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[1,4], 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[1,5], Paul A. Clemons[1], Bruno Melillo[3], C. Suk-Yee Hon[1], Johannes Ottl[6], Christoph E. Dumelin[6], Jonas V. Schaefer[6], Ann Marie E. Faust[2], Frédéric Berst[6], Stuart L. Schreiber[1,4], 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 Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA

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

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

Abstract

Diversity-oriented synthesis (DOS)is a powerful strategy to prepare molecules with underrepresented features in commercial screening collections, resulting in the elucidation of novel biological mechanisms. In parallel to the development of DOS, DNA-encoded libraries (DELs) have emerged as an effective, efficient screening strategy to identify protein binders. 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, generated using diverse skeleton architectures with varying exit vectors, derived from DOS, to achieve structural diversity beyond what is possible by varying

appendages alone. We will make this DEL 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 produced valuable probes of targets of interest, including compounds that exert their biological effects through novel mechanisms of action (nMoA)[1]. One approach to generate collections with broad molecular diversity is diversity-oriented synthesis (DOS)[2–6], which has produced numerous nMoA compounds through phenotypic screening campaigns[7–19]. Similar success with target-based screening approaches can be scarce in cases where detailed knowledge of the underlying biology is not available, where there is no knowledge of ligandable pockets or 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 the structural and especially the performance diversity of screening collections — few such collections are readily available to the scientific community from commercial vendors. Compounding the challenges to the widespread use of such collections is that these compounds generally have increased costs, with reduced or sporadic supply of key synthetic intermediates for the rapid confirmation and follow-up of identified hit compounds[21]. Consequently, access to such collections is restricted to researchers with substantial funding and/or synthetic chemistry resources, creating a barrier to new discovery in hit-generation sciences.

DNA-encoded library (DEL) technology is established as a screening tool in industry and academia to identify binders of soluble proteins via affinity-based screens. 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 the simultaneous parallel screening of entire libraries of compounds for binding to a target of interest[22]. The amount of material needed for these screens is greatly reduced compared to traditional high-throughput screening (HTS), as it can leverage the sensitivity and reliability of polymerase chain reaction (PCR) and next-generation sequencing (NGS) for data generation[23]. Various approaches that expand the applicability of DELs for synthetic chemistry compatibility have been developed successfully, thus expanding 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 DEL screening, including alternate coding schemes and novel procedures for

57 performing screens, that broaden the range of tractable targets and enhance the value of screening
58 information[29–32].

59 Several approaches are used for DEL generation. The “single pharmacophore library”[33] is the most common
60 strategy for synthesizing DELs. Here, successive steps of appendage diversification of a common skeleton
61 (pharmacophore) are encoded by enzymatic ligation that record the synthetic history for each individual
62 compound (**Fig. 1a**). The skeleton may be generated in situ or off-DNA, depending on the types of building
63 blocks used in the synthesis. Another strategy (**Fig. 1b**) does not use a fixed central skeleton, and instead
64 connects diverse building blocks in a linear fashion. An expanded strategy (**Fig. 1c**), employed in this study,
65 makes use of multiple skeletal elements that are consistent in reactivity to allow for simultaneous appendage
66 diversification using a common set of diverse appendages. Earlier work by Gerry et al.[33] highlighted a library
67 generated with this strategy; in this work, the library generation strategy is expanded with additional
68 multifunctional skeletons, additional building blocks, and diverse chemistry types. An attractive feature of this
69 strategy is that the common appendages vary greatly with respect to their relative exit vectors, allowing for
70 increased diversification.

71

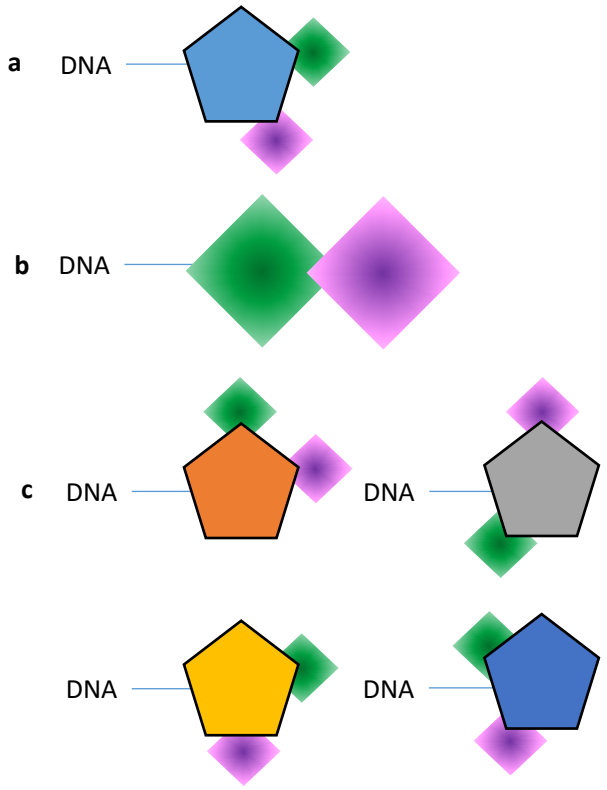

Fig. 1 | Conceptual comparison of DNA-encoded library synthesis strategies. Shapes with solid coloring represent multifunctional skeletons with defined exit vectors, and shapes with graduated coloring represent collections of building blocks with diverse pharmacophoric features. **a**, A single fixed skeleton library wherein appendage diversity results 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 DEL screening with compounds with diverse skeleton architectures that are readily derived through DOS chemistry. We call this approach diversity-oriented synthesis encoded by deoxyoligonucleotides (DOSEDO). In this work, we describe the design, synthesis, and validation experiments performed using this enhanced DEL. The DOSEDO library is used in screening campaigns to generate diverse “hits” from different protein targets, demonstrating a range of screening success. This approach is an expansion of prior work from this group, wherein a complete set of stereoisomers was encoded from a single scaffold series[33], and this approach is complimentary to approaches such as the ‘aldehyde explosion’ approach, where alternate reaction conditions are used to modify a single on-DNA

89 moiety to many different functional groups[34]. This library and resources for resynthesis are available to the
90 academic scientific community through a facilitated access approach, discussed later in this report[35].

91 **Results**

92 **Selection of components**

93 While DEL technology allows for the rapid generation of extremely large compound libraries, we prepared a
94 moderate-sized library compared to those described in some reports[36]. The reasons for this library size were
95 three-fold: 1) to ensure that reaction validation prior to library production could be performed with good
96 representation of the final library and manually controlled by a single operator with high reproducibility; 2) to
97 ensure that moderate-to-high coverage of library sequences was achievable in widely available and relatively
98 low-cost NGS experiments; and 3) to ensure a manageable cost for acquiring the raw materials to enable
99 library production and follow-up studies. We selected 61 multifunctional compounds to serve as skeletons[33,
100 37–41], all comprising secondary amines (Fmoc-, Boc- or Ns-protected) and an aryl halide (Br or I), which we
101 planned to use for on-DNA diversification (**Fig. 2**). Each skeleton bore either a carboxylic acid or primary
102 hydroxyl group, which were used as the site of DNA attachment and in some cases allowed for an alternative
103 DNA attachment point. Other DELs generated with similarly diverse skeletons have led to successful
104 screening campaigns[43]. We included as many discrete stereo- and position isomers of these skeletons as
105 was practical.

106 A branching rather than linear sequence library design was used so that the rigid central skeletons, containing
107 well-defined exit vectors, allowed for a broad distribution of appendages in three-dimensional space[44]. We
108 also focused on amine and aryl halide functional groups to enable the use of thoroughly validated DNA-
109 compatible diversification steps (acylation, sulfonylation, reductive amination, and Suzuki couplings)[24].
110 Building blocks to be appended to the core skeletons were selected through a semi-automated process
111 created in Knime Analytics Platform[45] to pick a diverse set from commercially available compounds within a
112 defined calculated property space, followed by a manual review (**Supplementary Fig. 21**). For our DNA
113 coding scheme, we used a modified version of the double-stranded DNA scheme described previously[46]. The
114 overall goals of these choices were to facilitate library construction and resynthesis and to maximize the
115 structural space encompassed by the library.

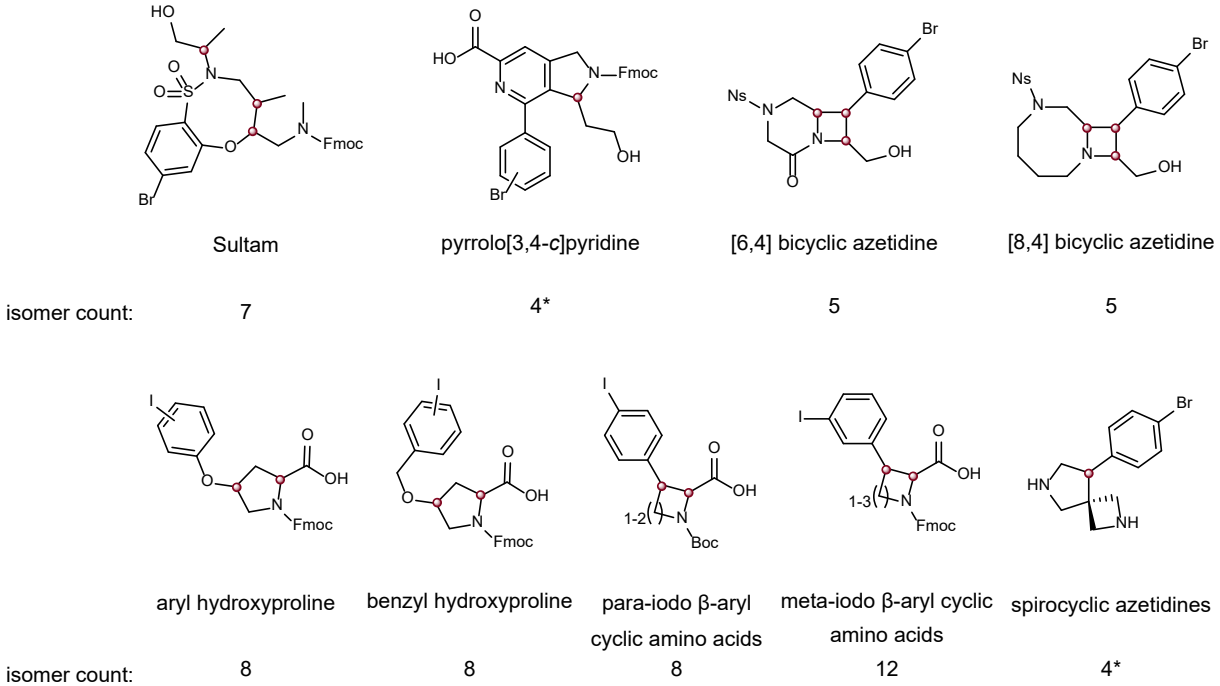

Fig. 2 | Overview of skeletons selected for inclusion in the DOSEDO library. Skeletons shown are indicative of the compounds loaded onto DNA. Isomer counts report the number of isomers (stereogenic and positional) for the indicated skeletons. Red spheres indicate two distinct isomers at each stereogenic 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.

