

## CHEMISTRY

# Predicting three-component reaction outcomes from ~40,000 miniaturized reactant combinations

Julian Götz<sup>1†</sup>, Euan Richards<sup>1</sup>, Iain A. Stepek<sup>1‡</sup>, Yu Takahashi<sup>1§</sup>, Yi-Lin Huang<sup>1¶</sup>, Louis Bertschi<sup>2</sup>, Bertran Rubi<sup>2#</sup>, Jeffrey W. Bode<sup>1\*</sup>

Efficient drug discovery depends on reliable synthetic access to candidate molecules, but emerging machine learning approaches to predicting reaction outcomes are hampered by poor availability of high-quality data. Here, we demonstrate an on-demand synthesis platform based on a three-component reaction that delivers drug-like molecules. Miniaturization and automation enable the execution and analysis of 50,000 distinct reactions on a 3-microliter scale from 193 different substrates, producing the largest public reaction outcome dataset. With machine learning, we accurately predict the result of unknown reactions and analyze the impact of dataset size on model training, both enabling accurate outcome predictions even for unseen reactants and providing a sufficiently large dataset to critically evaluate emerging machine learning approaches to chemical reactivity.

## INTRODUCTION

Accurately predicting the outcome of organic reactions through data-driven methods remains an unmet challenge, chiefly due to the paucity of large, unbiased datasets that cover a meaningful range of molecular properties and structural diversity (1, 2). Molecular assembly processes involving three or more components are uniquely suited to generating large libraries of organic molecules from a modest number of building blocks, rendering them ideal for producing the experimental data needed for evaluating and improving computational approaches to the prediction of reaction outcomes. For example, well-studied multicomponent reactions including Ugi condensations (3, 4) constitute the basis for work at the forefront of miniaturized library generation (5).

Inspired by such processes, our group has reported Synthetic Fermentation—the assembly of preformed building blocks into functionally and stereochemically rich oligomers without the need for exogenous reagents (6, 7). We recognized that similar processes could be leveraged to yield more drug-like molecules through three-component couplings, enabling a miniaturized approach to the production of hundreds of thousands of compounds from a limited selection of starting materials. This approach presented the opportunity to generate an extensive dataset comprising the outcome of tens of thousands of unique reactant combinations, thereby establishing an ideal training set for machine learning–assisted predictions of successful reactant combinations. The use of high-throughput experimentation to provide tailored datasets relevant to the prediction problem at hand has become a standard practice in the field (8–11), but it has remained limited to a few thousand reactions in

total and rarely more than a few dozen different substates (12). Despite progress toward predicting the reaction outcome for unknown substrates (11–13), extrapolation on multiple reactant dimensions remains a substantial challenge. In addition, the small number of test data points in extrapolative settings often hampers rigorous evaluation of model predictions. Assembling a dataset on the order of tens of thousands of reactions would enable testing of unconfirmed hypotheses in the field, including what quantity of data is necessary for predictive models and the dependence of extrapolation ability on model type and dataset size.

In this study, we report an on-demand reaction platform that uses three-component assembly to conduct 50,000 reactions on the microliter scale, each using a unique combination of substrates. Analyzing all reaction mixtures using liquid chromatography–high-resolution mass spectrometry (LC-HRMS), we track the formation of eight different products for each combination, resulting in a reaction outcome dataset of unprecedented size comprising ~40,000 reactant combinations, each with distinct products. Using machine learning, we accurately predict the reaction outcome for unseen reactant combinations and untested substrates. By investigating model performance at different dataset sizes, we can derive general conclusions for data efficiency of machine learning models in reaction prediction and the importance of training data generation.

## RESULTS

## Reaction development and automation

In this implementation of Synthetic Fermentation, we devised a three-component reaction where a potassium acyltrifluoroborate (KAT) initiator (I) reacts with 1 equiv of an isoxazolidine monomer (M) bearing a protecting group that reveals an  $\alpha$ -keto acid. This reacts with a thiohydrazide (TH) or aminobenzenethiol (ABT) terminator (T) to give the final product (Fig. 1A and figs. S1 and S2). Notably, the reaction proceeds in aqueous buffer without additional reagents or catalysts and yields benign by-products. The reaction conditions and concentrations are ideal for dilution with buffer and direct-to-biology screens (6, 14). To establish the synthetic scope and training set, we curated a building block collection initially comprising 78 initiators, 74 monomers, and 41 terminators (figs. S3 to S5). While some specialized building blocks were introduced to

Copyright © 2025 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

<sup>1</sup>Laboratory for Organic Chemistry, Department of Chemistry and Applied Biosciences, ETH Zürich, 8093 Zürich, Switzerland. <sup>2</sup>Molecular and Biomolecular Analysis Service (MoBIAS), Department of Chemistry and Applied Biosciences, ETH Zürich, 8093 Zürich, Switzerland.

\*Corresponding author. Email: bode@org.chem.ethz.ch

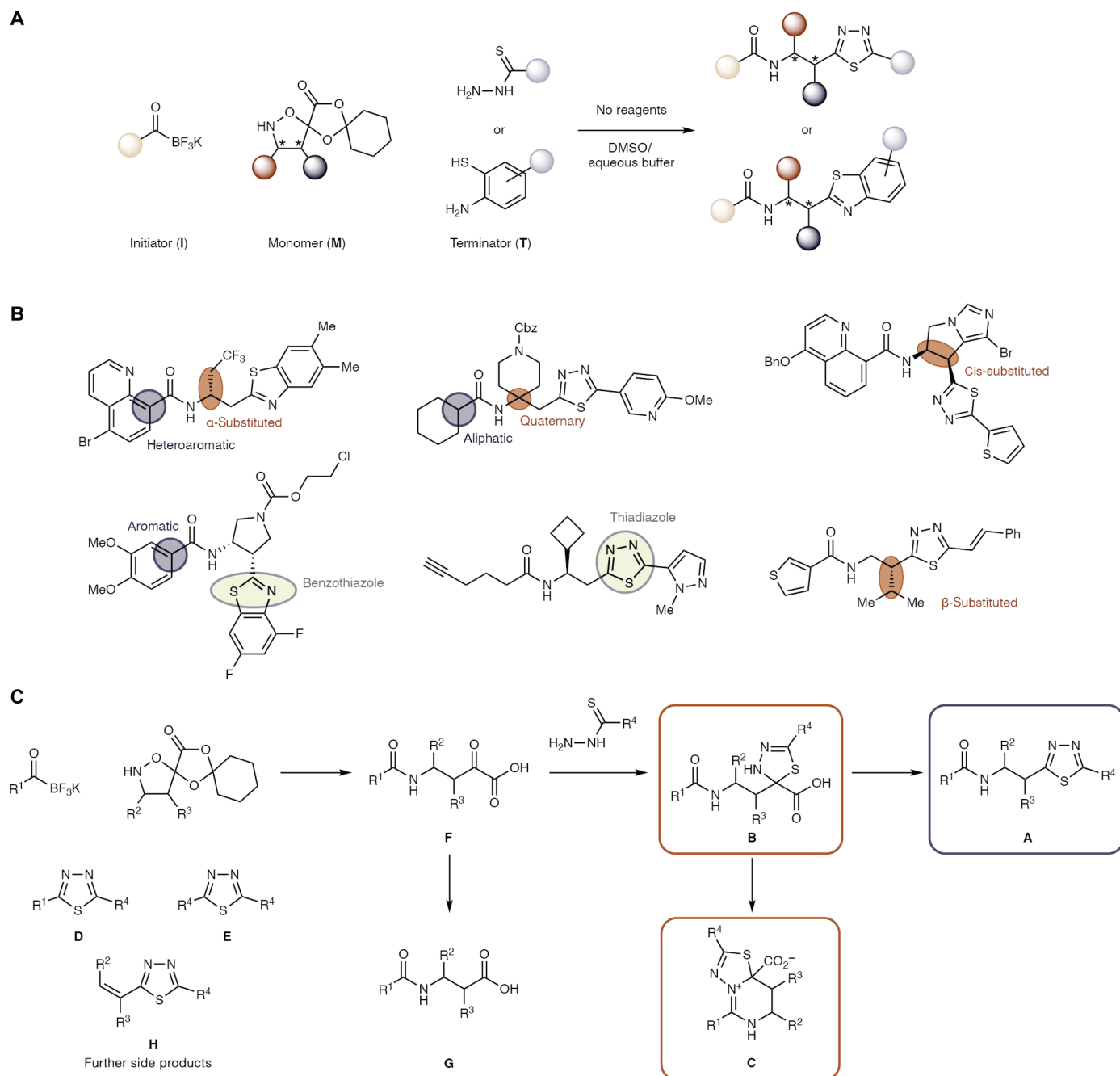
†Present address: Roche Pharma Research and Early Development, F. Hoffmann–La Roche AG, 4070 Basel, Switzerland.

‡Present address: Chemical Development, Novo Nordisk, 2880 Bagsværd, Denmark.

