



Harnessing DNA-Encoded Libraries to Discover Bioactive Small Molecules

Citation

Chan, Alix I. 2018. Harnessing DNA-Encoded Libraries to Discover Bioactive Small Molecules. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:41128464>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

**Harnessing DNA-Encoded Libraries to Discover Bioactive Small
Molecules**

A dissertation presented

by

Alix I. Chan

to

The Committee on Higher Degrees in Chemical Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Chemical Biology

Harvard University

Cambridge, Massachusetts

March 2018

© 2018 Alix I. Chan

All rights reserved

Harnessing DNA-Encoded Libraries to Discover Bioactive Small Molecules

Abstract

Given the ever-increasing number of proteins, nucleic acids, and metabolites implicated in human disease, it is highly desirable to develop small molecules to probe therapeutically relevant biological pathways and to serve as leads for the development of new medicines. Recently, DNA-encoded chemical libraries – solution-phase collections of compounds each covalently linked to and specifically barcoded by a unique DNA sequence – have played an increasingly important role in the discovery of such bioactive compounds. DNA-encoded chemical libraries have several advantages, leveraging the extremely high sensitivity afforded by DNA amplification and the remarkable accessibility of modern high-throughput DNA sequencing. Taken together, these properties enable the efficient synthesis of large, diverse DNA-linked compound collections and the facile discovery of novel molecular interactions from these libraries. *In vitro* affinity selections on DNA-encoded chemical libraries have led to the discovery of new classes of synthetic small-molecule ligands against a variety of protein targets.

Previous work in the Liu group validated our ability to identify potent probe molecules from our DNA-templated libraries, as well as explore the biology of their protein targets through chemical means. I have applied *in vitro* selections on our DNA-templated macrocycle libraries to a large number of proteins and protein complexes associated with

human disease. In one case, our efforts led to new inhibitors of IDE from a library of 256,000 macrocycles, thus validating this library as a source of new bioactive compounds. In addition, I describe our use of the IDUP system to simultaneously evaluate all possible protein-ligand interactions out of combined libraries of DNA-tagged proteins and DNA-encoded small molecules. Not only were we successful in recapitulating known binding interactions in this assay format, I also discovered a previously unknown covalent inhibitor, ethacrynic acid, of the human protein kinase MAP2K6. I further probed the mechanistic basis of this binding interaction and showed that inhibition is due in part to ethacrynic acid's ability to alkylate a nonconserved cysteine residue in MAP2K6. These results are illustrative of the potential of unbiased *in vitro* binding selections to uncover bioactive molecules with novel modes of protein target engagement.

For mom and dad.

Table of Contents

Abbreviations	vii
Acknowledgements	ix
Chapter 1: Using DNA-encoded libraries to discover bioactive molecules	1
1.1 Introduction	2
1.2 The construction of DNA-encoded libraries	3
1.3 <i>In vitro</i> selection methods for ligand discovery from DNA encoded libraries	10
1.4 Reaction discovery using DNA-encoded chemical libraries	19
1.5 Conclusions and thesis overview	22
1.6 References	24
Chapter 2: <i>In vitro</i> selections on libraries of DNA-templated macrocycles	31
2.1 Motivations	32
2.2 Validation of a 256,000-member macrocycle library through the discovery of new IDE inhibitors	32
2.3 Selections on biomedically important proteins using the first-generation DNA templated library	39
2.4 Selections on biomedically important proteins using the second-generation 256,000-member macrocycle library	43
2.5 Discussion and outlook	48
2.6 General protocol for affinity selections	49
2.7 Follow-up synthesis of macrocycles after DTS library selections	56
2.8 References	62
Chapter 3: Discovery of a covalent kinase inhibitor from a DNA-encoded small-molecule library x protein library selection	64
3.1 Interaction determination using unpurified proteins (IDUP)	65
3.2 Construction of a DNA-barcoded protein library	66
3.3 Synthesis of a DNA-encoded bioactive compound library	72
3.4 IDUP library x library experiment to detect protein-ligand interactions	75
3.5 Identification and biochemical characterization of ethacrynic acid as a covalent inhibitor of MAP2K6	79
3.6 Selectivity of ethacrynic acid for MAP2K6 over other MAP2K proteins	82
3.7 Discussion	84
3.8 Experimental Methods	85
3.9 References	110
Chapter 4: Future directions for DNA-templated libraries	113
4.1 A new DNA-templated reaction: Suzuki-Miyaura couplings	114
4.2 Outlook	119
4.3 Experimental Methods	120
4.4 References	122

Abbreviations

AcOH	Acetic acid
ADHP	2-Amino-4,6-dihydropyrimidine
Alloc	Allyloxycarbonyl protecting group
BET	Bromodomain and Extra-Terminal
bisX	Bisindolylmaleimide X
Boc	tert-Butyloxycarbonyl protecting group
BRD2/BRD3	Bromodomain-containing protein 2/3
BSA	Bovine serum albumin
BSOCOES	bis(2-(succinimidooxycarbonyloxy)ethyl)sulfone
cAMP	Cyclic adenosine monophosphate
CatD	Cathepsin D
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DCM	Dichloromethane
DFG	Asp-Phe-Gly motif
DIPEA	<i>N,N'</i> -diisopropylethylamine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DMTMM	4-(4,6-Dimethoxy-1,3,5,-triazin-2-yl)-4-methylmorpholinium
DNA	Deoxyribonucleic acid
DPAL	DNA-programmed affinity labeling
DTS	DNA-templated synthesis
DTT	Dithiothreitol
EA	Ethacrynic acid
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	Ethylenediaminetetraacetic acid
ESAC	Encoded self-assembling chemical
EtOAc	Ethyl acetate
Fmoc	Fluorenylmethyloxycarbonyl protecting group
FRET	Förster resonance energy transfer/Fluorescence resonance energy transfer
GC/MS	Gas chromatography/mass spectrometry
GO	Gene ontology
GPCR	G-protein-coupled receptor
GSK3 α	Glycogen synthase kinase 3 α
GST	Glutathione S-transferase
HA	Hemagglutinin
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate
HEK293	Human embryonic kidney cells 293
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High-performance liquid chromatography
IAA	Iodoacetamide
IC ₅₀	Half maximal inhibitory concentration
IDE	Insulin-Degrading Enzyme

IDPCR	Interaction-dependent PCR
IDT	Integrated DNA Technologies
IDUP	Interaction determination using unpurified proteins
IPTG	Isopropyl β -D-1-thiogalactopyranoside
LC/MS	Liquid chromatography mass spectrometry
LDLR	Low density lipoprotein receptor
MALDI	Matrix assisted laser desorption/ionization
MAP2K	Mitogen-activated protein kinase kinase
MIDA	Methyliminodiacetic acid
NEB	New England Biolabs
NK3	Neurokinin-3
NKCC	Na-K-Cl cotransporter
NMM	<i>N</i> -Methylmorpholine
NMP	<i>N</i> -Methylpyrrolidone
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PAM	Protospacer adjacent motif
PBST	Phosphate buffered saline with Tween-20
PC-DNA	Photocrosslinking DNA
PCR	Polymerase chain reaction
PCSK9	Proprotein convertase subtilisin/kexin type 9
PKI	Protein kinase inhibitor
PLA	Proximity ligation assay
PLATO	Parallel analysis of translated ORFs
PMSF	Phenylmethylsulfonyl fluoride
PPI	Protein-protein interaction
PRKX	Protein kinase, X-linked
qPCR	Quantitative polymerase chain reaction
RDPCR	Reaction-dependent PCR
SMI-Seq	Single-molecular-interaction sequencing
sNHS	<i>N</i> -hydroxysulfosuccinimide
TBST	Tris-buffered saline with Tween-20
TCEP	Tris(2-carboxyethyl)phosphine
TEA	Triethylamine
TFA	Trifluoroacetic acid
TOF	Time-of-flight (mass spectrometry)
TPPTS	Triphenylphosphine-3,3',3''-trisulfonic acid trisodium salt
TXPTS	Tris(2,4-dimethyl-5-sulfophenyl) phosphine trisodium salt
USER	Uracil-specific excision reagent

Acknowledgements

Though the PhD is often a self-directed pursuit, the completion of this dissertation would not have been possible without the scientific, financial, social, and emotional support of countless others.

Thank you to Prof. David Liu for over five years of research insights, financial support, and scientific mentorship. I don't know where I could have found the academic freedom and scientific opportunities I have enjoyed during my tenure in the lab. Thank you also for creating and fostering a lab environment that is endlessly supportive and creative.

Thank you to Prof. Stuart Schreiber, Prof. Emily Balskus, and Prof. Nathanael Gray for serving as my DAC for the past three years. I appreciate all the advice you provided as I attempted to navigate my way through graduate school. Thank you also to Prof. Alan Saghatelian for advice during my PQE and to Prof. Amit Choudhary for serving on my defense committee.

Thank you to Lynn McGregor for mentoring me in the early years of my PhD, as well as setting the foundations of the IDUP project. You really set me up for success and for that I cannot be more grateful.

Thank you to Juan Pablo Maianti for your guidance since day 1, literally, of my time in the Liu lab. I'm incredibly thankful for your mentorship in all aspects of science and research.

Thank you to Dmitry Usanov for your tireless research efforts, which opened up so many opportunities for myself and others. I am fortunate to have worked alongside you for the past few years.

Thank you to Phillip Lichtor and David Bryson for all the incredibly constructive scientific discussions and assistance. Thank you also for all the incredibly unproductive and distracting sidebar conversations, you made our office a place I always looked forward to coming into.

Thank you to other members of the DTS subgroup, past and present - Rick McDonald, Jia Niu, Zhen Chen, Beverly Mok, Alex Peterson, Jon Chen - for five years of stimulating discussions and scientific support, both during Millis and at all other hours.

Thank you to all the other members of the Liu group for the insights, inspiration, and for generally make the lab a fun place to work, particularly Ahmed Badran, Alexis Komor, Bill Kim, Brian Chaikind, Chris Podracky, Holly Rees, Jeff Bessen, Johnny Hu, Jon Levy, Luke Koblan, Nicole Gaudelli, Sherry Gao, Tim Roth, Tina Wang, Tony Huang, and Travis Blum. I have been lucky to benefit from your collective wisdom and friendship.

Thank you to Aleks Markovic for ensuring that the lab ran smoothly every single day, which is no small feat at all.

Thank you to Jason Millberg who has been a steady source of support since the day I was invited to interview for the Chemical Biology program. Thank you also to Suzanne Walker and Dan Kahne for their leadership of and advocacy for the Chemical Biology program. I am grateful to have ended up in a graduate program with such strong community and support.

Thank you to all members of my Boston social life for keeping me sane through all the ups and downs of graduate school. To Abraham Waldman, Alex Turner, Andrew Snavely, Angela She, Ayano Kohlgruber, Bennett Meier, Fred Rubino, Harriet Lau, Hubert Huang, Jack Nicoludis, Katie Schaefer, Liv Johannessen, Leigh Matano, Nav Ranu, Nitzan

Koppel, Sam Wellington, Sami Farhi, Su Vora, Tristan Owens, Vini Mani, and Wesley Chen – thank you for sharing laughter, food & drink, griping sessions, too much Gchatting, Type 1 and occasional Type 2 fun, and many more other gifts than I can enumerate. Special shout-out to Justin Feng for being a willing a co-conspirator in all manner of Fun Things; the last few years have been a blast and I owe you at least 50% of the credit.

Thank you to my all my undergraduate research advisors. Prof. Todd Clements was the first to introduce me to independent research. Thank you to Prof. Katherine Maloney for an exceptional summer research experience and for encouraging me to apply to the Harvard Chemical Biology program. Thank you to Prof. David Vosburg for being a wonderful undergraduate thesis mentor, both in teaching me about organic chemistry and guiding me in how to become a successful independent researcher.

Thank you to the Narwhals – Andrew Jennings, Heather Williams, Keiko Hiranaka, Kim Quach, Maia Valcarce, Mary Van Vleet, Nick Hill, Sarah Ferraro, Steve Matsumoto – for over a decade of late night diner runs, legitimately amazing karaoke to legitimately awful songs, and too many terrible inside jokes that I can barely remember the source of. I am grateful every day for your limitless support and exceptional camaraderie.

Thank you to Mr. Ross Ellison for introducing me to the logic and beauty of chemistry. I am endlessly thankful for the energy and rigor you brought to your teaching, which left me with the strongest possible foundation for my future education.

Thank you to Olivia Pogorelskin, for helping me survive everything from hypothetical gangrene to actual bears. I hope we're brought to tears by many more sunsets in the future.

Thank you to my Popo for always looking out for me, for helping make college a reality, and for teaching me to appreciate a home-cooked meal.

Thank you Mom for always doing everything in your power to give me the privileged life I have been lucky enough to lead, even when it meant losing sleep or sanity.

Thank you to Dad for all the sacrifices I know you made to get me here, and even more thanks for the sacrifices that I am not even aware of.

Thank you to Brandon for your unwavering confidence in my abilities, for commiserating when things go south, and for your insights and perspective. See you in the same timezone soon, hopefully.

Chapter 1: Using DNA-encoded libraries to discover bioactive molecules

Alix I. Chan, Lynn M. McGregor and David R. Liu

Adapted in part from: Chan, A. I.; McGregor, L. M.; Liu, D. R. "Novel Selection Methods for DNA-Encoded Chemical Libraries" *Curr. Opin. Chem. Biol.* **26**, 55-61 (2015).

