Diversity-oriented synthesis encoded by deoxyoligonucleotides

Liam Hudson,[1,2] Jeremy W. Mason,[1,2] Matthias V. Westphal,[1,2] Matthieu J. R. Richter,[1] Jonathan R. Thielman,[1] Bruce K. Hua,[1] Christopher J. Gerry,[1] Guoqin Xia,[3] Heather L. Osswald,[3] John M. Knapp,[1] Zher Yin Tan,[1] Praveen Kokkonda,[1] Ben I. C. Tresco,[1] Shuang Liu,[1,6] Andrew G. Reidenbach,[1] Katherine S. Lim,[1] Jennifer Poirier,[2] John Capece,[2] Simone Bonazzi,[2] Christian M. Gampe,[2] Nichola J. Smith,[2] James E. Bradner,[2] Connor W. Coley,[1,4] Paul A. Clemons,[1] Bruno Melillo,[3] Johannes Ottl,[5] Christoph E. Dumelin,[5] Jonas V. Schaefer,[5] Ann Marie E. Faust,[2] Frédéric Berst,[5] Stuart L. Schreiber,[1,6] Frédéric J. Zécri,[2*] Karin Briner[2]

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

[2] Novartis Institutes for BioMedical Research, 181 Massachusetts Avenue, Cambridge, MA 02139, USA

[3] Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA, 92037, USA

[4] Department of Chemical Engineering, MIT, Cambridge, Massachusetts 02139, USA

[5] Novartis Institutes for BioMedical Research, Novartis Pharma AG, Novartis Campus, CH-4002, Basel, Switzerland

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

Abstract

Diversity-oriented synthesis (DOS) is a powerful strategy to prepare molecules with underrepresented features in commercial screening collections, a direct result of which has been to elucidate novel biological mechanisms of action. Running in parallel to the development of DOS over the past few decades, DNA-encoded libraries (DELs) have emerged as an effective and efficient screening strategy to identify protein binders. Yet, despite recent advancements in this field most DEL syntheses are limited by the presence of sensitive DNA-based constructs. Here we describe the design, synthesis and validation experiments performed for a 3.7 million-member DEL using diverse skeleton architectures derived from DOS to achieve

structural diversity beyond that afforded by varying appendages alone. We will make this encoded collection available to the academic scientific community to increase access to novel structural features and accelerate early-phase drug discovery.

Introduction

Screening collections comprising novel chemical structures have been shown to afford valuable probes of novel targets, including ones that exert their biological effects through novel mechanisms of action (nMoA).[1] One approach to produce such collections with broad molecular diversity is DOS,[2–6] which has afforded numerous nMoA compounds through phenotypic screening campaigns.[7–19] Similar success with target-based screening approaches may be challenging in cases where detailed knowledge of the underlying biology is unavailable, where there is no knowledge of neither ligandable pockets, nor known ligands, or where desirable efficacy can only be achieved through polypharmacological modulation of the target system.[20] Despite the long-held goal of DOS — to increase structural and especially performance diversity in screening collections — there are few such collections readily available to the scientific community from commercial vendors. Compounding the challenges to widespread adoption of such collections is that these compounds generally garner increased cost, with reduced or sporadic supply of key synthetic intermediates for the rapid confirmation and follow-up of any identified hit compounds.[21] Consequently, access to such compounds is restricted to researchers with substantial funding and/or synthetic chemistry resources, representing a barrier to new discovery in hit-generation sciences.

DEL technology has been established as a general screening tool in industry and academia to identify binders of soluble proteins via affinity-based screening experiments. The general concept of DEL technology — to prepare large libraries of compounds through combinatorial split-and-pool chemistry with concurrent encoding of each individual molecule by a unique DNA barcode — facilitates simultaneous screening of entire libraries of compounds in parallel for binding to a target of interest.[22] The amount of material required for these screens is greatly reduced compared to traditional high-throughput screening (HTS) as it can leverage the sensitivity and reliability of the polymerase chain reaction (PCR) and next-generation sequencing (NGS) for data generation.[23] Various approaches have been successfully developed over the years that expand the applicability of DELs in terms of synthetic chemistry compatibility and thus expand the chemical space coverage of libraries (typically in an aqueous environment, restricted by the sensitivities of DNA to reagents,

heat, and pH).[24–28] In addition, progress has been made in screening DELs such as the use of alternate coding schemes and novel procedures for performing screens that broaden the range of tractable targets and enhance the value of information garnered.[29–32]

The “single pharmacophore library”[33] is the most common strategy for synthesizing DELs. Here, successive steps of appendage diversification of a common skeleton (pharmacophore) are encoded by enzymatic ligation that record the synthetic history for each individual compound (**Fig. 1a**). In this example the skeleton may be generated in situ or off-DNA depending on the type of building blocks to be used in synthesis. Another approach (**Fig. 1b**) does not include any fixed central skeleton, and instead directly connects diverse building blocks in a linear fashion. An expanded approach (**Fig. 1c**), which is the focus of this study, makes use of multiple skeletal elements that are consistent in reactivity to allow for simultaneous appendage diversification using a common set of diverse appendages. An attractive feature of this strategy is that the common appendages vary greatly with respect to their relative exit vectors.

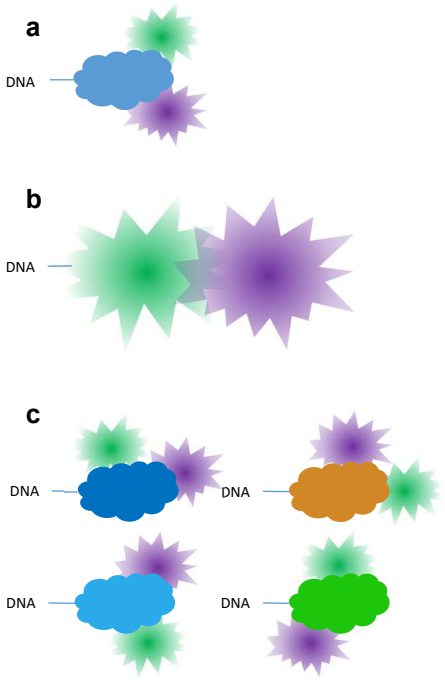

Fig. 1 | Conceptual comparison of DEL library strategies. Clouds with solid coloring represent multifunctional skeletons with defined exit vectors, spiked shapes with graduated coloring represent collections of building blocks with diverse pharmacophoric features. **a**, A single fixed skeleton library wherein appendage diversity emanates from fixed exit vectors of the central skeleton. **b**, A skeleton-free library constructed by direct linkage of diverse building blocks to each

other. **c**, A library comprising multiple skeletons bearing well-defined but variable exit vectors to collections of building blocks with diverse chemical features.

We aim to facilitate the screening of DELs comprising compounds with diverse skeleton architectures readily derived through DOS chemistry. We call this approach diversity-oriented synthesis encoded by deoxyoligonucleotides (DOSEDO). Here, we describe the design, synthesis and validation experiments performed using the first DOSEDO library. This collection is being made available to the academic scientific community through a mechanism discussed later in this report.[34]

Results

Selection of components.

While DEL technology allows for the rapid generation of extremely large compound libraries, we sought to prepare a moderate-sized library relative to those described in some industrial reports.[35] The reasons for this were three-fold: 1) to ensure that reaction validation prior to library production could be performed with good representation of the final library and be manually controlled by a single operator with high reproducibility. 2) to ensure that moderate-to-high coverage of library sequences was readily achievable in widely available and relatively low cost NGS experiments, and 3) to ensure a manageable cost for acquisition of raw materials to enable library production and follow-up studies. We therefore selected 61 multifunctional compounds to serve as skeletons,[36–42] all comprising secondary amines (Fmoc-, Boc- or Ns-protected) and an aryl halide (Br or I), which we envisioned using for on-DNA diversification (**Fig. 2**). In addition, each skeleton bore either a carboxylic acid or primary hydroxyl group, which were used as the site of DNA-attachment, and in some cases allowed for an alternative DNA-attachment point. We also sought to include as many discrete stereo- and position-isomers of these skeletons as practical.

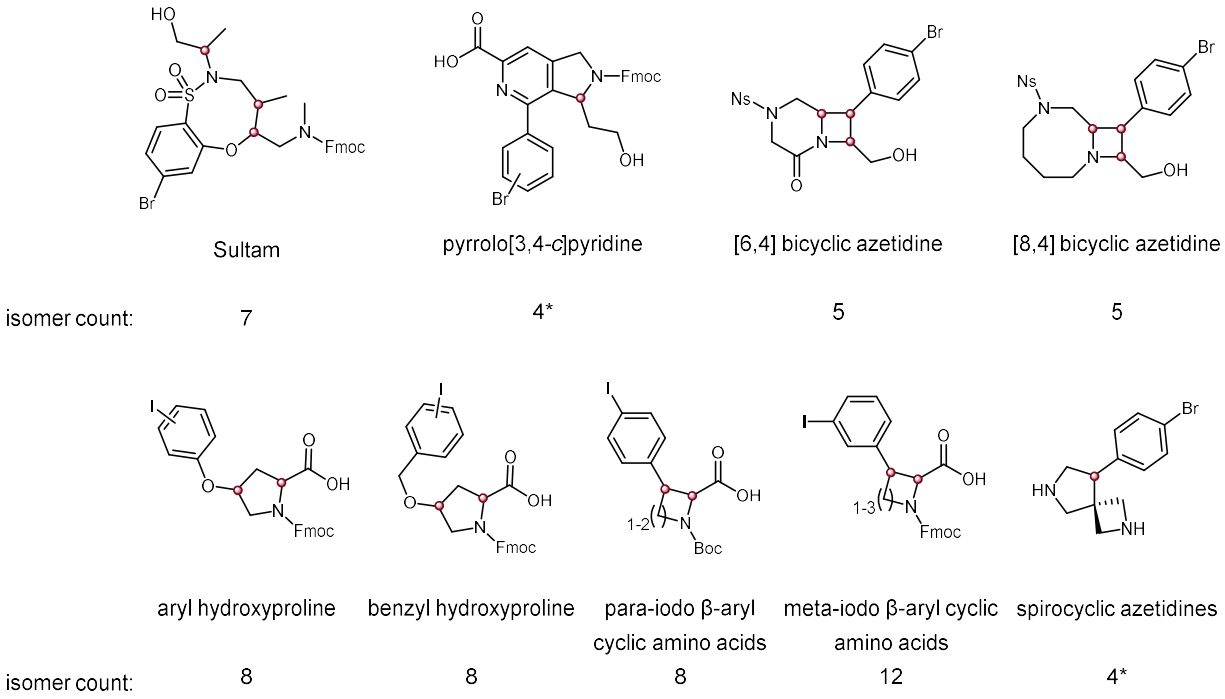

Fig. 2 | Overview of skeletons selected for inclusion in the DOSEDO library. Skeletons shown are indicative of the compounds loaded onto DNA. Isomer counts report the number of isomers (stereogenic and positional) for the indicated skeletons. Red spheres indicate two distinct isomers at each stereogenic center – in no cases were stereochemical mixtures or skeletons of unknown absolute configuration used in the library construction. Asterisks indicate skeleton series having alternate DNA-attachment points.

A branching rather than linear sequence library design was implemented such that the rigid central skeletons, adorned with well-defined exit vectors, afford a broad distribution of appendages in three-dimensional space.[43] We also focused on amine and aryl halide functional groups to enable use of thoroughly validated DNA-compatible diversification steps (acylation, sulfonylation, reductive amination, and Suzuki couplings respectively).[24]

Building blocks for appendage to the above-described skeletons were selected through a semi-automated process created in Knime Analytics Platform[44] to pick diversely from commercially available compounds within a defined calculated property space followed by manual review (**Supplementary Fig. 22**).

For our coding scheme, we opted to use a modified version of the double-stranded DNA scheme previously described by scientists at GlaxoSmithKline (GSK).[45]

Development of DNA-linkage chemistry. Skeletons attached to DNA through their carboxylic acids were coupled under modified literature conditions[45] and optimized to be sparing of material (**Supplementary Tables 6–8**). Skeletons attached to DNA through their primary hydroxyl group were coupled using an

optimized carbamate-linkage protocol (**Supplementary Section 3.2.1.2.**). The hydroxyl group was activated with *N*,*N′*-disuccinimidyl carbonate (DSC), filtered over silica to remove *N*-hydroxysuccinimide (NHS) and unreacted DSC, and incubated with amine-functionalized DNA. We also developed a DNA-compatible solution-phase Ns deprotection (**Supplementary Section 3.2.2.)** with similarities to a recently reported method,[46] rather than making use of previously described solid-supported processes.[47,48]

After linkage of each skeleton to DNA, purification was performed by preparative HPLC with concentration and desalting of the resulting HPLC eluents performed by EtOH precipitation and ultrafiltration. Chromatographic data for the purified constructs following reaction validation and library production are provided in **Supplementary Table 11**.