§Present address: Modality Research Laboratories, Kyowa Kirin Co., 411-8731 Shizuoka, Japan.

¶Present address: Quality Control, Bachem AG, 4416 Bubendorf, Switzerland.

#Present address: Biologics Process Development, LCMS R&D, Lonza AG, 3930 Visp, Switzerland.



**Fig. 1. Synthetic Fermentation.** (A) A KAT initiator (I) undergoes a reaction with an isoxazolidine monomer (M) to produce an  $\alpha$ -keto acid that subsequently condenses with a TH or ABT terminator (T) to form the final product. The reaction sequence proceeds without additional reagents or catalysts. (B) Examples of products obtained through Synthetic Fermentation with some common structural features highlighted. (C) The desired product **A** is formed in most reactions. Several side products were also identified. In some cases, the reaction stalled before the final oxidative decarboxylation at product **B**, which sometimes led to alternative, bicyclic product **C**. Further products that are occasionally observed are the intermediate  $\alpha$ -keto acid (**F**) and its oxidation product **G**, the I-T condensation product **D**, the T-T condensation product **E**, and the M-T elimination product **H**. The analogous products formed with ABT terminators are shown in fig. S1.

cover a broad set of motifs, most were readily made and have been previously reported (6, 7, 14–18). Three different initiator types (aliphatic, aromatic, and heteroaromatic), four different monomer types ( $\alpha/\beta$ -substituted vis-à-vis the nitrogen, cis-substituted, and quaternary), and two terminator types (ABT and TH) give rise to an on-demand library of ~236,000 products (Fig. 1B) that we call

“PRIME library”—i.e., the reactions to produce these structures can be executed immediately from the initially available stock of 193 building blocks.

This particular multicomponent assembly was selected not only for the attractiveness of the resulting products but also because it is not a perfect process and therefore representative of broader reaction

prediction problems. While the desired product (**A**) is typically favored, the reaction may stall ahead of the final oxidative decarboxylation (product **B**) or undergoes intramolecular cyclization instead (**C**). In certain cases, the **I-T** condensation product (**D**) and the terminator dimer (**E**) are also detectable. We further observed stalled **I-M** intermediate (**F**) and the corresponding carboxylic acid **G**. Last, an **M-T** elimination product (**H**) was formed in some test reactions (Fig. 1C). The various reaction outcomes, particularly the distinct reaction pathways leading to desired structure **A** versus alternative product **C**, rendered this system well suited both for high-throughput experimentation and as a testing ground for modern reaction prediction.

To render the multicomponent coupling amenable to high-throughput synthesis using acoustic dispensing, we modified our preferred conditions for Synthetic Fermentation by changing the solvent system from *t*-BuOH to dimethyl sulfoxide (DMSO)/1 M aq. oxalic acid and reducing the concentration of building block solutions to 50 mM to ensure solubility. Most building blocks are stable for months at  $-20^{\circ}\text{C}$  in DMSO (**I** and **M**) or in the reaction solvent (**T**). To reduce the consumption of building blocks and disposables, we adopted acoustic dispensing (4, 19) for the automated delivery of reaction components and were pleased to find that it performed well at preparing the 384-well plates used for the reactions, with each three-component reaction conducted in 3.3  $\mu\text{l}$  of 9:1 DMSO/aq. oxalic acid. Even at this low volume, no notable evaporation was observed during incubation at  $60^{\circ}\text{C}$ . It should be noted that these conditions are optimized for the highly parallel, automated synthesis of small amounts for screening. To conduct high-yielding synthesis of individual compounds in batch, other conditions may give better results. For example, cleaner product formation is observed with ammonium salt additives and product **B** can be transformed to **A** by prolonging the reaction time, increasing the temperature, or adding an oxidant (6, 7, 14).

### Execution and evaluation of 50,000 reactions

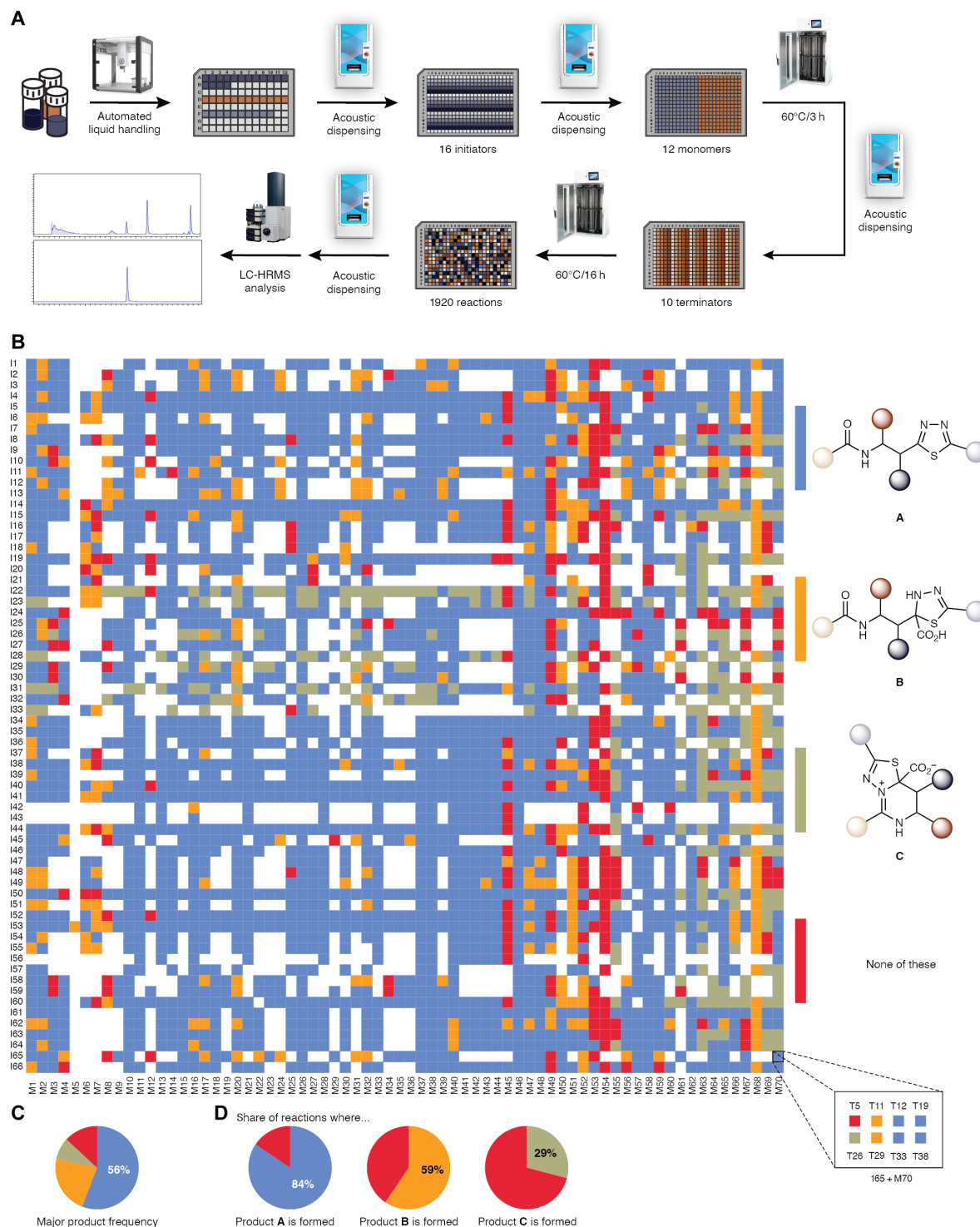
Library syntheses were conducted in batches of 1920 reactions using 16 initiators, 12 monomers, and 10 terminators per run, which required 22 hours for setup and conduction. Improvements in the automated workflow, namely using incubators within the automated system and conducting the previously manual preparation of source plates on a programmable robotic liquid handler (see movie S1), allowed us to prepare two batches or 3840 reactions in a 24-hour period. The full automation workflow is outlined in Fig. 2A. Stock solutions of building blocks were prepared freshly or used from previous batches, and aliquots were analyzed by liquid chromatography-mass spectrometry (LC-MS) for quality control. The source plates were transferred to an automation system where six 384-well synthesis plates served as the targets for acoustic dispensing (see movie S2). The plates containing initiator and monomer mixtures were incubated at  $60^{\circ}\text{C}$  for 3 hours, followed by addition of terminator solutions and incubation at  $60^{\circ}\text{C}$  overnight (16 hours). Dispensing, incubation, and all auxiliary steps (plate transfers, centrifuging, and sealing/desealing) proceeded without human intervention on the automated system.

