Continuous collective analysis of chemical reactions

Maowei Hu[1], Lei Yang[2], Nathaniel Twarog[3], Jason Ochoada[3], Yong Li[2], Brandon M. Young[4], Jeanine Price[4], Kevin McGowan[4], Theresa H. Nguyen[4], Zhe Shi[4], Mary Ashley Rimmer[2], Shea Mercer[5], Zoran Rankovic[2,4], Anang A. Shelat[3], Daniel J. Blair[1]*

[1]Department of Chemical Biology and Therapeutics, St Jude Children's Research Hospital, Memphis, TN 38105, United States.

[2]Analytical Technologies Center, Department of Chemical Biology and Therapeutics, St Jude Children's Research Hospital, Memphis, TN 38105, United States.

[3]Lead Discovery Informatics Center, Department of Chemical Biology and Therapeutics, St Jude Children's Research Hospital, Memphis, TN 38105, United States.

[4]Medicinal Chemistry Center, Department of Chemical Biology and Therapeutics, St Jude Children's Research Hospital, Memphis, TN 38105, United States.

[5]Program Management, Department of Chemical Biology and Therapeutics, St Jude Children's Research Hospital, Memphis, TN 38105, United States.

* e-mail: Daniel.Blair@stjude.org

Modularized synthesis of small organic molecules is transforming our capacity to create medicines and materials. Disruptive acceleration of this molecule building strategy will broadly unlock its functional potential and requires integration of many new assembly chemistries. Recent advances in high-throughput chemistry stand to enable selection of appropriate chemical reaction conditions from the vast range of potential options. However, a disconnect between the rates of exploration and evaluation has limited progress. Here we report how intrinsic fragmentation features of chemical building blocks generalizes their analysis to yield sub-second readouts of reaction outcomes. Central to this advance was identifying that groups typically attached to boron, nitrogen, and oxygen atoms fragment in a specific and selective manner by mass spectrometry, enabling target agnostic analysis. Combining these features with acoustic droplet ejection mass spectrometry we could eliminate slow chromatographic steps and continuously evaluate chemical reaction outcomes in multiplexed formats. This allowed rapid assignment of reaction conditions to molecules derived from ultra-high throughput chemical synthesis experiments.

Increasingly optimized methods for automated iterative chemical synthesis continue to advance the important goal of finding new molecular functions (Fig. 1a)[1,2,3]. Automated synthesis of small organic molecules has recently reached a level of sophistication which provides on-demand access to natural products, pharmaceuticals, and functional materials[4,5]. However, whilst the cycle time of these processes is shrinking[6], it fails to meet the requirements for efficient exploration of chemical space. Radical advancements in many other fields[7,8,9] have arisen from combining modular platforms with high-throughput analytics – therein contracting the time to data output and process optimization. Yet for small molecules asynchronous analytical processes limit progress toward this goal (Fig. 1b), so a significant increase in pace must be realized.

Generalized access to chemical space will stem from incorporation of a wide range of chemical processes into modularized automated formats. But the inherent complexity of forming chemical unions suggests that the concept of "generalized" conditions for many reactions is a foregone impossibility because of the vast range potential coupling partners[10]. Accordingly, automated high-throughput reaction scouting[11] has become the method of choice for obtaining a molecule of interest, as this allows multiple reaction conditions to be actioned in parallel. Despite these benefits, high-content chemical synthesis is analytically cumbersome on even modest scales (i.e. 2–3mins/sample by liquid chromatography = 2 days for ~1000 reactions)[11,12,13] and whilst leading solutions for high-throughput analysis have broken the 1s/sample barrier, they offer limited resolution providing global rather than local trends[14]. With predictive models becoming progressively more effective for structurally aware condition prioritization there is a present and growing need for accurate high-content chemical reaction data[10,15,16].

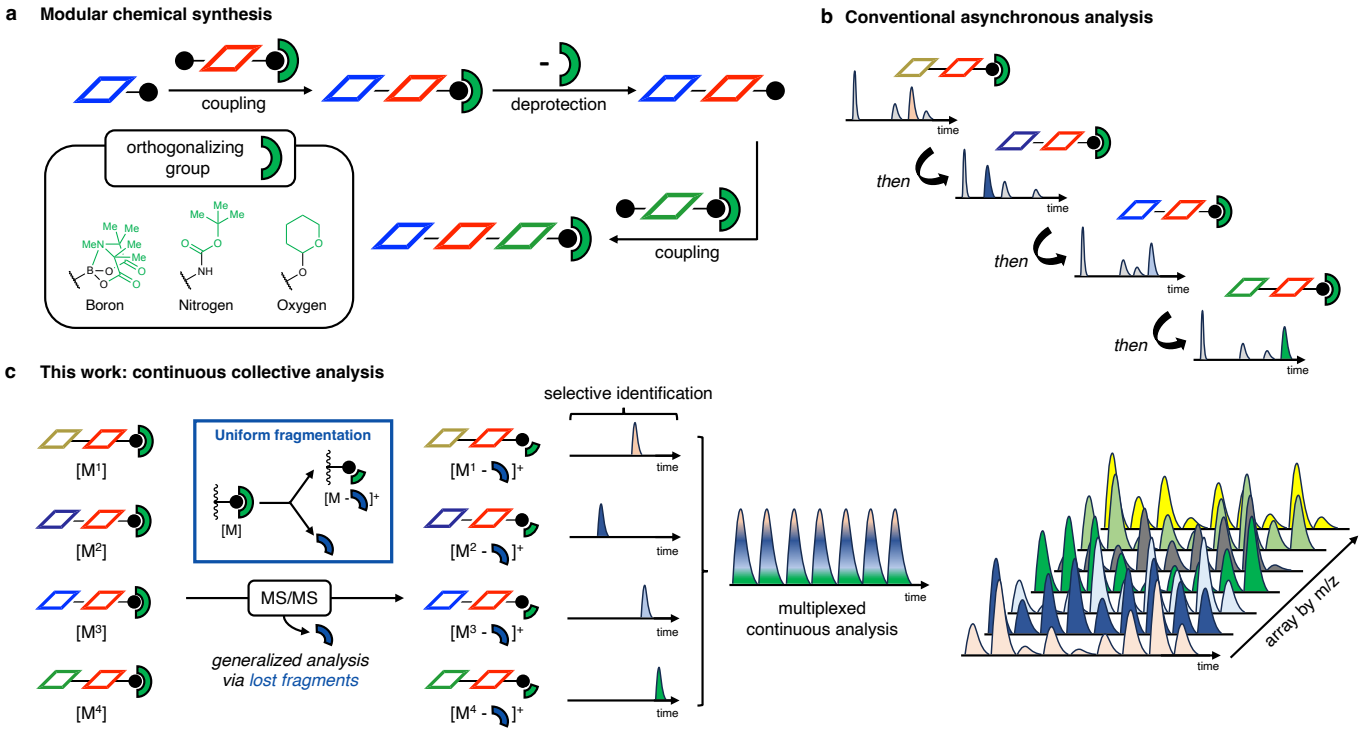


Fig. 1 Loss of common chemical pieces drives rapid analysis. a, Modular chemical synthesis utilizes common orthogonalizing groups to reversibly attenuate reactivity during iterative assembly sequences. **b**, Standard methods for analyzing chemical reactions typically follow one-at-a-time asynchronous workflows to chromatographically separate desired analytes from other components. **c**, Orthogonalizing groups used in modular synthesis possess diagnostic fragmentation patterns allowing direct, simple, and sensitive extraction of analyte signals by tandem mass spectrometry. Diagnostic fragmentation behavior eliminates the need for chromatography allowing analysis to operate in a continuous and multiplexed manner. MS = mass spectrometry.

A common strategy for fast target independent analysis can be found in proteomics[17], where the loss of common chemical pieces permits rapid readouts by tandem mass spectrometry. These methods forgo *a priori* information about analytes except that they contain fragmentation-compatible analytics-enabling chemical pieces. A powerful aspect of these techniques are that common lost chemical pieces provide a direct means to select the desired analyte from amongst complex mixtures. Other advances in tandem mass spectrometry have increased the pace of reaction mixture analysis[11,18,19] however progress in this space has not yielded generalizable solutions because of the requirement of prior knowledge about analytes or the need for analytical standards.

We envisioned that the strategy of common lost chemical pieces could be readily applied to high-throughput chemical synthesis should it be combined with modular chemical synthetic routes, because modular iterative chemistry universally features orthogonalizing groups attached to boron[4], nitrogen[20], or oxygen[21,22] atoms (Fig. 1a). In this way, orthogonalizing groups could serve as uniform handles for determining reaction outcomes *via* tandem mass spectrometry. Enablement of precise fragment directed target identification would unlock simple sample multiplexing (Fig. 1c) and obviate the need for slow chromatographic steps, overcoming the limitations of conventional asynchronous analysis (Fig. 1b). With this in mind, we set out to identify fragmentation behavior which would broadly underlie high-throughput analysis of modular chemical reaction sequences.

Neutral loss method development

To identify common lost chemical pieces associated with orthogonalizing groups used in modular chemical synthesis (Fig. 2a), we began by generating a series of 60 (tetramethyl)methylamino diacetic acid (TIDA) boronates derived from Buchwald-Hartwig reactions (Extended Data Fig. 1a). Fragmentation analysis revealed that for 58 of the TIDA boronates the loss of 86 Daltons (Da) was the dominant process (Fig. 2b, Extended Data Fig. 1b). We ascribed this neutral loss to scission along the N-B frustrated Lewis pair[4] axis releasing methyl acrylic acid. Analysis of collision energies for these 60 boronates indicated that this 86 Da loss was a mild process (Extended Data Fig. 1c).

