ABSTRACT: We previously reported interaction determination using unpurified proteins (IDUP), a method to selectively amplify DNA sequences encoding ligand:target pairs from a mixture of DNA-linked small molecules and unpurified protein targets in cell lysates. In this study, we applied IDUP to libraries of DNA-encoded bioactive compounds and DNA-tagged human kinases to identify ligand:protein binding partners out of 32,096 possible combinations in a single solution-phase library × library experiment. The results recapitulated known small molecule:protein interactions and also revealed that ethacrynic acid is a novel ligand and inhibitor of MAP2K6 kinase. Ethacrynic acid inhibits MAP2K6 in part through alkylation of a nonconserved cysteine residue. This work validates the ability of IDUP to discover ligands for proteins of biomedical relevance.

Discovering small molecules that specifically modulate the activity of proteins of biomedical interest remains a crucial activity in the life sciences. DNA-encoded chemical libraries have emerged as a rich source of such small molecules as biological probes and leads for therapeutics development, and are typically evaluated for binding to individual protein targets by affinity enrichment using immobilized, purified protein targets. The effectiveness of these methods is limited by artifactual enrichment of library members that bind the solid support or nonphysiologically relevant forms of target proteins, incomplete knowledge of the biological context necessary for the target to adopt its relevant form, and the inability to simultaneously explore interactions with multiple proteins of interest. Few methods of screening DNA-encoded libraries, such as selections on cell-surface displayed proteins, parallel selections under varied conditions, or the use of photo-cross-linking probes to perform selections on unmodified proteins, have begun to address these limitations.

To address some of these drawbacks, our group developed interaction determination using unpurified proteins (IDUP), a solution-phase method for in vitro identification of protein-binding ligands from combinations of ligands and unpurified proteins in a single experiment. In IDUP, binding of a DNA-tagged protein and a DNA-encoded ligand stabilizes the hybridization of short (6- to 8-nt) complementary regions at the 3’ ends of their associated DNA barcodes (Figure 1). The resulting short DNA duplex undergoes primer extension by a DNA polymerase, encoding both the small molecule and the protein it binds on a single oligonucleotide. Only these extended oligonucleotides with primer sequences from both libraries can undergo PCR amplification. Subsequent high-throughput DNA sequencing reveals the identities of all ligand:protein partner pairs. IDUP enables simultaneous evaluation of small molecule and protein libraries in a single experiment in cell lysate and leverages the efficiency of DNA-encoded libraries and high-throughput DNA sequencing. We previously validated IDUP’s ability to enrich DNA sequences encoding known binding pairs from an excess of mock barcodes not conjugated to small molecules or target proteins. In this study, we conducted a discovery-oriented IDUP experiment using libraries of DNA-barcoded proteins and small molecules to identify novel binding pairs.

The majority of our library of protein targets consisted of human kinases, many of which are of biomedical interest. The ability of IDUP to assess the selectivity of small molecules could, in principle, distinguish promiscuous and selective kinase ligands. To assemble this protein library, we identified a set of 289 cytosolic and soluble human kinase ORFs included in pDONR21 vectors for Gateway cloning (Harvard PlasmID Repository). The ORFs were subcloned into an N-terminal SNAP-tag fusion protein plasmid by Gateway cloning to enable DNA barcoding. The resulting pDEST-SNAP-kinase vectors were transiently transfected into HEK293T cells. The corresponding cell lysates were individually treated with 31-nt

Figure 1. Overview of IDUP. DNA-barcoded small molecules and proteins are combined in cell lysate. Primer extension, PCR and DNA sequencing reveal the identity of protein:ligand pairs.
benzylguanine-linked oligonucleotides (DNA-BG) that each contained a unique 6-nt barcode and the common 3’ 8-nt hybridization region required for IDUP. DNA-BG barcodes were validated computationally and in a mock IDUP experiment to remove sequences that were subject to positive or negative PCR bias. Unlabeled SNAP protein was quenched using SNAP-Cell Block (New England Biolabs) and the lysates were pooled to obtain 236 SNAP-tagged, DNA-barcoded target proteins. In parallel, an aliquot of pooled lysates was quenched with SNAP-Cell Block, then combined with pooled DNA-BG, to generate a non-DNA-tagged negative control sample.

