Article

DNA barcoding a complete matrix of stereoisomeric small molecules

Christopher J. Gerry, Mathias J. Wawer, Paul A Clemons, and Stuart L. Schreiber

Just Accepted

DNA barcoding a complete matrix of stereoisomeric small molecules

Christopher J. Gerry,[†,‡] Mathias J. Wawer,[‡] Paul A. Clemons,[‡,*] Stuart L. Schreiber[†,‡,*]

Affiliations:

[†]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA

[‡]Chemical Biology and Therapeutics Science Program, Broad Institute, 415 Main Street, Cambridge, Massachusetts 02142, USA

*To whom correspondence should be addressed

pclemons@broadinstitute.org; stuart_schreiber@harvard.edu

1

Abstract

It is challenging to incorporate stereochemical diversity and topographic complexity into DNA-encoded libraries (DELs) because DEL syntheses cannot fully exploit the capabilities of modern synthetic organic chemistry. Here, we describe the design, construction, and validation of DOS-DEL-1, a library of 107,616 DNA-barcoded chiral 2,3-disubsituted azetidines and pyrrolidines. We used stereospecific C–H arylation chemistry to furnish complex scaffolds primed for DEL synthesis, and we developed an improved on-DNA Suzuki reaction to maximize library quality. We then studied both the structural diversity of the library and the physicochemical properties of individual compounds using Tanimoto multi-fusion similarity analysis, among other techniques. These analyses revealed not only that most DOS-DEL-1 members have "drug-like" properties, but also that the library more closely resembles compound collections derived from diversity synthesis than those from other sources (*e.g.*, commercial vendors). Finally, we performed validation screens against horseradish peroxidase and carbonic anhydrase IX, and we developed a novel, Poisson-based statistical framework to analyze the results. A set of assay positives were successfully translated into potent carbonic anhydrase inhibitors (IC_{50} = 20.1–68.7 nM), which confirmed the success of the synthesis and screening procedures. These results establish a strategy to synthesize DELs with scaffold-based stereochemical diversity and complexity that does not require the development of novel DNA-compatible chemistry.

Introduction

Chemical probes can modulate protein activities selectively, rapidly, and reversibly, enabling them to interrogate a wide range of therapeutic hypotheses.[1] Many outstanding problems in biomedical research, however, cannot be addressed by currently available tool compounds, so small molecules that act *via* novel mechanisms of action (nMoAs) are needed.

DNA-encoded library (DEL) technology provides a means to identify small-molecule "binders" that interact with biomacromolecules in novel ways. DELs comprise small molecules that are covalently linked to unique DNA sequences that encode, most frequently, their synthetic histories.[2,3] Because the DNA barcode enables pooled storage and screening, large libraries can be screened in a single affinity-based experiment with an immobilized macromolecular target. Inhibitors of kinases,[4] phosphatases,[5] metalloproteases,[6] and hydrolases,[7] among others,[8–10] have emerged from DELs in recent years. Beyond inhibiting enzyme activity, binders may perturb biological systems by altering the interactomes of their protein binding partners (*i.e.*, inducing novel protein associations that rewire cellular circuitry). They may also change the dynamic properties of their targets, which can alter cellular half-lives (degraders and stabilizers), rates of post-translational modifications, or protein activities (*e.g.*, by gaining neo-functions).[11,12] Binders can modulate protein behavior both as monovalent compounds[13–15] and as bivalent molecules in which two binders are covalently tethered.[16–18]

Due to the challenges associated with DEL synthesis, many DNA-barcoded compounds exhibit a narrow range of structural features and architectures. For example, reactions commonly used in DEL syntheses (*e.g.*, cross-couplings, S_NAr reactions, and amidations) tend to generate libraries rich in sp^2-hybridized carbons and amide bonds.[19–21] Moreover, DELs are constructed using split–pool synthesis, so they are often enriched in compounds that resemble traditional combinatorial chemistry reaction products: high molecular weight, low rigidity, and few stereogenic elements.[10,22] Recent advances in DNA-compatible methodology have begun to expand the types of transformations, building blocks, and structural features that may be incorporated into DELs.[23–30] However, performing complexity-generating reactions in DEL syntheses remains difficult due to the inherent presence of both DNA and water during the reactions, low substrate concentrations, high yield and purity requirements, and other factors.[26,31] Approaches to improve the tolerance of the DNA barcode to high levels of organic solvent using cationic resins are currently being explored.[32] Though these methods can expand the scope of DNA-compatible chemistry dramatically, they present their own set of challenges when applied to library syntheses.

An alternative strategy for addressing the structural gaps in current DELs is to integrate the principles of diversity-oriented synthesis (DOS). DOS uses strategically designed sequences of complexity-generating transformations to yield compounds with structural features often found in naturally occurring small molecules ("natural products"), such as spirocyclic ring systems and series of contiguous stereocenters.[33–35] These compounds of "intermediate complexity"—more complex than the contents of many screening libraries but less complex than natural products— exhibit properties that are both correlated with success in the clinic and favorable for protein binding.[36–38] Phenotypic screening experiments with non-encoded libraries derived from DOS, whose members contain chemical features associated with natural products, have identified nMoA compounds that engage a diverse range of targets.[39] It is not yet known whether these types of compounds can similarly expand the capabilities of target-based screening by enabling DELs to identify binders of proteins that are currently challenging to engage.

Here, we describe the design, synthesis, structural analysis, and test screening of a library of 107,616 DNA-barcoded saturated *N*-heterocycles, which we refer to as "DOS-DEL-1." To synthesize DOS-DEL-1, we leveraged stereospecific C–H arylation chemistry to build complex scaffolds that were attached to DNA for DEL synthesis. We then examined the physicochemical

3

properties of the final compounds and compared DOS-DEL-1 to non-encoded small-molecule libraries. Finally, we performed affinity-based enrichment experiments to confirm the success of DOS-DEL-1's synthesis procedure. This study illustrates that a DEL comprising all possible stereoisomers of novel scaffolds can be constructed without the arduous development of new DNA-compatible chemistry. DOS-DEL-1 and the procedures used to characterize it enable the future exploitation of novel DELs and their comparative analyses to existing DELs.

Results

Library design. We aimed to synthesize a DEL featuring under-represented chemical features like stereochemical diversity, topographic complexity, and small rings. Compounds containing these features may be generated *via* Pd-mediated C–H arylation of azetidines and pyrrolidines,[40] but this reaction cannot currently be performed in the presence of DNA. We instead envisioned a pathway that would incorporate C–H arylation before DNA attachment, where all the tools of modern synthesis are available.

We designed DOS-DEL-1 to comprise a complete matrix of stereoisomers of 2,3-disubstituted azetidines and pyrrolidines (Fig. 1a). While azetidines and pyrrolidines have similar reactivities, their minor differences in ring size and geometry can alter compound topography dramatically.[41] Synthesizing all stereoisomers of these complex scaffolds facilitates the study from primary screening results of structure–activity relationships (SAR), including stereochemistry-based SAR; enrichment in a barcode associated with one stereoisomer but not others is evidence of a specific binding event. This approach has benefitted both non-encoded[42] and DNA-barcoded[43] screening libraries.

DOS-DEL-1 contains approximately 100,000 compounds. Current DELs can theoretically contain >10^{12} distinct compounds,[20] but evidence is lacking that larger numbers produce greater screening successes, and indeed there are suggestions that the opposite might be true.[44] We therefore focused on ensuring that our DNA barcodes were encoding the correct compounds (*i.e.,* not reaction by-products or unreacted starting materials). Furthermore, a smaller library size permits the library's inherently unequal distribution of barcodes to be considered during enrichment calculations. While this focus on quality often came at the expense of quantity, DOS-DEL-1 is similar in size to the original (non-encoded) Broad DOS library that has yielded many nMoA compounds despite its relatively modest size.[39]

Scaffold synthesis. All four stereoisomers of each of the azetidine and pyrrolidine scaffolds were synthesized and prepared for DNA attachment (Fig. 1b). To synthesize one set of enantiomers, D-azetidine-2-carboxylic acid (**1a**) and D-proline (**1b**) were first converted to aminoquinolines **2a** and **2b**.[40] To retain the *cis* stereodiad, **2a** was treated with excess Boc anhydride and catalytic DMAP to generate the corresponding bis-Boc product, which was then exposed to H_2O_2 and LiOH to afford carboxylic acid **3a**. The *trans* diastereomer **3b** was obtained *via* Boc protection of the amine, followed by treatment with excess NaOH at 110 °C. The proline derivatives **3c** and **3d** were synthesized in a similar manner that accounted for the increased steric hindrance around the amide. The corresponding enantiomers were synthesized analogously.

These eight scaffolds were then prepared for attachment to an amine-functionalized duplex DNA "headpiece"[21] (DNA-HP; Fig. S1). For each intermediate, a spacer was installed to distance the densely functionalized core from the site of DNA conjugation. The resulting NHS esters **4a-4d** and *ent*-**4a-4d** were readily attached to DNA (Table S1).

