Discovery of a Covalent Kinase Inhibitor from a DNA-Encoded Small-Molecule Library × Protein Library Selection

Alix I. Chan[ID], **Lynn M. McGregor**[ID,†], **Tara Jain**[‡], and **David R. Liu**[*,ID]

The Broad Institute of Harvard and MIT, Howard Hughes Medical Institute, and the Department of Chemistry and Chemical Biology, Harvard University, 75 Ames Street, Cambridge, Massachusetts 02142, United States

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,[1] and are typically evaluated for binding to individual protein targets by affinity enrichment using immobilized, purified protein targets.[2,3] 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

*Corresponding Author. drliu@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
†**Present Addresses**
Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA 94158, United States.
‡Harvard Medical School, Boston, MA 02115, United States.

proteins,[4] parallel selections under varied conditions,[3] or the use of photo-crosslinking probes to perform selections on unmodified proteins,[5] 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.[6] 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[7] in cell lysate[6] 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.[6] 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 pDONR221 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 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 nM)[8] in the IDUP format.[6] 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 cLog$P$ 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.[6] In addition, we observed high enrichment (89.5-fold) of the barcodes corresponding to PKI peptide (a cAMP-dependent kinase inhibitor[9]) and PRKX (a cAMP-dependent kinase[10]). 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 nM[11]) and BRD3 (15.2- or 19.4-fold, K_D = 60 nM[11]). 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 amplicons 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 nM) and GSK3α by bisindolylmaleimide X (bisX)[12] (4.7-fold IDUP enrichment, IC_{50} = 115 nM). Finally, we discovered that ethacrynic acid (EA) inhibits MAP2K6 (4.7-fold IDUP enrichment, Z′-LYTE IC_{50} = 4.5 μM).

EA is an FDA-approved loop diuretic that inhibits the NKCC symporter[13] and has not been previously reported to inhibit any kinases. EA contains a Michael acceptor that reacts with glutathione[13] and EA derivatives have been previously used as covalent bromodomain

inhibitors.[14] 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 function[17] 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 × library experiment, suggesting that IDUP can illuminate a compound's selectivity even within a protein family.

The one-pot experiment described here enriched barcodes encoding seven different binding interactions out of 32 096 possible combinations. Only one (Bak:Bcl-xL) had been previously validated in the IDUP format, showing that IDUP is a generalizable method for detecting binding interactions in complex mixtures. The inhibition of MAP2K6 by EA by targeting Cys38 was also unexpected. Such selective probes could be used to investigate the role of MAP2K6 in redox sensing, development, and cancer.[18] EA's inhibition of MAP2K6, a cellular target unrelated to current uses of EA to treat edema or as a probe for GST function[13] or Wnt signaling,[19] suggests that further studies of EA and related compounds as biological probes might be fruitful. The discovery of this novel ligand interaction site in MAP2K6 through IDUP highlights the potential of unbiased binding assays to reveal probes with unanticipated inhibition mechanisms.

To our knowledge, this work represents the first library × library DNA-encoded selection for the identification of previously unknown ligand:protein binding pairs. The approach described here, if applied to a genome-scale donor vector library,[20] could concurrently

evaluate binding of DNA-encoded libraries of small molecules to many human proteins. DNA-encoded small-molecule libraries containing thousands to billions of chemically diverse members have been reported,[1] and only limited work has used DNA-encoded libraries to reveal cysteine-reactive covalent ligands[21] 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 × library methods such as IDUP will find additional use in the rapid, unbiased discovery of small molecules capable of binding target proteins.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. (a) Zimmermann G, Neri D. Drug Discovery Today. 2016; 21:1828–1834. [PubMed: 27477486] (b) Goodnow RA Jr, Dumelin CE, Keefe AD. Nat. Rev. Drug Discovery. 2016; 16:131–147. [PubMed: 27932801] (c) Salamon H, Skopic MK, Jung K, Bugain O, Brunschweiger A. ACS Chem. Biol. 2016; 11:296–307. [PubMed: 26820267]

2. (a) Maianti JP, McFedries A, Foda ZH, Kleiner RE, Du XQ, Leissring MA, Tang W-J, Charron MJ, Seeliger MA, Saghatelian A, Liu DR. Nature. 2014; 511:94–98. [PubMed: 24847884] (b) Wichert M, Krall N, Decurtins W, Franzini RM, Pretto F, Schneider P, Neri D, Scheuermann J. Nat. Chem. 2015; 7:241–249. [PubMed: 25698334]