Reaction optimization. With purified skeleton–DNA conjugates in hand, we began optimization of the diversification steps of the DEL synthesis. Initial focus was given to reducing the residual amount of acetate derived from preparative HPLC mobile phase, which generally presented a chemoselectivity challenge in acylation reactions. Acylation conditions were optimized using DNA-conjugated proline as a representative amine. Following preparative HPLC, multiple washes of EtOH precipitates, ultrafiltration, or desalting on G-25 Sephadex led to excellent minimization of acetylation by-products in acylation reactions. We found that minor adaptations of published acylation,[45] reductive amination and sulfonylation conditions were effective across many substrate–reactant combinations.[36]

We opted to leverage Suzuki coupling reactions for the diversification of aryl bromides and iodides in the DOSEDO library. Searching for general coupling conditions appropriate for all targeted skeletons, we first performed several optimization reactions (**Supplementary Fig. 3**). Four palladium catalysts were screened in combination with two boronic acid pinacol esters dissolved in either EtOH, MeCN, or DMA. A total of 288 reactions were performed, leading to the selection of $PdCl_2(dppf)\cdot CH_2Cl_2$ as the preferred catalyst. This catalyst achieved the highest conversion to desired product at all temperatures and in all co-solvents relative to matched reaction sets. In addition, the $PdCl_2(dppf)\cdot CH_2Cl_2$ catalyst resulted in the least variable outcomes across a range of temperature and time combinations. We also found that MeCN was the preferred solvent in combination with any of the catalysts tested, though in practice we found that EtOH was complementary to MeCN for building block dissolution. A 1:1 mixture of EtOH and MeCN was used in subsequent reactions for increased homogeneity.

Building block validations. Amine-capping building blocks were profiled using HPLC-purified and desalted **Pro-DNA** as a model system following the general scheme shown in **Fig. 3a**. At the end of each reaction, one relative volume of 5% aqueous piperidine was added to quench residual electrophile, and an EtOH precipitation was performed prior to LCMS analysis. A tabulated dataset can be found in **Supplementary**

Table 12, and

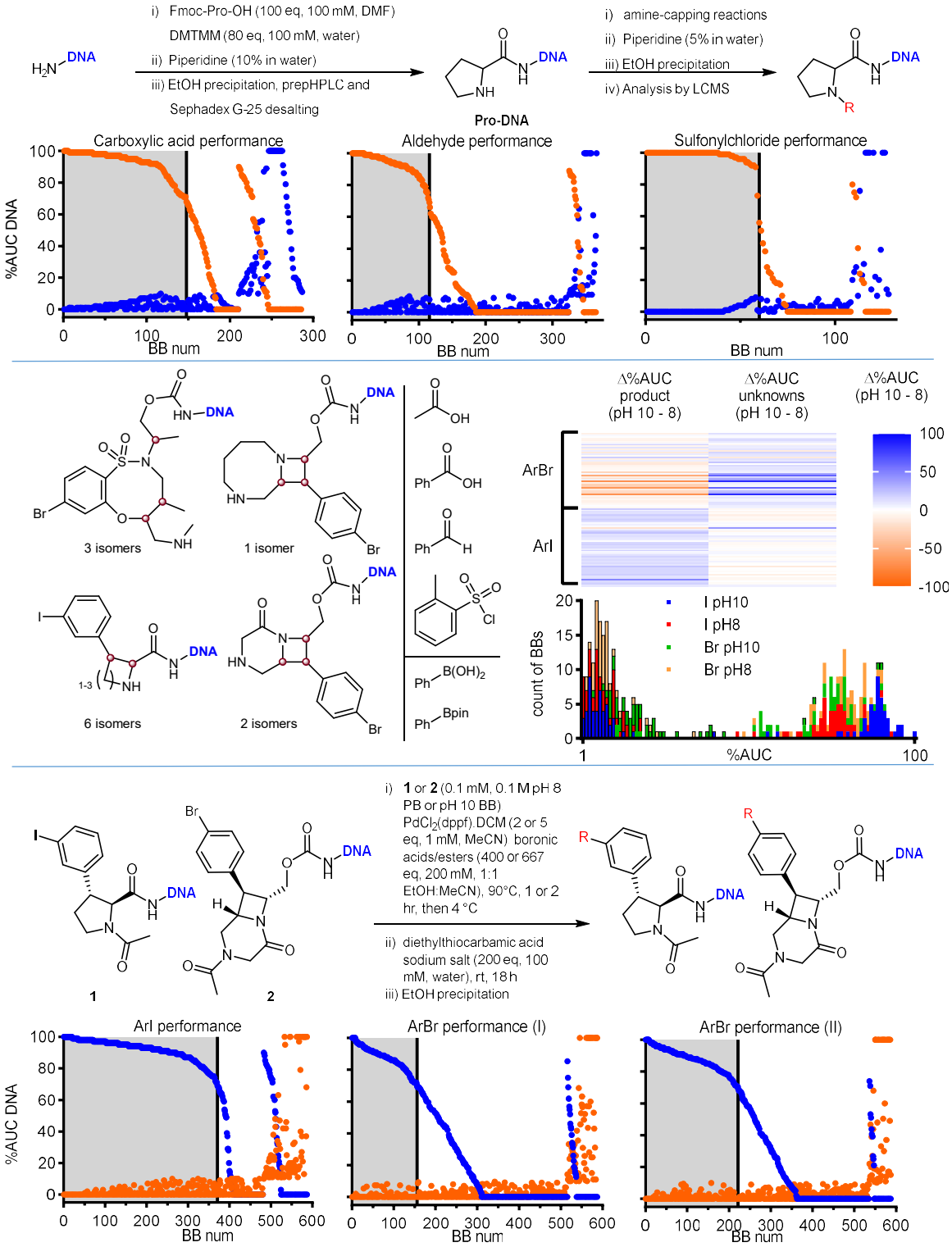
i) Fmoc-Pro-OH (100 eq, 100 mM, DMF)
DMTMM (80 eq, 100 mM, water)
ii) Piperidine (10% in water)
iii) EtOH precipitation, prepHPLC and Sephadex G-25 desalting
Pro-DNA
i) amine-capping reactions
ii) Piperidine (5% in water)
iii) EtOH precipitation
iv) Analysis by LCMS
Carboxylic acid performance
Aldehyde performance
Sulfonylchloride performance
%AUC DNA
BB num
3 isomers
1 isomer
6 isomers
2 isomers
Δ%AUC product (pH 10 - 8)
Δ%AUC unknowns (pH 10 - 8)
Δ%AUC (pH 10 - 8)
ArBr
ArI
count of BBs
%AUC
I pH10
I pH8
Br pH10
Br pH8
i) 1 or 2 (0.1 mM, 0.1 M pH 8 PB or pH 10 BB) PdCl2(dppf).DCM (2 or 5 eq, 1 mM, MeCN) boronic acids/esters (400 or 667 eq, 200 mM, 1:1 EtOH:MeCN), 90°C, 1 or 2 hr, then 4 °C
ii) diethylthiocarbamic acid sodium salt (200 eq, 100 mM, water), rt, 18 h
iii) EtOH precipitation
ArI performance
ArBr performance (I)
ArBr performance (II)

Fig. 3a reports the output for each reaction class. Only building blocks demonstrating >70% area-under-the curve (AUC) conversion to desired product with <10% AUC unknown species were considered for inclusion in the DOSEDO library.

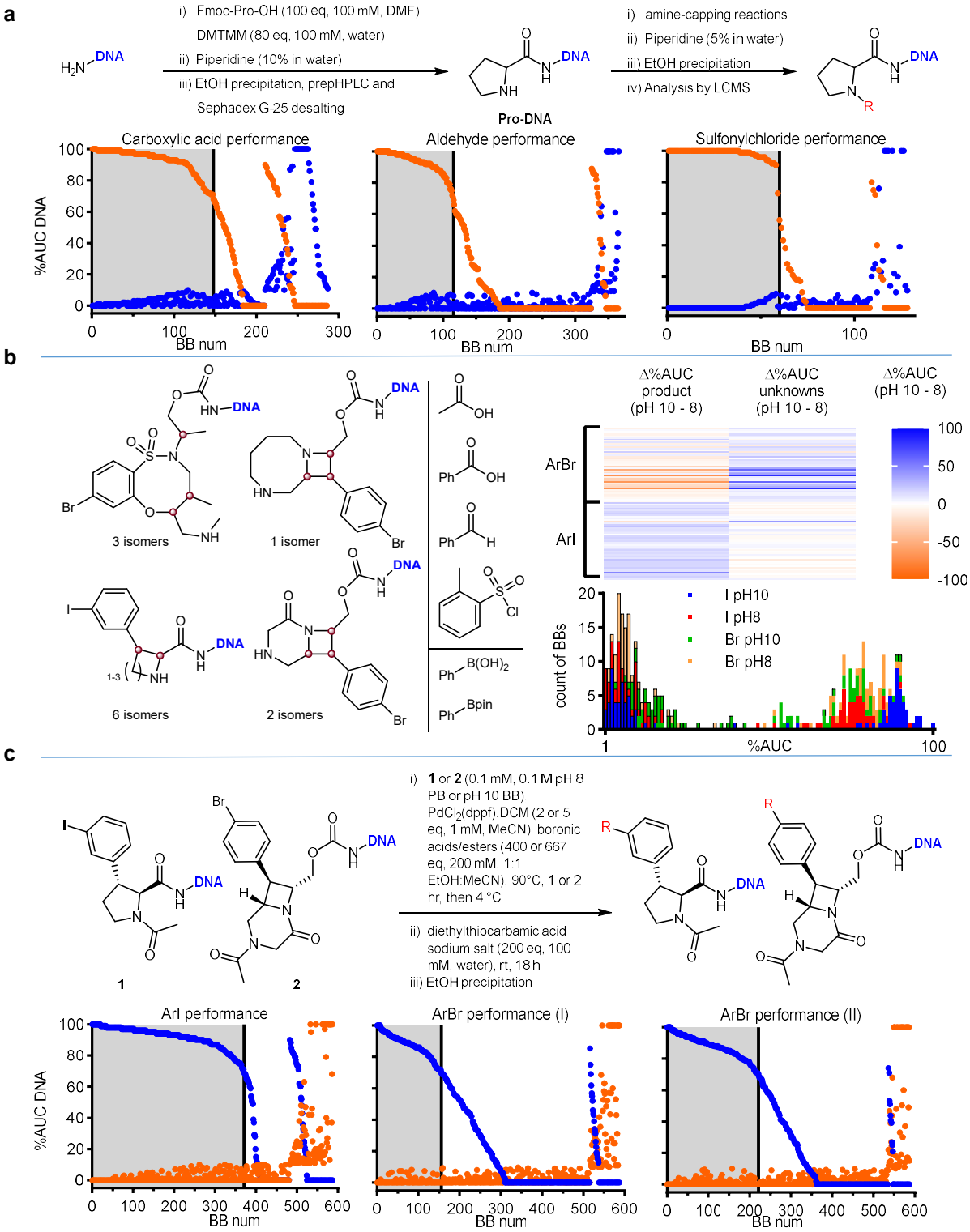

Fig. 3 | Summary of building block validations. a, Summary of amine-capping validation experiments, including a general synthetic scheme for preparation and use of a proline-based model construct (Pro-DNA), and charts showing the performance of acylation, reductive alkylation and sulfonylation. Blue points show %AUC corresponding to desired product species; orange points show %AUC of unknown DNA species as assessed by MaxEnt1 deconvoluted mass spectra. "BB num" is a ranking of compounds in two phases: i) decreasing %AUC product with increasing %AUC unknowns up to 10, then ii) decreasing %AUC product with >10 %AUC unknowns. The highlighted grey region of each chart denotes building blocks that met our inclusion criteria of >70%AUC product with <10%AUC unknowns. **b**, Effect of buffer on performance of Suzuki couplings comparing a subset of skeletons [with variably capped amines] bearing ArI versus ArBr. The skeleton series, amine capping, and model Suzuki coupling building blocks are shown, alongside a heat map representation of the difference in %AUC product and unknown species present when analogous reactions were performed at pH 10 vs. pH 8. Also displayed is a stacked bar chart comparing the performance of ArBr and ArI skeletons, on aggregate, with respect to product formation (bars without borders) and unknown species (bars with black borders) in different buffers. **c**, Summary of Suzuki coupling building-block validations including the constructs and conditions used and a comparative analysis of performance for a model ArI and ArBr (with the chart labelled "Arylbromide (II)" making use of higher catalyst and boronic acid/ester loading). Additional reaction time is also noted.