Analyzing 50,000 different reactions—each with distinct starting materials, products, and side products—presents a formidable task. The automated system was used to prepare plates for LC-HRMS analysis by aliquoting 30 nl of the reaction solution and diluting 1:1000 with a MeCN solution containing fenofibrate as an internal standard. Valuing data quality over speed, we sought the resolution

of all expected products by chromatography, leading to an LC-HRMS runtime of 14 min per sample, including column washes (see fig. S6 for an example trace). With a dual column setup, we could reduce the effective time to 7 min per sample or about 8 months of instrument time in total. Extensive scripting for the determination of required sum formulae of all expected products, including from conceivable deprotection reactions, using Python/RDKit, and near-complete automation of instrument operation kept human involvement to a minimum. The peak areas for all sum formulae were determined using a VBScript within Bruker Compass Data Analysis and normalized by dividing by the peak area of the internal standard (fenofibrate). Because of challenges of normalizing product responses for 50,000 different reaction combinations, the reaction outcome was assessed with a binary label depending on the presence or absence of a product or side product in the LC-HRMS trace. The major product was determined by scaling the peak area according to the relative ionizability of the product type and comparing the scaled value across products (see the Supplementary Materials). For inclusion in the dataset, each data point was checked against several quality criteria using collected metadata such as building block LC-MS, transfer logs from acoustic dispensing, and product LC-HRMS. These rigorous protocols ensured high data quality without experimental replicates and led to the exclusion of around 10,000 data points from the final dataset. By conducting a subset of the reactions in duplicate (377 reactions), we observed good reproducibility (98% accuracy for product **A** and 73% balanced accuracy for the major product; see fig. S7). The outcome for ~40,000 reactions determined to be operationally successful is shown in Fig. 2B (see figs. S8 to S10 for individual products). An interactive version of this heatmap showing the outcome for individual reactions is available at <https://jugoetz.com/synferm-heatmap>. In 84% of reactions, the expected product **A** is observed, and it is the major product in 56% of reactions. Product **B** is observed irregularly with few building blocks strongly favoring it. Formation of product **C** is more likely for aliphatic initiators (**I22** to **I33**) or monomers bearing a quaternary carbon (**M55** to **M70**) and occurs almost exclusively for TH terminators (46% for TH and 6% for ABT). Only few building blocks afford none of these products. For example, monomers with a 2-pyridyl or 4-pyridyl substituent in the  $\beta$ -position (**M53** and **M54**) degrade under the established reaction conditions.

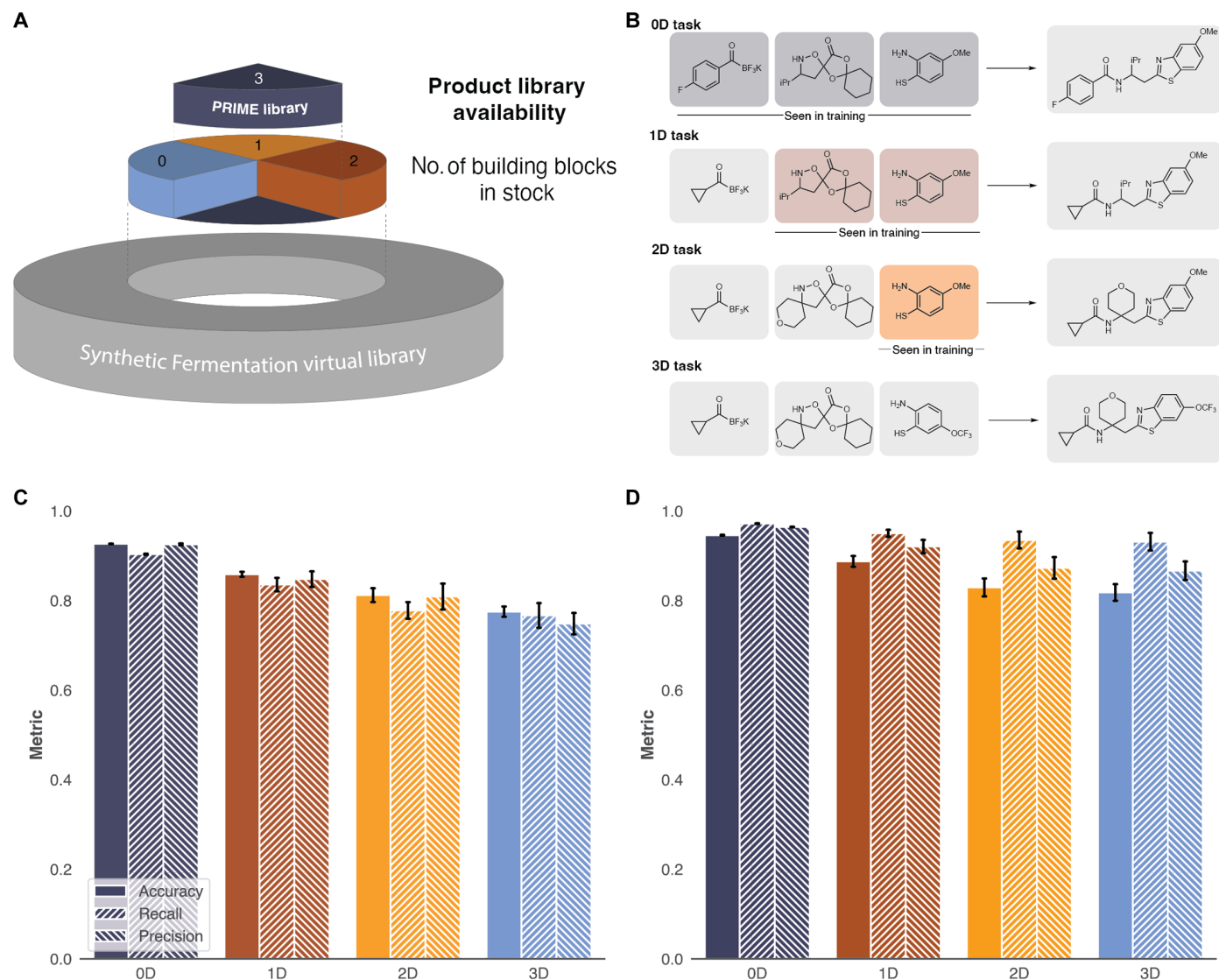
### Prediction of the reaction outcome

Having collected the reaction outcome data, we turned to machine learning to predict the outcome for arbitrary combinations of building blocks. Our use of a three-component reaction confronts machine learning models with four distinct prediction problems of increasing complexity, i.e., predicting for reactions with three, two, one, or zero previously seen building blocks—although not necessarily in the same reaction (see Fig. 3A). We trained and evaluated multitask binary classifiers for the formation of products **A**, **B**, and **C** independently for all four problems, referred to according to the number of independent dimensions between training and test data. The “zero-dimensional (0D) split” corresponds to zero independent dimensions or all three building blocks being present in the training data, whereas in the 3D split, none of the three building blocks would be present in the training data (Fig. 3B). The splitting method is described in detail in our previous work (20) and in the Supplementary Materials. Similar approaches have been used elsewhere (13, 21–23) to prevent information leakage. For each problem, we



**Fig. 2. Automated conduction and analysis of 50,000 reactions.** (A) The synthesis workflow begins with the preparation of the source plate on an OT-2 liquid handler. From the source plate, solutions of **I** and **M** are distributed to six 384-well reaction plates by acoustic dispensing and incubated within the automated system. **T** solutions are added by acoustic dispensing, and incubation is continued. Aliquots are removed for subsequent LC-HRMS analysis. h, hours. (B) Heatmap of major product (between **A**, **B**, and **C**) across all initiators and monomers, averaged over all terminators tried with this combination. The “None” class is assigned if none of the three products was present. For the combination of **I65** and **M70**, the square is expanded into all reactions with different terminators making up this square. Product **A**: blue; product **B**: orange; product **C**: khaki; None: red. White squares indicate that the combination was not attempted. For heatmaps for the formation of individual products, see figs. S8 to S10. An interactive version of this heatmap showing the outcome for individual reactions is available at <https://jugoetz.com/synferm-heatmap>. (C) Product **A** was the major product in 56%, product **B** in 23%, and product **C** in 9% of all reactions. The remaining 13% afforded none of these products. (D) Average observation ratio for each product (**A**: 84%; **B**: 59%; **C**: 29%).





**Fig. 3. Data-driven prediction of the reaction outcome.** (A) Different library spaces are differentially difficult to predict. The Synthetic Fermentation virtual library can be subdivided on the basis of the number of new building blocks needed to synthesize a VL member. With the number of new building blocks, the difficulty of predictions increases. The training data are a subset of the PRIME library. (B) Data splits created to simulate the different prediction problems faithfully. Building blocks in colored boxes are contained in the training data for each task, and building blocks in gray boxes are not contained in the training data. For example, a 1D split would leave out building blocks on one dimension (here, initiators) but contain all building blocks for the other two dimensions (monomers and terminators) in both training and test data. Using multiple folds, the three possible permutations (for 1D and 2D splits) of reactant types to be left out are investigated at the same time. (C) Performance (accuracy, precision, and recall) of the best model on each task, averaged over all products **A** to **C** and (D) for product **A**. The selected models were FFN/OHE for the 0D problem and XGB/FP for the other problems. Error bars indicate the standard error of the mean over nine random repetitions.

tried several different representations, spanning structural fingerprints (FP), RDKit, and quantum chemical (QC) properties; one-hot encoding (OHE) of reactants; or graph encoding of the reaction (CGR) (24). We tested model architectures comprising several graph neural networks (GNNs) (25–28), a feed-forward network (FFN), gradient boosting (XGB) (29), and logistic regression. Hyperparameters were optimized for each combination. The best classifier for each problem was selected by the area under the precision-recall curve (AUPRC) obtained on validation data, and computationally cheaper methods were favored if score differences were not significant.