3. Soutter HH, Centrella P, Clark MA, Cuozzo JW, Dumelin CE, Guie MA, Habeshian S, Keefe AD, Kennedy KM, Sigel EA, Troast DM, Zhang Y, Ferguson AD, Davies G, Stead ER, Breed J, Madhavapeddi P, Read JA. Proc. Natl. Acad. Sci. U. S. A. 2016; 113:E7880–E7889. [PubMed: 27864515]

4. Wu Z, Graybill TL, Zeng X, Platchek M, Zhang J, Bodmer VQ, Wisnoski DD, Deng J, Coppo FT, Yao G, Tamburino A, Scavello G, Franklin GJ, Mataruse S, Bedard KL, Ding Y, Chai J, Summerfield J, Centrella PA, Messer JA, Pope AJ, Israel DI. ACS Comb. Sci. 2015; 17:722–731. [PubMed: 26562224]

5. (a) Zhao P, Chen Z, Li Y, Sun D, Gao Y, Huang Y, Li X. Angew. Chem. Int. Ed. 2014; 53:10056–10059.(b) Li G, Zheng W, Chen Z, Zhou Y, Liu Y, Yang J, Huang Y, Li X. Chem. Sci. 2015; 6:7097–7104. [PubMed: 28757982]

6. McGregor LM, Jain T, Liu DR. J. Am. Chem. Soc. 2014; 136:3264–3270. [PubMed: 24495225]

7. Wu CY, Wang DH, Wang XB, Dixon SM, Meng LP, Ahadi S, Enter DH, Chen CY, Kato J, Leon LJ, Ramirez LM, Maeda Y, Reis CF, Ribeiro B, Weems B, Kung HJ, Lam KS. ACS Comb. Sci. 2016; 18:320–329. [PubMed: 27053324]

8. Petros AM, Nettesheim DG, Wang Y, Olejniczak ET, Meadows RP, Mack J, Swift K, Matayoshi ED, Zhang H, Thompson CB, Fesik SW. Protein Sci. 2000; 9:2528–2534. [PubMed: 11206074]

9. Glass DB, Lundquist LJ, Katz BM, Walsh DA. J. Biol. Chem. 1989; 264:14579–14584. [PubMed: 2760075]

10. Li X, Li HP, Amsler K, Hyink D, Wilson PD, Burrow CR. Proc. Natl. Acad. Sci. U. S. A. 2002; 99:9260–92655. [PubMed: 12082174]

11. Filippakopoulos P, Qi J, Picaud S, Shen Y, Smith WB, Fedorov O, Morse EM, Keates T, Hickman TT, Felletar I, Philpott M, Munro S, McKeown MR, Wang Y, Christie AL, West N, Cameron MJ, Schwartz B, Heightman TD, La Thangue N, French CA, Wiest O, Kung AL, Knapp S, Bradner JE. Nature. 2010; 468:1067–1073. [PubMed: 20871596]

12. Brehmer D, Godl K, Zech B, Wissing J, Daub H. Mol. Cell. Proteomics. 2004; 3:490–500. [PubMed: 14769951]

13. (a) Molnar J, Somberg JC. Am. J. Ther. 2009; 16:86–92. [PubMed: 19142159] (b) Maron, BA., Rocco, TP. Goodman & Gilman's: The Pharmacological Basis of Therapeutics. 12. Brunton, LL.Chabner, BA., Knollmann, BC., editors. McGraw-Hill; New York, NY: 2011.

14. Daguer J-P, Zambaldo C, Abegg D, Barluenga S, Tallant C, Müller S, Adibekian A, Winssinger N. Angew. Chem. Int. Ed. 2015; 54:6057–6061.

15. Raman M, Chen W, Cobb MH. Oncogene. 2007; 26:3100–3112. [PubMed: 17496909]

16. (a) Liu Q, Sabnis Y, Zhao Z, Zhang T, Buhrlage SJ, Jones LH, Gray NS. Chem. Biol. 2013; 20:146–159. [PubMed: 23438744] (b) Weerapana E, Wang C, Simon GM, Richter F, Khare S, Dillon MB, Bachovchin DA, Mowen K, Baker D, Cravatt BF. Nature. 2010; 468:790–795. [PubMed: 21085121]

17. Matsumoto T, Kinoshita T, Matsuzaka H, Nakai R, Kirii Y, Yokota K, Tada T. J. Biochem. 2012; 151:541–549. [PubMed: 22383536]