Development of DNA linkage chemistry

Skeletons attached to DNA through their carboxylic acids were coupled under conditions modified from earlier reports[46] and optimized to minimize material usage (**Supplementary Tables 6–8**). Skeletons attached to DNA through their primary hydroxyl groups were coupled using an optimized carbamate-linkage protocol (**Supplementary Section 2.3.1.2**). 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 2.2.2),** similar to a recently reported method[47], rather than using solid-supported processes[48,49]. After linking each skeleton to DNA, conjugates were purified by preparative HPLC. The resulting HPLC eluents were concentrated and desalted by EtOH precipitation and ultrafiltration. Chromatographic data for the purified constructs following reaction validation and library production are provided in **Supplementary Table 11**.

135 **Reaction optimization**

136 With purified skeleton–DNA conjugates in hand, we optimized the diversification steps of the DEL synthesis.
137 We focused initially on reducing the residual acetate derived from preparative HPLC mobile phase, which
138 generally presented a chemoselectivity challenge in acylation reactions. Acylation conditions were optimized
139 using DNA-conjugated proline as a representative amine. Following preparative HPLC, multiple washes of
140 EtOH precipitates, ultrafiltration, or desalting on G-25 Sephadex led to excellent minimization of acetylation
141 by-products in acylation reactions. Minor adaptations of published acylation[46], reductive amination and
142 sulfonylation conditions were effective across many substrate–reactant combinations[33].

143 We used Suzuki coupling reactions to diversify aryl bromides and iodides in the library. Using general
144 coupling conditions that could be applied to all skeletons, we performed several optimization reactions
145 (**Supplementary Fig. 2**). Four palladium catalysts were screened in combination with two boronic acid
146 pinacol esters that were dissolved in either EtOH, MeCN, or DMA. A total of 288 reactions were performed,
147 leading to the selection of $PdCl_2(dppf)\cdot CH_2Cl_2$ as the preferred catalyst. This catalyst had the highest
148 conversion to desired product at all temperatures and in all co-solvents compared to matched reaction sets. In
149 addition, the $PdCl_2(dppf)\cdot CH_2Cl_2$ catalyst resulted in the least variable outcomes across a range of
150 temperature and time conditions. We also found that MeCN was the preferred solvent in combination with any
151 of the catalysts tested, though in practice, EtOH was complementary to MeCN for building block dissolution.
152 Therefore, a 1:1 mixture of EtOH and MeCN was used in later reactions for increased homogeneity.

153 **Building block validations**

154 Amine-capping building blocks were profiled using HPLC-purified and desalted **Pro-DNA** as a model system,
155 following the general scheme shown in **Fig. 3a**. At the end of each reaction, one relative volume of 5%
156 aqueous piperidine was added to quench residual electrophile, and an EtOH precipitation was performed prior
157 to LCMS analysis. A tabulated dataset can be found in **Supplementary Table 12**, and **Fig. 3a** reports the
158 output for each reaction class. Only building blocks with >70% area-under-the curve (AUC) conversion to the
159 desired product and with <10% AUC unknown species were considered for inclusion in the DOSEDO library.
160 Prior to initiating a full Suzuki-coupling validation campaign, we applied the developed conditions to a subset
161 of DNA-conjugated skeletons with uncapped or protected amines (**Supplementary Fig. 3**) using three
162 building blocks that generally had high conversion to desired products. The DNA-conjugated skeletons (as

163 free amines) gave variable outcomes based on the skeleton class or stereoisomer. We also noted generally
164 poor conversion for these couplings, mirroring the conclusions drawn for related published skeletons[33].

165 We then assessed the effect of amine-capping mode on Suzuki coupling for different skeleton classes and to
166 assess the reactivity of aryl iodides versus bromides. We performed acylations with acetic acid and benzoic
167 acid, sulfonylations with tosyl chloride, and reductive alkylations with benzaldehyde on six DNA-conjugated
168 aryl iodides and six aryl bromides (**Fig. 3b**). The resulting 48 products were subjected to Suzuki coupling
169 conditions (using a 0.1 mM solution of DNA-conjugate in pH 10 bicarbonate buffer), with phenylboronic acid
170 or its pinacol ester. During early optimization experiments, we noticed some apparent degradation of [6,4]
171 bicyclic azetidine skeletons, possibly resulting from basic hydrolysis. Therefore, we repeated the same set of
172 reactions with a pH 8 phosphate buffer. Analysis of the resulting 192 LCMS traces led to several conclusions.
173 First, the use of boronic acid or pinacol ester had little differential effect on average %AUC product formation,
174 which were 78 and 76, respectively, with average %AUC unknown DNA species of 11 in both cases.
175 Likewise, the different amine capping modes had no substantial impact on the conversion to desired products
176 (with average %AUC {standard deviation} for benzoic acid = 80{10}; acetic acid = 77{11}; *o*-toluenesulfonyl
177 chloride = 77{20}; benzaldehyde = 73{11}). A key differentiating factor was the combination of buffer pH and
178 whether the skeleton bore a bromide or iodide. **Fig. 3b** shows the change in %AUC product and unknown
179 DNA species that occurred in response to the buffer change. In general, the iodide-containing group of
180 reactants (comprising *meta*-iodo phenyl cyclic amino acids based on azetidine, pyrrolidine and piperidine, as
181 either the *cis* or *trans* diastereomers) had preferred reaction profiles in pH 10 bicarbonate buffer, with
182 considerably greater %AUC product and less %AUC unknown species present versus the otherwise identical
183 pH 8 phosphate buffered conditions. The reverse was true for the aryl bromide group. We applied the
184 developed conditions with our full set of boronic acids and esters using exemplar ArI and ArBr model systems
185 (**1** and **2** respectively, **Fig. 3c**), and found that while reaction profiles were generally effective with respect to
186 unknown byproducts, changing the buffer alone did not give a comparable validation rate between iodide and
187 bromide model systems. A higher catalyst and boronic acid/ester loading, as well as a longer reaction time,
188 led to an acceptable number of validated building blocks, with good concordance between the iodide and
189 bromide validation campaigns.

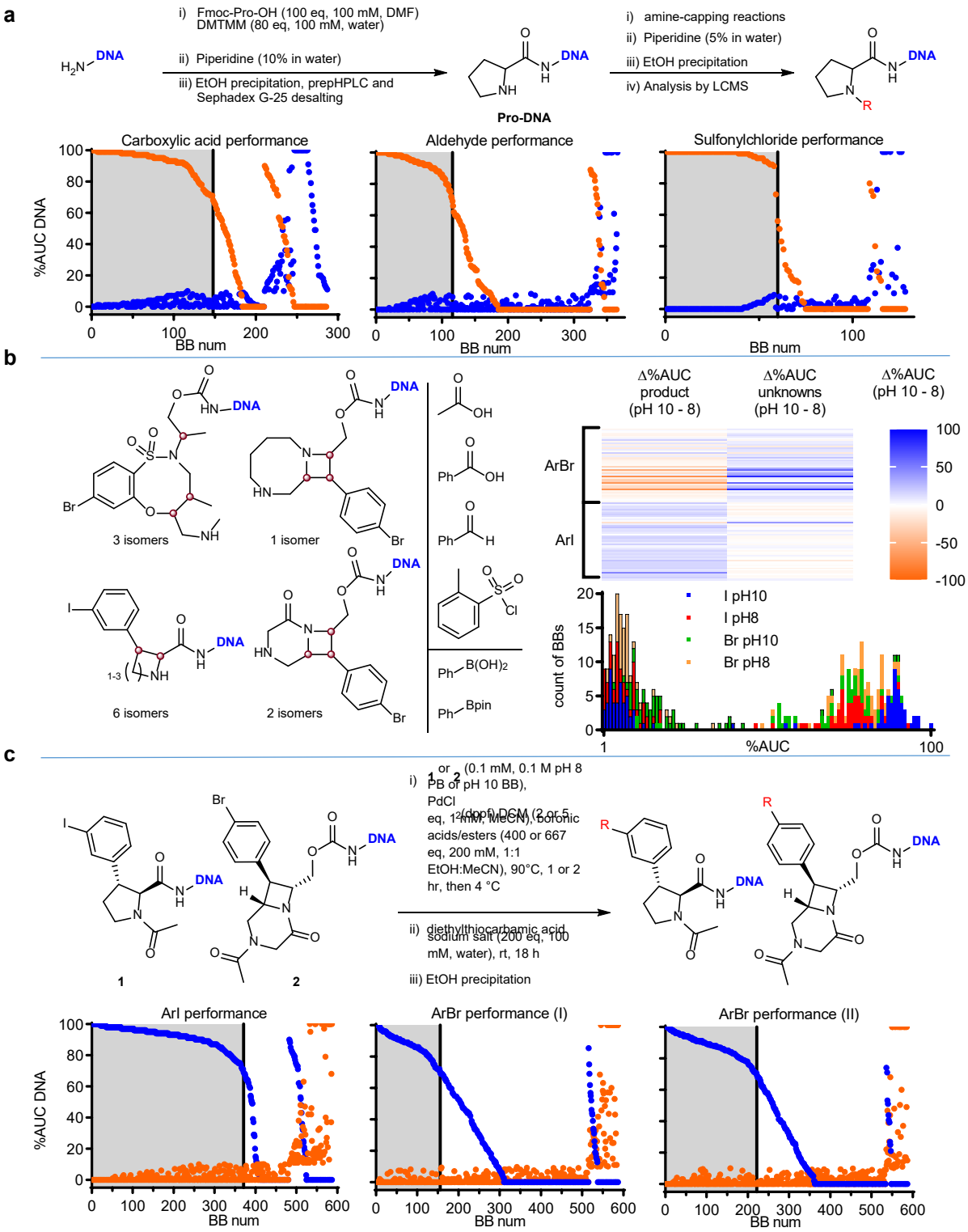

Fig. 3 | Summary of building block validations. a, Summary of amine-capping validation experiments, including a general synthetic scheme for the preparation and use of a proline-based model construct (Pro-DNA), and charts showing the performance of acylation, reductive alkylation, and sulfonylation. Blue points show %AUC (area under the curve) corresponding to desired product species; orange points show %AUC of unknown DNA species as assessed by MaxEnt1 deconvoluted mass spectra. "BB (building block) 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