Prior to initiating a full Suzuki-coupling validation campaign, we applied the developed conditions to a subset of DNA-conjugated skeletons bearing uncapped or protected amines (**Supplementary Fig. 4**) using three building blocks that generally afforded high conversion to desired products. Interestingly, we found that the DNA-conjugated skeletons (as free amines) gave variable outcomes based on the skeleton class or stereoisomer. We also noted generally poor conversion for these couplings, mirroring the conclusions drawn for related published skeletons.[36] We therefore wanted to assess the effect of amine-capping mode on Suzuki coupling for different skeleton classes and to assess the reactivity of aryl iodides versus bromides.

We performed acylations with acetic acid and benzoic acid, sulfonylations with tosyl chloride, and reductive alkylations with benzaldehyde on six DNA-conjugated aryl iodides and six aryl bromides (**Fig. 3b**). The resulting 48 products were then subjected to Suzuki coupling conditions as developed (i.e. using a 0.1 mM solution of DNA-conjugate in pH 10 bicarbonate buffer), with phenylboronic acid or its pinacol ester. During early optimization experiments, we noticed some apparent degradation of [6,4] bicyclic azetidine skeletons, possibly resulting from basic hydrolysis. Therefore, we replicated the above set of reactions with the only change being to use a pH 8 phosphate buffer.

Following analysis of the resulting 192 LCMS traces, several conclusions could be drawn. First, use of boronic acid or pinacol ester had little differential effect on average %AUC product formation, which were 78 and 76 respectively with average %AUC unknown DNA species of 11 in both cases. Likewise, the different amine capping modes had no substantial impact on the conversion to desired products (with average %AUC {standard deviation} for: benzoic acid = 80{10}; acetic acid = 77{11}; *o*-toluenesulfonyl chloride = 77{20}; benzaldehyde = 73{11}). A key differentiating factor was the combination of buffer pH and whether the skeleton bore a bromide or iodide. **Fig. 3b** shows the change in %AUC product and unknown DNA species

that occurred in response to the change of buffer. In general, the iodide-containing group of reactants (comprising *meta*-iodo phenyl cyclic amino acids based on azetidine, pyrrolidine and piperidine, as either the *cis* or *trans* diastereomers) gave preferred reaction profiles in pH 10 bicarbonate buffer, with considerably greater %AUC product and less %AUC unknown species present versus the otherwise identical pH 8 phosphate buffered conditions. The reverse was true for the aryl bromide group. We next applied the developed conditions with our full set of boronic acids and esters using exemplar ArI and ArBr model systems (**1** and **2** respectively, **Fig. 3c**), and found that while reaction profiles were generally effective with respect to unknown byproducts, changing the buffer alone did not afford a comparable validation rate between iodide and bromide model systems. A higher catalyst and boronic acid/ester loading, as well as a longer reaction time, did however lead to an acceptable number of validated building blocks with good concordance between the iodide and bromide validation campaigns.

Following validation of our complete collection of amine-capping reagents and boronic acids/esters, we checked the extent of available diversity that could be successfully incorporated into the final library design. In the most rudimentary sense, the expected reactivity was observed; for example: aryl and vinyl sulfonyl chlorides were considerably more productive than aliphatic sulfonyl chlorides (presumably due to rapid hydrolysis of the latter in aqueous mixtures), and sterically more hindered aldehydes and carboxylic acids tended to be less reactive. We then prepared, for each class of building block, a distance matrix generated from Morgan 2 fingerprints, and applied multidimensional scaling (MDS, by Sammon mapping, **Fig. 4**). Except for alkyl sulfonyl chlorides, we observed that in general positive reaction outcomes in amine capping reactions afforded good coverage of available chemical space across all reaction modes. However, for Suzuki couplings (**Fig. 4a**–**c**), it is noteworthy that the building blocks forming the core of the MDS map tend to perform better compared to the extremities, wherein the represented compounds comprise the most dissimilar building blocks to the rest of the collection. We conclude therefore that while our conditions do allow for reasonable inclusion of building blocks, they are biased against the less represented/more diverse chemotypes. Upon inspection we found that electron rich and/or hindered 5-membered heterocycles, as well as vinyl boronates, often constitute the bulk of poorly performing reactants in this process. Moreover, we noticed that seeking consensus positive outcomes in the ArI and ArBr Suzuki coupling validation campaigns led to an erosion of fingerprint diversity in the highest priority set of building blocks (**Fig. 4a**), but believe this sacrifice of appendage diversity for consistent synthetic outcomes to be a worthwhile compromise.

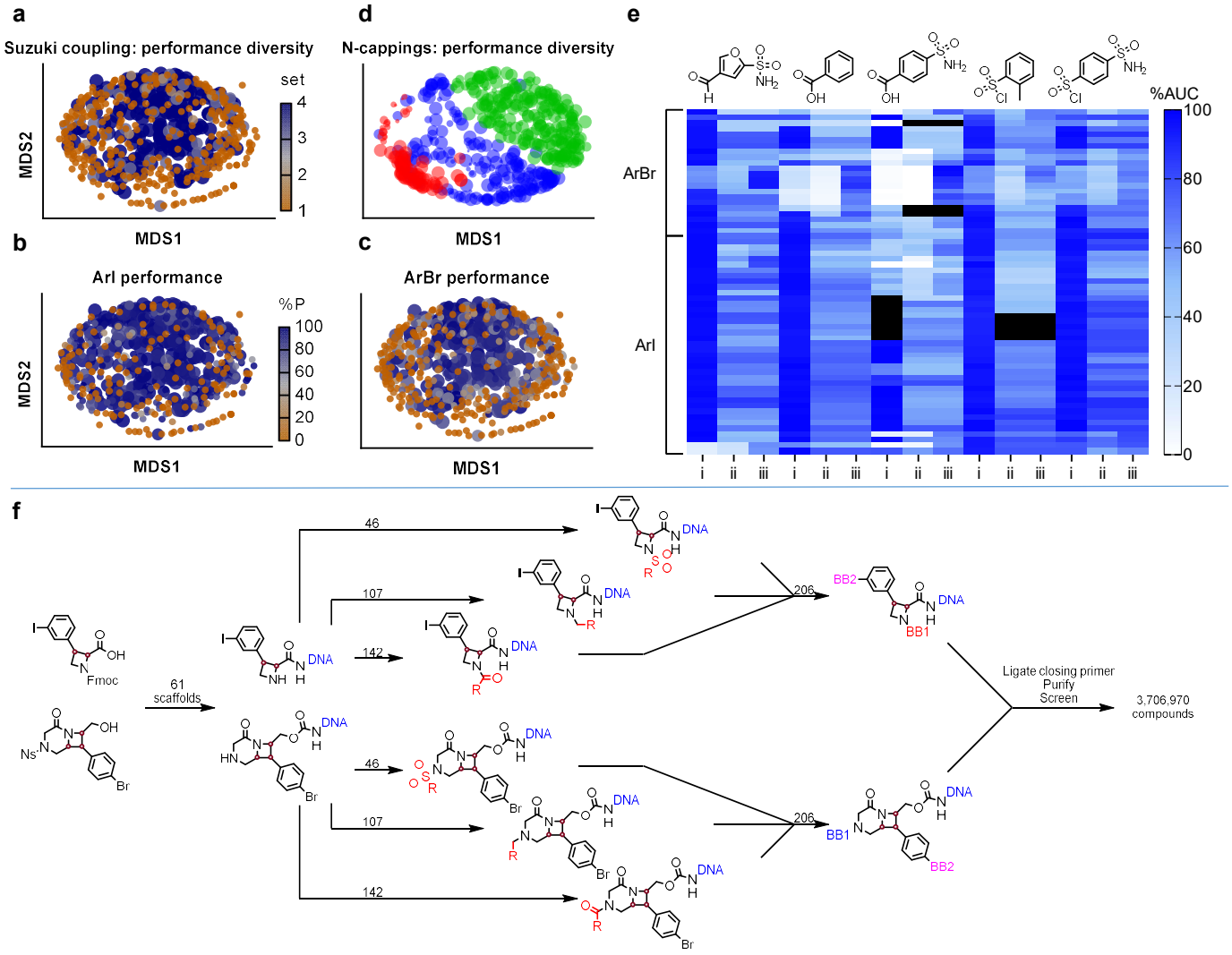

Fig. 4 | Assessment of reactivity in building block and skeleton validation experiments. a, Combined performance of ArI and ArBr Suzuki coupling reactions wherein each common boronic acid/ester moiety was replaced for a methyl group in silico, then visualized by Sammon mapping of a Morgan2 fingerprint distance matrix to two dimensions, coloring and sizing by preference for inclusion in DELs (4 being the highest), lowest preference BBs are layered atop higher preference for visualization. **b**, Performance of the same BBs represented in **a** specific to the ArI model system with size of points still indicative of consensus preference category but colored by adjusted %AUC product (%P). **c**, As for **b** showing data specific to the ArBr model system. **d**, Comparison of amine-capping performance by Sammon mapping of individual building blocks (without any in silico treatment of common reactive moieties), sized by %AUC product, blue – aldehydes, green – carboxylic acids, red - sulfonylchlorides. **e**, Heat-map representation of skeleton validation showing data for five different amine capping reactions followed by a single Suzuki coupling with (4-sulfamoylphenyl)boronic acid. Each row represents reactions of a specific skeleton. Column labelling indicates: i – adjusted %AUC for the desired amine-capped product; ii – adjusted %AUC of the desired Suzuki coupled product; iii – adjusted combined %AUC for the desired Suzuki coupled product as well as the Suzuki product of non-*N*-capped starting material. Black cells indicate either that the chromatographic data obtained was insufficient to resolve starting material and product peaks, or that no DNA-species appeared to be present based on chromatography (potentially due to loss of the DNA-pellet during EtOH precipitation, or an instrumentation error). **f**, Synthetic plan for DOSEDO library production using a single ArI and ArBr skeleton to represent the wider collection, with numbers indicating how many building blocks were employed at each step.

Based on the above-described analysis, we reasoned that the Suzuki coupling used in the library synthesis should be performed after amine-capping reactions, and under different conditions for the bromide and iodide-containing pools of intermediates. With those defined restrictions, we designed a synthetic plan accordingly (**Fig. 4f**) to maximize the performance and consistency of individual reactions. The synthesis began with attachment of skeletons to DNA-headpiece through either acylation or carbamylation. After amine deprotection, purification, and encoding of individual constructs, two separate sub-library pools were generated for ArI and ArBr-containing skeletons. These separate pools then underwent amine-capping and encoding reactions, followed by Suzuki coupling and encoding reactions. The two sub-libraries were pooled, closing sequences were ligated, and the sub-libraries were purified as a single encoded collection.

Skeleton validation. Prior to commencing synthesis of the DOSEDO library, we sought to validate that the conditions for amine capping, and Suzuki coupling were compatible with all skeletons planned for inclusion. We had hoped to identify any critical reactivity or stability issues; for example, to determine whether reactions are stereochemistry-dependent, which could have impact on the quality of the library or effect how we might interpret screening data or prioritize putative hits for resynthesis.

Detailed protocols for the skeleton validation campaign are described in **Supplementary Section 3.2.4**. To summarize: the concentration of each purified skeleton-DNA conjugate was normalized in nuclease-free water, the conjugates were subjected to five amine-capping reactions (at least one of each reaction class), a sample was taken for LCMS analysis, and the residual crude amine capping reaction mixture was subjected to a single Suzuki coupling (under skeleton-appropriate conditions) followed by a final LCMS and analysis (**Fig. 4e**)**.** We found that all skeleton–DNA constructs maintained their integrity throughout these reaction sequences, and performance was acceptable for all skeletons to be included in the DOSEDO library using optimized reaction conditions. We did however note substantive variability in the performance of specific reactions. For example, β-arylated cyclic amino acids in general behaved consistently – with high-yielding amine cappings and concurrent high-yielding Suzuki couplings – except in the case of *trans*-configured piperidines, which performed very poorly in reductive alkylations. In addition, while the sultam skeletons generally displayed slightly reduced reactivity towards amine-capping than other skeletons, they exhibited notably reduced reactivity across all seven stereoisomers in the panel towards acylation than both sulfonylation and reductive amination.

Based on the above-described findings, we chose to include all 61 skeletons, 295 amine-capping building blocks (comprising 142 carboxylic acids, 107 aldehydes and 46 sulfonyl chlorides), and 206 boronic acids/esters in the actual library synthesis. Moreover, we also decided to include two null-reaction conditions in the Suzuki coupling step ('subjected to reaction conditions without boronic acid/ester present', and 'not subjected to reaction conditions'), affording a theoretical combinatoric matrix of 3,706,970 encoded compounds.