DOS-DEL-1 library synthesis. The on-DNA synthetic pathway was designed to maximize the yield and purity of the final compounds. The Boc protecting group is more easily cleaved from the

4

azetidines than from the pyrrolidines (Tables S2–S3 and Figs. S2–S3), so we decided to perform this transformation directly after DNA attachment; this ordering allows each set of substrates to undergo deprotection separately before being pooled for appendage diversification. Reductive amination and sulfonylation were identified as high-yielding *N*-capping reactions (Tables S4–S5), and the Suzuki reaction was chosen to functionalize the aryl iodide. The Suzuki reaction proceeded in low yield and generated multiple by-products when performed in the presence of a free amine (Fig. S4), so it was designated as the last step in the pathway. Conditions based on known DNA-compatible Suzuki reactions resulted in unacceptably high dehalogenation of the aryl iodide substrate.[45] Pd(dppf)Cl$_2$ was discovered to be a superior alternative to Pd(PPh$_3$)$_4$, eventually affording higher yields and less deiodination with fewer equivalents of palladium after substantial optimization (Table S6).

We next determined the building blocks that would be used to realize the full library. Appendages contribute substantially to the physicochemical properties of DEL members,[3] so we avoided compounds with high molecular weight and lipophilicity when selecting candidate building blocks. Incomplete reactions and undesired by-products can compromise DEL screens,[46] so only building blocks that achieved ≥85% conversion with all tested scaffolds were included. The actual scaffolds to be used for library synthesis—rather than model systems—were used during these experiments. Overall, 42 sulfonyl chlorides, 72 aldehydes, and 117 boronic acids were validated (Tables S7–S9; Supplementary Dataset 1).

Following the successful synthesis of a 16-member pilot library (Figs. S5–S9), we performed a full-library synthesis (Fig. 1c). After scaffold encoding, DNA attachment, and Boc deprotection, the resulting free amines were pooled for appendage diversification *via N*-capping and again for cross-coupling, each step receiving a DNA barcode (Supplementary Dataset 2). The final library contains 107,616 distinct compounds derived from 8 scaffolds, 114 *N*-capping (BB1) elements, and 118 Suzuki-derived (BB2) elements (117 boronic acids plus one blank well).

Cheminformatic analysis of DOS-DEL-1. To determine the physicochemical properties of the compounds that compose DOS-DEL-1 and its subsets (Fig. 2a), all library members were enumerated computationally. The vast majority of compounds in DOS-DEL-1 fall within the "drug-likeness" guidelines outlined by Lipinski and Veber (Fig. 2b).[47,48] ≥90% of compounds meet benchmarks regarding LogP, topological polar surface area (TPSA), and numbers of hydrogen-bond donors, hydrogen-bond acceptors, and rotatable bonds. 58% of compounds meet all 6 cutoffs, and 88% meet 5 of 6 (Table S10).

In addition to analyzing DOS-DEL-1 as a whole, studying distinct sub-libraries (Supplementary Dataset 3) allowed us to determine the effects of synthetic decisions on physicochemical properties (see Fig. 2a). Two such decisions that considerably altered MW, LogP, and TPSA were 1) the choice of *N*-capping reagent class ("ald" vs. "sc"), and 2) the functionalization of the aryl iodide ("Suz" vs. "no Suz"). The sulfonyl group in all members of the "sc" sub-library contributed to a higher mean MW, lower mean LogP, and higher mean TPSA compared to the "ald" sub-library (see Fig. 2b). The shapes of the property distributions, however, differ between the two sub-libraries (Fig. 2c), which indicates that the sulfonyl group is not the lone contributor to these differences. Performing a Suzuki reaction tended to increase MW, LogP, and TPSA compared to the parent aryl iodides (see Fig. 2c), even though iodine is a heavy, lipophilic atom. Additional comparison plots (Figs. S10–S33 and Supplementary Dataset 4) show how these synthetic choices influenced additional physicochemical properties (*e.g.*, numbers of hydrogen-bond donors and acceptors) and the distributions of those properties in various DOS-DEL-1 sub-libraries.

Multi-fusion similarity analysis of DOS-DEL-1. Tanimoto-based multi-fusion similarity analysis can be used to calculate the structural relationships between sets of small molecules,[49] which

provide further insight into the consequences of synthetic decisions. First, we analyzed compound structures within DOS-DEL-1. Eight sub-libraries (see Fig. 2a) were defined according to combinations of reaction parameters; for example, the "az/sc/Suz" sub-library comprises azetidines with a sulfonyl chloride N-capping reagent and a functionalized aryl iodide. Within-group and cross-group multi-fusion similarities were calculated for each sub-library (Fig. S34). Overall, despite its use as a "diversification" step, the Suzuki reaction appears to increase similarities both within and across sub-libraries, though this may be partially explained by the larger size of the "Suz" sub-libraries.[50]

Next, each sub-library was compared to five off-DNA small-molecule collections (Supplementary Dataset 5): 2,477 natural products (NP),[36] 6,152 commercial compounds (CC),[36] 6,623 diversity-oriented synthesis (DOS1) compounds,[36] 12,011 known bioactives (BIO),[51] and another 19,149 DOS compounds (DOS2; Fig. 3).[51] The resulting average-maximum fusion similarity plots render these data in two dimensions, with each point representing the Tanimoto similarity of the nearest neighbor (vertical coordinate) and the average similarity (horizontal coordinate).

As mentioned above, we had sought to synthesize a DEL whose members more closely resembled compounds from DOS (DOS1 and DOS2) than those found in commercial collections (CC), which often exhibit high sp^2-content and low stereochemical complexity. According to mean-fusion similarity (Fig. 3a), DOS-DEL-1 compounds are most like DOS1 and DOS2, especially a subset of DOS1 in the middle of the plot. They are least similar to CC, though many members of DOS-DEL-1 find their most similar structure in CC or BIO. For comparison, mean-maximum similarity fusions between members of DOS-DEL-1 are also depicted.

We then compared each of the eight DOS-DEL-1 sub-libraries to these five reference compound collections (Fig. 3b). The highest maximum-fusion similarity was achieved by comparing CC and the "ald/Suz" sub-libraries, which is consistent with the set of reactions most commonly performed in industry.[52,53] Performing a Suzuki reaction appears to increase multi-fusion similarity to all five reference libraries broadly, though the effect on median-fusion similarity for CC and maximum-fusion similarity for NP is weak. Neither ring size nor N-capping reagent class produce a consistent effect across reference libraries, though the higher maximum-fusion similarity of CC to DOS-DEL-1 "pyr/Suz" sub-libraries highlights the dearth of azetidines in commercial collections.

Barcode validation. After DEL synthesis, next-generation DNA sequencing can identify which DNA barcodes are present and quantify their relative abundances. Deep sequencing (Table S11 and Fig. S35) revealed all but three of the 100,000+ expected barcodes; the three missing sequences were each detected in subsequent experiments (Fig. S36; Supplementary Dataset 6). When each set of tags was analyzed separately, the relative abundances of only two BB1 tags and three BB2 tags fell outside 0.5x–2.0x of their set's mean abundance (Figs. S37–S39). As described previously,[54,55] the resulting distribution of individual compound barcode frequencies is well-approximated by a negative binomial distribution, especially when barcodes containing the two most problematic BB2 tags are omitted (Fig. S40). Due to the size of DOS-DEL-1, these minor discrepancies can be accounted for during analysis of the screening data.

DEL affinity-based screens. To assess the quality of DOS-DEL-1, we performed test screening experiments with two targets often used for this purpose: horseradish peroxidase (HRP) and carbonic anhydrase IX (CA-IX).

Rather than rely upon raw sequencing counts in post-screen eluents, we sought to account for the non-uniform distribution of library barcodes and the possibility of compounds binding the immobilization matrix. To do this, we envisioned a "barcode enrichment" score for each barcode as the quotient of its post-screen abundance (the "after" distribution) and its abundance in a beads-only no-target control (the "before" distribution). Such a quotient, however, would have a

much lower level of confidence when the input representation was low than when the input representation was high. To account for this discrepancy, we instead computed ratios between confidence limits estimated from the data at fixed values of confidence (CI_{95}) using Poisson distributions modeled on the numbers of observed sequence reads. We defined "normalized fold change" as the ratio of a CI_{95} underestimate of the "after" distribution and a CI_{95} overestimate of the "before" distribution (see Supporting Information). Compared to a negative-binomial distribution (Fig. 4),[54,55] this Poisson-based method better represents both low- and high-abundance barcodes (under ~25 and over ~380 sequencing reads in Fig. 4, respectively) prior to screening.

DEL screens were first performed with HRP, which is commonly used in initial screening experiments.[19,56] We observed that many of the compounds with the highest normalized fold-change scores contain sulfonyl chloride–derived Michael acceptors (Fig. 5a). Our normalization procedure enabled analysis of the leftmost portion of the plot (those barcodes with the lowest "before" representation), which is typically challenging to interpret when enrichment scores are not normalized due to artificially high enrichment scores (Fig. S41).

We expanded this analysis to individual and paired diversity elements to determine if electrophiles were broadly enriched by exposure to HRP. To do this, we compared the representation of diversity elements among the top $k=2,389$ compounds ranked by normalized fold-change relative to that of the top $k=2,389$ compounds ranked by the input ("before") representation. We scored each diversity element or pair of elements using chigram analysis (see Supporting Information). This analysis revealed that the first-, second-, and fifth-most highly enriched BB1 DNA tags correspond to compounds containing sulfonyl chloride-derived Michael acceptors, and that their relative rank matches the order that would be predicted when considering reactivity principles (Fig. 5b). Neither *p*-ethylphenylsulfonamides (Fig. S42) nor *N*-phenylacrylamides (Fig. S43) were enriched, which suggest that an electrophilic motif is necessary but not sufficient for binding. Analyzing pairs of diversity elements revealed little influence of BB2 upon enrichment but more substantial effects due to scaffold identity (Fig. 5c). Finally, as an internal control, DNA tags encoding the same building block exhibited broadly similar behavior (Fig. S44).