1.1 Introduction

The rapidly expanding wealth of genomic, proteomic, and metabolomic data has led to the identification of many biological targets with therapeutic potential. Small molecules that can potently and specifically engage proteins or nucleic acids associated with disease are especially valuable as probes to validate the putative roles of targets in disease progression, or as potential leads for therapeutic development. Conventional efforts to discover small-molecule ligands frequently use high-throughput screening, in which thousands of compounds are individually exposed to a target of interest and assayed for bioactivity. Such screening methods, however, can be costly, time-consuming, and require major instrumentation and specialized expertise.

A complementary approach to evaluating synthetic small-molecule libraries uses *in vitro* selections to rapidly and simultaneously assess the ability of all library members to interact with targets of interest. This approach is especially amenable to the evaluation of large, chemically diverse, DNA-encoded libraries that have been described by several research groups in academia and industry [1-13]. DNA-encoded libraries consist of collections of molecules that are each covalently linked to a distinct DNA oligonucleotide. The sequence of the DNA acts as a unique barcode that the researcher designs to specifically correspond to each chemical structure. Because minute amounts of DNA can be readily replicated and sequenced, *in vitro* selections of DNA-encoded chemical libraries offer the major advantage of simultaneously evaluating up to billions of compounds for their ability to interact with target proteins in a single experiment. Selections significantly reduce the amount of compounds, target protein, time, and cost required to evaluate a library [14-15]. In addition, selections on such libraries can readily yield structure–activity

relationships that inform future medicinal chemistry efforts. Recently, DNA-encoded chemical libraries have played an increasingly large role in small molecule probe and drug discovery campaigns in both academic and industrial settings.

1.2 The construction of DNA-encoded libraries

The theoretical principle of using DNA to encode synthetic libraries was outlined by Brenner and Lerner's proposal [16] for the concurrent syntheses of polypeptide and oligonucleotide sequences on a chemically functionalized solid support. Because of the orthogonal chemistries necessary to synthesize these polymers, it was envisioned that such a system could be used to link an artificial genotype (the DNA sequence) and phenotype (the peptide sequence and its potential bioactivity). In addition, the repertoire of monomer building blocks would not be limited to those accepted by biological translation systems such as nucleic acid libraries (as used in SELEX methods) or phage-displayed peptides [17-20]. Indeed, bead-supported encoded libraries were soon synthesized. The first reports in 1993 from Needels *et al* [21] and Nielsen *et al* [22] showed that DNA-encoded peptides, cosynthesized on solid supports, could be selected for binding to cognate antibodies.

In 2004 came the first reports of combinatorially assembled synthetic small molecule libraries [23-25], wherein DNA tags were directly covalently linked to library members without a solid support. The field of DNA-encoded chemical libraries has been dominated by solution-phase libraries since (with some notable exceptions [26-27]), as the solid support is unnecessary in most cases for linking library member to encoding barcode. One main limitation of solution-phase assembly of DNA-encoded chemical libraries – that the combinatorial synthetic steps are often limited to water-compatible reactions that do not damage DNA – has been addressed by continuing efforts to develop and validate new

“DNA-compatible” reactions [28-30]. Indeed, using both commonly used, robust reactions (e.g. amide bond formation, reductive amination) and the expanding repertoire of DNA-compatible reactions, libraries of up to billions or trillions of members have been reported [31-33].

Broadly speaking, DNA-encoded combinatorial synthesis methods can be categorized as *DNA-recorded* or *DNA-directed* (Figure 1.1). In one round of DNA-recorded chemistry (representative scheme in Figure 1.1A), a DNA-linked small molecule scaffold is first coupled to another building block organic fragment. This is followed by tagging of the DNA barcode with short oligonucleotide whose sequence is specific to the reaction that was carried out. This DNA tagging can be carried out through enzymatic methods such as Klenow polymerase fill-in or T4 DNA ligase splinting [4,34]. Chemical ligation methods, employing reactions such as the alkyne-azide “click” cycloaddition or photoactivated crosslinking, have also been used [35]. In any case, 2-4 successive rounds of split-and-pool synthesis, alternating with DNA barcode elaboration, lead to the final combinatorially assembled library. The vast majority of DNA encoded chemical libraries, particularly those employed in industrial settings, are constructed in a DNA-recorded manner, due to the relative ease of synthesis and ability to access large compound collections [31-33].

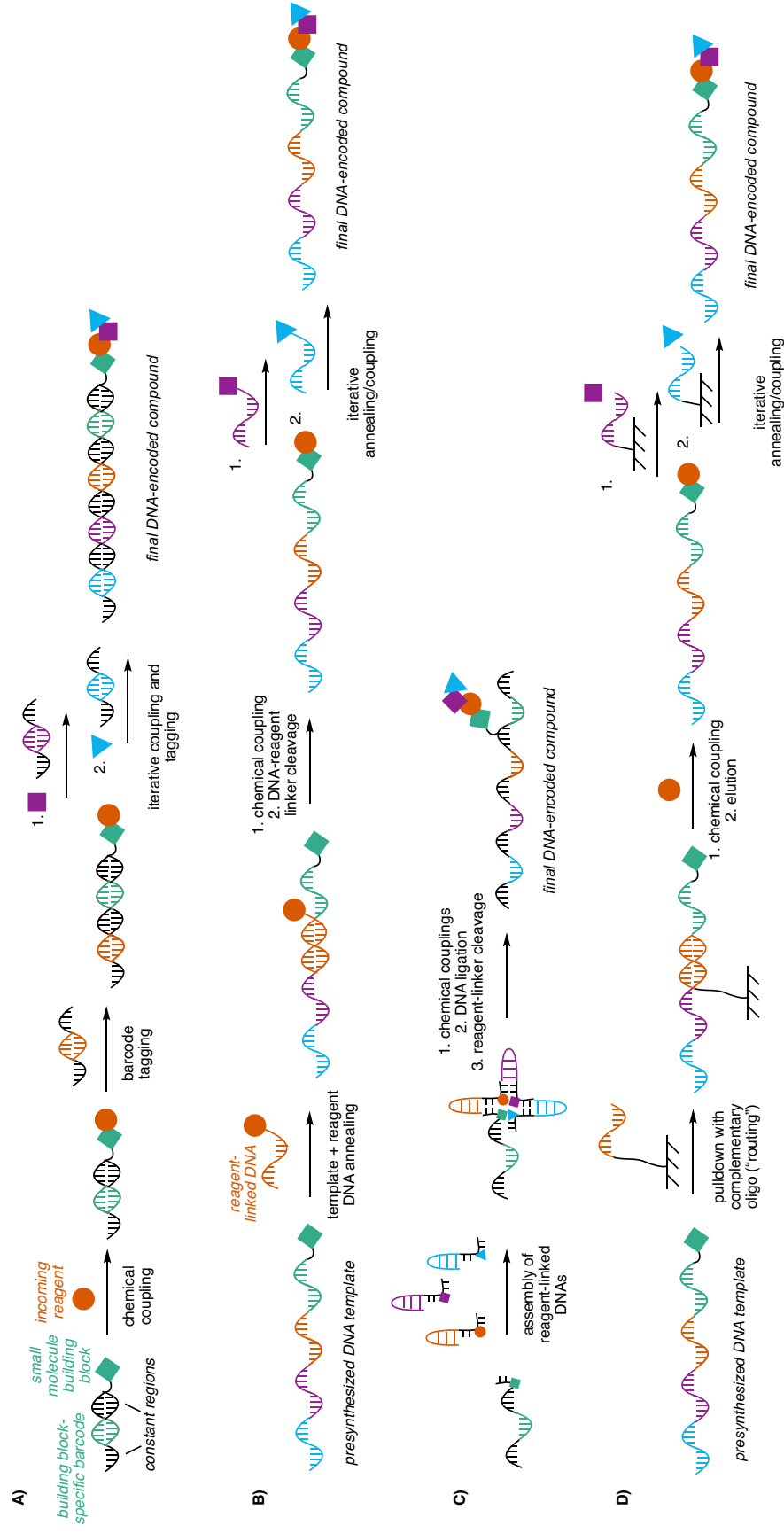


Figure 1.1. Synthetic schemes for the generation of DNA-tagged small molecules. Each of these strategies can combinatorially assemble large building block collections to synthesize entire libraries of diverse DNA-encoded compounds. (A) Molecules synthesized using *DNA-recorded chemistry* are constructed through iterative steps of solution-phase chemical coupling and DNA barcode tagging. (B)-(D) illustrate examples of *DNA-directed chemistry*. (B) In DNA-templated reactions, the presynthesized DNA oligonucleotides bear small molecules building blocks. Oligonucleotides with complementary sequences can hybridize and promote reactions between their small molecule cargos. (C) In YoctoReactor system, small molecule building blocks are assembled and coupled at a junction of DNA hairpins. (D) In DNA routing, template DNAs are isolated by pulldown onto complementary sequences immobilized on solid support. These spatially separated templates are subjected to independent organic reactions before being eluted and the cycle of annealing/coupling is repeated.

In contrast, in *DNA-directed* synthetic methods (Figure 1.1B-D), a presynthesized DNA sequence linked to the library starting fragment(s) determines the chemical transformations that are subsequently performed. Various methods have been reported for DNA-directed library synthesis, but all utilize base pairing between a chemically functionalized “template” DNA strand and “reagent” functionalized DNA strands that bear at least partially complementary sequences. The resulting DNA duplexes drive specific reactivity between the organic moieties linked to the template/reagent DNA pairs.

The Liu group first reported its version of DNA-directed chemistry, termed *DNA-templated synthesis* (DTS), starting in 2001 [36], in which one DNA-linked reactive moiety on the template strand (bearing a “codon” sequence) specifically hybridizes to another reagent DNA-linked reactant (the complementary “anticodon” sequence) via Watson-Crick base pairing (Figure 1.1B). Only under the higher effective molarity conditions induced by this DNA duplex does coupling of these two reactants occur. This allows for combinatorial synthesis of multiple combinations of codon/anticodon pairs in a single solution. The incoming building blocks are ultimately cleaved from the reagent DNA strand and transferred to the template strand, regenerating a reactive moiety (e.g. an amine) for subsequent DNA-templated chemical transformations.

DNA-templated chemistry has several advantages. Because reactions are determined by DNA complementarity, multiple distinct reactions can be specifically performed in a single pot, obviating the need to split-and-pool library members at each step of combinatorial assembly [23]. In addition, the high effective molarity of reagents induced by the DNA duplex can help favor otherwise unfavorable chemical

transformations. Finally, the Liu group has been able to take advantage of the DTS scheme to make and isolate DNA-linked peptide macrocycles in high purity, which is not possible through DNA-recorded reactions [2]. Using this technique, in 2010 the Liu group published the synthesis of a library of 13,824 DNA-templated peptidic macrocycles [2]. Our most recent efforts have culminated in the DNA-templated synthesis of a 256,000-member library of macrocycles with improved druglike properties [see Chapter 2].

The Li group developed a variation of DNA-templated synthesis [13] in which polyinosine stretches were included in the template DNA strand. This allows for DNA templated reactions to occur on a “universal template”, obviating the need for split-and-pool synthesis of the entire repertoire of DNA templates. Each oligonucleotide reagent DNA strand is ligated to the template oligonucleotide after coupling so that the final product can be decoded via DNA sequencing.

Other alternate methods for DNA-directed chemistry are conceptually similar to the strategy of Li *et al* [13], in that the DNA barcode for the final product is not predetermined. Instead, the hybridization events that drive coupling are possible for any template+reagent building block pairs. The barcoding portion of the DNA does not necessarily participate in duplex formation; instead, it records the reaction that occurred via a subsequent DNA reagent ligation. Vipergen developed the YoctoReactor system (Figure 1.1C) [1, 37], which assembles multiple building block-bearing DNA hairpins at a central DNA junction to drive reactivity. Turberfield and colleagues also recently reported a system for iterative hybridization of DNA hairpin-linked monomer units, where assembly is driven by short complementary sequences at each building block DNA’s termini [38]. Their system was used to synthesize polypeptide and polyolefin sequences in a DNA-templated manner.

Other routes for DNA-directed chemistry have also been developed that do not rely on the higher effective molarity of solution-phase hybridized template+reagent pairs to direct on-DNA reactions. In what the Harbury group termed *DNA routing* (Figure 1.1D) [39], presynthesized DNA templates are “routed” through a series of solid supports each functionalized with a different anticodon sequence. This process spatially separates the library members so that discrete chemical transformations can be performed on each isolated aliquot. Iterative steps of elution, pooling and rerouting lead to combinatorial assembly of the final library [25].

Regardless of the technique used to synthesize and encode these molecules, it has become obvious within the last decade that DNA-encoded chemical libraries have become a rich source of bioactive compounds. Researchers have continued to advance the synthetic methods available to construct DNA-encoded chemical libraries. The size of these libraries enables sampling of wider swaths of chemical space (billions or more compounds) than traditional high-throughput screening compound collections (limited to millions of molecules) [31]. Reports have described molecules that are more difficult for conventional flask syntheses to access such as macrocycles [2, 40, 41], molecules with unique functionality, such as targeting a protein-protein interaction [42], or electrophilic moieties that could irreversibly bind to protein targets [27]. A vast number of chemical probes and drug leads have been published (for the most recent, but already outdated, compilation of hit molecules see Goodnow *et al* 2017 [31]). Previous work in the Liu group identified inhibitors of Src kinase [43-44] and insulin-degrading enzyme (IDE) [45] from a library of relatively modest size (13,824 members) [2]. As of this writing, other hits from DNA-encoded libraries have been advanced as far as Phase I clinical trials [46]. With the large

amount of investment in biotechnology startups and within pharmaceutical companies in DNA-encoded library technology, it is likely that this technology will only play a larger role in drug discovery pipelines in the future.