Library synthesis. As a key goal of this work is to make this library available for community screening, we sought to ensure adequate scale of production to support that effort without introducing non-validated elements into our protocols. Therefore, the synthesis scale was limited by operational considerations rather than by available materials. A 15% excess of each skeleton was used at the outset of synthesis (100 nmol) relative to the limiting forward primer-binding site duplex. This ensured that variability in quantification of skeleton–headpiece conjugates was normalized during the first stage of synthesis. Skeleton identity was concurrently encoded by introduction of a cycle-1 tag in two–fold excess of the forward primer binding site duplex to ensure that all tandem ligations progressed to complete conversion (**Supplementary Fig. 7**).

After pooling iodides and bromides separately, we performed cycle-2 tag ligations with a slight molar excess (10%) relative to the cycle 1 tags used. Ligation efficiencies were assessed by analytical electrophoresis using pools of 12 samples per lane, and no evidence of individual ligation failure by native or denaturing methods was detected (**Supplementary Fig. 8 and 9**).

Amine-capping reactions were performed, then after quenching residual electrophiles with piperidine the products pooled and concentrated with respect to DNA species and cleansed of small-molecule components by ultrafiltration and EtOH precipitation. For continued synthesis, recovery was assumed to be quantitative. The pools of encoded intermediates were each first split, and then cycle-3 tag ligations were performed, again with a slight molar excess (approx. 10%) relative to the cycle-2 tag. Again, we monitored ligation efficiencies by analytical electrophoresis with 12 pooled samples per lane (**Supplementary Fig. 10**). Suzuki couplings were then performed according to the developed conditions above (**Fig. 3c**) and products were pooled as separate Br and I-derived sub-libraries.

Input analysis. Prior to a production-scale closing ligation, we engaged in a small test library closure and sequencing experiment – DOSEDO_v1. We assumed that efficiency of all processes leading up to this stage

were perfect and pooled a small quantity of the sub-libraries at a presumed 1:1 molar ratio at the individual encoded compound level. Closing sequences were ligated and purification was performed by continuous flow electrophoresis (**Supplementary Section 3.4**). The obtained 'yield' for DOSEDO_v1 closure over all steps of synthesis was between 19% (fluorometric quantification) and 37% (spectrophotometeric quantification) relative to the limiting forward primer binding site duplex.

DOSEDO_v1 was prepared for input sequence analysis according to the methods outlined in **Supplementary section 4,** with the resulting analysis shown in **Fig. 5a**. We deduced that pooling of sub-libraries was not equivalent; with I-derived compounds being in a 1.31–fold excess, on average, according to the two negative binomial distributions of the sub-libraries. We therefore adjusted the pooling ratio for a production-scale closure we termed DOSEDO_v2 (**Supplementary Section 3.5**). **Fig. 5b** shows the effect of input correction based on the v1 analysis. We also investigated the distribution of cycle-1, -2, and -3 tag ligations, on aggregate across all compounds, for both sub-libraries (**Supplementary Fig. 14 and 15**), and found that with very few exceptions distributions were tight, and concordance of ligation efficiency between sub-libraries was high. While depth of sequencing for the v1 library was not ideal (7% of expected sequences were not observed) we dedicated more reads to the v2 input analysis, and as a result could assess the unobserved sequences (0.03%; 1254 sequences) for potential trends of relevance to library use (**Fig. 5c**). This analysis presented evidence to suggest that a specific Cy2 tag ligation failed in many cases from the I-derived sub-library, but not the Br-derived pool, and in another example a combination of two Cy2 tags with a specific Cy3 tag was commonly unobserved, seemingly with the Cy1 identity having no impact on the outcome. With these two described exceptions, we assume that all ligations were similarly effective in both sub-libraries, and the v2 sub-library pooling leads to an equivalent representation of individual compounds in the final library.

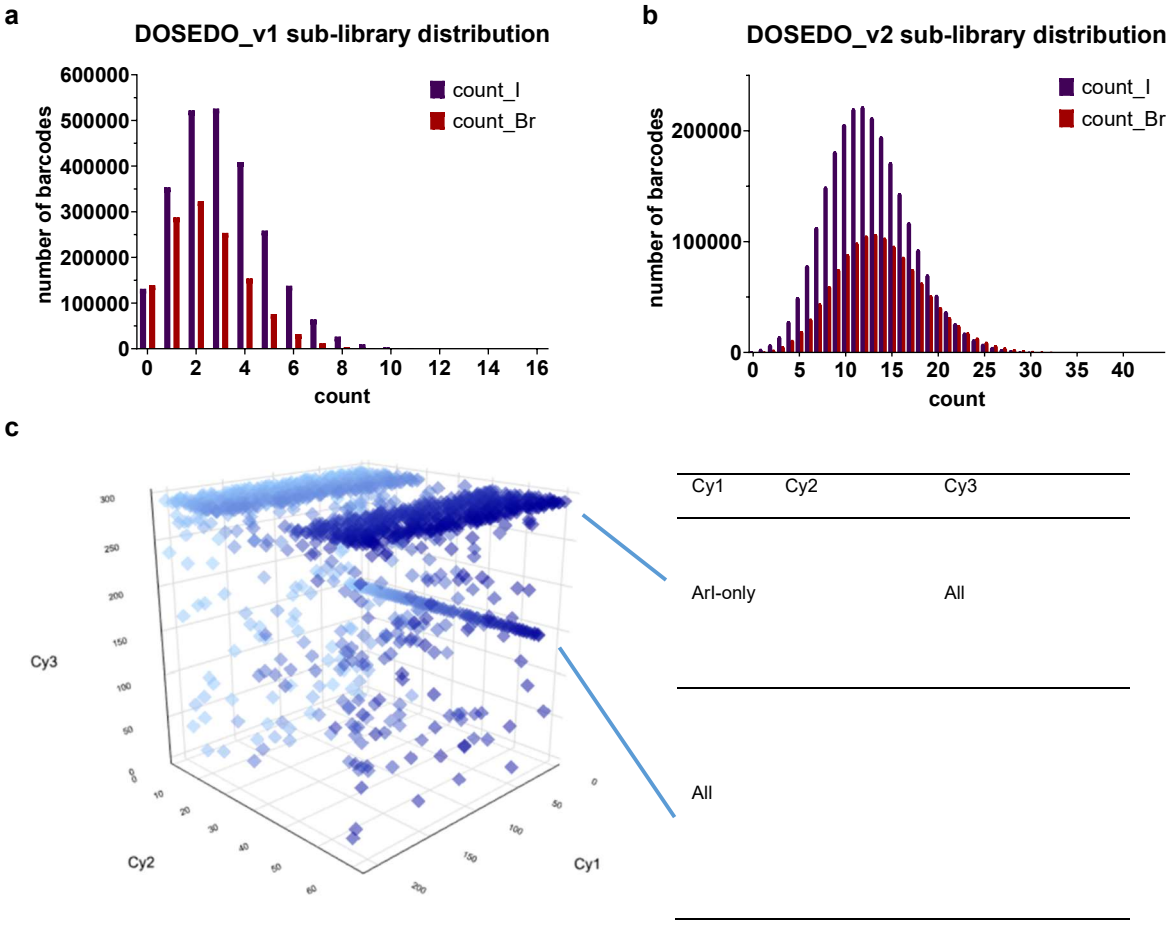

Fig. 5 | Assessment of input sequences. a, The number of barcodes observed at a given number of counts for the input DOSEDO_v1 library, showing the relative distribution of Br- and I-derived sub-libraries. **b**, As in **a** for DOSEDO_v2. **c.** unobserved sequences in DOSEDO_v2, with two sets of features highlighted in blue for being unobserved with unusual incidence as SAR (structure OR sequence activity relationship). Chart is graduated in blue in line with Cy2 identity to better perceive depth.

PoC/validation screening experiments. We next wanted to ensure that the synthetic chemistry during DEL synthesis performed as expected and was encoded as intended. To that end, we selected three protein targets to screen and subsequently synthesize putative hits off-DNA for validation purposes: carbonic anhydrase IX (CAIX), isocitrate dehydrogenase 1 (IDH1) R132H, and ubiquitin specific peptidase 7 (USP7).

CA isoforms are common model systems for the DEL field due to the primary sulfonamide structural motif being well known to bind – often with high potency and tolerance of other functionality in the ligands[49]. We therefore included at least one example of a primary sulfonamide in each reaction set performed. There are also potential clinical applications for selective inhibitors of CAIX.[50–52] IDH1 mutants have been implicated in a number of conditions, most notably in acute myeloid leukemia[53,54] and diffuse gliomas.[55] There are several

compounds known to bind IDH1 R132H's allosteric site,[56] some of which bear resemblance to compounds encoded within the DOSEDO library.[57] USP7 is a deubiquitinating enzyme (DUB) that can alter the degradation rate and cellular localization of specific protein substrates, some of which are of high interest in cancer progression.[61] We viewed mutant IDH1 and wild-type USP7 as more challenging proof-of-concept targets than CAIX, in that we had minimal prior knowledge of chemical features that promote binding and did not deliberately bias the collection towards potent chemotypes, as was the case for CAIX.

Screens were performed using an estimated 1 million copies of each encoded molecule per sample and using only a single cycle of panning the library over target protein. No dedicated optimization of conditions was undertaken to reflect more accurately the potential output of an initial experiment as it might be performed by external labs.

CAIX screening afforded enrichment of primary sulfonamide features above baseline noise (

Fig. 6). When including the additional variable of the skeletons some intriguing SAR presented itself – specifically a set of high-enrichment compounds and some low enrichment compounds of similar structure. This indicated a possible difference in binding based on subtle appendage changes or skeleton stereochemistry but could also have indicated a variable efficiency of on-DNA synthesis or bias induced by the screening protocol and NGS library preparation that was not fully captured by the prospective validation experiments. We therefore sought to probe these possibilities through off-DNA syntheses of the inferred compounds associated with high-barcode enrichment.

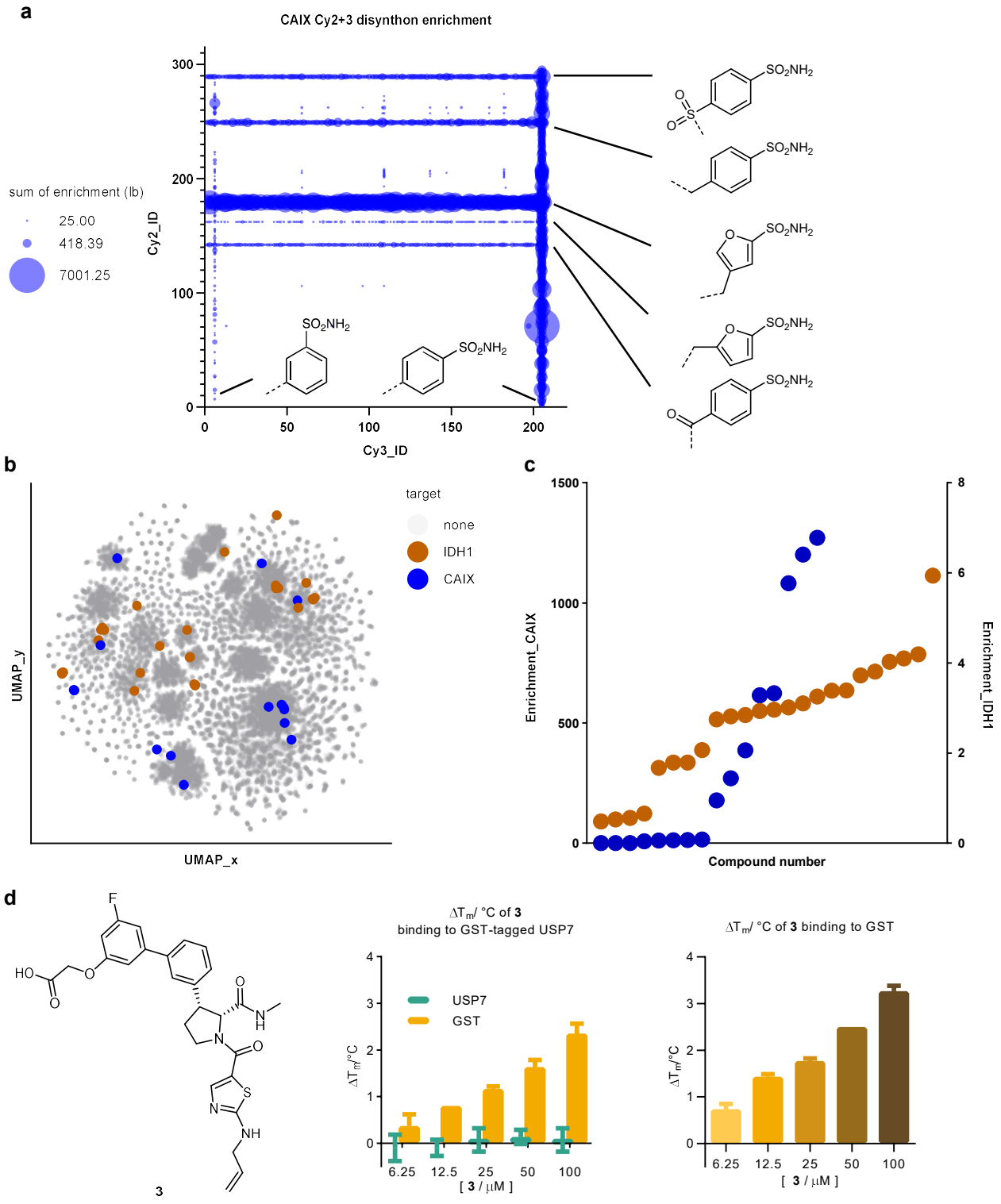

Fig. 6 | Analysis of validation screening exercises. a, CAIX screen output summarized by sum of lower bound (lb) calculated enrichments for all Cy2-Cy3 disynthons with all primary sulfonamide-containing building blocks included in the DOSEDO library shown (summed lower bound enrichments <25 was excluded from the plot for clarity). **b**, Chemical space map generated through UMAP multi-dimensional scaling of a Tanimoto distance matrix derived from Morgan