We then sought ways to study the chemistry at BB2. DOS-DEL-1 was designed to contain 1,824 compounds with an aryl primary sulfonamide motif at BB2. This sub-structure is known to promote binding to CA-IX,[19,57] so enrichment of primary sulfonamides upon exposure to CA-IX would suggest that the on-DNA Suzuki reactions had been successful.

The normalized fold-change plot revealed that many of the most highly enriched compounds contain an aryl primary sulfonamide (Fig. 5d). Most of these compounds contain *para* aryl sulfonamides and fewer contain *meta* aryl sulfonamides. This ordering is consistent with the increased accessibility of the *para* sulfonamide over the *meta* regioisomer to the enzyme active site. In this case, Poisson normalization revealed additional primary sulfonamides that would have otherwise been obscured (see Supplementary Dataset 6).

Chigram analysis revealed that primary sulfonamide–containing compounds were broadly enriched by exposure to CA-IX. The first- and second-most highly enriched BB2 DNA tags among the top $k=1,120$ compounds correspond to *para* and *meta* aryl primary sulfonamides, respectively (Fig. 5e), which reaffirms the normalized fold-change analysis. Expanding this analysis to pairs of diversity elements suggested generally little impact of scaffold and BB1 identity on the enrichment of sulfonamide-containing compounds, though there does appear to be modest complementarity between *para* primary sulfonamides and *trans* azetidine scaffolds (Fig. 5f). These results are consistent with previous observations that the primary sulfonamide motif dominates the binding interaction with CA-IX.[43] DNA tags encoding the same building block again behaved similarly (Fig. S45). Taken together, both sets of screening results confirm the robustness of the on-DNA

7

chemistry during library synthesis and the ability of our analysis method to uncover relationships between structures of highly enriched compounds.

CA-IX hit synthesis and evaluation. To validate compound–target interactions, positives from the CA-IX screens (and control compounds) were synthesized off DNA (Fig. 6a; Fig. S46) and subjected to a carbonic anhydrase inhibition assay (Fig. 6b; Figs. S47–S52). Every compound with a sulfonamide at the *para* position (**5**–**8**) exhibited potent inhibition (IC_{50} = 20.1–68.7 nM)— even Boc carbamates **8a**/**8b**. Therefore, the identity of BB1 largely does not influence protein binding, which was suggested by the screening data (see Fig. 5f). Moving the sulfonamide to the *meta* position (**8a/8b** vs. **9a/9b**) reduced inhibitory activity by ~15- to ~50-fold, which is consistent with the chigram analysis (see Figs. 5e and 5f). Stereochemistry did not influence IC_{50} in most cases, but pyrrole **7a** and carbamate **9b** were found to be ~3-fold and ~40-fold more potent than their enantiomers, respectively. No inhibitory activity was detected with biphenyl azetidines **10a** or **10b**, which lack primary sulfonamides.

Discussion

Phenotypic screening has benefitted tremendously from the inclusion of novel chemical structures, such as those that result from either biosynthetic pathways yielding natural products or DOS yielding synthetic compounds.[39] The application of the DOS approach to DEL-based screening, however, had not been explored before this study. We acknowledge that these initial findings have not yet illuminated the full consequences of incorporating under-represented chemical features into DELs, but the synthesis and characterization of DOS-DEL-1 are helpful first steps towards answering this important question in the future.

While compounds that contain common structural motifs like peptide bonds have had success in DEL screens,[9,58] the full potential of DELs likely remains untapped. We have shown that modern organic synthesis can furnish DEL scaffolds that yield products containing rare structural features. Instead of installing structural complexity on DNA, we performed our key complexity-generating transformations before DNA attachment. This approach has enabled for the first time, to our knowledge, the DNA barcoding of all possible stereoisomers of complex scaffolds containing multiple stereogenic elements. This design choice is supported by recent studies that describe the significant influence of scaffold geometry on the behavior of DNA-barcoded compounds.[43] We hope to leverage this insight towards the discovery of small-molecule binders of challenging therapeutic targets in the future—increasingly sophisticated DELs and screening efforts with that aim are ongoing.

DEL technology can generate exceedingly large libraries, but we consciously sacrificed library size to maximize the robustness of our synthesis and screening procedures. A recent analysis has questioned the relationship between DEL productivity and library size,[59] and results from the Liu laboratory have confirmed that libraries need not exceed tens or hundreds of thousands of members to find potent nMoA compounds.[60,61] Larger libraries are correlated with weaker signals in DEL screens,[21] and this effect appears to outweigh the benefits of increased appendage diversity in libraries comprising >10^8 members.[44] Furthermore, data from smaller libraries are more easily normalized because fewer sequencing reads are required to account for unequal barcode distributions, both at specific tagging steps and in the resulting distribution of full barcodes. This challenge is similar to well-studied sequencing biases in the context of differential gene-expression analysis.[62]

The use of split–pool synthesis places strict demands on DEL chemistry if both individual hits and broader trends are to be extracted from screening data reliably. As a result, we developed an

improved on-DNA Suzuki reaction that uses Pd(dppf)Cl$_2$ to minimize hydrodehalogenation by-products. We also tailored the on-DNA pathway to account for differences in reactivity between the azetidine and pyrrolidine scaffolds, and we set a high bar for building-block validation (≥85%). Our screens using HRP and CA-IX confirmed the robustness of our approach to on-DNA chemistry, and the choice of test proteins allowed us to assess the quality of the library in a manner that was agnostic to variable scaffold size and geometry. Phenotypic screening experiments have already demonstrated that the activity of 2,3-disubstituted azetidines can be affected by changes in stereochemistry and appendage identity,[63] so we are confident that we will observe these trends in future DEL screens.

We developed a framework for analyzing DEL screening data that both normalizes barcode enrichment values and identifies SAR in a systematic, unbiased manner. Previous uses of Poisson or negative binomial distribution statistics to model screening data assumed an equal distribution of library members,[54,55,64] but our method accounts for experimentally determined low-abundance tags and compounds when calculating barcode enrichment. It also enables the automation of unbiased SAR discovery as expressed by preferences in screening experiments for either individual diversity elements or pairs of diversity elements. These computational advances draw upon our experience in generating SAR rules for non-encoded DOS libraries,[65] and extend them using a Poisson-based barcode enrichment model suitable for a non-uniform input distribution including (possibly) rare tags or barcodes. This strategy should be broadly applicable to all DEL screening data in which the baseline distribution of barcodes (or a no-target control experiment) has also been measured.

DOS-DEL-1 also shows that a modestly sized DEL can include drug-like compounds with a wide range of physicochemical properties. Like other types of libraries built using combinatorial techniques,[22] DELs often contain compounds of high molecular weight and lipophilicity.[10] The vast majority of compounds in DOS-DEL-1, however, comply with traditional measures of drug-likeness. But because many orally available drugs violate these guidelines,[66] non-compliant compounds were also included.

Multi-fusion similarity analysis revealed that DOS-DEL-1 members are more like compounds derived from DOS than those from commercial vendors, despite the use of commonly used DNA-compatible transformations. Analyzing individual sub-libraries allowed us to study how individual synthetic decisions affected both physicochemical properties and similarity to off-DNA compound collections. These results demonstrate that an emphasis on scaffold complexity—installed with modern off-DNA synthesis—can imbue DELs with structural diversity. Combining these analyses with DEL screening data may facilitate improved understanding of the consequences of small-molecule properties and synthetic choices on protein binding.

Conclusion

DEL technology is attracting attention from academia and industry alike,[67] but substantial advances must be made for the reality in the laboratory to fulfill its promise. All too often, DEL screening with challenging targets fail to yield effective binders.[68,69] Despite recent breakthroughs in screening, barcoding, and data analysis, the search for small-molecule binders remains fundamentally limited by the types of chemical structures that can be attached to DNA. On-DNA methodologies and synthetic planning have lagged behind their increasingly sophisticated off-DNA counterparts, but the results described herein constitute one strategy for expanding the types of chemical structures found in DELs.

Supporting Information. The Supporting Information is available free of charge on the ACS Publications website.

- Detailed methods for all chemical, biological, and computational procedures, full characterization data (including NMR spectra) for all synthetic intermediates and final compounds, and additional details regarding reagents and instruments used.
- Supplementary Datasets 1-6 containing raw data regarding building block validation, DNA barcoding, physicochemical property calculations, DOS-DEL-1 and comparator library compound structures, and affinity-based screens with HRP and CA-IX.

Data availability. The data supporting the findings of this study are available within this paper's main text, Supporting Information, and Supplementary Datasets.

