Title

Building Block-Centric Approach to DNA-Encoded Library Design.

Permalink

https://escholarship.org/uc/item/0zg4n8mn

Journal

Journal of chemical information and computer sciences, 64(12)

Authors

Fitzgerald, Patrick

Dixit, Anjali

Zhang, Chris

et al.

Publication Date

2024-06-24

DOI

10.1021/acs.jcim.4c00232

Peer reviewed

Article

Building Block-Centric Approach to DNA-Encoded Library Design

Patrick R. Fitzgerald, Anjali Dixit, Chris Zhang, David L. Mobley, and Brian M. Paegel*
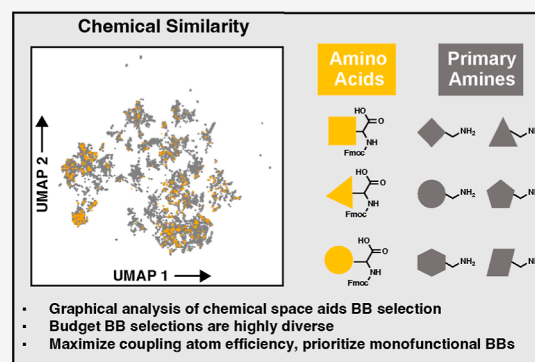
ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: DNA-encoded library technology grants access to nearly infinite opportunities to explore the chemical structure space for drug discovery. Successful navigation depends on the design and synthesis of libraries with appropriate physicochemical properties (PCPs) and structural diversity while aligning with practical considerations. To this end, we analyze combinatorial library design constraints including the number of chemistry cycles, bond construction strategies, and building block (BB) class selection in pursuit of ideal library designs. We compare two-cycle library designs (amino acid + carboxylic acid, primary amine + carboxylic acid) in the context of PCPs and chemical space coverage, given different BB selection strategies and constraints. We find that broad availability of amines and acids is essential for enabling the widest exploration of chemical space. Surprisingly, cost is not a driving factor, and virtually, the same chemical space can be explored with "budget" BBs.

- Graphical analysis of chemical space aids BB selection
- Budget BB selections are highly diverse
- Maximize coupling atom efficiency, prioritize monofunctional BBs

1. INTRODUCTION

DNA-encoded library (DEL) synthesis grants nearly limitless access to a large and novel chemical space—the raw materials of early drug discovery. DELs assign unique DNA barcodes to all possible combinations of small-molecule building blocks (BBs) during encoded split-and-pool combinatorial synthesis.[1] Then, by leveraging the massive throughput of genome sequencers, DELs are analyzed by affinity selection to identify all BB combinations that produce a ligand of the target protein. Numerically large libraries can be rapidly synthesized and screened to furnish hits that occupy a distinct chemical space.[2,3] DEL has identified novel small-molecule hits against many targets[4] and yielded clinical candidates.[5−9] These successes have driven a virtuous cycle of developing new capabilities in library design, synthesis, and analysis.[10]

Library design is a primary factor affecting the likelihood that DEL hits will lead to a development candidate. A successful DEL campaign will produce several hit series with attractive physicochemical properties (PCPs). As an example, Pfizer employs a DEL hit scoring metric that is highly influenced by molecular weight (MW), wherein >600 Da hits are ignored.[11] PCPs tend to inflate undesirably as the number of chemistry cycles increases in a combinatorial library synthesis.[12,13] DEL reaction development efforts have aimed to achieve robust, atom-efficient coupling, but the rate of DEL reaction development has seemingly outpaced the rate of implementation in library synthesis and selection.[14,15] DEL reaction development efforts often focus on synthetic yield and DNA compatibility,[16−18] which are critical features to consider during DEL synthesis. However, additional aspects, such as compatible BB availability, reaction scheme simplicity, and library member PCP space are not widely discussed. These considerations are critical to effective combinatorial library design, having figured prominently in GSK's retrosynthetic combinatorial analysis procedure.[19]

Early DEL reaction sequences provided critical insight for improving DEL design. These library designs,[1,6−8,20] which featured numerous cycles of chemistry (three to four), skewed library PCPs away from the attractive rule-of-5 (RO5)[21] space that is preferred as a starting point for lead optimization.[12] Additionally, use of a core scaffold increased MW and hydrophobicity while confining the library to a specific region of chemical space.[1,20,22] Libraries from this era successfully identified leads that developed into clinical candidates against RIP1[5,6] and sEH,[7,8] but both clinical candidates are truncates from three and four-cycle libraries, respectively. Identifying active truncates is a challenge that has been addressed with the development of on-DNA hit validation methods[23−26] but still requires significant downstream investment in hit evaluation.[10] However, looking for truncates in libraries with many (>3) cycles of chemistry is a low signal-to-noise proposition: library size increases exponentially with the number of chemistry cycles,[12] leading to increased false-negative rates.[27] Therefore, screening a DEL with fewer coupling steps is likely advantageous.[22]

In this work, we specify the principles for building DELs that maximize the likelihood of yielding hits with desirable PCPs for future development. These principles are consistent with our previous assertion that DEL technology, like "click chemistry," should explore easily accessible regions of chemical space.[18] Expanding upon this concept, DEL should prioritize reactions that use widely available BB sets. To demonstrate, we visualize BB coverage of chemical space using a dimensionality reduction technique, uniform manifold approximation and projection (UMAP).[28] From UMAP projections of commonly used BB sets, we demonstrate the effect that various BB selection strategies have on the properties of final output libraries. Collectively, these analyses provide a framework for the computational design and characterization of BBs for DEL synthesis.

2. RESULTS AND DISCUSSION

2.1. Library Design. To illustrate our library design principles, we compare a previously employed two-cycle library design with a plausible alternative two-cycle library design. In our previous library design, DEL1,[29−31] an amine-functionalized *o*-nitroveratryl photocleavable linker, was acylated with Fmoc−amino acids (Fmoc−AAs) in the first chemistry cycle, then the pendant Fmoc was removed, and carboxylic acids were coupled in the second cycle (Figure 1A). DEL1 required four synthetic manipulations, and each library compound photocleaved from beads possessed a pendant primary amide and a BB-connecting amide. In the proposed alternative design, DEL2, an aldehyde-functionalized photocleavable linker is coupled with primary amines in the first cycle through reductive amination, and the resulting secondary amine is acylated with carboxylic acids in the second cycle (Figure 1B). When liberated from the bead, library products contain only a BB-connecting amide.[32] The previous library design effectively contributes an extra −CONH$_2$ from the photocleavage product, which adds two hydrogen bond acceptors, one hydrogen bond donor, and 44 Da.

The DEL2 synthesis scheme offers several advantages over the DEL1 scheme. DEL2 requires fewer synthetic manipulations during library synthesis, introduces less bond construction "baggage," and is overall more atom efficient. Additionally, hits from DEL2 can be synthesized in solution from commercially available materials in a single step, allowing for facile exploration of additional derivatives.[33] DEL2 is likely to be productive against many targets as amides are the most common functional group in bioactive compounds,[34] and their formation is the most prominent reaction in medicinal chemistry hit optimization campaigns.[35] Previous DEL selections highlight the utility of libraries like DEL2.[36−39]

2.2. Building Block UMAP Comparison. We posited that by tapping amines, a substantially larger BB pool than Fmoc−AAs, DEL2 would sample additional valuable chemotypes. To assess the diversity of the BB pools used in DEL1 and DEL2, we analyzed Enamine's commercial Fmoc−AA, primary amine, and carboxylic acid collections using UMAP. We refer to these collections as BB sets. First, we truncated BB sets by the replacement of defining functional groups (NH$_2$, COOH, and *N*-Fmoc) with −H (Table S1). Next, we calculated Morgan ECFP6 fingerprints (radius = 3) for each truncate structure and used the fingerprint to generate a 2D-Tanimoto similarity matrix for all BB truncates. Finally, we employed UMAP to convert a high-dimensional representation of the similarity matrix to a 2D visualization. As the similarity matrix was
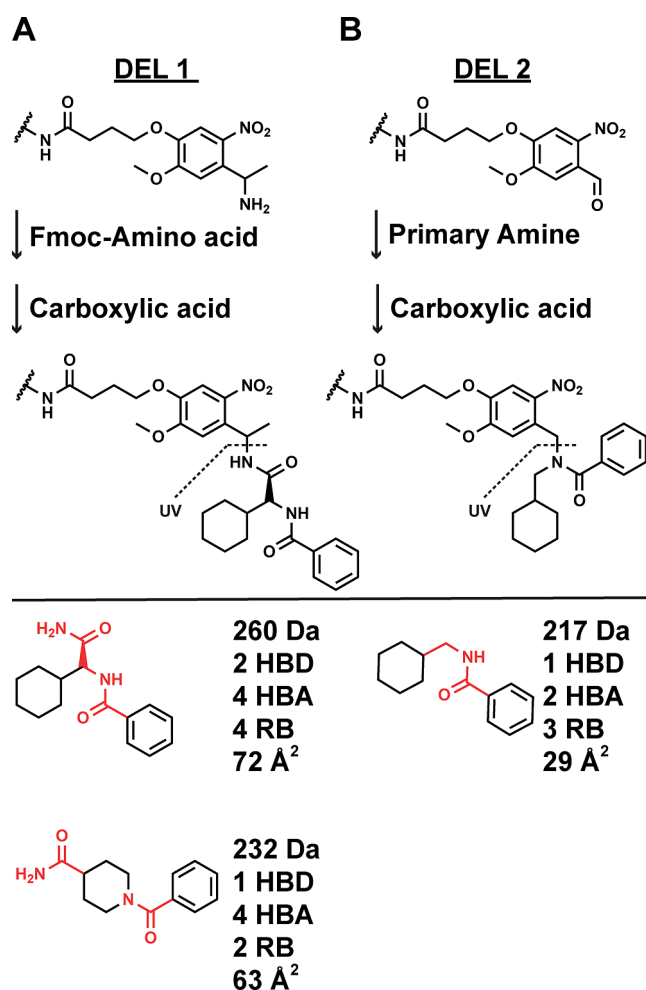

Figure 1. Property-focused analysis of DEL synthesis strategies. (A) Synthesis of DEL1 proceeds on an amine-functionalized photocleavable linker. Fmoc−AA coupling occurs in the first chemistry cycle, followed by carboxylic acid coupling in the second step. Compounds photocleave at the indicated bond (UV) as primary amides. Primary amine and secondary amine containing Fmoc−AAs are depicted along with their calculated properties. The constant library scaffold is indicated (red). (B) DEL2 synthesis starts from an aldehyde-functionalized photocleavable linker. Primary amines are coupled by reductive amination in the first chemistry cycle, followed by carboxylic acid coupling in the second step. Compounds photocleave at the indicated bond (UV) as secondary amides. An example compound is illustrated along with its calculated properties according to RO5 criteria.

generated from all three BB sets, the projection of chemical space was the same for each set, though the projected UMAP distances may not directly correlate with Tanimoto similarity scoring difference. Implementing our UMAP analysis without first truncating the BB sets yielded higher Tanimoto similarity scores and more densely populated UMAP plots, particularly for carboxylic acid and Fmoc−AA BBs (Figure S1).

