

Cite as: S. Lin *et al.*, *Science*
10.1126/science.aar6236 (2018).

Mapping the dark space of chemical reactions with extended nanomole synthesis and MALDI-TOF MS

Shishi Lin,¹ Sergei Dikler,² William D. Blincoe,³ Ronald D. Ferguson,¹ Robert P. Sheridan,⁴ Zhengwei Peng,⁴ Donald V. Conway,⁵ Kerstin Zawatzky,³ Heather Wang,³ Tim Cernak,^{6*} Ian W. Davies,^{7†} Daniel A. DiRocco,⁷ Huaming Sheng,^{3‡} Christopher J. Welch,^{3‡§||} Spencer D. Dreher^{1‡}

¹Chemistry Capabilities and Screening, Merck & Co., Inc., Kenilworth, NJ 07033, USA. ²Bruker Daltonics Inc., Billerica, MA 01821, USA. ³Analytical Research & Development, Merck & Co., Inc., Rahway, NJ 07065, USA. ⁴Modeling and Informatics, Merck & Co., Inc., Kenilworth, NJ 07033, USA. ⁵Discovery Sample Management, Merck & Co., Inc., Kenilworth, NJ 07033, USA. ⁶Discovery Chemistry, Merck & Co., Inc., Boston, MA 02115 USA. ⁷Process Research & Development, Merck & Co., Inc., Rahway, NJ 07065, USA.

*Present address: Department of Medicinal Chemistry, Department of Chemistry, Program in Chemical Biology, University of Michigan, Ann Arbor, MI 48109, USA. †Present address: Princeton Catalysis Initiative, Princeton University, Princeton, NJ 08544 USA.

‡Corresponding author. Email: huaming.sheng@merck.com (H.S.); christopher@welchinnovation.com (C.J.W.); spencer_dreher@merck.com (S.D.D.)

§Present address: Indiana Consortium for the Analytical Sciences, Indianapolis, IN 46202, USA. ||Welch Innovation, LLC, Cranbury NJ 08512, USA.

Understanding the practical limitations of chemical reactions is critically important for efficiently planning the synthesis of compounds in pharmaceutical, agrochemical and specialty chemical research and development. However, literature reports of the scope of new reactions are often cursory and biased toward successful results, severely limiting the ability to predict reaction outcomes for untested substrates. We herein illustrate strategies for carrying out large scale surveys of chemical reactivity using a material-sparing nanomole-scale automated synthesis platform with greatly expanded synthetic scope combined with ultra-high throughput (uHT) matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS).

The field of synthetic organic chemistry has spawned a vast array of creative reactions that can be logically combined to prepare nearly any molecule. However, efficient selection of the precise reaction sequence that leads to a particular product remains a challenge, as poorly understood substrate-specific interactions often necessitate laborious screening of combinations of catalysts, reagents and conditions. High-throughput experimentation (HTE) chemistry facilitates these investigations by increasing the pace (*1–7*), and an emerging strategy uses machine-learning (ML) data generated with these screening tools to create predictive models for specific problems (*8, 9*). Data mining surveys of existing published or proprietary databases containing information on hundreds of millions of reactions have been somewhat encouraging (*10–14*), but a pervasive, longstanding bias toward reporting successful results limits the utility of this information for model-building. Additionally, these data have not been collected under controlled experimental conditions, and most of the substrates are not representative of the complexity that arises in applied synthetic problems. Hence, no existing large, structured datasets meet all of the requirements for effective reaction modeling. Consequently, intentional surveys are required to systematically map reactivity patterns, with the rapid identification of inaccessible regions, the dark space of chemical reactivity, being particularly important for focusing subsequent ML investigations.

Chemical reactivity space is vast, even for a single synthetic transformation, with permutations of possible substrate structures, catalysts and reaction conditions soon proliferating into unmanageable experimental complexity. Large scale reactivity surveys have thus far been limited by the throughput of reaction screening and analysis technologies. Additionally, the scarcity and high cost of relevant complex substrates requires that screening be carried out on as small a scale as possible. Existing miniaturized HTE chemistry tools have only worked under a narrow range of conditions, and have thus far not been applicable to the vast majority of high quality synthetic protocols. In this work, we describe how engineering advances in automated, miniaturized reaction experimentation and the use of fast MALDI-MS for routine reaction analysis affords an opportunity to rapidly survey the chemical reactivity landscape. We demonstrate how these advances can be used to systematically reveal diverse sources of reaction failure within large substrate sets, information that is largely missing from current chemistry data sources, representing an important advance toward broadly effective reaction profiling.

Extending nanomole synthesis for systematic reaction profiling

Our interest in reaction profiling has grown from an internal effort termed the C-N Coupling Initiative, which aims to systematically build tools to completely understand and leverage

a given synthetic transformation. A recent examination of our electronic notebooks (ELN) shows C-N coupling to be a hit or miss affair (~10,000 examples, ~35% failure rate) that, despite its potential to render drug-like molecules, may be underutilized due to its higher demand for optimization than, for example, the Suzuki reaction (~50,000 example, 18% failure). Attempts to use this ELN data set for creating predictive models have been disappointing, as the data are fractured, with many different classes of amines that would be expected to have different reactivity patterns (primary, secondary, NH-heterocyclic, amides, sulfonamides) while also containing deep channels of project-driven substrate similarity in the aryl halide coupling partners. Additionally, these data are collected under a wide array of synthetic protocols, with specific conditions seldom being repeated. Given this lack of a suitable dataset, a real and urgent goal of applied research is to increase the understanding of C-N coupling reactions by building an experimental process suitable for fast reaction mapping. Clearly, this systematic reaction profiling capability will also be of general value for surveying other reaction types where available knowledge is scarce.

Failure in metal-catalyzed reactions can result from a variety of causes, including inhibition of catalysts due to substrate binding, competitive side-reactions, local steric and electronic interactions and substrate decomposition caused by harsh reaction conditions. In addition, small perturbations in molecular structure may result in changes in reactivity, especially for complex substrates. We perceived that mapping reaction failure arising from this diverse patchwork of mechanisms would require a large collection of structurally diverse analogs as well as a tremendous number of experiments. Emerging nanomole scale tools have the material-sparing miniaturized scale and ultra-high-throughput (uHT) required for broad profiling, however, existing experimental strategies involve simplified engineering and are inadequate for examining most high-quality synthetic protocols. The first generation nanomole synthesis platform that we developed (1) was limited to plastic compatible, homogenous, ambient temperature reactions in low volatility polar, aprotic solvents, encompassing only a small subset of organic transformations. The recent work by Perera *et al.* (2), extends nanomole synthesis capabilities in a flow chemistry format, allowing heating and the use of diverse solvents; however, this format also requires that reactions remain homogeneous, necessitating the use of very high dilution and limiting study to extremely fast kinetics that can keep up with serial HPLC analysis. Rather than accepting such constraints, we systematically engineered and validated new plate-based nanomole synthesis tools with a general ability to carry out a wide variety of typical synthetic reactions. We identified effective chemically-compatible glass microplate reactors, fast 384-tip dosing of

reagent solutions in volatile solvents, and designed aluminum sealing blocks that can retain volatile solvents on heating. In addition, we used LabRam™ resonant acoustic mixing to both agitate reactions and to create milky slurries of solid inorganic bases that can be added to reactions by parallel liquid handling. Finally, we created a nanomole scale photochemistry tool to enable reaction evaluation in the rapidly-growing field of photoredox catalysis. We describe the use of these tools in reaction profiling experiments below, but they are expected to benefit extensions to nanoscale chemistry/ biological evaluation as well (15).