fingerprints of enumerated encoded compounds. A random sample of 50,000 compounds from the enumerated DOSEDO library was used as background chemical space along with 44 selected compounds for resynthesis. These are colored by their binding target. **c**, Calculated enrichments of selected resynthesis target compounds from CAIX- and IDH1 R132H-validation screens. **d**, Structure of **3** alongside DSF results with *N*-His-GST tagged USP7 as well as His tagged GST (5 μM) in 50 mM Tris-HCl, pH 7.4. Two peaks belonging to GST (T_m 54.6 °C) and USP7 (T_m 44.9 °C) respectively were observed in the melt curves of GST-tagged USP7. No thermal stabilization of USP7 by **3** was observed by tracking the USP7 peak; stabilization of GST by **3** was observed by tracking the GST peak. **3** induces thermal stabilization of His-tagged GST in a dose-dependent manner, up to 3.3 °C at 100 μM. Data are mean ± SD, n=3 technical replicates.

An IDH1 R132H screen performed using similar conditions yielded far lower dynamic range in counts and calculated enrichment ratios than the CAIX screen. Nevertheless, we identified apparent SAR across multiple series encoded within the library.

To validate the library synthesis, we selected compounds for resynthesis covering all reaction types, skeleton series, and levels of enrichment obtained from each screen. Wherever possible we also captured close structural analogues with markedly different enrichment values to probe the variability of the generated sequencing data relative to off-DNA validation (

Fig. 6b and c).

Upon commencement of off-DNA resynthesis we assumed that encoded compounds were the true compounds of interest, and for the sake of simplicity only targeted those compounds rather than attempting to recreate exactly the DEL synthesis conditions to capture byproducts or truncates. Compounds were prepared following closely related, published conditions,[36,38–40,62] often switching the order of steps relative to the DEL synthesis where it was deemed helpful on a case-by-case basis. We modified all carboxylic acids used for DNA-attachment as the methyl amides and left corresponding hydroxyl groups used for carbamate DNA-linkage unmodified.

Anticipated CA binders were principally assessed using a colorimetric enzyme-inhibition assay. While selected primary sulfonamides did tend to validate as active it is noteworthy that the anticipated non-binders did often inhibit CA function quite potently. Given the high enrichment ratios typically observed for validated CA binders, these false negative results are likely indicative of poor synthetic performance for specific reaction combinations in the pool (**Supplementary Table 10**).

Compounds prepared as putative IDH1 R132H binders were selected based on lower signal-to-noise sequencing data, and as such we anticipated lower correlation of enrichment to off-DNA binding. This did hold true, though 9 of the 25 compounds tested were validated as binders by SPR (**Supplementary Fig. 18**). Of the four skeleton series included in the resynthesized hits, we were able to identify at least two binding series, the sultam series being the more potent. We then went on to show that only the tightest binding sultam (K_D = 0.89 μM, also the most highly enriched compound from the NGS data) was able to effectively stabilize IDH1 R132H and display robust dose-dependent enzymatic inhibition (**Supplementary Fig. 20 and 21).** It is noteworthy that the specific stereochemical features of the identified hit align with the previously published IDH1 R132H inhibitor BRD2879,[57] bearing different substituents installed through an alternate reaction scheme. BRD2879 was previously identified using a conventional high-throughput screen and thus provides support for the overall synthesis, encoding, screening, barcode amplification, and next-generation sequencing steps of the experiments reported here.

For the USP7 screen, contrary to CAIX and IDH1 R132H screens, only a single compound was nominated for off-DNA synthesis (**3**), primarily because the level of enrichment was considerably lower relative to the matrix only control screen. USP7 was screened as an N-terminally His-tagged GST fusion protein, and we found that **3** selectively bound the GST portion of the construct. This finding was confirmed by DSF with isolated GST as compared to the GST-USP7 fusion (**Fig. 6d**), and then further by SPR studies (**Supplementary Fig. 19**).

Discussion

In conclusion, we have synthesized a DNA-encoded collection of compounds using a strategy based on both skeletal and appendage diversification and validated that the library can be used to identify binders to targets of interest. We have confidence that the on-DNA synthetic chemistry, as well as the concurrent encoding steps, was satisfactory, but our study also highlights the impact of noise in DEL screening data as well as the critical importance of the protein constructs used. We also have confidence that the encoded compounds will be of interest to the hit-discovery community since underrepresented chemical features are present, including novel and rigid skeleton architectures and well-defined stereogenic elements.

We will make this library available to the academic scientific community for screening by mirroring Novartis' existing Facilitated Access to Screening Technologies lab (FASTlab) framework.[34] At the conclusion of a

FASTlab DOSEDO screen, the academic researcher can freely use all data generated for their own research, which could potentially include further collaboration with Novartis.

It is our hope that this newly described DOSEDO resource will not only lead to the discovery of useful tool compounds for novel and challenging targets, but also foster collaboration for the broader scientific community to benefit.

Methods

Analytical and purification

Nuclear magnetic resonance (NMR) spectroscopy

NMR spectra were recorded on Bruker AV-III HD 2-channel spectrometers operating at a frequency of 400.14 MHz (^{1}H) and 100.61 MHz (^{13}C) and equipped with either a 5 mm BBO-F conventional (rt) probehead or a 5 mm BBO-F cryoprobe. Unless indicated otherwise, data was acquired at 25 °C using the software ICON-NMR under TopSpin program control and processed with MestReNova (Version 12.0.2-20910). Spectra were referenced to residual solvent resonance according to literature[58]. For spectra recorded at elevated temperatures, the same literature values of residual solvent resonances (at ambient temperature) were used for referencing, neglecting temperature-dependent peak shifts. Resonance signals are reported as chemical shifts (δ) in ppm with multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet or unresolved, br = broad signal), couplings constant(s) in Hertz (Hz) and integral (not for ^{13}C). ^{13}C and ^{19}F NMR spectra were recorded with broadband ^{1}H decoupling.

Reaction monitoring by ultra-performance liquid chromatography-mass spectrometry (UPLC-MS)

Samples were resolved on various Waters ACQUITY systems equipped with C18 columns (ACQUITY UPLC BEH C18 1.7 µm, 2.1x30 mm, ACQUITY UPLC BEH C18 1.7 µm, 2.1x50 mm, or ACQUITY UPLC CSH C18 1.7 µm, 2.1x50 mm) kept at 50 °C. UV signal was recorded with ACQUITY UPLC PDA detectors (210–400 nm). Light scattering signal was recorded with ACQUITY UPLC ELSD or SofTA 1100 ELSD detectors. Low-resolution mass spectra (LRMS) of eluting species were recorded with Waters SQD, SQD-2 or QDa detectors (electron spray ionization (positive and negative modes), scan time 0.3 sec, scanning range 120-1250 Da (QDa), 120-1600 Da (SQD) or 120-2850 Da (SQD-2)). High-resolution mass spectra (HRMS) of

eluting species were recorded on systems equipped with Waters Xevo G2 Qtof or Waters Xevo G2-XS Tof detectors (electron spray ionization (positive mode), scan time 0.2 sec, scanning range 100-2050 Da).

Solvent A1	0.1% formic acid in water
Solvent B1	0.1% formic acid in acetonitrile
Solvent A2	5 mM ammonium hydroxide in water
Solvent B2	5 mM ammonium hydroxide in acetonitrile
Solvent A3	0.05% trifluoroacetic acid in water
Solvent B3	0.05% trifluoroacetic acid in acetonitrile

RXNMON-Acidic

flow: 1.0 mL/min, runtime: 2.0 min

column: ACQUITY UPLC BEH C18 1.7 µm, 2.1x30 mm

time	%A (solvent A1)	%B (solvent B1)
0.00	98	2
0.10	98	2
1.50	2	98
1.80	2	98
1.90	98	2
2.00	98	2

RXNMON-Basic

flow: 1.0 mL/min, runtime: 2.0 min

column: ACQUITY UPLC BEH C18 1.7 µm, 2.1x30 mm

time	%A (solvent A2)	%B (solvent B2)
0.00	98	2
0.10	98	2
1.50	2	98
1.80	2	98
1.90	98	2

2.00	98	2

ProductAnalysis-Acidic

flow: 1.0 mL/min, runtime: 5.2 min

column: ACQUITY UPLC BEH C18 1.7 µm, 2.1x50 mm

time	%A (solvent A1)	%B (solvent B1)
0.00	98	2
4.40	2	98
5.15	2	98
5.19	98	2

ProductAnalysis-Basic

flow: 1.0 mL/min, runtime: 5.2 min

column: ACQUITY UPLC BEH C18 1.7 µm, 2.1x50 mm

time	%A (solvent A2)	%B (solvent B2)
0.00	98	2
4.40	2	98
5.15	2	98
5.19	98	2

HRMS

flow: 1.0 mL/min, runtime: 2.2 min

column: ACQUITY UPLC CSH C18 1.7 µm, 2.1x50 mm

time	%A (solvent A3)	%B (solvent B3)
0.00	98	2
0.06	98	2
1.76	2	98
2.00	2	98
2.16	98	2

Ion-pair chromatographic analysis of DNA-conjugates by ultra-performance liquid chromatography mass spectrometry (UPLC-MS)

Samples were resolved on a Waters ACQUITY system equipped with a C18 column (ACQUITY Oligonucleotide BEH C18 1.7 µm, 2.1x50 mm, part #186003949) kept at 50 °C. UV signal was recorded with an ACQUITY TUV detector (260 nm, sampling rate 20 points/sec). Mass spectra of eluting species were recorded with a Waters SQ Detector 2 connected to a ZSpray™ source (negative ion mode, scan time 0.2 sec, scanning range 500-3000 Da, MaxEnt1 deconvolution, processing 2-20 kDa).

Standard method (5-50% B)

flow: 0.5 mL/min, runtime: 6.0 min, injection volume: 10 µL, temp: 50 °C

column: ACQUITY Oligonucleotide BEH C18 1.7 µm, 2.1x50 mm

solvent A: 250 mM hexafluoroisopropanol and 8 mM NEt_3 in water

solvent B: methanol

time	%A (solvent A)	%B (solvent B)
0.00	95	5
4.00	50	50
4.50	5	95
5.00	5	95
5.10	95	5
6.00	95	5

Preparative high-performance liquid chromatography (prep-HPLC)

Purifications were performed on a Waters system equipped with Waters 515 pumps, a Waters 2545 binary gradient module, a Waters Acquity QDa detector, a Waters 2998 PDA detector and a Waters 2767 sample manager. The column (Waters XBridge Prep C18 OBD, 5 µm, 30 mm (inner diameter) x 50 mm) was kept at rt. Material was injected as solution in methanol/water mixtures (1.5 mL) and eluted with gradients of solvent A (water) and solvent B (acetonitrile), both modified with either 0.1% formic acid or 5 mM ammonium hydroxide (75 mL/min flowrate). Fraction collection was triggered by UV signal and mass (TIC) of eluting

species. Predefined gradients were picked from a list of methods optimized for target retention time as observed during reaction monitoring by UPLC-MS.