Fmoc−AAs, the smallest BB set, covered the narrowest area of chemical space (Figure 2A). Primary amines were far more numerous, and their truncates covered the areas of UMAP space occupied by Fmoc−AA truncates while densely populating previously unpopulated regions (Figure 2B). Additionally, when analyzed by direct comparison of truncate structures, very few (*n* = 78) truncate structures were unique from the Fmoc−AA set (Figure S2). Carboxylic acids, upon
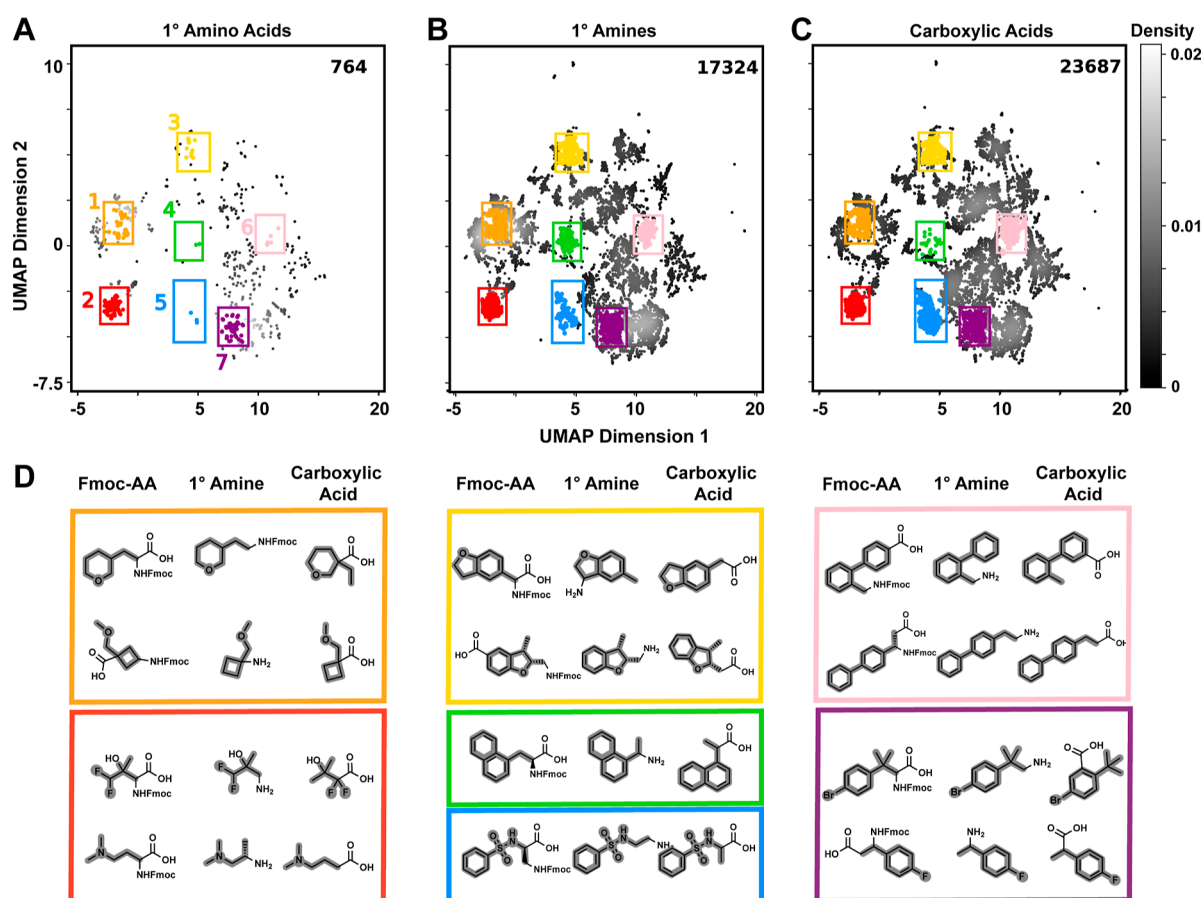
Figure 2. UMAP analysis of Fmoc−amino acid, primary amine, and carboxylic acid BBs from Enamine as truncates. Density plots arranged by chemical similarity are compared for (A) Fmoc−amino acid, (B) primary amine, and (C) carboxylic acid truncates. Grayscale intensity denotes the probability density of points. BB truncate counts in each pool are indicated in the top right of each plot. (D) Example BBs from Fmoc−AA, primary amine, and carboxylic acid sets are depicted for specific regions of UMAP space [colored points/rectangular outlines in (A−C)], and their truncates are indicated (bold). Fmoc−AAs cover a limited portion of the UMAP space relative to primary amines or carboxylic acids.

truncation, covered similar areas of chemical space as both Fmoc−AA and primary amine truncates (Figure 2C). The increased coverage observed for amines and carboxylic acids relative to that for amino acids appeared to be a reflection of increased breadth as well as depth (Figure S3). UMAP analysis readily identified the most prevalent chemotypes, supporting the assumption that distance on the UMAP plot corresponded to chemical similarity (Figure 2D). As examples of overlapping chemotype representation, all three BB sets thoroughly sampled linear aliphatics, cycloalkanes, and substituted benzene chemotypes (regions 1, 2, and 7, respectively). On the other hand, regions represented by 2,3-dihydrobenzofuran, naphthalene, benzylsulfonamide, and biphenyl chemotypes (regions 3, 4, 5, and 6, respectively) are poorly represented by the Fmoc−AA set while being well represented by the primary amine and carboxylic acid BB sets.

The large enhancement in the chemical space coverage of primary amine or carboxylic acid BB sets is a prime advantage of using monofunctional BB sets. Previous analysis established the difference in the quantity of these and other BB classes,[40] yet a direct comparison of their chemical similarity has not been conducted. Our analysis hinted at structural limitations resulting from designing a DEL that includes Fmoc−AAs. Namely, such a library is likely to be biased toward linear aliphatic, cyclic aliphatic, and benzene derivatives as the three main chemotypes. Notably, inhibitors of sEH and Wip1

derived from DEL screening both feature cycloalkane Fmoc−AA BBs.[8,41] The BB chemotypes that are more highly represented in amine and carboxylic acid sets represent valuable materials for exploring chemical space. The substituted dihydrobenzofuran moiety is a key pharmacophore in a potent bromodomain 2 inhibitor.[42,43] Naphthyl groups feature in inhibitors of ADAMTS5,[44] SARS-CoV-2 PLpro,[45] and SARS CoV2 mPro.[46] Sulfonamides are found in multiple approved drugs. Finally, biphenyls are of clear interest to DEL practitioners, given the heavy investment in DNA-compatible Suzuki reaction development.[20,31] These observations in concert with library selection analyses[47] and recent modeling experiments[3] indicate that it is advantageous to design DELs that draw from deep commercial pools.

2.3. BB Diversity as a Function of BB Cost. We next sought to understand how BB cost constraints might influence the accessibility of diverse chemotypes. The cost of Fmoc−AA, primary amine, and carboxylic acid BB sets all conform to roughly bimodal distributions (Figure S4). We stratified BBs by applying cost cutoffs (≤$100, ≤ $250, and ≤$500 per 250 mg) and then generated corresponding UMAP plots (Figure 3) and Venn diagrams (Figure S5). While the density of UMAP plots decreases for all three BB classes with stricter cost filtering, cost most influences Fmoc−AAs (Figure 3A). The two most represented chemotypes within Fmoc−AAs—linear aliphatics (region 2) and substituted benzenes (region 7)—
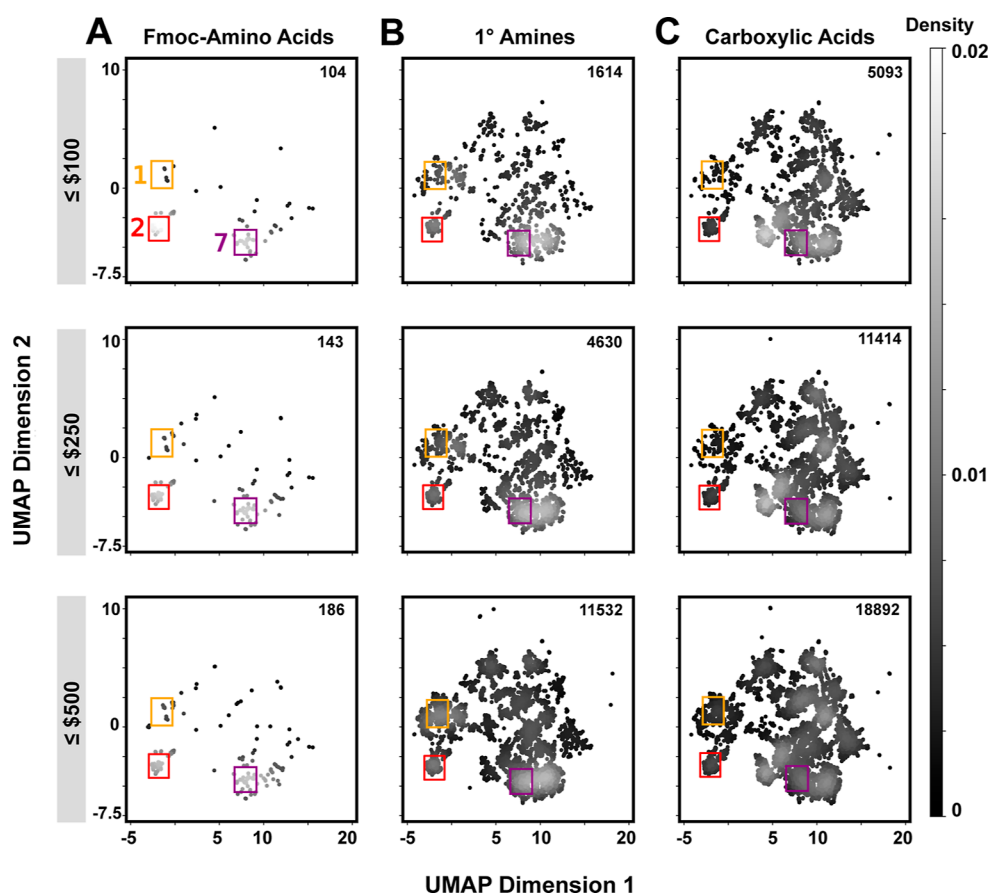
Figure 3. Availability of BB truncates following cost filtering. UMAP analyses are shown for three different cost thresholds (≤$100, ≤$250, and ≤$500 per 250 mg) for (A) Fmoc−AAs, (B) primary amines, and (C) carboxylic acids. Grayscale intensity denotes the probability density of points. BB truncate counts in each pool are indicated on the top right of each plot. Primary amines and carboxylic acids cover more diverse chemical space compared to Fmoc−amino acids at all cost points, especially at the lowest cost cutoff.

remain accessible at the lowest cost filter, but most other regions do not. The third most common grouping, cyclo-alkanes (region **1**), are only accessed above $500/250 mg. For primary amines, the majority of UMAP space remains accessible even at the lowest price cutoff, albeit slightly more sparsely (Figure 3B). Additionally, upon comparison of identical truncate scaffolds (Figure S6), amines are substantially cheaper than their Fmoc−AA counterparts (median cost difference for 250 mg = $509). Even after stringent cost filtering, the carboxylic acid set included in both DEL1 and DEL2 is the most abundant BB class and broadly covers the UMAP space (Figure 3C). At all the selected cost cutoffs, primary amines and carboxylic acids more thoroughly sample the UMAP space compared to Fmoc−AAs.