MALDI-TOF MS reaction analysis

Owing to the sheer number of experiments necessary, surveying chemical reactivity can be limited by the rate of analytical measurement. Despite considerable recent progress, HPLC-MS assays typically used for such screening are constrained by the fastest speed with which samples can be mechanically taken up and injected (~10 s/sample) (16). Spectroscopic techniques can be faster, but often lack the resolving power needed to study a diverse collection of molecules. High throughput mass spectrometry provides excellent mass to charge ratio (*m/z*) separation of individual products (17–20) and MALDI-MS has been intensively optimized in recent years for rapid and robust, molecule-specific MS imaging of biological tissue slices. The new generation of MALDI-TOF instruments equipped with a 10 kHz scanning beam laser, substantially faster X,Y stage and faster plate loading/unloading cycle dramatically improves sample throughput to hundreds of thousands of measurements per day. We were therefore intrigued by the possibility of applying this fast analysis approach (several samples/s) to high throughput reaction screening. Although MALDI is commonly used for high throughput analysis, applications are typically limited to either quantitative biochemical assays that feature a single substrate/product within each well (21–28) or chemistry discovery using a MS-active tag or special MALDI plate to ensure MS detection of all labeled products (29–31). Neither of these approaches enabled the rapid, label-free analysis of a diverse collection of products that we required.

Direct MS analysis of small molecules can be vulnerable to interference from reaction components that would typically be removed by chromatography. To analyze samples by MALDI-TOF, we found a suitable ionization matrix and instrument settings afforded excellent performance for diverse pharmaceutically relevant products, with the addition of an ionizing internal standard helping to normalize the effects of different reaction conditions on product ionization and MALDI spot formation. Screening and optimization of MALDI ionization matrices was carried out using a library of drug-like informer compounds (32), leading to the identifica-

tion of α -cyano-4-hydroxycinnamic acid (CHCA) as a generally preferred matrix (4 mg/ml in 50% H₂O/ACN with 0.1% TFA) affording optimal S/N across a wide range of protonated model compounds. A general protocol including MALDI plate preparation, automated data acquisition and data processing was established for MALDI analysis, in which 175 nL of quenched reaction mixture in DMSO with added internal standard was directly spotted on a HTS MALDI target plate in 1536 format using a 16-channel positive displacement liquid handling robot. In our experiments, the spotting time for a single plate is about 30 min, but MALDI plate preparation time can be shortened using a liquid handling robot with 1536-channel pipetting head. The resulting DMSO on the target plate can be evaporated either under vacuum at 0.01 bar (2 min) or at atmospheric pressure (1 hour). 150 nL of CHCA matrix solution is then applied onto the dried reaction spots through the same liquid dispensing procedure. The readout speed for each 1536 MALDI plate ranges from 8-11 min depended on the number of laser shots per spectrum. All automated MALDI runs were set up, triggered and processed using dedicated uHTS software for MALDI-TOF instruments (Bruker MALDI PharmaPulse 2.0). In the whole molecule analysis experiment a custom processing script was used to simultaneously extract nearly 400 masses, peak intensities and peak areas for reaction products and the internal standard into a single spreadsheet in addition to standard MALDI PharmaPulse processing. A signal to noise-ratio of 5:1 was used in the software to extract the final MALDI-MS data.

uHT fragment poisons analysis

In order to test our MALDI analytical approach for general reaction profiling, we began with an analytically simplified screening protocol developed by Glorius and co-workers for identifying reaction poisons, i.e., functional groups and reagents that interfere with a given reaction (Fig. 1) (33, 34). Since the same reaction is investigated in each experimental well, a single product standard can be used throughout, greatly simplifying analysis. We realized that the throughput of the MALDI technique could enable screening of much larger sets of steric and electronic functional variants and higher-order polyfunctional molecular fragments to precisely identify specific deleterious binding motifs and cooperative interactions. At the same time, we could evaluate the ability of the MALDI system to provide accurate product quantification in the presence of different reaction conditions and a wide variety of additives.

We selected 4 different C-N coupling conditions (Fig. 1A, Ir/Ni and Ru/Ni photoredox (35) as well as Pd and Cu conditions (32) that were previously demonstrated to perform well using our chemistry informer libraries approach. Each method was evaluated using the new nanomole experimental

platforms in the presence of 383 polar molecular fragments (with one control reaction containing no additive) ranging from simple compounds containing a single functional group (such as amides, esters and N-heterocycles) to more complex polyfunctional compounds. The resulting 1536 fragment additive reactions were analyzed in a single MALDI experiment in 11 min, with the settings described above and using a deca-deuterated product analog, added to each well after reaction completion, as an internal standard (Fig. 1A). The reactions were also analyzed using a conventional 2-min UPLC-MS method. The raw MALDI product response shows only marginal correlation with the UPLC-MS/UV data for the same reactions (extracted ion count, EIC, $R^2 = 0.24$, UV210, $R^2 = 0.16$), however, normalization of the MALDI signal with the closely related internal standard signal affords quite good correlation (EIC, $R^2 = 0.85$) showing that chemistry-specific effects on MALDI signal intensity can be mitigated by the use of internal standards (Fig. 1B). The normalized MALDI product response metric, ranked from high to low for all 4 methods, is shown in Fig. 1C, with the corresponding UPLC-MS EIC data shown below for comparison. This analytical approach is similar to the application of MALDI MS for high throughput screening of enzyme inhibitors, but demonstrates here that MALDI, in conjunction with an appropriate product standard, can be used for routine reaction screening in the presence of a diversity of interfering elements.

To simplify MALDI-based assessments in fragment additive screening, we binned the normalized MALDI product response for each reaction into poisons (>50% signal knockdown) and non-poisons (<50% signal knockdown) for each fragment and synthetic method (Fig. 2). Across all 4 methods, the MALDI assignment matched the UPLC-MS EIC signal ~95% of the time and UPLC-UV210 data ~88% of the time (we found that UPLC-UV210 nm data often show peaks that interfere with the product peak). The fragment screening data showed the expected outcome that each of the synthetic methods is susceptible to poisoning by particular single functional groups (see figs. S19 and S20 for complete lists). We were also able to observe that when these single-functional poisons are incorporated as parts of polyfunctional fragments, reaction poisoning almost always occurs (>90% of the time). Most interestingly, polyfunctional fragments composed of functional groups that are not themselves poisons often (~50% of the time) act as poisons as well. The structural details of this poisoning are very specific. For example, we observed cases where the combination of N-heterocycle and ketone strongly poisons a given catalytic system, while other similar structural combinations do not (Fig. 2). This work gives some indication of how screening large permutations of multifunctional structures can identify precise motifs that lead to reaction failure. We were gratified to find that with large number of experiments, the effects of false negatives

and positives on identifying trends are to some degree offset by the internal validation coming from using multiple fragments that are structurally related. The fragment based analysis herein reveals Cu to be remarkably more resilient to single and poly-functional poisoning than the other methods evaluated, which should support its wider use in complex, densely functionalized substrates. The Ir/Ni and Ru/Ni photoredox methods were very similar to each other in their inhibition profiles, and also generally performed very similarly to Pd. Most importantly, the specific problematic functional arrays that poison each method can be roughly mapped, and subsequent detailed nanomole chemistry exploration with structural variants can provide focused data for effective ML predictive modeling [as in ref. (8)].

uHT simplest partner whole molecule analysis

We next investigated the much more challenging use of MALDI MS to characterize the effects present in large whole molecule substrate sets (Fig. 3). Whereas the fragments approach is useful for rapid identification of problematic functionality, classifying local steric and electronic as well as bulk molecular effects such as solubility can only be accomplished using whole molecule substrates. Previous work in our labs explored the effects of pharma-relevant whole molecule informer compounds (32) on the performance of different synthetic methods. Though useful to begin to systematically evaluate different synthetic methods, the 18 compounds in these test sets do not provide enough structural diversity to begin to assign general causes of reaction failure. In addition, the impact of amines in complex reactivity space was not previously taken into account. Hence, we envisioned a much larger virtual array of 192 N-heterocycle-containing aryl bromides crossed with 192 cyclic secondary amines (Fig. 3A), both of increasing molecular complexity, representing 36,864 unique potential products. This array, if evaluated with the four aforementioned C-N coupling protocols, would result in 147,456 experiments. Although experimental assessment of all substrates and protocols is conceivable, we explored a systematic “simplest partner” approach to study the isolated structural effects of individual building blocks in this space. In this protocol, each bromide is subjected to reaction with the simplest amine in the set and each amine with the simplest bromide. Applying the four different synthetic methods to this 384-substrate test set affords a total of 1536 experiments, or approximately 1% of the experimental space. In order to ensure that every reaction in the set would have a strong MALDI-MS response, the simplest coupling partners were chosen to incorporate a basic nitrogen atom, thereby ensuring MS detectability in the positive ion mode. While measuring products in negative ion mode was not the focus of this work, analysis of small molecules by MALDI-TOF in

negative ion mode can be successfully achieved when appropriate sample preparation methods are used (36–38). Clearly, this strategy would be poorly suited for the synthesis of hydrocarbons or other species having poor MS ionization properties.