Eager to explore the potential of TIDA boronates diagnostic fragmentation behavior to enable rapid analysis of reaction products we turned to neutral loss mass spectrometry[23] (Extended Data Fig. 2a). In neutral loss mass spectrometry, the first (Q1) and third (Q3) quadrupoles of a tandem mass spectrometer scan at a predetermined fixed offset, and data are collected when signals separated by this offset is detected. Using this method, common lost chemical fragments enable samples of different molecular weights to be analyzed without creating individualized analysis profiles (i.e. we can operate in

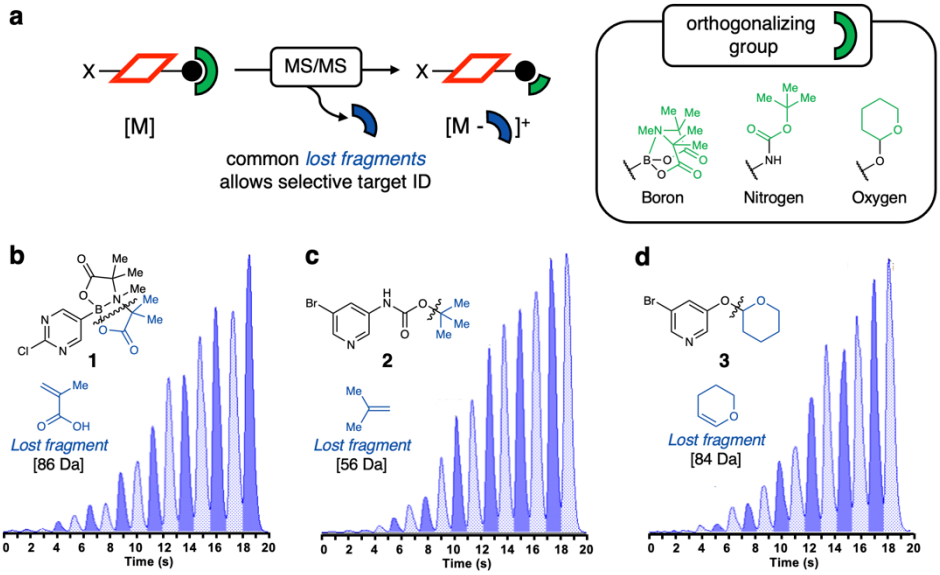


Fig. 2 Fragmentation of modular building blocks enables single second analysis a, Orthogonalizing groups commonly employed in iterative chemical synthesis sequences can serve as functional handles for rapid identification *via* tandem mass spectrometry through the loss of neutral fragments. When combined with acoustic ejection mass spectrometry this allows 1.2 s per sample readout speeds. Groups commonly attached to boron (**b**), nitrogen (**c**) and oxygen (**d**) atoms exhibit fragmentation patterns enabling neutral loss acoustic ejection mass spectrometry (*NL*-ADE-MS). MS = mass spectrometry.

a target agnostic manner). In this way no prior information is required about the analyte of interest except that it contains a common fragmentable component - offering a sample density limited only by the resolution of the mass spectrometer.

To efficiently execute neutral loss analysis, we coupled it with rapid sampling *via* acoustic droplet ejection mass spectrometry (ADE-MS)[18,19] which allows direct introduction of nanoliter volumes into a mass spectrometer, eliminating any slow chromatographic steps (Extended Data Fig. 2b). Initial experiments using neutral loss acoustic droplet ejection mass spectrometry (*NL*-ADE-MS) revealed that the neutral loss of methyl acrylic acid from TIDA boronates was highly sensitive and linear at analytical concentrations (R^2(avg) = 0.99) (Extended Data Fig. 2c, d). The relative signal for each analyte using *NL*-ADE-MS is a product of relative ionizability and fragmentability, so mass spectrometry responses could vary dramatically, nevertheless for equimolar concentrations of our 60 TIDA boronates their mass spectrometry responses were within a single order of magnitude (Extended Data Fig. 2d).

Off-target loss of neutral fragments could limit the utility of *NL*-ADE-MS. To test this, we selected a group of 100 compounds from the St Jude compound collection[24] spanning mW, LogP, and polar surface area (Extended Data Fig. 3a, b, c). Analysis of these compounds by *NL*-ADE-MS yielded very few false positives (2%, Extended Data Fig. 3d). Reported small molecule neutral loss data[25,26] were similarly supportive that loss of 86 Da was distinctly diagnostic of TIDA boronates.

Conventional liquid chromatography mass spectrometry (LC-MS) workflows involve creation of customized profiles to accurately identify the signals of interests. On an individualized basis this is practical and achievable, but in the context of high-throughput synthesis this customized approach transitions from enabling to limiting. Contrastingly, the direct identification of small molecules by neutral loss sets the stage for a generalized analytical method should a single set of parameters be applicable. To validate the generality of our approach, we established that *NL*-ADE-MS data were largely independent of fragmentation parameters[27,28] (Extended Data Fig. 4a, b), free from sample-to-sample cross-talk (Extended Data Fig 4d), and were uninfluenced by the contents of the sample mixture (Extended Data Fig. 4e). Additionally, we could accommodate samples spanning a wide range of molecular weights (200 Da) and increase the pace of sample-to-sample analysis from 3 to 1.2 seconds without loss of performance (Extended Data Fig 4c, f). Collectively these data validated that *NL*-ADE-MS has capacity to allow large groups of small molecules, related only by their propensity to lose chemical fragments, to be analyzed in a single pass.

Simplified parameterization

Considering that our intention in developing high-throughput analysis was to simplify the interrogation of reaction conditions, we recognized that tandem mass spectrometry analysis of starting materials could inform *NL*-ADE-MS method development (Fig. 2a). Accordingly, we found that tandem-mass spectrometry data for halo-TIDA boronate building block **1** was sufficient to generate the parameters for developing a *NL*-ADE-MS method (Fig. 2b). In addition to boron atoms, nitrogen and oxygen atoms serve as common connecting points for modular small molecule synthesis (Fig. 1a). Tandem mass spectrometry analysis of a building blocks **2** and **3** containing the nitrogen orthogonalizing group tert-butyloxycarbonyl (Boc, Fig. 2c)[29], and the oxygen orthogonalizing group tetrahydropyran (THP, Fig. 2d)[30] showed that each readily lost diagnostic chemical fragments (56 and 84 Da respectively) and were thus *NL*-ADE-MS compatible. Subsequent *NL*-ADE-MS analysis revealed that these common functional handles enabled single sample per second analysis by loss of chemical fragments (Fig. 2c, d). These types of fragmentation signatures are certainly intrinsic features of many other chemical functional groups and building blocks, however, to keep this study bounded we chose to demonstrate the utility of this strategy by focusing on functional groups pertinent to iterative assembly sequences.

Reaction mixture assessment

To test the performance of *NL*-ADE-MS at scrutinizing the outcomes of chemical reactions derived from orthogonalized boron, nitrogen, and oxygen atoms we performed a series of miniaturized Buchwald-Hartwig reactions[11]. Using a range of halogenated building blocks [BTIDA (**1**), Boc (**2**), and THP (**3**)], 6 aryl amines (**4-9**), and 64 reaction conditions (4 palladium catalysts, 4 bases, 4 solvents) we generated 384 chemical reactions for each functional handle (Fig. 3a). We were able to leverage the fragmentation signatures associated with TIDA boronates (**10-15**), Boc amines (**16-21**), and THP alcohols (**22-27**) to ascertain the relative product output of 384 chemical reactions in only 7.68 minutes each by *NL*-ADE-MS. This took an equivalent amount of time as collecting two LC-MS samples. *NL*-ADE-MS data displayed excellent point-to-point correlation against LC-MS (R^2 = 0.89-0.95) (Fig. 3b, c, d, Extended Data Fig. 5). Given the high accuracy of our method fine-grained features were readily distinguished

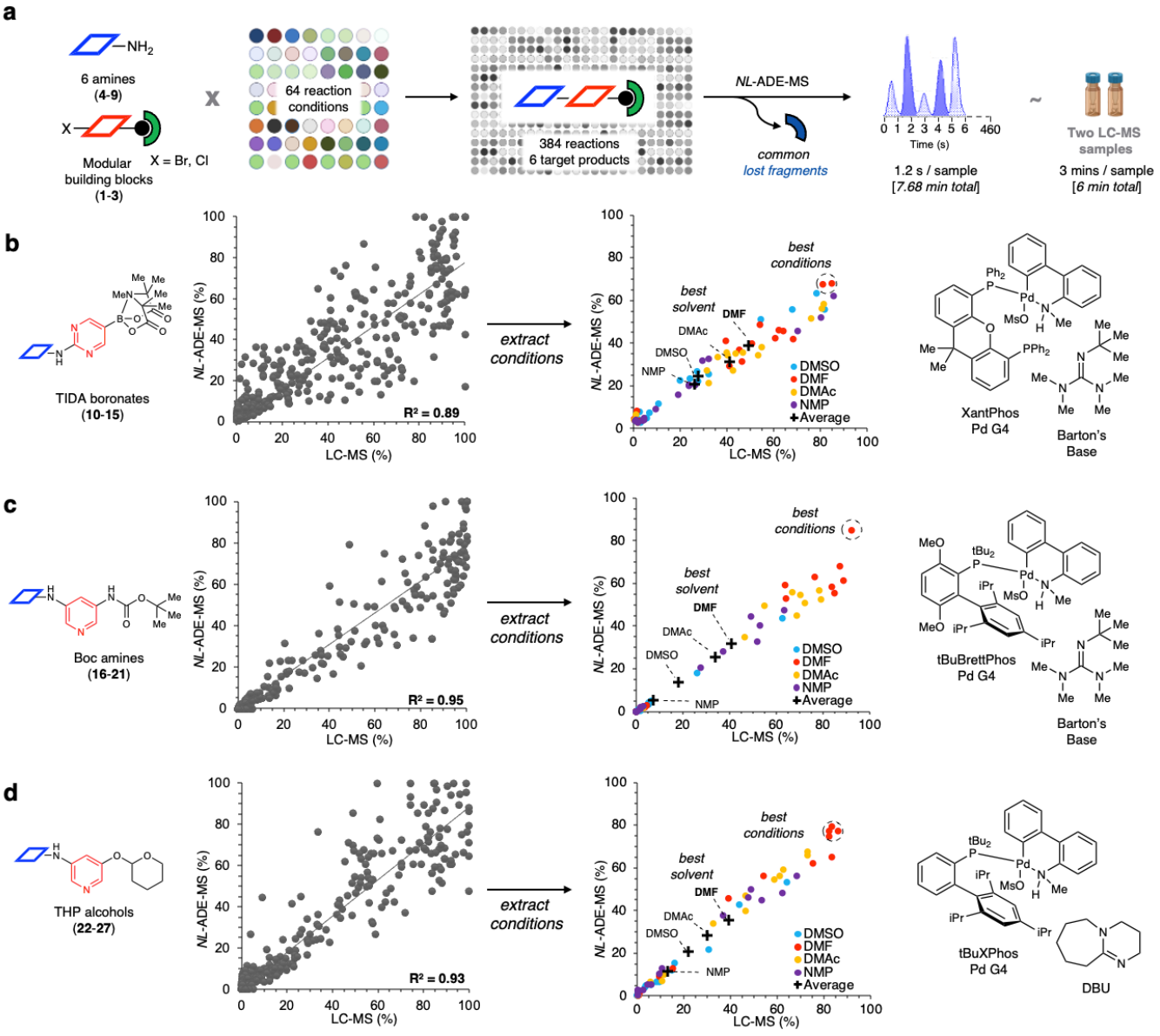


Fig. 3 Generalized determination of chemical reaction outcomes. a, Loss of common chemical pieces enables the rapid analysis of chemical reaction products featuring orthogonalizing groups for boron (**b**), nitrogen (**c**), and oxygen (**d**) by using *NL*-ADE-MS. Optimal reaction conditions and solvents could be assigned using *NL*-ADE-MS across 384 chemical reactions spanning 64 separate reaction conditions by comparison of relative reaction outcomes, requiring only 7.68 minutes of data collection**.** LC-MS, liquid chromatography mass spectrometry; DMSO, dimethylsulfoxide; DMF, dimethylformamide; DMAc, Dimethylacetamide; NMP, *N*-Methyl-2-pyrrolidone; Ph, Phenyl; Ms, Methane sulfonyl; MTBD, 7-methyl-1,5,7-triazabicyclo(4.4.0)dec-5-ane; tBu, tertiary butyl; Cy, cyclohexyl.

across all 64 possible reaction conditions. Using these data, we were able to accurately derive the optimal solvent for a specific substrate class by examining the average relative product intensity (Fig. 3 b, c, d). From here we could then select the best performing catalyst and base combination in all cases. Additional trends could be extracted from these data, for example, *NL*-ADE-MS analysis of THP containing products (**22**-**27**) selected the same best palladium catalyst (*t*BuXPhos Pd G4) but was insensitive to the choice of base with all 4 bases clustering within 4% relative output (Fig. 3d).