Our analysis emphasizes the importance of designing library reaction sequences that can tap large and economical BB pools while also minimizing the atoms involved in bond construction and simplifying the interpretation of synthesis products. Library synthesis and hit synthesis budgets may vary drastically between different DEL practitioners, but understanding the effect of cost on potential library outcomes is universally important. Even under generous cost constraints (<$500/250 mg), the Fmoc−AA set samples only 186 truncates and cost-limited Fmoc−AAs even more sparsely sample chemical space. Although our analysis was limited to Fmoc−AAs, primary amines, and carboxylic acids available from one vendor, these BB classes are among the most common in DEL synthesis.[18]

More generally, monofunctional BB classes tend to have more members than bifunctional BB classes,[13] so the trends observed here are likely general. Implementing economical BB classes with robust and broad "click-like" scope[18] also eases the prioritization of hit synthesis and validation, a major bottleneck for DEL technology.[23−26] Looking forward, these simplified library designs incorporating economical BBs further position DEL screening output to feed into direct-to-biology platforms for higher throughput structure exploration and lead optimization via cellular activity assays.[48−50]

2.4. BB Diversity as a Function of Molecular Weight. We suspected that PCP filtering—a common aspect of DEL workflows—would also influence library structural diversity to different extents based on numerical BB set size. Fmoc−AA (ignoring Fmoc), primary amine, and carboxylic acid BB sets have similar MW but different clog P distributions [median MW = 171, 174, and 220 Da, and median clog P = −1.8, 0.69, and 1.1, respectively (Figure S7)]. To analyze the relationship between MW and chemical space, we apply three MW cutoffs (≤200, ≤250, and ≤300 Da) and visualize the resulting subsets using UMAP or Venn diagram (Figures 4 and S8). For the Fmoc−AAs, 70, 94, and 98% remain after filtering at ≤200, ≤250, and ≤300 Da cutoffs, respectively. Similarly, for primary amines, 65, 89, and 95% remain at the same cutoffs. Carboxylic acids experience a more drastic reduction, with 32, 67, and 88% remaining at these cutoffs. For all three sets, no specific regions are disproportionately affected by MW filtering. At all
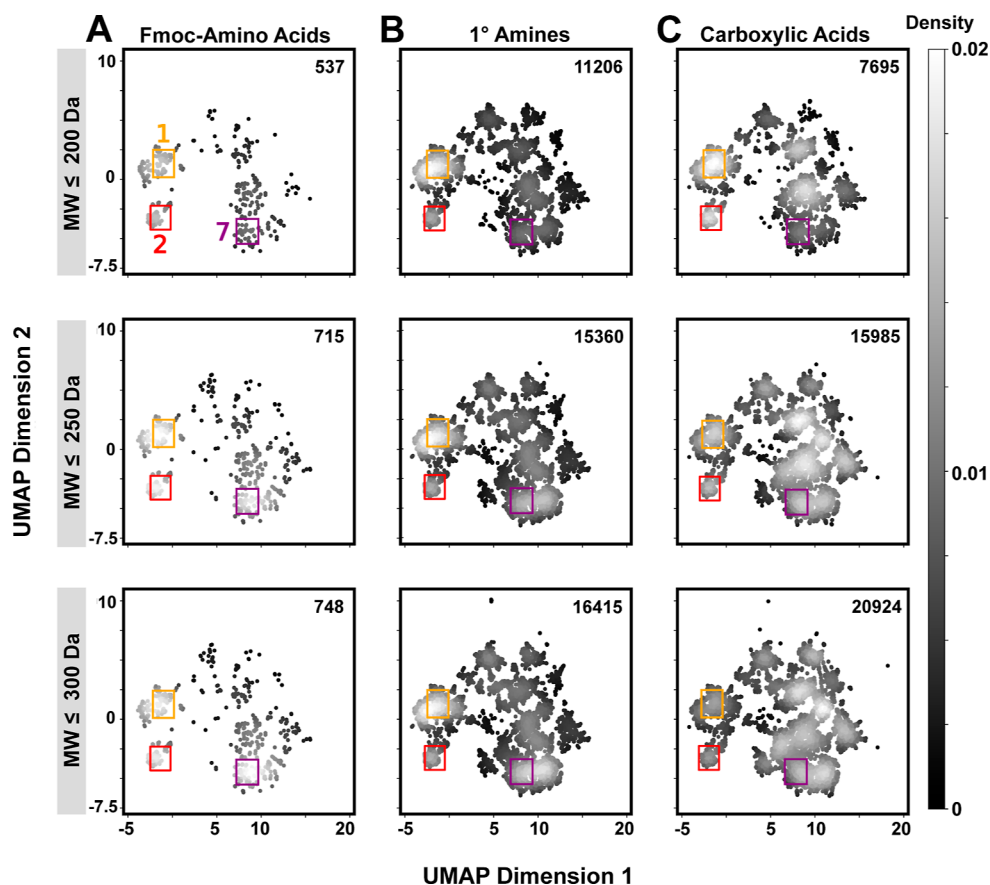
Figure 4. Availability of BB truncates following MW filtering. UMAP analyses are shown for three different MW thresholds (≤200, ≤250, and ≤300 Da) for (A) Fmoc−AAs, (B) primary amines, and (C) carboxylic acids. Grayscale intensity denotes the probability density. BB truncate counts in each pool are indicated on the top right of each plot. Filtering at these MW thresholds does not substantially alter the accessibility of chemical diversity for any of the BB sets. These regions (color coded) are shown to aid in comparison of identical UMAP regions between BB sets and MW filters.
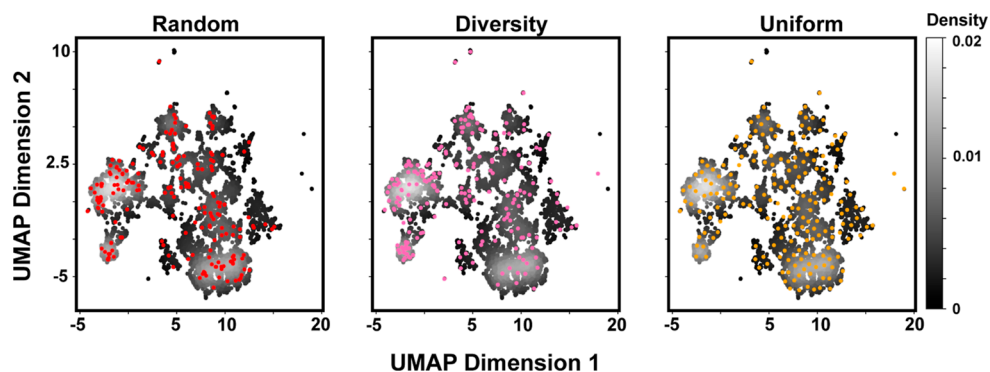

Figure 5. Impact of different BB selection strategies on chemical space sampling. Density plots of primary amine truncates arranged by chemical similarity are overlaid with BB selections (colored points) generated by random, diversity, and uniform methods. Grayscale intensity denotes the probability density of points. Random selection is biased toward dense UMAP regions. Diversity selection improves the sampling of sparsely populated regions but remains biased toward densely populated regions. Uniform selection maximizes the coverage of the UMAP space.

MW cutoffs, primary amine and carboxylic acid sets recapitulate the chemotypes present in the corresponding Fmoc−AA set while also offering numerous additional unique truncates.

UMAP analysis shows that MW filtering does not affect the availability of chemical diversity for any of the BB sets in this study. For the 2-cycle libraries DEL1 and DEL2, the ≤200, ≤250, and ≤300 Da MW cutoffs roughly correspond to library products within the druglike PCP RO5 constraints. These are