Traditional exploration of complex substrate space using HPLC analysis requires the preparation of product standards to obtain response factors for quantification and to confirm product identity for every new molecule of interest, an untenable proposition for our envisioned experiment. The prospect of adding an internal standard that is tailored to each product for MALDI normalization is equally daunting. We reasoned that in our search for causes of reaction failure, a focus on common structural reactivity trends rather than absolute product yields might afford valuable insights without the traditional need for individual product standards. Also, we hoped that the use of a single internal standard in all wells, while clearly not controlling for structure-based variations in MALDI signal intensity, might still be useful for normalizing the effects of MALDI spotting variability, thereby vastly simplifying experimental execution. Hence, we endeavored to use rapid MALDI responses for 384 different products coupled with parameterized and clustered molecular descriptors (such as functional group counts, steric hindrance, and bulk properties such as H-bond donors/acceptors and polar surface area) to identify fundamental incompatibilities between given sets of reaction conditions and the diverse structural properties within an enormous compound set. Once identified, such reactivity trends can be confirmed by conventional analytical methods. The MALDI signal for all 1536 reactions was acquired using a single internal standard for normalization across all reactions, as well as a conventional 2-min UPLC-MS analysis for comparison. Two-dimensional scatter plots for all 1536 reactions (Fig. 3B) reveal that the normalized product MALDI responses are poorly correlated vs UPLC-MS (EIC, $R^2 = 0.33$) and UPLC-UV data (total wavelength chromatogram, TWC, $R^2 = 0.30$, we found an averaged wavelength to be more useful in this diverse compounds set than a single wavelength, such as UV210). However, we found that we could use the MALDI data to create binary thresholds for reaction success/ failure that permitted us to identify general structural entities that correlate with reaction failure within a large whole molecule set. To this end, we set an arbitrary pass/ fail threshold of 20% of the average MALDI value across the entire set. When the binned data for each synthetic method are clustered as % of failed reactions for different structural parameters that have at least 10 examples within the set (Fig. 4A), the data from the normalized MALDI experiment show essentially the same trends revealed in the UPLC-MS data (both EIC and TWC responses), suggesting that MALDI can be used to stack-rank the structural entities

that most often cause reaction failure for each set of conditions (see figs. S26 to S33 for complete lists). Again, any of these identified problematic structural parameters can be promoted for higher-order systematic ML studies.

Although lists of potential poisons for specific synthetic methods are useful, understanding the comparative reactivity of different synthetic methods for problematic functionality is arguably even more important. Figure 4B shows how the clustered binary pass/ fail data for all 4 synthetic methods can be combined into a single graph that reveals structural entities that are generally problematic for bromides and amines across all methods. The data can serve as a valuable map for academic research focused on creating new methods that can overcome these common problems. At the same time this graph also reveals which method affords the lowest failure rate for each parameter. Figure 4B clearly shows the synthetic advantage of the Cu system in nearly all parameters used to describe aryl bromides (the most problematic are listed in Fig. 4C), which again substantiates its value in complex synthesis; at the same time, for a number of parameters within the amine structural space (Fig. 4B, in red) Pd is the best catalyst. Diving deeper (Fig. 4D), we can identify very specific structural features within amines that are problematic for Cu and where Pd can provide a synthetic advantage. Particularly interesting is the reproducibly higher performance of Pd versus Cu for 3-substituted-4-N-carbonyl piperazines and 3,4-N-heterocycle-appended piperazines, revealing subtle conformational effects several bonds removed from the reactive amine that are not considered in the current understanding of C-N coupling reactivity. These systems were explored at larger reaction scale (10 μ mol, 40x scale-up) with additional structural examples to confirm the trends (see figs. S37 to S39 for further details on Cu versus Pd amine trends).

The fragment additive and whole molecule studies both point to Cu having a significant advantage over the other methods examined in this study with respect to single and polyfunctional group tolerance and scope. These observations hold with aryl bromides, but Cu also appears to have specific liabilities in the amine structural space. This study demonstrates that uHT fragment and whole molecule approaches can provide complementary reactivity information, and both approaches will likely be inextricably linked in the future of predictive chemistry. Given the observed importance of precisely defined structures evident in this work, and a shortage of diverse test substrates, chemists will likely alternate between the two approaches to refine their maps of holistic reactivity effects. The most important message from this work is that, rather than accurately determining reaction yields, broad uHT mapping of substrate space with no isolated product standards can reveal specific molecular interactions using adequately discriminating binary analytical

approaches that simply separate successes from failures. The reactivity trends identified by MALDI in these experiments were confirmed with conventional UV/ MS scoring, suggesting that MALDI alone can be used to uncover results previously only accessible by slower analytical methods. Looking forward, we hope that the convenience, robustness and speed of MALDI will allow for creation of large structure/ mass response data sets to enable quantitative yield prediction to further increase value of the MALDI-MS approach.

Full factorial whole molecule space

The remaining 99% of the full factorial whole molecule cross-over substrate coupling space may contain additional information on higher order bi-molecular interactions that can also influence reaction performance. The simplest partner approach can be used to triage the experiments required to map the enormity of the remaining space. We reasoned that if either compound in a complex pair performs poorly in the simplest partner test, then structural poisoning will be likely in the more complex combination. Likewise, when both partners perform well with simple partners, the combination is likely to work well together. We investigated 288 experiments in this cross-over space and found that we could assign >50% of the space into hits/ misses (with 90% accuracy, see SI for details). The remaining space, with substrates having middling MALDI responses, could not be effectively resolved. This pruning approach allows us to chart complex bimolecular space using the minimal coverage provided by the simplest partner experiments, thereby eliminating >70,000 experiments that do not need to be performed.

Toward predictive informatics

The miniaturized uHT reaction engineering and MALDI TOF-MS analytical advancements described in this work enable generation of large structured datasets that pinpoint problematic structural elements within complex substrate space. Identified problems can then be promoted for more detailed mapping using uHT experimentation coupled with atomic and quantum molecular physical descriptors to enable structure-based machine-learning. In this work we have run over 3000 experiments studying the effects of just 4 synthesis conditions on a relatively narrow area of chemistry space (the coupling of cyclic secondary amines and N-heterocycle containing aryl bromides). The real power in this approach will harness the tremendous analytical speed of MALDI analysis for the iterative evaluation of large substrate arrays against diverse catalysts, bases, solvents, reagent stoichiometries and temperatures, enabling big data chemistry informatics in the search for general solutions to problematic areas in organic synthesis.

A persistent focus on revealing structural limitations in chemistry is a win-win scenario for academic researchers and

synthetic practitioners alike. Dark space that remains inaccessible across synthetic methods becomes fodder for new experimental research of known value, and specific knowledge of limitations will lead to predictivity and understanding that will increase the speed and success of the design-make-test cycle, helping to remove synthetic problem-solving from the critical path. Even with optimally efficient tools and strategies in place, mapping the entire landscape of useful chemical reactivity is currently beyond the reach of any single organization. Given the tremendous enabling value of such a massive survey, a precompetitive public-private partnership to address this gap might even be an achievable and worthwhile goal.

