Article

Identifying general reaction conditions by bandit optimization

https://doi.org/10.1038/s41586-024-07021-y

Received: 31 July 2023

Accepted: 3 January 2024

Published online: 28 February 2024

Check for updates

Jason Y. Wang (王亿珩)[1,2], Jason M. Stevens[3], Stavros K. Kariofillis[1,2,8,12], Mai-Jan Tom[2,12], Dung L. Golden[3,12], Jun Li[4], Jose E. Tabora[4], Marvin Parasram[1,9], Benjamin J. Shields[1,10], David N. Primer[3,11], Bo Hao[5], David Del Valle[4], Stacey DiSomma[4], Ariel Furman[4], G. Greg Zipp[6], Sergey Melnikov[7], James Paulson[4] & Abigail G. Doyle[1,2]✉

Reaction conditions that are generally applicable to a wide variety of substrates are highly desired, especially in the pharmaceutical and chemical industries[1–6]. Although many approaches are available to evaluate the general applicability of developed conditions, a universal approach to efficiently discover these conditions during optimizations is rare. Here we report the design, implementation and application of reinforcement learning bandit optimization models[7–10] to identify generally applicable conditions by efficient condition sampling and evaluation of experimental feedback. Performance benchmarking on existing datasets statistically showed high accuracies for identifying general conditions, with up to 31% improvement over baselines that mimic state-of-the-art optimization approaches. A palladium-catalysed imidazole C–H arylation reaction, an aniline amide coupling reaction and a phenol alkylation reaction were investigated experimentally to evaluate use cases and functionalities of the bandit optimization model in practice. In all three cases, the reaction conditions that were most generally applicable yet not well studied for the respective reaction were identified after surveying less than 15% of the expert-designed reaction space.

Chemists have long sought robust synthetic methods that can be applied to a wide variety of substrates[11–13]. However, these methods are generally developed and optimized with only one or a few model substrates. These 'optimized' conditions are subsequently applied to a substrate scope, usually with higher yielding substrates preferentially reported. However, optimal reaction conditions for one substrate are not guaranteed to be applicable to other molecules. Despite the increased efficiency of reaction optimization enabled by automated reaction systems[14–20] and optimization algorithms[21–30], this phenomenon still substantially hampers the adoption of newly developed methodologies in synthetic chemistry[31,32]. Further optimization for different target substrates is typically required, and pharmaceutically relevant molecules with high structural complexity might not be compatible with the existing conditions at all[33]. Most work so far has focused on retroactively evaluating the general applicability of developed methodologies using substrate scope design or additive screening[34–37].

Nevertheless, post hoc analyses of applicability do not change the reaction conditions derived from antecedent optimization. De novo optimization processes that can directly yield generally applicable conditions are highly sought. Recent advances in asymmetric catalysis have started to address this problem, in which chiral catalysts that enable highly stereoselective transformations for a broad range of substrates were identified through multi-substrate screening[1–4]. However, despite advances in high-throughput experimentation (HTE), exhaustive examination of high-dimensional reaction conditions for a sizable scope of diverse substrates remains analytically difficult and experimentally expensive to carry out. Judicious selection of experiments is, therefore, imperative to efficiently explore a reaction space during optimization[38]. A notable recent example from Burke, Aspuru-Guzik and Grzybowski aimed to find more general sets of conditions for a Suzuki–Miyaura cross-coupling reaction with aryl halides and aryl *N*-methyliminodiacetic acid (MIDA) boronates[5] using Bayesian optimization. After the initial benchmarking and downselection of reaction conditions before optimization, exploration of more than 50% of the reaction space identified conditions more general than a previously published standard condition. This important advance notwithstanding, a universal reaction optimization model targeting general applicability, especially one with an efficient experiment selection strategy that can also be easily incorporated into the workflow of bench chemists, has not yet been realized.

In this study, we show that reinforcement learning models can effectively guide chemists to the most generally applicable conditions for a given substrate scope without previous experimental data on the reaction system. We designed a discrete optimization framework with experiment selection strategies that target condition generality, as quantified by average reactivity (albeit other distribution metrics can

[1]Department of Chemistry, Princeton University, Princeton, NJ, USA. [2]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA, USA. [3]Chemical Process Development, Bristol Myers Squibb, Summit, NJ, USA. [4]Chemical Process Development, Bristol Myers Squibb, New Brunswick, NJ, USA. [5]Janssen Research and Development, Spring House, PA, USA. [6]Discovery Synthesis, Bristol Myers Squibb, Princeton, NJ, USA. [7]Spectrix Analytical Services, North Haven, CT, USA. [8]Present address: Department of Chemistry, Columbia University, New York, NY, USA. [9]Present address: Department of Chemistry, New York University, New York, NY, USA. [10]Present address: Molecular Structure and Design, Bristol Myers Squibb, Cambridge, MA, USA. [11]Present address: Loxo Oncology at Lilly, Louisville, CO, USA. [12]These authors contributed equally: Stavros K. Kariofillis, Mai-Jan Tom, Dung L. Golden. ✉e-mail: agdoyle@chem.ucla.edu

be used). Through performance benchmarking on four existing reaction datasets, we demonstrate that the implemented reinforcement learning model and its underlying algorithms reach high accuracies for identifying optimal general conditions in all cases, while being adaptable, scalable and data efficient. To further substantiate the optimization framework, we validated the learning model on three unseen chemical transformations.

Model design and development

The multi-armed bandit problem[7–10] is a reinforcement learning problem that resembles many characteristics of the generality optimization problem in chemistry. In the classic stochastic formulation, a casino player is presented with a series of slot machines, each with a fixed but different reward distribution that is initially unknown. With a limited budget, the objective of the player is to maximize overall winnings by recognizing and playing the slot machine with better payouts. To do so, the player efficiently allocates the limited resources to balance the exploration of rarely played machines and the exploitation of current best options. In a reaction optimization campaign, chemists need to choose from many options for reaction conditions to maximize certain objectives with limited initial knowledge of how they will perform on a wide range of substrates (Fig. 1a). Finite experimental resources must be efficiently allocated to each reaction condition in consideration of a similar exploration–exploitation tradeoff: current best conditions derived from empirical knowledge are usually exploited, whereas new conditions are explored in hopes of discovering previously unknown and more effective methods. The similar characteristics of both problems prompted us to adapt solutions to the multi-armed bandit problem (often called bandit optimization algorithms) for generality optimization in chemistry.

The multi-armed bandit problem has been previously studied in chemistry contexts for autonomous drug design and reaction condition discovery[39,40]. In the latter case, an information-directed adaptive sampling algorithm was designed to sample conditions for a single reaction to maximize information gains and reaction yields[41]. Whereas condition arms are dropped in this example after they are sampled once for each reaction, we hypothesized that repeated sampling of distribution of each condition arm over a substrate scope (the underlying population for each arm) guided by bandit algorithms would enable the prediction of condition generality across substrates, a main contrast with the previous work (Fig. 1c). Using reaction yield as an example of an optimization objective, the same substrate scope is expected to exhibit different reactivities under different conditions, resulting in unique reward distributions for each arm (Fig. 1b). The treatment of condition variables as discrete arms enable flexible interpretation of conditions. For example, arms can cover one condition dimension (for example, solvent) or many dimensions (for example, combinations of catalyst, ligand, base and solvent). Incorporating substrates into a distribution also means no explicit search space needs to be defined, and the algorithm can adjust its estimation of the distribution of each condition by continuing to sample that condition. This feature enables both the elimination of ineffective arms and the expansion of substrate scope on the fly during optimization. The latter is especially important in application, as the generality of a reaction condition is highly dependent on the scope it is applied to.