197 highlighted grey region of each chart denotes building blocks that met our inclusion criteria of >70%AUC product with
198 <10%AUC unknowns. **B**, Effect of buffer on performance of Suzuki couplings comparing a subset of skeletons [with
199 variably capped amines] bearing ArI (aryliodide) versus ArBr (arylbromide). The skeleton series, amine capping, and
200 model Suzuki coupling building blocks are shown, alongside a heat map representation of the difference in %AUC product
201 and unknown species present when analogous reactions were performed at pH 10 vs. pH 8. Also displayed is a stacked
202 bar chart comparing the performance of ArBr and ArI skeletons, on aggregate, with respect to product formation (bars
203 without borders) and unknown species (bars with black borders) in different buffers. **C**, Summary of Suzuki coupling
204 building-block validations including the constructs and conditions used and a comparative analysis of performance for a
205 model ArI and ArBr (with the chart labelled "Arylbromide (II)" making use of higher catalyst and boronic acid/ester
206 loading). Additional reaction time is also noted.

207

208 Following validation of our complete collection of amine-capping reagents and boronic acids/esters, we tested
209 the extent of available diversity that could be incorporated into the final library design. In a basic sense, the
210 expected reactivity was observed; for example: aryl and vinyl sulfonyl chlorides were considerably more
211 productive than aliphatic sulfonyl chlorides (presumably due to rapid hydrolysis of the latter in aqueous
212 mixtures), and sterically more hindered aldehydes and carboxylic acids tended to be less reactive. For each
213 class of building block, we prepared a distance matrix generated from Morgan 2 fingerprints and applied
214 multidimensional scaling (MDS, by Sammon mapping, **Fig. 4**). Except for alkyl sulfonyl chlorides, we
215 generally observed that positive reaction outcomes in amine capping reactions resulted in good coverage of
216 available chemical space across all reaction modes. However, for Suzuki couplings (**Fig. 4a**–**c**), the building
217 blocks forming the core of the MDS map tend to perform better than those at the extremities, which are
218 populated with compounds comprising the most dissimilar building blocks in the collection. We conclude that
219 while our conditions allow for reasonable inclusion of building blocks, they are biased against less
220 represented and more diverse chemotypes. We found that electron-rich and/or -hindered 5-membered
221 heterocycles, as well as vinyl boronates, often constituted the bulk of poorly performing reactants. Moreover,
222 we noticed that seeking consensus positive outcomes in the ArI and ArBr Suzuki coupling validation
223 campaigns led to an erosion of fingerprint diversity in the highest priority set of building blocks (**Fig. 4a**), but
224 we believe this sacrifice of appendage diversity for consistent synthetic outcomes to be a worthwhile
225 compromise.

226

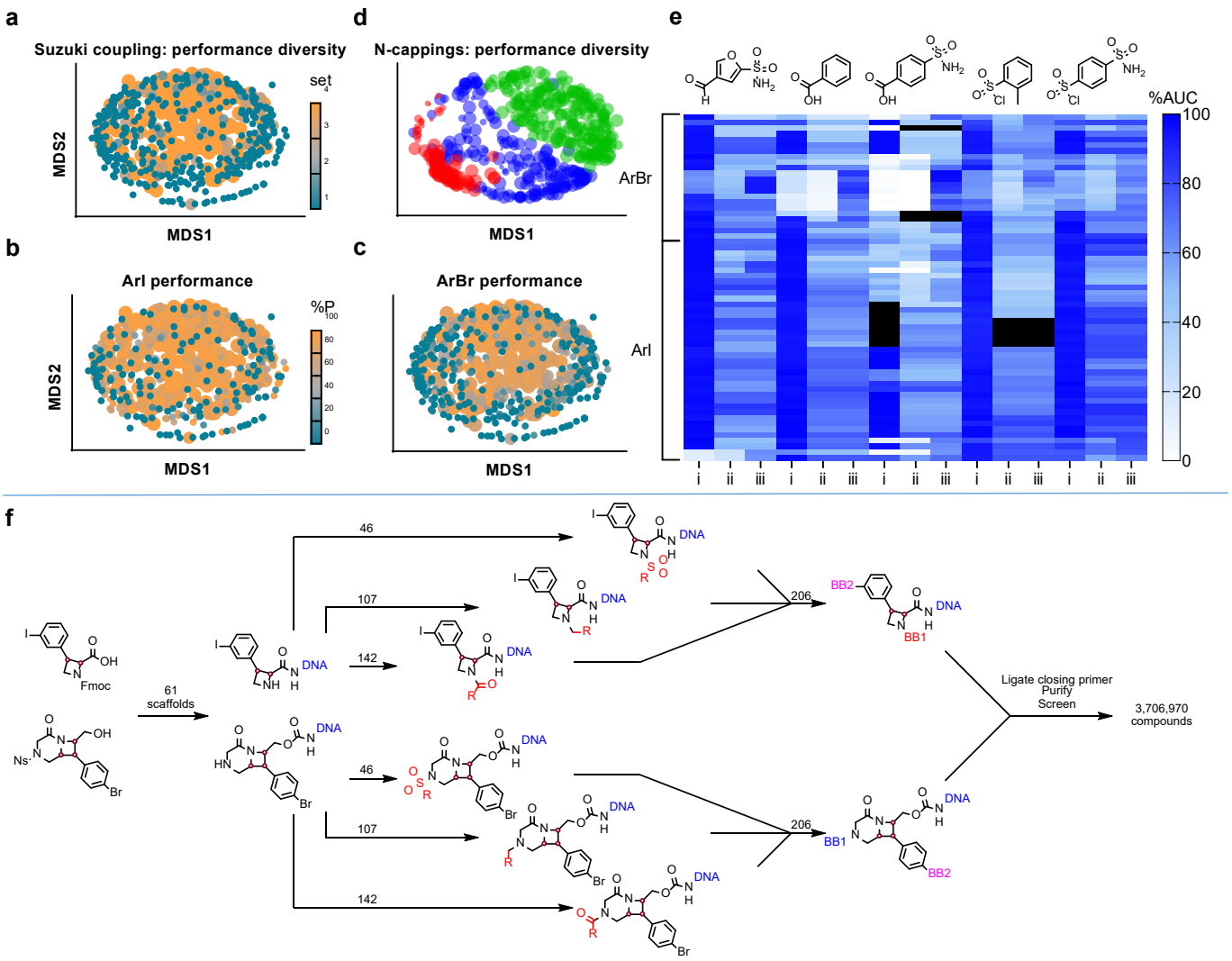

Fig. 4 | Assessment of reactivity in building block and skeleton validation experiments. A, Combined performance of ArI (aryliodide) and ArBr (arylbromide) Suzuki coupling reactions, where each common boronic acid/ester moiety was replaced with a methyl group in silico, then visualized by Sammon mapping of a Morgan2 fingerprint distance matrix to two dimensions, coloring and sizing by preference for inclusion in DNA-encoded libraries (DELs) (4 = highest), lowest preference building blocks (BBs) are layered above higher preference BBs for visualization. **b**, Performance of the same BBs represented in **a** specific to the ArI model system, with point size still indicating consensus preference category but colored by adjusted %AUC (area under the curve) 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 (with no 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 were present based on chromatography (potentially due to loss of the DNA pellet during ethanol precipitation or 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 used for each step.

248

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

258 **Skeleton validation**

259 Prior to DOSEDO library synthesis, we validated the conditions for amine capping and Suzuki coupling to
260 ensure that they were compatible with all skeletons planned for inclusion in the library. We aimed to identify
261 any critical reactivity or stability issues, such as determining whether reactions are dependent on
262 stereochemistry, that could impact the quality of the library or affect the interpretation of screening data or
263 prioritization of putative hits for resynthesis. Detailed protocols for the skeleton validation campaign are
264 described in **Supplementary Section 2.2.4**. The concentration of each purified skeleton-DNA conjugate was
265 normalized in nuclease-free water. Conjugates were subjected to five amine-capping reactions (at least one
266 of each reaction class). A sample was taken for LCMS analysis, and the residual crude amine capping
267 reaction mixture was subjected to a single Suzuki coupling (under skeleton-appropriate conditions) followed
268 by a final LCMS and analysis (**Fig. 4e**)**.** All skeleton–DNA constructs maintained their integrity throughout
269 these reaction sequences, and performance was acceptable for all skeletons to be included in the DOSEDO
270 library using optimized reaction conditions. We did note variability in the performance of specific reactions.
271 For example, β-arylated cyclic amino acids behaved consistently – with high-yielding amine cappings and
272 concurrent high-yielding Suzuki couplings – except in the case of *trans*-configured piperidines, which
273 performed very poorly in reductive alkylations. In addition, while sultam skeletons generally displayed slightly
274 reduced reactivity towards amine-capping than other skeletons, they had notably reduced reactivity across all
275 seven stereoisomers in the panel towards acylation compared to sulfonylation and reductive amination.