18. (a) Diao Y, Liu W, Wong CC, Wang X, Lee K, Cheung PY, Pan L, Xu T, Han J, Yates JR 3rd, Zhang M, Wu Z. Proc. Natl. Acad. Sci. U. S. A. 2010; 107:20974–20979. [PubMed: 21078955] (b) Warr N, Siggers P, Carre GA, Wells S, Greenfield A. Biol. Reprod. 2016; 94:103. [PubMed: 27009039] (c) Mitra AP, Almal AA, George B, Fry DW, Lenehan PF, Pagliarulo V, Cote RJ, Datar RH, Worzel WP. BMC Cancer. 2006; 6doi: 10.1186/1471-2407-6-159(d) Parray AA, Baba RA, Bhat HF, Wani L, Mokhdomi TA, Mushtaq U, Bhat SS, Kirmani D, Kuchay S, Wani MM, Khanday FA. Cancer Invest. 2014; 32:416–422. [PubMed: 25019214]

19. Lu D, Liu JX, Endo T, Zhou H, Yao S, Willert K, Schmidt-Wolf IG, Kipps TJ, Carson DA. PLoS One. 2009; 4:e8294. [PubMed: 20011538]

20. Yang X, Boehm JS, Yang X, Salehi-Ashtiani K, Hao T, Shen Y, Lubonja R, Thomas SR, Alkan O, Bhimdi T, Green TM, Johannessen CM, Silver SJ, Nguyen C, Murray RR, Hieronymus H, Balcha D, Fan C, Lin C, Ghamsari L, Vidal M, Hahn WC, Hill DE, Root DE. Nat. Methods. 2011; 8:659–661. [PubMed: 21706014]