We implemented the optimization framework in Python centred around a reaction scope object that can create substrate scopes with possible conditions, interface with bandit algorithms, propose and record experimental results, predict yields for unrun reactions and recommend general conditions (Extended Data Fig. 2). We implemented numerous stochastic bandit algorithms for both binary rewards (for example, reactivity thresholds) and continuous rewards (for example, numeric reaction yields). Effective algorithm classes were identified through extensive benchmarking with synthetic data

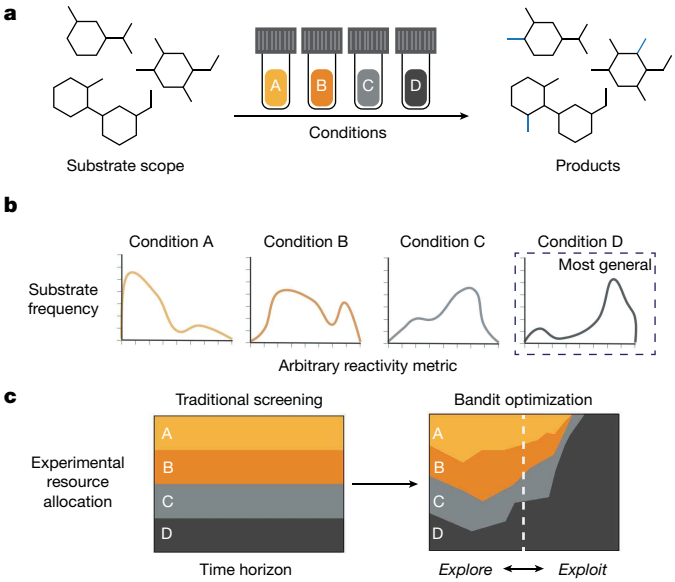


Fig. 1 | Optimization of the most general conditions with bandit algorithms. a, Illustration of a generality optimization problem, in which multiple conditions are evaluated with a substrate scope. **b**, Illustration of substrate reactivity distributions of each condition. More general conditions (for example, D) will have more substrates exhibiting higher reactivity and their distributions will be negatively skewed. The *x*-axis represents an arbitrary reactivity metric and the *y*-axis represents frequency (or count, for a histogram) of substrates. **c**, Illustration of experimental resource allocations of traditional screening approach and bandit optimization. Bandit optimization allocates more resources for the better conditions during later stages of optimization.

as well as empirical modifications and hyperparameter selections that are beneficial to algorithm performance. The Bayes UCB (Upper Confidence Bound) algorithm[42] with tuned parameters mostly offered the best performance, whereas the UCB1-Tuned algorithm[43] is preferred in practice because of the absence of tunable parameters and generally satisfactory performance. Multiple approaches to support batch proposing and updating were also implemented to allow parallel experimentation in practice (see Supplementary Information for details on algorithm benchmarking and development). Unlike optimization frameworks that involve costly fitting of Gaussian processes and neural networks as surrogate models[44], our framework is also lightweight and computationally efficient with minimal software dependencies. This advantage not only enhances software performance in a production environment but also enables us to extensively simulate the learning model with existing datasets to statistically evaluate its effectiveness.

Performance testing with chemistry reaction datasets

We simulated the optimization model on three previously published chemistry reaction datasets consisting of a variety of conditions applied to a broad scope of substrates: a nickel-catalysed borylation dataset previously investigated by Bristol Myers Squibb (BMS)[45], a deoxyfluorination dataset from the Doyle group[46] and a Buchwald–Hartwig C–N cross-coupling dataset[47], all with the aim of finding the most general conditions with different reactivity metrics (Fig. 2a). For every dataset, the most general conditions were first determined through analyses of reaction yield distributions (Fig. 2c; see Supplementary Information for detailed yield analyses on all datasets). Optimization runs were then simulated by iteratively allowing the bandit algorithms to propose experiments and providing the algorithms with actual experimental results. For all three reactions, we used the Bayes UCB algorithm with beta prior for binary metrics and Gaussian prior for continuous metrics (see Supplementary Information section 8 for performance

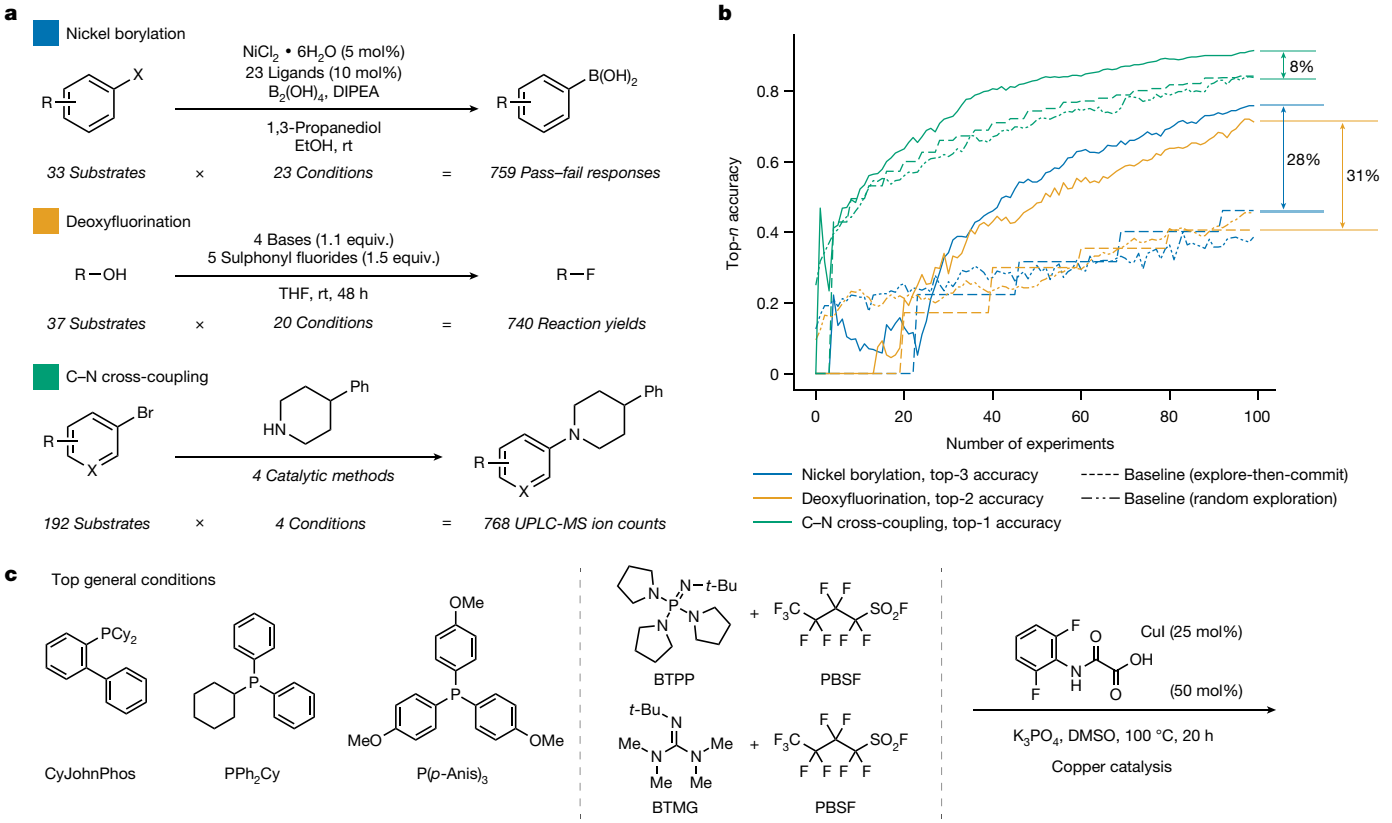


Fig. 2 | Testing the bandit optimization framework on three datasets with different objectives and condition complexities. a, General reaction schemes of the three datasets tested. Details of the reaction scope can be found in the Supplementary Information. **b**, Accuracy of identifying top-*n* optimal conditions for all three datasets tested. Accuracy at experiment *i* is defined as the relative frequency (across 500 simulations with random starts) of the algorithm correctly selecting actual top conditions as optimal condition, based on all experimental results up to experiment *i*. Details on algorithms: nickel borylation dataset was run with Bayes UCB (beta prior); deoxyfluorination dataset was run with Bayes UCB (Gaussian prior); and C–N cross-coupling dataset was run with Bayes UCB (Gaussian prior). **c**, Top general conditions for each dataset that were used as objectives during generality optimization. DIPEA, diisopropylethylamine; THF, tetrahydrofuran; UPLC-MS, ultra-high performance liquid chromatography-mass spectrometry; DMSO, dimethyl sulfoxide.

comparison with other algorithms). After each round, the learning model updated its beliefs for the reaction scope, and this process was continued until a specified number of experiments was reached. This simulation process was repeated many times (for example, 500) and the top-*n* accuracy was used as a metric to compare algorithm performances. Top-*n* accuracy was calculated as the relative frequency of the model correctly identifying the top-*n* conditions with data collected up to time point *t* across all simulations.