Multiplexed rank ordering of reaction conditions

Mixing of multiple analytes within a single sample (multiplexing) is an efficient strategy to increase throughput. As *NL*-ADE-MS directly extracts peaks of interest based on common lost fragments, sample multiplexing is simple provided analytes are distributed in the mass domain (Fig. 1c). By pooling prior test compounds, we found we could mix 8 analytes together to achieve a rate of 7 samples/second.

Armed with a rapid multiplexed analytical pipeline we sought to pressure test *NL*-ADE-MS across structure property space. We recognized that direct synthesis of a diverse suite of 100s of products in a highly pure form would be impractical. Instead, we leveraged high throughput miniaturized chemical synthesis to efficiently access a large array of chemical reactions. We defined a bounded but large region of chemical space using ~600 aryl amine building blocks from our in-house reagent collection and 8 halo-TIDA boronate building blocks (**1**, **28**-**34**). This would cover ~5000 prospective products arising from Buchwald-Hartwig couplings. From these 600 amines we selected 96 distinct structures (**35**-**130**, Extended Data Fig. 6a). Using automated liquid handling we prepared the 768 target products in a high-throughput manner (Fig. 4a). To assess the reaction condition ranking potential of *NL*-ADE-MS we actioned 16 different conditions (4 palladium catalysts and 4 organic bases, Extended Data Fig. 6b, c) for each target, for a total of 12288 reactions across thirty-two 384-well microtiter plates.

Multiplexing was achieved by using acoustic liquid handling, allowing all 12,288 reaction mixtures to occupy a single 1536-well microtiter analysis plate in an 8-plex format. At a pace of 1.2 seconds/sample, reaction outcome data were collected in just 32 minutes by *NL*-ADE-MS. Collecting this same data set using conventional LC-MS methods would require 25.6 days of continuous operation at 3 min/sample. To rapidly process *NL*-ADE-MS data we developed a software package which allows data extraction and analysis in the time domain rather than generating extracted ion chromatograms. In this way data were directly and accurately assigned to their parent wells. Comparison of individual and multiplexed *NL*-ADE-MS data provided near identical results (R^2 =0.98), allaying concerns arising from analyte crosstalk within these multiplexed samples (Extended Data Fig. 7).

The motivation for generating high-throughput reaction data is extraction of key features to guide decision making when navigating chemical space. As a case study to evaluate the performance of our high-throughput data we envisaged three plausible objectives of these types of high-content experimental campaigns which could be applied to many different reaction types: a discovery chemistry approach for diversifying a single chemical scaffold (*Scenarios 1* and *2*), and a search for the highest output conditions for a particular reaction across large areas of chemical space i.e. generalized conditions (*Scenario 3*).

Scenario 1: Conditions for a specific halide. Examination of halo-TIDA boronate **1** across all 96 amines (**35**-**130**) accurately prioritized XantPhos Pd G4 and DBU as the optimal reaction conditions when compared against LC-MS data which showed excellent rank ordering efficiency (R^2= 0.98) (Fig. 4b). Feature extraction for TIDA boronates **28**, **29** and **30** was similarly effective (Extended Data Fig. 8).

Scenario 2: Conditions for a specific amine. Analysis of amine **82** against all 8 TIDA boronates (**1**, **28**-**34**) revealed excellent performance (R^2 = 0.98) and selected the best reaction condition as tBuXPhos Pd G4 and DBU (Fig. 4c). Data across 24 amines (**59**-**82**) showed excellent correlation (R^2(avg) = 0.91) and from these data we were able to directly identify the best conditions in one third of cases. Recognizing that an important outcome of such studies is to rapidly identify high-output reaction

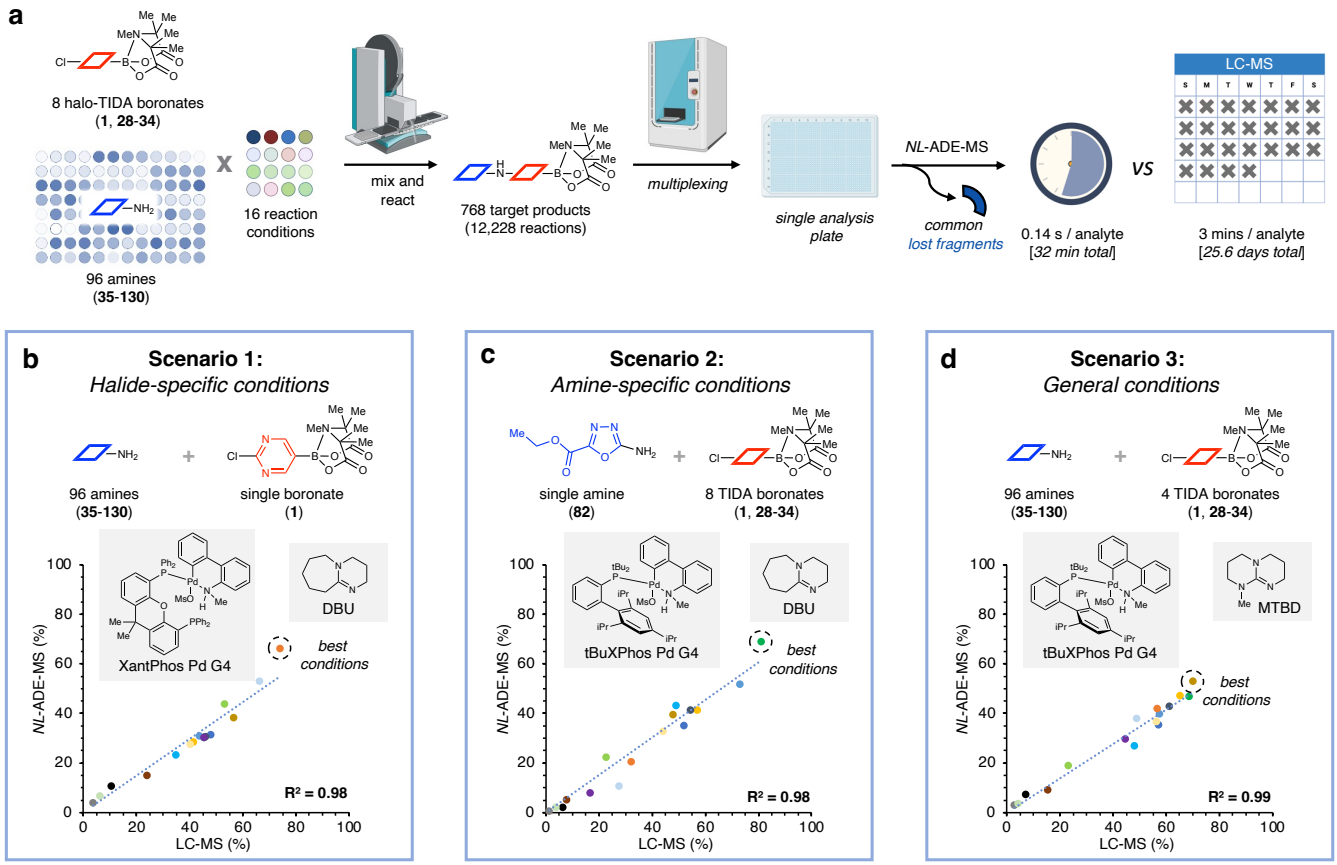


Fig. 4 Ultra-high throughput multiplexed analysis of chemical reaction mixtures. a, Automated liquid handling enabled creation of 12,288 reaction mixtures spanning 738 potential chemical products. Acoustic liquid handling allowed analytical samples to be reformatted into a single 1536-well plate for analysis. Multiplexed *NL*-ADE-MS determined the relative outcomes of these 12,288 reactions in 32 minutes. Comparative LC-MS data revealed that multiplexed *NL*-ADE-MS selects the best reaction conditions with comparable efficiency for specific halides (**b**), amines (**c**), or across substrates (**d**). Data point color scheme represents each of 16 separate reaction conditions, see supporting information for additional details. LC-MS = liquid chromatography mass spectrometry.

conditions we found that for 92% of *NL*-ADE-MS experiments the best condition was within 5% of the top 3 conditions (Extended Data Fig. 9).

Scenario 3. Interrogating generality. Evaluation of large volumes of reaction conditions can maximize the aggregated product output of a synthetic method by identifying general conditions[31]. Direct comparison of our *NL*-ADE-MS data against a practically accessible subset of LC-MS data (50%) revealed uniformly consistent behavior, prioritizing conditions across 384 potential reaction products (Fig. 4d). Here, our *NL*-ADE-MS data rank ordered catalysts and bases in a near identical manner to LC-MS (R^2 = 0.99) correctly identifying tBuXPhos Pd G4 and MTBD as optimal.

Encouraged by the performance of TIDA boronates in multiplexed analyses we re-visited our data for products derived from Boc amine **2** (**16**-**21**) and THP alcohol **3** (**22**-**27**) in a 6-plex format. In this multiplexed format we obtained reaction outcome data for 384 reactions in only 1.28 minutes. Multiplexed data matched individual data (Extended Data Fig. 10), selecting the same best catalysts and bases, affirming the generality of lost common chemical pieces for enabling rapid readouts of chemical reactions.

Data availability

All data are available in the main text of the supplementary materials.