Purification of target compounds described in Section 2.3. was performed using a Teledyne ISCO ACCQPrep HP150 equipped with a XBridge C18 column (19 x 250 mm, 5 µm), eluting with 10–100% MeCN [+1% HCO_2H] in H_2O [+1% HCO_2H])

Preparative high-performance liquid chromatography of oligonucleotide conjugates (microprep-HPLC)

Purifications were performed on an Agilent Infinity system equipped with an Infinity 1260 Bio Quat Pump (pump system, G5611A), an Infinity 1260 HiP Bio ALS (autosampler, G1330B), an Infinity 1290 TCC (thermostatted column compartment, G5667A), an Infinity 1260 DAD (diode array detector, G4212B), and an Infinity 1260 Bio FC-AS (fraction collector, G5664A) under Agilent OpenLab CDS (C.01.07 SR1) software control. The column (Waters XBridge BEH C18 OBD, 130 Å, 5 µm, 10x150 mm) was kept at 50 °C. Material was injected as aqueous solutions (up to 100 µL per injection) and eluted with gradients of solvent A (50 mM triethylammonium acetate in water) and solvent B (MeOH). Fraction collection was triggered by UV (260 nm).

5-50% method

flow: 5.0 mL/min, runtime: 20 min

column: Waters XBridge BEH C18 OBD, 130 Å, 5 µm, 10x150 mm

time	%A (solvent A)	%B (solvent B)
0.0	95	5
15.0	50	50
16.0	5	95
18.0	5	95
18.1	95	5
20.0	95	5

Following separation desired species-containing fractions were pooled and concentrated by ultrafiltration (Millipore Amicon Ultra-15/4 centrifugal filters; Ultracel 10K/3K) according to manufacturer guidelines. Three cycles of sample concentration/buffer dilution with nuclease free water were employed.

Other methods and equipment

Centrifugation was performed with a Beckman Coulter Allegra X-15R or Beckman Coulter Microfuge 18. Aqueous solutions of DNA were quantified photospectrometrically with a Thermo Scientific Nanodrop One. Electrophoresis was performed with an Invitrogen E-GEL Power Snap device using commercial gels (E-Gel EX, 4% agarose) according to manufacturer guidelines. Denaturing electrophoresis was performed using Novex™ TBE-Urea Gels (15%, Invitrogen, EC68855BOX) prepacked gels and Novex™ TBE-Urea Sample Buffer (2X, Invitrogen, LC6876) according to manufacturer guidelines. Completed gels were stained with SYBR™ Gold Nucleic Acid Gel Stain (Invitrogen, S11494) according to manufacturer directions.

Preparative continuous flow electrophoresis

A Bio-Rad Model 491 Prep Cell was used according to manufacturer guidelines for the purification of crude DNA-encoded libraries following the ligation of oligonucleotides bearing a closing primer binding site sequence. The gel assembly tube (small or large) was cast with 4% agarose in 1x TBE buffer to a height of 8.5 – 9.5 cm and left to solidify over 2 hr while circulating rt water through the cooling core at approx 100 mL/min, carefully overlaying 2 mL of a 1:1 mixture of IPA and water on the top of the gel, then standing overnight at rt without water circulation through the cooling core. No stacking gel was cast. After standard assembly of the Prep Cell (see: http://www.bio-rad.com/webroot/web/pdf/lsr/literature/M1702925.pdf) and addition of 1x TBE buffer, the crude DEL sample was loaded in Novex™ Hi-Density TBE Sample Buffer (5X, Invitrogen LC6678), and electrophoresis was run at 12 W constant power. Fraction collection was performed using a Masterflex L/S easy-load II (Cole Parmer, 77200-50) running at approx 1 mL/min and BioFrac™ fraction collector (Bio-Rad) with a fraction time window of 2 min, i.e., approx 2 mL fractions). Total electrophoresis time was roughly 3.5 hr. Fractions were analyzed by E-GEL Power Snap device (Invitrogen, G8100) using commercial gels (Invitrogen, E-Gel EX, 4% agarose). Product containing fractions of adequate purity were pooled, then concentrated/desalted using Millipore Amicon Ultra-15/4 centrifugal filters; Ultracel 10K according to manufactures guidelines, washing at least twice with nuclease free water.

Flash chromatography

Automated flash chromatography was performed on Teledyne ISCO CombiFlash® systems equipped with TUV and ELSD detectors using prepacked columns (RediSep® Rf, prepacked with 4 g, 12 g, 24 g, 40 g, 80 g, 120 g, or 330 g silica, 35-70 µm). Crude material was usually adsorbed on diatomaceous earth (Biotage Isolute HM-N) and then subjected to chromatography (dry-loading) eluting with gradients of ethyl acetate in heptane, methanol in CH_2Cl_2, or as specified. Fractions containing homogeneous material according to detection method (TUV/ELSD) and/or thin layer chromatography (TLC) were pooled, and solvents were removed by rotary evaporation under reduced pressure at 30–45 °C to afford target compounds.

Compound synthesis and characterization

All chemical synthesis procedures and compound characterization data are provided in the Supplementary Information in Sections 1, 2, 7, and 8, and all DOSEDO library member structures can be found in **Supplementary Files 2** and **3.**

Preparation of next generation sequencing libraries

Barcodes of a DEL sample (input or eluted after screening) were PCR amplified using 3 µL i5 index primer (10 µM stock in water), 3 µL i7 index primer (10 µM stock in water), 19 µL cleaned up elution samples, and 25 µL Invitrogen Platinum™ Hot Start PCR Master Mix (2×) (Invitrogen 13000012). The PCR method is as follows: 95 °C for 2 min; 22 cycles of 95 °C (15 s), 55 °C (15 s), 72 °C (30 s); 72 °C for 7 min; hold at 4 °C. The PCR products were cleaned up using the ChargeSwitch PCR Clean-Up Kit, pooled in equimolar amounts, and the 187 bp amplicon was gel purified using a 2% E-Gel™ EX Agarose Gels (ThermoFisher Scientific G401002) and QIAquick Gel Extraction Kit (Qiagen 28704). The DNA concentration was measured using the Qubit dsDNA BR assay kit and sequenced using a HiSeq SBS v4 50 cycle kit (Illumina FC-401-4002) and HiSeq SR Cluster Kit v4 (Illumina GD-401-4001) on a HiSeq 2500 instrument (Illumina) in a single 50-base read with custom primer CTTAGCTCCCAGCGACCTGCTTCAATGTCGGATAGTG and 8-base index read 32 using custom primer CTGATGGAGGTAGAAGCCGCAGTGAGCATGGT (**Supplementary Figure 12**).

DEL screening

The protein targets and beads-only control (B buffer replacing the protein) were screened in duplicate using a KingFisher Duo Prime (Thermo Scientific) in a 96-well deepwell plate (Thermo Scientific 95040452) at room temperature. C-His tagged IDH1 R132H (Met1-Leu414) was purchased from G-Biosciences (BAN1708, 50

μg). N-His and GST tagged USP7 (Lys208-Glu560) were purchased from SinoBiological (11681-H20B, 100 μg) and reconstituted in 400 μL water. The buffers used are 'B Buffer' containing 25 mM HEPES pH 7.4, 150 mM NaCl, 0.05% Tween-20 (w/v), and 'S Buffer' containing 25 mM HEPES, pH 7.4, 150 mM NaCl, 0.05% Tween-20 (w/v) and 0.3 mg/mL Ultrapure Salmon Sperm DNA (ThermoFisher Scientific 15632011). His-Tag Dynabeads (Invitrogen 10103D, 10 μL per sample) was washed three times with B buffer before protein immobilization. The proteins were diluted to 1.2 μM in B buffer (100 μL per sample) and immobilized to the washed beads (1 hr, medium mix). The beads were washed once with B buffer (200 μL) and twice with S buffer (200 μL) (3 min each, medium mix). The beads were then transferred to the DEL library (1 million copies per library member, 100 μL in S buffer) and incubated (1 hr, medium mix). The beads with protein (except beads-only control) and DEL bound were washed once with S buffer (200 μL) and twice with B buffer (200 μL) (3 min each, medium mix). The beads were transferred to B buffer (100 μL) and heated (90 °C, 10 min) to elute DEL compounds into the supernatant. 20 μL of the elution was restriction digested by 0.1 μL StuI (NEB R0187) in 56.5 μL 1×SmartCutter buffer (NEB B7204S) per sample (37 °C, 1 hr) and cleaned up using the ChargeSwitch PCR Clean-Up Kit (Thermo Scientific CS12000) (**Supplementary Figure 15**).

Assay details

Carbonic anhydrase

Carbonic anhydrase II (CAII) inhibition assays were performed using a commercially available kit (BioVision, #K473-100), following the manufacturer's protocol. For each assayed DOSEDO compound, 11 concentrations (half-log dilutions from 10 μM to 100 pM) were tested, with three technical replicates performed on each of two separate 384-well μClear® medium-binding, flat-bottom polystyrene microplates (Greiner Bio-One, #82051-294) (n=6 for each compound concentration); for **53** only, six technical replicates (n=6) were all performed on the same plate for each compound concentration. Each reaction well contained 3 μL of compound (prepared in 10% DMSO, 90% kit buffer), 25.5 μL of enzyme/buffer mix (1.4 μL enzyme, 24.1 μL kit buffer), and 1.5 μL of substrate, for a total volume of 30 μL. Absorbance at 405 nm was read on a Molecular Devices SpectraMax i3x. In addition to the experimental conditions, a DMSO control and a no-enzyme control were each performed with three technical replicates on each plate, except for the experiment with **53**, where six technical replicates for each control were performed on the same plate. The average change in absorbance for the no-enzyme control was subtracted from the change in absorbance over the same time interval for each experimental condition and the DMSO control on the same plate. Then for each

concentration of each compound, relative activity was calculated separately for each replicate based on the average change in absorbance for the DMSO control on the same plate. Dose-response curves were fitted and IC50 values and 95% confidence intervals were calculated using GraphPad Prism v9.2.0. Curves were fitted using GraphPad Prism's log(inhibitor) vs. response – variable slope nonlinear regression equation, with the top asymptote constrained to 100. Outliers were detected and removed via GraphPad Prism's ROUT method, with a Q-value of 1%. Error bars (only shown if greater than the height of the symbol) represent 95% confidence intervals. See **Supplementary Figure 16**.

Recombinant homodimeric IDH1 R132H production

The plasmid pET-22b(+) containing the recombinant IDH1 R132H gene (Met1-Leu414) fused with a His6 tag at the C terminus, via NdeI and XhoI restriction sites, was expressed in the E. coli BL21(DE3) strain. A starter culture was grown in 100 mL of LB media supplemented with ampicillin (50 µg/mL) at 37 °C overnight. The starter culture (1:100 v/v) was used to inoculate 3 liters of LB media supplemented with ampicillin (50 µg/mL) and grown at 37 °C to an OD600 nm= 0.6. Expression was induced with 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) at 20 °C for 14 hr. The cells were collected by centrifugation (Thermo Scientific™ Sorvall™ LYNX 6000 centrifuge) at 8,000 × g for 10 min. The pellets were suspended in 30 mL of lysis buffer containing 20 mM Tris-HCl, 500 mM NaCl, pH 7.4, 1% Triton X-100, lysozyme (1 mg/mL), benzonase (10 µL), 1 mM TCEP and 1 mM phenylmethylsulfonyl fluoride (0.1 M solution in ethanol, 300 uL). The cells were lysed on ice by sonication (50% amplitude, 3 s pulse on and 3 s pulse off, for total 7 min 'on' time). The cell debris was precipitated by centrifugation (Thermo Scientific™ Sorvall™ LYNX 6000 centrifuge) at 30,000 × g for 60 min. The cell lysates were filtered through a 0.45 µm syringe filter and purified by affinity chromatography using a 5 mL HisTrap HP column (GE Healthcare Life Sciences) with wash buffer (20 mM Tris-HCl, 500 mM NaCl, pH 7.4) and a gradient of 4-100% elution buffer (20 mM Tris-HCl, 500 mM NaCl, 250 mM imidazole, pH 7.4) over 10 CV. The fractions containing IDH1 R132H were pooled, concentrated and further purified using a Superdex S200 26/600 200 pg column in 20 mM Tris-HCl, 100 mM NaCl, pH 7.4. Pure fractions of IDH1 R132H were pooled, concentrated, aliquoted and flash frozen before storage at -80 °C.

IDH1 R132H and USP7 surface plasmon resonance

IDH1 R132H: SPR experiments were performed on a Biacore T200 instrument (GE Healthcare) at 25 °C using a running buffer containing 50 mM HEPES pH 7.5, 50 mM KCl, 0.005% tween-20, 1 mM DTT, 2% DMSO. The compounds were 2-fold serial diluted from 20 µM to 78 nM with a 2% DMSO only control.