We found that for the 0D split (i.e., all building blocks have been seen in the training data), the FFN/OHE model was superior, significantly outperforming all other combinations. For the 1D, 2D, and 3D splits, the XGB/FP models were optimal, although the computationally more expensive directed message passing neural network (D-MPNN)/CGR combination showed similar performance on most problems (see fig. S11 for all models). Evaluating the selected models on test data, we found the 0D model to have an accuracy of  $93 \pm 0.2\%$  for predicting products **A** to **C**. The 1D model still showed a good accuracy of  $86 \pm 1.6\%$ . The 2D model was correct for  $81 \pm 4.4\%$  of reactions and the 3D model for  $78 \pm 3.3\%$ .

These metrics and the respective precision and recall scores as well as metrics for product **A** only are shown in Fig. 3.

While these results are satisfactory, it is worth highlighting their limitations. For product **A** specifically, which was formed in more than four of five reactions, only the 0D and 1D models are meaningfully more accurate than a naïve classifier that always predicts that the expected product is formed. The monotonous decline of accuracy from the 0D model to the 3D model is expected and reflects the relative difficulty of learning chemical information in contrast to learning the combinatorial structure of the dataset. Still, on the 3D split, the “chemistry-aware” XGB/FP is significantly and markedly better than the chemistry-unaware FFN/OHE model, which (by design of the split) is only as good as random guesses.

### Experimental validation

In addition to the computational validation, models were tested on prospective tasks reflecting real-world use cases (Fig. 4A). As we use the 0D model to predict reaction outcomes for the PRIME library (the VL including predictions is provided in data S4), we tested the precision of this model by preparing one plate of compounds predicted to be synthesizable. Of 149 valid reactions that were predicted to yield product **A** by the model, it was found in 147 of them, corresponding to a 99% precision score in this prospective evaluation (Fig. 4B). For the secondary objectives, predicting formation of products **B** and **C**, the prospective accuracy values were 87 and 86%, respectively.

The second task reflects a question that may arise when using Synthetic Fermentation for compound optimization. If an initial hit from the PRIME library was promising but analogs are desired for optimization, one may synthesize new initiators not contained in the PRIME library. To determine whether these building blocks are worth procuring, one would ask the 1D model for the likely reaction outcome. To test this scenario, we curated a diverse set of 12 initiators that were not used in the library synthesis. These new initiators were reacted with random monomers and terminators from the library. For 593 valid reactions, the model successfully predicted whether product **A** is formed for 533 (90%) of them (Fig. 4C).

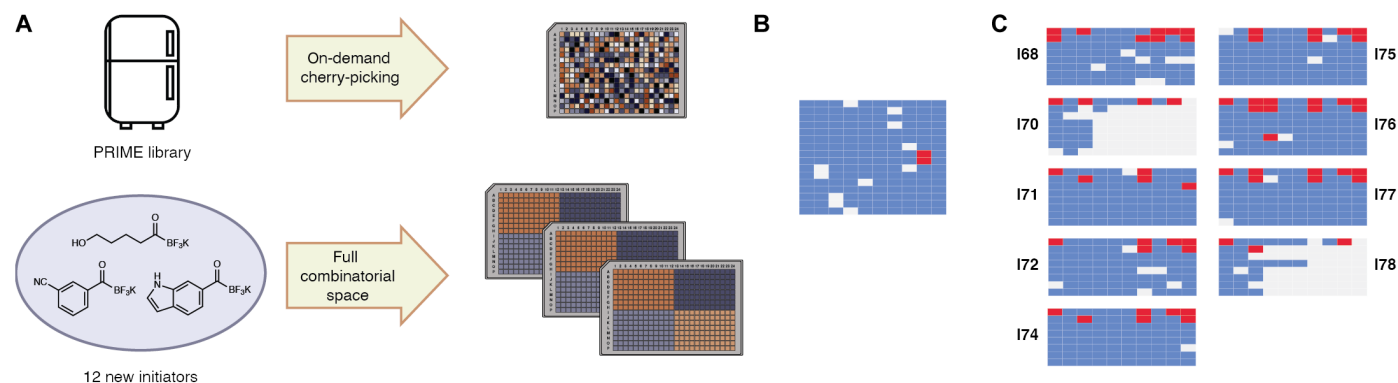
Notably, most errors were false negatives and in only six cases was product **A** not detected despite being predicted to form, corresponding to 98.9% precision. As the main objective in this scenario is avoiding costly synthesis of unproductive building blocks, the model proved highly suitable.

The strong prospective performance, combined with the previous computational evaluation, demonstrates the practical usefulness of the models. In a discovery setting, these models could be used to efficiently prioritize and select compounds for synthesis, reducing waste of material and time in unsuccessful reactions.

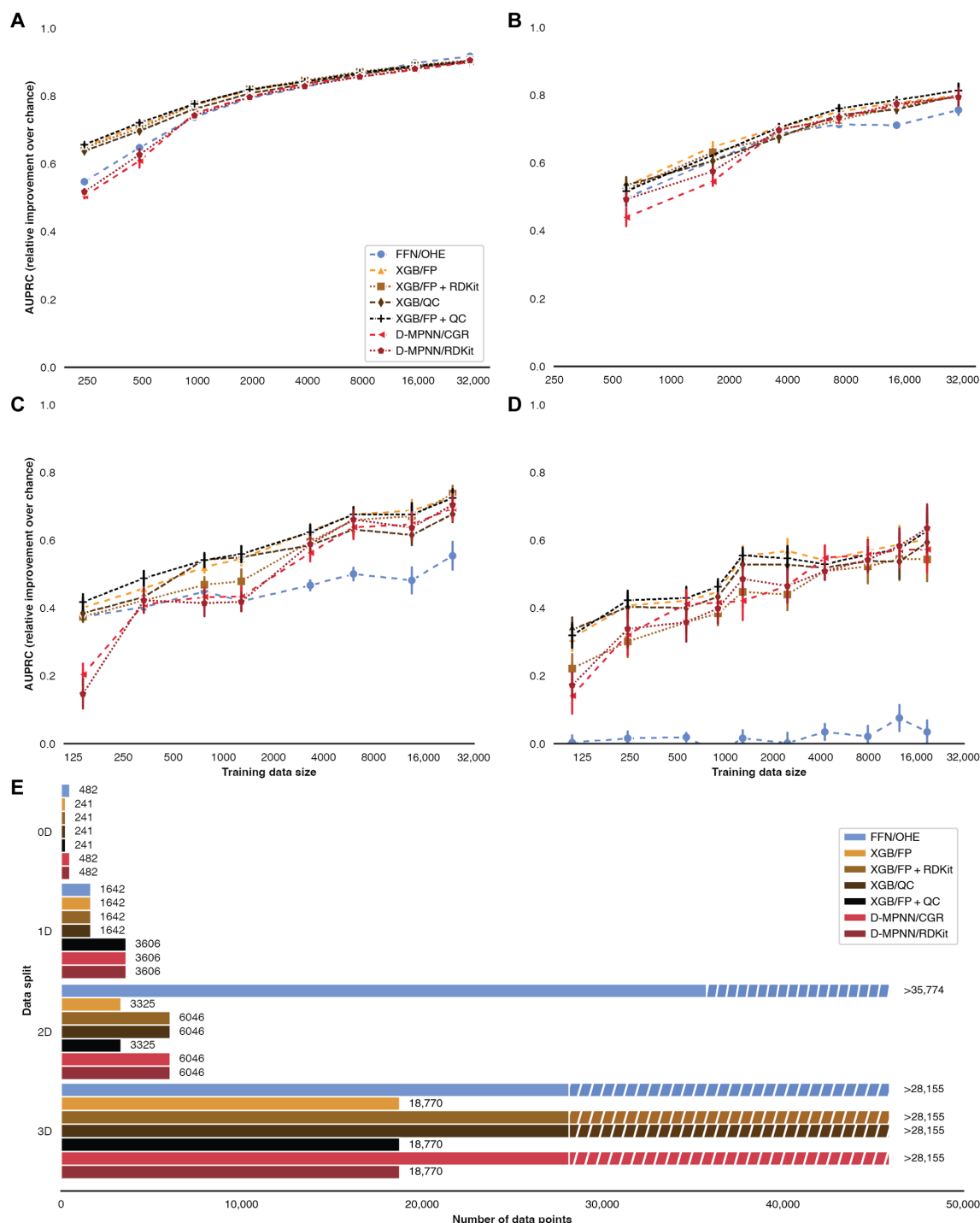
### Data and model requirements for reaction prediction

The unprecedentedly large dataset provides the opportunity to investigate the influence of dataset size on the performance of different model types. For each of the problems outlined above, we created truncated splits with a reduced number of data points in the training set. We selected the three model types that showed the best performance for the full dataset (FFN/OHE, XGB, and D-MPNN). For the chemistry-aware models, we used either structural features (FP for XGB and CGR for D-MPNN) or additional property features (RDKit). The five model types were trained (including full hyperparameter tuning) for all truncated splits.