Code availability
Scripts associated with data extraction from *NL*-ADE-MS experiments are available on request.

Acknowledgements
We would like to thank the American Lebanese Syrian Associated Charities (ALSAC) and St. Jude Children's Research Hospital. We thank S. D. Dreher for discussions related to executing high-throughput chemical synthesis.

Author contributions
Project was designed and managed by D.J.B. The manuscript and figures were composed by D.J.B. and M.H. Acoustic droplet ejection mass spectrometry experiments, LC-MS data, high-throughput chemical synthesis, and preparation of TIDA boronate standards were performed by M.H. Building blocks were prepared by M.H., B.Y., J.P., T.N., Z.S., K.M., and D.J.B. Tandem mass spectrometry and QC data were collected and processed by L.Y. and Y.L. Data processing tools were created by N.T. and M.A.R. with input from L.Y. and A.S. Chemoinformatic data were generated by J.O. with A.S. Expedient delivery of data from our departmental centers was coordinated by S.M., Z.R., and A.S.

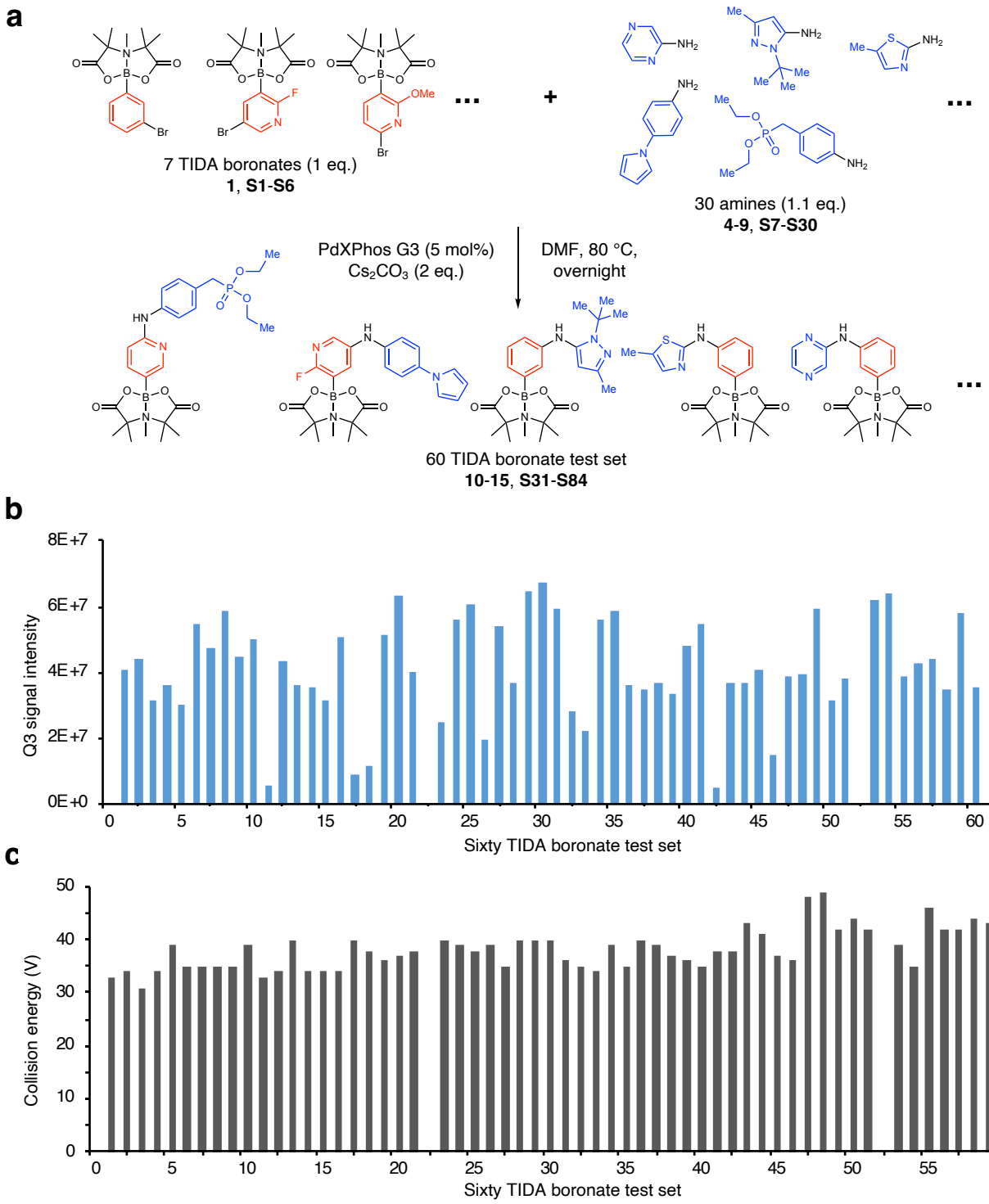
Competing interests
The authors declare no competing interests.

Correspondence and requests for materials should be addressed to Daniel J. Blair.

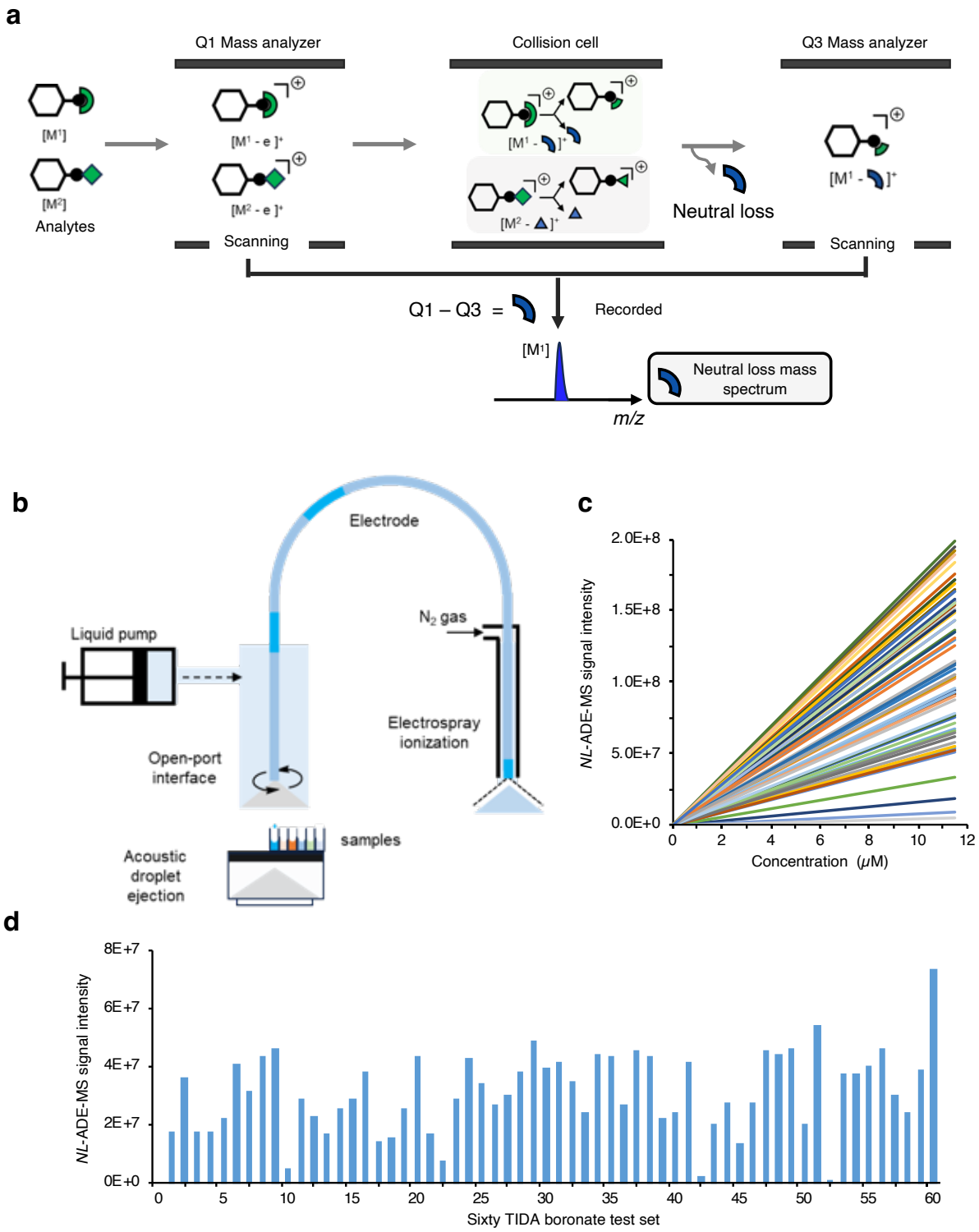
References

1. J. W. Lehmann, D. J. Blair, M. D. Burke. Toward the generalized iterative synthesis of small molecules. *Nature Reviews Chemistry* **2**, 0115 (2018) doi: 10.1038/s41570-018-0115
2. D. Salley, J. S. Manzano, P. J. Kitson, L. Cronin. Robotic Modules for the programmable chemputation of molecules and materials. *ACS Cent. Sci.* **9**, 1525-1537 (2023) doi: 10.1021/acscentsci.3c00304
3. M. Abolhasani, E. Kumacheva, The rise of self-driving labs in chemical and materials sciences. *Nature Synthesis* **2**, 483-492 (2023). doi: 10.1038/s44160-022-00231-0
4. D. J. Blair, S. Chitti, M. Trobe, D. M. Kostyra, H. M. S. Haley, R. L. Hansen, S. G. Ballmer, T. J. Woods, W. Wang, V. Mubayi, M. J. Schmidt, R. W. Pipal, G. F. Morehouse, A. M. E. Palazzolo Ray, D. L. Gray, A. L. Gill, M. D. Burke. Automated iterative Csp3-C bond formation. *Nature*, **604**, 92-97 (2022) doi: 10.1038/s41586-022-04491-w
5. P. J. Kitson, G. Marie, J.-P. Francoia, S. S. Zalesskiy, R .C. Sigerson, J. S. Mathieson, L. Cronin Digitization of multistep organic synthesis in reactionware for on-demand pharmaceuticals. *Science* **359**, 314-319, (2018) doi: 10.1126/science.aao3466
6. W. Wang, N. Angello, D. J. Blair, K. Medine, T. Tyrikos-Ergas, A. Laporte, M. D. Burke. Rapid automated iterative small molecule synthesis. *ChemRxiv* (2023) doi: 10.26434/chemrxiv-2023-qpf2x
7. M. T. Bonde, S. Kosuri, H. J. Genee, K. Sarup-Lytzen, G. M. Church, M. O. A. Sommer, H. H. Wang, Directed mutagenesis of thousands of genomic targets using microarray-derived oligonucleotides ACS Synth. Biol. 4, 17-22, (2015). doi: 10.1021/sb5001565
8. J. Chang, P. Nikolaev, J. Carpena-Núñez, R. Rao, K. Decker, A. E. Islam, J. Kim, M. A. Pitt, J. I. Myung, B. Maruyama. Efficient closed-loop maximization of carbon nanotube growth rate using Bayesian optimization. *Sci. Rep.* **10**, 9040, (2020). doi: 10.1038/s41598-020-64397-3
9. A. Thompson, J. Schäfer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, C. Hamon, Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS *Anal. Chem.* **75**, 1895-1904, (2003). doi: 10.1021/ac0262560
10. N. I. Rinehart, R. K. Saunthawal, 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) doi: 10.1126/science.adg2114
11. A. B. Santanilla, E. L. Regalado, T. Pereira, M. Shevlin, K. Bateman, L-C. Capeau, J. Schneeweis, S. Berritt, Z.-C. Shi, P. Nanatermet, Y. Lui,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 (2014). doi: 10.1126/science.125920