important thresholds to consider when balancing chemical diversity against PCPs (e.g., MW). Previous analysis established that reasonable library PCPs are most readily obtained with optimized library design scheme rather than with strict BB filtering.[13] These observations, along with the identification of high MW BBs in previously identified DEL hits have motivated some to prescribe including high MW BBs, though upper limits were not explicitly stated.[2,51] Conversely, others have implemented BB filtering strategies to align their
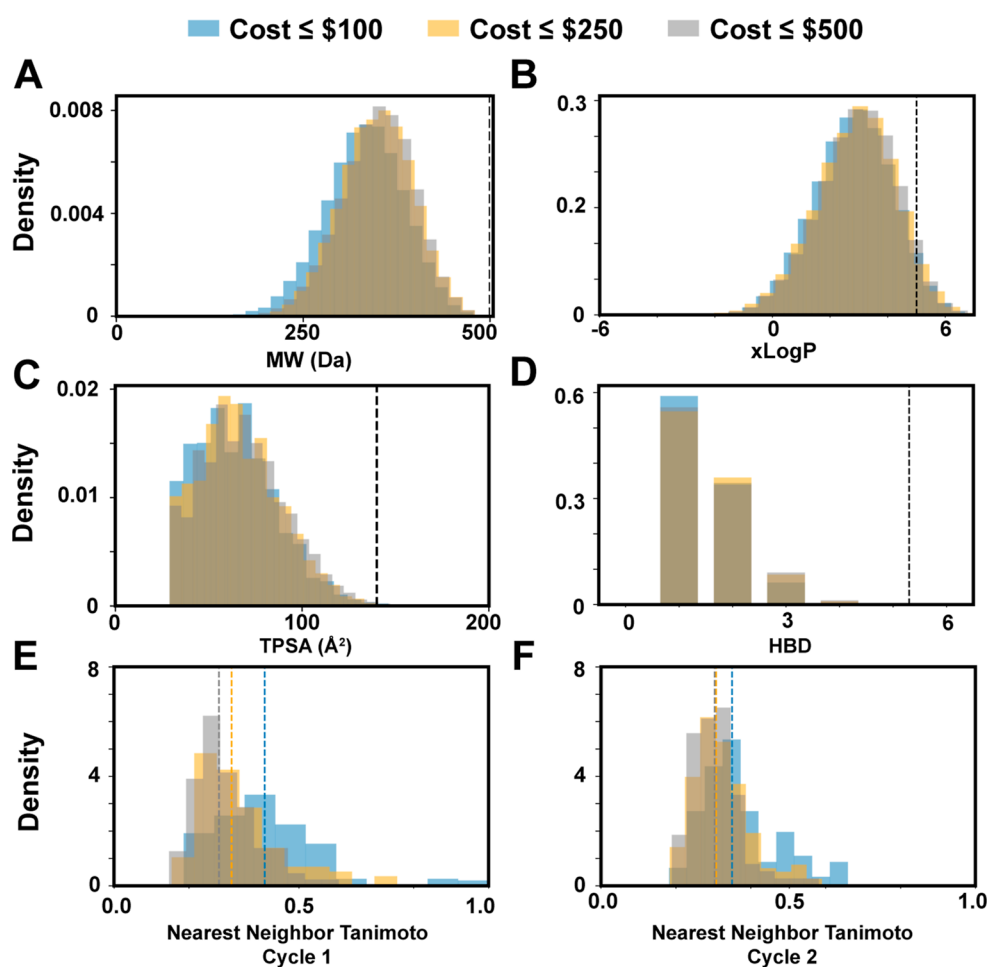
Figure 6. PCP and diversity summary statistics for three two-cycle libraries (condensation of 192 primary amines with 192 carboxylic acids) with different cost constraints. BBs are filtered by cost ($\leq$\$100, $\leq$\$250, and $\leq$500 per 250 mg) and then randomly selected until reaching the intended split size (192). Libraries are enumerated, and their properties are compared. The PCPs of enumerated products including (A) MW, (B) *n*-octanol−water partition coefficient (*x* Log *P*), (C) total polar surface area, and (D) HBD count are calculated and plotted as density plots. Black dashed lines indicate RO5 thresholds. The complete similarity matrix of all BBs within a given library is calculated, from which nn-Tanimoto scores for (E) the primary amine BBs and (F) the carboxylic acid BBs are determined and plotted.

libraries with RO5 PCPs,[52] and separately, medicinal chemists have adopted a "rule of 2" BB criteria during lead optimization.[53] Our analysis supports implementing a MW cutoff during BB selection, as this does not appear to impact the sampling of chemical space. However, several important caveats bear consideration. First, our constraints focus on an analysis of 2-cycle libraries. Additional considerations likely play into BB election and scheme optimization for 3-cycle libraries. Second, the UMAP approach is a powerful tool for high-level visualization of chemical space, but may not capture more detailed features. Along these lines, truncates from a wide range of chemotypes remain represented at even the $\leq$200 Da cutoff, but these are less densely functionalized than their higher MW counterparts.

2.5. BB Diversity as a Function of the BB Selection Method. Having observed the effects of cost and property constraints on BB diversity, we sought to evaluate how different BB selection strategies impact chemical sampling. When sampled 192 times (2 × 96-well plates), the collection of amines containing ~15,000 compounds offers an unfathomably large number of options $\binom{15000}{192}$. We explore three methods for sampling of BBs from very large pools: random,

diversity, and uniform. We applied each strategy to select 192 BBs from the primary amine (Figure 5) and carboxylic acid pools (Figure S9). The random method selects 192 points from the collection of available BBs after all relevant filters have been applied. Random sampling selects more points from the most densely populated regions of chemical space, which follows trends in the commercial collection. The "diversity" method selects BBs with minimized nearest-neighbor Tanimoto similarity scores. We seed the selection with a randomly selected BB, find the most dissimilar BBs, and iterate until reaching 192 BB selections. UMAP analysis of diversity selection shows sampling of regions not covered by random selection, including singletons. Lastly, the "uniform" selection method uses the distances between points in the UMAP space rather than Tanimoto distance, minimizing sampling of densely populated regions of the UMAP space.

While it is tempting to prescribe uniform selection, since it maximizes UMAP coverage, each BB selection method raises distinct considerations. Random selection is biased toward the most abundant commercially available compound classes. Alternatively, strategic selection can increase the number of unique chemical fingerprints sampled while minimizing compound similarity.[54] Minimizing similarity is not necessarily

advantageous. The diversity method intentionally selects exotic and potentially undesirable compounds, such as silicates or perfluorocarbons. Selecting compounds from a similar compound class can be advantageous as certain regions of chemical space correlate with higher proclivity for protein binding. These "privileged" structures are frequently observed in drugs.[55] The GSK benzimidazole scaffold is one such example specific from DEL.[56] Similarity-based selection also biases against small changes to a scaffold, such as the introduction of a single methyl group. These subtle changes can profoundly alter the binding affinity.[57,58] BB similarity across different positions (such as the inclusion of similar chemotypes in position 1 and position 2) is another design consideration. Completely overlapping chemotypes across both positions is likely unwise as this reduces overall library structural diversity. However, duplication of chemotypes across multiple positions could be beneficial for addressing symmetric protein targets[59] or for increasing confidence in hit determination by observing similar chemotypes in multiple positions.[29] Regardless of the selection strategy, high reaction yield should be a prerequisite for BB inclusion.[16,17]

2.6. Library Comparisons. Having evaluated the effects of cost considerations, MW filtering, and selection method on BB diversity, we examine their combined effects during library design. We explore three 192 × 192 libraries following the DEL2 library design (primary amine + carboxylic acid) using a random BB picking strategy. These libraries shared the same PCP cutoffs (MW $\leq$ 250 Da) but differ in their cost constraints ($\leq$\$100, $\leq$\$250, and $\leq$\$500 per 250 mg). We enumerate all three libraries and calculate the PCPs of every library product. The distributions of MW, predicted n-octanol−water partitioning coefficients ($x \log P$), polar surface area, and hydrogen bond donor (HBD) count were nearly identical for all three libraries (Figure 6A−D). We use 2D nearest-neighbor Tanimoto scores to characterize intralibrary BB similarity as a measure of chemical diversity. Analyzing the primary amines (Figure 6E), the cheapest set has higher nearest-neighbor (nn) scores (median nn-Tanimoto$_{amines \leq \$100}$ = 0.40) compared with the middle and highest cost sets, which were similar (median nn-Tanimoto$_{amines \leq \$250}$ = 0.31 and median nn-Tanimoto$_{amines \leq \$500}$ = 0.28). Analyzing the carboxylic acid set (Figure 6F), a similar trend was observed, with the lowest cost set having the highest nearest-neighbor (nn) scores (median nn-Tanimoto$_{COOH \leq \$100}$ = 0.35) compared with the middle and highest cost sets (median nn-Tanimoto$_{COOH \leq \$250}$ = 0.31 and median nn-Tanimoto$_{COOH \leq \$500}$ = 0.30). These trends in nn-Tanimoto scores as a function of BB-cost were reproducible following replicate sampling (Figure S10).The total cost of BBs in each cycle scales proportionally to the maximum allowed BB cost in the library. For BB costs restricted at $\leq$\$100, $\leq$\$250, and $\leq$\$500 per 250 mg, the total cost is roughly \$10,000, \$25,000, and \$50,000, respectively, for both cycle 1 and cycle 2 BBs. We compared the enumerated library generated by random selection with a proportional sampling of an Enamine lead-like set and found that the DEL set had minimal chemical similarity to the lead-like set. Additionally, DEL members share higher nearest-neighbor Tanimoto similarity with other DEL members than lead-like compounds do with other lead-like compounds (Figure S11).

PCP distributions of enumerated libraries were minimally affected by cost filtering of BB sets, while intralibrary BB similarity followed a law of diminishing returns with respect to cost. These observations are specific to the split-size and BB

sets examined but highlight important potential consequences of cost filtering. For large BB sets, cost filtering does not appear to impact PCP space accessibility as all three libraries shared similar PCPs. In contrast, BB diversity noticeably improved by increasing the cost filter from $\leq$\$100 to $\leq$\$250. However, continuing to increase the cost threshold to $\leq$\$500 did not offer substantial improvement. Our comparison of the enumerated DEL with a comparative lead-like set illustrated important differences between DEL chemical space and compound screening collections. DEL compounds had higher intralibrary similarity, as expected from combinatorial library products featuring overlapping BBs. While ChEMBL has been used as a biologically active reference set to map DEL compounds,[3,40] it is unclear how desirable the overlap of these chemical spaces should be because lead novelty is one of the technology's primary advantages. DELs are orders of magnitude larger than HTS collections,[4] DEL selection routinely identifies hits dissimilar from HTS screening sets,[5,6,36,47] and DEL hits frequently access unexpected binding modes.[60]

2.7. Outlook. The BB-centric approach to the DEL design of this study ultimately aimed to understand ideality in the context of chemistry and BB selection. Our library design goals included minimizing nondiversity elements, limiting DELs to two to three couplings, leveraging numerically large BB sets, and prioritizing facile hit synthesis off-DNA. These principles should be broadly applicable, although this initial study considers a narrower scope of unbiased two-cycle libraries using only three BB classes. While this approach is readily extended to other common DEL BB classes, it does not account for stereochemistry, 3D structure,[61] or experimentally determined BB coupling yield. Additionally, it is still unclear which areas of chemical space will be the most productive as DEL often uncovers hit compounds with unexpected binding modes.[60] Further work is necessary to develop guidelines for three-cycle libraries or modes to identify covalent inhibitors,[62] protein−protein interaction disruptors,[63] or molecular degraders[64,65] to accommodate their expanded PCPs. Additional studies comparing different dimensionality reduction techniques for DEL analysis, such as TMAP,[66] GTM,[40] or UMAP,[67] are also likely to be valuable. Finally, while model library size was constrained to 192 × 192 (~37k numerical diversity), which is at least an order of magnitude smaller than some DELs, the field is migrating toward smaller libraries with more lead-like to drug-like properties.[36,68] These trends have accompanied experimental design innovations, enabling the detection of weaker binding events from lower cycle libraries through photoactivatable handles.[69−71]

Computational tools to understand and manage diversity are critical given the near limitless expanses of chemical space that are accessible to DEL. Estimates for the size of drug-like chemical space range from 10^{23} to 10^{60} depending on the computational method employed.[72] Practical DEL sizes sample only a minuscule fraction of this space. One possible avenue for computational intervention is the use of theoretical database mining to identify unexplored regions of chemical space.[73] Current methods have been employed to select BBs that narrow the PCP space of DELs,[52] enumerate libraries,[74] and assess DEL chemical diversity.[40] One such modeling experiment highlighted library architecture and BB class as key drivers of library PCP and diversity, respectively.[3] From a collection of 2497 DELs, the top 5 most diverse libraries were three-cycle libraries generated from robust coupling reactions

with diverse BBs. Conversely, the least diverse libraries were generated from multiple heterocyclization reactions. While this computational assessment of diversity does not directly predict library productivity, simple library designs with diverse BBs have proven highly productive.[2,18,47]

Moving forward, improvements in DEL computational analysis tools and experimental approaches will empower new hit finding capabilities. Meta-analysis comparing the productivity of different library designs across several targets[10,47,75] could provide valuable insights for the DEL community without requiring the disclosure of sensitive intellectual property. Such disclosures could address fundamental DEL design questions regarding optimal library size, productivity of two- vs three-cycle libraries, productivity of different BB classes, and the impact of different BB selection strategies. As DEL is arguably guided by BB diversity, new methods to synthesize novel BBs efficiently should be a high priority.[53,76,77] Additional on-DNA reactions that utilize abundant BBs in creative ways[36] and further development of broad-scope cross-coupling reactions[78,79] are also likely to be highly productive. Such advances, in concert with evolving computational tools, library design schemes, selection methods, and screening strategies, will continue to hone the technology's already indelible impact on small molecule discovery.