1.3 *In vitro* selection methods for ligand discovery from DNA encoded libraries

The most common selections performed on DNA-encoded chemical libraries are binding selections on purified target proteins (Figure 1.2A). A target protein is immobilized by affinity tag or covalent attachment to a solid support and then incubated with a DNA-encoded chemical library. Alternatively, the library can be incubated with free protein in solution and either captured with immobilized, target-specific antibodies or pulled down directly via an affinity handle [1,47,48]. After washing [7], the bound protein and library members capable of binding the target are eluted and subjected to additional round(s) of selection or PCR amplification of their associated DNA templates and massively parallel high-throughput DNA sequencing. Depending on the total library size, putative hits can be identified from amplified sequences either by comparing the sequence abundance to that in the starting library [45] or by fitting observed sequence counts to a negative binomial distribution [8] or to the Poisson distribution [49].

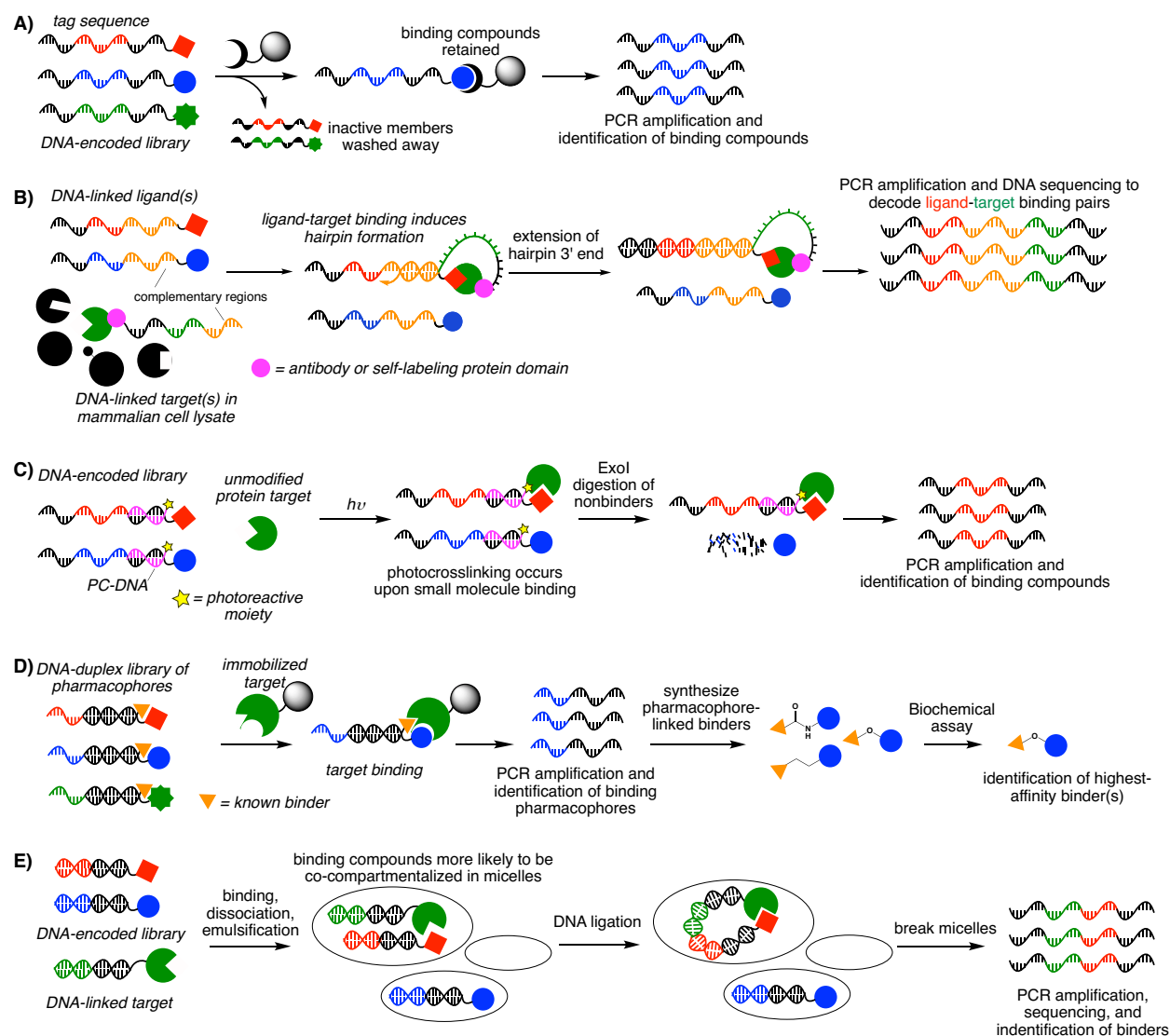


Figure 1.2. Technologies for the *in vitro* selection of DNA-encoded chemical libraries. (A) In solid-phase selection, all library members are simultaneously evaluated. Library members that are capable binding of an immobilized target protein are physically separated from inactive library members. (B) Interaction determination using unpurified proteins (IDUP) takes advantage of the selective formation of a DNA hairpin between binding species to encode both protein and ligand on the same DNA strand. (C) DNA-programmed affinity labeling (DPAL) makes use of an oligonucleotide-coupled photoreactive group (PC-DNA) that covalently labels only proteins bound by a library member. These complexes protect the corresponding DNA tag from digestion and allow its subsequent library sequencing and decoding. (D) Encoded self-assembling chemical libraries (ESAC) use DNA hybridization to display multiple pharmacophores for interaction with a target protein. Tightly binding pharmacophore combinations guide the synthesis of high-affinity ligands based on the pharmacophore fragments. (E) Library selections using binder trap enrichment utilize micelles to co-compartmentalize DNA tags associated with binding species.

As DNA-encoded library selections have become more widespread in academia and industry within the past few years, the experimental protocols for performing such selections have been highly streamlined. Automation [50] has become commonplace, and the ability to run selections in parallel allows for rapid assessment of the ligandability of multiple protein targets [51]. Even when performed manually, a researcher can easily perform selections on a single protein target under a variety of conditions (e.g. varying target concentration, or including or excluding known protein cofactors) to gain information about enriched species' binding modes and affinities directly from post-selection sequencing data [48].

While one can perform selections for binding affinity on a large number of human proteins (and this method is applicable regardless of a protein's specific biological fold or function), such target-based selections on DNA encoded chemical libraries have largely been limited to soluble, purified protein targets. Naturally, there are many classes of proteins that do not express or maintain their physiologically relevant states *in vitro*. For example, though G-protein coupled receptors (GPCRs) represent one of the largest classes of drug targets [52], these membrane-bound targets are not generally amenable to *in vitro* selection under standard conditions. Groups have successfully performed selections against GPCRs by mutationally thermostabilizing the protein targets [53] or solubilizing the recombinant protein with detergent [54]. Wu, Israel, and colleagues reported a cell-based method that allows for affinity selections to be performed on membrane proteins [55]. They first overexpressed the NK3 receptor (a member of the tachykinin family of GPCRs) on HEK293 cells and used the cells directly as bait for affinity selection against a number of DNA-encoded libraries. These selections required a large amount of starting library (~2

nmol) and iterative rounds of selection, but they were nevertheless able to yield known and novel NK3 ligands with affinities as low as sub-nanomolar.

While selections on solid support-immobilized protein targets offer dramatically increased efficiency compared to screens, they still for the most part must be performed on a single, purified target protein of interest. Target immobilization may result in artefactual binding or in the loss of native conformational properties that are required for bona fide binding to native targets. In addition, washing and elution steps required for selections using immobilized targets may also remove active library members or fail to result in the isolation of desired species. Thus, alternate methods that attempt to circumvent these limitations have been developed to identify binders from DNA encoded libraries.

Interaction-dependent PCR (IDPCR) [56] was developed by the Liu group to address these limitations and enable simultaneous evaluation of binding between all members of combined libraries of DNA-linked targets and compounds in a single solution (Figure 1.2B). In IDPCR, protein-small molecule binding brings encoding DNA sequences into close proximity and promotes DNA hybridization of a self-priming hairpin. Polymerase-catalyzed primer extension produces a selectively amplifiable DNA sequence encoding both members of the protein-small molecule complex.

The IDPCR approach was extended to detect ligand binding to unpurified proteins in a method called interaction determination using unpurified proteins (IDUP) [57]. By operating in cell lysates, IDUP preserves post-translational modifications and interactions with endogenous binding partners, enabling the study of difficult-to-purify targets and increasing the potential biological relevance of detected interactions. During IDUP, target proteins are associated with DNA oligonucleotide tags either noncovalently using a DNA-

linked antibody or covalently using a self-labelling protein domain such as a SNAP-tag. In a model library x library binding experiment using combined libraries of 262 DNA-linked small molecules and 256 cell lysates expressing SNAP-tagged targets, IDUP enriched all five known interactions highly, despite having affinities varying from 0.2 nM to 3.2 mM [57]. This method provides an efficient approach for rapidly evaluating the binding of ligand libraries in cases in which purified proteins are not available or differ significantly from their native cellular counterparts. Chapter 3 of this thesis describes the latest iteration of the IDUP as it is applied to a campaign to discover previously unknown binders from real libraries of proteins and small molecules.

Li and coworkers developed another method for assaying binding of small-molecules libraries to unmodified, non-immobilized proteins by DNA-programmed affinity labeling (DPAL) [58]. In DPAL (Figure 1.2C), unmodified proteins are mixed with small DNA oligonucleotides bearing a 5'-azidophenyl photocrosslinking moiety (PC-DNA). When a DNA-tagged small molecule binds the target, DNA hybridization to a complementary region on the short PC-DNA brings the photo-reactive group into proximity with the protein target. UV irradiation leads to covalent attachment of the PC-DNA to the target protein, which can be identified by mass spectrometry. The DPAL technique was adapted to identify binders from libraries of small molecules by taking advantage of hybridization between the PC-DNA and the binding ligands' DNA tags [59]. The DNA tags of non-binding small molecules are digested by Exonuclease I, but the hybridized DNA tags of bound molecules are protected from digestion. In addition, DPAL is compatible with targets in cell lysates, a condition in which targets may closely mimic their native state, and with large excesses of non-binding DNA-linked small molecules. DPAL can also perform iterative

rounds of selection to greatly amplify DNA sequences corresponding to binding small molecules.

Photoaffinity probes, such as diazirines, benzophenones, and phenyl azides, stabilize interactions between protein targets and DNA-linked small molecules [58,60]. Using DPAL, Li and coworkers enriched DNA sequences corresponding to interactions as weak as 14 mM. Especially when combined with multivalent ligand display [60], the use of photoaffinity probes could stabilize weak interactions enough to enable DNA-encoded fragment-based screens [61].

In the most recent iteration of this method, Li and coworkers reported a “ligate-cross-link-purify” strategy that could be theoretically applied to existing DNA-encoded libraries [62]. A short PC-DNA with complementarity to the small molecule-proximal region of the DNA tag is ligated to each library member. The resulting hairpin structure brings the photocrosslinking moiety (a 3'phenylazide in this case) in close proximity to any bound protein targets. After UV irradiation, the covalent protein-DNA adducts can be isolated and the bound library members identified through high-throughput sequencing. This approach was shown to enrich even moderate-affinity (36-89 μ M) macrocyclic binders to avidin.

Denton and Krusemark also reported a similar strategy to crosslink DNA-linked binders to their protein targets using electrophilic or photoreactive groups [63]. Creating this covalent linkage between target-ligand pairs allowed for more stringent washing of affinity-immobilized proteins after library incubation and crosslinking. Such a method could theoretically enrich binders with high dissociation constants or lower overall affinity.

In an alternate approach to fragment-based selections, Neri and coworkers

developed a DNA-hybridization based approach for performing selections using encoded self- assembled chemical libraries (ESAC) [24]. In an ESAC selection (Figure 1.2D), a DNA-linked small-molecule ligand is hybridized to a library of DNA-encoded pharmacophores. The hybridized mixture is subjected to *in vitro* selection either to identify fragments that improve binding of a known ligand (affinity maturation) or to discover novel synergistically binding ligands. These fragments are then covalently coupled using a variety of linkers, and the resulting compounds are assayed without the DNA barcodes to discover linker architectures optimal for binding. Using ESAC, Neri and coworkers discovered higher-affinity inhibitors of trypsin [64] and MMP-3 [9].

When DNA-linked antibodies are bound to a protein or protein complex, the DNA tags are brought into close proximity and can be linked by ligation [65] or extension [66] to give a selectively amplifiable DNA sequence. Landegren and coworkers have validated the proximity ligation assay (PLA) for the quantification of biomarkers by high-throughput sequencing [67] and for the analysis of protein–protein interactions and protein subcellular localization or post-translational modification [68,69]. PLA was also applied to the study of small molecules capable of disrupting the interaction between VEGF-A and its receptors VEGFR-1 and VEGFR-2 [70].