On the 0D split (Fig. 5A), i.e., a pure interpolation task, the performance for all model types converges above 2000 data points. Despite the small difference, the FFN/OHE model is significantly better than all other model types at large dataset sizes. While the predictive power keeps rising even up to ~30,000 training data points, working with a much sparser dataset of a few thousand samples yields satisfactory models. The differentiation below 2000 samples, where XGB models outperform neural network methods, hints at the higher data efficiency of the former. On the 1D split (Fig. 5B), a similar picture emerges. For small training datasets, XGB is the best choice, but with increasing dataset size (>4000 samples), the advantage over neural networks vanishes. Unexpectedly, the chemistry-unaware model (FFN/OHE) is indistinguishable from the chemistry-aware methods up to training dataset sizes of ~8000 samples. On the 2D split (Fig. 5C), scores are lower across model types, but the advantage of



**Fig. 4. Experimental validation.** (A) Tasks posed in experimental validation. One plate was created by cherry-picking compounds from the in-stock PRIME library that the 0D model predicted to work. Three plates were created on the basis of 12 new initiators previously not contained in the PRIME library, paired with 10 previously used monomers and 8 previously used terminators. Reaction outcomes for these plates were predicted with the 1D model. (B) Heatmap of the reaction outcome for the cherry-picked plate. Blue squares indicate that product **A** was found (i.e., the 0D model was correct). Red squares indicate that it was not found. Gray squares were excluded from the analysis because they failed previously established quality control criteria. (C) Heatmap showing whether the model predicted the correct outcome for the plates with new initiators, grouped by the initiator. Blue squares indicate a correct prediction, and red squares indicate an incorrect prediction. Gray squares were excluded from the analysis because they failed previously established quality control criteria.



**Fig. 5. Data efficiency of machine learning models.** Relative improvement over the chance level for the AUPRC on test data depending on the size of the dataset used in training. As relative improvement is given, a random classifier has an expected value of 0 and a perfect classifier has an expected value of 1. All error bars correspond to the standard error of the mean over nine random repetitions. **(A)** On the 0D split, differentiation is small but significant at large dataset sizes. Below a few thousand training samples, XGB markedly outperforms even a simple neural network and outperforms GNNs by a wider margin, irrespective of the tested features. **(B)** On the 1D split, XGB models are best for most dataset sizes, but GNNs catch up for large training datasets. Only at large sizes, the chemistry-aware models outperform the chemistry-unaware model (FFN/OHE). **(C)** On the 2D split, XGB models are preferred at all training dataset sizes, but GNNs can outperform the chemistry-unaware model after a few thousand data points. **(D)** On the 3D split, the chemistry-unaware model is not better than chance, while chemistry-aware models are moderately predictive and do not consistently differentiate in performance. **(E)** Number of training data points required for a 60% relative improvement over the chance level depending on the model, featurization, and split dimension. Dashed bars indicate that the estimator has not reached the required score even at the largest tested training data size, which is indicated by the solid portion of the bar. Data for the FFN/OHE model on the 3D split are included for completeness, but the model can only give random predictions on this split by design.

XGB models for small datasets is reproduced and extends to even larger sizes. The chemistry-unaware model becomes unfavorable at smaller sizes (a few thousand data points). Last, on the 3D split (Fig. 5D), the chemistry-unaware model gives random predictions (as expected by design). The markedly increased variance of the results leads to no consistent differentiation among the chemistry-aware models. Performance remains moderate but much better than chance up to the maximum dataset size.

The results show that at small dataset sizes, QC features can have a minor positive effect on prediction performance. For the XGB/QC model, this coincides with a performance penalty at larger dataset sizes. The penalty can be mitigated by combining Morgan fingerprints and QC features in the same model, but the marginal improvement over the less expensive XGB/FP featurization may not warrant the additional computational cost at inference time.

These observations imply that while predicting within a combinatorial reaction dataset (equivalent to the 0D split) is possible even from a sparsely sampled combinatorial space (i.e., ~1% of ~200,000 possible data points), there is no reason to use chemistry-aware models in such cases. Even when extrapolating for one of the three reactants (1D split), chemistry awareness became relevant only beyond 10,000 training data points in our case, which is an order of magnitude larger than previously examined datasets. If chemistry-aware models are used for the 0D and 1D splits, we saw only disadvantages for more complicated GNNs over XGB, especially at the small dataset sizes that are common in chemistry. Figure 5E shows how the data demand increases with the difficulty of the problem and how data efficient the tested model types were.

It is worth asking to what extent these results transfer to other reaction types. The modelability of molecular properties depends on the roughness of the structure-property landscape (30). Arguably, the same holds for the structure-reactivity landscape in predicting reaction properties. Qualitatively, the Synthetic Fermentation dataset exhibits moderate roughness with many building blocks following simple patterns (e.g., formation of product **C** mainly for aliphatic initiators and quaternary monomers), combined with a few reactivity cliffs (e.g., the pyridyl-substituted monomers **M53** and **M54**). We therefore expect our results to transfer to many other chemical reactions with moderately rough structure-property landscapes, such as nucleophilic substitutions and Suzuki couplings. For reaction types with a substantially rougher property landscape (e.g., Ni-catalyzed cross couplings), more data may be required for accurate predictions. Our results are also limited in that they are specific to reactions with three independent variables. However, many prediction problems from reactions with more independent variables are typically framed as three-dimensional problems (e.g., two substrates and a ligand for cross couplings while other variables such as temperature, solvent, duration, and precatalyst are kept constant or one-hot encoded).

## DISCUSSION

In summary, we have developed an automated, microscale synthesis platform for the on-demand synthesis of vast numbers of drug-like compounds. On the basis of a three-component reaction, now, 220,990 VL members are immediately accessible from 188 curated building blocks. We conducted and analyzed 50,000 reactions affording distinct products, generating, to our knowledge, the first public reaction dataset of this magnitude. We used these data to

train machine learning classifiers for reaction outcome with excellent accuracy within the PRIME library and notable generalization ability to new building blocks. The usefulness of the classifiers to prioritize synthesizable molecules was corroborated by prospective experimental validation. While many high-throughput campaigns in recent years have obtained models for reaction predictions, the unprecedented size of the present dataset allowed us to study the data requirements of different machine learning models under a set of interpolation and extrapolation tasks, revealing that reaction outcome prediction problems are best approached as interpolation tasks but that sparsely (~1%) sampling the combinatorial space suffices for training predictive models, substantiating prior conjectures (31). For either interpolation or extrapolation tasks, we find that simpler models outperform frequently used neural network-based approaches at dataset sizes previously available in the field. Combined, our observations suggest that the optimal approach to reliable reaction outcome predictions is building simple, local models with sparse datasets; elaborate model architectures and featurization will not suffice to construct general models of chemical reactivity, at least not in the absence of new modes of data collection.

## MATERIALS AND METHODS

### Library syntheses

#### Preparations

For plate-based library syntheses, building blocks were dissolved in DMSO (**I** and **M**) or in 9:1 DMSO/1 M aq. oxalic acid (**T**) at 55.6 mM (**I** and **M**) and 50 mM (**T**) concentrations, respectively. Building block solutions were stored at  $-20^{\circ}\text{C}$  in Greiner Bio-One Masterblock 96 well, 2 ml, PP, V-bottom plates (catalog no. 780270). Echo source plates (384-well PP 2.0 Microplate Echo qualified, catalog no. PP-0200) were prepared from stock solutions using an OT-2 liquid handler. For each initiator (16 total), two wells were filled with 63  $\mu\text{l}$  of DMSO solution and a third with 35  $\mu\text{l}$  of DMSO solution. For each monomer (12 total), three wells were filled with 63  $\mu\text{l}$  of DMSO solution and a fourth well was filled with 25  $\mu\text{l}$  of DMSO solution. The bottom row (P) of the source plate was filled with 1 M aq. oxalic acid. After the transfers, the Echo source plate was covered with a Beckman Microclime Environmental lid (catalog no. 001-5719) conditioned with DMSO. Source plates were used within ~1 hour. After source plate preparation, QC solutions of the building blocks were prepared using the OT-2 liquid handler by diluting aliquots 1:1000 with MeCN. QC solutions were submitted for LC-MS.