3. METHODS

3.1. Building Block Acquisition. We acquired separate BB catalog files (in .sdf format) for purchasable Fmoc-protected amino acid BBs, primary amine BBs, and carboxylic acid BBs directly through correspondence with a representative at Enamine. Updated BB lists can also be downloaded by creating an account on the Enamine web site or through the database search feature on DataWarrior (V5.5.0).[80] Each catalog file contains information regarding basic PCPs, cost, IUPAC name, catalog information, and RDKit molecule rendering for BBs. We provide the exact raw structure files we used on our GitHub, denoted with the file name following the format: "[functional group]_stock.sdf".

3.2. Building Block Truncation. For each Enamine BB stock file, we used the OELibraryGen function from Open-Eye's OEChem module (v.2021.1.1)[81] to generate truncated versions of each BB via reaction SMIRKS. For Fmoc-protected amino acid BBs, we replaced the—Fmoc, $-NH_2$, and $-COOH$ groups with $-H$. For amine BBs, we replaced $-NH_2$ groups with $-H$. Lastly, for carboxylic acid BBs, we replaced $-COOH$ groups with $-H$ (Figure S1). The SMIRKS patterns and associated file preparation scripts can be found on our GitHub.

3.3. Building Block Filtering. We applied a SMIRKS filter to remove BBs containing secondary amines, secondary amino acids, and aniline nitrogens. To filter BBs by cost, we used pandas (v.1.2.1)[82] to write queries based on our established cost thresholds. For PCP filtering, we calculated the relevant properties for each BB before writing queries to filter based on RO3 (MW < 300, cLog $P \leq 3$, HBD ≤ 3, HBA ≤ 3) or RO5 (MW < 500, cLog P < 5, HBD ≤ 5, HBA ≤ 10) specifications. We exported files containing the canonical SMILES, Enamine ID, cost, and PCP of each BB as well as the canonical SMILES of its corresponding truncate as .csv files for further analysis. We provide these cleaned structure files in our GitHub and denote each file as "[functional group]_df.csv".

3.4. Similarity Calculation. We used RDKit (v.2020.09.01)[83] to calculate the 2D Tanimoto similarity among all truncates. We used the canonical SMILES for each truncate to generate an RDKit molecule object and transformed the molecule into a Morgan fingerprint with radius of 3 bonds and 2048 bits. For each truncate, we calculated its similarity to all other truncates, yielding an all-by-all similarity matrix. We provide a script for this procedure on our GitHub.textbfChemical space projections. To visualize the coverage of chemical space of different truncate selections, we used a dimensionality reduction technique known as UMAP (v.0.5.3).[28] UMAP assigns coordinates to each truncate based on the truncate's chemical distance to other truncates. We converted the calculated all-by-all similarity matrix into an all-by-all distance matrix and input this into the UMAP algorithm. Here, we define chemical distance by subtracting the 2D Tanimoto similarity from 1, leading to distance values ranging from 0 (very similar) to 1 (very dissimilar). The resultant output of UMAP was a set of 2D coordinates for each truncate, where chemically similar BBs were closer together in the UMAP projection and more dissimilar BBs were further away from each other, providing a depiction of the chemical space. We combined these projections with a Gaussian kernel density estimate from SciPy (v.1.5.3) to illustrate the coverage of the chemical space in this work.

3.5. Building Block Selection Strategies. We investigated three different BB selection strategies in this work: random, diversity, and uniform. For random selection, we used pandas (v.1.2.1) to randomly sample a specified number of BBs. In diversity selection, we used the MinMaxPicker function from RDKit to select a maximally diverse set of compounds given an initial seed and a specified number of compounds to select. The algorithm performs similarity calculations based on the underlying Morgan fingerprints. Lastly, in uniform selection, we sample from the density of the UMAP projection. In this final selection strategy, we establish a minimum distance threshold in the UMAP space and, given an initial truncate, identify truncates that are at least the established distance away in UMAP space.

3.6. Library Enumeration and Analysis. Once we had selected a set of BBs for each library cycle, we performed a library enumeration using the OELibraryGen function from OpenEye's OEChem module (v.2021.1.1).[81] In this work, we limited our analysis to two-cycle libraries where BBs in cycle 1 contain an amine group and BBs in cycle 2 contain a carboxylic acid group. Thus, our only "BB-linking" reaction was amide formation, which we represented as a SMIRKS reaction string. Once we enumerated all library products, we calculated various PCPs using OpenEye's OEMolProp module (v.2021.1.1).[81]

■ ASSOCIATED CONTENT

Data Availability Statement

All original data files and code for this analysis are made publicly available on GitHub at https://github.com/MobleyLab/DEL_BB_design.

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.4c00232.

> BB truncation, overlap between full BB truncates, overlap between truncates subset by cost, overlap between truncates subset by MW, and selection strategies for selecting carboxylic acids (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Brian M. Paegel − *Department of Pharmaceutical Sciences, University of California, Irvine, California 92697, United States; Department of Chemistry, University of California, Irvine, California 92697, United States;* ⓞ orcid.org/0000-0002-6531-6693; Email: bpaegel@uci.edu

Authors

Patrick R. Fitzgerald − *Skaggs Doctoral Program in the Chemical and Biological Sciences, Scripps Research, La Jolla, California 92037, United States;* ⓞ orcid.org/0000-0002-9324-2134

Anjali Dixit − *Department of Pharmaceutical Sciences, University of California, Irvine, California 92697, United States;* ⓞ orcid.org/0000-0002-4997-8005

Chris Zhang − *Department of Chemistry, University of California, Irvine, California 92697, United States;* ⓞ orcid.org/0000-0002-7298-7145

David L. Mobley − *Department of Pharmaceutical Sciences, University of California, Irvine, California 92697, United States; Department of Chemistry, University of California, Irvine, California 92697, United States;* ⓞ orcid.org/0000-0002-1083-5533

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jcim.4c00232

Notes

The authors declare the following competing financial interest(s): D.L.M. serves on the scientific advisory boards of Anagenex and OpenEye Scientific Software, Cadence Molecular Sciences. He is also an Open Science Fellow with Psivant.

■ ABBREVIATIONS

DEL, DNA-encoded library; BB, building block; PCP, physicochemical property; HTS, high-throughput screening; HBD, hydrogen bond donor; HBA, hydrogen bond acceptor

■ REFERENCES

(1) Clark, M. A.; Acharya, R. A.; Arico-Muendel, C. C.; Belyanskaya, S. L.; Benjamin, D. R.; Carlson, N. R.; Centrella, P. A.; Chiu, C. H.; Creaser, S. P.; Cuozzo, J. W.; Davie, C. P.; Ding, Y.; Franklin, G. J.; Franzen, K. D.; Gefter, M. L.; Hale, S. P.; Hansen, N. J. V.; Israel, D. I.; Jiang, J.; Kavarana, M. J.; Kelley, M. S.; Kollmann, C. S.; Li, F.; Lind, K.; Mataruse, S.; Medeiros, P. F.; Messer, J. A.; Myers, P.; O'Keefe, H.; Oliff, M. C.; Rise, C. E.; Satz, A. L.; Skinner, S. R.; Svendsen, J. L.; Tang, L.; van Vloten, K.; Wagner, R. W.; Yao, G.; Zhao, B.; Morgan, B. A.; Van Vloten, K.; Wagner, R. W.; Yao, G.; Zhao, B.; Morgan, B. A.; van Vloten, K.; Wagner, R. W.; Yao, G.; Zhao, B.; Morgan, B. A. Design, Synthesis and Selection of DNA-Encoded Small-Molecule Libraries. *Nat. Chem. Biol.* **2009**, *5*, 647−654.

(2) Satz, A. L. What Do You Get from DNA-Encoded Libraries? *ACS Med. Chem. Lett.* **2018**, *9*, 408−410.

(3) Pikalyova, R.; Zabolotna, Y.; Horvath, D.; Marcou, G.; Varnek, A. Chemical Library Space: Definition and DNA-Encoded Library Comparison Study Case. *J. Chem. Inf. Model.* **2023**, *63*, 4042−4055.

(4) Sunkari, Y. K.; Siripuram, V. K.; Nguyen, T. L.; Flajolet, M. High-Power Screening (HPS) Empowered by DNA-Encoded Libraries. *Trends Pharmacol. Sci.* **2022**, *43*, 4−15.

(5) Harris, P. A.; King, B. W.; Bandyopadhyay, D.; Berger, S. B.; Campobasso, N.; Capriotti, C. A.; Cox, J. A.; Dare, L.; Dong, X.; Finger, J. N.; Grady, L. C.; Hoffman, S. J.; Jeong, J. U.; Kang, J.; Kasparcova, V.; Lakdawala, A. S.; Lehr, R.; McNulty, D. E.; Nagilla, R.; Ouellette, M. T.; Pao, C. S.; Rendina, A. R.; Schaeffer, M. C.; Summerfield, J. D.; Swift, B. A.; Totoritis, R. D.; Ward, P.; Zhang, A.; Zhang, D.; Marquis, R. W.; Bertin, J.; Gough, P. J. DNA-Encoded Library Screening Identifies Benzo[b][1,4]oxazepin-4-ones as Highly Potent and Monoselective Receptor Interacting Protein 1 Kinase Inhibitors. *J. Med. Chem.* **2016**, *59*, 2163−2178.

