr60307a002

6 Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **98**, 5648-5652 (1993). https://doi.org:10.1063/1.464913

7 Stephens, P. J., Devlin, F. J., Chabalowski, C. F. & Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *The Journal of Physical Chemistry* **98**, 11623-11627 (1994). https://doi.org:10.1021/j100096a001

8 Hehre, W. J., Ditchfield, R. & Pople, J. A. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *The Journal of Chemical Physics* **56**, 2257-2261 (1972). https://doi.org:10.1063/1.1677527