#### Automated synthesis

The Echo source plate was loaded into an automation system controlled by Cellario, served by a HighRes ACell robotic arm. Six Labcyte 384-well low dead volume Echo qualified microplates (catalog no. LP-0200; subsequently called “reaction plates”) were loaded onto the system. On the system, the source plate lid was removed, and the plate was subjected to centrifugation on a BioNex HiG automated centrifuge and loaded onto a Labcyte Echo 655 dispenser. The building block solutions were dispensed across the six reaction plates such that each reaction plate well received 990 nl of each **I** and **M** solution and 220 nl of oxalic acid solution. Initiators were dispensed along rows (e.g., **I1** would fill all wells in row A for all six plates). Monomers were dispensed in half a plate each (e.g., **M1** would fill the rectangle spanned by wells A3 and P12 in one of the six reaction plates). Columns 1, 2, 23, and 24 were always left empty (space for controls if the plates are used in assays). Upon completion of all transfers to a



reaction plate, it was removed from the Echo, sealed using an Agilent PlateLoc Thermal Microplate Sealer with Agilent Peelable Aluminium RT seals (catalog no. 24214-00), and moved to a HighRes Steristore D incubator at 60°C. Each plate was scheduled independently and was removed from the incubator after 3 hours.

During the incubation period, the source plate was removed from the system and charged with terminator solutions on the OT-2 liquid handler. For each terminator (10 total), four wells were filled with 65  $\mu$ l of the solution in 9:1 DMSO/1 M aq. oxalic acid and a fifth well with 30  $\mu$ l of the solution in 9:1 DMSO/1 M aq. oxalic acid. The source plate was moved back to the automation system before the end of the 3-hour incubation period.

After the incubation period, the source plate was delidded, subjected to centrifugation, and loaded into the Echo. The reaction plate was removed from the incubator, desealed using a Brooks XPeel Automated Plate Seal Remover, and loaded into the Echo. Terminator solutions were dispensed along columns such that each reaction plate well received 1100 nl of T solution (e.g., T1 would fill all wells in columns 3 and 13 of all six plates). After completion of the transfers, each reaction plate was sealed as described above and incubated at 60°C for 16 hours.

Note that the procedure describes one batch of 1920 compounds. Two batches can be prepared simultaneously by scheduling transfers for the second batch during incubation of the first batch and vice versa.

#### Preparation of analysis plates

The internal standard stock solution was prepared by dissolving 63 mg of fenofibrate in 100 ml of MeCN (LC-MS grade) with 0.1% formic acid. This stock solution was stored for up to 1 year at 4°C in the dark. The internal standard solution ( $c = 21.8 \mu\text{M}$ ) was prepared fresh by diluting 4.0 ml of the stock solution with 316 ml of MeCN (LC-MS grade). The automation system was loaded with the six reaction plates and six Eppendorf 384-well twin.tec PCR plates (catalog no. 0030128508; subsequently called “analysis plates”). The analysis plates were filled with 30-nl aliquots from the reaction plates on the Echo, retaining the layout. Reaction plates were sealed (for seals, see above) under argon for long-term storage at 4°C. Analysis plates were diluted with 30  $\mu$ l of internal standard solution per well using an Agilent BioTek EL406 Washer Dispenser and sealed in air (for seals, see above). Analysis plates were submitted to LC-MS characterization. Log files of Echo surveys and transfers were gathered to identify transfer errors (see the “Data filtering” section in the Supplementary Materials).

#### LC-MS analysis

Analyses were conducted on an Agilent 1290 Infinity II high-performance liquid chromatography system (Agilent Ltd., Germany) coupled with a Bruker maXis II mass spectrometer (Bruker Daltonics, Germany). The high-performance liquid chromatography system was equipped with a binary solvent delivery system, a diode array detector, a well plate autosampler, and reversed-phase 50 mm by 3.0 mm–inside diameter Agilent Zorbax Eclipse Plus C18 3.5- $\mu$ m columns in a dual-column setup. The mobile phase consisted of (A) acetonitrile with 0.1% formic acid and (B) water with 0.1% formic acid. Starting with isocratic conditions of 2% A at a flow rate of 0.6 ml min<sup>−1</sup> for 1 min, the gradient started at 2% A and was linearly increased to 98% in 5 min at a flow rate of 0.6 ml min<sup>−1</sup>, followed by isocratic conditions of 98% A at a flow rate of 0.6 ml min<sup>−1</sup> for 0.5 min. The total run time was 7 min. The injection volume was 4  $\mu$ l. The ultraviolet chromatogram was collected on a diode array detector using a microflow cell at 220, 254, 280, and 365 nm.

The mass spectrometer was operated in wide pass quadrupole mode, with the time-of-flight (TOF) data being collected between  $m/z$  50 and 1000 with a low-collision energy (in-source collision-induced dissociation) of 0 eV and an ion energy offset of 4 eV. The optimized source conditions were drying gas (10 liter min<sup>−1</sup>; nitrogen, 99.99% purity) at a temperature of 250°C; a nebulizer pressure of 1.6 bar; capillary and endplate voltages of 500 and 4500 V, respectively; a TOF flight tube voltage of 12,000 V; a reflection voltage of 3183 V; a pusher voltage of 1700 V; and a detector voltage of 2268 V. The resolving power of the instrument was around 45,000 with a 4-Hz spectra rate, depending on the sample concentration and peak width. The electrospray ionization TOF mass spectrometer was calibrated in the positive mode using a solution of sodium formate in isopropanol/water 1:1 [ $\text{Na}(\text{NaCOOH})_n^+$  cluster] on the enhanced quadratic algorithmic mode. Data were centroided during acquisition using independent reference lock-mass ion (hexamethoxyphosphazene,  $m/z$  322.0481) via the Lock-Mass interface to ensure mass accuracy and reproducibility.

The accurate mass and composition were automatically calculated using Data Analysis 5.3 software (Bruker Daltonics, Germany) with an automated scripting procedure. This scripting procedure obtained the accurate mass calibration, created specific extracted ion chromatograms (EICs) from given sum formulas, identified compounds, and created reports.

#### Building block syntheses

For information on building block syntheses, please refer to the Supplementary Materials.

#### Selection of experiments

Building block selection from the library was random with some constraints for experimental feasibility: Building blocks were randomly grouped into groups of 16 (initiators), 12 (monomers), and 10 (terminators). From all possible combinations of these groups, the experiment batches were drawn without replacement before the onset of the experimental campaign. If a building block was missing at the time of the experiment for any reason, it was substituted by a different, random, previously unused building block. In a few cases, when no adequate stock of an unused building block could be found, a random previously used building block was swapped in, giving rise to the 377 duplicate reactions in the dataset.

#### LC-MS data extraction

Processing of spectra was automated to determine the peak area in the EICs for all expected species, aggregating over different ionizations ( $\text{H}^+$  and  $\text{Na}^+$ ). For each reaction, at least nine EICs were extracted: the expected product (A), the known side products (B to H), and the internal standard. In addition, all EICs for sum formulae arising from potential deprotections were extracted if any expected product contained an *N*-boc or *N*-cbz protecting group, a *t*-Bu ester, or a trimethylsilyl (TMS)–protected alkyne. The summary report after processing contained the peak area and count for each sum formula that was searched. Simultaneously, a more extensive PDF report was automatically prepared per sample. It contained the full 2D chromatogram and all identified peaks with ultraviolet and mass spectra and was used for manual checks of automatically extracted data.

The summary report from LC-MS analysis contained the peak area and count for each sum formula that was searched. Summary reports were automatically extracted using a Python script. Contributions

from different sum formulae corresponding to the same product were aggregated (e.g., area of product **A** and area of product **A** minus TMS plus **H** if **A** contained a TMS-protected alkyne). The peak areas were normalized per sample by dividing through the peak area of the internal standard.

The LC-MS responses were scaled per product type by dividing by the 85th percentile for this product type. This is to account for the differential ionizability of the core scaffolds. We chose the 85th percentile for robust scaling because at higher percentile values, outliers start dominating the scaling.

### Data analysis

For additional information on data analysis, please refer to the Supplementary Materials.

### Training of models on truncated datasets

Truncated datasets with a reduced number of training points were constructed by applying the same split logic as for the main splits described in the “Data splitting” section in the Supplementary Materials by varying the train-test ratios. For the 1D splits, only one dimension was chosen for all splits (initiators) and, for the 2D splits, only one set of two dimensions was chosen (initiators and monomers). This deviation from the previous procedure was made to reduce the variance in the results. For each of the truncated splits, the full model fitting procedure as described above, including hyperparameter optimization, was conducted for seven model types: FFN/OHE, XGB/FP, XGB/FP + RDKit, XGB/QC, XGB/FP + QC, D-MPNN/CGR, and D-MPNN/CGR + RDKit. These model types were selected because they were previously found to perform well across all splits, as a baseline (FFN/OHE), or because they are frequently used in the literature (QC features). For a better comparison across dataset sizes, the achieved AUPRC was normalized using the chance level (i.e., a random classifier would have a normalized AUPRC of 0.0, whereas a perfect classifier would score 1.0). The Wilcoxon signed-rank test was used to confirm the statistical significance of improvement over another model. We use  $P = 0.05$  as the threshold for significance.