To confirm that meaningful learning took place with the developed model, we established baselines for comparison of each dataset. The first is a pure exploration baseline in which the conditions are randomly selected for evaluation. The other baseline strategy, explore-then-commit (ETC), tries each condition during the exploration stage and exploits by committing to the best option from exploration. To compare with other algorithms, at any given time point, the best empirical option from all previous, completed exploration rounds is identified. After a new round of exploration is complete, ETC re-evaluates and chooses a new option that appears best with the inclusion of new data, and its accuracy is also updated accordingly, yielding a stepwise accuracy baseline. The pure exploration and ETC baselines exhibit similar accuracies in practice because of the similar concept of uniform exploration, with ETC being less noisy because of the more structured exploration by round. These two baseline strategies mimic the state-of-the-art multi-substrate screening approaches, in which different combinations of substrates and conditions are evaluated, and the most general condition is chosen based on the average performance using all available data. Compared with ETC baselines, the bandit algorithms achieved substantial improvements in accuracies for all three datasets (28%, 31% and 8%) within 100 experiments (Fig. 2b). An accuracy improvement of 30% indicates that the probability of finding general conditions within a relatively low experimental budget is better when pursuing the bandit strategy compared with the baselines. For the C–N cross-coupling dataset, the ETC strategy reached high accuracy (>80%) because each round of exploration costs at most four experiments. Despite the high baseline accuracy, the highest-performing bandit algorithm still achieved an 8% improvement in accuracy. To evaluate the data efficiency of the bandit algorithms, we simulated a palladium-catalysed C–N cross-coupling reaction dataset with more than 3,600 experiments (Extended Data Fig. 1a,b)[48]. The best-performing Bayes UCB algorithm achieved more than 90% accuracy after exploring only 2% of the reaction scope (72 reactions) (Extended Data Fig. 1c). We also visualized the experiments selected by the Bayes UCB algorithm at different time points in a single optimization run (Extended Data Fig. 1d) to illustrate the general behaviour of bandit algorithms (further discussion can be found in Supplementary Information section 8.5). Taken together, these results validated that the bandit algorithms can be successfully translated to chemistry reaction data and are accurate in finding the most general conditions for various reactions, condition precisions and optimization objectives.

Optimization study 1: C–H arylation reaction

Next, we set out to evaluate the bandit algorithms on unseen data for distinct chemical transformations. A reaction dataset with many diverse

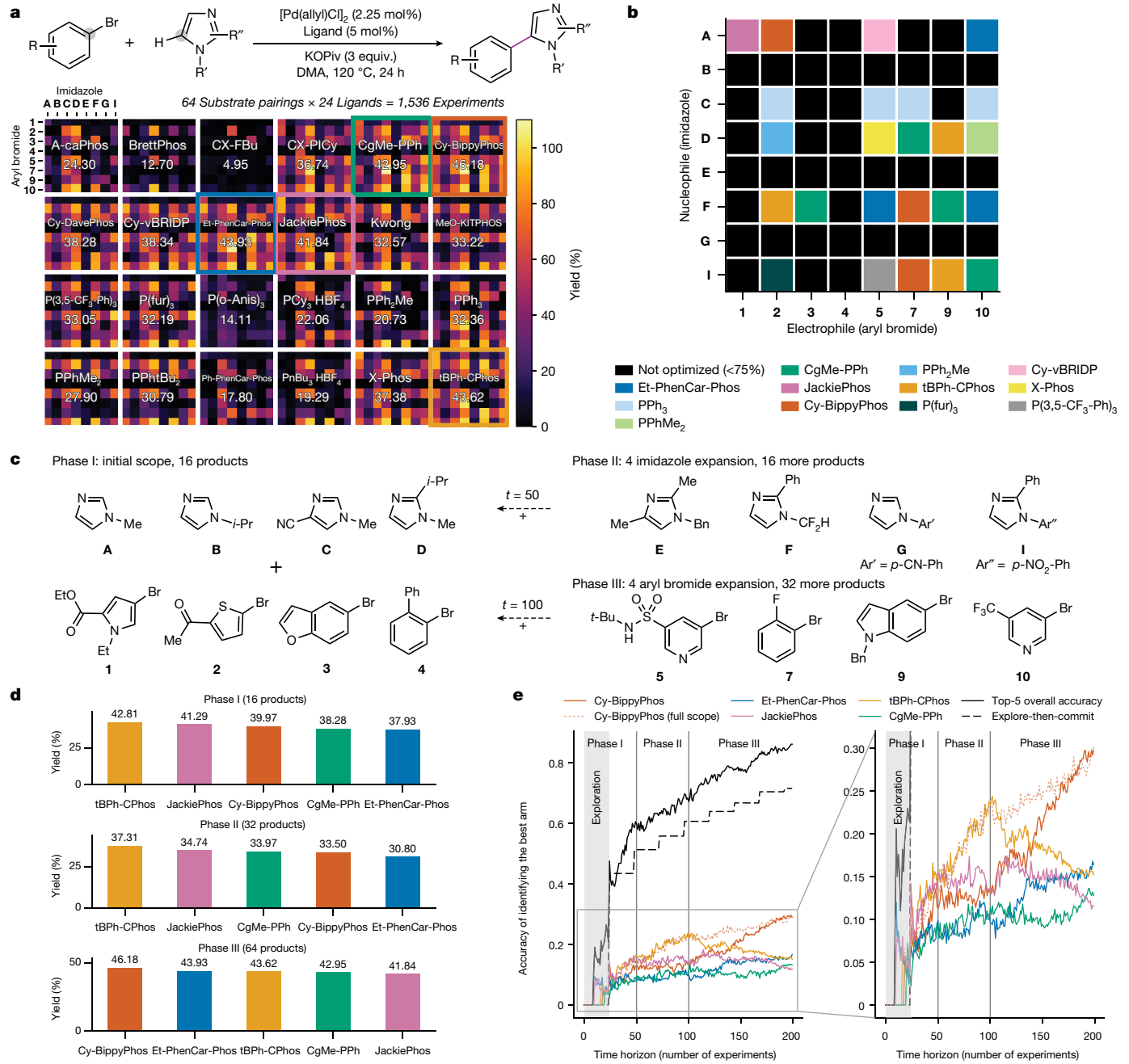


Fig. 3 | Optimization studies of a palladium-catalysed C–H arylation reaction. a, General reaction scheme and HTE results for palladium-catalysed C–H arylation of imidazoles with aryl bromides. Average yields across all 64 products for each ligand are shown in white. Structures of all 24 ligands are included in the Supplementary Information. **b**, Ligand optimization results using a model substrate approach. The ligand that gives the highest yield (that is also >75% yield) for each of the 64 products is selected as the optimal ligand. Substrate combinations are considered as not optimized if no ligand surpasses the 75% reactivity threshold. **c**, Substrate scope search space expansion scheme. Phase I (imidazoles **A, B, C** and **D** and aryl bromides **1, 2, 3** and **4**) was expanded with four more imidazoles (**E, F, G** and **I**) after 50 experiments and four more aryl bromides (**5, 7, 9** and **10**) after 100 experiments. **d**, Average yield distributions for top-5 (overall) ligands during three phases. **e**, Average accuracy of identifying each of the five most general ligands as the optimal ligand over time during different phases (UCB1-Tuned algorithm, 500 random starts). Overall top-5 accuracy (black, solid), top-5 explore-then-commit baseline accuracy (black, dashed) and the accuracy of identifying Cy-BippyPhos as optimal with full scope of data available from the start (red, dotted) are also shown. KOPiv, potassium pivalate; DMA, dimethylacetamide.

substrate pairings and calibrated reaction yields for all products under the same environment, that which is also sufficiently large for modelling, would be ideal to evaluate the performance of generality optimization algorithms in a regime in which multiple substrate dimensions simultaneously interact with conditions. Owing to the lack of these datasets in the literature, we decided to collect a palladium-catalysed imidazole direct C5-arylation dataset that satisfies these requirements. This dataset builds on a C–H arylation dataset investigated in a previous collaboration between the Doyle group and BMS[25], in which the conditions were extensively surveyed with a single pair of substrates. However, in this case, we expanded the substrate dimensions of both imidazoles and aryl bromides and specifically studied ligand effects with an expanded ligand scope. A total of 64 unique C5-arylated imidazole products were generated from eight imidazoles and eight aryl bromides, each evaluated with 24 ligands yielding 1,536 total reactions (Fig. 3a).