21. Zambaldo C, Daguer JP, Saarbach J, Barluenga S, Winssinger N. MedChemComm. 2016; 7:1340–1351.

22. Backus KM, Correia BE, Lum KM, Forli S, Horning BD, González-Páez GE, Chatterjee S, Lanning BR, Teijaro JR, Olson AJ, Wolan DW, Cravatt BF. Nature. 2016; 534:570–574. [PubMed: 27309814]
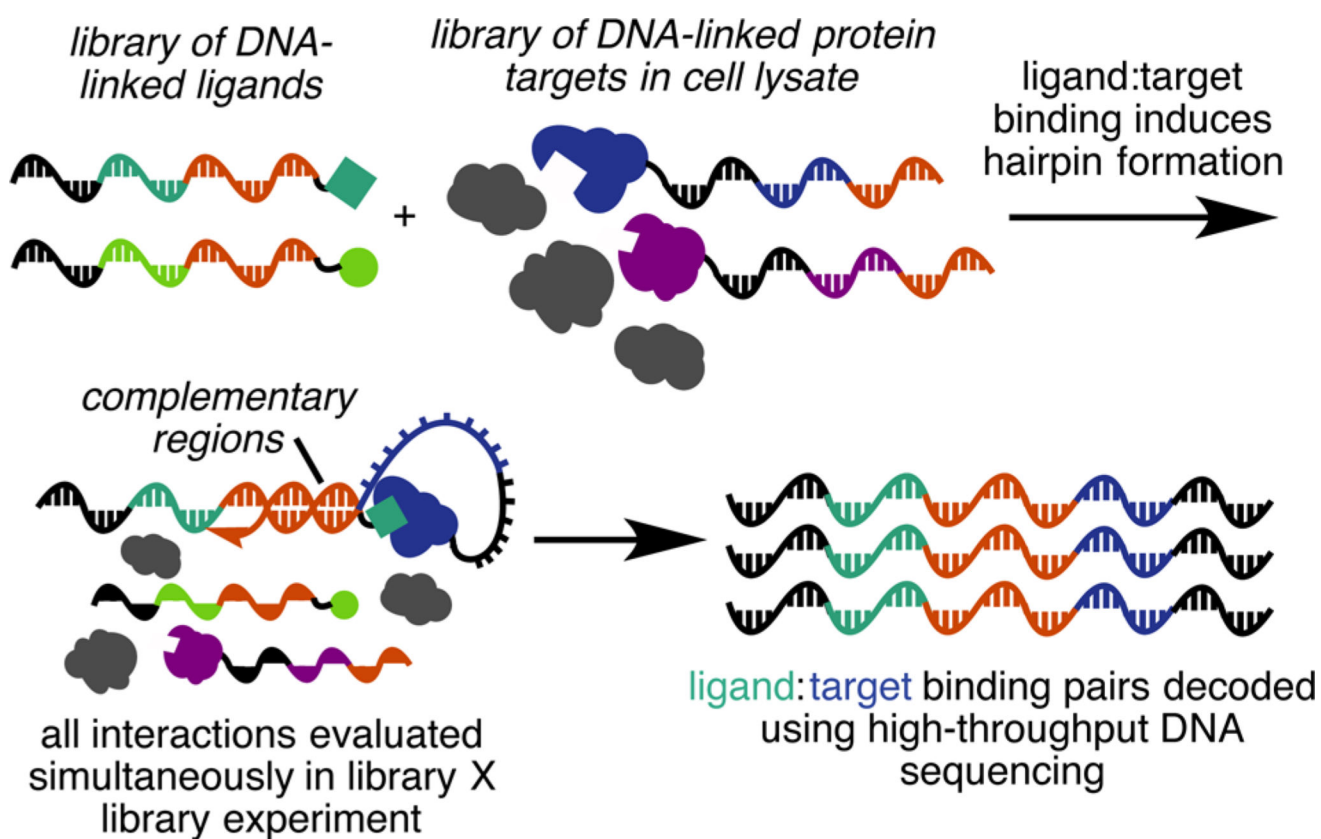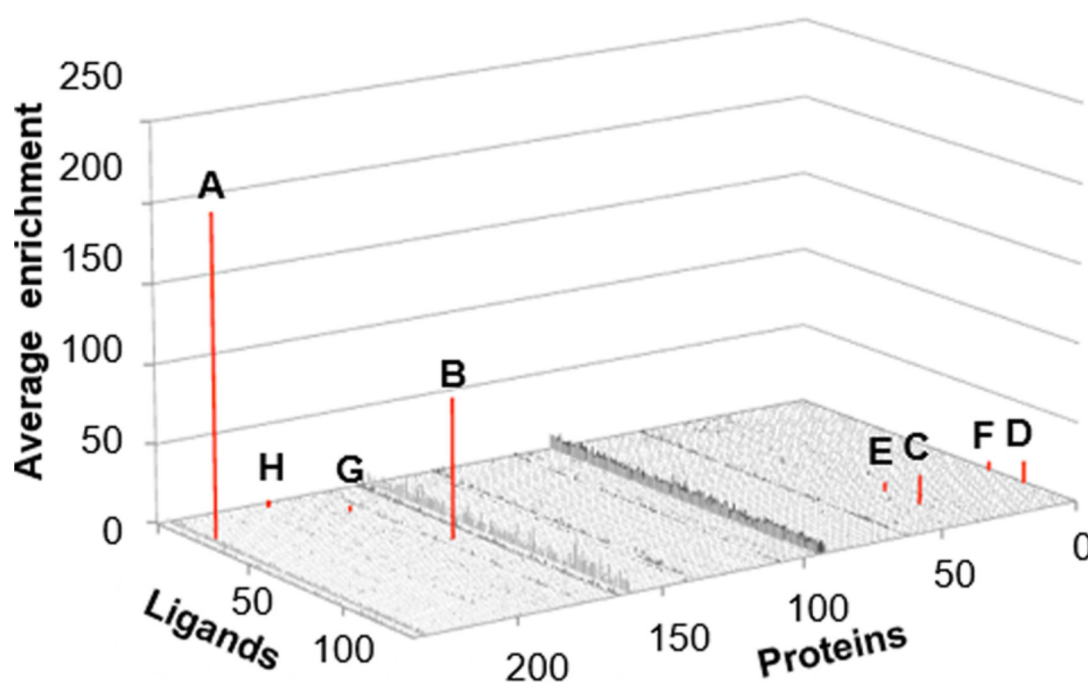
Figure 1.
Overview of IDUP. DNA-barcoded small molecules and proteins are combined in cell lysate.
Primer extension, PCR and DNA sequencing reveal the identity of protein:ligand pairs.