Methods summary

The specific extended nanomole-scale chemistry protocols used for the 4 described C-N coupling synthetic methods (Ir/Ni and Ru/Ni photoredox, Cu and Pd) as well as the MALDI-MS and comparative UPLC-MS analysis protocols are described in detail in the Supplementary Information. A brief summary is provided below.

Description of extended nanomole-scale chemistry platform

Nanomole-scale chemistry reactions, to this point, have been run using plastic plates with low-volatility, plastic compatible solvents (polar aprotics, DMSO and NMP) and homogeneous components (bases and catalysts) at room temperature. In this work, reaction engineering advancements have enabled a much wider scope of potential reaction conditions. The use of glass 1536 well-plates and rapid 384-tip pipetting enables use of volatile, non-plastic compatible solvents (dioxane demonstrated, herein) and substrates (piperidine). Development of a resonant acoustic LabRam grinding protocol to produce long-lasting slurries that can be dosed with liquid handling robotics has enabled the use of heterogeneous inorganic bases (Cs_2CO_3 and K_3PO_4). An aluminum reactor block was designed that can provide a high-quality seal to prevent solvent loss, and a heating mechanism was devised that can be equipped to heat the aluminum block inside the LabRam, which provides a robust mechanism for parallel reaction agitation. Finally, a nanomole-scale photochemistry tool was created using a modified aluminum reactor with an acrylic plastic bottom that allows uniform light penetration.

Description of MALDI-TOF analysis

All reaction mixtures were quenched and diluted to standard UPLC-MS concentration. These quenched reactions were then spotted using a Mosquito HTS liquid handling robot on Bruker HTS MALDI targets with barcodes (1.0 mm thickness) in 1536 format [HTS MALDI plate 1.0 mm, BC, Part# 1833280]. The targets were mounted on Bruker HTS MALDI

adapters (Part# 1847571). 175 nL of reaction mixture was deposited and allowed to dry followed by depositing 150 nL of a 4 mg/mL solution of α -cyano-4-hydroxycinnamic acid in 0.1% TFA, 50% acetonitrile/ H_2O . These targets were analyzed on a Bruker Rapiflex MALDI-TOF/TOF.

Description of comparative UPLC-MS analysis

The quenched reactions were also monitored using a Waters Acquity UPLC I-Class system (Waters Corp., Milford, MA, USA) equipped with a binary pump, a FTN sampler, column manager, a photodiode array detector, SQD detector 2 with electrospray ionization (ESI) source in the positive mode and MassLynx® software. Separations were performed on a Waters CORTECS UPLC C18+ column (30 \times 2.1mm, 1.6 μm).

REFERENCES AND NOTES

1. A. Buitrago Santanilla, E. L. Regalado, T. Pereira, M. Shevlin, K. Bateman, L. C. Campeau, J. Schneeweis, S. Bertritt, Z. C. Shi, P. Nantermet, Y. Liu, R. Helmy, C. J. Welch, P. Vachal, I. W. Davies, T. Cernak, S. D. Dreher, Nanomole-scale high-throughput chemistry for the synthesis of complex molecules. *Science* **347**, 49–53 (2015). [doi:10.1126/science.1259203](https://doi.org/10.1126/science.1259203) Medline
2. D. Perera, J. W. Tucker, S. Brahmabhatt, C. J. Helal, A. Chong, W. Farrell, P. Richardson, N. W. Sach, A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow. *Science* **359**, 429–434 (2018). [doi:10.1126/science.aap9112](https://doi.org/10.1126/science.aap9112) Medline
3. S. W. Kraska, D. A. DiRocco, S. D. Dreher, M. Shevlin, The Evolution of Chemical High-Throughput Experimentation To Address Challenging Problems in Pharmaceutical Synthesis. *Acc. Chem. Res.* **50**, 2976–2985 (2017). [doi:10.1021/acs.accounts.7b00428](https://doi.org/10.1021/acs.accounts.7b00428) Medline
4. V. I. Martin, J. R. Goodell, O. J. Ingham, J. A. Porco Jr., A. B. Beeler, Multidimensional reaction screening for photochemical transformations as a tool for discovering new chemotypes. *J. Org. Chem.* **79**, 3838–3846 (2014). [doi:10.1021/jo500190b](https://doi.org/10.1021/jo500190b) Medline
5. K. Troshin, J. F. Hartwig, Snap deconvolution: An informatics approach to high-throughput discovery of catalytic reactions. *Science* **357**, 175–181 (2017). [doi:10.1126/science.aan1568](https://doi.org/10.1126/science.aan1568) Medline
6. A. McNally, C. K. Prier, D. W. MacMillan, Discovery of an α -amino C-H arylation reaction using the strategy of accelerated serendipity. *Science* **334**, 1114–1117 (2011). [doi:10.1126/science.1213920](https://doi.org/10.1126/science.1213920) Medline
7. M. N. Hopkinson, A. Gómez-Suárez, M. Teders, B. Sahoo, F. Glorius, Accelerated Discovery in Photocatalysis using a Mechanism-Based Screening Method. *Angew. Chem. Int. Ed.* **55**, 4361–4366 (2016). [doi:10.1002/anie.201600995](https://doi.org/10.1002/anie.201600995) Medline
8. 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). [doi:10.1126/science.aar5169](https://doi.org/10.1126/science.aar5169) Medline
9. A. Milo, A. J. Neel, F. D. Toste, M. S. Sigman, A data-intensive approach to mechanistic elucidation applied to chiral anion catalysis. *Science* **347**, 737–743 (2015). [doi:10.1126/science.1261043](https://doi.org/10.1126/science.1261043) Medline
10. C. W. Coley, R. Barzilay, T. S. Jaakkola, W. H. Green, K. F. Jensen, Prediction of Organic Reaction Outcomes Using Machine Learning. *ACS Cent. Sci.* **3**, 434–443 (2017). [doi:10.1021/acscentsci.7b00064](https://doi.org/10.1021/acscentsci.7b00064) Medline
11. S. Szymkuć, E. P. Gajewska, T. Klucznik, K. Molga, P. Dittwald, M. Startek, M. Bajczyk, B. A. Grzybowski, Computer-Assisted Synthetic Planning: The End of the Beginning. *Angew. Chem. Int. Ed.* **55**, 5904–5937 (2016). [doi:10.1002/anie.201506101](https://doi.org/10.1002/anie.201506101) Medline
12. A. Lavecchia, Machine-learning approaches in drug discovery: Methods and applications. *Drug Discov. Today* **20**, 318–331 (2015). [doi:10.1016/j.drudis.2014.10.012](https://doi.org/10.1016/j.drudis.2014.10.012) Medline
13. M. H. Todd, Computer-aided organic synthesis. *Chem. Soc. Rev.* **34**, 247–266 (2005). [doi:10.1039/b104620a](https://doi.org/10.1039/b104620a) Medline