We first retrospectively analysed the dataset by mimicking a traditional model substrate approach, in which the ligands are screened with a model substrate (or product) to identify the highest-performing ligand as optimal. For each of the 64 products in the scope, we filtered out products (40 out of 64) that did not achieve a reaction yield above 75% (these reactions can usually be considered as 'not optimized' in practice). For the rest of the products, the highest-yielding ligand was selected (Fig. 3b). Twelve out of 24 ligands in the scope can be considered as 'optimal' with different substrate pairings. Most of these ligands, however, are non-optimal when considering all 64 products. The most notable example, PPh_3, is the optimal ligand for imidazole **C** with multiple aryl bromides, but its average yield over all products is only 32.4%, compared with 46.2% for CyBippyPhos. Moreover, our previous HTE study of C–H arylation[25], in which imidazole **C** and aryl bromide **7** were used as model substrates to evaluate 1,984 different reaction conditions including 14 monophosphine ligands, identified CgMe-PPh as the optimal ligand almost exclusively (19 out of top 20 conditions, with the only other ligand being PPh_3). These analyses highlight that a traditional screening approach with a model substrate, even after extensive exploration of the condition space, does not usually produce a satisfying general condition. By contrast, simulating the bandit model with this dataset showed an 85% top-5 accuracy (Fig. 3e, compared with the 71% explore-then-commit baseline) and a > 95% top-9 accuracy on average after 200 experiments (see Supplementary Information for detailed simulation studies of this reaction). Non-optimal ligands, such as PPh_3, are almost always excluded from consideration by the model, thus reducing bias when choosing general conditions.

A key advantage of the bandit optimization model is that no search space needs to be explicitly defined. Reactivity responses from various substrates are treated as feedback from the environment that the algorithm is learning from. This means that the substrate scope, as part of a dynamic environment, can arbitrarily change on the fly and the model can learn these changes continuously from the feedback it receives during optimization. It is common in practice to expand the substrate scope and further evaluate the use of a developed method, which can affect how generally applicable a condition is and the ability of the optimization model to select these conditions. For this problem setting, we designed a test scenario in which both the imidazole and aryl bromide scopes available to the algorithm were restricted at first and expanded during optimization. Four imidazoles (**A, B, C** and **D**) and four aryl bromides (**1, 2, 3** and **4**) constituted the initial scope, defined as phase I. After 50 experiments in phase I, the imidazole scope was expanded to include four additional imidazoles (**E, F, G** and **I**), creating 16 new potential products in phase II. After 50 experiments in phase II, the aryl bromide scope was expanded again to include four more aryl bromides (**5, 7, 9** and **10**), creating 32 new potential products in phase III (Fig. 3c). Although phases I and II experience similar rankings for the top-5 ligands, the relative order changes in phase III after the addition of four aryl bromides (Fig. 3d). During optimization simulations, the individual accuracies over time for each of the top-5 ligands were tracked and compared (Fig. 3e). The model correctly identified the initial ligand reactivity rankings in phases I and II. When the reactivity ranking was changed in phase III, the algorithm did not overcommit and successfully adjusted its belief in ligand performance by increasingly sampling Cy-BippyPhos (red) and Et-PhenCarPhos (blue), the top-2 performing ligands. The previous top ligands, tBPh-CPhos (orange) and JackiePhos (purple), were downgraded by the algorithm in phase III. We also compared the accuracy of Cy-BippyPhos under a substrate expansion regime with the accuracy of Cy-BippyPhos obtained from a separate optimization simulation in which the full substrate scope is always available for the algorithm to sample from. Although the initial accuracies understandably differed because of the different reactivity distributions in phases I and II, the end accuracies at experiment 200 are similar despite the differences in the initial sampling pools.

Optimization study 2: amide coupling reaction

Owing to the prevalence of amide bond structures in biological systems and pharmaceutical compounds, amide coupling reactions are the most commonly used reactions in medicinal and process chemistry[49]. Carboxylic acids are often preferred as inexpensive and abundant starting materials. Their chemical stability, while desirable on account of the ease of handling on scale, necessitates activation by coupling reagents, usually through in situ formation of an acid halide or anhydride. Despite the vast number of activators (>200) developed for amide coupling reactions[50], chemists often resort to a few routine reagents on the basis of their proven reliabilities[51]. However, the efficacy of these coupling reagents when applied to specific target substrates is still difficult to assess a priori, especially for the challenging coupling with weakly nucleophilic anilines. Aniline deactivation from the aromatic system, as well as accompanying steric and electronic demands from various substituents, complicates the selection of productive coupling reagents. Other aspects of reaction conditions, such as bases and solvents, can also affect reactivity.

Using the late-stage functionalization of indomethacin, a commonly prescribed nonsteroidal anti-inflammatory drug (NSAID), as an example, we sought to demonstrate the ability of the bandit model to identify generally applicable amide coupling conditions when faced with a diverse scope of aniline substrates and reaction conditions (Fig. 4a). For the defined reaction scope, we attempted to identify the most general activator–base combinations. Not expecting a notable solvent effect between the three solvents chosen (THF, MeCN and DMF), we prioritized activators and bases because they often work in tandem and generate reactive intermediates, which can affect amide coupling reactivity. We first aimed to filter out less-effective activators by setting the optimization objective to activators alone. Unlike simulation studies in which real-time feedback was immediately provided for each proposed experiment, experiments proposed in batch are necessary in practice to maximize time efficiency, resulting in a delayed feedback setting. Similar to a kriging believer[52,53] in a sequential optimization problem, our implementation of batched bandit optimization uses a separately trained random forest prediction model with existing data. Both the optimization model and the prediction model were updated when experimental feedback became available. After eight rounds of initial experiments (five experiments per round), activators were ranked by reactivity based on the beliefs of the model, and the bottom four activators (PFTU, HOTU, HATU and PyBOP) were eliminated. For the four remaining activators (DPPCl, BOP-Cl, TCFH and TFFH), the optimization objective was modified to activator–base combinations. Relevant data for the four activators retained were recycled and incorporated as knowledge of the new objective by the optimization model. After 16 additional rounds of experiments, all activator–base combinations were again ranked by projected reactivity (top nine conditions are shown in Fig. 4b). Overall, about 12% of the reaction scope were experimentally explored following the suggestions of the model.

To conclusively evaluate the resulting rankings from our model, we collected experimental results for all remaining reactions not explored during optimization and analysed true reactivity rankings for activators and activator–base combinations for comparison. The model correctly identified and ranked the top three activators during the activator selection phase. For activator–base combinations, top nine out of 10 combinations were identified, with the top four correctly ranked. Interestingly, HATU–DIPEA, one of the most commonly applied amide coupling activator–base combinations[54], was the only condition not selected in top 10 as HATU was eliminated in the initial rounds. Use of DPPCl (diphenylphosphinic chloride) with NMM or DIPEA yielded the most effective general reaction conditions, ranking number one and two, respectively. Using HATU–DIPEA as a benchmark, the average yields over three solvents (THF, MeCN and DMF) for DPPCl–NMM and DPPCl–DIPEA for each aniline substrate were also analysed (Fig. 4d).