Code availability. Code used to perform cheminformatic analyses and analyze sequencing results, including figure preparation, are available *via* GitHub (https://github.com/remontoire-pac/ngc-dosdel/).
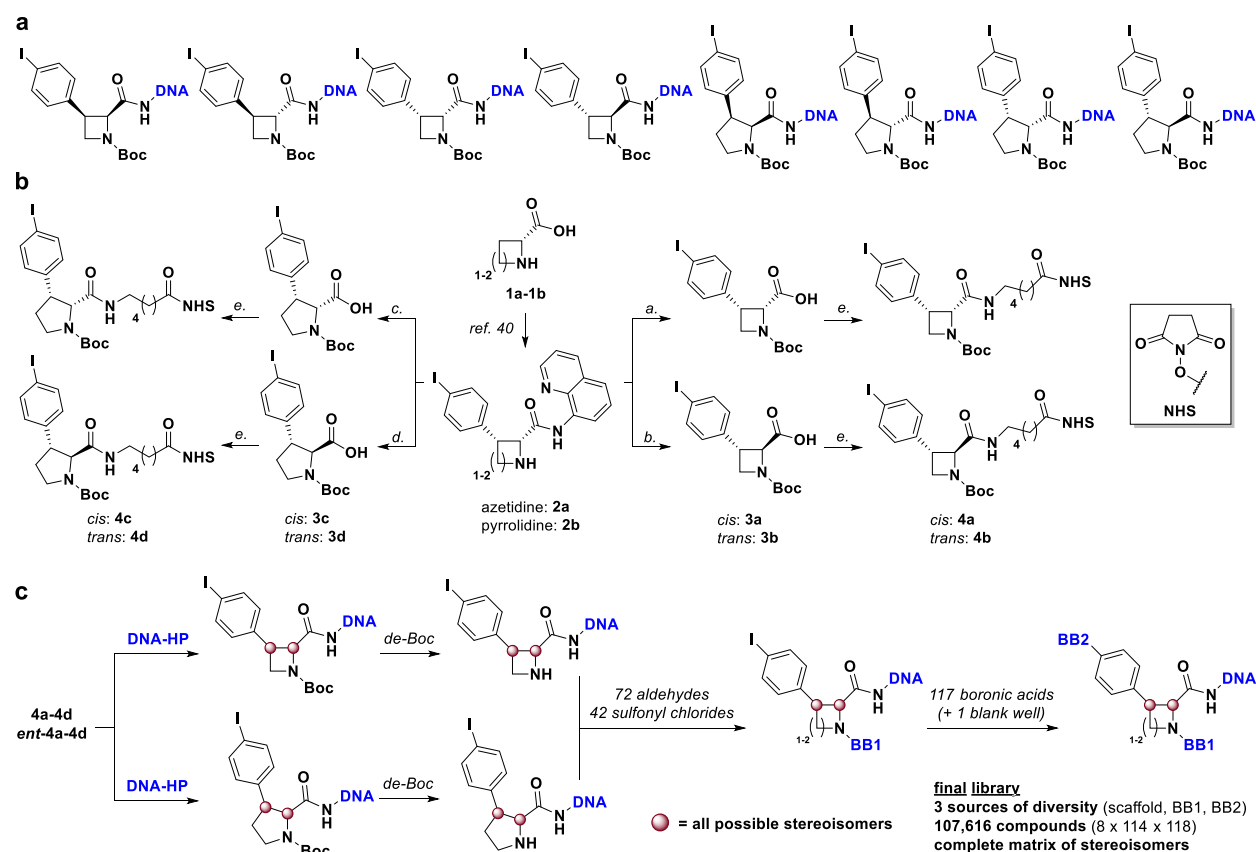
Figures