Using a microarray to quantify unique DNA tags for each antibody, PLA can evaluate all pairwise protein–protein interactions (PPIs) within a larger set of proteins [71]. Recently, Huang and coworkers applied PLA to interrogate 1204 PPIs in a one-by-one fashion [72], resulting in the identification of hundreds of previously validated PPIs. Tagging individual antibodies with unique DNA barcodes could enable readout of PPIs by high- throughput DNA sequencing and expedite PPI cataloging efforts.

Whereas the previously described methods use DNA hybridization to promote amplification of active library members, *in vitro* selection typically takes advantage of the spatial separation of active library members from inactive species. To select binding molecules from their DNA-encoded chemical library [1], Vipergen developed a technique called binder trap enrichment (Figure 1.2E) [73]. After exposing a DNA-encoded chemical library to a target of interest, binding pairs are compartmentalized in water and oil emulsion droplets. Statistically, binding ligands are more efficiently co-compartmentalized with the target protein than nonfunctional library members. Enzymatic ligation of the co-compartmentalized oligonucleotides encoding the ligand and target is followed by PCR amplification and high-throughput DNA sequencing to identify the binding entities.

The vast majority of DNA-encoded library screening is restricted to binding assays, with the assumption that affinity can often be used as a proxy for biochemical efficacy. Indeed, the low concentrations of each individual library member and the inability to test members in isolation preclude any sort of functional or phenotypic assay (that would be analogous to traditional high-throughput screens). However, Paegel and coworkers have reported a means to perform functional assays on DNA-encoded compounds by integrating on-bead synthesized libraries with microfluidic systems [74]. Like the Vipergen system, this selection method encapsulates library members into aqueous droplets in an oil suspension, though the Paegel group's libraries are synthesized and encapsulated on a solid bead support. However, in this case, every library members is co-incubated with an enzyme target and its substrate within the droplet. Once encapsulated, compounds are liberated from the bead support (via photocleavage) and can be assessed individually, in solution, for inhibition of the enzyme target's biochemical activity. In one example, Paegel

and colleagues were able to detect and isolate beads encoding a known inhibitor, pepstatin A, of Cathepsin D (CatD) within individual droplets. Pepstatin A inhibits CatD's proteolytic activity on a fluorogenic peptide that is co-incubated within the droplets; thus, droplets encapsulating the inhibitor exhibited lower fluorescence and could be isolated from non-inhibitor-containing droplets using a fluorescence-detecting droplet sorting system. This system, combining DNA-encoded library synthesis on solid support with droplet encapsulation and microfluidic sorting, is uniquely able to perform functional selections on DNA-encoded libraries. No other DNA-encoded library selection method is able to directly test for inhibition of a target enzyme's bioactivity. Though this method requires specialized equipment and expertise, and the throughput may be lower than traditional solid-supported affinity selections, this capability to test biochemical efficacy is not currently possible with all other solution-phase DNA-encoded library selections.

1.4 Reaction discovery using DNA-encoded chemical libraries

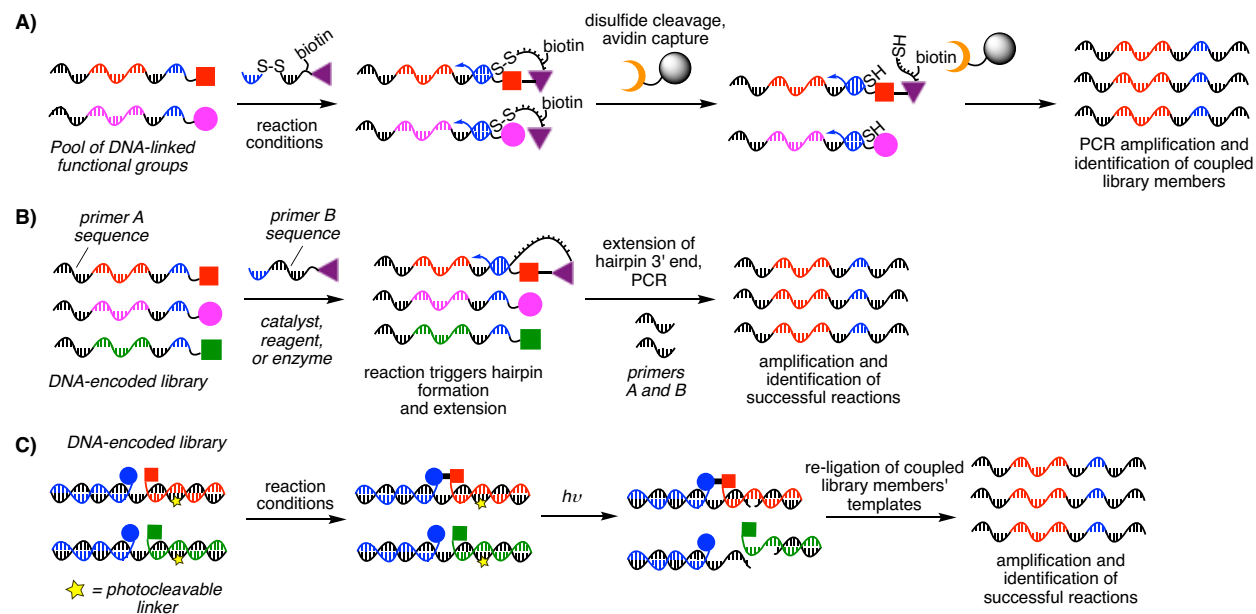


Figure 1.3. Approaches for DNA-encoded discovery of bond-forming reactions. (A) The first reported reactions discovered from DNA-encoded chemical libraries used disulfide cleavage and streptavidin capture to isolate the DNA encoding pairs of substrates that had undergone bond formation under a given set of reaction conditions. (B) Reactivity-dependent PCR (RDPCR) is a one-pot method that uses the formation of a self-priming hairpin between coupled library members' DNA barcodes to selectively amplify DNA encoding bond-forming substrate combinations. (C) Bond formation can also be detected by installing a photocleavable linker in a DNA template used for reagent hybridization. Only covalently coupled reagents' templates are re-ligated, amplified, and sequenced.

In addition to using DNA-encoded libraries for discovery of bioactive ligands, similar techniques have been developed for the high throughput selection of chemical reactions. Interest in the development of novel chemical transformations has led to the development of unbiased reactivity screens based on LC/MS [75], GC/MS [76], and sandwich immunoassay [77]. Reaction discovery using DNA-encoded chemical libraries can dramatically streamline the reaction discovery process by enabling simultaneous evaluation of all possible combinations of potential reactants among substrate library members. The first DNA-encoded reaction discovery schemes by Liu and coworkers relied

on solid-phase separation of bond-forming pairs of library members and their encoding DNA from inactive species (Figure 1.3A) and resulted in the discovery of a Pd(II)-mediated alkynamide-alkene coupling reaction [78], a Au(III)-catalyzed and acid-catalyzed alkene hydroarylation [79] and a biocompatible, Ru(II)-catalyzed azide reduction induced by visible light [80].

In reactivity-dependent PCR (RDPCR) (Figure 1.3B), covalent bond formation results in hybridization of a self-priming hairpin that can be extended to encode the identity of both substrates and that is selectively amplified in PCR [81]. RDPCR can also be configured to detect bond cleavage that results in the unmasking of a reactive functional group. In a proof-of-principle example, a DNA-linked peptide reacted with an activated DNA-linked carboxylate only if the peptide was first treated with the protease subtilisin (generating a free amino terminus) [82]. Li and coworkers developed another approach to report covalent bond formation using DNA to bring together substrates that hybridize to a common template (Figure 1.3C) [83]. After hybridizing the substrates and allowing them to react, a photolabile base in the template strand is cleaved. If covalent bond formation has joined the two substrates, their pendant DNA sequences serve as splints for a ligation reaction that restores the cleaved template strand and enables its PCR amplification.

Methods with the potential to evaluate DNA-encoded chemical libraries A variety of novel DNA-based methods for detecting binding between proteins and small molecules rely on the ability of protein-small molecule complexes to prevent interactions between enzymes and encoding DNA. Complexes that prevent digestion by *Escherichia coli* Exonuclease I [83-85] or Exonuclease III [86] could in principle allow detection of DNA sequences encoding active small molecules by high-throughput sequencing, without relying

on spatial separation or primer extension to further amplify the codes of active library members. These potential selection methods would not require solid-phase immobilization, washing, or elution steps, but would render the DNA-encoded libraries single-use. Indeed, DPAL, IDUP, and the proximity extension assay all take advantage of protein-small molecule complexes' ability to protect DNA from exonucleases to decrease the signal arising from inactive library members [57,59,66]. Complexes that prevent DNA from interacting with T4 DNA ligase [87] or T7 RNA polymerase [88] could also be adapted to evaluate DNA-encoded chemical libraries. These approaches would have the significant disadvantage, however, of operating by a loss-of-signal mechanism in which the sequences corresponding to inactive compounds are selectively amplified.

Novel DNA sequencing techniques can also drive advances in library evaluation. Colonies are clusters of DNA immobilized in a polyacrylamide gel that can be sequenced in high-throughput by single base extension or ligation-based sequencing [89]. Single-molecular-interaction sequencing (SMI-Seq), an approach developed by Church and coworkers, uses colony sequencing to identify the complexes of DNA-linked proteins by analyzing the degree of co-localization of colonies corresponding to each protein [90]. SMI-Seq was used to evaluate the selectivity of each member of a library of 200 single-chain variable fragment (scFv) variants in its ability to bind to each member of a library of 55 human antigens.

In the strand-displacement competition assay developed by Gothelf and coworkers [91], a DNA-linked small molecule is hybridized with a complementary oligonucleotide. Protein-small molecule binding decreases the affinity of this duplex, enabling toehold displacement of the complementary strand. In the proof-of-principle study, this toehold

displacement resulted in a signal that could be analyzed by PAGE (polyacrylamide gel electrophoresis) or FRET (Förster resonance energy transfer). In principle, this approach could also be adapted to a high-throughput sequencing-based readout, and to the evaluation of DNA-encoded chemical libraries.

Solid-phase selections have also been adapted to the study of protein–protein interactions. In parallel analysis of translated ORFs (PLATO), Elledge and coworkers use an immobilized protein to capture binding proteins that are linked to their encoding DNA by ribosome display [92]. PLATO could in principle be adapted to evaluate small molecules linked to beads for target identification, or when interfaced with a DNA-encoded chemical library, could potentially enable simultaneous evaluation of all interactions with a set of ribosome-displayed protein targets.

1.5 Conclusions and thesis overview

Selections on DNA-encoded chemical libraries have recently resulted in the discovery of new classes of synthetic small-molecule ligands against a variety of protein targets including several associated with human disease [4,7,45,93-95]. Given the ever-increasing number of proteins, nucleic acids, and metabolites implicated in human disease, innovations in the field of DNA-encoded libraries and their rapid *in vitro* selection are likely to play an increasingly important role in the discovery of small molecules with the potential to probe therapeutically relevant biological pathways or to serve as leads for the development of new medicines. Creative approaches for evaluating DNA-encoded chemical libraries are likely to continue to leverage the ease of designing complementary DNA strands, the extremely high sensitivity afforded by DNA amplification, and the remarkable efficiency of modern high-throughput DNA sequencing to facilitate the increasingly

efficient and applicable discovery of novel molecular interactions.

Previous work in the Liu group identified inhibitors of Src kinase [43-44] and insulin-degrading enzyme [45] from our 13,800-member DNA-templated macrocycle library. This work validated our ability to identify potent probe molecules from our DNA-templated libraries, as well as explore the biology of these protein targets through chemical means. Chapter 2 of this thesis focuses on further efforts to discover bioactive compounds using *in vitro* selections on DNA-templated macrocycle libraries. I describe efforts to identify new inhibitors of IDE from a library of 256,000 macrocycles, thus validating this library as a source of new bioactive compounds. In addition, I summarize and highlight efforts to discover compounds from both our 13,800- and 256,000-member libraries that can bind proteins such as Cas9, PCSK9, and the BAF complex through *in vitro* selections.

Chapter 3 of this dissertation describes our use of the IDUP system to simultaneously evaluate all possible protein-ligand interactions out of combined libraries of DNA-tagged proteins and DNA-encoded small molecules. Not only were we successful in recapitulating known binding interactions in this assay format, I also discovered a previously unknown covalent inhibitor, ethacrynic acid, of the human protein kinase MAP2K6. I further probed the mechanistic basis of this binding interaction and showed that inhibition is due in part to ethacrynic acid's ability to alkylate a nonconserved cysteine residue in MAP2K6. These results are illustrative of the potential of unbiased *in vitro* binding selections to uncover bioactive molecules with novel modes of protein target engagement.