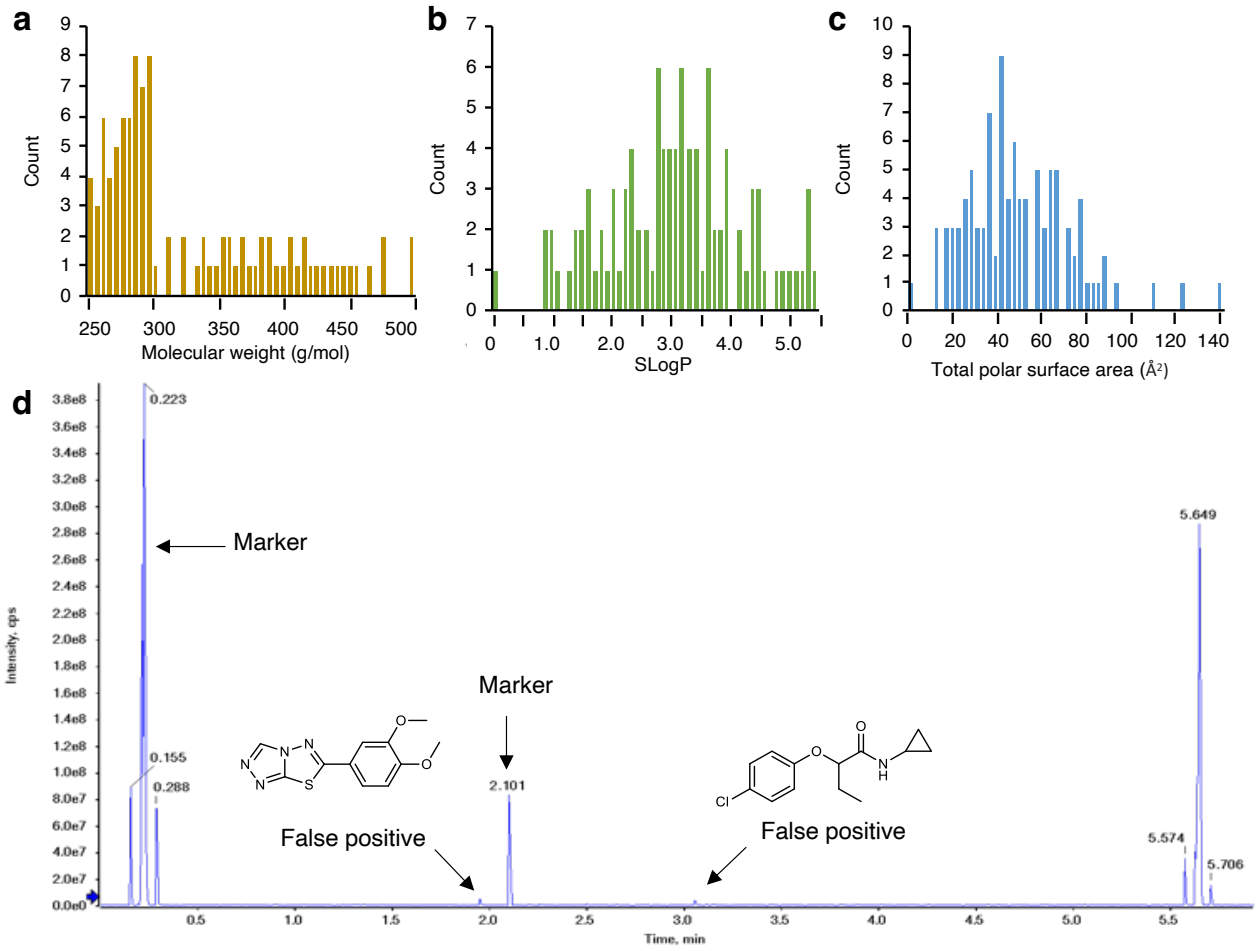
12. B. Mahjour, R. Zhang, Y. Shen, A. McGrath, R. Zhao, O. G. Mohamed, Y. Lin, Z. Zhang, J. L Douthwaite, A. Tripathi, T. Cernak, Rapid planning and alaysis of high-throughput experiment arrays for reaction discovery. *Nature Communications* **14**, 3924, (2023), doi: 10.1038/s41467-023-39531-0
13. A. Osipyan, S. Shaabani, R. Warmerdam, S. V. Shishikina, H. Boltz, A. Dömling. Automated, accelerated nanoscale synthesis of iminopyrrolidines. *Angew. Chem. Int. Ed.* **59**, 12423-12427, (2020) doi: 10.1002/anie.202000887
14. S. Lin, S. Dikler, W. D. Blincoe, R. D. Ferguson, R. P. Sheridan, Z. Peng, D. V. Conway, K. Zawatzky, H. Wang, T. Cernak, I. W. Davies, D. A. Dirocco, H. Sheng, C. J. Welch, S. D. Dreher. Mapping the dark space of chemical reactions with extended nanomole synthesis and MALDI-TOF MS. *Science* **361**, eaar6236, (2018) doi: 10.1126/science.aar6236
15. 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. *Nature Chemistry* (2024) doi: 10.1038/s41557-023-01393-w
16. N. H. Angello, V. Rathore, W. Berker, 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 condtions for heteroaryl Suzuki-Miyaura coupling. *Science* **378**, 399-405 (2022) doi: 10.1126/science.adc8743
17. Q. Yu, X. Liu, M. P. Keller, J. Navarrete-Perea, T. Zhang, S. Fu. L. P. Vaites, S. R. Shuken, E. Schmid, G. R. Keele, J. Li, E. L. Huttlin, E. H. Rashan, J. Simcox. G A. Churchill, D. K. Schweppe, A. D. Attie, J. A. Paulo, S. P. Gygi. Sample multiplexing-based targeted pathway proteomics with real-time analytics reveals the impact of genetic variation on protein expression *Nature Communications* **14**, 555 (2023) doi: 10.1038/s41467-023-36269-7
18. H. Hu, A. Singh, D. Lehnherr, V. Mdluli, S. W. Chun, A. M. Makarewicz, J. R. Houker, O. Ukaegbu, S. Li, X. Wen, D. G. McLaren, J. E. Velasquez, J. C. Moore, S. Galanie, E. Appiah-Amponsah, E. K. Regalado, Accelerating pharmaceutical process development with an acoustic droplet ejection-multiple reaction monitoring-mass spectrometry workflow *Analytical Chemistry* **96**, 1138-1146, (2024) doi: 10.1021/acs.analchem.3c04211
19. K. J. DiRico, W. Hua, C. Liu, J. W. Tucker, A. S. Ratnayake, M. E. Flanagan, M. D. Troutman, M. C. Noe, H. Zhang, Ultra-high-throughput acoustic droplet ejection-open port interface-mass spectrometry for parallel medicinal chemistry. *ACS Med Chem. Lett.* **11**, 1101-1110, (2020) doi: 10.1021/acsmedchemlett.0c00066
20. Y.-C. Lin, F. Schneider, K. J. Eberle, D. Chiodi, H. Nakamura, S. h. Reisberg, J. Chen, M. Saito, P. S. Baran, Atroposelective total synthesis of darobactin A, *J. Am. Chem. Soc.* **144**, 14458-14462 (2022). doi: 10.1021/jacs.2c05892
21. I. B. Seiple, Z. Zhang, P. Jakubec, A. Langlois-Mercier, P. M. Wright, D. T. Hog, K. Yabu, S. R. Allu, T. Fukuzaki, P. N. Carlsen, Y. Kitamura, X. Zhou, M. L. Condakes, F. T. Szczypiński, W. D. Green, A. G. Myers, A platform for the discovery of new macrolide antibiotics. *Nature* **533**, 338-345, (2016). doi: 10.1038/nature17967
22. Mori, Y., Nogami, K., Hayashi, H. & Noyori, R. Sulfonyl-stabilized oxiranyllithium-based approach to polycyclic ethers. Convergent synthesis of the ABCDEF-ring system of yessotoxin and adriatoxin. *J. Org. Chem.* **68**, 9050–9060 (2003).
23. A. E. Schoen, J. W. Amy, J. D. Ciupek, R. G. Cooks, P. Dobberstein, G. Jung, A hybrid BEQQ mass spectrometer. *Int. J. Mass Spectrom. Ion Proc.* **65**, 125-140 (1985), doi: 10.1016/0168-1176(85)85059-X
24. G. Nishiguchi, S. Das, J. Ochoada, H. Long, R. E. Lee, Z. Rankovic, A. A. Shelat, Evaluating and evolving a screening library in academia: the St Jude Approach. *Drug Discovery Today* **26**, 1060-1069, (2021). doi: 10.1016/j.drudis.2021.01.005
25. A. Aisporna, H. P. Benton, A. Chen, R. J. E. Berks, J. M. Galano, M. Biera, G. Siuzdak, Neutral loss mass spectral data enhances molecular similarity analysis in METLIN *J Am Soc Mass Spectrom* **33**, 530-534, (2022) doi: 10.1021/jasms.1c00343
26. Y. Ma, T. Kind, D. Yang, C. Leon, O. Fiehn, MS2Analyzer: A software for small molecule substructure annotations from accurate tandem mass spectra. *Anal. Chem.* **86**, 10724-10731, (2014). doi: 10.1021/ac502818e
27. C. A. Sherwood, A. Eastham, L. W. Lee, J. Risler, H. Mirzaei, J. A. Falkner, D. B. Martin, Rapid optimization of MRM-MS instrument parameters by subtle alteration of precursor and product m/z targets. *J. Proteome Res.* **8**, 3746-3751, (2009) doi: 10.1021/pr801122b
28. R. Dubey, D. W. Hull, S. Lai, C. Ming-Hui, D. F. Grant, Correction of precursor and product ion relative abundances in order to standardize CID spectra and improve Ecom50 accuracy for non-targeted metabolomics. *Metabolomics* **11**, 753-763, (2015) doi: 10.1007/s11306-014-0732-0
29. G. Verardo, A. Gorassini, Characterization of N-Boc/Fmoc/Z-N-formyl-gem-diaminoalkyl derivatives using electrospray ionization multi-stage mass spectrometry. *J. Mass. Spectrom.* **48**, 1136-1149, (2013). doi: 10.1002/jms.3278
30. J. Ruzicka, C. Weisbecker, A. B. Attygalle, Collision-induced dissociation mass spectra of positive ions derived from tetrahydropyranyl (THP) ethers of primary alcohols. *J. Mass. Spectrom.* **46**, 12-23, (2011). doi: 10.1002/jms.1849
31. C. M. Prieto Kullmer, J. A. Kautzky, S. W. Krska, T. Nowak, S. D. Dreher, D. W. C. MacMillan, Accelerating reaction generality and mechanistic insight through additive mapping, *Science* **376**, 532-539, (2022), doi: 10.1126/science.abn1885