Biotinylated homodimeric IDH1 (RU ~3500) was immobilized to the streptavidin sensor chip (Cytiva BR100531) preconditioned with 1M NaCl and 40 mM NaOH (60 s ×3 cycles, 100 μL/min) and running buffer (60 s, 100 μL/min). The run has a startup run of 12 cycles using the running buffer, flow rate of 50 μL/min, contact time of 60 s and dissociation time of 150—500 s depending on the compound, with two negative controls using the running buffer between each compound. The syringe was washed with 50% DMSO between injections. Data was analyzed using the affinity mode of the Biacore T200 Evaluation Software and is presented in **Error! Reference source not found.**.

USP7 and GST: SPR experiments were performed on a Biacore T200 instrument (GE Healthcare) at 25 °C using a running buffer containing 20 mM HEPES pH 7.5, 100 mM NaCl, 0.005% tween-20, 0.2 mM TCEP, 1% DMSO. The compounds were 2-fold serial diluted from 20 μM to 78 nM with a 1% DMSO only control. His-GST-tagged USP7 (SinoBiological 11681-H20B) and His-tagged GST (Abcam ab89494) were immobilized to the streptavidin sensor chip (Cytiva BR100531) via two different channels. The streptavidin sensor chip was preconditioned with 1M NaCl and 40 mM NaOH (60 s ×3 cycles, 100 μL/min), running buffer (60 s, 100 μL/min), and functionalized to capture His-tagged proteins by 350 mM EDTA (60 s), 1 μM Tris-NTA biotin trifluoroacetate salt solution (Sigma-Aldrich 75543) (100 s), 500 μM $NiCl_2$ (100 s), 3 mM EDTA (60 s) (for all functionalizing steps: flow rate 100 μL/min, reagents in running buffer). The run has a startup run of 12 cycles using the running buffer, flow rate of 50 μL/min, contact time of 60 s and dissociation time of 150 s. The syringe was washed with 50% DMSO between injections. Data was analyzed using the affinity mode of the Biacore T200 Evaluation Software and is shown in **Supplementary Fig. 18**.

IDH1 R132H differential scanning fluorimetry

Thermal shift experiments were carried out using a QuantStudio™ 7 Flex system (Applied Biosystems) in a MicroAmp Optical 384-well plate (Applied Biosystems 4309849)[59]. Each well contained 10 μL of 5 μM protein with 10× Sypro Orange in 50 mM Tris-HCl, pH 7.4. Various concentrations of compounds (6.25 – 100 μM, 1% DMSO final) were mixed with the protein for thermal stabilization studies. The samples in triplicates were subjected to temperature increases from 25 °C to 95 °C at 0.02 °C s^{-1}, with optical filters of x1-m3 corresponding to excitation 470 nm and emission 586 nm respectively. Protein Thermal Shift Software v1.3 (Applied Biosystems) was used to determine the melting temperature, T_m, from the derivative of the melt curve. Data are shown in **Error! Reference source not found.**.

IDH1 R132H enzymatic inhibition

The inhibitory activity of the off-DNA compounds against IDH1 R132H were assessed by absorbance assays similarly to the reported procedures[60], using IDH1 R132H purified in-house as described above. The conversion of 2-oxoglutarate (2OG) and NADPH to 2-hydroxyglutarate (2HG) and $NADP^+$ catalyzed by IDH1 R132H were monitored by measuring NADPH absorbance at 340 nm in 96-well half area clear microtiter plates (Greiner Bio-One 675001) in a final volume of 100 µL continuously over 1 hr at room temperature. The assay buffer consists of 100 mM Tris-HCl, 100 mM NaCl, 10 mM $MgCl_2$, 0.005% Tween-20, pH 7.5, 0.1 mg/mL bovine serum albumin and 0.2 mM dithiothreitol (DTT).

Percentage inhibition of IDH1 R132H was measured by diluting 1 mM compound stock in DMSO to 40 µM in the assay buffer (25 µL, 10 µM compound and 1% DMSO final), incubated with IDH1 R132H (25 µL, 30 nM final) for 12 min, followed by the addition of 2OG (25 µL, 1.5 mM final) and NADPH (25 µL, 50 µM final) to initiate the reaction. The difference in absorbance, ΔA_{340}, in the linear range of the reaction profile was converted to % residual activity with the DMSO control being the 100% residual activity reference. The % inhibition is calculated by (1 − activity with inhibitor/activity with DMSO control) × 100%. IC_{50} of the compounds were measured in a similar manner with a 3-fold serial dilution of the compounds from 10 µM to 0.169 nM (final concentration) plus a 1% DMSO control. % residual activities at 11 compound concentrations were fitted using GraphPad Prism to obtain the IC_{50} value. Data are shown in **Supplementary Figure 20**.

In silico

Building block selection

Building block selection was performed through a semi-automated process, meant to serve the group as a whole, and not just the DOSEDO library synthesis specifically. Building blocks were sourced from immediately available internal stocks of commercial building blocks, as well as newly ordered compounds (primarily from Enamine). Knime was used to filter to a diverse set of building blocks in a given desired reactivity class for manual review and selection for purchasing. The detailed workflow and explanation are shown in **Supplementary Figure 21**.

A vendor catalogue was used as input to the described workflow, with structures in SMILES format. For the purpose of the workflow illustration, the input structures were all ‘amines’ in the Enamine catalogue at the time (28976 compounds). Input structures were desalted, converted to canonical smiles, and replicate molecules

were removed (leaving 27296 compounds). A manual definition of a target functional group was then defined as substructure filter, in this example a primary amine moiety without other restrictions using the SMARTS query: [#7;h2] (retaining 15700 compounds). A set of hierarchical substructure filters were then applied to remove functionality undesirable in the specified building block set. Filters included various alkyl halides, carboxylic acids, thiols, thioethers, hydrazines, amino alcohols, and in the case of primary amines, for example, any additional secondary amine was excluded (leaving 11289 compounds). A 'dummy' library was enumerated using this filtered set of building blocks by fixing the skeleton and 'other' building block of a three-component combinatoric library as representative members. As an example the reaction SMARTS for the set of primary amines in this example was: [#7;h2:1] >>[#6]-[#7]-[#6](=O)-[#6]-1-[#6](-[#6]-[#6]-[#7]-1-[#6]-[#6]-1=[#6]-[#6](C#N)=[#6](-[#6]-2-[#6]-[#6]-2)-[#6](F)=[#6]-1)-[#6]-1=[#6]-[#6](-[#7;h1:1])=[#6]-[#6]=[#6]-1. Molecular descriptors were calculated for this dummy library, including exactMW, SLogP, HBD, HBA, fsp3, TPSA etc. Compounds were further filtered by hard cut-offs of SLogP <6.0 and exactMW <600; which in this 'primary amine' example reduced the available pool of building blocks to 7699. RDKit Diversity Picker was employed under default settings (based on Morgan 2 fingerprint diversity) to select 500 compounds from the pool. These 500 were naturally biased towards larger building blocks, so in an effort to enrich the building block selection towards those affording more desirable physicochemical properties, a second round of selection was performed. The second round focused on building blocks with SLogP <5 and exactMW <450 (when incorporated as dummy enumerated library members). Typically, as many as available of these second round building blocks were selected for inclusion in the final list of selected building blocks; however, if several hundred met the selection criteria, a set of 200 were chosen, using the initial 500 as template structures to bias away from. Final lists were written to a file and their structures were reviewed manually before ordering. In general, any labile moieties were omitted, and racemic building blocks were only included if they were deemed interesting by the reviewing chemist(s).

NOTE: Once delivered, selected building blocks were split into multiple 2D-barcoded matrix tubes to facilitate development of optimal conditions for a given building block's use in different contexts. This was achieved through a semi-automated process of Tecan-guided dissolution in a highly volatile solvent (or solvent mixture), followed by transfer and gentle solvent removal using a Genevac HT-12 centrifugal evaporator. Building blocks were stored in a Hamilton robotic tube handler as dry powder/oil in roughly 1 mg portions.

Library enumeration

Enumeration of the entire DEL (i.e., final targets associated to the merged Cy1-4 sequences) was performed using Knime Analyics Platform. To represent the assumed pharmacophore of any future putative hit, the skeletons were normalized in a number of ways, as detailed below:

1) Skeletons containing alkyl linkers to their DNA attachment site (e.g., O=C(O)CCCC(=O)N1C[C@H](c2ccc(I)cc2)C2(CNC2)C1) were truncated down to the corresponding methyl amides, marking the DNA-attachment site with a tritium atom (e.g., [3H]CC(=O)N1C[C@H](c2ccc(I)cc2)C2(CNC2)C1).
2) Skeletons linked to DNA via a carbamate linkage (e.g., OC[C@@H]1[C@H](c2ccc(Br)cc2)[C@H]2CNCCCCN12) were represented in the enumerated library as methyl ethers, again marking the DNA-attachment point with a tritium atom (e.g., [3H]COC[C@@H]1[C@H](c2ccc(Br)cc2)[C@H]2CNCCCCN21).
3) Skeletons linked to DNA via an amide linkage (e.g., O=C(O)[C@@H]1C[C@H](Oc2ccc(I)cc2)CN1) were represented in the enumerated library as methyl amides, again marking the DNA-attachment point with a tritium atom (e.g., [3H]CNC(=O)[C@@H]1C[C@H](Oc2ccc(I)cc2)CN1).
4) Skeletons containing a functionality known to completely and cleanly hydrolyze during synthesis were modified accordingly (e.g., the ethyl ester of CCOC(=O)c1cc2c(c(-c3cccc(Br)c3)n1)[C@H](CCO)NC2 becomes [H]OC(=O)c1cc2c(c(-c3cccc(Br)c3)n1)[C@H](CCOC[3H])NC2).

Amine-capping reactions were performed in silico on the pool of modified skeletons according to the below reaction definitions, followed by filtering to products derived from reactivity at the most nucleophilic N-atom in each skeleton.

N-capping reaction definitions (as SMARTS):

Sulfonylation: [#7:1].[Cl:2][S:3](=[O:4])=[O:5]>>[#7:1][S:3](=[O:4])=[O:5]

Reductive amination: [#7:1].[*:4]-[#6:2]=[O:3]>>[#7:1]-[#6;h2:2]-[*:4]

Acylation: [#7:1].[#8;h1:2]-[#6:3]=[O:4]>>[#7:1]-[#6:3]=[O:4]

Encoding information was joined to structure based on canonical smiles representations of the skeletons and building blocks used for each ‘reaction’. Products of N-capping reactions were labelled as ‘intermediate’.

Suzuki couplings were performed in silico on the pool of intermediates according to the reaction definitions shown below, again joining encoding information to enumerated structures based on canonical smiles representations of molecules used in each 'reaction'. Null reactions (i.e., no boronic acid/ester present and either with Pd-catalyst and heated, or without Pd-catalyst and not heated) were also encoded accordingly.

Suzuki coupling reaction definitions (as SMARTS):

I[*:1].[#8]-[#5](-[#8])-[*:2]>>[#6:2]-[*:1]

Br[*:1].[#8]-[#5](-[#8])-[*:2]>>[#6:2]-[*:1]

No effort was made to directly enumerate all likely side products deriving from each building block (e.g., carboxylic ester and nitrile hydrolysis and TFA-protecting group removal).

Complete enumeration afforded 3724965 unique barcodes, representing 3688975 unique compounds. Duplicate structures arose from: i) 'null Suzuki couplings', and ii) two analogous Suzuki couplings using a boronic acid and its pinacol ester (CC1(C)OB(c2ccc(S(C)(=O)=O)cc2)OC1(C)C and CS(=O)(=O)c1ccc(B(O)O)cc1).

A selection of properties was calculated for the enumerated library and plotted for a quick overview of the whole collection (Error! Reference source not found.).

Acknowledgements

We thank all members of the Novartis DNA-encoded library platform group for many fruitful discussions and their valuable input; Ritesh Tichkule, Thomas Dice, Aaron Hohos, Greg Wendel and Scott Bowes for their guidance with compound management; Phil Michaels, Max Cushner and Carmelina Rakiec for analytical support; Rishi Arora for his SPR advice, as well as Patricia Horton and Michael Romanowski for providing biotinylated IDH1 R132H; Ellen Crawford and Ayako Honda for support with outsourcing of chemical synthesis. The research was supported in part by the National Institute of General Medical Sciences (R35GM127045 awarded to S.L.S.) and by the NIBR Scholar's Program.

References

1. Gerry, C. J. & Schreiber, S. L. Recent achievements and current trajectories of diversity-oriented synthesis. *Current Opinion in Chemical Biology* **56**, 1–9 (2020).