276 Based on these findings, we included all 61 skeletons, 295 amine-capping building blocks (142 carboxylic
277 acids, 107 aldehydes, and 46 sulfonyl chlorides), and 206 boronic acids/esters in the library synthesis. We
278 also included two null-reaction conditions in the Suzuki coupling step ('subjected to reaction conditions
279 without boronic acid/ester present', and 'not subjected to reaction conditions'), resulting in a theoretical
280 combinatoric matrix of 3,706,970 encoded compounds.

281 **Library synthesis**

282 As a key goal of this work was to make this library available for screening by the research community, we
283 sought to ensure adequate scale of production to support that effort without introducing non-validated
284 elements into our protocols. Therefore, the synthesis scale was limited by operational considerations rather
285 than by available materials. A 15% excess of each skeleton was used at the outset of synthesis (100 nmol)
286 relative to the limiting forward primer-binding site duplex. This decision ensured that variability in the
287 quantification of skeleton–headpiece conjugates was normalized during the first stage of synthesis. Skeleton
288 identity was concurrently encoded by introducing a cycle-1 tag in two–fold excess of the forward primer
289 binding site duplex, to ensure that all tandem ligations progressed to complete conversion (**Supplementary**
290 **Fig. 6**). After pooling iodides and bromides separately, we performed cycle-2 tag ligations with a slight molar
291 excess (10%) relative to the cycle 1 tags. Ligation efficiencies were assessed by analytical electrophoresis
292 using pools of 12 samples per lane, and we detected no evidence of individual ligation failure by native or
293 denaturing methods (**Supplementary Figs. 7 and 8**). Amine-capping reactions were performed, and after
294 quenching residual electrophiles with piperidine, the products were pooled and concentrated with respect to
295 DNA species and cleansed of small-molecule components by ultrafiltration and EtOH precipitation. For
296 continued synthesis, we assumed that recovery was quantitative. The pools of encoded intermediates were
297 first split, and then cycle-3 tag ligations were performed with a slight molar excess (approx. 10%) relative to
298 the cycle-2 tag. We monitored ligation efficiencies by analytical electrophoresis with 12 pooled samples per
299 lane (**Supplementary Fig. 9**). Suzuki couplings were then performed according to the developed conditions
300 (**Fig. 3c**), and products were pooled as separate Br and I-derived sub-libraries.

301 **Input analysis**

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

304 perfect and pooled a small quantity of the sub-libraries at a presumed 1:1 molar ratio at the individual
305 encoded compound level. Closing sequences were ligated, and purification was performed by continuous flow
306 electrophoresis (**Supplementary Section 2.4**). The 'yield' for DOSEDO_v1 closure over all steps of synthesis
307 ranged from 19% (fluorometric quantification) to 37% (spectrophotometric quantification) relative to the
308 limiting forward primer binding site duplex. DOSEDO_v1 was prepared for input sequence analysis according
309 to the methods outlined in **Supplementary Section 3,** with the resulting analysis shown in **Fig. 5a**. We
310 determined that pooling of sub-libraries was not equivalent; I-derived compounds were present in a 1.31–fold
311 excess, on average, according to the two negative binomial distributions of the sub-libraries. We adjusted the
312 pooling ratio for a production-scale closure that we termed DOSEDO_v2 (**Supplementary Section 2.5**). **Fig.**
313 **5b** shows the effect of input correction based on the v1 analysis. We also investigated the distribution of
314 cycle-1, -2, and -3 tag ligations, on aggregate across all compounds, for both sub-libraries (**Supplementary**
315 **Figs. 13 and 14**), and found that with very few exceptions, distributions were tight, with high concordance of
316 ligation efficiency between sub-libraries. While sequencing depth for the v1 library was not ideal (7% of
317 expected sequences were not observed), we dedicated more reads to the v2 input analysis, and as a result
318 could assess the unobserved sequences (0.03%; 1254 sequences) for potential trends relevant to library use
319 (**Fig. 5c**). This analysis suggested that in many cases, a specific Cy2 tag ligation failed from the I-derived pool
320 but not the Br-derived pool. In another example, a combination of two Cy2 tags with a specific Cy3 tag was
321 commonly not observed, with the Cy1 identity having no apparent impact on the outcome. With these two
322 exceptions, we assume that all ligations were similarly effective in both sub-libraries, and the v2 sub-library
323 pooling leads to an equivalent representation of individual compounds in the final library.

324

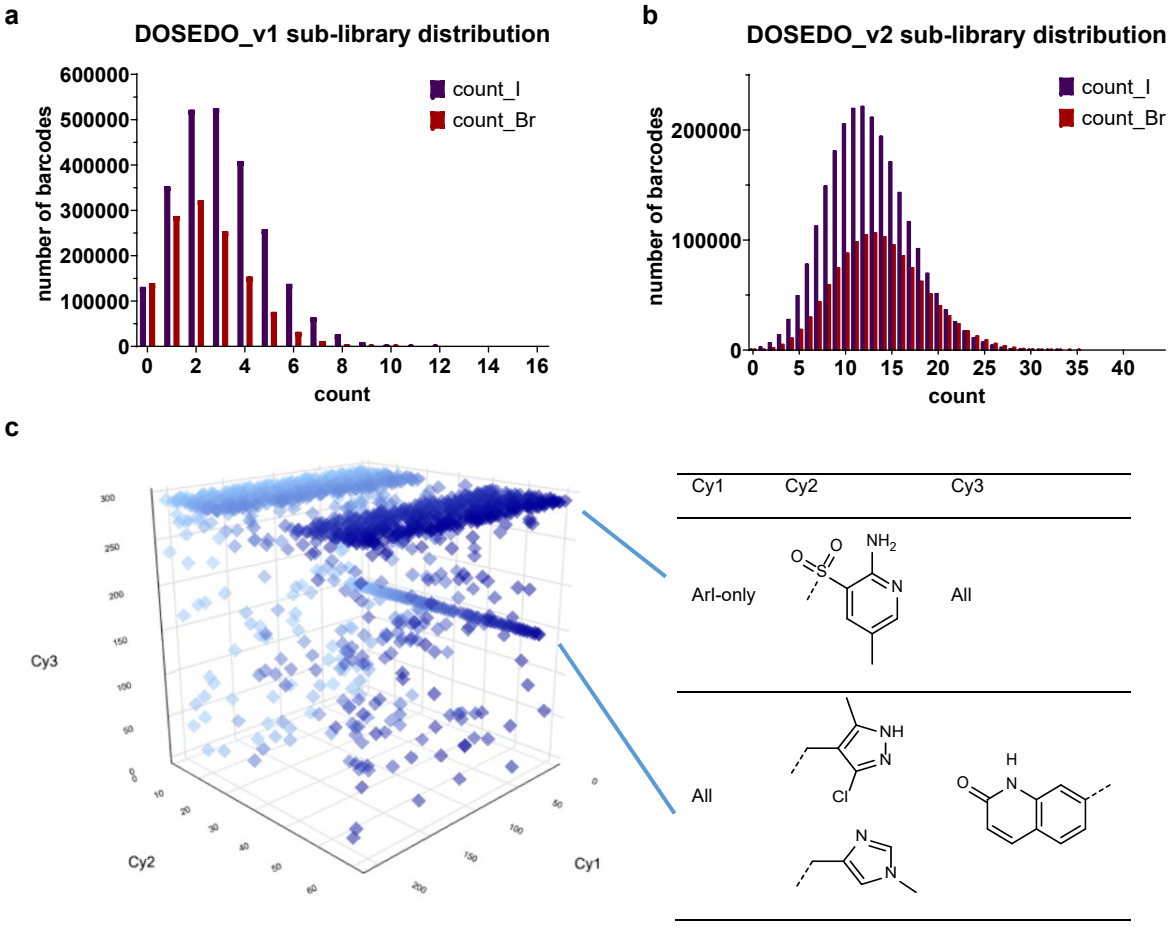

Fig. 5 | Assessment of input sequences. a, The number of barcodes observed at a given number of counts for the input DOSEDO_v1 library, showing the relative distribution of Br- and I-derived sub-libraries. **b**, As in **a** for DOSEDO_v2. **c.** unobserved sequences in DOSEDO_v2, with two sets of features 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. We selected three protein targets to screen and then 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 because the primary sulfonamide structural motif is known to bind, often with high potency and tolerance of other functionality in the ligands[50]. Therefore, we included at least one example of a primary sulfonamide as a likely binder in each reaction set. There are also potential clinical applications for selective inhibitors of CAIX[51–53]. IDH1 mutants have been implicated in several medical conditions, including acute myeloid leukemia[54,55] and

347 diffuse gliomas[56]. There are several compounds known to bind IDH1 R132H in its allosteric site[57], some of
348 which resemble compounds encoded within the DOSEDO library[58]. USP7 is a deubiquitinating enzyme (DUB)
349 that can alter the degradation rate and cellular localization of specific protein substrates, some of which are of
350 high interest in cancer progression[59]. We viewed mutant IDH1 and wild-type USP7 as more challenging proof-
351 of-concept targets than CAIX, in that we had minimal prior knowledge of the chemical features that promote
352 binding and did not deliberately bias the collection towards potent chemotypes, as was the case for CAIX. All
353 screens were performed using an estimated 1 million copies of each encoded molecule per sample and using
354 only a single cycle of panning the library over the target protein. Initial screening conditions were not
355 optimized, to more accurately reflect the potential output of an initial experiment as it might be performed by
356 an external research group. Thus, a diverse set of targets with potentially large differences in hit rates was
357 chosen to test the DOSEDO library, under conditions that would be relevant to a wide range of research
358 capabilities.