a

b

c

Fig. 1 | Design and synthesis of DOS-DEL-1. a, The eight azetidine- and pyrrolidine-based scaffolds that constitute the basis of DOS-DEL-1 can be accessed from commercially available chiral α-amino acids. **b**, Synthesis of NHS esters **4a-4d** for scaffold attachment to DNA. Azetidine **2a** and pyrrolidine **2b** were synthesized from D-azetidine-2-carboxylic acid (**1a**) and D-proline (**1b**), respectively according to the procedures described by Maetani et al.[40] Reagents and conditions can be found in the Supporting Information. **c**, Synthesis of DOS-DEL-1. After scaffold attachment (*via* NHS esters **4a-4d** and ***ent*-4a-4d**) to a double-stranded DNA "headpiece" (DNA-HP; Fig. S1), the azetidines and pyrrolidines undergo Boc deprotection separately. Then, all eight scaffolds are pooled for two rounds of DNA-encoded split–pool synthesis: 114 *N*-capping reactions and 118 Suzuki reactions. The resulting library has three sources of diversity, comprises 107,616 unique compounds, and includes all possible stereoisomers of the azetidine and pyrrolidine scaffolds (as indicated by the red balls).
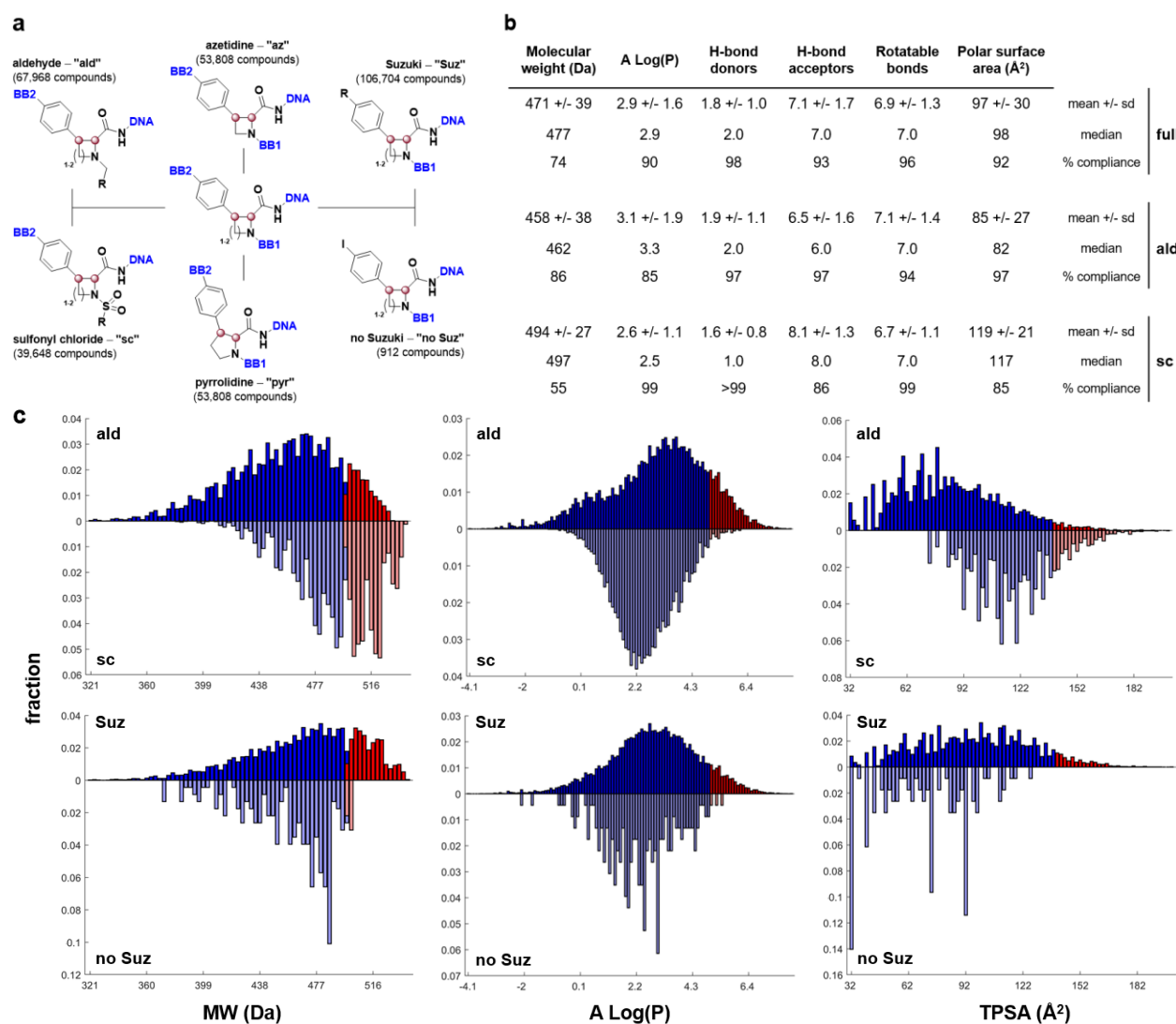
Fig. 2 | DOS-DEL-1 synthetic decisions influence small-molecule physicochemical properties. a, DOS-DEL-1 sub-libraries corresponding to divergent synthetic decisions. Subsets of DOS-DEL-1 may be defined according to the size of the heterocyclic scaffold (azetidine="az" or pyrrolidine="pyr"), the identity of the *N*-capping reagent (aldehyde="ald" or sulfonyl chloride="sc"), or the functionalization of the aryl iodide (Suzuki="Suz" or no Suzuki="no Suz"). Red balls indicate both possible configurations at that position. **b,** Distribution statistics for physicochemical properties corresponding to Lipinski's "rule of 5" (ref. 47) or Veber's bioavailability guidelines (ref. 48). Values depict mean ± SD, median, or % compliance with "rules" across all members of the library or indicated sub-library. **c,** Physicochemical property distributions of DOS-DEL-1 sub-libraries. Histogram bins are colored according to adherence to (blue) or violation of (red) Lipinski's (MW < 500 Da; LogP ≤ 5) or Veber's (TPSA < 140 Å^2) guidelines. Bin height or depth represents the fraction of the sub-library's compounds that fall within the bin's boundaries. Sub-libraries that represent divergent synthetic decisions are juxtaposed to highlight the consequences of those decisions on physicochemical properties.
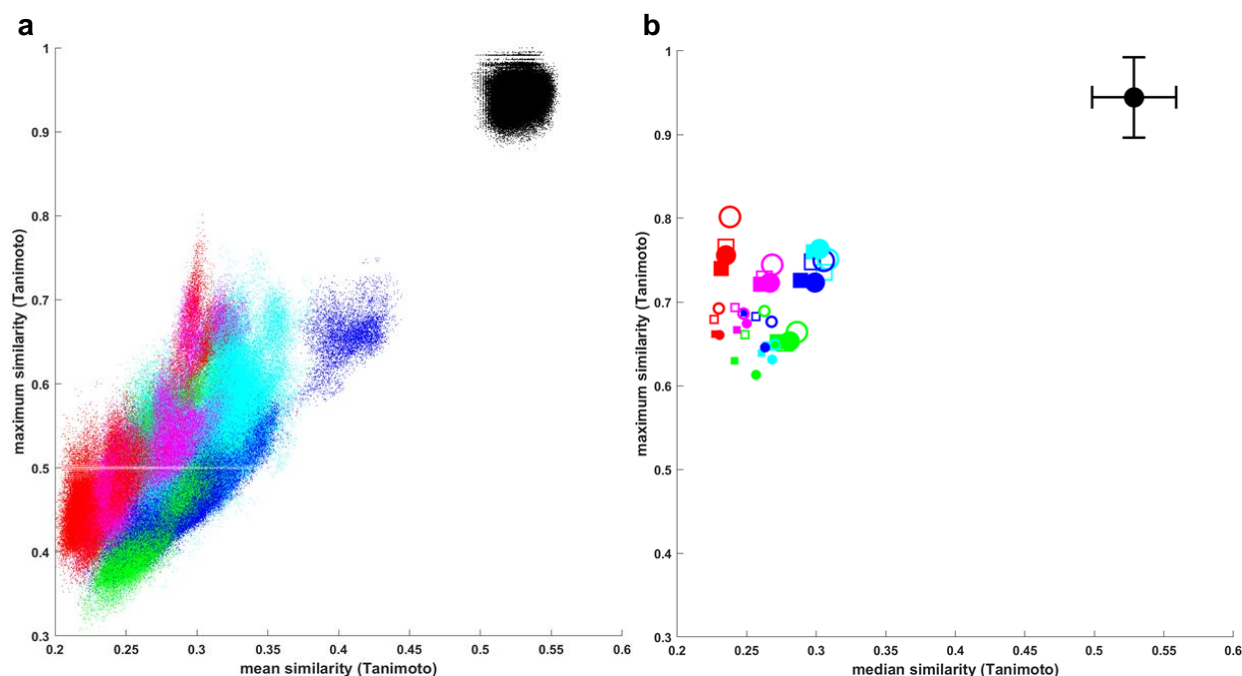
Fig. 3 | DOS-DEL-1 members more closely resemble compounds from diversity-oriented synthesis than those from other sources. Tanimoto-based multi-fusion (average vs. maximum) similarity representations. **a,** Comparison of DOS-DEL-1 to itself (black) and to each of 5 published small-molecule collections: 6,152 commercial compounds (red), 2,477 natural products (green), and 6,623 diversity-oriented synthesis compounds (blue) from ref. 36.; plus 12,011 known actives and PubChem screening hits (magenta) and 19,149 diversity-oriented synthesis compounds (cyan) from ref. 51. **b,** Comparison of 8 library subsets defined by azetidine (square) or pyrrolidine (circle) skeleton, aldehyde (open) or sulfonyl chloride (filled) *N*-capping, and functionalization (large) or not (small) of the aryl iodide using a Suzuki reaction, to the same published collections. For reference, the intra-library multi-fusion similarities among all DOS-DEL-1 members are represented as a mean (black circle) ±3 SD (black bars) of multi-fusion (median vs. maximum) similarities.
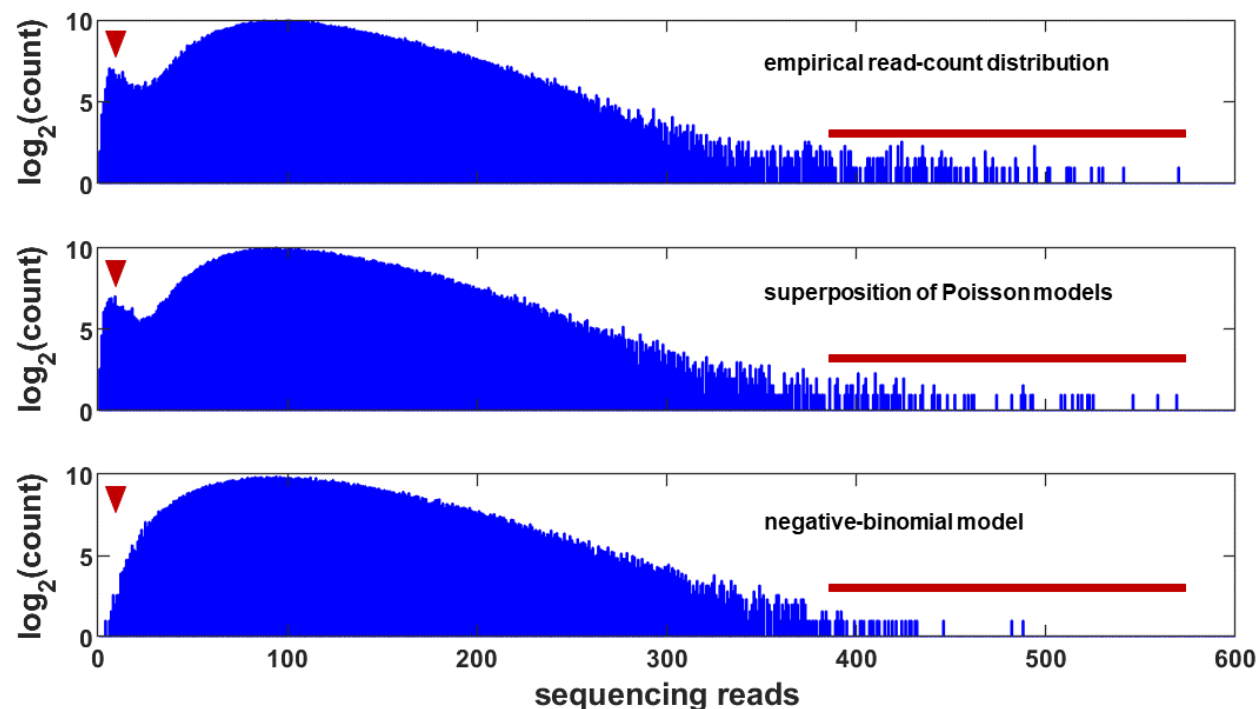
Fig. 4 | The distribution of DOS-DEL-1 barcodes is more accurately modeled by a Poisson-derived frequency distribution than by a negative-binomial distribution. The empirical read-count distribution (top) for DOS-DEL-1 is compared to a frequency distribution created by sampling from a set of Poisson-distribution models representing each tag, fit using all compounds containing that tag (middle). The corresponding negative-binomial distribution (bottom) is shown for comparison, which fails to account for tags that were inefficiently incorporated (red arrows) and appears to understate the most efficient tagging (red bars).
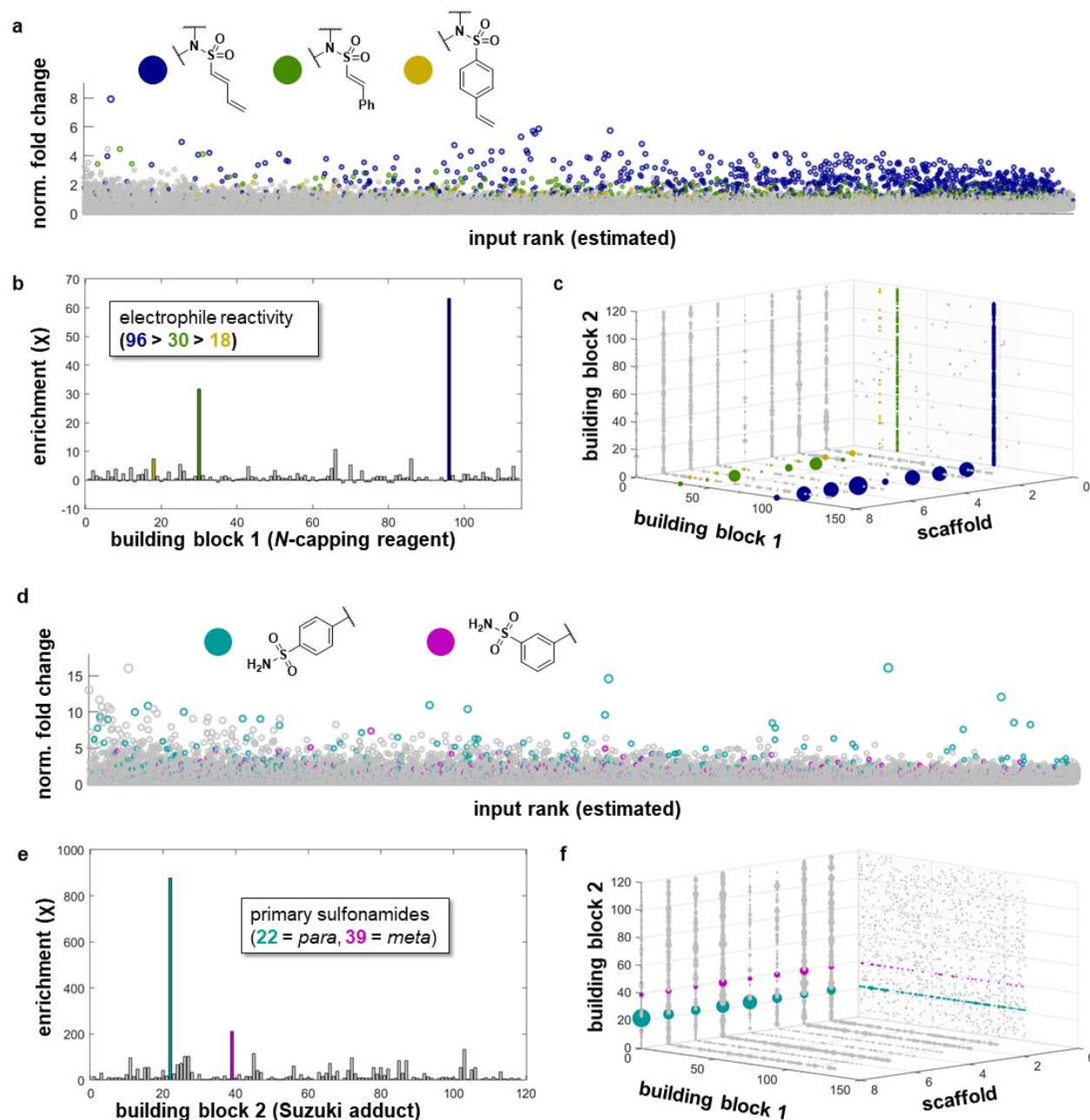
Fig. 5 | DOS-DEL-1 screens using test proteins reveal preferential binding as a function of building blocks and building-block combinations. a, Normalized fold-change plot for DOS-DEL-1 members binding horseradish peroxidase as a function of input representation rank, and depicting compounds containing select significantly enriched electrophilic building blocks (blue: strong; green: intermediate; gold: weak). **b,** Enrichment plot for each of 114 *N*-capping building blocks (building block 1), expressed as standardized residuals (χ) relative to expected occurrence among compounds within the top 2,389 normalized fold-change scores. **c,** Enrichment scores, expressed as standardized residuals (χ) by size, for co-occurrence of pairs of building blocks relative to expected co-occurrence among compounds within the top 2,389 normalized fold-change scores. **d,** Normalized fold-change plot for DOS-DEL-1 members binding carbonic anhydrase IX as a function of input representation rank, and depicting compounds containing select significantly enriched individual building blocks (teal: primary *para* sulfonamide; magenta:

primary *meta* sulfonamide). **e,** Enrichment plot for each of 119 Suzuki adduct building blocks (building block 2), expressed as standardized residuals (χ) relative to expected occurrence among compounds within the top 1,120 normalized fold-change scores. **f,** Enrichment scores, expressed as standardized residuals (χ) by size, for co-occurrence of pairs of building blocks relative to expected co-occurrence among compounds within the top 1,120 normalized fold-change scores. In panels 5c and 5f, all three combinations of pairs of building blocks (8 x 114, 8 x 119, 114 x 119) are presented on the "faces" of a single 3D plot for compactness.

a

(R,R): **5a**
(S,S): **5b**

(R,R): **6a**
(S,S): **6b**

(R,R): **7a**
(S,S): **7b**

(R,R): **8a**
(S,S): **8b**

(R,R): **9a**
(S,S): **9b**

(R,R): **10a**
(S,S): **10b**

b

	carbonic anhydrase IC$_{50}$ (95% CI)
5a	24.3 nM (21.5–27.5 nM)
5b	25.6 nM (20.2–33.0 nM)
6a	23.7 nM (17.5–32.7 nM)
6b	20.1 nM (15.5–27.9 nM)
7a	20.8 nM (17.4–24.9 nM)
7b	68.7 nM (52.4–87.2 nM)
8a	24.7 nM (21.7–28.3 nM)
8b	21.4 nM (18.7–24.6 nM)
9a	12.1 µM (7.71–38.6 µM)
9b	297 nM (241–367 nM)
10a	>100 µM
10b	>100 µM
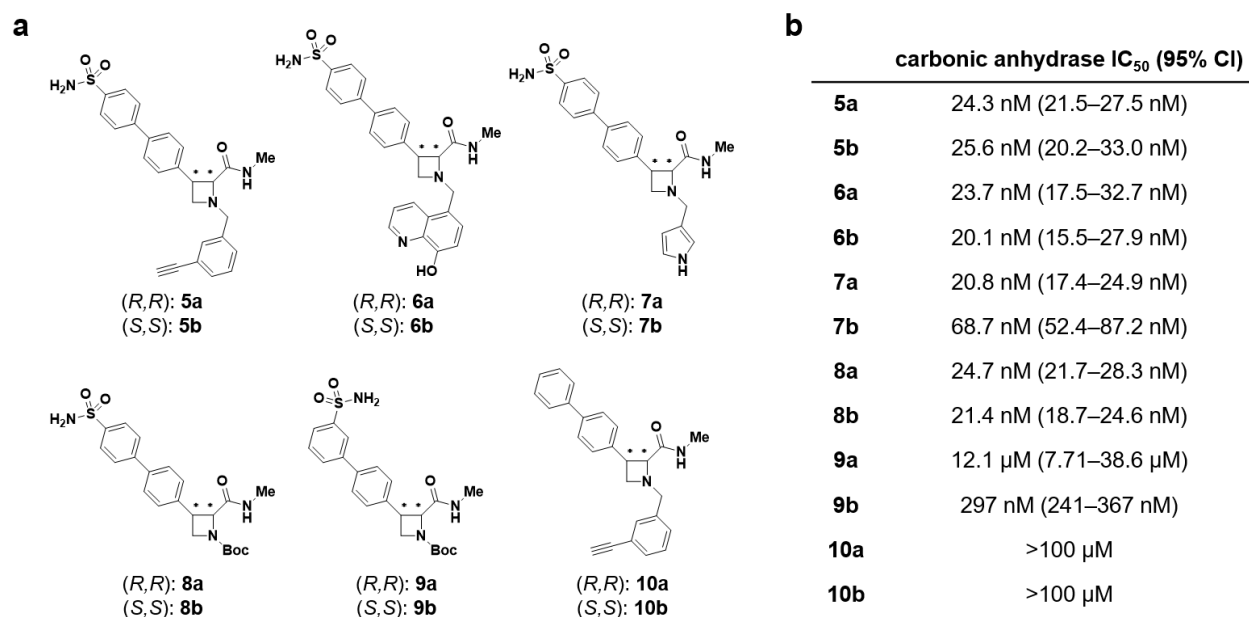
Fig. 6 | Biochemical assay data are consistent with CA-IX screening results. a, Chemical structures of off-DNA compounds synthesized to assess screening results. Azetidines **5a**, **5b**, **6a**, **6b**, **7a**, and **7b** correspond to on-DNA assay positives; the remaining compounds facilitated deeper validation of the screening data and the success of the on-DNA chemistry. Asterisks indicate atoms of variable stereochemical configuration (as defined below each structure). Synthetic details are provided in the Supporting Information (see Fig. S46; Supporting Information: "Synthesis of off-DNA CA-IX hits"). **b**, In vitro carbonic anhydrase inhibition assay results. Compounds depicted in Fig. 6a were subjected to a colorimetric carbonic anhydrase inhibition assay. IC$_{50}$ values and 95% confidence intervals are shown. Statistics were calculated using GraphPad Prism software.

References

(1)   Garbaccio, R. M.; Parmee, E. R. The Impact of Chemical Probes in Drug Discovery: A

       Pharmaceutical Industry Perspective. *Cell Chem. Biol.* **2016**, *23* (1), 10–17.

(2)   Brenner, S.; Lerner, R. A. Encoded Combinatorial Chemistry. *Proc. Natl. Acad. Sci.* **1992**,

       89 (12), 5381–5383.

(3)   Franzini, R. M.; Neri, D.; Scheuermann, J. DNA-Encoded Chemical Libraries: Advancing

       beyond Conventional Small-Molecule Libraries. *Acc. Chem. Res.* **2014**, *47* (4), 1247–

       1255.

(4)   Yang, H.; Medeiros, P. F.; Raha, K.; Elkins, P.; Lind, K. E.; Lehr, R.; Adams, N. D.;

       Burgess, J. L.; Schmidt, S. J.; Knight, S. D.; Auger, K. R.; Schaber, M. D.; Franklin, G. J.;

       Ding, Y.; DeLorey, J. L.; Centrella, P. A.; Mataruse, S.; Skinner, S. R.; Clark, M. A.;

       Cuozzo, J. W.; Evindar, G. Discovery of a Potent Class of PI3Kα Inhibitors with Unique

       Binding Mode via Encoded Library Technology (ELT). *ACS Med. Chem. Lett.* **2015**, *6* (5),

       531–536.

(5)   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.

(6)   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. (Sargent); 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* (16), 7061–7079.

(7)   Litovchick, A.; Dumelin, C. E.; Habeshian, S.; Gikunju, D.; Guié, M.-A.; Centrella, P.; Zhang, Y.; Sigel, E. A.; Cuozzo, J. W.; Keefe, A. D.; Clark, M. A. Encoded Library Synthesis Using Chemical Ligation and the Discovery of SEH Inhibitors from a 334-Million Member Library. *Sci. Rep.* **2015**, *5* (1), 10916.