14. R. Sheridan, W. Schafer, P. Piras, K. Zawatzky, E. C. Sherer, C. Roussel, C. J. Welch, Toward structure-based predictive tools for the selection of chiral stationary phases for the chromatographic separation of enantiomers. *J. Chromatogr. A* **1467**, 206–213 (2016). [doi:10.1016/j.chroma.2016.05.066](https://doi.org/10.1016/j.chroma.2016.05.066) Medline
15. N. J. Gesmundo, B. Sauvagnat, P. J. Curran, M. P. Richards, C. L. Andrews, P. J. Dandliker, T. Cernak, Nanoscale synthesis and affinity ranking. *Nature* **557**, 228–232 (2018). [doi:10.1038/s41586-018-0056-8](https://doi.org/10.1038/s41586-018-0056-8) Medline
16. C. J. Welch, Are We Approaching a Speed Limit for the Chromatographic Separation of Enantiomers? *ACS Cent. Sci.* **3**, 823–829 (2017). [doi:10.1021/acscentsci.7b00250](https://doi.org/10.1021/acscentsci.7b00250) Medline
17. M. Wlekinski, B. P. Loren, C. R. Ferreira, Z. Jaman, L. Avramova, T. J. P. Sobreira, D. H. Thompson, R. G. Cooks, High throughput reaction screening using desorption electrospray ionization mass spectrometry. *Chem. Sci.* **9**, 1647–1653 (2018). [doi:10.1039/C7SC04606E](https://doi.org/10.1039/C7SC04606E) Medline
18. S. Sun, T. R. Slaney, R. T. Kennedy, Label free screening of enzyme inhibitors at femtomole scale using segmented flow electrospray ionization mass spectrometry. *Anal. Chem.* **84**, 5794–5800 (2012). [doi:10.1021/ac3011389](https://doi.org/10.1021/ac3011389) Medline
19. K. J. Boggio, E. Obasuyi, K. Sugino, S. B. Nelson, N. Y. Agar, J. N. Agar, Recent advances in single-cell MALDI mass spectrometry imaging and potential clinical impact. *Expert Rev. Proteomics* **8**, 591–604 (2011). [doi:10.1586/epr.11.53](https://doi.org/10.1586/epr.11.53) Medline
20. X. W. Diefenbach, I. Farasat, E. D. Guetschow, C. J. Welch, R. T. Kennedy, S. Sun, J. C. Moore, Enabling Biocatalysis by High-Throughput Protein Engineering Using Droplet Microfluidics Coupled to Mass Spectrometry. *ACS Omega* **3**, 1498–1508 (2018). [doi:10.1021/acsomega.7b01973](https://doi.org/10.1021/acsomega.7b01973)
21. M. S. Ritorto, R. Ewan, A. B. Perez-Oliva, A. Knebel, S. J. Buhrlage, M. Wightman, S. M. Kelly, N. T. Wood, S. Virdee, N. S. Gray, N. A. Morrice, D. R. Alessi, M. Trost, Screening of DUB activity and specificity by MALDI-TOF mass spectrometry. *Nat. Commun.* **5**, 4763 (2014). [doi:10.1038/ncomms5763](https://doi.org/10.1038/ncomms5763) Medline
22. C. Haslam, J. Hellicar, A. Dunn, A. Fuetterer, N. Hardy, P. Marshall, R. Paape, M. Pemberton, A. Resemann, M. Leveridge, The Evolution of MALDI-TOF Mass Spectrometry toward Ultra-High-Throughput Screening: 1536-Well Format and Beyond. *J. Biomol. Screen.* **21**, 176–186 (2016). [doi:10.1177/1087057115608605](https://doi.org/10.1177/1087057115608605) Medline
23. M. Winter, T. Bretschneider, C. Kleiner, R. Ries, J. P. Hehn, N. Redemann, A. H. Luippold, D. Bischoff, F. H. Büttner, Establishing MALDI-TOF as Versatile Drug Discovery Readout to Dissect the PTP1B Enzymatic Reaction. *SLAS Discov.* **10.1177/2472555218759267** (2018). [doi:10.1177/2472555218759267](https://doi.org/10.1177/2472555218759267) Medline
24. K. Beeman, J. Baumgärtner, M. Laubenheimer, K. Hergesell, M. Hoffmann, U. Pehl, F. Fischer, J.-C. Pieck, Integration of an In Situ MALDI-Based High-Throughput Screening Process: A Case Study with Receptor Tyrosine Kinase c-MET. *SLAS Discov.* **22**, 1203–1210 (2017). [doi:10.1177/2472555217727701](https://doi.org/10.1177/2472555217727701) Medline
25. R. E. Heap, A. G. Hope, L. A. Pearson, K. M. S. E. Reyskens, S. P. McElroy, C. J. Hastie, D. W. Porter, J. S. C. Arthur, D. W. Gray, M. Trost, Identifying Inhibitors of Inflammation: A Novel High-Throughput MALDI-TOF Screening Assay for Salt-Inducible Kinases (SIKs). *SLAS Discov.* **22**, 1193–1202 (2017). [doi:10.1177/2472555217717473](https://doi.org/10.1177/2472555217717473) Medline
26. T. Si, B. Li, T. J. Comi, Y. Wu, P. Hu, Y. Wu, Y. Min, D. A. Mitchell, H. Zhao, J. V. Sweedler, Profiling of Microbial Colonies for High-Throughput Engineering of Multistep Enzymatic Reactions via Optically Guided Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *J. Am. Chem. Soc.* **139**, 12466–12473 (2017). [doi:10.1021/jacs.7b04641](https://doi.org/10.1021/jacs.7b04641) Medline
27. M. W. Duncan, H. Roder, S. W. Hunsucker, Quantitative matrix-assisted laser desorption/ionization mass spectrometry. *Brief. Funct. Genomics* **7**, 355–370 (2008). [doi:10.1093/bfpg/eln041](https://doi.org/10.1093/bfpg/eln041) Medline
28. P. Wang, R. W. Giese, Recommendations for quantitative analysis of small molecules by matrix-assisted laser desorption ionization mass spectrometry. *J. Chromatogr. A* **1486**, 35–41 (2017). [doi:10.1016/j.chroma.2017.01.040](https://doi.org/10.1016/j.chroma.2017.01.040) Medline
29. J. R. Cabrera-Pardo, D. I. Chai, S. Liu, M. Mrksich, S. A. Kozmin, Label-assisted mass spectrometry for the acceleration of reaction discovery and optimization. *Nat. Chem.* **5**, 423–427 (2013). [doi:10.1038/nchem.1612](https://doi.org/10.1038/nchem.1612) Medline
30. T. J. Montavon, J. Li, J. R. Cabrera-Pardo, M. Mrksich, S. A. Kozmin, Three-component reaction discovery enabled by mass spectrometry of self-assembled monolayers. *Nat. Chem.* **4**, 45–51 (2012). [doi:10.1038/nchem.1212](https://doi.org/10.1038/nchem.1212) Medline
31. A. B. Diagne, S. Li, G. A. Perkowski, M. Mrksich, R. J. Thomson, SAMDI Mass Spectrometry-Enabled High-Throughput Optimization of a Traceless Pétasis Reaction. *ACS Comb. Sci.* **17**, 658–662 (2015). [doi:10.1021/acscombsci.5b00131](https://doi.org/10.1021/acscombsci.5b00131) Medline
32. P. S. Kutchukian, J. F. Dropinski, K. D. Dykstra, B. Li, D. A. DiRocco, E. C. Streckfuss, L. C. Campeau, T. Cernak, P. Vachal, I. W. Davies, S. W. Krska, S. D. Dreher, Chemistry informer libraries: A chemoinformatics enabled approach to evaluate and advance synthetic methods. *Chem. Sci.* **7**, 2604–2613 (2016). [doi:10.1039/C5SC04751J](https://doi.org/10.1039/C5SC04751J) Medline
33. K. D. Collins, F. Glorius, A robustness screen for the rapid assessment of chemical reactions. *Nat. Chem.* **5**, 597–601 (2013). [doi:10.1038/nchem.1669](https://doi.org/10.1038/nchem.1669) Medline
34. J. Richardson, J. C. Ruble, E. A. Love, S. Berritt, A Method for Identifying and Developing Functional Group Tolerant Catalytic Reactions: Application to the Buchwald-Hartwig Amination. *J. Org. Chem.* **82**, 3741–3750 (2017). [doi:10.1021/acs.joc.7b00201](https://doi.org/10.1021/acs.joc.7b00201) Medline
35. E. B. Corcoran, M. T. Pirnot, S. Lin, S. D. Dreher, D. A. DiRocco, I. W. Davies, S. L. Buchwald, D. W. MacMillan, Aryl amination using ligand-free Ni(II) salts and photoredox catalysis. *Science* **353**, 279–283 (2016). [doi:10.1126/science.aag0209](https://doi.org/10.1126/science.aag0209) Medline
36. A. R. Korte, Y. J. Lee, MALDI-MS analysis and imaging of small molecule metabolites with 1,5-diaminonaphthalene (DAN). *J. Mass Spectrom.* **49**, 737–741 (2014). [doi:10.1002/jms.3400](https://doi.org/10.1002/jms.3400) Medline
37. L. J. Soltzberg, P. Patel, Small molecule matrix-assisted laser desorption/ionization time-of-flight mass spectrometry using a polymer matrix. *Rapid Commun. Mass Spectrom.* **18**, 1455–1458 (2004). [doi:10.1002/rcm.1505](https://doi.org/10.1002/rcm.1505) Medline
38. R. Shi, X. Dai, W. Li, F. Lu, Y. Liu, H. Qu, H. Li, Q. Chen, H. Tian, E. Wu, Y. Wang, R. Zhou, S. T. Lee, Y. Lifshitz, Z. Kang, J. Liu, Hydroxyl-Group-Dominated Graphite Dots Reshape Laser Desorption/Ionization Mass Spectrometry for Small Biomolecular Analysis and Imaging. *ACS Nano* **11**, 9500–9513 (2017). [doi:10.1021/acsnano.7b05328](https://doi.org/10.1021/acsnano.7b05328) Medline