### Inference with the models

For inference, the correct model to apply is chosen by the number of previously seen building blocks. For prediction tasks with three previously seen building blocks, the 0D model is used. For prediction tasks with two previously seen building blocks, the 1D model is used. For prediction tasks with one previously seen building block, the 2D model is used. For prediction tasks with no previously seen building blocks, the 3D model is used. All predictions are made by a committee of nine models (from the nine training folds), aggregated by majority voting. For the 1D and 2D models only, the folds are not equivalent (see the “Data splitting” section in the Supplementary Materials). In this case, only the three models obtained on the three relevant folds are used for the prediction (i.e., for a prediction task where the initiator was not previously seen, only the three folds where the splitting dimension was initiators are used).

### Construction of the PRIME virtual library

In the final published version of the PRIME virtual library, we included all products theoretically accessible from the 66 valid initiators used to generate the reaction dataset and the 11 valid initiators used in validation, as well as the 70 valid monomers and 41 valid terminators used. This gives rise to 220,990 library members enumerated from

these building blocks using Python. A total of 84% of the PRIME VL is predicted to be synthesizable by our ML models. Note that the difference to the initially mentioned 236,000 library members is due to the removal of invalid building blocks over the course of the project. To add reaction predictions, we use the appropriate model, i.e., the 0D model for reactions where all building blocks were seen in the training data and the 1D model for reactions where the initiator was not seen (validation initiators).

### Property predictions for the PRIME virtual library

We used Schrödinger version 2023-3 to predict physicochemical properties of all library members. Members were prepared for predictions using LigPrep with Epik ionization. Property predictions were run using QikProp. The results are available from Zenodo at <https://zenodo.org/doi/10.5281/zenodo.13769229> as compressed SDF files containing chunks of 10,000 library members in a tar archive. The total number of structures exceeds 10,000 per chunk as the ionization states generated by LigPrep/Epik have separate entries.

### Further methods

For additional information on validation plates, training of machine learning models, and the virtual library, please refer to the Supplementary Materials.

### Supplementary Materials

#### The PDF file includes:

Supplementary Text  
Figs. S1 to S18  
Legends for movies S1 and S2  
Legends for data S1 to S4  
References

#### Other Supplementary Material for this manuscript includes the following:

Movies S1 and S2  
Data S1 to S4

### REFERENCES AND NOTES

- W. Beker, R. Roszak, A. Wołos, N. H. Angello, V. Rathore, M. D. Burke, B. A. Grzybowski, Machine learning may sometimes simply capture literature popularity trends: A case study of heterocyclic Suzuki–Miyaura coupling. *J. Am. Chem. Soc.* **144**, 4819–4827 (2022).
- M. Saebi, B. Nan, J. E. Herr, J. Wahlers, Z. Guo, A. M. Zurański, T. Kogej, P.-O. Norrby, A. G. Doyle, N. V. Chawla, O. Wiest, On the use of real-world datasets for reaction yield prediction. *Chem. Sci.* **14**, 4997–5005 (2023).
- I. Ugi, C. Steinbrückner, Über ein neues Kondensations-Prinzip. *Angew. Chem.* **72**, 267–268 (1960).
- F. Sutanto, S. Shaabani, C. G. Neochoritis, T. Zarganes-Tzitzikas, P. Patil, E. Ghonchepour, A. Dömling, Multicomponent reaction–derived covalent inhibitor space. *Sci. Adv.* **7**, eabd9307 (2021).
- Z. Wang, S. Shaabani, X. Gao, Y. L. D. Ng, V. Sapozhnikova, P. Mertins, J. Krönke, A. Dömling, Direct-to-biology, automated, nano-scale synthesis, and phenotypic screening-enabled E3 ligase modulator discovery. *Nat. Commun.* **14**, 8437 (2023).
- Y. L. Huang, J. W. Bode, Synthetic fermentation of bioactive non-ribosomal peptides without organisms, enzymes or reagents. *Nat. Chem.* **6**, 877–884 (2014).
- I. A. Stepek, T. Cao, A. Koetemann, S. Shimura, B. Wollscheid, J. W. Bode, Antibiotic discovery with synthetic fermentation: Library assembly, phenotypic screening, and mechanism of action of  $\beta$ -peptides targeting penicillin-binding proteins. *ACS Chem. Biol.* **14**, 1030–1040 (2019).
- D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher, A. G. Doyle, Predicting reaction performance in C–N cross-coupling using machine learning. *Science* **360**, 186–190 (2018).
- N. H. Angello, V. Rathore, W. Beker, A. Wołos, E. R. Jira, R. Roszak, T. C. Wu, C. M. Schroeder, A. Aspuru-Guzik, B. A. Grzybowski, M. D. Burke, Closed-loop optimization of general reaction conditions for heteroaryl Suzuki–Miyaura coupling. *Science* **378**, 399–405 (2022).