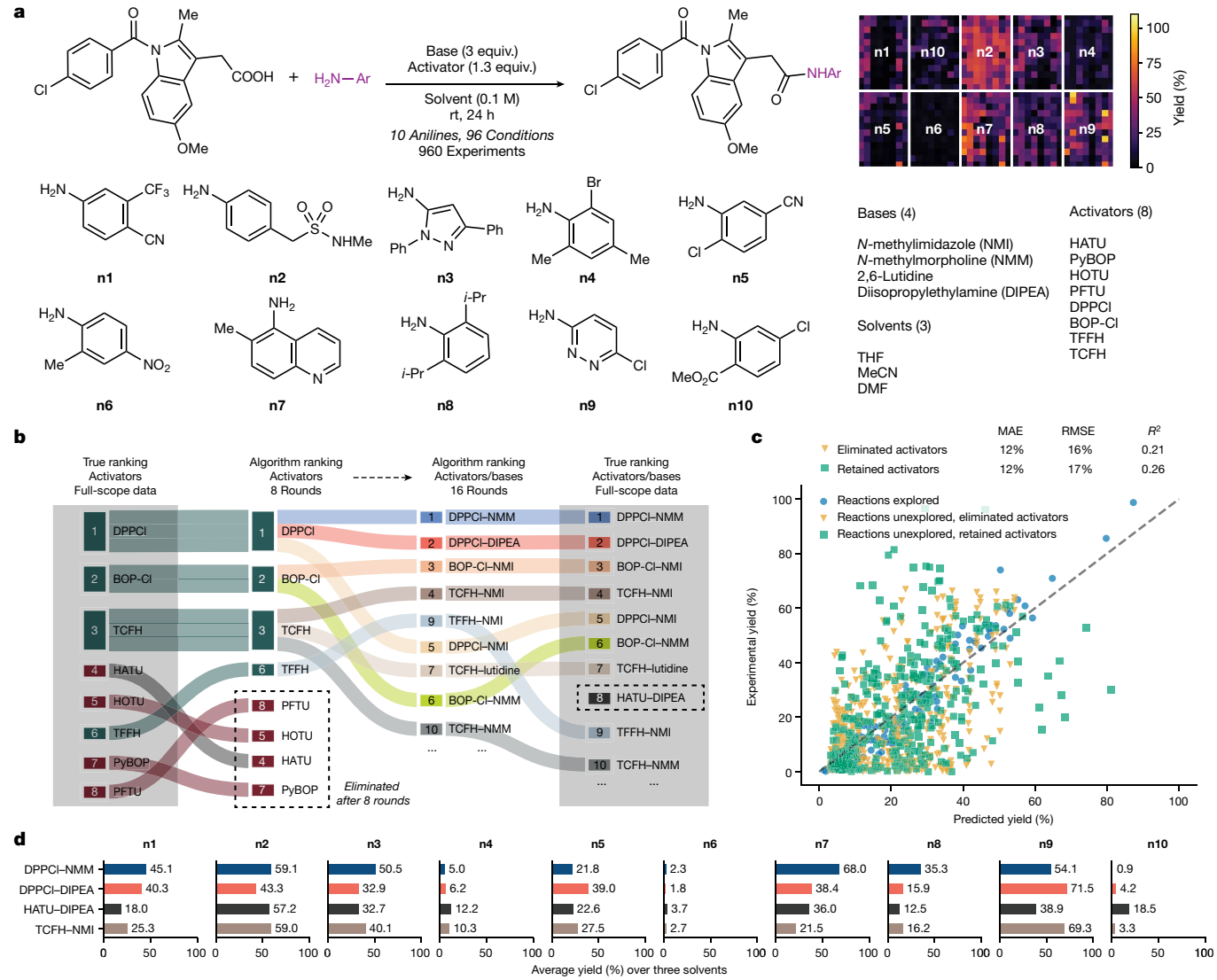


Fig. 4 | Optimization studies of an amide coupling reaction with anilines. **a**, The substrate and condition scope for the amide coupling reaction. The structures of bases and activators are included in Supplementary Information. **b**, Algorithm rankings for activators after eight rounds of experiments (five experiments per round) and algorithm rankings for activator–base combinations after 16 rounds of experiment (five experiments per round) using UCB1-Tuned as the selection algorithm. True rankings for activators and activator–base combinations from all experimental yields collected using HTE are shown in grey boxes for comparison. **c**, The performance of random forest prediction model trained with results from 24 experimental rounds. Predicted yields for the entire scope, further divided into three groups, were compared with true experimental yields. MAE, mean absolute error; RMSE, root mean square error; and R^2, coefficient of determination. **d**, Average yields over three solvents (THF, MeCN and DMF) for identified conditions of DPPCI–NMM and DPPCI–DIPEA when applied to all 10 aniline nucleophiles. HATU–DIPEA and TCFH–NMI were used as baseline comparisons.

DPPCI–NMM significantly outperformed, or at least matched, HATU–DIPEA for most anilines (except **n10**), including highly deactivated anilines (**n1**) and sterically hindered anilines (**n8**). When compared with TCFH–NMI, a reagent combination developed by BMS for challenging amide coupling reaction with non-nucleophilic amines[55], DPPCI also exhibited superior reactivities for selected anilines (for example, **n7**). Although not a commonly used amide coupling reagent, the optimization results suggest that DPPCI can be effective for amide coupling with anilines. Effective amide couplings using DPPCI have been separately investigated by BMS[56]. The desirability of DPPCI-mediated amide coupling in commercial routes, owing to its exceptional thermal stability[57] and improved atom economy compared with the mechanistically similar but much more common activator T3P, has also been demonstrated on multi-kilo scales[58].

Finally, we evaluated the accuracy of the final prediction model from the last round of optimization with measured ground truth data for the full scope. The random forest model was only trained with 12.5% of the data from the reaction scope explored during optimization but exhibits good prediction accuracy for unexplored experiments involving both activators retained and eliminated after initial experimental rounds (12% mean absolute error for both, Fig. 4c). The good accuracy of the prediction model under a low-data regime further validates the approach of using a supervised machine learning model to predict experimental results in a delayed feedback setting during optimization.

Optimization study 3: phenol alkylation reaction

The prevalence of alkyl aryl ethers in natural products and pharmaceuticals has prompted developments in mild and general syntheses of these products. Despite advances in transition-metal catalysed C–O cross-coupling reactions[59], traditional approaches, such as Williamson

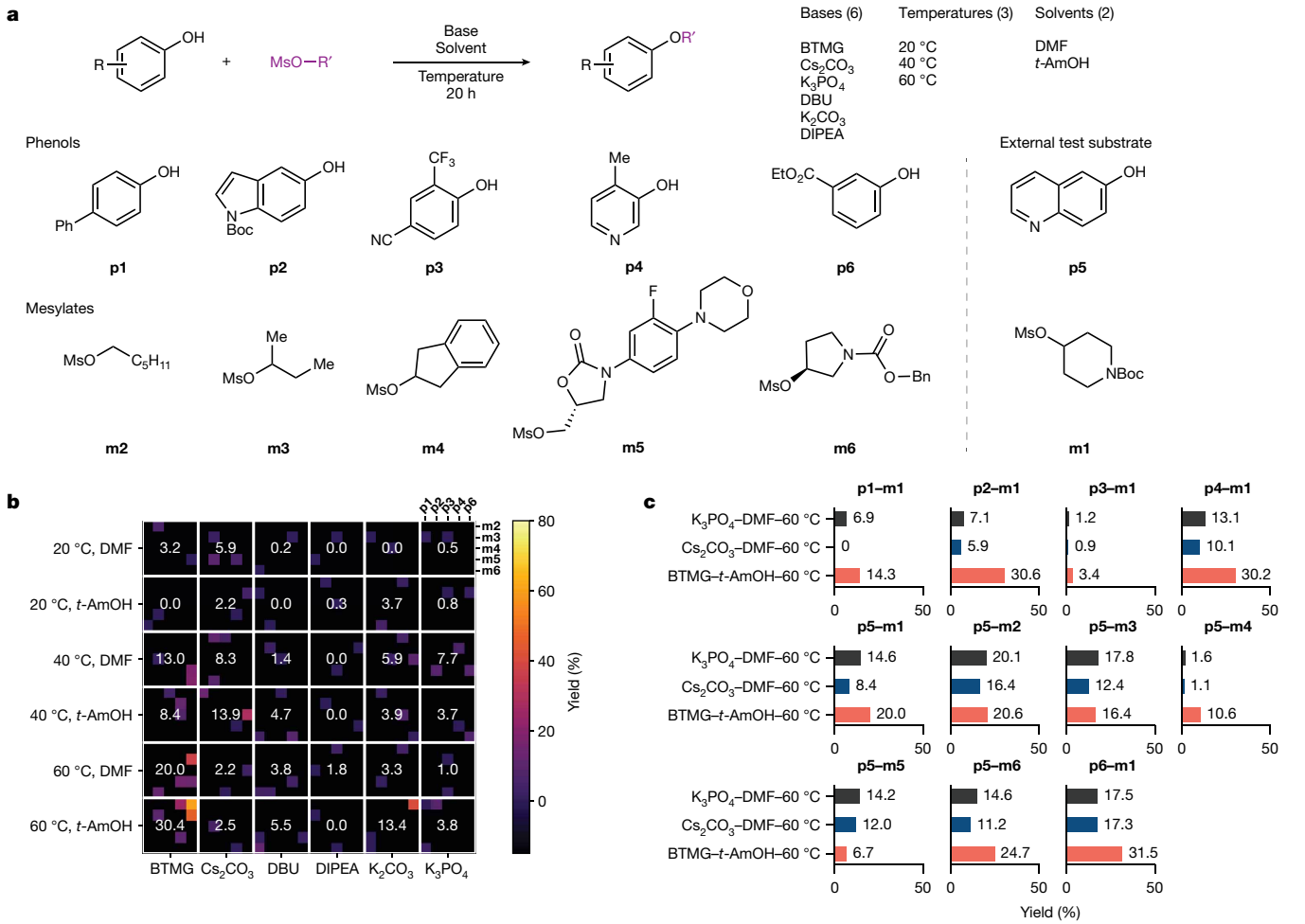


Fig. 5 | Optimization studies of phenol alkylation with mesylates. a, The substrate and condition scope for the phenol alkylation reaction, with two external test substrates not included in optimization highlighted. **b**, Summary of experiments conducted after four rounds of optimization (90 experiments). For each condition, different substrate combinations were selected to test by the UCB1-Tuned algorithm, with the yields for each individual reaction shown with a colour scale. The white numbers represent the current average yields of all conditions based on reactions that have been run. **c**, Performance comparison of the optimal condition identified, BTMG–*t*-AmOH–60 °C, with two most commonly used phenol alkylation conditions at BMS (K_3PO_4–DMF–60 °C and Cs_2CO_3–DMF–60 °C) on 11 unseen substrate pairings. BTMG, 2-tert-butyl-1,1,3,3-tetramethylguanidine; DBU, 1,8-diazabicyclo[5.4.0]undec-7-ene; DMF, dimethylformamide; t-AmOH, tert-amyl alcohol.