1.6 References

1. Hansen, M. H. *et al.* A yoctoliter-scale DNA reactor for small-molecule evolution. *J. Am. Chem. Soc.* **131**, 1322–1327 (2009).
2. Tse, B. N., Snyder, T. M., Shen, Y. & Liu, D. R. Translation of DNA into a library of 13 000 synthetic small-molecule macrocycles suitable for *in vitro* selection. *J. Am. Chem. Soc.* **130**, 15611–15626 (2008).
3. Weisinger, R. M., Wrenn, S. J. & Harbury, P. B. Highly parallel translation of DNA sequences into small molecules. *PLoS ONE* **7**, e28056 (2012).
4. Clark, M. A. *et al.* Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nat. Chem. Biol.* **5**, 647–654 (2009).
5. He, Y. & Liu, D. R. A sequential strand-displacement strategy enables efficient six-step DNA-templated synthesis. *J. Am. Chem. Soc.* **133**, 9972–9975 (2011).
6. He, Y. & Liu, D. R. Autonomous multistep organic synthesis in a single isothermal solution mediated by a DNA walker. *Nat. Nanotechnol.* **5**, 778–782 (2010).
7. Leimbacher, M. *et al.* Discovery of small-molecule interleukin-2 inhibitors from a DNA-encoded chemical library. *Chemistry* **18**, 7729–7737 (2012).
8. Buller, F. *et al.* High-throughput sequencing for the identification of binding molecules from DNA-encoded chemical libraries. *Bioorg. Med. Chem. Lett.* **20**, 4188–4192 (2010).
9. Scheuermann, J. *et al.* DNA-encoded chemical libraries for the discovery of MMP-3 inhibitors. *Bioconjug. Chem.* **19**, 778–785 (2008).
10. Novoa, A., Machida, T., Barluenga, S., Imberty, A. & Winssinger, N. PNA-encoded synthesis (PES) of a 10 000-member hetero-glycoconjugate library and microarray analysis of diverse lectins. *ChemBioChem* **15**, 2058–2065 (2014).
11. Kawakami, T. *et al.* *In vitro* selection of multiple libraries created by genetic code reprogramming to discover macrocyclic peptides that antagonize VEGFR2 activity in living cells. *ACS Chem. Biol.* **8**, 1205–1214 (2013).
12. Schlippe, Y. V. G., Hartman, M. C. T., Josephson, K. & Szostak, J. W. *In vitro* selection of highly modified cyclic peptides that act as tight binding inhibitors. *J. Am. Chem. Soc.* **134**, 10469–10477 (2012).
13. Li, Y., Zhao, P., Zhang, M., Zhao, X. & Li, X. Multistep DNA-templated synthesis using a universal template. *J. Am. Chem. Soc.* **135**, 17727–17730 (2013).
14. Zhu, Z. & Cuzzo, J. Review article: high-throughput affinity-based technologies for

- small-molecule drug discovery. *J. Biomol. Screen.* **14**, 1157–1164 (2009).
15. Kleiner, R. E., Dumelin, C. E. & Liu, D. R. Small-molecule discovery from DNA-encoded chemical libraries. *Chem. Soc. Rev.* **40**, 5707–5717 (2011).
 16. Brenner, S. & Lerner, R. A. Encoded combinatorial chemistry. *P. Nat. Acad. Sci. USA.* **89**, 5381–5383 (1992).
 17. Ellington, A. D. & Szostak, J. W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818 (1990).
 18. Scott, J. K. & Smith, G. P. Searching for peptide ligands with an epitope library. *Science* **249**, 386–390 (1990).
 19. Cwirla, S. E., Peters, E. A., Barrett, R. W. & Dower, W. J. Peptides on phage: a vast library of peptides for identifying ligands. *P. Nat. Acad. Sci. USA.* **87**, 6378–6382 (1990).
 20. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).
 21. Needels, M. C. *et al.* Generation and screening of an oligonucleotide-encoded synthetic peptide library. *P. Nat. Acad. Sci. USA.* **90**, 10700–10704 (1993).
 22. Nielsen, J., Brenner, S., Janda, K. Synthetic methods for the implementation of encoded combinatorial chemistry. *J. Am. Chem. Soc.* **115**, 9812–9813 (1993).
 23. Gartner, Z. J. *et al.* DNA-templated organic synthesis and selection of a library of macrocycles. *Science* **305**, 1601–1605 (2004).
 24. Melkko, S., Scheuermann, J., Dumelin, C. E. & Neri, D. Encoded self-assembling chemical libraries. *Nat. Biotechnol.* **22**, 568 (2004).
 25. Halpin, D. R. & Harbury, P. B. DNA display II. Genetic manipulation of combinatorial chemistry libraries for small-molecule evolution. *PLoS Biology* **2**, e174 (2004).
 26. MacConnell, A. B., McEnaney, P. J., Cavett, V. J. & Paegel, B. M. DNA-encoded solid-phase synthesis: encoding language design and complex oligomer library synthesis. *ACS Comb. Sci.* **17**, 518–534 (2015).
 27. Zambaldo, C., Daguer, J. P., Saarbach, J., Barluenga, S. & Winssinger, N. Screening for covalent inhibitors using DNA-display of small molecule libraries functionalized with cysteine reactive moieties. *MedChemComm* **7**, 1340–1351 (2016).
 28. Malone, M. L. & Paegel, B. M. What is a ‘DNA-compatible’ reaction? *ACS Comb. Sci.* **18**, 182–187 (2016).
 29. Luk, K.-C. & Satz, A.L. “DNA-compatible chemistry.” In *A Handbook for DNA-Encoded*

Chemistry: Theory and Applications for Exploring Chemical Space and Drug Discovery, First Edition. 67-98 (John Wiley & Sons, Inc. 2014).

30. Satz, A. L. *et al.* DNA compatible multistep synthesis and applications to DNA encoded libraries. *Bioconjug. Chem.* **26**, 1623–1632 (2015).
31. Goodnow, R. A., Dumelin, C. E. & Keefe, A. D. DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nat. Rev. Drug Discov.* **16**, 131–147 (2017).
32. Salamon, H., Klika Škopić, M., Jung, K., Bugain, O. & Brunschweiler, A. Chemical biology probes from advanced DNA-encoded libraries. *ACS Chem. Biol.* **11**, 296–307 (2016).
33. Neri, D. & Lerner, R. A. DNA-encoded chemical libraries: a selection system based on endowing organic compounds with amplifiable information. *Annu. Rev. Biochem.* **87**, 5.1-5.24 (2018).
34. Mannocci, L. *et al.* High-throughput sequencing allows the identification of binding molecules isolated from DNA-encoded chemical libraries. *P. Nat. Acad. Sci. U.S.A.* **105**, 17670–17675 (2008).
35. Keefe, A. D., Clark, M. A., Hupp, C. D., Litovchick, A. & Zhang, Y. Chemical ligation methods for the tagging of DNA-encoded chemical libraries. *Curr. Opin. Chem. Biol.* **26**, 80–88 (2015).
36. Gartner, Z. J. & Liu, D. R. The generality of DNA-templated synthesis as a basis for evolving non-natural small molecules. *J. Am. Chem. Soc.* **123**, 6961–6963 (2001).
37. Blakskjaer, P., Heitner, T. & Hansen, N. J. V. Fidelity by design: YoctoReactor and binder trap enrichment for small-molecule DNA-encoded libraries and drug discovery. *Curr. Opin. Chem. Biol.* **26**, 62–71 (2015).
38. Meng, W. *et al.* An autonomous molecular assembler for programmable chemical synthesis. *Nat. Chem.* **8**, 542 (2016).
39. Halpin, D. R. & Harbury, P. B. DNA display I. sequence-encoded routing of DNA populations. *PLOS Biology* **2**, e173 (2004).
40. Connors, W. H., Hale, S. P. & Terrett, N. K. DNA-encoded chemical libraries of macrocycles. *Curr. Opin. Chem. Biol.* **26**, 42–47 (2015).
41. Zhu, Z. *et al.* Design and application of a DNA-encoded macrocyclic peptide library. *ACS Chem. Biol.* **13**, 53–59 (2018).
42. Kollmann, C. S. *et al.* Application of encoded library technology (ELT) to a protein-protein interaction target: discovery of a potent class of integrin lymphocyte function-associated antigen 1 (LFA-1) antagonists. *Bioorgan. Med Chem.* **22**, 2353–

2365 (2014).

43. Kleiner, R. E., Dumelin, C. E., Tiu, G. C., Sakurai, K. & Liu, D. R. *In vitro* selection of a DNA-templated small-molecule library reveals a class of macrocyclic kinase inhibitors. *J. Am. Chem. Soc.* **132**, 11779–11791 (2010).
44. Georghiou, G., Kleiner, R. E., Pulkoski-Gross, M., Liu, D. R. & Seeliger, M. A. Highly specific, bisubstrate-competitive Src inhibitors from DNA-templated macrocycles. *Nat. Chem. Biol.* **8**, 366–374 (2012).
45. Maianti, J. P. *et al.* Anti-diabetic activity of insulin-degrading enzyme inhibitors mediated by multiple hormones. *Nature* **511**, 94–98 (2014).
46. Mullard, A. “DNA-Encoded drug libraries come of age.” *Nat. Biotechnol.* **34**, 450–451(2016)
47. Wrenn, S. J., Weisinger, R. M., Halpin, D. R. & Harbury, P. B. Synthetic ligands discovered by *in vitro* selection. *J. Am. Chem. Soc.* **129**, 13137–13143 (2007).
48. Cuozzo, J. W. *et al.* Discovery of a potent BTK inhibitor with a novel binding mode by using parallel selections with a DNA-encoded chemical library. *ChemBioChem* **18**, 864–871 (2017).
49. DeGroot MH, Schervish MJ: Probability and statistics. Addison- Wesley; 2012.
50. Decurtins, W. *et al.* Automated screening for small organic ligands using DNA-encoded chemical libraries. *Nat. Protoc.* **11**, 764 (2016).
51. Machutta, C. A. *et al.* Prioritizing multiple therapeutic targets in parallel using automated DNA-encoded library screening. *Nat. Commun.* **8**, 16081 (2017).
52. Hauser, A. S. *et al.* Pharmacogenomics of GPCR drug targets. *Cell* **172**, 41–54.e19 (2018).
53. Brown, D. G. *et al.* Agonists and antagonists of protease-activated receptor 2 discovered within a DNA-encoded chemical library using mutational stabilization of the target. *SLAS Discov.* 2472555217749847 (2018).
54. Ahn, S. *et al.* Allosteric ‘beta-blocker’ isolated from a DNA-encoded small molecule library. *P. Natl. Acad. Sci. U.S.A.* **114**, 1708–1713 (2017).
55. Wu, Z. *et al.* Cell-based selection expands the utility of DNA-encoded small-molecule library technology to cell surface drug targets: identification of novel antagonists of the NK3 tachykinin receptor. *ACS Comb. Sci.* **17**, 722–731 (2015).
56. McGregor, L. M., Gorin, D. J., Dumelin, C. E. & Liu, D. R. Interaction-dependent PCR: identification of ligand–target pairs from libraries of ligands and libraries of targets in a single solution-phase experiment. *J. Am. Chem. Soc.* **132**, 15522–15524 (2010).

57. McGregor, L. M., Jain, T. & Liu, D. R. Identification of ligand-target pairs from combined libraries of small molecules and unpurified protein targets in cell lysates. *J. Am. Chem. Soc.* **136**, 3264–3270 (2014).
58. Li, G. *et al.* Photoaffinity labeling of small-molecule-binding proteins by DNA-templated chemistry. *Angew. Chem. Int. Ed. Engl.* **52**, 9544–9549 (2013).
59. Zhao, P. *et al.* Selection of DNA-encoded small molecule libraries against unmodified and non-immobilized protein targets. *Angew. Chem. Int. Ed. Engl.* **53**, 10056–10059 (2014).
60. Li, G., Liu, Y., Yu, X. & Li, X. Multivalent photoaffinity probe for labeling small molecule binding proteins. *Bioconjug. Chem.* **25**, 1172–1180 (2014).
61. Scott, D. E., Coyne, A. G., Hudson, S. A. & Abell, C. Fragment-based approaches in drug discovery and chemical biology. *Biochemistry* **51**, 4990–5003 (2012).
62. Shi, B., Deng, Y., Zhao, P. & Li, X. Selecting a DNA-encoded chemical library against non-immobilized proteins using a ‘ligate-cross-link-purify’ strategy. *Bioconjug. Chem.* **28**, 2293–2301 (2017).
63. Denton, K. E. & Krusemark, C. J. Crosslinking of DNA-linked ligands to target proteins for enrichment from DNA-encoded libraries. *MedChemComm* **7**, 2020–2027 (2016).
64. Melkko, S., Zhang, Y., Dumelin, C. E., Scheuermann, J. & Neri, D. Isolation of high-affinity trypsin inhibitors from a DNA-encoded chemical library. *Angew. Chemie Int. Ed.* **119**, 4755–4758 (2007).
65. Fredriksson, S. *et al.* Multiplexed protein detection by proximity ligation for cancer biomarker validation. *Nat. Meth.* **4**, 327–329 (2007).
66. Lundberg, M., Eriksson, A., Tran, B., Assarsson, E. & Fredriksson, S. Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. *Nucleic Acids Res.* **39**, e102–e102 (2011).
67. Darmanis, S. *et al.* ProteinSeq: high-performance proteomic analyses by proximity ligation and next generation sequencing. *PLoS ONE* **6**, e25583 (2011).
68. Söderberg, O. *et al.* Direct observation of individual endogenous protein complexes *in situ* by proximity ligation. *Nat. Meth.* **3**, 995 (2006).
69. Blokzijl, A., Friedman, M., Pontén, F. & Landegren, U. Profiling protein expression and interactions: proximity ligation as a tool for personalized medicine. *J. Intern. Med.* **268**, 232–245 (2010).
70. Gustafsdottir, S. M. *et al.* Use of proximity ligation to screen for inhibitors of interactions between Vascular Endothelial Growth Factor A and its receptors. *Clin.*