We constructed a library of DNA-linked compounds with annotated bioactivity, hypothesizing that those compounds may have more favorable solubility, stability, or protein-binding properties. We identified a candidate set of 500 carboxylic acid-containing compounds and 250 aliphatic primary amines within the databases of the Broad Institute and Harvard’s Department of Chemistry and Chemical Biology. By inspection, we removed compounds containing functional groups that would interfere with DNA conjugation and compounds from overrepresented structural classes (e.g., quinolones, cephalosporins, or penicillins), arriving at a set of 177 carboxylate- and 87 amine-containing compounds.

Each small molecule’s 43-nt DNA barcode included an internal 7-nt barcode and a constant 3’ 8-nt hybridization region complementary to that of the protein library. Carboxylic acids were coupled to a 3’-amine-linked DNA oligonucleotide using DMT-MM or EDC and purified by HPLC, resulting in 97 DNA-linked compounds. Amine-containing small molecules were coupled to 3’-thiol-functionalized DNA using a heterobifunctional cross-linker containing both a maleimide and an NHS ester (Thermo Scientific Pierce), yielding an additional 39 DNA-linked compounds. We included the Bak peptide as a positive control, as we previously detected its binding to Bcl-xL protein ($K_D = 480 \text{ nM}$) in the IDUP format. The final library contained an equimolar mixture of each of the 136 DNA-linked compounds. The molecules span a range of chemical properties, including molecular weight (123 to 2222 Da, mean = 357 Da), lipophilicity (calculated cLogP of −9.8 to +7.4, mean = 1.8), and number of H-bond donors (1 to 40, mean = 4.7) (Figure S1).

We combined 2 pmol of the DNA-linked small-molecule library with the DNA-tagged protein library and performed IDUP primer extension (see Supporting Information p. S30). Extended products, encoding protein:ligand pairs, were selectively amplified by PCR and analyzed by high-throughput DNA sequencing. The abundance of each barcode out of the 32,096 possible ligand:protein combinations was compared to its frequency in the control IDUP experiment to define an enrichment value for each possible combination (Figure 2). Across seven technical replicates, the most significantly enriched sequence corresponded to Bcl-xL:Bak binding (205-fold average enrichment), the only interaction tested that we previously validated in an IDUP experiment. In addition, we observed high enrichment (89.5-fold) of the barcodes corresponding to PKI peptide (a cAMP-dependent kinase inhibitor) and PRKX (a cAMP-dependent kinase). Two different barcodes corresponding to variants of the BET inhibitor JQ1 enriched for binding to BET family proteins BRD2 (6.8- or 6.9-fold, $K_D = 128 \text{ nM}$) and BRD3 (15.2- or 19.4-fold, $K_D = 60 \text{ nM}$). Although DNA-encoded library selections can suffer from interference between the DNA and binding of a library member to a protein, this possibility did not preclude enrichment of these ligand:protein partners. We did not observe a strong correlation between DNA-free binding affinity and IDUP enrichment, potentially due to factors such as the DNA tag affecting IDUP enrichment positively or negatively.

Next, we evaluated if protein:small-molecule combinations encoded by other enriched amplitcons corresponded to bona fide protein:ligand pairs. We tested 11 interactions encoded by enriched barcodes in either kinase activity or binding assays using the corresponding non-DNA tagged ligands (Table S3). Using Z′-LYTE assays (Invitrogen), we measured the inhibition of PRKX by PKI ($IC_{50} = 52 \text{ nM}$) and GSK3α by bisindolylmaleimide X (bisX) (4.7-fold IDUP enrichment, $IC_{50} = 115 \text{ nM}$). Finally, we discovered that ethacrynic acid (EA) inhibits MAP2K6 (4.7-fold IDUP enrichment, Z′-LYTE $IC_{50} = 4.5 \mu M$).