ether synthesis[60], Mitsunobu etherification[61] and nucleophilic aromatic substitution (SNAr), are still widely used because of their simplicity. However, these reactions usually have limited functional group compatibility. We decided to investigate a base-promoted phenol alkylation reaction with alkyl mesylates, which also suffers from similar substrate applicability issues, with the objective of identifying a more general condition.

Six mesylates and six phenols were selected from commercial databases as substrates with varying structural motifs and complexities. We randomly left out one phenol (**p5**) and one mesylate (**m1**) as external testing substrates and did not include them in the optimization process. As a result, 25 substrate pairings (five phenols × five mesylates) were sampled by the algorithm during optimization, and 11 unseen pairings (those with **p5** and **m1**, including **p5**–**m1**) were tested after to externally validate the generality of the identified conditions. Six bases (inorganic and organic), two solvents and three temperatures were selected as the condition scope, totaling 36 overall conditions (Fig. 5a). Conditions selected by expert medicinal and process chemists at BMS and their corresponding reactivity data were used as a benchmark for the decisions of the bandit algorithm and optimization performance.

Using UCB1-Tuned algorithm, we conducted four rounds of optimization with a total of 90 experiments (36, 18, 18 and 18 for each round; all conducted experiments are included in the Supplementary Information section 11.3). The first round of experiments is a uniform exploration of all conditions required by UCB-type algorithms. All conditions were sequentially explored with randomly sampled substrate pairings (21 out of 25 were sampled at this stage). Subsequent rounds of experiments were chosen by the algorithm evaluating different conditions and substrate pairings. After 90 experiments, or 10% of the available reaction scope, the average yields and number of samples for each condition were analysed (Fig. 5b and Supplementary Fig. 118). Notable base (BTMG) and temperature (60 °C) effects on reactivity were observed, with BTMG–*t*-AmOH–60 °C identified as the most generally applicable condition, achieving an average yield of 30.4% over five substrate pairs tested. Two conditions most commonly used and most successful in past HTE datasets at BMS, Cs_2CO_3–DMF–60 °C and K_3PO_4–DMF–60 °C were selected as benchmark conditions for comparison (Supplementary Information section 11.4). These three conditions were tested on 11 unseen substrate pairings that involve phenol **p5** and mesylate **m1** (Fig. 5c). Compared with the benchmark conditions, the algorithmically derived condition, BTMG–*t*-AmOH–60 °C, performed better (or at least comparably) in all except one substrate pairing (**p5**–**m5**). These results showed that bandit algorithms are compatible with continuous parameter optimization and can be used with batch sizes amenable to HTE.

Furthermore, validation with unseen substrate pairings showed that the condition identified by the bandit algorithm during optimization is more generally applicable for the reaction scope, even when compared with conditions selected by practicing chemists that performed well in historical datasets.

Discussion

Our learning model can achieve data-efficient learning at high accuracies and has unique functionalities that we substantiated through the experimental investigations of three chemical transformations. Despite its advances, the optimization framework still has limitations and can be improved in a few areas. Given the typical experimental budget (100–1,000 experiments) and the efficiency of optimization (2–10% exploration of the scope needed), our approach is not suitable for the evaluation of a scope with thousands of possible conditions. Rather, the condition scope needs to be reduced by expert chemists to selective conditions that show reactivity initially, so that more experimental resources can be spent on sampling substrates. Furthermore, the treatment of reaction conditions as independent arms in a stochastic multi-armed bandit problem setting means that there is no sharing of structural information between arms. Although effective in all our test cases, this approach can be inefficient when more than 100 conditions need to be simultaneously evaluated and significant correlations between conditions are present. Elimination of less effective conditions, as demonstrated in the amide coupling example (optimization study 2), can attenuate this problem. Alternatively, suitable descriptors for conditions could be used to transfer knowledge between similar conditions, but the choice of descriptors is difficult to determine a priori. Finally, although we showed that the learning model can successfully adjust to a changing environment with unseen substrates and correctly identify most general conditions, addition of any new conditions will require additional sampling for the model to have an accurate estimation of their performance. This issue was partially addressed by the inclusion of a real-time supervised learning model, which can be used to extrapolate to unseen conditions and predict their effectiveness, but a more direct approach with knowledge transfer between arms is still desired.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-07021-y.