(8)   Kollmann, C. S.; Bai, X.; Tsai, C.-H.; Yang, H.; Lind, K. E.; Skinner, S. R.; Zhu, Z.; Israel, D. I.; Cuozzo, J. W.; Morgan, B. A.; Yuki, K.; Xie, C.; Springer, T. A.; Shimaoka, M.; Evindar, G. Application of Encoded Library Technology (ELT) to a Protein–Protein Interaction Target: Discovery of a Potent Class of Integrin Lymphocyte Function-Associated Antigen 1 (LFA-1) Antagonists. *Bioorg. Med. Chem.* **2014**, *22* (7), 2353–2365.

(9)   Salamon, H.; Klika Škopić, M.; Jung, K.; Bugain, O.; Brunschweiger, A. Chemical Biology Probes from Advanced DNA-Encoded Libraries. *ACS Chem. Biol.* **2016**, *11* (2), 296–307.

(10)   Franzini, R. M.; Randolph, C. Chemical Space of DNA-Encoded Libraries. *J. Med. Chem.* **2016**, *59* (14), 6629–6644.

(11)   Schreiber, S. A Chemical Biology View of Bioactive Small Molecules and a Binder-Based Approach to Connect Biology to Precision Medicines. *bioRxiv* **2018**, DOI: 10.1101/386904.

(12)   Schreiber, S. L. A Chemical Biology View of Bioactive Small Molecules and a Binder-Based Approach to Connect Biology to Precision Medicines. *Isr. J. Chem.* **2019**, *59* (1–2), 52–59.

(13)   Schreiber, S. L.; Crabtree, G. R. The Mechanism of Action of Cyclosporin A and FK506.

Immunol. Today **1992**, *13* (4), 136–142.

(14)   de Waal, L.; Lewis, T. A.; Rees, M. G.; Tsherniak, A.; Wu, X.; Choi, P. S.; Gechijian, L.;

Hartigan, C.; Faloon, P. W.; Hickey, M. J.; Tolliday, N.; Carr, S. A.; Clemons, P. A.;

Munoz, B.; Wagner, B. K.; Shamji, A. F.; Koehler, A. N.; Schenone, M.; Burgin, A. B.;

Schreiber, S. L.; Greulich, H.; Meyerson, M. Identification of Cancer-Cytotoxic Modulators

of PDE3A by Predictive Chemogenomics. *Nat. Chem. Biol.* **2016**, *12* (2), 102–108.

(15)   Jones, L. H. Small-Molecule Kinase Downregulators. *Cell Chem. Biol.* **2018**, *25* (1), 30–

35.

(16)   Spencer, D.; Wandless, T.; Schreiber, S.; Crabtree, G. Controlling Signal Transduction

with Synthetic Ligands. *Science (80-.).* **1993**, *262* (5136), 1019–1024.

(17)   Stanton, B. Z.; Chory, E. J.; Crabtree, G. R. Chemically Induced Proximity in Biology and

Medicine. *Science (80-.).* **2018**, *359* (6380), eaao5902.

(18)   Toure, M.; Crews, C. M. Small-Molecule PROTACS: New Approaches to Protein

Degradation. *Angew. Chemie Int. Ed.* **2016**, *55* (6), 1966–1973.

(19)   Li, Y.; De Luca, R.; Cazzamalli, S.; Pretto, F.; Bajic, D.; Scheuermann, J.; Neri, D.

Versatile Protein Recognition by the Encoded Display of Multiple Chemical Elements on a

Constant Macrocyclic Scaffold. *Nat. Chem.* **2018**, *10* (4), 441–448.

(20)   Zhu, Z.; Shaginian, A.; Grady, L. C.; O'Keeffe, T.; Shi, X. E.; Davie, C. P.; Simpson, G. L.;

Messer, J. A.; Evindar, G.; Bream, R. N.; Thansandote, P. P.; Prentice, N. R.; Mason, A.

M.; Pal, S. Design and Application of a DNA-Encoded Macrocyclic Peptide Library. *ACS*

Chem. Biol. **2018**, *13* (1), 53–59.

(21)   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. Design, Synthesis and Selection of DNA-Encoded Small-Molecule Libraries. *Nat. Chem. Biol.* **2009**, *5* (9), 647–654.

(22)   Feher, M.; Schmidt, J. M. Property Distributions: Differences between Drugs, Natural Products, and Molecules from Combinatorial Chemistry. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (1), 218–227.

(23)   Lu, X.; Fan, L.; Phelps, C. B.; Davie, C. P.; Donahue, C. P. Ruthenium Promoted On-DNA Ring-Closing Metathesis and Cross-Metathesis. *Bioconjug. Chem.* **2017**, *28* (6), 1625–1629.

(24)   Ruff, Y.; Berst, F. Efficient Copper-Catalyzed Amination of DNA-Conjugated Aryl Iodides under Mild Aqueous Conditions. *Medchemcomm* **2018**, *9* (7), 1188–1193.

(25)   Du, H.-C.; Huang, H. DNA-Compatible Nitro Reduction and Synthesis of Benzimidazoles. *Bioconjug. Chem.* **2017**, *28* (10), 2575–2580.

(26)   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.* **2018**, *115* (28), E6404–E6410.

(27)   Kölmel, D. K.; Loach, R. P.; Knauber, T.; Flanagan, M. E. Employing Photoredox Catalysis for DNA-Encoded Chemistry: Decarboxylative Alkylation of α-Amino Acids. *ChemMedChem* **2018**, *13* (20), 2159–2165.

(28) Wang, X.; Sun, H.; Liu, J.; Dai, D.; Zhang, M.; Zhou, H.; Zhong, W.; Lu, X. Ruthenium-Promoted C–H Activation Reactions between DNA-Conjugated Acrylamide and Aromatic Acids. *Org. Lett.* **2018**, *20* (16), 4764–4768.

(29) Gerry, C. J.; Yang, Z.; Stasi, M.; Schreiber, S. L. DNA-Compatible [3 + 2] Nitrone–Olefin Cycloaddition Suitable for DEL Syntheses. *Org. Lett.* **2019**, *21* (5), 1325–1330.

(30) Phelan, J. P.; Lang, S. B.; Sim, J.; Berritt, S.; Peat, A. J.; Billings, K.; Fan, L.; Molander, G. A. Open-Air Alkylation Reactions in Photoredox-Catalyzed DNA-Encoded Library Synthesis. *J. Am. Chem. Soc.* **2019**, *141* (8), 3723–3732.

(31) Blakemore, D. C.; Castro, L.; Churcher, I.; Rees, D. C.; Thomas, A. W.; Wilson, D. M.; Wood, A. Organic Synthesis Provides Opportunities to Transform Drug Discovery. *Nat. Chem.* **2018**, *10* (4), 383–394.

(32) Flood, D. T.; Asai, S.; Zhang, X.; Wang, J.; Yoon, L.; Adams, Z. C.; Dillingham, B. C.; Sanchez, B.; Vantourout, J. C.; Flanagan, M. E.; Piotrowski, D. W.; Richardson, P.; Green, S.; Shenvi, R.; Chen, J.; Baran, P.; Dawson, P. Expanding Reactivity in DNA-Encoded Library Synthesis via Reversible Binding of DNA to an Inert Quaternary Ammonium Support. *ChemRxiv* **2019**.

(33) Schreiber, S. L. Target-Oriented and Diversity-Oriented Organic Synthesis in Drug Discovery. *Science* **2000**, *287* (5460), 1964–1969.

(34) Burke, M. D.; Schreiber, S. L. A Planning Strategy for Diversity-Oriented Synthesis. *Angew. Chemie - Int. Ed.* **2004**, *43* (1), 46–58.

(35) Nielsen, T. E.; Schreiber, S. L. Towards the Optimal Screening Collection: A Synthesis Strategy. *Angew. Chemie - Int. Ed.* **2008**, *47* (1), 48–56.

(36) Clemons, P. A.; Bodycombe, N. E.; Carrinski, H. A.; Wilson, J. A.; Shamji, A. F.; Wagner,

B. K.; Koehler, A. N.; Schreiber, S. L. Small Molecules of Different Origins Have Distinct Distributions of Structural Complexity That Correlate with Protein-Binding Profiles. *Proc. Natl. Acad. Sci.* **2010**, *107* (44), 18787–18792.

(37)   Lovering, F.; Bikker, J.; Humblet, C. Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success. *J. Med. Chem.* **2009**, *52* (21), 6752–6756.

(38)   Lovering, F. Escape from Flatland 2: Complexity and Promiscuity. *Med. Chem. Commun.* **2013**, *4* (3), 515.

(39)   Gerry, C. J.; Schreiber, S. L. Chemical Probes and Drug Leads from Advances in Synthetic Planning and Methodology. *Nat. Rev. Drug Discov.* **2018**, *17* (5), 333–352.

(40)   Maetani, M.; Zoller, J.; Melillo, B.; Verho, O.; Kato, N.; Pu, J.; Comer, E.; Schreiber, S. L. Synthesis of a Bicyclic Azetidine with In Vivo Antimalarial Activity Enabled by Stereospecific, Directed C(Sp 3)–H Arylation. *J. Am. Chem. Soc.* **2017**, *139* (32), 11300–11306.

(41)   Burke, M. D.; Berger, E. M.; Schreiber, S. L. Generating Diverse Skeletons of Small Molecules Combinatorially. *Science (80-.).* **2003**, *302* (5645), 613–618.