	Interaction	Avg. enrichment (7 replicates)	Affinity (off DNA)
A	Bcl-xL + Bak	205.4	480 nM (K_D)[8]
B	PRKX + PKI	89.5	52 nM (IC_{50})
C	BRD3 + JQ1-NH_2	19.4	60 nM (K_D)[11]
D	BRD2 + JQ1-NH_2	15.2	128 nM (K_D) [11]
E	BRD3 + JQ1-CO_2H	6.9	60 nM (K_D)[11]
F	BRD2 + JQ1-CO_2H	6.8	128 nM (K_D)[11]
G	GSK3α + bisindolylmaleimide X	4.7	115 nM (IC_{50})
H	MAP2K6 + ethacrynic acid	4.7	4.5 µM (IC_{50})

Figure 2.
Results of seven replicate IDUP experiments of combined protein and small molecule libraries. Eight DNA barcodes corresponding to bona fide interactions (Bak:Bcl-xL, PKI:PRKX, JQ1:BRD2, JQ1:BRD3, bisX:GSK3α, EA:MAP2K6) enriched (red) out of 32 096 possibilities. K_D and IC_{50} values are for compounds and proteins not linked to DNA. Nonspecific amplification across some protein barcodes may arise from poor expression of those targets (see Figure S2).
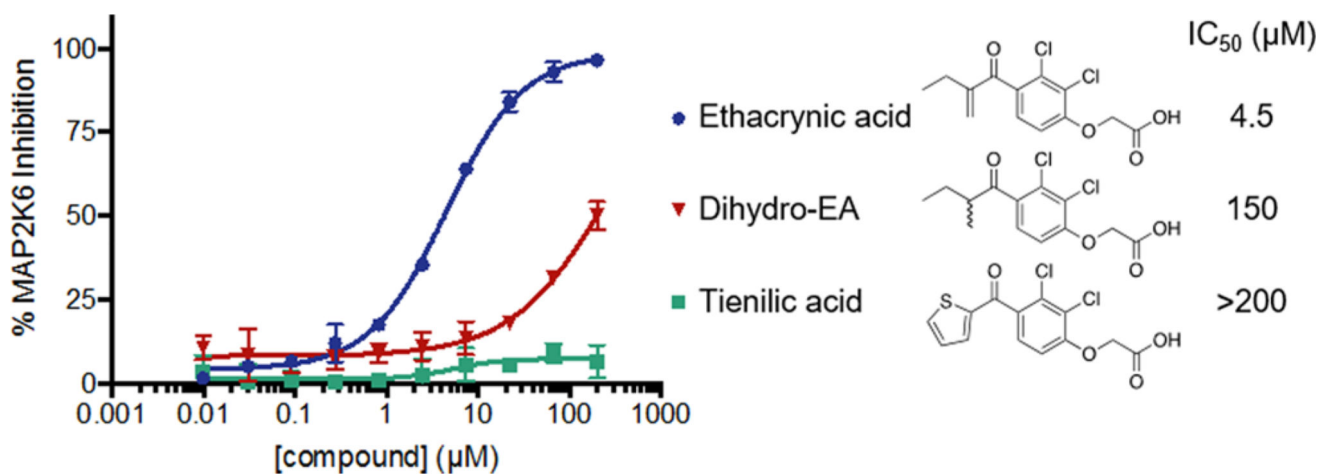