(6) Harris, P. A.; Berger, S. B.; Jeong, J. U.; Nagilla, R.; Bandyopadhyay, D.; Campobasso, N.; Capriotti, C. A.; Cox, J. A.; Dare, L.; Dong, X.; Eidam, P. M.; Finger, J. N.; Hoffman, S. J.; Kang, J.; Kasparcova, V.; King, B. W.; Lehr, R.; Lan, Y.; Leister, L. K.; Lich, J. D.; MacDonald, T. T.; Miller, N. A.; Ouellette, M. T.; Pao, C. S.; Rahman, A.; Reilly, M. A.; Rendina, A. R.; Rivera, E. J.; Schaeffer, M. C.; Sehon, C. A.; Singhaus, R. R.; Sun, H. H.; Swift, B. A.; Totoritis, R. D.; Vossenkämper, A.; Ward, P.; Wisnoski, D. D.; Zhang, D.; Marquis, R. W.; Gough, P. J.; Bertin, J. Discovery of a First-in-Class Receptor Interacting Protein 1 (RIP1) Kinase Specific Clinical Candidate (GSK2982772) for the Treatment of Inflammatory Diseases. *J. Med. Chem.* **2017**, *60*, 1247−1261.

(7) Belyanskaya, S. L.; Ding, Y.; Callahan, J. F.; Lazaar, A. L.; Israel, D. I. Discovering Drugs with DNA-Encoded Library Technology: From Concept to Clinic with an Inhibitor of Soluble Epoxide Hydrolase. *ChemBioChem* **2017**, *18*, 837−842.

(8) Ding, Y.; Belyanskaya, S.; DeLorey, J. L.; Messer, J. A.; Joseph Franklin, G.; Centrella, P. A.; Morgan, B. A.; Clark, M. A.; Skinner, S. R.; Dodson, J. W.; Li, P.; Marino, J. P.; Israel, D. I. Discovery of Soluble Epoxide Hydrolase Inhibitors Through DNA-Encoded Library Technology (ELT). *Bioorg. Med. Chem.* **2021**, *41*, 116216.

(9) Cuozzo, J. W.; Clark, M. A.; Keefe, A. D.; Kohlmann, A.; Mulvihill, M.; Ni, H.; Renzetti, L. M.; Resnicow, D. I.; Ruebsam, F.; Sigel, E. A.; Thomson, H. A.; Wang, C.; Xie, Z.; Zhang, Y. Novel Autotaxin Inhibitor for the Treatment of Idiopathic Pulmonary Fibrosis: A Clinical Candidate Discovered Using DNA-Encoded Chemistry. *J. Med. Chem.* **2020**, *63*, 7840−7856.

(10) Reiher, C. A.; Schuman, D. P.; Simmons, N.; Wolkenberg, S. E. Trends in Hit-to-Lead Optimization Following DNA-Encoded Library Screens. *ACS Med. Chem. Lett.* **2021**, *12*, 343−350.

(11) Foley, T. L.; Burchett, W.; Chen, Q.; Flanagan, M. E.; Kapinos, B.; Li, X.; Montgomery, J. I.; Ratnayake, A. S.; Zhu, H.; Peakman, M. C. Selecting Approaches for Hit Identification and Increasing Options by Building the Efficient Discovery of Actionable Chemical Matter from DNA-Encoded Libraries. *SLAS Discovery* **2021**, *26*, 263−280.

(12) Franzini, R. M.; Randolph, C. Chemical Space of DNA-Encoded Libraries: Miniperspective. *J. Med. Chem.* **2016**, *59*, 6629−6644.

(13) Zabolotna, Y.; Volochnyuk, D. M.; Ryabukhin, S. V.; Horvath, D.; Gavrilenko, K. S.; Marcou, G.; Moroz, Y. S.; Oksiuta, O.; Varnek, A. A Close-up Look at the Chemical Space of Commercially Available Building Blocks for Medicinal Chemistry. *J. Chem. Inf. Model.* **2022**, *62*, 2171−2185.

(14) Götte, K.; Chines, S.; Brunschweiger, A. Reaction Development for DNA-encoded Library Technology: From Evolution to Revolution? *Tetrahedron Lett.* **2020**, *61*, 151889.

(15) Fair, R. J.; Walsh, R. T.; Hupp, C. D. The Expanding Reaction Toolkit for DNA-Encoded Libraries. *Bioorg. Med. Chem. Lett.* **2021**, *51*, 128339.

(16) Malone, M. L.; Paegel, B. M. What is a "DNA-Compatible" Reaction? *ACS Comb. Sci.* **2016**, *18*, 182−187.

(17) Ratnayake, A. S.; Flanagan, M. E.; Foley, T. L.; Smith, J. D.; Johnson, J. G.; Bellenger, J.; Montgomery, J. I.; Paegel, B. M. A Solution Phase Platform to Characterize Chemical Reaction Compatibility with DNA-Encoded Chemical Library Synthesis. *ACS Comb. Sci.* **2019**, *21*, 650−655.

(18) Fitzgerald, P. R.; Paegel, B. M. DNA-Encoded Chemistry: Drug Discovery from a Few Good Reactions. *Chem. Rev.* 2021, *121*, 7155–7177.

(19) Lewell, X. Q.; Judd, D. B.; Watson, S. P.; Hann, M. M. RECAP Retrosynthetic Combinatorial Analysis Procedure: A Powerful New Technique for Identifying Privileged Molecular Fragments with Useful Applications in Combinatorial Chemistry. *J. Chem. Inf. Comput. Sci.* 1998, *38*, 511–522.

(20) Ding, Y.; Franklin, G. J.; DeLorey, J. L.; Centrella, P. A.; Mataruse, S.; Clark, M. A.; Skinner, S. R.; Belyanskaya, S. Design and Synthesis of Biaryl DNA-Encoded Libraries. *ACS Comb. Sci.* 2016, *18*, 625–629.

(21) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Delivery Rev.* 1997, *23*, 3–25.

(22) Zhang, Y.; Franzini, R. M. Design Considerations in Constructing and Screening DNA-Encoded Libraries. In *DNA Encoded Libraries*; Brunschweiger, A., Young, D. W., Eds.; Springer International Publishing: Cham, 2022; Chapter 4, pp 123–143.

(23) Prati, L.; Bigatti, M.; Donckele, E. J.; Neri, D.; Samain, F. On-DNA Hit Validation Methodologies for Ligands Identified From DNA-Encoded Chemical Libraries. *Biochem. Biophys. Res. Commun.* 2020, *533*, 235–240.

(24) Su, W.; Ge, R.; Ding, D.; Chen, W.; Wang, W. W.; Yan, H.; Wang, W. W.; Yuan, Y.; Liu, H.; Zhang, M.; Zhang, J.; Shu, Q.; Satz, A. L.; Kuai, L. Triaging of DNA-Encoded Library Selection Results by High-Throughput Resynthesis of DNA-Conjugate and Affinity Selection Mass Spectrometry. *Bioconjugate Chem.* 2021, *32*, 1001–1007.

(25) Xia, B.; Franklin, G. J.; Lu, X.; Bedard, K. L.; Grady, L. C.; Summerfield, J. D.; Shi, E. X.; King, B. W.; Lind, K. E.; Chiu, C.; Watts, E.; Bodmer, V.; Bai, X.; Marcaurelle, L. A. DNA-Encoded Library Hit Confirmation: Bridging the Gap between On-DNA and Off-DNA Chemistry. *ACS Med. Chem. Lett.* 2021, *12*, 1166–1172.

(26) Ratnayake, A. S.; Flanagan, M. E.; Foley, T. L.; Hultgren, S. L.; Bellenger, J.; Montgomery, J. I.; Lall, M. S.; Liu, B.; Ryder, T.; Kölmel, D. K.; Shavnya, A.; Feng, X.; Lefker, B.; Byrnes, L. J.; Sahasrabudhe, P. V.; Farley, K. A.; Chen, S.; Wan, J. Toward the Assembly and Characterization of an Encoded Library Hit Confirmation Platform: Bead-Assisted Ligand Isolation Mass Spectrometry (BALI-MS). *Bioorg. Med. Chem.* 2021, *41*, 116205.

(27) Satz, A. L.; Hochstrasser, R.; Petersen, A. C. Analysis of Current DNA Encoded Library Screening Data Indicates Higher False Negative Rates for Numerically Larger Libraries. *ACS Comb. Sci.* 2017, *19*, 234–238.

(28) McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* 2018, *3*, 861.

(29) Cochrane, W. G.; Malone, M. L.; Dang, V. Q.; Cavett, V. J.; Satz, A. L.; Paegel, B. M. Activity-Based DNA-Encoded Library Screening. *ACS Comb. Sci.* 2019, *21*, 425–435.

(30) Hackler, A. L.; FitzGerald, F. G.; Dang, V. Q.; Satz, A. L.; Paegel, B. M. Off-DNA DNA-Encoded Library Affinity Screening. *ACS Comb. Sci.* 2020, *22*, 25–34.

(31) Fitzgerald, P. R.; Cochrane, W. G.; Paegel, B. M. Dose–Response Activity-Based DNA-Encoded Library Screening. *ACS Med. Chem. Lett.* 2023, *14*, 1295–1303.

(32) Minkwitz, R.; Meldal, M. Application of a Photolabile Backbone Amide Linker for Cleavage of Internal Amides in the Synthesis towards Melanocortin Subtype-4 Agonists. *QSAR Comb. Sci.* 2005, *24*, 343–353.

(33) Sadybekov, A. A.; Sadybekov, A. V.; Liu, Y.; Iliopoulos-Tsoutsouvas, C.; Huang, X.-P.; Pickett, J.; Houser, B.; Patel, N.; Tran, N. K.; Tong, F.; Zvonok, N.; Jain, M. K.; Savych, O.; Radchenko, D. S.; Nikas, S. P.; Petasis, N. A.; Moroz, Y. S.; Roth, B. L.; Makriyannis, A.; Katritch, V. Synthon-Based Ligand Discovery in Virtual Libraries of Over 11 Billion Compounds. *Nature* 2022, *601*, 452–459.

(34) Ertl, P.; Altmann, E.; McKenna, J. M. The Most Common Functional Groups in Bioactive Molecules and How Their Popularity Has Evolved over Time. *J. Med. Chem.* 2020, *63*, 8408–8418.

(35) Tomberg, A.; Boström, J. Can Easy Chemistry Produce Complex, Diverse, and Novel Molecules? *Drug Discovery Today* 2020, *25*, 2174–2181.

(36) Nissink, J. W. M.; Bazzaz, S.; Blackett, C.; Clark, M. A.; Collingwood, O.; Disch, J. S.; Gikunju, D.; Goldberg, K.; Guilinger, J. P.; Hardaker, E.; Hennessy, E. J.; Jetson, R.; Keefe, A. D.; McCoull, W.; McMurray, L.; Olszewski, A.; Overman, R.; Pflug, A.; Preston, M.; Rawlins, P. B.; Rivers, E.; Schimpl, M.; Smith, P.; Truman, C.; Underwood, E.; Warwicker, J.; Winter-Holt, J.; Woodcock, S.; Zhang, Y. Generating Selective Leads for Mer Kinase Inhibitors-Example of a Comprehensive Lead-Generation Strategy. *J. Med. Chem.* 2021, *64*, 3165–3184.