(42)   Kato, N.; Comer, E.; Sakata-Kato, T.; Sharma, A.; Sharma, M.; Maetani, M.; Bastien, J.; Brancucci, N. M.; Bittker, J. A.; Corey, V.; Clarke, D.; Derbyshire, E. R.; Dornan, G. L.; Duffy, S.; Eckley, S.; Itoe, M. A.; Koolen, K. M. J.; Lewis, T. A.; Lui, P. S.; Lukens, A. K.; Lund, E.; March, S.; Meibalan, E.; Meier, B. C.; McPhail, J. A.; Mitasev, B.; Moss, E. L.; Sayes, M.; Van Gessel, Y.; Wawer, M. J.; Yoshinaga, T.; Zeeman, A.-M.; Avery, V. M.; Bhatia, S. N.; Burke, J. E.; Catteruccia, F.; Clardy, J. C.; Clemons, P. A.; Dechering, K. J.; Duvall, J. R.; Foley, M. A.; Gusovsky, F.; Kocken, C. H. M.; Marti, M.; Morningstar, M. L.; Munoz, B.; Neafsey, D. E.; Sharma, A.; Winzeler, E. A.; Wirth, D. F.; Scherer, C. A.; Schreiber, S. L. Diversity-Oriented Synthesis Yields Novel Multistage Antimalarial

23

Inhibitors. *Nature* **2016**, *538* (7625), 344–349.

(43) Favalli, N.; Biendl, S.; Hartmann, M.; Piazzi, J.; Sladojevich, F.; Gräslund, S.; Brown, P. J.; Näreoja, K.; Schüler, H.; Scheuermann, J.; Franzini, R.; Neri, D. A DNA-Encoded Library of Chemical Compounds Based on Common Scaffolding Structures Reveals the Impact of Ligand Geometry on Protein Recognition. *ChemMedChem* **2018**, *13* (13), 1303–1307.

(44) 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* (4), 234–238.

(45) Ding, Y.; Clark, M. A. Robust Suzuki–Miyaura Cross-Coupling on DNA-Linked Substrates. *ACS Comb. Sci.* **2015**, *17* (1), 1–4.

(46) Satz, A. L. Simulated Screens of DNA Encoded Libraries: The Potential Influence of Chemical Synthesis Fidelity on Interpretation of Structure–Activity Relationships. *ACS Comb. Sci.* **2016**, *18* (7), 415–424.

(47) 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 Deliv. Rev.* **1997**, *23* (1–3), 3–25.

(48) Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.* **2002**, *45* (12), 2615–2623.

(49) Medina-Franco, J. L.; Maggiora, G. M.; Giulianotti, M. A.; Pinilla, C.; Houghten, R. A. A Similarity-Based Data-Fusion Approach to the Visual Characterization and Comparison of Compound Databases. *Chem. Biol. Drug Des.* **2007**, *70* (5), 393–412.

(50) Swamidass, S. J.; Baldi, P. Mathematical Correction for Fingerprint Similarity Measures to Improve Chemical Retrieval. *J. Chem. Inf. Model.* **2007**, *47* (3), 952–964.

(51) Wawer, M. J.; Li, K.; Gustafsdottir, S. M.; Ljosa, V.; Bodycombe, N. E.; Marton, M. A.; Sokolnicki, K. L.; Bray, M.-A.; Kemp, M. M.; Winchester, E.; Taylor, B.; Grant, G. B.; Hon, C. S.-Y.; Duvall, J. R.; Wilson, J. A.; Bittker, J. A.; Dančík, V.; Narayan, R.; Subramanian, A.; Winckler, W.; Golub, T. R.; Carpenter, A. E.; Shamji, A. F.; Schreiber, S. L.; Clemons, P. A. Toward Performance-Diverse Small-Molecule Libraries for Cell-Based Phenotypic Screening Using Multiplexed High-Dimensional Profiling. *Proc. Natl. Acad. Sci.* **2014**, *111* (30), 10911–10916.

(52) Brown, D. G.; Boström, J. Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone? *J. Med. Chem.* **2016**, *59* (10), 4443–4458.

(53) Boström, J.; Brown, D. G.; Young, R. J.; Keserü, G. M. Expanding the Medicinal Chemistry Synthetic Toolbox. *Nat. Rev. Drug Discov.* **2018**, *17* (10), 709–727.

(54) Buller, F.; Zhang, Y.; Scheuermann, J.; Schäfer, J.; Bühlmann, P.; Neri, D. Discovery of TNF Inhibitors from a DNA-Encoded Chemical Library Based on Diels-Alder Cycloaddition. *Chem. Biol.* **2009**, *16* (10), 1075–1086.

(55) Faver, J. C.; Riehle, K.; Lancia, D. R.; Milbank, J. B. J.; Kollmann, C. S.; Simmons, N.; Yu, Z.; Matzuk, M. M. Quantitative Comparison of Enrichment from DNA-Encoded Chemical Library Selections. *ACS Comb. Sci.* **2019**, *21* (2), 75–82.

(56) Zimmermann, G.; Rieder, U.; Bajic, D.; Vanetti, S.; Chaikuad, A.; Knapp, S.; Scheuermann, J.; Mattarella, M.; Neri, D. A Specific and Covalent JNK-1 Ligand Selected from an Encoded Self-Assembling Chemical Library. *Chem. - A Eur. J.* **2017**, *23* (34), 8152–8155.

25

(57) Buller, F.; Steiner, M.; Frey, K.; Mircsof, D.; Scheuermann, J.; Kalisch, M.; Bühlmann, P.;

Supuran, C. T.; Neri, D. Selection of Carbonic Anhydrase IX Inhibitors from One Million

DNA-Encoded Compounds. *ACS Chem. Biol.* **2011**, *6* (4), 336–344.

(58) Goodnow, R. A.; Dumelin, C. E.; Keefe, A. D. DNA-Encoded Chemistry: Enabling the

Deeper Sampling of Chemical Space. *Nat. Rev. Drug Discov.* **2017**, *16* (2), 131–147.

(59) Eidam, O.; Satz, A. L. Analysis of the Productivity of DNA Encoded Libraries. *Med.

Chem. Commun.* **2016**, *7* (7), 1323–1331.

(60) Maianti, J. P.; McFedries, A.; Foda, Z. H.; Kleiner, R. E.; Du, X. Q.; Leissring, M. A.;

Tang, W.-J.; Charron, M. J.; Seeliger, M. A.; Saghatelian, A.; Liu, D. R. Anti-Diabetic

Activity of Insulin-Degrading Enzyme Inhibitors Mediated by Multiple Hormones. *Nature*

2014, *511* (7507), 94–98.

(61) Usanov, D. L.; Chan, A. I.; Maianti, J. P.; Liu, D. R. Second-Generation DNA-Templated

Macrocycle Libraries for the Discovery of Bioactive Small Molecules. *Nat. Chem.* **2018**,

10 (7), 704–714.

(62) Finotello, F.; Di Camillo, B. Measuring Differential Gene Expression with RNA-Seq:

Challenges and Strategies for Data Analysis. *Brief. Funct. Genomics* **2015**, *14* (2), 130–

142.

(63) Melillo, B.; Zoller, J.; Hua, B. K.; Verho, O.; Borghs, J. C.; Nelson, S. D.; Maetani, M.;

Wawer, M. J.; Clemons, P. A.; Schreiber, S. L. Synergistic Effects of Stereochemistry and

Appendages on the Performance Diversity of a Collection of Synthetic Compounds. *J.

Am. Chem. Soc.* **2018**, *140* (37), 11784–11790.

(64) Kuai, L.; O'Keeffe, T.; Arico-Muendel, C. Randomness in DNA Encoded Library Selection

Data Can Be Modeled for More Reliable Enrichment Calculation. *SLAS Discov.* **2018**, *23*

(5), 405–416.

(65)   Wawer, M. J.; Jaramillo, D. E.; Dančík, V.; Fass, D. M.; Haggarty, S. J.; Shamji, A. F.;

Wagner, B. K.; Schreiber, S. L.; Clemons, P. A. Automated Structure–Activity

Relationship Mining. *J. Biomol. Screen.* **2014**, *19* (5), 738–748.

(66)   Doak, B. C.; Over, B.; Giordanetto, F.; Kihlberg, J. Oral Druggable Space beyond the

Rule of 5: Insights from Drugs and Clinical Candidates. *Chem. Biol.* **2014**, *21* (9), 1115–

1142.

(67)   Yuen, L. H.; Franzini, R. M. Achievements, Challenges, and Opportunities in DNA-

Encoded Library Research: An Academic Point of View. *ChemBioChem* **2017**, *18* (9),

829–836.

(68)   Kleiner, R. E.; Dumelin, C. E.; Tiu, G. C.; Sakurai, K.; Liu, D. R. In Vitro Selection of a

DNA-Templated Small-Molecule Library Reveals a Class of Macrocyclic Kinase

Inhibitors. *J. Am. Chem. Soc.* **2010**, *132* (33), 11779–11791.

(69)   Favalli, N.; Bassi, G.; Scheuermann, J.; Neri, D. DNA-Encoded Chemical Libraries -

Achievements and Remaining Challenges. *FEBS Lett.* **2018**, *592* (12), 2168–2180.

For Table of Contents Only

100k+ DNA-barcoded compounds | 8 cyclic scaffolds | all stereoisomers

cheminformatic analysis

affinity-based screening