- Chem.* **54**, 1218–1225 (2008).
71. Hammond, M., Nong, R. Y., Ericsson, O., Pardali, K. & Landegren, U. Profiling cellular protein complexes by proximity ligation with dual tag microarray readout. *PLoS ONE* **7**, e40405 (2012).
 72. Chen, T.-C. *et al.* Using an in situ proximity ligation assay to systematically profile endogenous protein-protein interactions in a pathway network. *J. Proteome Res.* **13**, 5339–5346 (2014).
 73. Blakskjaer, P., Heitner, T. & Hansen, N. J. V. Fidelity by design: Yoctoreactor and binder trap enrichment for small-molecule DNA-encoded libraries and drug discovery. *Curr. Opin. Chem. Biol.* **26**, 62–71 (2015).
 74. MacConnell, A. B., Price, A. K. & Paegel, B. M. An integrated microfluidic processor for DNA-encoded combinatorial library functional screening. *ACS Comb. Sci.* **19**, 181–192 (2017).
 75. Beeler, A. B., Su, S., Singleton, C. A. & Porco, J. A. Discovery of Chemical Reactions through Multidimensional Screening. *J. Am. Chem. Soc.* **129**, 1413–1419 (2007).
 76. McNally, A., Prier, C. K. & MacMillan, D. W. C. Discovery of an α -amino C-H arylation reaction using the strategy of accelerated serendipity. *Science* **334**, 1114–1117 (2011).
 77. Quinton, J. *et al.* Reaction discovery by using a sandwich immunoassay. *Angew. Chem. Int. Ed. Engl.* **51**, 6144–6148 (2012).
 78. Kanan, M. W., Rozenman, M. M., Sakurai, K., Snyder, T. M. & Liu, D. R. Reaction discovery enabled by DNA-templated synthesis and *in vitro* selection. *Nature* **431**, 545–549 (2004).
 79. Rozenman, M. M., Kanan, M. W. & Liu, D. R. Development and initial application of a hybridization-independent, DNA-encoded reaction discovery system compatible with organic solvents. *J. Am. Chem. Soc.* **129**, 14933–14938 (2007).
 80. Chen, Y., Kamlet, A. S., Steinman, J. B. & Liu, D. R. A biomolecule-compatible visible-light-induced azide reduction from a DNA-encoded reaction-discovery system. *Nat. Chem.* **3**, 146–153 (2011).
 81. Gorin, D. J., Kamlet, A. S. & Liu, D. R. Reactivity-dependent PCR: direct, solution-phase *in vitro* selection for bond formation. *J. Am. Chem. Soc.* **131**, 9189–9191 (2009).
 82. Li, Y., Zhang, M., Zhang, C. & Li, X. Detection of bond formations by DNA-programmed chemical reactions and PCR amplification. *Chem. Commun. (Camb.)* **48**, 9513–9515 (2012).

83. Cao, Y. et al. Protein detection based on small molecule-linked DNA. *Anal. Chem.* **84**, 4314–4320 (2012).
84. Wu, Z., Zhen, Z., Jiang, J.-H., Shen, G.-L. & Yu, R.-Q. Terminal protection of small-molecule-linked DNA for sensitive electrochemical detection of protein binding via selective carbon nanotube assembly. *J. Am. Chem. Soc.* **131**, 12325–12332 (2009).
85. Xu, Y. et al. Terminal protection of small molecule-linked ssDNA for label-free and sensitive fluorescent detection of folate receptor. *Talanta* **128**, 237–241 (2014).
86. Wu, Z. et al. Terminal protection of small molecule-linked DNA: a versatile biosensor platform for protein binding and gene typing assay. *Anal. Chem.* **83**, 3104–3111 (2011).
87. Sugita, R., Mie, M., Funabashi, H. & Kobatake, E. Evaluation of small ligand–protein interaction by ligation reaction with DNA-modified ligand. *Biotechnol. Lett.* **32**, 97–102 (2009).
88. Mie, M., Sugita, R., Endoh, T. & Kobatake, E. Evaluation of small ligand–protein interactions by using T7 RNA polymerase with DNA-modified ligand. *Anal. Biochem.* **405**, 109–113 (2010).
89. Shendure, J. et al. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732 (2005).
90. Gu, L. et al. Multiplex single-molecule interaction profiling of DNA-barcoded proteins. *Nature* **515**, 554–557 (2014).
91. Zhang, Z., Hejesen, C., Kjelstrup, M. B., Birkedal, V. & Gothelf, K. V. A DNA-mediated homogeneous binding assay for proteins and small molecules. *J. Am. Chem. Soc.* **136**, 11115–11120 (2014).
92. Zhu, J. et al. Protein interaction discovery using parallel analysis of translated ORFs (PLATO). *Nat. Biotechnol.* **31**, 331–334 (2013).
93. Buller, F. et al. Discovery of TNF inhibitors from a DNA-encoded chemical library based on Diels-Alder cycloaddition. *Chem. Biol.* **16**, 1075–1086 (2009).
94. Podolin, P. L. et al. *In vitro* and *in vivo* characterization of a novel soluble epoxide hydrolase inhibitor. *Prostag. Oth. Lipid Med.* **104-105**, 25–31 (2013).
95. Deng, H. et al. Discovery of highly potent and selective small molecule ADAMTS-5 inhibitors that inhibit human cartilage degradation via encoded library technology (ELT). *J. Med. Chem.* **55**, 7061–7079 (2012).

Chapter 2: *In vitro* selections on libraries of DNA-templated macrocycles

Alix I. Chan, Dmitry L. Usanov, Juan Pablo Maianti, Beverly Mok, Zhen Chen, and David R. Liu.

I performed and analyzed data from all selections on DNA encoded libraries. Experiments to validate IDE inhibitors from the new library of 256,000 macrocycles were done in close collaboration with Dmitry Usanov (selections and data analysis) and Juan Pablo Maianti (fluorogenic peptide cleavage assays). Testing of enriched hits from *in vitro* selections in biochemical or cellular assays were performed in collaboration with Holly Rees (Cas9), Beverly Mok (BE3), Zhen Chen (PCSK9), Dmitry Usanov (PCSK9) and members of Prof. Cigall Kadoch's lab (BAF).

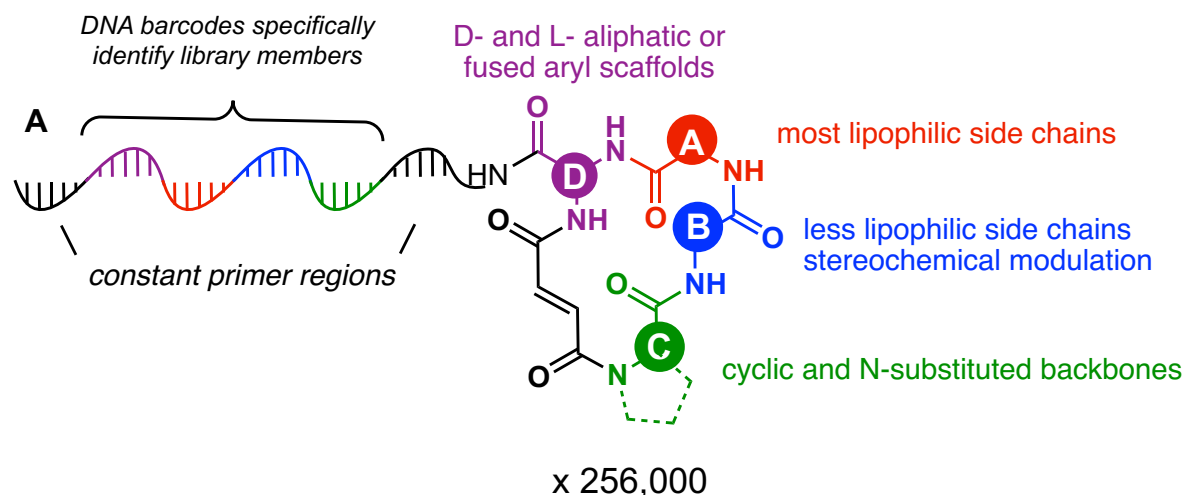
Some portions of this work are included in: Usanov, D.L.; Chan, A.I.; Maianti, J.P.; Liu, D.R. "Second-Generation DNA-Templated Macrocycle Libraries for the Discovery of Bioactive Small Molecules." *Nat. Chem.* In press (2018).

2.1 Motivations

Prior efforts in the Liu group led to the identification of new potent inhibitors of Src kinase [1-2] and insulin-degrading enzyme [3] from *in vitro* selections. These inhibitors validated our 13,824-member macrocycle library [0] as a source of new compounds that could be used as *in vivo* chemical probes [3] to elucidate biological phenomena. Thus, a major goal of my graduate work was to continue to perform *in vitro* selections on the lab's DNA encoded libraries to discover new bioactive molecules. Major efforts focused on validating that our new library of 256,000 DNA-templated macrocycles [4] could also be used as a source of bioactive molecules (Section 2.2). In addition, selections on other biomedically important proteins were performed using both the 13,824-member library previously reported [0] as well as the newer 256,000-member library [4] (Sections 2.3-2.4).

2.2 Validation of a 256,000-member macrocycle library through the discovery of new IDE inhibitors

Dmitry Usanov synthesized a new 256,000-member DNA-templated macrocycle library. This library was designed to have improved "beyond rule-of-5" properties for improved druglikeness [0]. These parameters should bias the library towards orally available compounds that could also be immediately useful as chemical probes. The general structure of each macrocyclic compound in the new library, consisting of four amino acid building blocks (A-D) linked to a DNA barcode at the 5' terminus, is shown in Figure 2.1. Varied amino acids were included in each of the building block positions in the DNA-templated syntheses (Figure 2.1A). Each macrocycle is specifically encoded by a DNA sequence that alternates constant (black) and variable barcode regions (colored to correspond to each building block position) (Figure 2.1B).



B

5' -CCCTGTACAC-NNNNNN-AAGTT-NNNNNN-ATGAT-NNNNNN-CTA-NNNN-CATCCCACTC-3'-OH

Figure 2.1 (A) General structure of DNA-templated macrocycles in the new 256,000-member library. Building blocks were chosen to maximize chemical diversity and druglikeness of the final library. (B) DNA sequence of each DNA template. Macrocycles are covalently attached to the DNA barcode at the 5' terminus through a 5' Amino Modifier 5 (Glen Research). Colored variable regions (Ns) correspond to barcodes for each of the four building block positions.

When analyzing the DNA sequences, we assigned shorthand notations for each macrocycle/barcode that are either 4 or 5 letters, e.g. ABCD or ABCXX – the first three letters each correspond to the first 3 amino acid building blocks (20 building blocks/barcodes for each), and the 4th or 4th and 5th letter correspond to the last building block (32 building blocks/barcodes). The structures of each building block and shorthand lettering are shown in Figure 2.2.

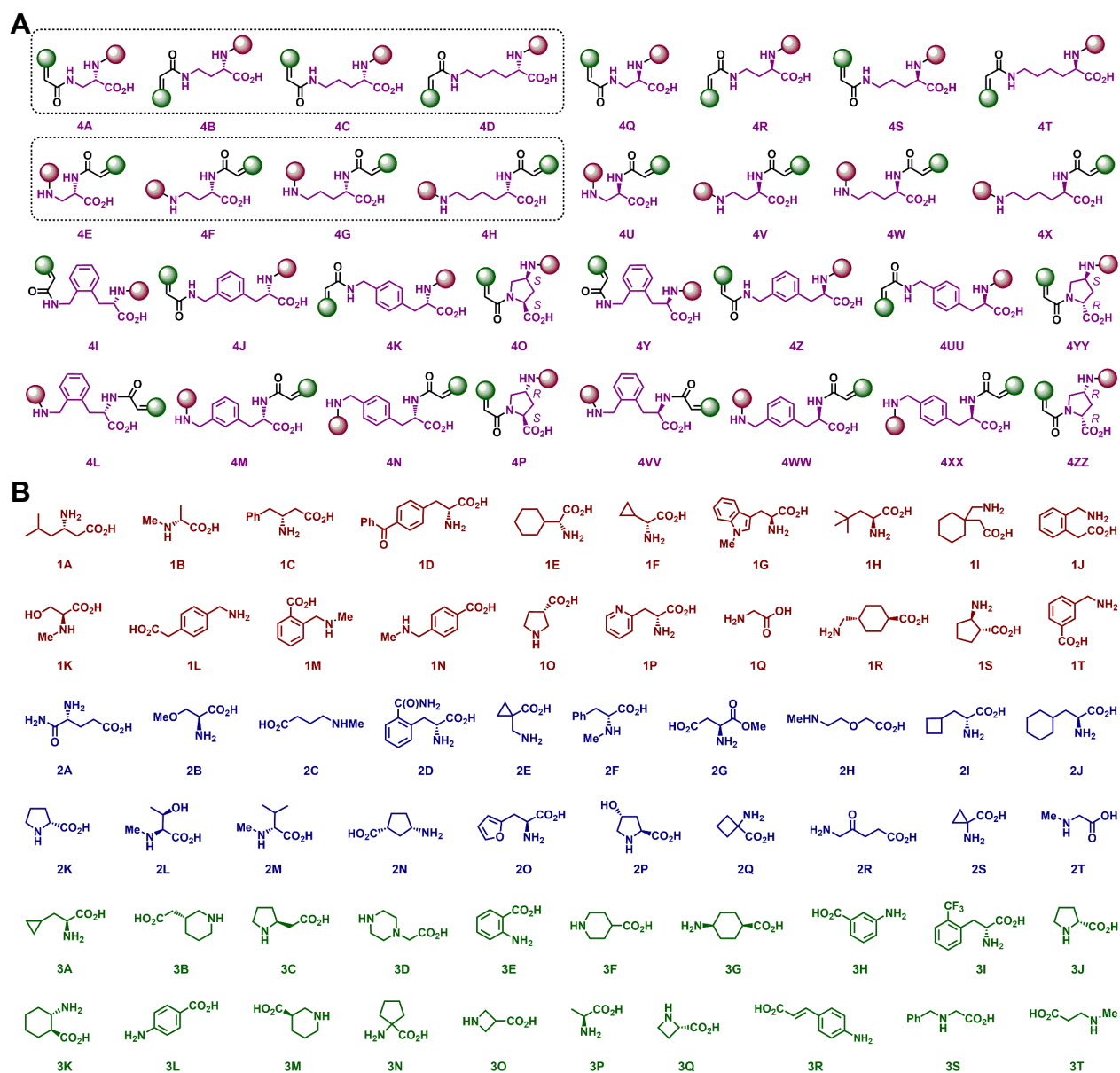


Figure 2.2. Building blocks for the second-generation DNA-templated macrocycle library. (A) Scaffold building blocks used in the second-generation library of macrocycles. Red and green spheres represent connectivity with building blocks 1 (red in (B)) and 3 (green in (B)). Scaffolds 4A-4H (dashed boxes) were used in the first-generation library. (B) Selected building blocks used in the synthesis of the 256,000-member library.