(37) Ryan, M. D.; Parkes, A. L.; Corbett, D.; Dickie, A. P.; Southey, M.; Andersen, O. A.; Stein, D. B.; Barbeau, O. R.; Sanzone, A.; Thommes, P.; Barker, J.; Cain, R.; Compper, C.; Dejob, M.; Dorali, A.; Etheridge, D.; Evans, S.; Faulkner, A.; Gadouleau, E.; Gorman, T.; Haase, D.; Holbrow-Wilshaw, M.; Krulle, T.; Li, X.; Lumley, C.; Mertins, B.; Napier, S.; Odedra, R.; Papadopoulos, K.; Roumpelakis, V.; Spear, K.; Trimby, E.; Williams, J.; Zahn, M.; Keefe, A. D.; Zhang, Y.; Soutter, H. T.; Centrella, P. A.; Clark, M. A.; Cuozzo, J. W.; Dumelin, C. E.; Deng, B.; Hunt, A.; Sigel, E. A.; Troast, D. M.; DeJonge, B. L. M. Discovery of Novel UDP-N-Acetylglucosamine Acyltransferase (LpxA) Inhibitors with Activity against Pseudomonas aeruginosa. *J. Med. Chem.* 2021, *64*, 14377–14425.

(38) Veerman, J. J. N.; Bruseker, Y. B.; Damen, E.; Heijne, E. H.; van Bruggen, W.; Hekking, K. F. W.; Winkel, R.; Hupp, C. D.; Keefe, A. D.; Liu, J.; Thomson, H. A.; Zhang, Y.; Cuozzo, J. W.; McRiner, A. J.; Mulvihill, M. J.; van Rijnsbergen, P.; Zech, B.; Renzetti, L. M.; Babiss, L.; Müller, G. Discovery of 2,4–1H-Imidazole Carboxamides as Potent and Selective TAK1 Inhibitors. *ACS Med. Chem. Lett.* 2021, *12*, 555–562.

(39) Lee, E. C. Y.; McRiner, A. J.; Georgiadis, K. E.; Liu, J.; Wang, Z.; Ferguson, A. D.; Levin, B.; von Rechenberg, M.; Hupp, C. D.; Monteiro, M. I.; Keefe, A. D.; Olszewski, A.; Eyermann, C. J.; Centrella, P.; Liu, Y.; Arora, S.; Cuozzo, J. W.; Zhang, Y.; Clark, M. A.; Huguet, C.; Kohlmann, A. Discovery of Novel, Potent Inhibitors of Hydroxy Acid Oxidase 1 (HAO1) Using DNA-Encoded Chemical Library Screening. *J. Med. Chem.* 2021, *64*, 6730–6744.

(40) Pikalyova, R.; Zabolotna, Y.; Volochnyuk, D. M.; Horvath, D.; Marcou, G.; Varnek, A. Exploration of the Chemical Space of DNA-encoded Libraries. *Mol. Inf.* 2022, *41*, 2100289.

(41) Gilmartin, A. G.; Faitg, T. H.; Richter, M.; Groy, A.; Seefeld, M. A.; Darcy, M. G.; Peng, X.; Federowicz, K.; Yang, J.; Zhang, S.-Y.; Minthorn, E.; Jaworski, J.-P.; Schaber, M.; Martens, S.; McNulty, D. E.; Sinnamon, R. H.; Zhang, H.; Kirkpatrick, R. B.; Nevins, N.; Cui, G.; Pietrak, B.; Diaz, E.; Jones, A.; Brandt, M.; Schwartz, B.; Heerding, D. A.; Kumar, R. Allosteric Wip1 phosphatase inhibition through flap-subdomain interaction. *Nat. Chem. Biol.* 2014, *10*, 181–187.

(42) Yu, Z.; Ku, A. F.; Anglin, J. L.; Sharma, R.; Ucisik, M. N.; Faver, J. C.; Li, F.; Nyshadham, P.; Simmons, N.; Sharma, K. L.; Nagarajan, S.; Riehle, K.; Kaur, G.; Sankaran, B.; Storl-Desmond, M.; Palmer, S. S.; Young, D. W.; Kim, C.; Matzuk, M. M. Discovery and Characterization of Bromodomain 2−Specific Inhibitors of BRDT. *Proc. Natl. Acad. Sci. U.S.A.* 2021, *118*, No. e2021102118.

(43) Lucas, S. C. C.; Atkinson, S. J.; Chung, C.-w.; Davis, R.; Gordon, L.; Grandi, P.; Gray, J. J. R.; Grimes, T.; Phillipou, A.; Preston, A. G.; Prinjha, R. K.; Rioja, I.; Taylor, S.; Tomkinson, N. C. O.; Wall, I.; Watson, R. J.; Woolven, J.; Demont, E. H. Optimization of a Series of 2,3-Dihydrobenzofurans as Highly Potent, Second Bromodomain (BD2)-Selective, Bromo and Extra-Terminal Domain (BET) Inhibitors. *J. Med. Chem.* 2021, *64*, 10711–10741.

(44) Deng, H.; O'Keefe, H.; Davie, C. P.; Lind, K. E.; Acharya, R. A.; Franklin, G. J.; Larkin, J.; Matico, R.; Neeb, M.; Thompson, M. M.; Lohr, T.; Gross, J. W.; Centrella, P. A.; O'Donovan, G. K.; Bedard, K. L. S.; van Vloten, K.; Mataruse, S.; Skinner, S. R.; Belyanskaya, S. L.; Carpenter, T. Y.; Shearer, T. W.; Clark, M. A.; Cuozzo, J. W.; Arico-

Muendel, C. C.; Morgan, B. A. Discovery of Highly Potent and Selective Small Molecule ADAMTS-5 Inhibitors That Inhibit Human Cartilage Degradation via Encoded Library Technology (ELT). *J. Med. Chem.* **2012**, *55*, 7061−7079.

(45) Fu, Z.; Huang, B.; Tang, J.; Liu, S.; Liu, M.; Ye, Y.; Liu, Z.; Xiong, Y.; Zhu, W.; Cao, D.; Li, J.; Niu, X.; Zhou, H.; Zhao, Y. J.; Zhang, G.; Huang, H. The Complex Structure of GRL0617 and SARS-CoV-2 PLpro Reveals a Hot Spot for Antiviral Drug Discovery. *Nat. Commun.* **2021**, *12*, 488.

(46) Jimmidi, R.; Chamakuri, S.; Lu, S.; Ucisik, M. N.; Chen, P.-J.; Bohren, K. M.; Moghadasi, S. A.; Versteeg, L.; Nnabuife, C.; Li, J.-Y.; Qin, X.; Chen, Y.-C.; Faver, J. C.; Nyshadham, P.; Sharma, K. L.; Sankaran, B.; Judge, A.; Yu, Z.; Li, F.; Pollet, J.; Harris, R. S.; Matzuk, M. M.; Palzkill, T.; Young, D. W. DNA-encoded Chemical Libraries Yield Non-Covalent and Non-Peptidic SARS-CoV-2 Main Protease Inhibitors. *Commun. Chem.* **2023**, *6*, 164.

(47) Eidam, O.; Satz, A. L. Analysis of the Productivity of DNA Encoded Libraries. *MedChemComm* **2016**, *7*, 1323−1331.

(48) Thomas, R. P.; Heap, R. E.; Zappacosta, F.; Grant, E. K.; Pogány, P.; Besley, S.; Fallon, D. J.; Hann, M. M.; House, D.; Tomkinson, N. C. O.; Bush, J. T. A Direct-to-Biology High-Throughput Chemistry Approach to Reactive Fragment Screening. *Chem. Sci.* **2021**, *12*, 12098−12106.

(49) Hendrick, C. E.; Jorgensen, J. R.; Chaudhry, C.; Strambeanu, I. I.; Brazeau, J.-F.; Schiffer, J.; Shi, Z.; Venable, J. D.; Wolkenberg, S. E. Direct-to-Biology Accelerates PROTAC Synthesis and the Evaluation of Linker Effects on Permeability and Degradation. *ACS Med. Chem. Lett.* **2022**, *13*, 1182−1190.

(50) Mahjour, B.; Zhang, R.; Shen, Y.; McGrath, A.; Zhao, R.; Mohamed, O. G.; Lin, Y.; Zhang, Z.; Douthwaite, J. L.; Tripathi, A.; Cernak, T. Rapid Planning and Analysis of High-Throughput Experiment Arrays for Reaction Discovery. *Nat. Commun.* **2023**, *14*, 3924.

(51) Zhang, Y.; Clark, M. Design Concepts for DNA-Encoded Library Synthesis. *Bioorg. Med. Chem.* **2021**, *41*, 116189.

(52) Zhu, H.; Flanagan, M. E.; Stanton, R. V. Designing DNA Encoded Libraries of Diverse Products in a Focused Property Space. *J. Chem. Inf. Model.* **2019**, *59*, 4645−4653.

(53) Goldberg, F. W.; Kettle, J. G.; Kogej, T.; Perry, M. W. D.; Tomkinson, N. P. Designing Novel Building Blocks Is an Overlooked Strategy To Improve Compound Quality. *Drug Discovery Today* **2015**, *20*, 11−17.

(54) Shi, Y.; von Itzstein, M. How Size Matters: Diversity for Fragment Library Design. *Molecules* **2019**, *24*, 2838.

(55) Carbery, A.; Skyner, R.; von Delft, F.; Deane, C. M. Fragment Libraries Designed to Be Functionally Diverse Recover Protein Binding Information More Efficiently Than Standard Structurally Diverse Libraries. *J. Med. Chem.* **2022**, *65*, 11404−11413.

(56) Ding, Y.; Chai, J.; Centrella, P. A.; Gondo, C.; DeLorey, J. L.; Clark, M. A. Development and Synthesis of DNA-Encoded Benzimidazole Library. *ACS Comb. Sci.* **2018**, *20*, 251−255.

(57) Schönherr, H.; Cernak, T. Profound Methyl Effects in Drug Discovery and a Call for New C−H Methylation Reactions. *Angew. Chem., Int. Ed.* **2013**, *52*, 12256−12267.

(58) Pinheiro, P. d. S. M.; Franco, L. S.; Fraga, C. A. M. The Magic Methyl and Its Tricks in Drug Discovery and Development. *Pharmaceuticals* **2023**, *16*, 1157.

(59) Kempf, D. J.; Sham, H. L.; Marsh, K. C.; Flentge, C. A.; Betebenner, D.; Green, B. E.; McDonald, E.; Vasavanonda, S.; Saldivar, A.; Wideburg, N. E.; Kati, W. M.; Ruiz, L.; Zhao, C.; Fino, L.; Patterson, J.; Molla, A.; Plattner, J. J.; Norbeck, D. W. Discovery of Ritonavir, a Potent Inhibitor of HIV Protease with High Oral Bioavailability and Clinical Efficacy. *J. Med. Chem.* **1998**, *41*, 602−617.