359 CAIX screening resulted in the expected enrichment of primary sulfonamide features above baseline noise
360 (**Fig. 6**). With the added variable of the skeletons, intriguing SAR was noted – specifically a set of high-
361 enrichment compounds and some low-enrichment compounds with similar structures. This finding indicated a
362 possible difference in binding based on subtle appendage changes or skeleton stereochemistry but could also
363 indicate a variable efficiency of on-DNA synthesis or bias induced by the screening protocol and NGS library
364 preparation that was not fully captured by the prospective validation experiments. We probed these
365 possibilities through off-DNA syntheses of the inferred compounds associated with high barcode enrichment.

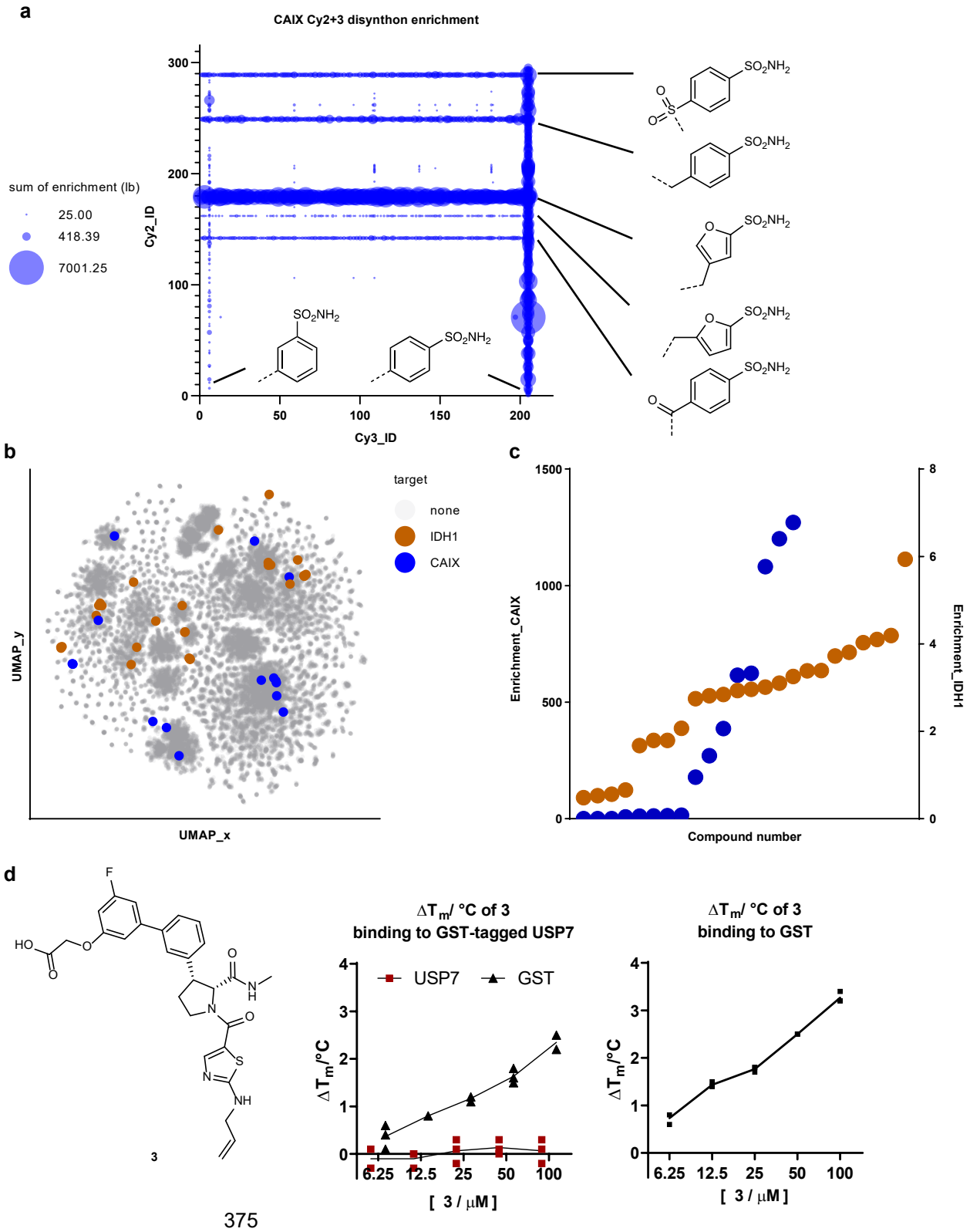

Fig. 6 | Analysis of validation screening exercises. a, Carbonic anhydrase IX (CAIX) screen output summarized by sum of lower bound (lb) calculated enrichments for all Cy2-Cy3 disynthons with all primary sulfonamide-containing building blocks included in the DOSEDO library shown (summed lower bound enrichments <25 was excluded from the

379 plot for clarity). **b**, Chemical space map generated through UMAP multi-dimensional scaling of a Tanimoto distance matrix
380 derived from Morgan fingerprints of enumerated encoded compounds. A random sample of 50,000 compounds from the
381 enumerated DOSEDO library was used as background chemical space, along with 44 selected compounds for
382 resynthesis. These compounds are colored by their binding target. **c**, Calculated enrichments of selected resynthesis
383 target compounds from CAIX- and isocitrate dehydrogenase 1 (IDH1) R132H-validation screens. **d**, Structure of **3**
384 alongside differential scanning fluorimetry (DSF) results with *N*-His-GST (glutathione S-transferase) tagged USP7 as well
385 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 ubiquitin specific
386 peptidase 7 (USP7) (T_m 44.9 °C) were observed in the melt curves of GST-tagged USP7. No thermal stabilization of
387 USP7 by **3** was observed by tracking the USP7 peak; GST stabilization by **3** was observed by tracking the GST peak. **3**
388 induces thermal stabilization of His-tagged GST in a dose-dependent manner, up to 3.3 °C at 100 μM. Data are mean ±
389 SD, n=3 technical replicates.
390

391 An IDH1 R132H screen was performed using similar conditions; this screen yielded a far lower dynamic range
392 of counts and calculated enrichment ratios than the CAIX screen. Nevertheless, we identified apparent SAR
393 across multiple series encoded within the library. To validate the library synthesis, we selected compounds for
394 resynthesis covering all reaction types, skeleton series, and levels of enrichment obtained from each screen.
395 Wherever possible, we also captured close structural analogues with markedly different enrichment values to
396 probe the variability of the generated sequencing data relative to off-DNA validation (**Fig. 6b and c**).

397 For off-DNA resynthesis, we assumed that encoded compounds were the true compounds of interest, and for
398 the sake of simplicity, synthesized those specific compounds rather than replicating the DEL synthesis
399 conditions to capture byproducts or truncates. Compounds were prepared following closely related published
400 conditions[33,38–40,60], often switching the order of steps relative to the DEL synthesis where it was deemed
401 helpful. We modified all carboxylic acids that were used for DNA attachment as the methyl amides and left
402 unmodified the corresponding hydroxyl groups that were used for carbamate DNA linkage.

403 Anticipated CA binders were principally assessed using a colorimetric enzyme-inhibition assay. While
404 selected primary sulfonamides tended to validate as active, the anticipated non-binders often inhibited CA
405 function quite potently. Given the high enrichment ratios typically observed for validated CA binders, these
406 false negative results likely indicate poor synthetic performance for specific reaction combinations in the pool
407 (**Supplementary Table 10**). Compounds prepared as putative IDH1 R132H binders were selected based on
408 lower signal-to-noise sequencing data, and as such, we expected lower correlation of enrichment to off-DNA

409 binding. This prediction held true, though 9 of the 25 compounds tested were validated as binders by SPR
410 (**Supplementary Fig. 17**). Of the four skeleton series included in the resynthesized hits, we identified at least
411 two binding series, with the sultam series being the more potent series. We then showed that only the tightest
412 binding sultam (K_D = 0.89 μM, also the most highly enriched compound from the NGS data) effectively
413 stabilized IDH1 R132H and displayed robust dose-dependent enzymatic inhibition (**Supplementary Figs. 19**
414 **and 20).** The specific stereochemical features of the identified hit aligned with the previously published IDH1
415 R132H inhibitor BRD2879[58], bearing different substituents installed through an alternate reaction scheme.
416 BRD2879 was identified using a conventional high-throughput screen, providing support for the overall
417 synthesis, encoding, screening, barcode amplification, and NGS steps of these experiments. For the USP7
418 screen, a single compound was nominated for off-DNA synthesis (**3**), primarily because the level of
419 enrichment was considerably lower relative to the matrix only control screen. USP7 was screened as an N-
420 terminally His-tagged glutathione S-transferase (GST) fusion protein, and we found that **3** selectively bound
421 the GST portion of the construct. This finding was confirmed by DSF with isolated GST as compared to the
422 GST-USP7 fusion (**Fig. 6d**), and further by SPR studies (**Supplementary Fig. 18**). The range of screening
423 outcomes for this library was in line with the distinct characteristics of the protein targets screened.

424 **Discussion**

425 In this work, we synthesized a DNA-encoded collection of compounds using a strategy based on skeleton and
426 appendage diversification, extending and expanding on earlier work[33]. In particular, the expanded number
427 and structural features of skeletons with varied exit vectors allowed us to achieve a branching library
428 synthesis scheme to cover more three-dimensional space. The library size was sufficiently large to be
429 inclusive of a large chemical space, but the relatively small size allowed us to perform robust analytics and
430 have confidence in the barcode identification of hits. We screened the library against several distinct protein
431 targets and validated that the library can be used to identify binders to targets of interest. We are confident
432 that the on-DNA synthetic chemistry, as well as the concurrent encoding steps, was satisfactory, but our study
433 highlights the impact of noise in DEL screening data and the critical importance of the protein targets used.
434 We believe that the encoded library will be of interest to the hit-discovery community because
435 underrepresented chemical features are present in the library, including novel and rigid skeleton architectures
436 and well-defined stereogenic elements. The use of defined novel skeleton structures whose synthesis can be

437 scaled up in preparation for hit validation, along with optimized downstream reactions with commercially
438 available building blocks, will facilitate the process of hit validation.