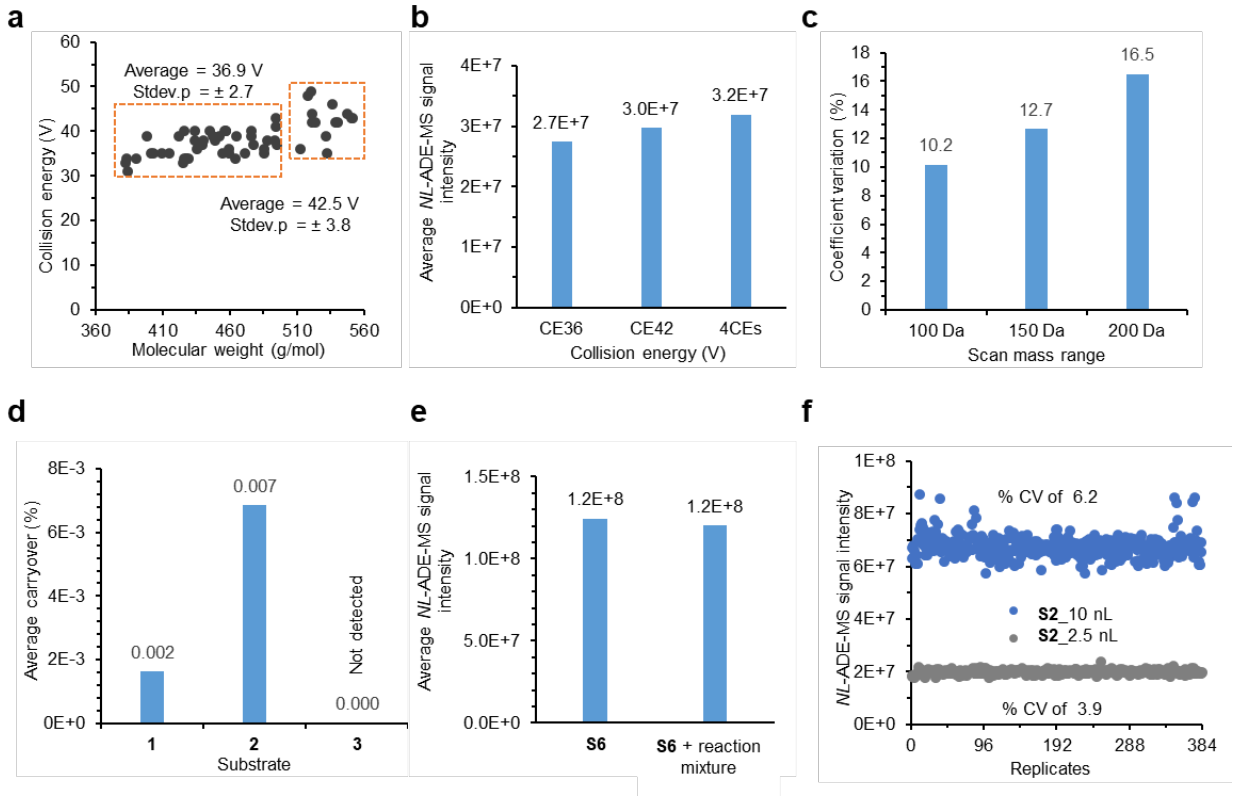
Extended Data Fig. 1 Synthesis and fragmentation analysis of TIDA boronates. **a**, A set of 60 TIDA boronates (**10-15**, **S31-S84**) were prepared from halo-TIDA boronates (**1**, **S1-S6**) and amines (**4-9**, **S7-S30**) *via* Buchwald-Hartwig coupling reactions and purified by preparative reverse phase liquid-chromatography mass spectrometry. **b**, Product ion scanning tandem mass spectrometry analysis of these 60 TIDA boronates (**10-15**, **S31-S84**) revealed consistent loss of 86 Da. **c**, Collision induced dissociation of these 60 TIDA boronates (**10-15**, **S31-S84**) follows a mild and consistent collision energy.



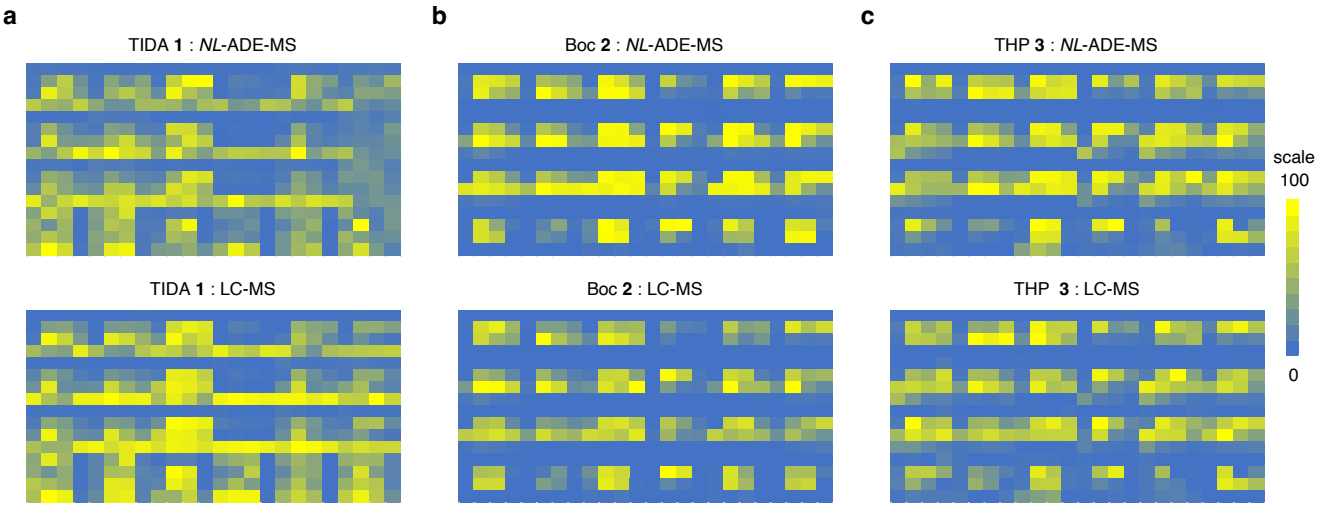
Extended Data Fig. 2 Neutral loss analysis of TIDA boronates. **a**, Schematic representation of neutral loss mass spectrometry, analyte MS data are collected only when parent Q1 ions and daughter Q3 ions are separated by the mass of the desired neutral lost fragment. **b**, Schematic representation of acoustic droplet ejection mass spectrometry. Nanoliter droplets are directly introduced into a mass spectrometry *via* acoustic ejection into an open port interface. **c**, Neutral loss acoustic droplet ejection mass spectrometry (*NL*-ADE-MS) data for 60 TIDA boronates (**10-15**, **S31-S84**) using loss of 86 Da fragments shows a linear response R^2(avg) = 0.99. **d**, Relative signal intensity for *NL*-ADE-MS data collected for 60 TIDA boronates (**10-15**, **S31-S84**) at equimolar concentrations (10 uM) revealed all signals to be within approximately 1 order of magnitude.



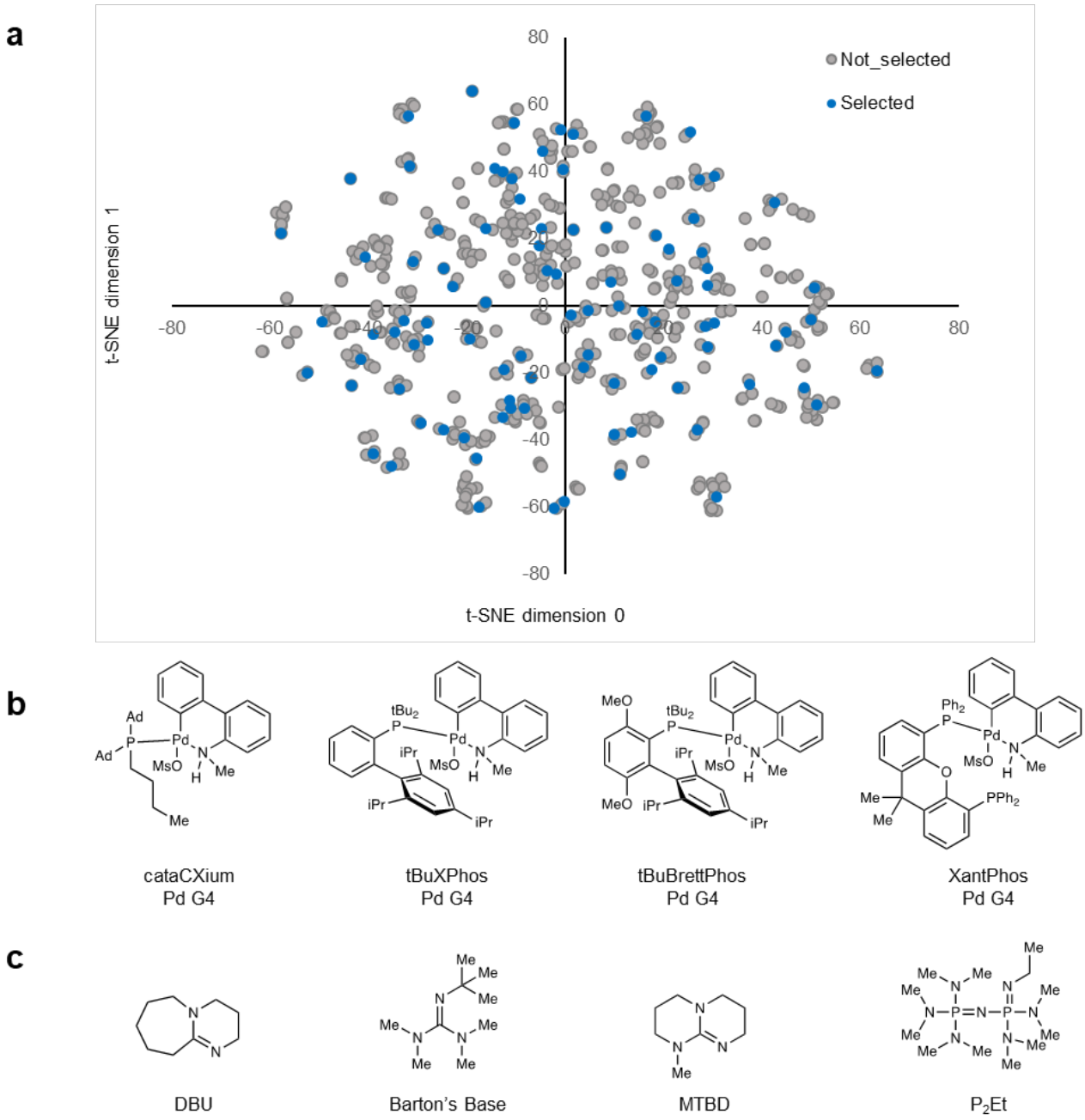
Extended Data Fig. 3 Evaluation of false positive signals arising from *NL*-ADE-MS. 100 diverse compounds spanning **a**, mW, **b,** LogP, and **c,** polar surface area were selected to validate the specificity of TIDA boronate fragmentation. **d**, *NL*-ADE-MS analysis of these 100 compounds revealed only two weak hits for the neutral loss of 86 Da demonstrative of high specificity of TIDA boronate fragmentation.