10. J. J. Dotson, L. van Dijk, J. C. Timmerman, S. Grosslight, R. C. Walroth, F. Gosselin, K. Püntener, K. A. Mack, M. S. Sigman, Data-driven multi-objective optimization tactics for catalytic asymmetric reactions using bisphosphine ligands. *J. Am. Chem. Soc.* **145**, 110–121 (2023).
11. D. F. Nippa, K. Atz, R. Hohler, A. T. Müller, A. Marx, C. Bartelmuß, G. Wuitschik, I. Marzuoli, V. Jost, J. Wolfard, M. Binder, A. F. Stepan, D. B. Konrad, U. Grether, R. E. Martin, G. Schneider, Enabling late-stage drug diversification by high-throughput experimentation with geometric deep learning. *Nat. Chem.* **16**, 239–248 (2024).
12. E. King-Smith, S. Berritt, L. Bernier, X. Hou, J. L. Klug-McLeod, J. Mustakis, N. W. Sach, J. W. Tucker, Q. Yang, R. M. Howard, A. A. Lee, Probing the chemical 'reactome' with high-throughput experimentation data. *Nat. Chem.* **16**, 633–643 (2024).
13. N. I. Rinehart, R. K. Saunthwal, J. Wellauer, A. F. Zahrt, L. Schlemper, A. S. Shved, R. Bigler, S. Fantasia, S. E. Denmark, A machine-learning tool to predict substrate-adaptive conditions for Pd-catalyzed C–N couplings. *Science* **381**, 965–972 (2023).
14. J. G. Hubert, I. A. Stepek, H. Noda, J. W. Bode, Synthetic fermentation of  $\beta$ -peptide macrocycles by thiadiazole-forming ring-closing reactions. *Chem. Sci.* **9**, 2159–2167 (2018).
15. G. Erös, Y. Kushida, J. W. Bode, A reagent for the one-step preparation of potassium acyltrifluoroborates (KATs) from aryl- and heteroarylhalides. *Angew. Chem. Int. Ed.* **53**, 7604–7607 (2014).
16. I. A. Stepek, "Synthetic fermentation as a platform for library synthesis, drug discovery and chemical outreach," thesis, ETH Zürich (2019).
17. S. Yu, H. Ishida, M. E. Juarez-Garcia, J. W. Bode, Unified synthesis of enantiopure  $\beta$ 2h,  $\beta$ 3h and  $\beta$ 2,3-amino acids. *Chem. Sci.* **1**, 637–641 (2010).
18. G. M. Fischer, M. K. Klein, E. Daltrozzi, A. Zumbusch, Pyrrolopyrrole cyanines: Effect of substituents on optical properties. *European J. Org. Chem.* **2011**, 3421–3429 (2011).
19. B. Hadimioglu, R. Stearns, R. Elson, Moving liquids with sound: The physics of acoustic droplet ejection for robust laboratory automation in life sciences. *J. Lab. Autom.* **21**, 4–18 (2016).
20. J. Götz, M. K. Jackl, C. Jindakun, A. N. Marziale, J. André, D. J. Gosling, C. Springer, M. Palmieri, M. Reck, A. Luneau, C. E. Brocklehurst, J. W. Bode, High-throughput synthesis provides data for predicting molecular properties and reaction success. *Sci. Adv.* **9**, ead32314 (2023).
21. A. F. Zahrt, J. J. Henle, S. E. Denmark, Cautionary guidelines for machine learning studies with combinatorial datasets. *ACS Comb. Sci.* **22**, 586–591 (2020).
22. A. Vall, S. Hochreiter, G. Klambauer, "BioassayCLR: Prediction of biological activity for novel bioassays based on rich textual descriptions," in *ELLIS Machine Learning for Molecule Discovery Workshop 2021* (ML4Molecules, 2021).
23. R. Joeres, D. B. Blumenthal, O. V. Kalinina, DataSAIL: Data splitting against information leakage. *bioRxiv* 566305 [Preprint] (2023); <https://doi.org/10.1101/2023.11.15.566305>.
24. E. Heid, W. H. Green, Machine learning of reaction properties via learned representations of the condensed graph of reaction. *J. Chem. Inf. Model.* **62**, 2101–2110 (2022).
25. T. N. Kipf, M. Welling, "Semi-supervised classification with graph convolutional networks," in *5th International Conference on Learning Representations (ICLR)* (2017).
26. W. L. Hamilton, R. Ying, J. Leskovec, Inductive representation learning on large graphs. *arXiv:1706.02216* (2018); <https://doi.org/10.48550/arXiv.1706.02216>.
27. K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea, A. Palmer, V. Settels, T. Jaakkola, K. Jensen, R. Barzilay, Analyzing learned molecular representations for property prediction. *J. Chem. Inf. Model.* **59**, 3370–3388 (2019).
28. Z. Xiong, D. Wang, X. Liu, F. Zhong, X. Wan, X. Li, Z. Li, X. Luo, K. Chen, H. Jiang, M. Zheng, Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *J. Med. Chem.* **63**, 8749–8760 (2020).
29. T. Chen, C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (ACM, 2016)*, pp. 785–794.
30. M. Aldeghi, D. E. Graff, N. Frey, J. A. Morrone, E. O. Pyzer-Knapp, K. E. Jordan, C. W. Coley, Roughness of molecular property landscapes and its impact on modelability. *J. Chem. Inf. Model.* **62**, 4660–4671 (2022).
31. P. Raghavan, B. C. Haas, M. E. Ruos, J. Schleinitz, A. G. Doyle, S. E. Reisman, M. S. Sigman, C. W. Coley, Dataset design for building models of chemical reactivity. *ACS Cent. Sci.* **9**, 2196–2204 (2023).
32. N. Mita, O. Tamura, H. Ishibashi, M. Sakamoto, Nucleophilic addition reaction of 2-trimethylsilyloxyfuran to *N*-gulosyl-C-alkoxymethylnitrones: Synthetic approach to polyoxin C. *Org. Lett.* **4**, 1111–1114 (2002).
33. A. Chiotellis, H. Ahmed, T. Betzel, M. Tanriver, C. J. White, H. Song, S. D. Ros, R. Schibli, J. W. Bode, S. M. Ametamey, Chemoselective  $^{18}\text{F}$ -incorporation into pyridyl acyltrifluoroborates for rapid radiolabelling of peptides and proteins at room temperature. *Chem. Commun.* **56**, 723–726 (2020).
34. M. Tanriver, Y.-C. Dzeng, S. Da Ros, E. Lam, J. W. Bode, Mechanism-based design of quinoline potassium acyltrifluoroborates for rapid amide-forming ligations at physiological pH. *J. Am. Chem. Soc.* **143**, 17557–17565 (2021).
35. S. Da Ros, "Kinetic and mechanistic investigation of the potassium acyltrifluoroborate ligation reaction," thesis, ETH Zurich (2018).
36. D. Wu, N. A. Fohn, J. W. Bode, Catalytic synthesis of potassium acyltrifluoroborates (KATs) through Chemoselective cross-coupling with a bifunctional reagent. *Angew. Chem. Int. Ed.* **58**, 11058–11062 (2019).
37. D. Wu, "Synthesis of potassium acyltrifluoroborates with transfer reagents and their applications to 3D photopatterning of hydrogels," thesis, ETH Zurich (2019).
38. A. M. Dumas, J. W. Bode, Synthesis of acyltrifluoroborates. *Org. Lett.* **14**, 2138–2141 (2012).
39. S. M. Liu, D. Wu, J. W. Bode, One-step synthesis of aliphatic potassium acyltrifluoroborates (KATs) from organocuprates. *Org. Lett.* **20**, 2378–2381 (2018).
40. S. Liu, "Efforts toward the synthesis of potassium acyltrifluoroborates," thesis, ETH Zurich (2018).
41. S. M. Liu, D. Mazunin, V. R. Pattabiraman, J. W. Bode, Synthesis of bifunctional potassium acyltrifluoroborates. *Org. Lett.* **18**, 5336–5339 (2016).
42. D. Mazunin, "Formation and functionalization of hydrogels with the potassium acyltrifluoroborate (KAT) ligation," thesis, ETH Zurich (2016).
43. A. O. Gálvez, C. P. Schaack, H. Noda, J. W. Bode, Chemoselective acylation of primary amines and amides with potassium acyltrifluoroborates under acidic conditions. *J. Am. Chem. Soc.* **139**, 1826–1829 (2017).
44. H. Noda, G. Erös, J. W. Bode, Rapid ligations with equimolar reactants in water with the potassium acyltrifluoroborate (KAT) amide formation. *J. Am. Chem. Soc.* **136**, 5611–5614 (2014).
45. M. K. Jackl, A. Schuhmacher, T. Shiro, J. W. Bode, Synthesis of *N,N*-alkylated  $\alpha$ -tertiary amines by coupling of  $\alpha$ -aminoalkyltrifluoroborates and Grignard reagents. *Org. Lett.* **20**, 4044–4047 (2018).
46. A. Schuhmacher, S. J. Ryan, J. W. Bode, Catalytic synthesis of potassium acyltrifluoroborates (KATs) from boronic acids and the thioimide KAT transfer reagent. *Angew. Chem. Int. Ed.* **60**, 3918–3922 (2021).
47. D. Schauenburg, "Potassium acyltrifluoroborates (KATs) for bio-macromolecular chemistry," thesis, ETH Zurich (2021).
48. Y.-L. Huang, "Synthetic fermentation of chemical libraries without organisms, enzymes or reagents," thesis, ETH Zurich (2014).
49. A. R. Todorov, M. Nieger, J. Helaja, Tautomeric switching and metal-cation sensing of ligand-equipped 4-hydroxy-/4-oxo-/1,4-dihydroquinolines. *Chem. A Eur. J.* **18**, 7269–7277 (2012).
50. S. Miyazaki, Y. Kurosaki, M. Inui, M. Kishida, K. Suzuki, M. Izumi, K. Soma, A. Pinkerton, "Oxazepine derivatives having trap inhibitory activity," WO2018119449A1 (2018).
51. X. Zhang, S. Chang, D. Ye, Y. Wang, Q. Li, Y. Liu, H. Sun, Z. Liu, J. Yang, L. Li, "Nitrogen-substituted aminocarbonate thiophene compound and use thereof," WO2022100625A1 (2022).
52. Y. Shirasaki, H. Miyashita, M. Nakamura, J. Inoue, "Alpha-ketoamide derivative, and production method and use thereof," WO2005056519A1 (2005).
53. A. Fawcett, M. J. Keller, Z. Herrera, J. F. Hartwig, Site selective chlorination of C(sp<sup>3</sup>)–H bonds suitable for late-stage functionalization. *Angew. Chem. Int. Ed.* **60**, 8276–8283 (2021).

**Acknowledgments:** We thank C. Brocklehurst and K. Tan for fruitful discussions. We acknowledge M. Meier of the Molecular and Biomolecular Analysis Service (MoBiAS) of the Department of Chemistry and Applied Biosciences at ETH Zürich for mass spectrometry. We acknowledge E. Riegler, T. Booi, and D. Keller of NEXUS at ETH Zürich for help with high-throughput experiments. We are grateful to G. Erös, A. Dumas, S. Da Ros, A. Gálvez, A. Schuhmacher, M. Tanriver, D. Schauenburg, P. Schilling, S. Liu, D. Wu, D. Mazunin, F. Masero, J. Hubert, T. Shiro, and Y.-C. Dzeng for contributing materials. **Funding:** This work was supported by Novartis Global Scholars Program (to J.W.B.) and ETH Zürich (to J.W.B.). **Author contributions:** Conceptualization: J.G., I.A.S., Y.-L.H., and J.W.B. Methodology: J.G., E.R., I.A.S., Y.T., Y.-L.H., L.B., and B.R. Software: J.G. and L.B. Formal analysis: J.G. Investigation: J.G., E.R., I.A.S., Y.T., Y.-L.H., and L.B. Data curation: J.G., E.R., I.A.S., Y.T., and L.B. Visualization: J.G. Validation: J.G., E.R., and L.B. Funding acquisition: J.W.B. Project administration: J.G., B.R., and J.W.B. Supervision: J.W.B. Writing—original draft: J.G. and J.W.B. Writing—review and editing: J.G., E.R., I.A.S., Y.T., Y.-L.H., L.B., B.R., and J.W.B. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Property predictions for the virtual library are available from Zenodo at <https://zenodo.org/doi/10.5281/zenodo.13769229>. The code is also available from GitHub at <https://github.com/jugoetz/library-generation> (data processing and analysis) and <https://github.com/jugoetz/synform-predictions> (machine learning). Versions archived at the time of publication are available from Zenodo at <https://doi.org/10.5281/zenodo.11121448> and <https://doi.org/10.5281/zenodo.11121455>.

Submitted 8 February 2025

Accepted 25 April 2025

Published 28 May 2025

10.1126/sciadv.adw6047