439 We aim to facilitate DOSEDO library screening for the academic scientific community through FAST-DEL[35]. It
440 is our hope that this newly described DOSEDO resource will lead to the discovery of useful tool compounds
441 for challenging targets and targets of high interest and will foster collaboration for the benefit of the broader
442 scientific community.

443 **Methods**

444 ## Analytical and purification

445 **NMR spectroscopy**

446 Nuclear magnetic resonance (NMR) spectra were recorded on Bruker AV-III HD 2-channel spectrometers
447 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
448 conventional (rt) probehead or a 5 mm BBO-F cryoprobe. Unless indicated otherwise, data were acquired at
449 25 °C using the software ICON-NMR under TopSpin program control and processed with MestReNova
450 (Version 12.0.2-20910). Spectra were referenced to residual solvent resonance[56]. For spectra recorded at
451 elevated temperatures, the same literature values of residual solvent resonances (at ambient temperature)
452 were used for referencing, neglecting temperature-dependent peak shifts. Resonance signals are reported as
453 chemical shifts (δ) in ppm with multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet or
454 unresolved, br = broad signal), couplings constant(s) in Hertz (Hz) and integral (not for ^{13}C). ^{13}C and ^{19}F NMR
455 spectra were recorded with broadband ^{1}H decoupling.

456 **Reaction monitoring by UPLC-MS**

457 Samples were resolved on Waters ACQUITY ultra-performance liquid chromatography-mass spectrometry
458 (UPLC-MS) systems equipped with C18 columns (ACQUITY UPLC BEH C18 1.7 µm, 2.1x30 mm, ACQUITY
459 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
460 signal was recorded with ACQUITY UPLC PDA detectors (210–400 nm). Light scattering signal was recorded
461 with ACQUITY UPLC ELSD or SofTA 1100 ELSD detectors. Low-resolution mass spectra (LRMS) of eluting
462 species were recorded with Waters SQD, SQD-2 or QDa detectors (electron spray ionization (positive and
463 negative modes), scan time 0.3 sec, scanning range 120-1250 Da (QDa), 120-1600 Da (SQD) or 120-2850

464 Da (SQD-2)). High-resolution mass spectra (HRMS) of eluting species were recorded on systems equipped
465 with Waters Xevo G2 Qtof or Waters Xevo G2-XS Tof detectors (electron spray ionization (positive mode),
466 scan time 0.2 sec, scanning range 100-2050 Da).

467 Solvent A1 0.1% formic acid in water
468 Solvent B1 0.1% formic acid in acetonitrile
469 Solvent A2 5 mM ammonium hydroxide in water
470 Solvent B2 5 mM ammonium hydroxide in acetonitrile
471 Solvent A3 0.05% trifluoroacetic acid in water
472 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

473

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

474

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

475

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

476

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

477 **Ion-pair chromatographic analysis of DNA-conjugates by UPLC-MS**

478 Samples were resolved on a Waters ACQUITY system equipped with a C18 column (ACQUITY
479 Oligonucleotide BEH C18 1.7 µm, 2.1x50 mm, part #186003949) kept at 50 °C. UV signal was recorded with
480 an ACQUITY TUV detector (260 nm, sampling rate 20 points/sec). Mass spectra of eluting species were
481 recorded with a Waters SQ Detector 2 connected to a ZSpray™ source (negative ion mode, scan time
482 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

483

484 **Preparative HPLC**

485 Purifications were performed on a Waters high-performance liquid chromatography (HPLC) system equipped
486 with Waters 515 pumps, a Waters 2545 binary gradient module, a Waters Acquity QDa detector, a Waters
487 2998 PDA detector and a Waters 2767 sample manager. The column (Waters XBridge Prep C18 OBD, 5 µm,
488 30 mm (inner diameter) x 50 mm) was kept at rt. Material was injected as solution in methanol/water mixtures
489 (1.5 mL) and eluted with gradients of solvent A (water) and solvent B (acetonitrile), both modified with either
490 0.1% formic acid or 5 mM ammonium hydroxide (75 mL/min flowrate). Fraction collection was triggered by UV

491 signal and mass (TIC) of eluting species. Predefined gradients were picked from a list of methods optimized
492 for target retention time as observed during reaction monitoring by UPLC-MS.

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

496 **Preparative HPLC of oligonucleotide conjugates**

497 Purifications were performed on an Agilent Infinity system equipped with an Infinity 1260 Bio Quat Pump
498 (pump system, G5611A), an Infinity 1260 HiP Bio ALS (autosampler, G1330B), an Infinity 1290 TCC
499 (thermostatted column compartment, G5667A), an Infinity 1260 DAD (diode array detector, G4212B), and an
500 Infinity 1260 Bio FC-AS (fraction collector, G5664A) under Agilent OpenLab CDS (C.01.07 SR1) software
501 control. The column (Waters XBridge BEH C18 OBD, 130 Å, 5 µm, 10x150 mm) was kept at 50 °C. Material
502 was injected as aqueous solutions (up to 100 µL per injection) and eluted with gradients of solvent A (50 mM
503 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

504

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

508 **Other methods and equipment**

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

516 **Preparative continuous flow electrophoresis**

517 A Bio-Rad Model 491 Prep Cell was used according to manufacturer guidelines to purify crude DNA-encoded
518 libraries following the ligation of oligonucleotides bearing a closing primer binding site sequence. The gel
519 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
520 to solidify over 2 hr while circulating rt water through the cooling core at approx. 100 mL/min, carefully
521 overlaying 2 mL of a 1:1 mixture of IPA and water on the top of the gel, then standing overnight at rt without
522 water circulation through the cooling core. No stacking gel was cast. After standard assembly of the Prep Cell
523 (see: http://www.bio-rad.com/webroot/web/pdf/lsr/literature/M1702925.pdf) and addition of 1x TBE buffer, the
524 crude DEL sample was loaded in Novex™ Hi-Density TBE Sample Buffer (5X, Invitrogen LC6678), and
525 electrophoresis was run at 12 W constant power. Fraction collection was performed using a Masterflex L/S
526 easy-load II (Cole Parmer, 77200-50) running at approx 1 mL/min and BioFrac™ fraction collector (Bio-Rad)
527 with a fraction time window of 2 min, i.e., approx 2 mL fractions). Total electrophoresis time was roughly 3.5
528 hr. Fractions were analyzed by E-GEL Power Snap device (Invitrogen, G8100) using commercial gels
529 (Invitrogen, E-Gel EX, 4% agarose). Product containing fractions of adequate purity were pooled, then
530 concentrated/desalted using Millipore Amicon Ultra-15/4 centrifugal filters; Ultracel 10K according to
531 manufactures guidelines, washing at least twice with nuclease-free water.

532 **Flash chromatography**

533 Automated flash chromatography was performed on Teledyne ISCO CombiFlash® systems equipped with
534 TUV and ELSD detectors using prepacked columns (RediSep® Rf, prepacked with 4 g, 12 g, 24 g, 40 g, 80 g,
535 120 g, or 330 g silica, 35-70 µm). Crude material was usually adsorbed on diatomaceous earth (Biotage

536 Isolute HM-N) and subjected to chromatography (dry-loading), eluting with gradients of ethyl acetate in
537 heptane, methanol in CH_2Cl_2, or as specified. Fractions containing homogeneous material according to
538 detection method (TUV/ELSD) and/or thin layer chromatography (TLC) were pooled, and solvents were
539 removed by rotary evaporation under reduced pressure at 30–45 °C to yield target compounds.

540 **Compound synthesis and characterization**

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

544 Preparation of NGS libraries

545 Barcodes of a DEL sample (input or eluted after screening) were PCR amplified using 3 µL i5 index primer
546 (10 µM stock in water), 3 µL i7 index primer (10 µM stock in water), 19 µL cleaned up elution samples, and 25
547 µL Invitrogen Platinum™ Hot Start PCR Master Mix (2×) (Invitrogen 13000012). The PCR method is as
548 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.
549 The PCR products were cleaned up using the ChargeSwitch PCR Clean-Up Kit, pooled in equimolar
550 amounts, and the 187 bp amplicon was gel purified using a 2% E-Gel™ EX Agarose Gels (ThermoFisher
551 Scientific G401002) and QIAquick Gel Extraction Kit (Qiagen 28704). The DNA concentration was measured
552 using the Qubit dsDNA BR assay kit and sequenced using a HiSeq SBS v4 50 cycle kit (Illumina FC-401-
553 4002) and HiSeq SR Cluster Kit v4 (Illumina GD-401-4001) on a HiSeq 2500 instrument (Illumina) in a single
554 50-base read with custom primer CTTAGCTCCCAGCGACCTGCTTCAATGTCGGATAGTG and 8-base index
555 read 32 using custom primer CTGATGGAGGTAGAAGCCGCAGTGAGCATGGT (**Supplementary Fig. 11**).

556 DEL screening

557
558 The protein targets and beads-only control (B buffer replacing the protein) were screened in duplicate using a
559 KingFisher Duo Prime (Thermo Scientific) in a 96-well deepwell plate (Thermo Scientific 95040452) at room
560 temperature. C-His tagged IDH1 R132H (Met1-Leu414) was purchased from G-Biosciences (BAN1708, 50
561 µg). N-His and GST tagged USP7 (Lys208-Glu560) were purchased from SinoBiological (11681-H20B, 100
562 µg) and reconstituted in 400 µL water. The buffers used are 'B Buffer' containing 25 mM HEPES pH 7.4, 150
563 mM NaCl, 0.05% Tween-20 (w/v), and 'S Buffer' containing 25 mM HEPES, pH 7.4, 150 mM NaCl, 0.05%
564 Tween-20 (w/v) and 0.3 mg/mL Ultrapure Salmon Sperm DNA (ThermoFisher Scientific 15632011). His-Tag

565 Dynabeads (Invitrogen 10103D, 10 μL per sample) was washed three times with B buffer before protein
566 immobilization. The proteins were diluted to 1.2 μM in B buffer (100 μL per sample) and immobilized to the
567 washed beads (1 hr, medium mix). The beads were washed once with B buffer (200 μL) and twice with S
568 buffer (200 μL) (3 min each, medium mix). The beads were transferred to the DEL library (1 million copies per
569 library member, 100 μL in S buffer) and incubated (1 hr, medium mix). The beads with protein (except beads-
570 only control) and DEL bound were washed once with S buffer (200 μL) and twice with B buffer (200 μL) (3 min
571 each, medium mix). The beads were transferred to B buffer (100 μL) and heated (90 °C, 10 min) to elute DEL
572 compounds into the supernatant. 20 μL of the elution was restriction digested by 0.1 μL StuI (NEB R0187) in
573 56.5 μL 1×SmartCutter buffer (NEB B7204S) per sample (37 °C, 1 hr) and cleaned up using the
574 ChargeSwitch PCR Clean-Up Kit (Thermo Scientific CS12000) (**Supplementary Fig. 15**).