1. Wagen, C. C., McMinn, S. E., Kwan, E. E. & Jacobsen, E. N. Screening for generality in asymmetric catalysis. *Nature* **610**, 680–686 (2022).
2. Rein, J. et al. Generality-oriented optimization of enantioselective aminoxyl radical catalysis. *Science* **380**, 706–712 (2023).
3. Betinol, I. O., Lai, J., Thakur, S. & Reid, J. P. A data-driven workflow for assigning and predicting generality in asymmetric catalysis. *J. Am. Chem. Soc.* **145**, 12870–12883 (2023).
4. Kim, H. et al. A multi-substrate screening approach for the identification of a broadly applicable Diels–Alder catalyst. *Nat. Commun.* **10**, 770 (2019).
5. Angello, N. H. et al. Closed-loop optimization of general reaction conditions for heteroaryl Suzuki-Miyaura coupling. *Science* **378**, 399–405 (2022).
6. Rinehart, N. I. et al. A machine-learning tool to predict substrate-adaptive conditions for Pd-catalyzed C–N couplings. *Science* **381**, 965–972 (2023).
7. Lattimore, T. & Szepesvári, C. *Bandit Algorithms* (Cambridge Univ. Press, 2020).
8. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* 2nd edn (Bradford Books, 2018).
9. Slivkins, A. Introduction to multi-armed bandits. Preprint at arxiv.org/abs/1904.07272v7 (2019).
10. White, J. M. *Bandit Algorithms for Website Optimization: Developing, Deploying, and Debugging* (O'Reilly Media, 2013).
11. Ruiz-Castillo, P. & Buchwald, S. L. Applications of palladium-catalyzed C–N cross-coupling reactions. *Chem. Rev.* **116**, 12564–12649 (2016).
12. Ogba, O. M., Warner, N. C., O'Leary, D. J. & Grubbs, R. H. Recent advances in ruthenium-based olefin metathesis. *Chem. Soc. Rev.* **47**, 4510–4544 (2018).
13. Kolb, H. C., VanNieuwenhze, M. S. & Sharpless, K. B. Catalytic asymmetric dihydroxylation. *Chem. Rev.* **94**, 2483–2547 (1994).
14. Chatterjee, S., Guidi, M., Seeberger, P. H. & Gilmore, K. Automated radial synthesis of organic molecules. *Nature* **579**, 379–384 (2020).
15. Echtermeyer, A., Amar, Y., Zakrzewski, J. & Lapkin, A. Self-optimisation and model-based design of experiments for developing a C–H activation flow process. *Beilstein J. Org. Chem.* **13**, 150–163 (2017).
16. Coley, C. W., Abolhasani, M., Lin, H. & Jensen, K. F. Material-efficient microfluidic platform for exploratory studies of visible-light photoredox catalysis. *Angew. Chem. Int. Ed.* **56**, 9847–9850 (2017).
17. 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).
18. Hsieh, H.-W., Coley, C. W., Baumgartner, L. M., Jensen, K. F. & Robinson, R. I. Photoredox iridium-nickel dual catalyzed decarboxylative arylation cross-coupling: from batch to continuous flow via self-optimizing segmented flow reactor. *Org. Process Res. Dev.* **22**, 542–550 (2018).
19. Schweidtmann, A. M. et al. Machine learning meets continuous flow chemistry: automated optimization towards the Pareto front of multiple objectives. *Chem. Eng. J.* **352**, 277–282 (2018).
20. Burger, B. et al. A mobile robotic chemist. *Nature* **583**, 237–241 (2020).
21. Häse, F., Aldeghi, M., Hickman, R. J., Roch, L. M. & Aspuru-Guzik, A. Gryffin: an algorithm for Bayesian optimization of categorical variables informed by expert knowledge. *Appl. Phys. Rev.* **8**, 031406 (2021).
22. Taylor, C. J. et al. Accelerated chemical reaction optimization using multi-task learning. *ACS Cent. Sci.* **9**, 957–968 (2023).
23. Zhou, Z., Li, X. & Zare, R. N. Optimizing chemical reactions with deep reinforcement learning. *ACS Cent. Sci.* **3**, 1337–1344 (2017).
24. Torres, J. A. G. et al. A multi-objective active learning platform and web app for reaction optimization. *J. Am. Chem. Soc.* **144**, 19999–20007 (2022).
25. Shields, B. J. et al. Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **590**, 89–96 (2021).
26. Häse, F., Roch, L. M., Kreisbeck, C. & Aspuru-Guzik, A. Phoenics: a Bayesian optimizer for chemistry. *ACS Cent. Sci.* **4**, 1134–1145 (2018).
27. Clayton, A. D. et al. Algorithms for the self-optimisation of chemical reactions. *React. Chem. Eng.* **4**, 1545–1554 (2019).
28. Reker, D., Hoyt, E. A., Bernardes, G. J. L. & Rodrigues, T. Adaptive optimization of chemical reactions with minimal experimental information. *Cell Rep. Phys. Sci.* **1**, 100247 (2020).
29. Shim, E. et al. Predicting reaction conditions from limited data through active transfer learning. *Chem. Sci.* **13**, 6655–6668 (2022).
30. Gao, H. et al. Using machine learning to predict suitable conditions for organic reactions. *ACS Cent. Sci.* **4**, 1465–1476 (2018).
31. Kozlowski, M. C. On the topic of substrate scope. *Org. Lett.* **24**, 7247–7249 (2022).
32. Gensch, T. & Glorius, F. The straight dope on the scope of chemical reactions. *Science* **352**, 294–295 (2016).
33. Dreher, S. D. Catalysis in medicinal chemistry. *React. Chem. Eng.* **4**, 1530–1535 (2019).
34. Kariofillis, S. K. et al. Using data science to guide aryl bromide substrate scope analysis in a Ni/photoredox-catalyzed cross-coupling with acetals as alcohol-derived radical sources. *J. Am. Chem. Soc.* **144**, 1045–1055 (2022).
35. Dreher, S. D. & Krska, S. W. Chemistry informer libraries: conception, early experience, and role in the future of cheminformatics. *Acc. Chem. Res.* **54**, 1586–1596 (2021).
36. Collins, K. D. & Glorius, F. A robustness screen for the rapid assessment of chemical reactions. *Nat. Chem.* **5**, 597–601 (2013).
37. Kullmer, C. N. P. et al. Accelerating reaction generality and mechanistic insight through additive mapping. *Science* **376**, 532–539 (2022).
38. Taylor, C. J. et al. A brief introduction to chemical reaction optimization. *Chem. Rev.* **123**, 3089–3126 (2023).
39. Svensson, H. G., Bjerrum, E. J., Tyrchan, C., Engkvist, O. & Chehreghani, M. H. Autonomous drug design with multi-armed bandits. In *2022 IEEE International Conference on Big Data* 5584–5592 (IEEE, 2022).
40. Romeo Atance, S., Viguera Diez, J., Engkvist, O., Olsson, S. & Mercado, R. *De novo* drug design using reinforcement learning with graph-based deep generative models. *J. Chem. Inf. Model.* **62**, 4863–4872 (2022).
41. Xu, Z., Shim, E., Tewari, A. & Zimmerman, P. Adaptive sampling for discovery. In *Proc. Advances in Neural Information Processing System* Vol. 35, 1114–1126 (NeurIPS, 2022).
42. Kaufmann, E., Cappe, O. & Garivier, A. On Bayesian upper confidence bounds for bandit problems. In *Proc. Machine Learning Research* Vol. 22, 592–600 (PMLR, 2012).
43. Auer, P., Cesa-Bianchi, N. & Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.* **47**, 235–256 (2002).
44. Snoek, J. et al. Scalable Bayesian optimization using deep neural networks. In *Proc. Machine Learning Research* Vol. 27, 2171–2180 (PMLR, 2015).
45. Stevens, J. M. et al. Advancing base metal catalysis through data science: insight and predictive models for Ni-catalyzed borylation through supervised machine learning. *Organometallics* **41**, 1847–1864 (2022).
46. Nielsen, M. K., Ahneman, D. T., Riera, O. & Doyle, A. G. Deoxyfluorination with sulfonyl fluorides: navigating reaction space with machine learning. *J. Am. Chem. Soc.* **140**, 5004–5008 (2018).
47. Lin, S. et al. Mapping the dark space of chemical reactions with extended nanomole synthesis and MALDI-TOF MS. *Science* **361**, eaar6236 (2018).
48. 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).
49. Brown, D. G. & Boström, J. Analysis of past and present synthetic methodologies on medicinal chemistry: where have all the new reactions gone? *J. Med. Chem.* **59**, 4443–4458 (2016).

50. El-Faham, A. & Albericio, F. Peptide coupling reagents, more than a letter soup. *Chem. Rev.* **111**, 6557–6602 (2011).
51. Dombrowski, A. W., Aguirre, A. L., Shrestha, A., Sarris, K. A. & Wang, Y. The chosen few: parallel library reaction methodologies for drug discovery. *J. Org. Chem.* **87**, 1880–1897 (2022).
52. Matheron, G. Principles of geostatistics. *Econ. Geol.* **58**, 1246–1266 (1963).
53. Zimmerman, D., Pavlik, C., Ruggles, A. & Armstrong, M. P. An experimental comparison of ordinary and universal kriging and inverse distance weighting. *Math. Geol.* **31**, 375–390 (1999).
54. Magano, J. Large-scale amidations in process chemistry: practical considerations for reagent selection and reaction execution. *Org. Process Res. Dev.* **26**, 1562–1689 (2022).
55. Beutner, G. L. et al. TCFH–NMI: direct access to *N*-acyl imidazoliums for challenging amide bond formations. *Org. Lett.* **20**, 4218–4222 (2018).
56. Stevens, J. M. et al. Leveraging high-throughput experimentation to drive pharmaceutical route invention: a four-step commercial synthesis of branebrutinib (BMS-986195). *Org. Process Res. Dev.* **26**, 1174–1183 (2022).
57. Sperry, J. B. et al. Thermal stability assessment of peptide coupling reagents commonly used in pharmaceutical manufacturing. *Org. Process Res. Dev.* **22**, 1262–1275 (2018).
58. Zheng, B. et al. Preparation of the HIV attachment inhibitor BMS-663068. Part 6. Friedel–Crafts acylation/hydrolysis and amidation. *Org. Process Res. Dev.* **21**, 1145–1155 (2017).
59. Krishnan, K. K., Ujwaldev, S. M., Sindhu, K. S. & Anilkumar, G. Recent advances in the transition metal catalyzed etherification reactions. *Tetrahedron* **72**, 7393–7407 (2016).
60. Fuhrmann, E. & Talbiersky, J. Synthesis of alkyl aryl ethers by catalytic Williamson ether synthesis with weak alkylation agents. *Org. Process Res. Dev.* **9**, 206–211 (2005).
61. Swamy, K. C. K., Kumar, N. N. B., Balaraman, E. & Kumar, K. V. P. P. Mitsunobu and related reactions: advances and applications. *Chem. Rev.* **109**, 2551–2651 (2009).

Methods

Detailed descriptions of bandit optimization algorithms implemented in this study, benchmark simulation testing of algorithms with synthetic data, optimization model design for chemistry reaction data and global analysis and simulation of various reaction datasets can be found in the Supplementary Information. Dataset designs, procedures of high-throughput experimentation, authentic product syntheses and characterizations for the palladium-catalysed imidazole C–H arylation reaction, amide coupling reaction and phenol alkylation reaction are also included in the Supplementary Information.

Data availability

All reaction datasets evaluated in simulation studies and the two newly collected reaction datasets (the palladium-catalysed C–H arylation reaction and the amide coupling reaction) are available at GitHub (https://github.com/doyle-lab-ucla/bandit-optimization). Raw data logs from simulation studies with both synthetic data and chemistry reaction data are available at Zenodo (https://doi.org/10.5281/zenodo.8170874).