ACKNOWLEDGMENTS

We thank A. Donofrio from Merck & Co., Inc., Kenilworth, NJ, USA, for helpful discussions. **Funding:** S.L. and K.Z. are grateful to the MRL Postdoctoral Research Fellows program for financial support. Financial support was provided by Merck & Co., Inc., Kenilworth, NJ 07033, USA. **Author contributions:** S.L., R.P.S., Z.P., I.W.D., D.D. and S.D.D. designed the chemistry experiments. S.L., R.D.F., H.S., and S.D.D. performed the chemistry experiments. S.L., R.D.F., D.C., T.C., I.W.D., D.D., and S.D.D. developed the extended nanochemistry tools. S.L., S.D., W.D.B., H.W., H.S., and C.J.W. developed the MALDI-MS analytical platform. S.L., K.Z., and S.D.D. analyzed comparative UPLC-MS data. S.L., R.P.S., Z.P., S.D.D. analyzed the reactivity trends. Z.P. conducted ELN data mining. S.L., S.D., T.C., I.W.D., H.S., C.J.W. and S.D.D. wrote the manuscript. **Competing interests:** None declared. **Data and materials availability:** All data described in this work are included in the Excel file associated with the supplementary information.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/cgi/content/full/science.aar6236/DC1

Materials and Methods

Figures S1 to S55

Tables S1 to S8

References

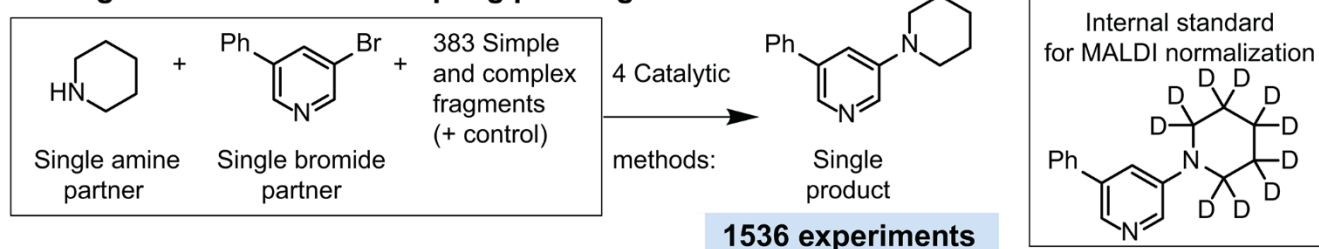
Data S1 to S5

1 December 2017; accepted 15 May 2018

Published online 24 May 2018

10.1126/science.aar6236

A. Fragment additive C-N coupling profiling:



4 Catalytic methods

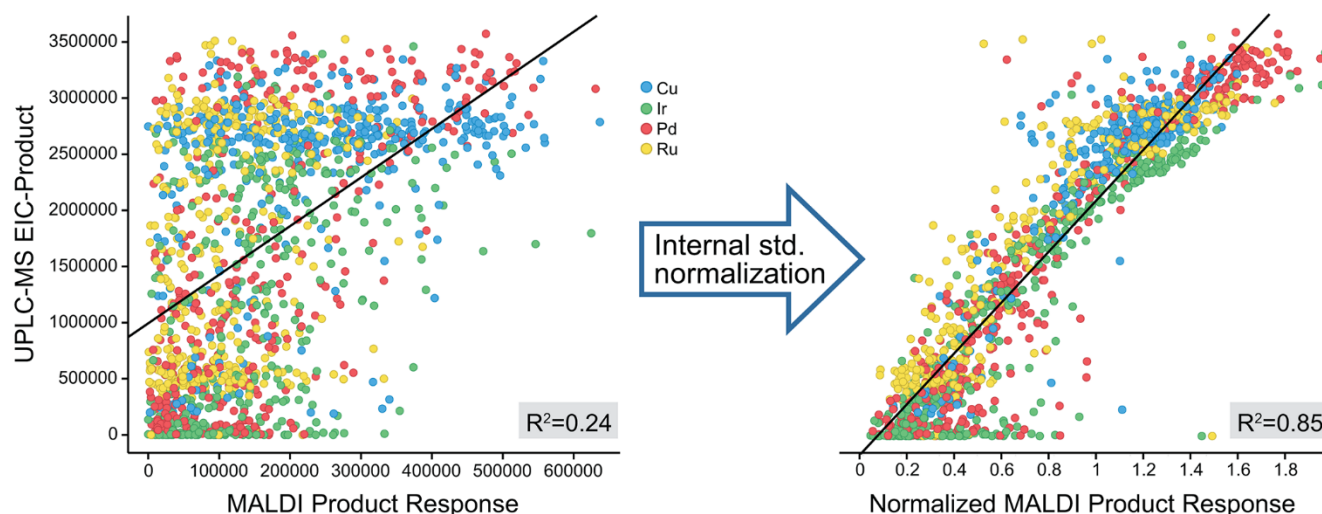
Cu: DiF oxamate ligand (50%), CuI (25%), K₃PO₄, DMSO, 100°C, 20h

Pd: RuPhos G2 precatalyst (10%), Dioxane, Cs₂CO₃, 80°C, 20h

Ir/Ni (photoredox): Ir(dFCF₃-ppy)₂(dtbbpy)(PF₆) (0.05%), NiCl₂ glyme (5%), DABCO, DMSO, 55°C, blue light, 20h

Ru/Ni (photoredox): Ru(bpy)₃(PF₆)₂ (0.1%), NiCl₂ glyme (5%), DABCO, DMSO, 55°C, blue light, 20h

B. Comparative MALDI versus MS-EIC for all data



C. Comparative MALDI versus MS-EIC for each method

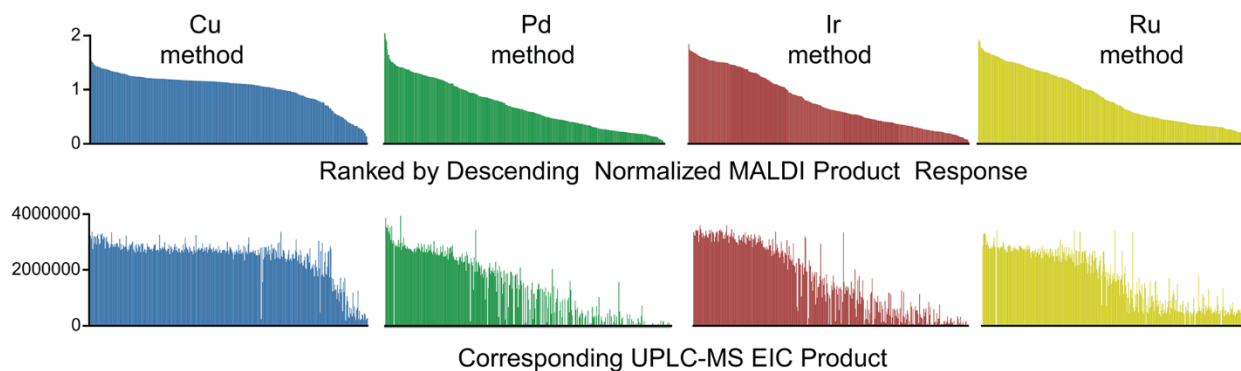


Fig. 1. uHT MALDI MS approach to screening potential reaction poisons. (A) 1536 reactions were run with 4 leading C-N coupling methods evaluating the impact of 383 single and polyfunctional fragments on one simple coupling reaction, requiring 11 min of MALDI acquisition time. (B) Raw MALDI product response is not well correlated with UPLC-MS or UPLC-UV measures, but normalization using an internal standard reveals very good correlation. (C) The ranked, normalized MALDI product response data with corresponding UPLC-MS (EIC) data below for comparison.