575 ## Assay details
576

577 **Carbonic anhydrase**

578 Carbonic anhydrase II (CAII) inhibition assays were performed using a commercially available kit (BioVision,
579 #K473-100), following the manufacturer's protocol. For each assayed DOSEDO compound, 11 concentrations
580 (half-log dilutions from 10 μM to 100 pM) were tested, with three technical replicates performed on each of
581 two separate 384-well μClear® medium-binding, flat-bottom polystyrene microplates (Greiner Bio-One,
582 #82051-294) (n=6 for each compound concentration); for **53** only, six technical replicates (n=6) were
583 performed on the same plate for each compound concentration. Each reaction well contained 3 μL of
584 compound (prepared in 10% DMSO, 90% kit buffer), 25.5 μL of enzyme/buffer mix (1.4 μL enzyme, 24.1 μL
585 kit buffer), and 1.5 μL of substrate, for a total volume of 30 μL. Absorbance at 405 nm was read on a
586 Molecular Devices SpectraMax i3x. In addition to the experimental conditions, a DMSO control and a no-
587 enzyme control were each performed with three technical replicates on each plate, except for the experiment
588 with **53**, where six technical replicates for each control were performed on the same plate. The average
589 change in absorbance for the no-enzyme control was subtracted from the change in absorbance over the
590 same time interval for each experimental condition and the DMSO control on the same plate. For each
591 concentration of each compound, relative activity was calculated separately for each replicate based on the
592 average change in absorbance for the DMSO control on the same plate. Dose-response curves were fitted
593 and IC50 values and 95% confidence intervals were calculated using GraphPad Prism v9.2.0. Curves were
594 fitted using GraphPad Prism's log(inhibitor) vs. response – variable slope nonlinear regression equation, with

595 the top asymptote constrained to 100. Outliers were detected and removed via GraphPad Prism's ROUT
596 method, with a Q-value of 1%. Error bars (only shown if greater than the height of the symbol) represent 95%
597 confidence intervals. See **Supplementary Fig. 16**.

598 **Recombinant homodimeric IDH1 R132H production**

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

616 **IDH1 R132H and USP7 surface plasmon resonance**

617 IDH1 R132H: SPR experiments were performed on a Biacore T200 instrument (GE Healthcare) at 25 °C
618 using a running buffer containing 50 mM HEPES pH 7.5, 50 mM KCl, 0.005% tween-20, 1 mM DTT, 2%
619 DMSO. The compounds were 2-fold serial diluted from 20 µM to 78 nM with a 2% DMSO only control.
620 Biotinylated homodimeric IDH1 (RU ~3500) was immobilized to the streptavidin sensor chip (Cytiva
621 BR100531) preconditioned with 1M NaCl and 40 mM NaOH (60 s ×3 cycles, 100 µL/min) and running buffer
622 (60 s, 100 µL/min). The run has a startup run of 12 cycles using the running buffer, flow rate of 50 µL/min,
623 contact time of 60 s and dissociation time of 150—500 s depending on the compound, with two negative

624 controls using the running buffer between each compound. The syringe was washed with 50% DMSO
625 between injections. Data was analyzed using the affinity mode of the Biacore T200 Evaluation Software and
626 is presented in **Supplementary Fig. 17**.

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

639 **IDH1 R132H differential scanning fluorimetry**

640 Thermal shift experiments were carried out using a QuantStudio™ 7 Flex system (Applied Biosystems) in a
641 MicroAmp Optical 384-well plate (Applied Biosystems 4309849)[57]. Each well contained 10 μL of 5 μM protein
642 with 10× Sypro Orange in 50 mM Tris-HCl, pH 7.4. Various concentrations of compounds (6.25 – 100 μM, 1%
643 DMSO final) were mixed with the protein for thermal stabilization studies. The samples in triplicates were
644 subjected to temperature increases from 25 °C to 95 °C at 0.02 °C s^{-1}, with optical filters of x1-m3
645 corresponding to excitation 470 nm and emission 586 nm respectively. Protein Thermal Shift Software v1.3
646 (Applied Biosystems) was used to determine the melting temperature, T_m, from the derivative of the melt
647 curve. Data are shown in **Supplementary Fig. 19.**

648 **IDH1 R132H enzymatic inhibition**

649 The inhibitory activity of the off-DNA compounds against IDH1 R132H were assessed by absorbance assays
650 similarly to reported procedures[58], using IDH1 R132H purified in-house as described above. The conversion
651 of 2-oxoglutarate (2OG) and NADPH to 2-hydroxyglutarate (2HG) and $NADP^+$ catalyzed by IDH1 R132H were
652 monitored by measuring NADPH absorbance at 340 nm in 96-well half area clear microtiter plates (Greiner

653 Bio-One 675001) in a final volume of 100 μL continuously over 1 hr at room temperature. The assay buffer
654 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
655 serum albumin and 0.2 mM dithiothreitol (DTT).

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

665 ## In silico

666 ### Building block selection

667 Building block selection was performed through a semi-automated process. Building blocks were sourced
668 from available internal stocks of commercial building blocks and newly ordered compounds (primarily from
669 Enamine). Knime was used to filter to a diverse set of building blocks in each desired reactivity class for
670 manual review and selection for purchasing. The detailed workflow and explanation are shown in
671 **Supplementary Fig. 21**.

672 A vendor catalogue was used as input to the described workflow, with structures in SMILES format. In the
673 workflow illustration, the input structures were all 'amines' in the Enamine catalogue at the time (28,976
674 compounds). Input structures were desalted, converted to canonical SMILES, and replicate molecules were
675 removed (27,296 compounds). A manual definition of a target functional group was defined as substructure
676 filter, in this example, a primary amine moiety without other restrictions using the SMARTS query: [#7;h2]
677 (15,700 compounds). A set of hierarchical substructure filters were applied to remove functionality that would
678 be undesirable in the specified building block set. Filters included various alkyl halides, carboxylic acids,
679 thiols, thioethers, hydrazines, amino alcohols, and in the case of primary amines, for example, any additional
680 secondary amine was excluded (11,289 compounds). A 'dummy' library was enumerated using this filtered

681 set of building blocks by fixing the skeleton and 'other' building block of a three-component combinatoric
682 library as representative members. As an example, the reaction SMARTS for the set of primary amines was:
683 [#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)-
684 [#6](F)=[#6]-1)-[#6]-1=[#6]-[#6](-[#7;h1:1])=[#6]-[#6]=[#6]-1. Molecular descriptors were calculated for this
685 dummy library, including exactMW, SLogP, HBD, HBA, fsp3, and TPSA. Compounds were further filtered by
686 hard cut-offs of SLogP <6.0 and exactMW <600, which in this 'primary amine' example reduced the available
687 pool of building blocks to 7,699. RDKit Diversity Picker was employed under default settings (based on
688 Morgan 2 fingerprint diversity) to select 500 compounds from the pool. These 500 were naturally biased
689 towards larger building blocks, so to enrich the building block selection towards those with more desirable
690 physicochemical properties, a second round of selection was performed. The second round focused on
691 building blocks with SLogP <5 and exactMW <450 (when incorporated as dummy enumerated library
692 members). Typically, as many as available of the second-round building blocks were selected for inclusion in
693 the final list of selected building blocks; however, if several hundred met the selection criteria, a set of 200
694 was chosen, using the initial 500 as template structures to bias away from. Final lists were written to a file,
695 and their structures were reviewed manually before ordering. In general, labile moieties were omitted, and
696 racemic building blocks were only included if they were considered interesting by the reviewing chemist(s).
697 Once delivered, selected building blocks were split into multiple 2D-barcoded matrix tubes to facilitate
698 development of optimal conditions for a given building block's use in different contexts. This organization was
699 achieved through a semi-automated process of Tecan-guided dissolution in a highly volatile solvent (or
700 solvent mixture), followed by transfer and gentle solvent removal using a Genevac HT-12 centrifugal
701 evaporator. Building blocks were stored in a Hamilton robotic tube handler as dry powder/oil in roughly 1 mg
702 portions.

703 **Library enumeration**

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

707 1) Skeletons containing alkyl linkers to their DNA attachment site (e.g.,
708 O=C(O)CCCC(=O)N1C[C@H](c2ccc(I)cc2)C2(CNC2)C1) were truncated down to the corresponding

709 methyl amides, marking the DNA attachment site with a tritium atom (e.g.,
710 [3H]CC(=O)N1C[C@H](c2ccc(I)cc2)C2(CNC2)C1).
711 2) Skeletons linked to DNA via a carbamate linkage (e.g.,
712 OC[C@@H]1[C@H](c2ccc(Br)cc2)[C@H]2CNCCCCN12) were represented in the enumerated
713 library as methyl ethers, again marking the DNA attachment point with a tritium atom (e.g.,
714 [3H]COC[C@@H]1[C@H](c2ccc(Br)cc2)[C@H]2CNCCCCN21).
715 3) Skeletons linked to DNA via an amide linkage (e.g., O=C(O)[C@@H]1C[C@H](Oc2ccc(I)cc2)CN1)
716 were represented in the enumerated library as methyl amides, again marking the DNA attachment
717 point with a tritium atom (e.g., [3H]CNC(=O)[C@@H]1C[C@H](Oc2ccc(I)cc2)CN1).
718 4) Skeletons containing a functionality known to completely and cleanly hydrolyze during synthesis were
719 modified accordingly (e.g., the ethyl ester of CCOC(=O)c1cc2c(c(-c3cccc(Br)c3)n1)[C@H](CCO)NC2
720 becomes [H]OC(=O)c1cc2c(c(-c3cccc(Br)c3)n1)[C@H](CCOC[3H])NC2).