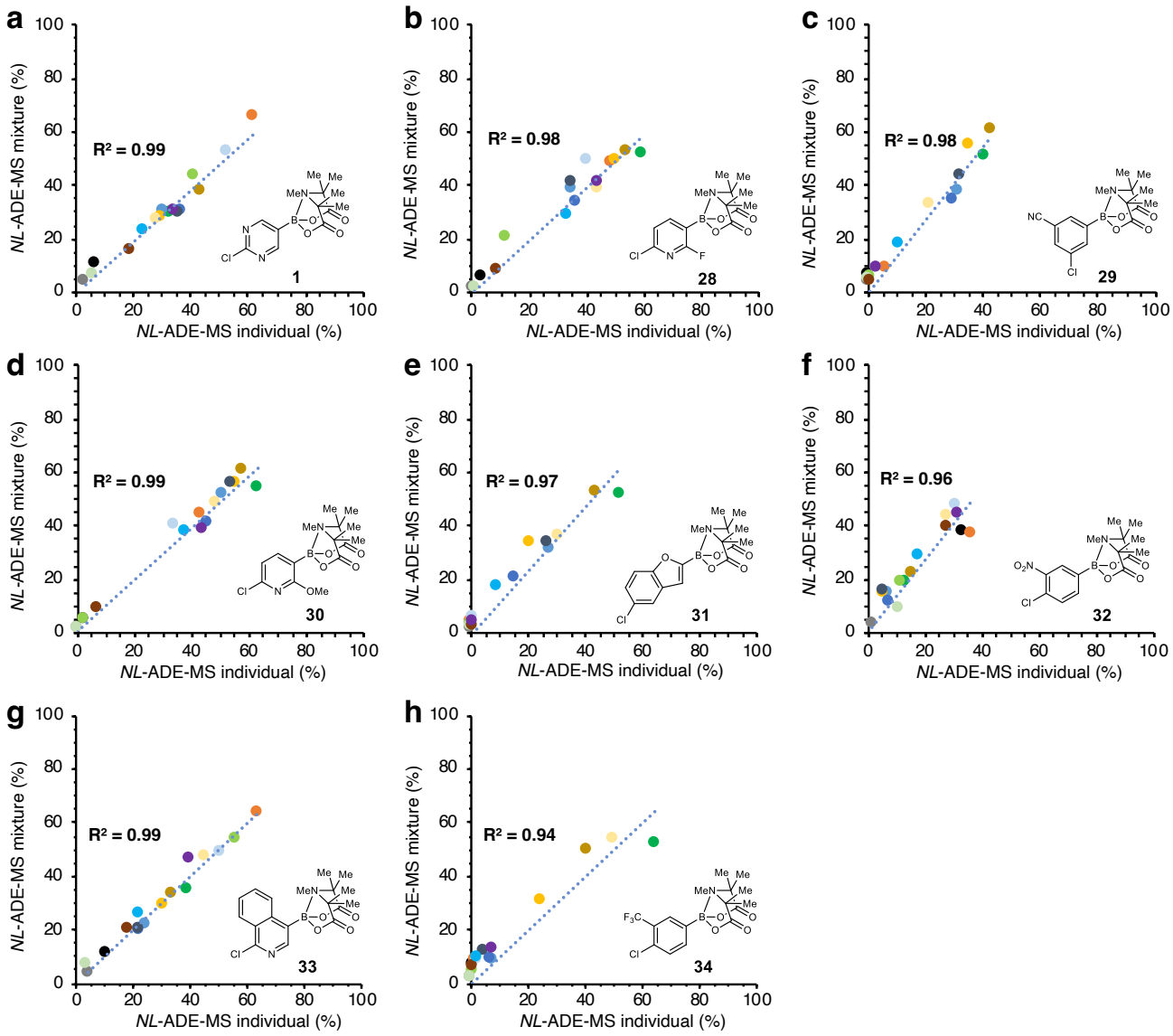
Extended Data Fig. 4 Neutral loss acoustic droplet ejection mass spectrometry method optimization. **a**, Collision energy scales with molecular weight across our 60 TIDA boronate test set, heavier molecules will require more kinetic energy to effect a similar fragmentation efficiency. **b**, To test the influence of non-idealized collision energies we grouped TIDA boronates in narrow and wide molecular weight ranges. *NL*-ADE-MS data were collected for scan ranges covering each grouping using the averaged collision energy for each group. Peak product signal intensities were tested at the wide range for 36 V and 42 V collision energy. Comparison against narrower scan range data with range optimized collision energies showed that non-idealized collision energies minimally impacted the average signal intensity across our TIDA boronate test set. **c**, Influence of mass scan range on sample to sample *NL*-ADE-MS data collection. **d**, Alternating injections of either TIDA boronate **1**, Boc **2**, or THP **3**, and blank wells allowed assessment of carry-over. Peaks separated by 1.2s demonstrated no significant residual signal. **e**, Authentic samples of TIDA boronate **S6** were mixed with reaction mixture components derived from Buchwald-Hartwig coupling reactions and analyzed by *NL*-ADE-MS. Signal intensities were essentially identical in the presence or absence of these reaction components, demonstrating that matrix effect have limited influence on *NL*-ADE-MS. **f**, Assessment of the optimal injection volume for *NL*-ADE-MS analysis. Using boronate **S2** ejection of a single 2.5 nL droplet was the most reproducible, whereas ejection of 10 nL (4 droplets) was more variable.



Extended Data Fig. 5 Head-to-head comparison of *NL*-ADE-MS and LC-MS. Relative product outputs were determined for products derived from building blocks **a**, **1**, **b**, **2**, and **c**, **3** by both *NL*-ADE-MS and LC-MS. These data show excellent agreement between *NL*-ADE-MS and LC-MS affirming the ability of *NL*-ADE-MS to rapidly acquire accurate reaction outcome data. Each individual 384-well plate required only 7.68 minutes of data collection by *NL*-ADE-MS, whereas the equivalent LC-MS data set needed 19.2 h.

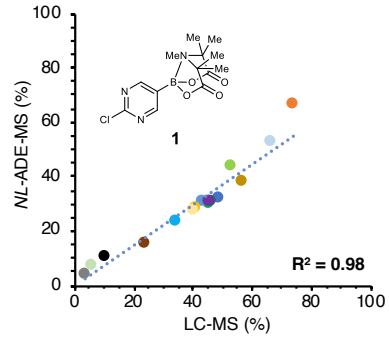


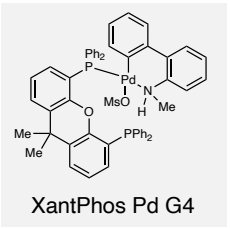
Extended Data Fig. 6 Amine selection for high-throughput experimentation. a, Representative t-SNE plot showing distribution of amines across structure property space, see supporting information for selection parameters. Selected amines shown in blue, non-selected cluster members in grey. **b**, Structures of palladium catalysts utilized in high-throughput experiment. **c**, Structures of organic bases utilized in high-throughput experiment. Ad, adamantyl; t-Bu, tertiary-butyl; iPr, isopropyl; Ms, Methane sulfonyl; Ph, phenyl.



Extended Data Fig. 7 Comparison of multiplexed and individual *NL*-ADE-MS data. a, Boronate **1**. **b**, Boronate **28**. **c**, Boronate **29**. **d**, Boronate **30**. **e**, Boronate **31**. **f**, Boronate **32**. **g**, Boronate **33**. **h**, Boronate **34**. Colors reflect individual specific reaction conditions.