Single and polyfunctional fragment additive effects on each method

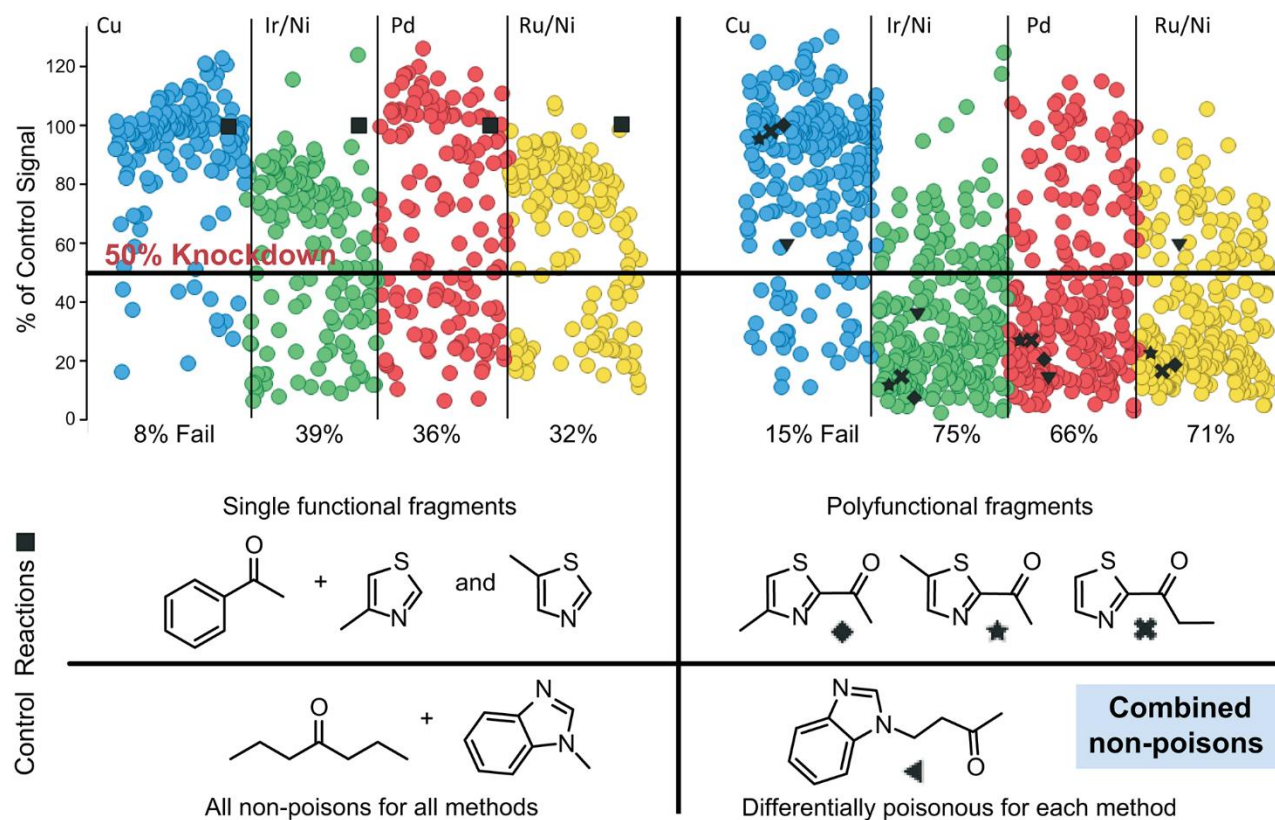
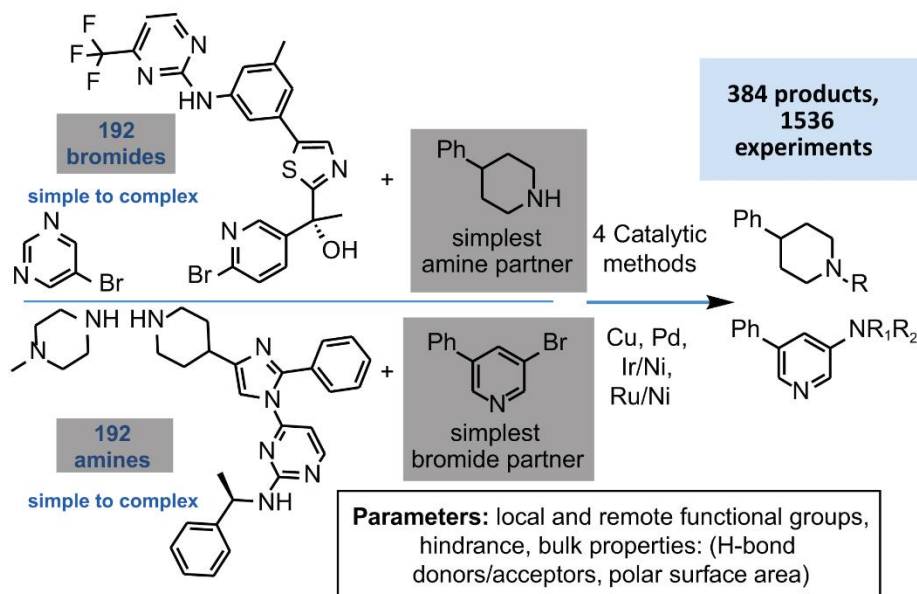


Fig. 2. Fragment additives poisoning study. Comparative rates of poisoning determined by MALDI analysis for single functional fragments and polyfunctional fragments for the 4 tested methods are displayed. This study reveals that Cu is well-suited to managing diverse functionality. Polyfunctional fragments that are composed of linked, single functional fragments that are not poisons themselves often lead to differential poisoning.