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

724 N-capping reaction definitions (as SMARTS):

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

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

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

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

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

734 Suzuki coupling reaction definitions (as SMARTS):

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

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

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

739 Complete enumeration afforded 3,724,965 unique barcodes, representing 3,688,975 unique compounds.
740 Duplicate structures arose from: i) 'null Suzuki couplings', and ii) two analogous Suzuki couplings using a
741 boronic acid and its pinacol ester (CC1(C)OB(c2ccc(S(C)(=O)=O)cc2)OC1(C)C and
742 CS(=O)(=O)c1ccc(B(O)O)cc1).

743 A selection of properties was calculated for the enumerated library and plotted for a quick overview of the
744 whole collection (**Supplementary Fig. 22**).

745 **Data Availability**

746 The datasets generated and/or analyzed during the current study are available from the corresponding author
747 upon reasonable request.

748 **Acknowledgements**

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

756 **Author Contributions**

757 L.H., B.K.H., C.J.G., H.L.O., S.L., A.G.R., S.B., J.E.B., P.A.C., B.M., C.S-Y.H, F.B., S.L.S., F.J.Z, and K.B.
758 contributed to the study design. L.H., J.W.M., C.J.G., and G.X. performed skeleton/scaffold synthesis. L.H.,
759 J.W.M., B.K.H., and C.J.G. performed on-DNA synthesis. M.J.R.R., J.R.T., J.M.K., Z.Y.T., P.K., and B.I.C.T.
760 performed hit synthesis. S.L., A.G.R., and K.S.L. performed screens. M.V.W., H.L.O., J.P., J.C., and C.M.G.
761 contributed to data generation. M.V.W. and K.S.L. performed data analysis. L.H., J.R.T., B.K.H., S.L., A.G.R.,
762 S.B., N.J.S., and C.W.C. performed data analysis and interpretation. J.O., C.E.D., J.V.S., and A.M.E.F. are

763 responsible for technical enablement of the work described. L.H., S.L., A.G.R., and K.S.L. drafted the
764 manuscript and Supplemental Information. L.H., S.B., J.V.S., F.B., and A.M.E.F. revised the manuscript and
765 Supplementary Information.

766 **Competing Interests**

767 C.J.G. is a shareholder in and employee of Kisbee Therapeutics. C.W.C. is an advisor to Anagenex. P.A.C. is
768 an advisor to nference, Inc., Pfizer, Inc., and Belharra Therapeutics. B.M. is a scientific advisor to Magnet
769 Biomedicine. S.L.S. is a shareholder and serves on the Board of Directors of Jnana Therapeutics and Kojin
770 Therapeutics; is a shareholder and advises Kisbee Therapeutics, Belharra Therapeutics, Magnet
771 Biomedicine, Exo Therapeutics, and Eikonizo Therapeutics; advises Vividian Therapeutics, Eisai Co., Ltd.,
772 Ono Pharma Foundation, F-Prime Capital Partners, and the Genomics Institute of the Novartis Research
773 Foundation; and is a Novartis Faculty Scholar.

774 **References**

775 1. Gerry, C. J. & Schreiber, S. L. Recent achievements and current trajectories of diversity-
776 oriented synthesis. *Curr Opin Chem Biol* **56**, 1–9 (2020).

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

779 3. Kato, N. *et al.* Diversity-oriented synthesis yields novel multistage antimalarial inhibitors. *Nature*
780 **538**, 344–349 (2016).

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

784 5. Gerry, C. J. & Schreiber, S. L. Chemical probes and drug leads from advances in synthetic
785 planning and methodology. *Nat Rev Drug Discov* **17**, 333–352 (2018).

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

788 7. Swinney, D. C. & Anthony, J. How were new medicines discovered? *Nat Rev Drug Discov* **10**,
789 507–519 (2011).

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

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

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

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

800 12. Chou, D. H. C. *et al.* Synthesis of a novel suppressor of β-cell apoptosis via diversity-oriented
801 synthesis. *ACS Med Chem Lett* **2**, 698–702 (2011).

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

804 14. Dückert, H. *et al.* Natural product-inspired cascade synthesis yields modulators of centrosome
805 integrity. *Nat Chem Biol* **8**, 179–184 (2012).

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

808 16. Huryn, D. M. *et al.* Chemical methodology as a source of small-molecule checkpoint inhibitors
809 and heat shock protein 70 (Hsp70) modulators. *Proc Natl Acad Sci U S A* **108**, 6757–6762
810 (2011).

811 17. Maetani, M. *et al.* Discovery of Antimalarial Azetidine-2-carbonitriles That Inhibit P. falciparum
812 Dihydroorotate Dehydrogenase. *ACS Med Chem Lett* **8**, 438–442 (2017).

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

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

817 20. Croston, G. E. The utility of target-based discovery. *Expert Opinion on Drug Discovery* vol. 12
818 427–429 Preprint at https://doi.org/10.1080/17460441.2017.1308351 (2017).

819 21. Baell, J. B. Broad coverage of commercially available lead-like screening space with fewer than
820 350,000 compounds. *J Chem Inf Model* **53**, 39–55 (2013).

821 22. Brenner, S. & Lerner, R. A. Encoded combinatorial chemistry. *Proc Natl Acad Sci U S A* **89**,
822 5381–5383 (1992).

823 23. Franzini, R. M., Neri, D. & Scheuermann, J. DNA-encoded chemical libraries: Advancing
824 beyond conventional small-molecule libraries. *Acc Chem Res* **47**, 1247–1255 (2014).

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

827 25. Flood, D. T. *et al.* Expanding Reactivity in DNA-Encoded Library Synthesis via Reversible
828 Binding of DNA to an Inert Quaternary Ammonium Support. *J Am Chem Soc* **141**, 9998–10006
829 (2019).

830 26. Ruff, Y. *et al.* An Amphiphilic Polymer-Supported Strategy Enables Chemical Transformations
831 under Anhydrous Conditions for DNA-Encoded Library Synthesis. *ACS Comb Sci* **22**, 120–128
832 (2020).

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

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

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

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

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

843 32. Neri, D. & Lerner, R. A. DNA-Encoded Chemical Libraries: A Selection System Based on
844 Endowing Organic Compounds with Amplifiable Information. *Annu Rev Biochem* **87**, 479–502
845 (2018).

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

848 34. Paciaroni, N. G., Ndungu, J. M. & Kodadek, T. Solid-phase synthesis of DNA-encoded libraries:
849 Via an 'aldehyde explosion' strategy. *Chemical Communications* **56**, 4656–4659 (2020).

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

852 36. Goodnow, R. A., Dumelin, C. E. & Keefe, A. D. DNA-encoded chemistry: enabling the deeper
853 sampling of chemical space. *Nat Rev Drug Discov* **16**, 131–147 (2016).

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

856 38. Marcaurelle, L. A. *et al.* An aldol-based build/couple/pair strategy for the synthesis of medium-
857 and large-sized rings: Discovery of macrocyclic histone deacetylase inhibitors. *J Am Chem Soc*
858 **132**, 16962–16976 (2010).

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

861 40. Lowe, J. T. *et al.* Synthesis and profiling of a diverse collection of azetidine-based scaffolds for
862 the development of CNS-focused lead-like libraries. *Journal of Organic Chemistry* **77**, 7187–
863 7211 (2012).

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

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

868 43. Deng, H. *et al.* Discovery, SAR, and X-ray Binding Mode Study of BCATm Inhibitors from a
869 Novel DNA-Encoded Library. *ACS Med Chem Lett* **6**, 919–924 (2015).

870 44. Dickson, P. & Kodadek, T. Chemical composition of DNA-encoded libraries, past present and
871 future. *Org Biomol Chem* **17**, 4676–4688 (2019).

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

874 46. Clark, M. A. *et al.* Design, synthesis and selection of DNA-encoded small-molecule libraries.
875 *Nat Chem Biol* **5**, 647–654 (2009).

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

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

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

883 50. Vullo, D. *et al.* Carbonic anhydrase inhibitors: Inhibition of the tumor-associated isozyme IX
884 with aromatic and heterocyclic sulfonamides. *Bioorg Med Chem Lett* **13**, 1005–1009 (2003).

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

887 52. Benej, M., Pastorekova, S. & Pastorek, J. Carbonic anhydrase IX: Regulation and role in
888 cancer. *Subcell Biochem* **75**, 199–219 (2014).

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

892 54. Cerchione, C. *et al.* IDH1/IDH2 Inhibition in Acute Myeloid Leukemia. *Frontiers in Oncology* vol.
893 11 345 Preprint at https://doi.org/10.3389/fonc.2021.639387 (2021).

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

896 56. Han, S. *et al.* IDH mutation in glioma: molecular mechanisms and potential therapeutic targets.
897 *British Journal of Cancer* vol. 122 1580–1589 Preprint at https://doi.org/10.1038/s41416-020-
898 0814-x (2020).

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

902 58. Law, J. M. *et al.* Discovery of 8-Membered Ring Sulfonamides as Inhibitors of Oncogenic
903 Mutant Isocitrate Dehydrogenase 1. *ACS Med Chem Lett* **7**, 944–949 (2016).

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

906 60. Maetani, M. *et al.* Synthesis of a bicyclic azetidine with in vivo antimalarial activity enabled by
907 stereospecific, directed C(sp3)-H arylation. *J Am Chem Soc* **139**, 11300−11306 (2017).

908