To validate our ability to identify binders of a protein target from the new 256,000-member library, we performed affinity selections against His-tag immobilized IDE. Solid-supported affinity selections for this and other targets were generally carried out by first

immobilizing recombinantly expressed, purified, affinity-tagged (e.g. polyhistidine- or GST-tagged) protein targets on an appropriate magnetic bead affinity resin. Some attempts were made to perform selections on larger protein complexes or samples purified from more complex biological samples. After loading protein onto solid support, the target is exposed to the entire DNA-templated macrocycle library. After co-incubation, nonbinding library members are removed in multiple wash steps. Theoretically, true ligands to the protein preferentially remain and can be recovered either by eluting the entire protein off the solid support using or by denaturing the protein to release bound species. The eluent can be PCR amplified and barcoded for high-throughput sequencing, which I performed on Illumina MiSeq, NextSeq, or HiSeq systems.

From prior studies, we knew that the combination of D-benzophenoyl-alanine (1D) and L-cyclohexylalanine (2J) was an effective pharmacophore for IDE ligands, and we also included one previously validated binder of IDE (known as **5b** [3]) in the library. However, when we originally performed selections on IDE using this library, we did not observe obvious enrichment of the positive control or any other IDE binders (Figure 2.3A).

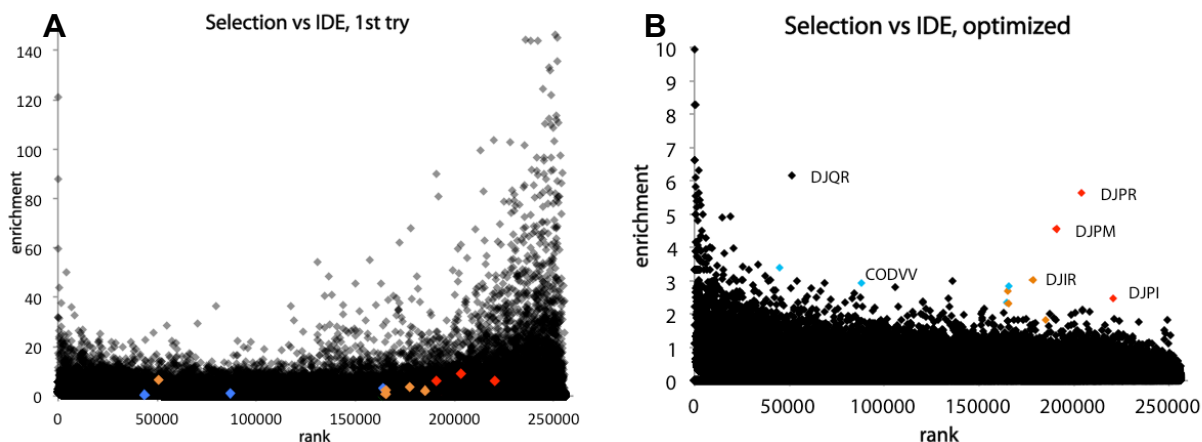


Figure 2.3 *In vitro* selections on IDE using a 256,000-member macrocycle library. Enrichment is calculated as post-selection frequency of each barcode from high throughput sequencing divided by the pre-selection frequency. Rank is determined by increasing pre-selection frequency, from left to right (1 to 256,000). (A) First attempt at *in vitro* selection with the 256,000-member library. True binders or related family members (colored points) are buried within the noise level. (B) Selection performed under optimized experimental and data processing conditions. Colored points correspond to (A). True binders (labeled) enrich clearly above the noise level.

We initially hypothesized that enrichment of the family of expected binders might be obscured by the high enrichment of covalent binders to IDE, as every member of the macrocycle library contains a potentially electrophilic fumarate moiety. However, after performing selections with this library on numerous other targets, we realized that codons corresponding to generally hydrophobic building blocks (especially those with backbone aryl rings) tended to enrich. The macrocycles with these building blocks (1J, 1L, 1M, 1N, 1T, 3E, 3H, 3L, 3R), though enriched in selection, did not display binding affinity when resynthesized and tested off-DNA. Thus, we chose to computationally filter out these ‘promiscuous’ building blocks from this and future selections.

Historically our selections protocol had included bovine serum albumin (BSA) as a reagent to block nonspecific library member + protein interactions. However, given the propensity of hydrophobic library members to enrich, and the fact that no other published

selections from other groups utilize BSA in their selections, I repeated the selection on IDE in the absence of BSA in the blocking and library incubation buffers. By excluding BSA and with the computational filtering discussed above, we were finally able to see true IDE binders in the selections data (Figure 2.3 B).

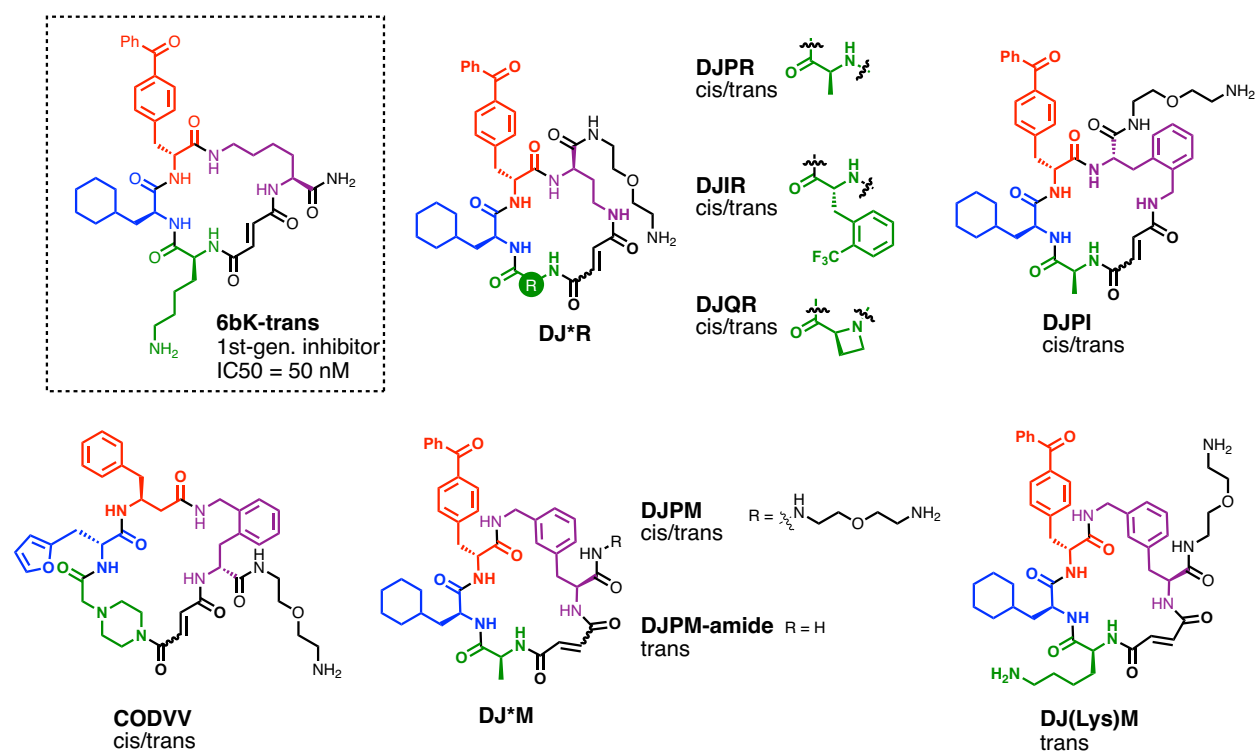


Figure 2.4. Macrocycles resynthesized and tested for IDE inhibition *in vitro*.

Macrocycle	IC ₅₀ (μM)
6bK	0.049
DJPR-trans	2.6
DJPR-cis	0.34
DJIR-trans	135
DJIR-cis	0.034
DJQR-trans	2.4
DJQR-cis	2
DJPI-trans	0.062
DJPI-cis	0.36
CODVV-trans	~4000
CODVV-cis	21
DJPM-trans	0.068
DJPM-cis	0.96
DJPM-amide	0.078
DJLysM	0.096

Table 2.1 Potencies of each macrocyclic IDE inhibitor in the fluorogenic peptide cleavage assay.

I resynthesized many macrocycles that enriched from this final IDE selection so that we could measure their inhibitory activity using a fluorogenic decapeptide cleavage assay [3,4] (Figure 2.4, Table 2.1). I synthesized many macrocycles of the **DJ**** family, as these displayed the combination of D-benzophenoyl-alanine (corresponding to barcode 1D) and L-cyclohexylalanine (barcode 2J) building blocks that we knew to be crucial for the ability of **6bK** to bind IDE. Some of these macrocycles had IC₅₀ values that were similar or even slightly better than **6bK**, such as **DJIR-cis** (IC₅₀ = 34 nM). We were particularly surprised that the cis-alkene variant of **DJIR** was a better inhibitor than the trans-alkene macrocycle (34 nM vs. 135 μM IC₅₀), as previous IDE hits had all been more potent as the trans-isomers. Given that **6bK**'s physicochemical properties were improved by installation of the 3rd-position lysine [3], I synthesized a **DJ*M** family variant, **DJLysM** that included this moiety. However, this analog was not notably more potent (IC₅₀ =96 nM) than the

compounds found from our library. One potential reason is that the exocyclic bis-aminoethyl linker also displays an aliphatic primary amine. I made the carboxamide linker variant **DJPM-amide** to test if a primary amine is helpful for bioactivity, though we did not observe a large effect on potency ($IC_{50} = 78 \text{ nM}$). We also found that although the previously optimized inhibitor **6bK** is a 20-atom macrocycle, other-sized macrocycles (21 atoms for **DJPM-trans**, $IC_{50} = 68 \text{ nM}$; 18 atoms for **DJPR-cis**, $IC_{50} = 340 \text{ nM}$) were also potent inhibitors. The alternate macrocycle sizes and conformations likely allowed the molecules to sample a variety of distinct binding modes.

In addition to **DJ**** family variants, we also tested a member of a totally distinct macrocycle family. The 24-member macrocycle **CODVV-cis** similarly displays aromatic and hydrophobic side chains at the first and second amino acid positions and indeed is an inhibitor of IDE ($IC_{50} = 21 \text{ }\mu\text{M}$) though it is notably less potent than the **DJ**** macrocycles. Nevertheless, these molecules demonstrate that new chemical series of inhibitors could be identified from *in vitro* selection on this new library of 256,000 DNA-templated macrocycles.

2.3 Selections on biomedically important proteins using the first-generation DNA templated library

In addition to validating the new library of 256,00 macrocycles, I also performed selections of the lab's ~13,800-member macrocycle library against over 40 proteins of biomedical interest (see Table 2.2). These targets are associated with a wide range of biological and disease functions (validated through genetic methods), but most lacked high quality, selective chemical probes. Protein targets were also chosen based on recommendations from other research groups or literature reports and commercial availability. Most of these selections did not yield enriched barcode families that warranted

further follow-up studies. One example of a promising selection, against Cas9, is shown below.

AfAGM1	CARD9	CrNleB	HsHat	LDHA	RNAP elo
AfGNA1	Cas9	CSF1R	HsOGT	Mgat5	RNAP holo
AfNMT1	Cas9 + gRNA	CTSB	HTRA1	MICB	RNAP holo-Bt
AfRho1	CDB456	DnaK	HXK2	PARP-1	TNFRSF4
AfUAP1	ClpB	EED	IKBKE	PatoxG	TNFSF12
Bcl11A	ClpB + ATP	GacA	IKKB	PD-L1	USP2a
Beclin-1	Cre	HIPK2	IL-17RA	RNAP core	USP9x

Table 2.2. Protein targets subjected to *in vitro* selection against a 13,800-member macrocycle library. These targets are associated with a wide range of diseases/biological functions.