A. C-N coupling whole molecule reaction profiling



B. Normalized MALDI versus UPLC-MS EIC data for all data

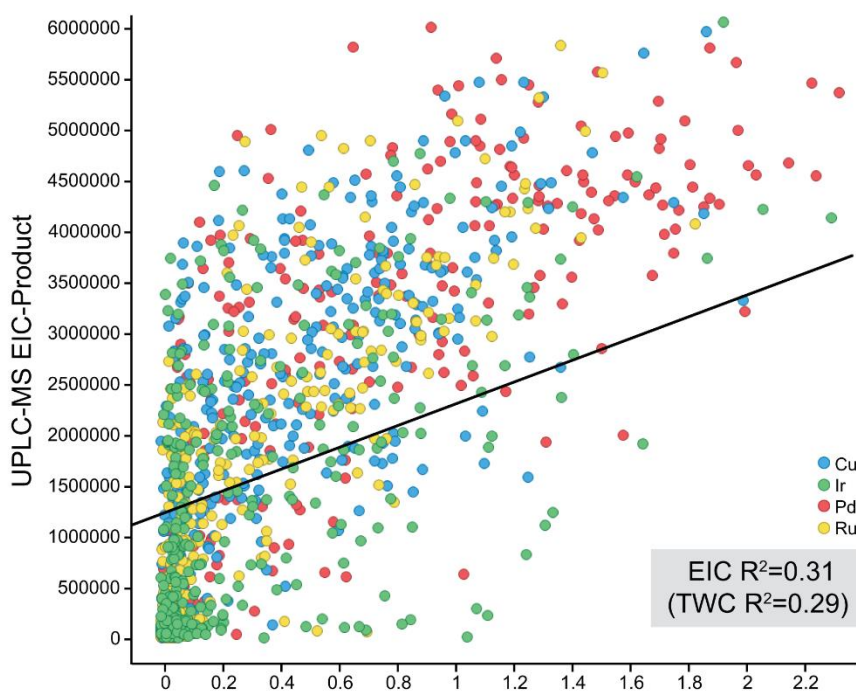
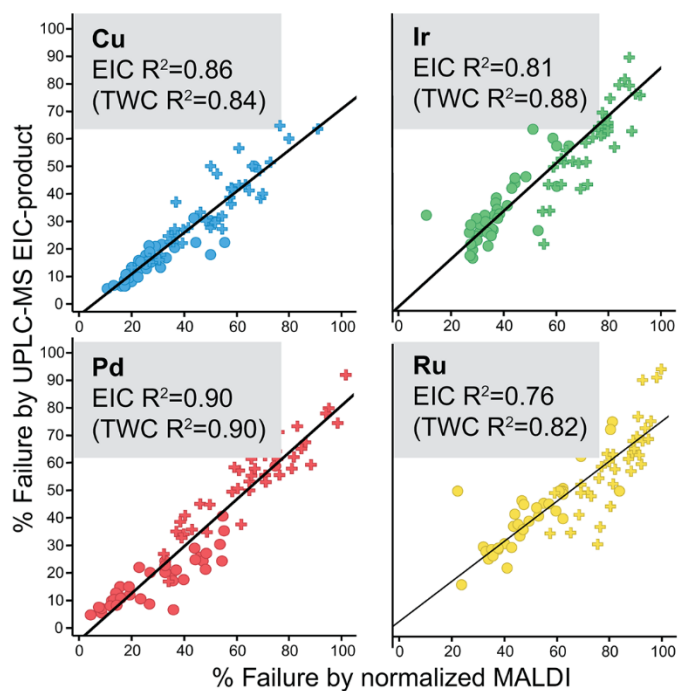
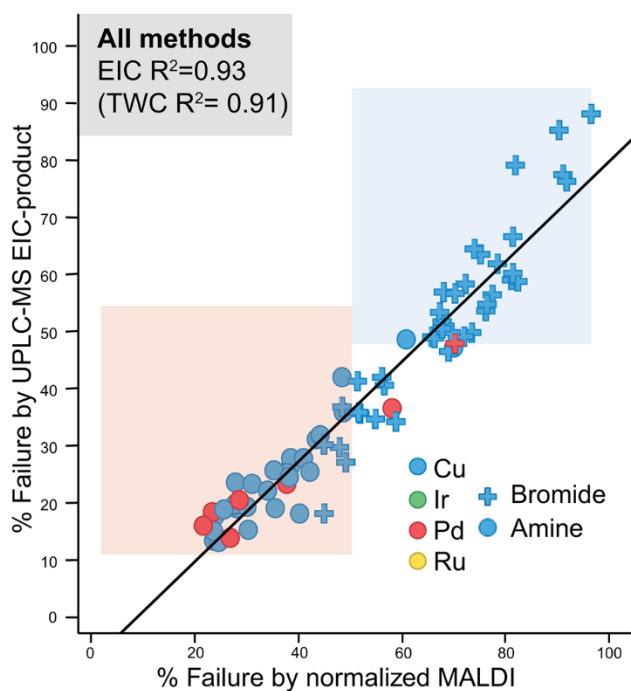


Fig. 3. Whole molecule simplest partner evaluation with MALDI analysis. (A) 192 bromides and 192 amines are each crossed with a simple, mass-active coupling partner under the 4 previously described (Fig. 1A) catalytic methods. (B) Using a single internal standard, the normalized MALDI data show poor correlation with UPLC-MS and UPLC-UV metrics.

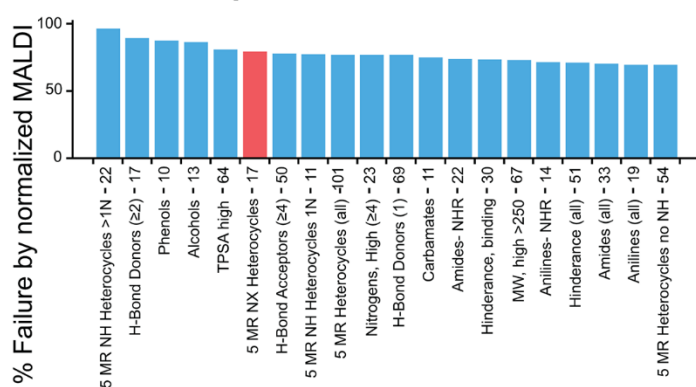
A. Parameter % failure for each method



B. Parameter % failure across methods



C. Problematic parameters in bromides



D. Problematic parameters in amines

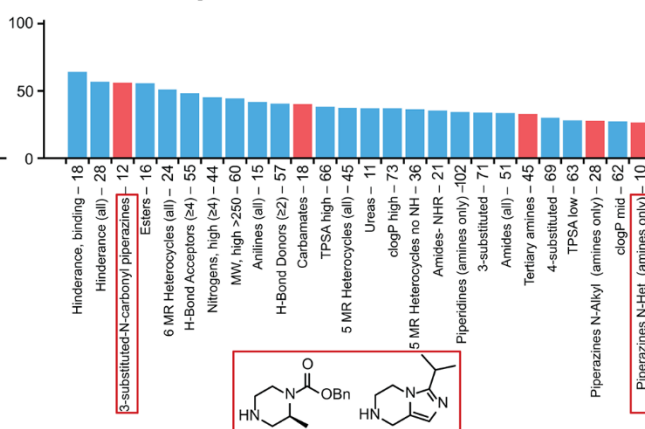


Fig. 4. Whole molecule reactivity trends from binary thresholding analysis. (A) % Failure for different clustered parameters determined by pass/fail binary binning using thresholded MALDI data correlates very well with similarly binned UPLC-MS EIC and UPLC-UV TWC data. Each point on the graph represents the failure rate of different specific aggregated structural parameters, such as NH-heterocycles or H-bond donors, for each synthetic method. (B) The average failure rate for clustered parameters for all 4 methods reveals the functionality that is problematic across methods. The color of each data point reveals which method has the lowest failure rate. (C and D) The most problematic parameters for aryl bromides (blue insert from Fig. 4B) and amines (red insert in Fig. 4B) are listed in descending order of failure rates across methods, along with the number of examples in the test set. In nearly all cases, Cu is the preferred method in these difficult reactions. Several very specific structural types in amines are found to be problematic for Cu, and show greater reactivity with Pd.

Mapping the dark space of chemical reactions with extended nanomole synthesis and MALDI-TOF MS

Shishi Lin, Sergei Dikler, William D. Blincoe, Ronald D. Ferguson, Robert P. Sheridan, Zhengwei Peng, Donald V. Conway, Kerstin Zawatzky, Heather Wang, Tim Cernak, Ian W. Davies, Daniel A. DiRocco, Huaming Sheng, Christopher J. Welch and Spencer D. Dreher

published online May 24, 2018

ARTICLE TOOLS

<http://science.sciencemag.org/content/early/2018/05/23/science.aar6236>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2018/05/23/science.aar6236.DC1>

REFERENCES

This article cites 38 articles, 7 of which you can access for free
<http://science.sciencemag.org/content/early/2018/05/23/science.aar6236#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science* is a registered trademark of AAAS.