Figure 3.
Inhibition of MAP2K6 by EA is dependent on the Michael acceptor. The nonelectrophilic analogs shown are >30-fold less potent.
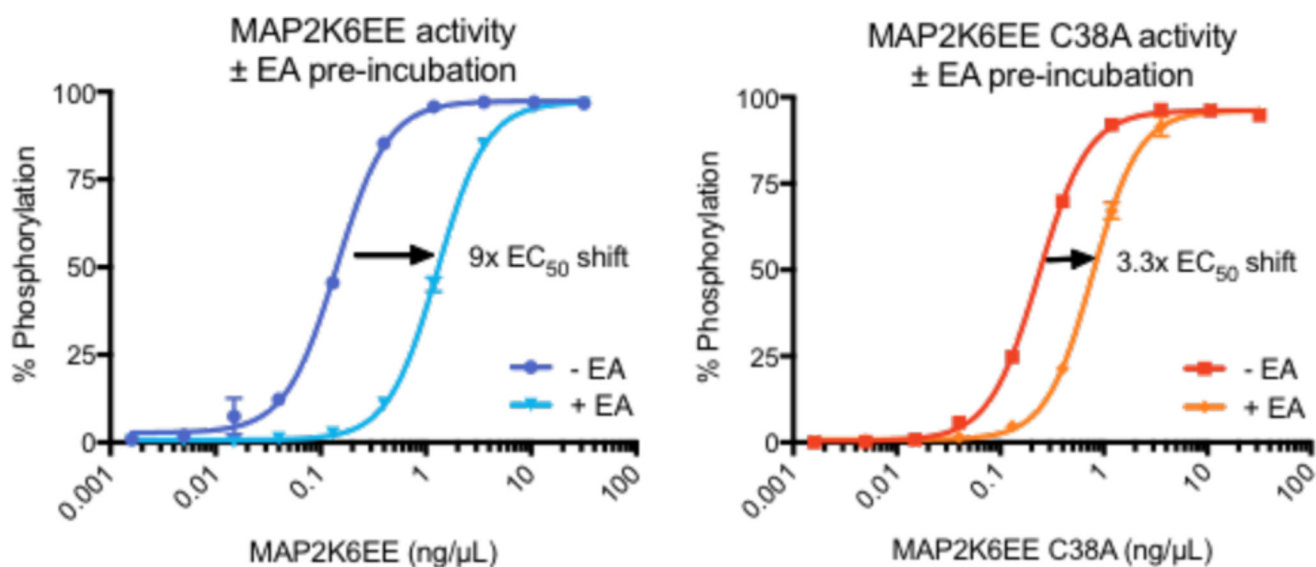
Figure 4.
Kinases were incubated with 100 μM EA (+EA) or DMSO (−EA), then dialyzed to remove free EA. Activity was assayed as a function of kinase concentration. Ethacrynic acid inhibits MAP2K6 to a greater degree when Cys38 is present (left) than when this residue is mutated to an alanine (right).
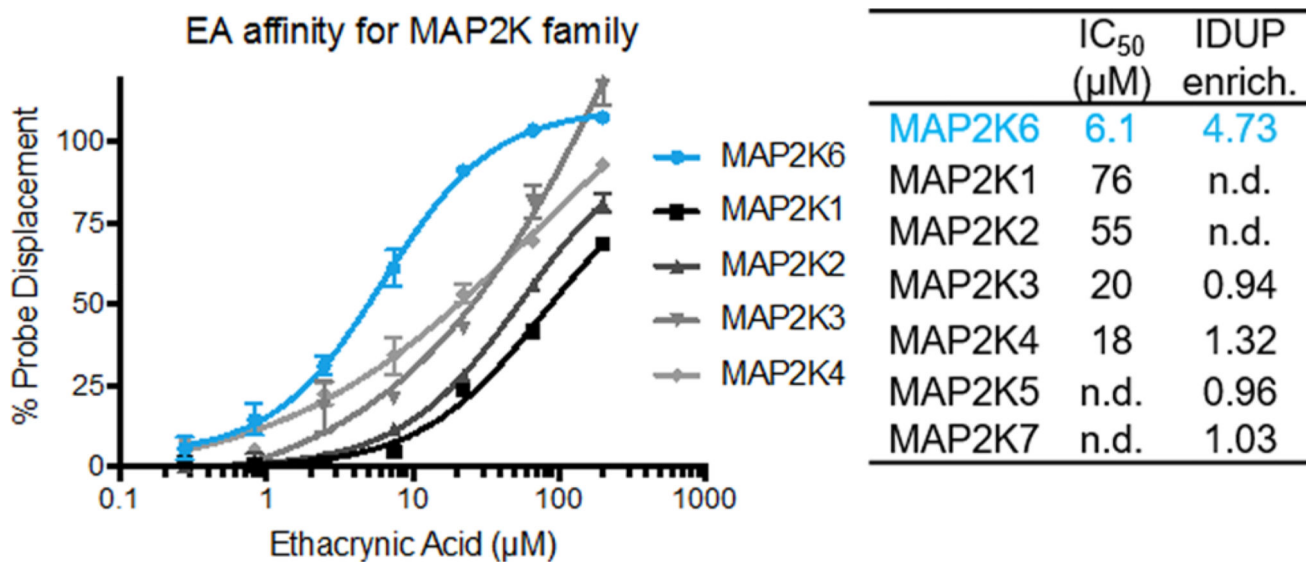
EA affinity for MAP2K family

	IC$_{50}$ (μM)	IDUP enrich.
MAP2K6	6.1	4.73
MAP2K1	76	n.d.
MAP2K2	55	n.d.
MAP2K3	20	0.94
MAP2K4	18	1.32
MAP2K5	n.d.	0.96
MAP2K7	n.d.	1.03

Figure 5.
EA has higher affinity for MAP2K6 than the related MAP2K family members, as measured in both LanthaScreen Eu assays and our initial IDUP assay. n.d. = not determined.