2. Schreiber, S. L. Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science* **287**, 1964–1969 (2000).

3. Kato, N. *et al.* Diversity-oriented synthesis yields novel multistage antimalarial inhibitors. *Nature* **538**, 344–349 (2016).

4. Galloway, W. R. J. D., Isidro-Llobet, A. & Spring, D. R. Diversity-oriented synthesis as a tool for the discovery of novel biologically active small molecules. *Nature Communications* vol. 1 1–13 Preprint at https://doi.org/10.1038/ncomms1081 (2010).

5. Gerry, C. J. & Schreiber, S. L. Chemical probes and drug leads from advances in synthetic planning and methodology. *Nature Reviews Drug Discovery* **17**, 333–352 (2018).

6. Beckmann, H. S. G. *et al.* A strategy for the diversity-oriented synthesis of macrocyclic scaffolds using multidimensional coupling. *Nat Chem* **5**, 861–867 (2013).

7. Swinney, D. C. & Anthony, J. How were new medicines discovered? *Nature Reviews Drug Discovery* **10**, 507–519 (2011).

8. Swinney, D. C. Phenotypic vs. Target-based drug discovery for first-in-class medicines. *Clinical Pharmacology and Therapeutics* vol. 93 299–301 Preprint at https://doi.org/10.1038/clpt.2012.236 (2013).

9. Wagner, B. K. The resurgence of phenotypic screening in drug discovery and development. *Expert Opinion on Drug Discovery* vol. 11 121–125 Preprint at https://doi.org/10.1517/17460441.2016.1122589 (2016).

10. Yu, C. *et al.* High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines. *Nature Biotechnology* **34**, 419–423 (2016).

11. Schaefer, G. I. *et al.* Discovery of small-molecule modulators of the sonic hedgehog pathway. *J Am Chem Soc* **135**, 9675–9680 (2013).

12. Chou, D. H. C. *et al.* Synthesis of a novel suppressor of β-cell apoptosis via diversity-oriented synthesis. *ACS Medicinal Chemistry Letters* **2**, 698–702 (2011).

13. Kuo, S. Y. *et al.* Small-molecule enhancers of autophagy modulate cellular disease phenotypes suggested by human genetics. *Proc Natl Acad Sci U S A* **112**, E4281–E4287 (2015).

14. Dückert, H. *et al.* Natural product-inspired cascade synthesis yields modulators of centrosome integrity. *Nature Chemical Biology* **8**, 179–184 (2012).

15. Santagata, S. *et al.* Tight coordination of protein translation and HSF1 activation supports the anabolic malignant state. *Science (1979)* **341**, (2013).

16. Huryn, D. M. *et al.* Chemical methodology as a source of small-molecule checkpoint inhibitors and heat shock protein 70 (Hsp70) modulators. *Proc Natl Acad Sci U S A* **108**, 6757–6762 (2011).

17. Maetani, M. *et al.* Discovery of Antimalarial Azetidine-2-carbonitriles That Inhibit P. falciparum Dihydroorotate Dehydrogenase. *ACS Medicinal Chemistry Letters* **8**, 438–442 (2017).

18. Plouffe, D. M. *et al.* High-Throughput Assay and Discovery of Small Molecules that Interrupt Malaria Transmission. *Cell Host and Microbe* **19**, 114–126 (2016).

19. Aldrich, L. N. *et al.* Discovery of a small-molecule probe for V-ATPase function. *J Am Chem Soc* **137**, 5563–5568 (2015).

20. Croston, G. E. The utility of target-based discovery. *Expert Opinion on Drug Discovery* vol. 12 427–429 Preprint at https://doi.org/10.1080/17460441.2017.1308351 (2017).

21. Baell, J. B. Broad coverage of commercially available lead-like screening space with fewer than 350,000 compounds. *Journal of Chemical Information and Modeling* **53**, 39–55 (2013).

22. Brenner, S. & Lerner, R. A. Encoded combinatorial chemistry. *Proc Natl Acad Sci U S A* **89**, 5381–5383 (1992).

23. Franzini, R. M., Neri, D. & Scheuermann, J. DNA-encoded chemical libraries: Advancing beyond conventional small-molecule libraries. *Accounts of Chemical Research* **47**, 1247–1255 (2014).

24. Shi, Y., Wu, Y. R., Yu, J. Q., Zhang, W. N. & Zhuang, C. L. DNA-encoded libraries (DELs): a review of on-DNA chemistries and their output. *RSC Advances* **11**, 2359–2376 (2021).

25. Flood, D. T. *et al.* Expanding Reactivity in DNA-Encoded Library Synthesis via Reversible Binding of DNA to an Inert Quaternary Ammonium Support. *J Am Chem Soc* **141**, 9998–10006 (2019).

26. Ruff, Y. *et al.* An Amphiphilic Polymer-Supported Strategy Enables Chemical Transformations under Anhydrous Conditions for DNA-Encoded Library Synthesis. *ACS Combinatorial Science* **22**, 120–128 (2020).

27. Westphal, M. V. *et al.* Water-compatible cycloadditions of oligonucleotide-conjugated strained allenes for dna-encoded library synthesis. *J Am Chem Soc* **142**, 7776–7782 (2020).

28. Gerry, C. J., Yang, Z., Stasi, M. & Schreiber, S. L. DNA-Compatible [3 + 2] Nitrone-Olefin Cycloaddition Suitable for DEL Syntheses. *Organic Letters* **21**, 1325–1330 (2019).

29. MacConnell, A. B., Price, A. K. & Paegel, B. M. An Integrated Microfluidic Processor for DNA-Encoded Combinatorial Library Functional Screening. *ACS Combinatorial Science* **19**, 181–192 (2017).

30. Hackler, A. L., Fitzgerald, F. G., Dang, V. Q., Satz, A. L. & Paegel, B. M. Off-DNA DNA-Encoded Library Affinity Screening. *ACS Combinatorial Science* **22**, 25–34 (2020).

31. Petersen, L. K. *et al.* Screening of DNA-encoded small molecule libraries inside a living cell. *J Am Chem Soc* **143**, 2751–2756 (2021).

32. Neri, D. & Lerner, R. A. DNA-Encoded Chemical Libraries: A Selection System Based on Endowing Organic Compounds with Amplifiable Information. *Annual Review of Biochemistry* vol. 87 479–502 Preprint at https://doi.org/10.1146/annurev-biochem-062917-012550 (2018).

33. Neri, D. & Lerner, R. A. DNA-Encoded Chemical Libraries: A Selection System Based on Endowing Organic Compounds with Amplifiable Information. *Annu Rev Biochem* **87**, 479–502 (2018).

34. For additional information about availability and logistics for screening the DOSEDO library, please contact the FASTDEL team at fast.del@novartis.com.

35. Goodnow, R. A., Dumelin, C. E. & Keefe, A. D. DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nature Reviews Drug Discovery* **16**, 131–147 (2016).

36. Gerry, C. J., Wawer, M. J., Clemons, P. A. & Schreiber, S. L. DNA Barcoding a Complete Matrix of Stereoisomeric Small Molecules. *J Am Chem Soc* **141**, 10225–10235 (2019).

37. Pandya, B. A. *et al.* Practical asymmetric synthesis of β-hydroxy γ-amino acids via complimentary aldol reactions. *Tetrahedron* **67**, 6131–6137 (2011).

38. Marcaurelle, L. A. *et al.* An aldol-based build/couple/pair strategy for the synthesis of medium- and large-sized rings: Discovery of macrocyclic histone deacetylase inhibitors. *J Am Chem Soc* **132**, 16962–16976 (2010).

39. Gerard, B. *et al.* Synthesis of a stereochemically diverse library of medium-sized lactams and sultams via S NAr cycloetherification. *ACS Combinatorial Science* **13**, 365–374 (2011).

40. Lowe, J. T. *et al.* Synthesis and profiling of a diverse collection of azetidine-based scaffolds for the development of CNS-focused lead-like libraries. *Journal of Organic Chemistry* **77**, 7187–7211 (2012).

41. Martín, R., Moyano, A., Pericàs, M. A. & Riera, A. A concise enantioselective entry to the synthesis of deoxy-azasugars. *Organic Letters* **2**, 93–95 (2000).

42. Drouillat, B., Couty, F. & Marrot, J. Chirality transfer in azetidinium ylides: An enantioselective route to α-quaternary azetidines. *Synlett* **2009**, 0767–0770 (2009).

43. Dickson, P. & Kodadek, T. Chemical composition of DNA-encoded libraries, past present and future. *Organic and Biomolecular Chemistry* **17**, 4676–4688 (2019).

44. Berthold, M. R. *et al.* KNIME: The Konstanz Information Miner. *Studies in Classification, Data Analysis, and Knowledge Organization* 319–326 (2008) doi:10.1007/978-3-540-78246-9_38.

45. Clark, M. A. *et al.* Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nature Chemical Biology* **5**, 647–654 (2009).

46. Stress, C. J., Sauter, B., Schneider, L. A., Sharpe, T. & Gillingham, D. A DNA-Encoded Chemical Library Incorporating Elements of Natural Macrocycles. *Angewandte Chemie - International Edition* **58**, 9570–9574 (2019).

47. Halpin, D. R., Lee, J. A., Wrenn, S. J. & Harbury, P. B. DNA Display III. Solid-Phase Organic Synthesis on Unprotected DNA. *PLoS Biology* **2**, e175 (2004).

48. FRANCH, T. *et al.* Enzymatic Encoding Methods For Efficient Synthesis Of Large Libraries. *WO/2007/062664* (2017).

49. Vullo, D. *et al.* Carbonic anhydrase inhibitors: Inhibition of the tumor-associated isozyme IX with aromatic and heterocyclic sulfonamides. *Bioorganic and Medicinal Chemistry Letters* **13**, 1005–1009 (2003).

50. Becker, H. M. Carbonic anhydrase IX and acid transport in cancer. *British Journal of Cancer* vol. 122 157–167 Preprint at https://doi.org/10.1038/s41416-019-0642-z (2020).

51. Benej, M., Pastorekova, S. & Pastorek, J. Carbonic anhydrase IX: Regulation and role in cancer. *Sub-Cellular Biochemistry* **75**, 199–219 (2014).

52. Wichert, M. & Krall, N. Targeting carbonic anhydrase IX with small organic ligands. *Current Opinion in Chemical Biology* vol. 26 48–54 Preprint at https://doi.org/10.1016/j.cbpa.2015.02.005 (2015).

53. Cerchione, C. *et al.* IDH1/IDH2 Inhibition in Acute Myeloid Leukemia. *Frontiers in Oncology* vol. 11 345 Preprint at https://doi.org/10.3389/fonc.2021.639387 (2021).

54. Liu, X. & Gong, Y. Isocitrate dehydrogenase inhibitors in acute myeloid leukemia. *Biomarker Research* vol. 7 Preprint at https://doi.org/10.1186/s40364-019-0173-z (2019).

55. Han, S. *et al.* IDH mutation in glioma: molecular mechanisms and potential therapeutic targets. *British Journal of Cancer* vol. 122 1580–1589 Preprint at https://doi.org/10.1038/s41416-020-0814-x (2020).

56. Ma, T. *et al.* Inhibitors of Mutant Isocitrate Dehydrogenases 1 and 2 (mIDH1/2): An Update and Perspective. *Journal of Medicinal Chemistry* vol. 61 8981–9003 Preprint at https://doi.org/10.1021/acs.jmedchem.8b00159 (2018).

57. Law, J. M. *et al.* Discovery of 8-Membered Ring Sulfonamides as Inhibitors of Oncogenic Mutant Isocitrate Dehydrogenase 1. *ACS Medicinal Chemistry Letters* **7**, 944–949 (2016).

58. Fulmer, G. R. *et al.* NMR chemical shifts of trace impurities: Common laboratory solvents, organics, and gases in deuterated solvents relevant to the organometallic chemist. *Organometallics* **29**, 2176–2179 (2010).

59. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2**, 2212–2221 (2007).

60. Liu, S. *et al.* Roles of metal ions in the selective inhibition of oncogenic variants of isocitrate dehydrogenase 1. *Commun. Biol. 2021 41* **4**, 1–16 (2021).

61. Schauer, N. J. *et al.* Selective USP7 inhibition elicits cancer cell killing through a p53-dependent mechanism. *Scientific Reports 2020 10:1* **10**, 1–15 (2020).

62. Maetani, M. *et al.* Synthesis of a bicyclic azetidine with in vivo antimalarial activity enabled by stereospecific, directed C(sp 3)-H arylation. *J Am Chem Soc* **139**, 11300−11306 (2017).