Code availability

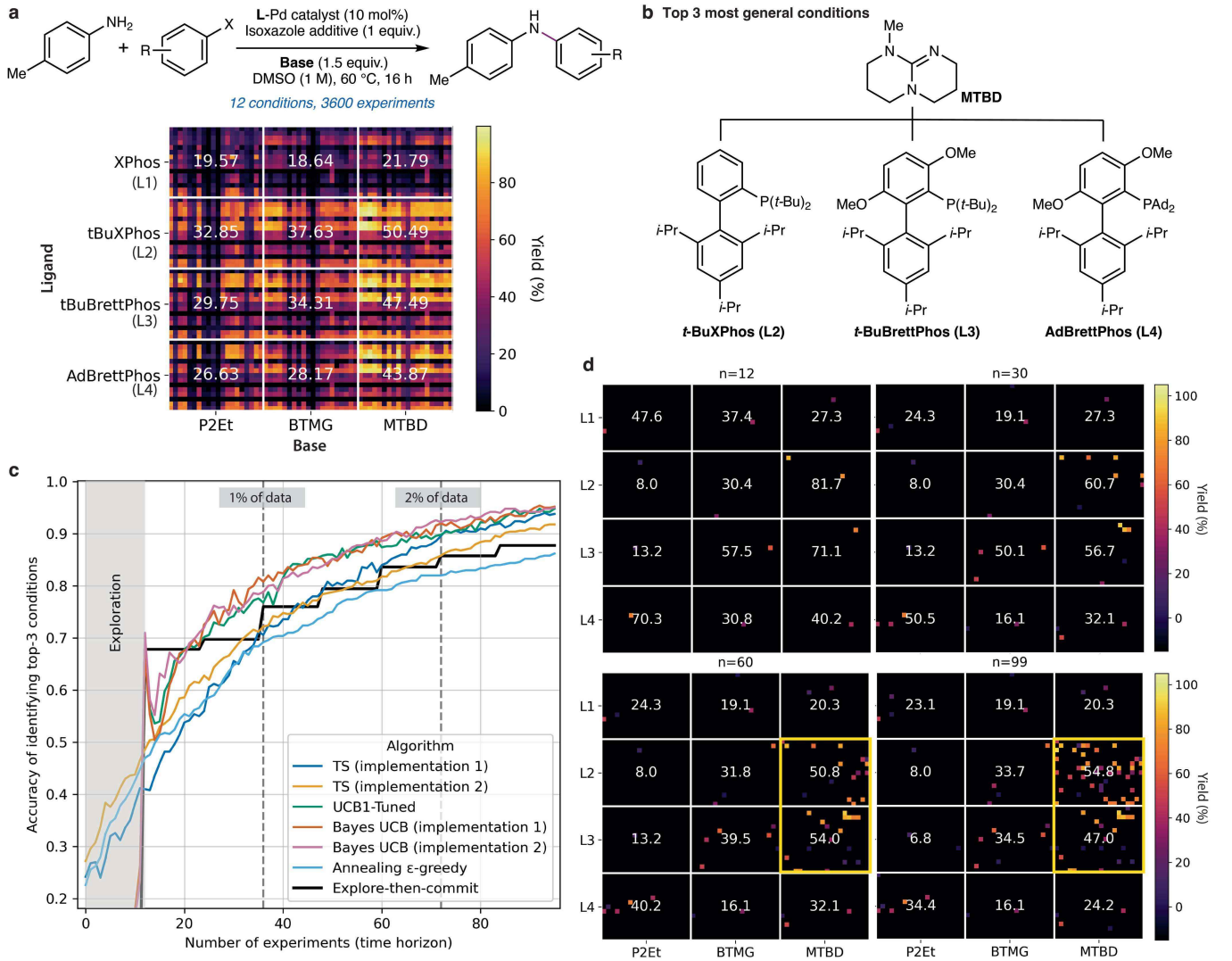
All source codes for implemented optimization algorithms and models, simulation methods for synthetic data and chemistry reaction dataset and analysis functions for data logs and optimization results are available at GitHub (https://github.com/doyle-lab-ucla/bandit-optimization). The current release of the software is also available at Zenodo (https://doi.org/10.5281/zenodo.8181283).

Acknowledgements The financial support for this study was provided by BMS, the Princeton Catalysis Initiative, the NSF under the CCI Center for Computer Assisted Synthesis (CHE-2202693) and the Dreyfus Program for Machine Learning in the Chemical Sciences and Engineering. J.Y.W. acknowledges support from the BMS Graduate Fellowship in Synthetic Organic Chemistry. S.K.K. acknowledges support from the NSF Graduate Research Fellowship Program under grant no. DGE-1656466. M.P. acknowledges support from the NIH F32 Ruth L. Kirschstein NRSA Fellowship (1F32GM129910-01A1). We thank J. Raab, M. Ruos and S. Gandhi for reviewing the Supplementary Information.

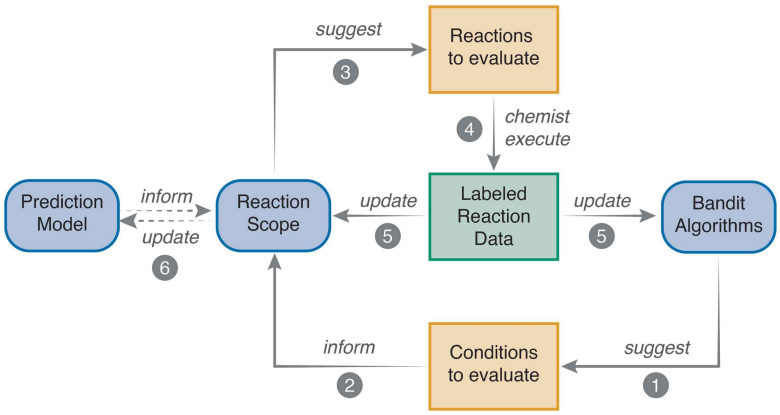
Author contributions J.Y.W. and A.G.D. designed the overall research project. J.Y.W. designed and implemented optimization models and algorithms with inputs from J.M.S., J.L., J.E.T., B.J.S. and A.G.D.; J.M.S., B.J.S., J.L., J.E.T., J.Y.W. and A.G.D. designed and planned reaction scopes for the C–H arylation reaction, the amide coupling reaction and the phenol alkylation reaction. J.M.S., S.K.K., M.-J.T., D.L.G., M.P., D.N.P., B.H., D.D., S.D., A.F., G.G.Z., S.M. and J.P. carried out high-throughput experiments and authentic product synthesis for the three reactions. J.Y.W. wrote the paper with inputs from all authors.

Competing interests The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-07021-y.
Correspondence and requests for materials should be addressed to Abigail G. Doyle.
Peer review information *Nature* thanks Jolene Reid and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



Extended Data Fig. 1 | Testing the bandit algorithms on a previously published C–N cross-coupling reaction dataset. a, General reaction scheme of the C–N cross-coupling reaction and reactivity heatmap grouped by base and ligand, with average yields for each base/ligand combination shown in white text. Structures for all substrates and conditions in the scope are included in the Supplementary Information. **b**, Top three most general base–ligand conditions for the dataset. **c**, Average accuracies of identifying top-3 conditions with various algorithms across 500 simulations with random starts. Exploration refers to the uniform exploration required by some algorithms, during which each condition is sequentially selected once. Different implementations of TS and Bayes UCB algorithms were used and differentiated by implementation 1 and 2 for simplicity. This plot is reproduced in Fig. S83, with the details of the algorithms included in the legend. TS: Thompson Sampling; UCB: upper confidence bound. **d**, Real-time optimization progress for simulation 0 (the first simulation) of a Bayes UCB (implementation 2) algorithm at n = 12, 30, 60, 99. Squares with different colors represent all reactions that have been suggested and evaluated by the algorithm at the time. The real-time empirical average for each base/ligand combination is shown in white texts.



Extended Data Fig. 2 | Model architecture and workflow of bandit algorithms during reaction optimization. The bandit algorithm suggests a condition (an arm) to evaluate first. The chemist-designed reaction scope suggests a reaction to evaluate with the selected condition. The suggested reaction is tested experimentally, and the result is used to update both the reaction scope and the bandit algorithm for the next round of proposal. Finally, a prediction model, separately trained with existing experimental results, is optionally used to propose reactions to evaluate via other mechanisms (e.g., batch proposal).