EA is an FDA-approved loop diuretic that inhibits the NKCC symporter and has not been previously reported to inhibit any kinases. EA contains a Michael acceptor that reacts with glutathione and EA derivatives have been previously used as covalent bromodomain inhibitors. We investigated whether it inhibits MAP2K6 by forming a covalent adduct with the protein. Nonelectrophilic analogs of EA (dihydro-ethacrynic acid and tienilic acid) exhibited >30-fold weaker inhibition of MAP2K6 (Figure 3). Incubating MAP2K6 with EA yielded a +303 adduct in the intact protein mass spectrum, consistent with covalent modification by EA (Figure S3B). Sequential treatment of MAP2K6 with EA and then iodoacetamide (IAA), a cysteine alkylating agent, resulted in modification by IAA at...
only five of MAP2K6's six cysteines (Figure S3D), suggesting that EA modifies MAP2K6 at a cysteine residue. We analyzed EA-treated MAP2K6 by tryptic digest and MALDI-TOF and observed only one peptide (residues 37–49) with a modification consistent with EA adduct formation (Figure S4A). This peptide contains a single cysteine residue, and fragmentation of this peptide by tandem mass spectrometry confirmed that Cys38 was the site of covalent modification (Figure S4C).

To better assess the mechanism of EA inhibition, we incubated with EA a constitutively active MAP2K6 mutant containing phosphomimetic S207E and T211E mutations (MAP2K6EE), dialyzed the protein into EA-free buffer, and observed 9-fold apparent loss of kinase activity in the Z'-LYTE assay. Preincubation of EA with a C38A point mutant of MAP2K6EE resulted in a smaller loss in inhibition potency of ~3.3-fold (Figure 4). Together, these results suggest that covalent modification of MAP2K6 by EA at Cys38 is partially, but not solely, responsible for kinase inhibition.

A member of the MAP2K family, MAP2K6 activates p38 MAP kinase in response to environmental stresses.15 Previous cheminformatic and proteome-wide studies implicated Cys128 (in the Gatekeeper region) or Cys196 (adjacent to the DFG motif) as more accessible or reactive toward small-molecule electrophiles.16 In contrast, Cys38 is located within a nonactive site region with poorly understood function17 and is not conserved among other MAP2Ks (Figure S5). We confirmed that EA has higher affinity for MAP2K6 than other MAP2Ks (Figure 5). These trends are consistent with the results of the IDUP library experiment, suggesting that IDUP can reveal cysteine-reactive covalent ligands21 such as ethacrynic acid. Such a library of electrophiles could be used as covalent fragment leads against the proteome, analogous to current mass spectrometry-based activity based protein profiling methods.22 Given the vast size of small molecule:protein interaction space that could be explored by integrating these existing resources, we anticipate that DNA-encoded library methods such as IDUP will be explored by integrating these existing resources, we anticipate that DNA-encoded library methods such as IDUP will find additional use in the rapid, unbiased discovery of small molecules capable of binding target proteins.

**ASSOCIATED CONTENT**

* Supporting Information
The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b04880.

Supplemental figures and experimental methods (PDF)

**AUTHOR INFORMATION**

* Corresponding Author
  *drlu@fas.harvard.edu

**ORCID**

Alix I. Chan: 0000-0002-2698-4031
Lynn M. McGregor: 0000-0003-4310-2384
David R. Liu: 0000-0002-9943-7557
Communication

Present Addresses
1Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA 94158, United States.
2Harvard Medical School, Boston, MA 02115, United States.

Notes
The authors declare the following competing financial interest(s): D.R.L. is a founder of Ensemble Therapeutics, a company that uses DNA-templated synthesis for drug discovery.

ACKNOWLEDGMENTS

JQ1-CO₂H and JQ1-NH₂ were gifts from Dr. James Bradner. We thank Dr. Sunia Trauger for assistance with mass spectrometry. This work was supported by NIH R01 EB022376 (formerly R01 GM065400), NIH R35 GM118062, DARPA N66001-14-2-4053, and the Howard Hughes Medical Institute.

REFERENCES