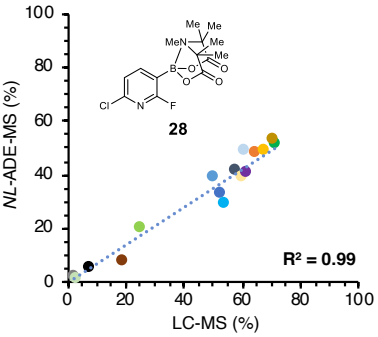
a

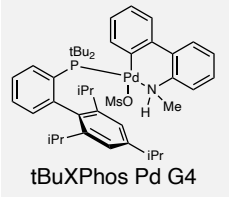




Rank	Base	NL-ADE -MS (%)	LC-MS (%)
1	DBU	66	74
2	MTBD	53	66
3	Barton's Base	44	53

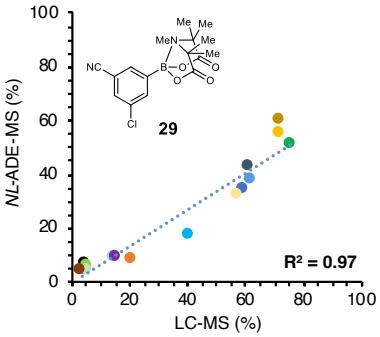
b

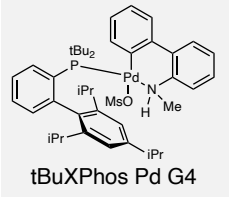




Rank	Base	NL-ADE -MS (%)	LC-MS (%)
1	MTBD	53	71
2	DBU	52	72
3	Barton's Base	49	68

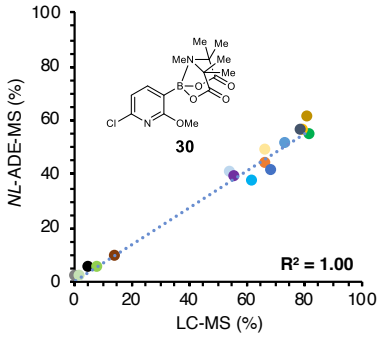
c

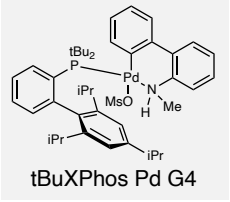




Rank	Base	NL-ADE -MS (%)	LC-MS (%)
1	MTBD	60	72
2	Barton's Base	55	72
3	DBU	51	76

d

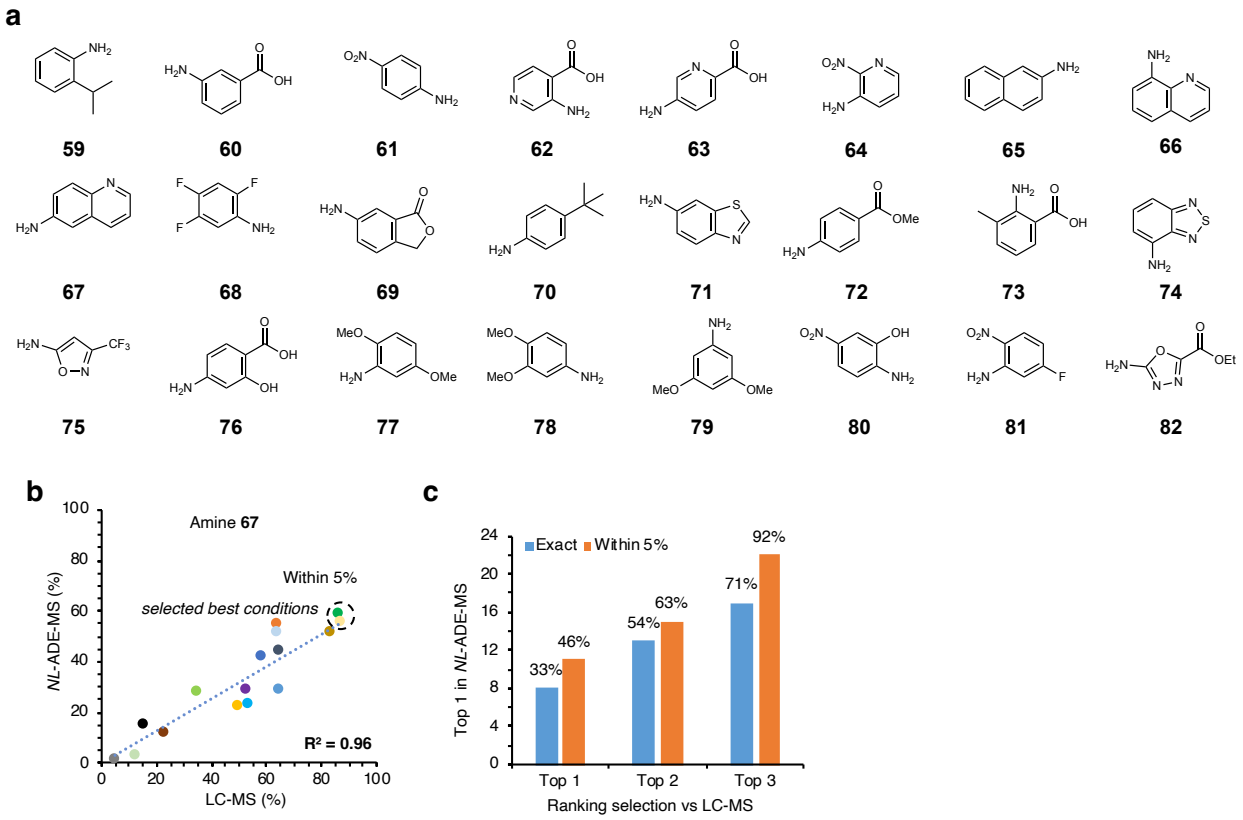




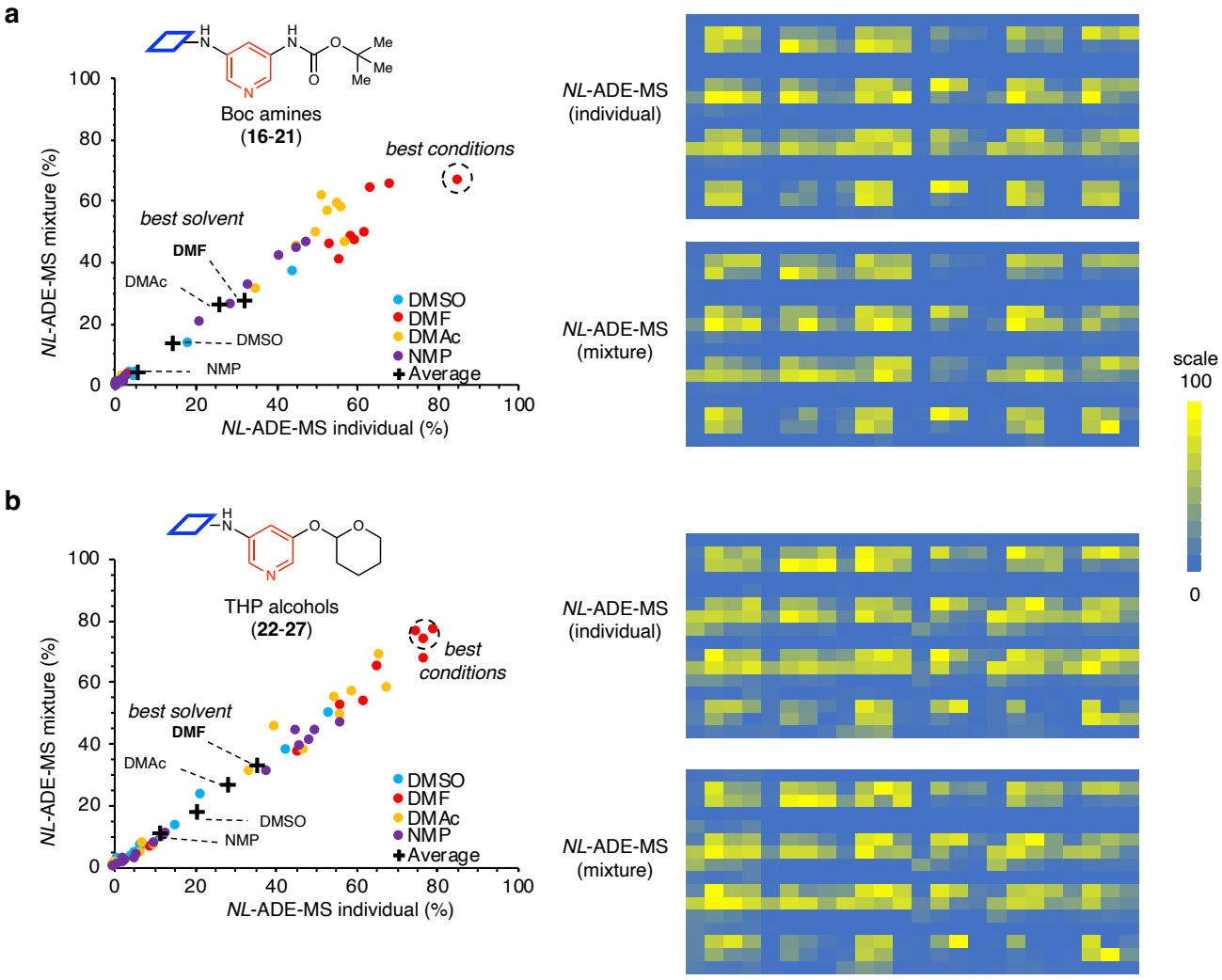
Rank	Base	NL-ADE -MS (%)	LC-MS (%)
1	MTBD	61	81
=2	DBU	56	80
=2	Barton's Base*	56	79

* tBuBrettPhos Pd G4

Extended Data Fig. 8 Rank ordering of reaction conditions for individual boronates. Using 4 TIDA boronates **a**, **1**, **b**, **28**, **c**, **29**, **d**, **30**, and all 96 amines (**35**-**130**) as representative examples, *NL*-ADE-MS allows reaction conditions to be prioritized for specific boronates directly from ultra-high throughput experimentation. These data show excellent correlation against LC-MS data (R^2 (avg) = 0.98). In all cases the best condition was selected or was within 5% where selection of either condition would provide a similar outcome. This subset of the *NL*-ADE-MS data required only 7.68 minutes whereas the equivalent LC-MS dataset required 6.4 days of continuous data collection a 1,200-fold increase in pace.



Extended Data Fig. 9 Rank ordering of reaction conditions for individual amines. Using all 8 TIDA boronates (**1**, **28-34**) and amines (**59-82**) as a representative example, *NL*-ADE-MS allowed reaction conditions to be prioritized for specific amines directly from ultra-high throughput experimentation (structures of amines shown in panel **a**). These data show excellent correlation against LC-MS data (R^2(avg)= 0.91, see supplementary tabular data sheet 22). In 8 of 24 cases the best condition was directly selected, however in many cases condition data were close to one another (i.e. panel **b**). With the goal of assessing the utility of these data for selecting the highest output chemical reaction conditions we applied a nearest neighbor constraint of "within 5%" (panel **c**). This allowed much more effective determination of reaction data, and by using these constrains we found that in 22 of 24 cases (92%) the best conditions selected by *NL*-ADE-MS were within 5% of the top 3 conditions identified by LC-MS.



Extended Data Fig. 10 Comparison of multiplexed and individual *NL*-ADE-MS data. In a 6-plex format products derived from **a**, Boc **2** (**16**-**21**), and **b**, THP **3** (**22**-**27**), were compared across all 64 conditions. These data were in excellent agreement. Multiplexing reduced the data acquisition time from 7.68 mins to 1.28 mins per set of 384 reactions. Shown inset are heatmaps generated from experiments on these two timescales.