(60) Collie, G. W.; Clark, M. A.; Keefe, A. D.; Madin, A.; Read, J. A.; Rivers, E. L.; Zhang, Y. Screening Ultra-Large Encoded Compound Libraries Leads to Novel Protein−Ligand Interactions and High Selectivity. *J. Med. Chem.* **2024**, *67*, 864−884.

(61) Sauer, W. H. B.; Schwarz, M. K. Molecular Shape Diversity of Combinatorial Libraries: A Prerequisite for Broad Bioactivity. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 987−1003.

(62) Guilinger, J. P.; Archna, A.; Augustin, M.; Bergmann, A.; Centrella, P. A.; Clark, M. A.; Cuozzo, J. W.; Däther, M.; Guié, M. A.; Habeshian, S.; Kiefersauer, R.; Krapp, S.; Lammens, A.; Lercher, L.; Liu, J.; Liu, Y.; Maskos, K.; Mrosek, M.; Pflügler, K.; Siegert, M.; Thomson, H. A.; Tian, X.; Zhang, Y.; Konz Makino, D. L.; Keefe, A. D. Novel Irreversible Covalent BTK Inhibitors Discovered Using DNA-Encoded Chemistry. *Bioorg. Med. Chem.* **2021**, *42*, 116223.

(63) Silvestri, A. P.; Zhang, Q.; Ping, Y.; Muir, E. W.; Zhao, J.; Chakka, S. K.; Wang, G.; Bray, W. M.; Chen, W.; Fribourgh, J. L.; Tripathi, S.; He, Y.; Rubin, S. M.; Satz, A. L.; Pye, C. R.; Kuai, L.; Su, W.; Schwochert, J. A. DNA-Encoded Macrocyclic Peptide Libraries Enable the Discovery of a Neutral MDM2−p53 Inhibitor. *ACS Med. Chem. Lett.* **2023**, *14*, 820−826.

(64) Disch, J. S.; Duffy, J. M.; Lee, E. C. Y.; Gikunju, D.; Chan, B.; Levin, B.; Monteiro, M. I.; Talcott, S. A.; Lau, A. C.; Zhou, F.; Kozhushnyan, A.; Westlund, N. E.; Mullins, P. B.; Yu, Y.; von Rechenberg, M.; Zhang, J.; Arnautova, Y. A.; Liu, Y.; Zhang, Y.; McRiner, A. J.; Keefe, A. D.; Kohlmann, A.; Clark, M. A.; Cuozzo, J. W.; Huguet, C.; Arora, S. Bispecific Estrogen Receptor α Degraders Incorporating Novel Binders Identified Using DNA-Encoded Chemical Library Screening. *J. Med. Chem.* **2021**, *64*, 5049−5066.

(65) Chen, Q.; Liu, C.; Wang, W.; Meng, X.; Cheng, X.; Li, X.; Cai, L.; Luo, L.; He, X.; Qu, H.; Luo, J.; Wei, H.; Gao, S.; Liu, G.; Wan, J.; Israel, D. I.; Li, J.; Dou, D. Optimization of PROTAC Ternary Complex Using DNA Encoded Library Approach. *ACS Chem. Biol.* **2023**, *18*, 25−33.

(66) Probst, D.; Reymond, J.-L. Visualization of Very Large High-dimensional Data Sets as Minimum Spanning Trees. *J. Cheminf.* **2020**, *12*, 12.

(67) Zhang, C.; Pitman, M.; Dixit, A.; Leelananda, S.; Palacci, H.; Lawler, M.; Belyanskaya, S.; Grady, L.; Franklin, J.; Tilmans, N.; Mobley, D. L. Building Block-Based Binding Predictions for DNA-Encoded Libraries. *J. Chem. Inf. Model.* **2023**, *63*, 5120−5132.

(68) McCoull, W.; Boyd, S.; Brown, M. R.; Coen, M.; Collingwood, O.; Davies, N. L.; Doherty, A.; Fairley, G.; Goldberg, K.; Hardaker, E.; He, G.; Hennessy, E. J.; Hopcroft, P.; Hodgson, G.; Jackson, A.; Jiang, X.; Karmokar, A.; Lainé, A. L.; Lindsay, N.; Mao, Y.; Markandu, R.; McMurray, L.; McLean, N.; Mooney, L.; Musgrove, H.; Nissink, J. W. M.; Pflug, A.; Reddy, V. P.; Rawlins, P. B.; Rivers, E.; Schimpl, M.; Smith, G. F.; Tentarelli, S.; Travers, J.; Troup, R. I.; Walton, J.; Wang, C.; Wilkinson, S.; Williamson, B.; Winter-Holt, J.; Yang, D.; Zheng, Y.; Zhu, Q.; Smith, P. D. Optimization of an Imidazo[1,2-a]pyridine Series to Afford Highly Selective Type I1/2 Dual Mer/Axl Kinase Inhibitors within VivoEfficacy. *J. Med. Chem.* **2021**, *64*, 13524−13539.

(69) Sannino, A.; Gironda-Martínez, A.; Gorre, É. M. D.; Prati, L.; Piazzi, J.; Scheuermann, J.; Neri, D.; Donckele, E. J.; Samain, F. Critical Evaluation of Photo-cross-linking Parameters for the Implementation of Efficient DNA-Encoded Chemical Library Selections. *ACS Comb. Sci.* **2020**, *22*, 204−212.

(70) Ma, H.; Murray, J. B.; Luo, H.; Cheng, X.; Chen, Q.; Song, C.; Duan, C.; Tan, P.; Zhang, L.; Liu, J.; Morgan, B. A.; Li, J.; Wan, J.; Baker, L. M.; Finnie, W.; Guetzoyan, L.; Harris, R.; Hendrickson, N.; Matassova, N.; Simmonite, H.; Smith, J.; Hubbard, R. E.; Liu, G. PAC-FragmentDEL − Photoactivated Covalent Capture of DNA-Encoded Fragments for Hit Discovery. *RSC Med. Chem.* **2022**, *13*, 1341−1349.

(71) Wu, X.; Chen, Y.; Lu, W.; Jin, R.; Lu, X. Quantitative Validation and Application of the Photo-Cross-Linking Selection for Double-Stranded DNA-Encoded Libraries. *Bioconjugate Chem.* **2022**, *33*, 1818−1824.

(72) Polishchuk, P. G.; Madzhidov, T. I.; Varnek, A. Estimation of the Size of Drug-Like Chemical Space Based on GDB-17 Data. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 675−679.

(73) Buehler, Y.; Reymond, J.-L. Molecular Framework Analysis of the Generated Database GDB-13s. *J. Chem. Inf. Model.* **2023**, *63*, 484−492.

(74) Martín, A.; Nicolaou, C. A.; Toledo, M. A. Navigating the DNA Encoded Libraries Chemical Space. *Commun. Chem.* **2020**, *3*, 127.

(75) Machutta, C. A.; Kollmann, C. S.; Lind, K. E.; Bai, X.; Chan, P. F.; Huang, J.; Ballell, L.; Belyanskaya, S.; Besra, G. S.; Barros-Aguirre, D.; Bates, R. H.; Centrella, P. A.; Chang, S. S.; Chai, J.; Choudhry, A. E.; Coffin, A.; Davie, C. P.; Deng, H.; Deng, J.; Ding, Y.; Dodson, J. W.; Fosbenner, D. T.; Gao, E. N.; Graham, T. L.; Graybill, T. L.; Ingraham, K.; Johnson, W. P.; King, B. W.; Kwiatkowski, C. R.; Lelièvre, J.; Li, Y.; Liu, X.; Lu, Q.; Lehr, R.; Mendoza-Losana, A.; Martin, J.; McCloskey, L.; McCormick, P.; O'Keefe, H. P.; O'Keeffe, T.; Pao, C.; Phelps, C. B.; Qi, H.; Rafferty, K.; Scavello, G. S.; Steiginga, M. S.; Sundersingh, F. S.; Sweitzer, S. M.; Szewczuk, L. M.; Taylor, A.; Toh, M. F.; Wang, J.; Wang, M.; Wilkins, D. J.; Xia, B.; Yao, G.; Zhang, J.; Zhou, J.; Donahue, C. P.; Messer, J. A.; Holmes, D.; Arico-Muendel, C. C.; Pope, A. J.; Gross, J. W.; Evindar, G. Prioritizing Multiple Therapeutic Targets in Parallel Using Automated DNA-Encoded Library Screening. *Nat. Commun.* **2017**, *8*, 16081.

(76) Helal, C. J.; Bundesmann, M.; Hammond, S.; Holmstrom, M.; Klug-Mcleod, J.; Lefker, B. A.; McLeod, D.; Subramanyam, C.; Zakaryants, O.; Sakata, S. Quick Building Blocks (QBB): An Innovative and Efficient Business Model to Speed Medicinal Chemistry Analog Synthesis. *ACS Med. Chem. Lett.* **2019**, *10*, 1104−1109.

(77) Wang, Y.; Haight, I.; Gupta, R.; Vasudevan, A. What is in Our Kit? An Analysis of Building Blocks Used in Medicinal Chemistry Parallel Libraries. *J. Med. Chem.* **2021**, *64*, 17115−17122.

(78) Wang, J.; Lundberg, H.; Asai, S.; Martín-Acosta, P.; Chen, J. S.; Brown, S.; Farrell, W.; Dushin, R. G.; O'Donnell, C. J.; Ratnayake, A. S.; Richardson, P.; Liu, Z.; Qin, T.; Blackmond, D. G.; Baran, P. S. Kinetically Guided Radical-Based Synthesis of $C(sp^3)-C(sp^3)$ linkages on DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E6404−E6410.

(79) Kölmel, D. K.; Meng, J.; Tsai, M. H.; Que, J.; Loach, R. P.; Knauber, T.; Wan, J.; Flanagan, M. E. On-DNA Decarboxylative Arylation: Merging Photoredox with Nickel Catalysis in Water. *ACS Comb. Sci.* **2019**, *21*, 588−597.

(80) Sander, T.; Freyss, J.; Von Korff, M.; Rufener, C. DataWarrior: An Open-Source Program for Chemistry Aware Data Visualization and Analysis. *J. Chem. Inf. Model.* **2015**, *55*, 460−473.

(81) *OpenEye Toolkits*, version 1.1; OpenEye Scientific, 2021, http://www.eyesopen.com.

(82) Reback, J.; McKinney, W. *Pandas-Dev/pandas: Pandas 1.2.1*; Zenodo, 2021..

(83) Landrum, G. *RDKit/RDKit: 2020_09_1 (Q3 2020)*; Zenodo, 2020..