2.3.1 Selections on Cas9

One specific concern in performing selections on DNA-encoded libraries is the possibility of selecting for binding to the oligonucleotide barcodes rather than to the small molecule which the DNA barcodes. However, proteins that naturally bind DNA, such as transcription factors, often have large biomedical relevance [0] so it would be worthwhile to be able to select for binders to targets in such protein classes. Recently, genome editing proteins such as Cas9 (which necessarily bind nucleic acids) have reached prominence as tools for genetic manipulation *in vitro* and *in vivo* [7]. Because of our lab's expertise in using this protein, and the potential benefits of a small-molecule controllable gene editing tool, I performed selections using the 13,800 member library against recombinant Cas9 protein.

Our standard selection protocol included the incubation of target protein with a large excess of yeast RNA to block nonspecific nucleic acid interactions. I tested whether this strategy is sufficient to block protein + oligonucleotide barcode interactions by performing *in vitro* affinity selections against Cas9, both with and without its native guide RNA (gRNA) substrate precomplexed. The selections against Cas9 showed enrichment for a

family of barcode sequences regardless of the inclusion (Figure 2.4A) or exclusion (Figure 2.4B) of the native gRNA substrate.

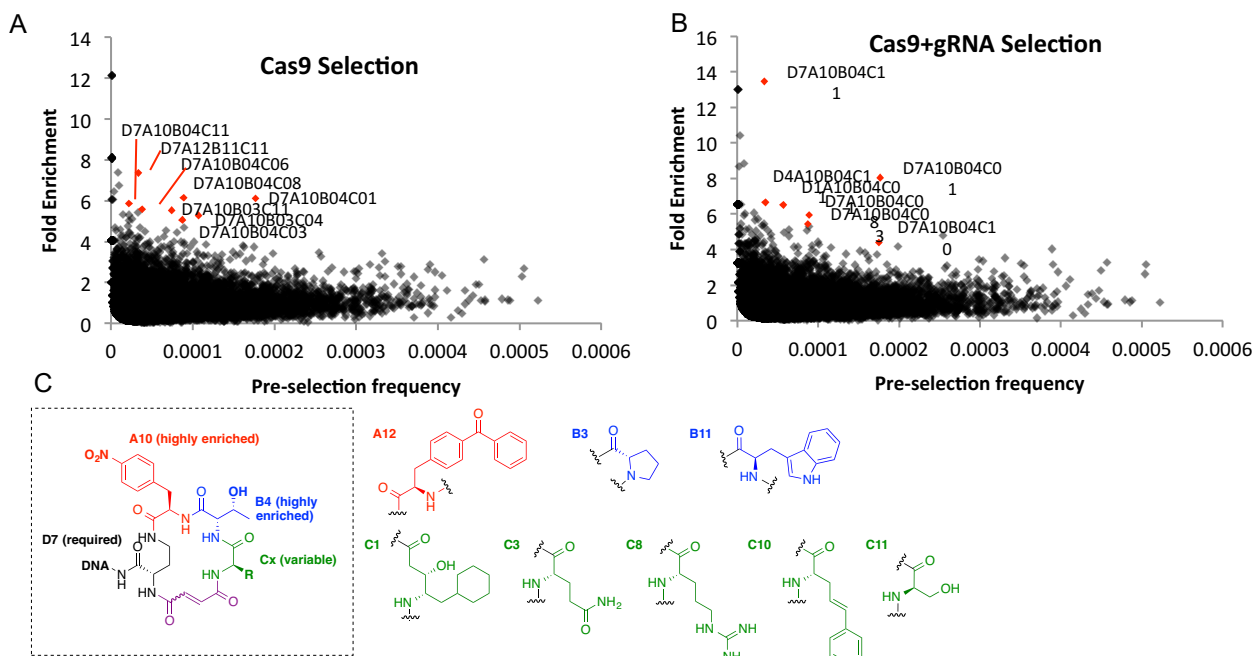


Figure 2.4 (A) Selection on Cas9 alone and (B) with pre-complexed guide RNA yielded a number of highly related library members. (C) Structures corresponding to these enriched species.

The consensus sequence of the enriched barcodes has no obvious double-stranded GG motif (to mimic a PAM sequence which is required for Cas9 + gRNA binding [7]) or complementarity to the guide RNA included in one of the selection conditions. Thus, we thought the likelihood of these sequences' enrichment was not likely due to direct DNA-protein binding interactions. I synthesized a few hits from this selection (including both cis/trans isomers of **D7A10B4C8**/**D7A10B4C11**) using solid phase peptide synthesis. Weak inhibition was observed in a DNA cleavage assay and weak binding ($>10 \mu\text{M } K_D$) was measured using fluorescence polarization. However, we later tested some of these compounds against BE3, a fusion protein of Cas9 with the APOBEC1 and UGI [0] that is capable of specifically converting cytosine into thymine bases. Two molecules from the

Cas9 selection, **A10B4C3D7-cis** (Tse 2c) and **A10B4C8D7-trans** (Tse 3t), showed modest dose-dependent inhibition of this C→T conversion by BE3 *in vitro* (up to 50 % inhibition of C→T conversion). This indicates that these compounds may interfere with Cas9 binding to a target DNA site. Unfortunately, the effect size is small and not observed in cell culture assays. However, this result still validates our ability to perform selections for binders to a protein with native nucleic acid affinity from a DNA-encoded library.

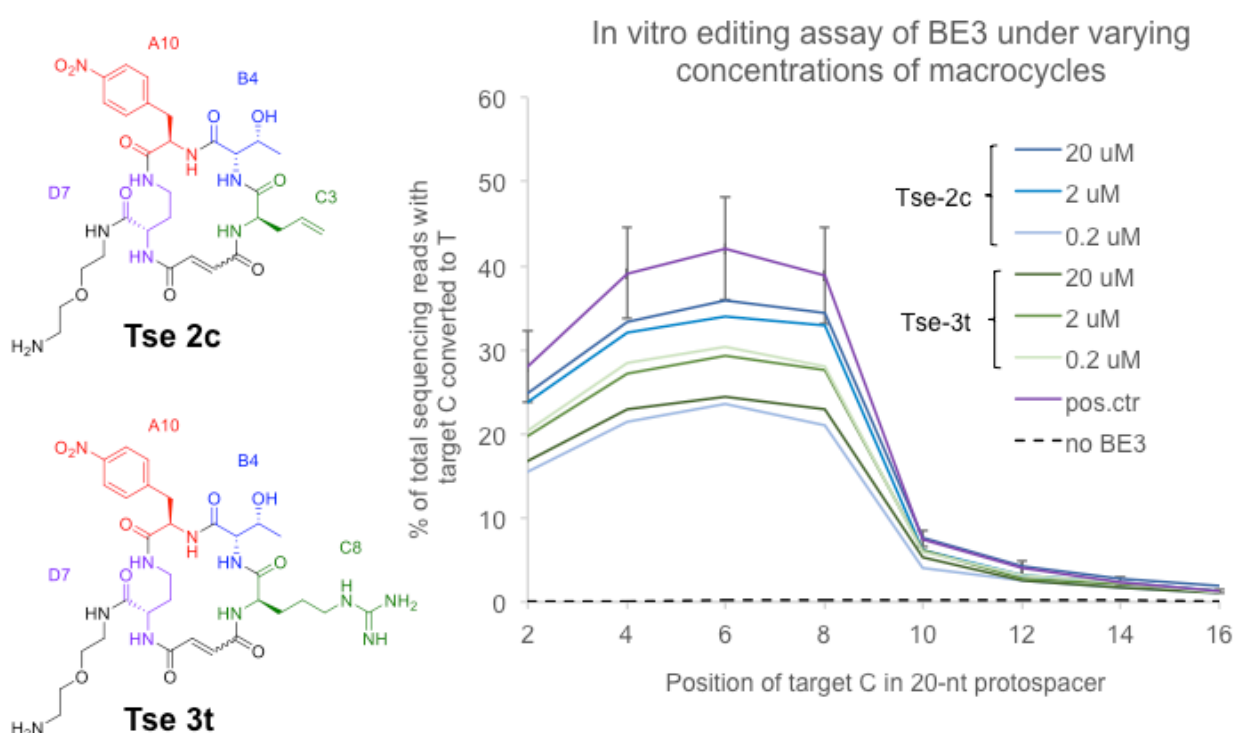


Figure 2.5. Testing macrocycles from Cas9 selection for inhibition of BE3 activity. Error bars reflect s.d. of three technical replicates performed on same day.

2.4 Selections on biomedically important proteins using the second-generation 256,000-member macrocycle library

Once we had our 2nd-generation 256,000-member macrocycle library I hand, I used it to perform *in vitro* selections on a number of protein targets. Most of these targets were provided by other academic labs. Highlighted examples of proteins are shown in Table 2.3

Targets	Biomedical relevance	Collaborating lab
Abl, Brk, PD-1/PD-L1, B7H4, Hck	cancer	Markus Seeliger
CypD	necrosis	Markus Seeliger
ApoE2/3/4	Alzheimer's	Brad Hyman
CARD9, TRIM62, CTD	IBD	Ramnik Xavier
NS5, NS3	Dengue	Priscilla Yang
Zika E	Zika	Priscilla Yang
HAT, DESC1	influenza	Donald Ingber
ClpP1P2, ClpC, ClpX	Mtb proteasome	Fred Goldberg
PrP	human prion protein	Stuart Schreiber
BAF, SSX	cancer	Cigall Kadoch
Cas9, BE3	genome editing	
PCSK9	cholesterol homeostasis	

Table 2.3. Protein targets which selections were performed on.

Selections on some of these proteins showed initially promising results. A few examples are discussed below.

2.4.1 Selections on PCSK9

The PCSK9 protein plays a crucial role in LDL regulation and has been strongly implicated as a target for therapeutic intervention for many years [0]. However, despite the evidence that knockout or downregulation of the protein leads to dramatically lower LDL levels and treat or prevent cardiovascular disease [0], no small molecule inhibitors of PCSK9 have been made into therapeutics. Thus, we performed selections against His-tagged PCSK9 using the new 256,000-member macrocycle library to search for potential

binders. We observed several families of enriched barcodes after selection (Figure 2.6) that warranted synthesis and testing of the associated macrocycles.

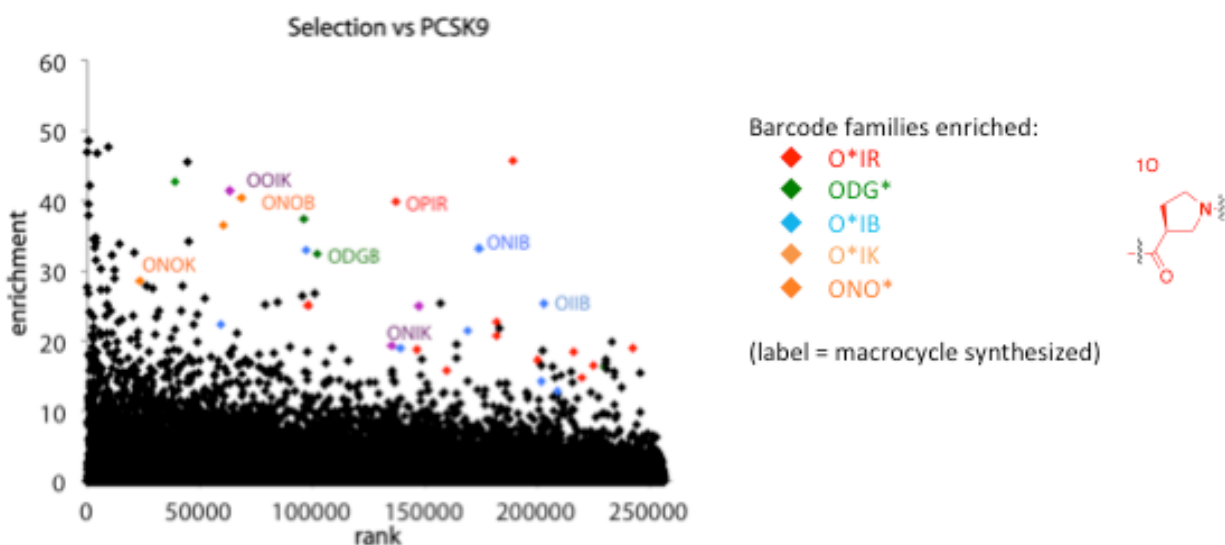


Figure 2.6. Results from *in vitro* selection of the 256,000-member macrocycle library versus PCSK9. Macrocycles with labels were synthesized and tested for binding to PCSK9 in surface plasmon resonance assays.

I synthesized a representative set of macrocycles corresponding to enriched barcode families from this selection. Compounds were tested both for direct binding to PCSK9 as well for disruption of the native PCSK9+ LDLR protein-protein interaction using surface plasmon resonance (SPR). Only one compound, **OOIK-trans**, showed bioactivity (Figure 2.7), with an $IC_{50} \sim 5 \mu M$ for disruption of PCSK9 + LDLR in SPR. However, we also observed aggregation of the compound at these concentrations. All other compounds were inactive in both binding and disruption assays at concentrations up to $\sim 30 \mu M$.

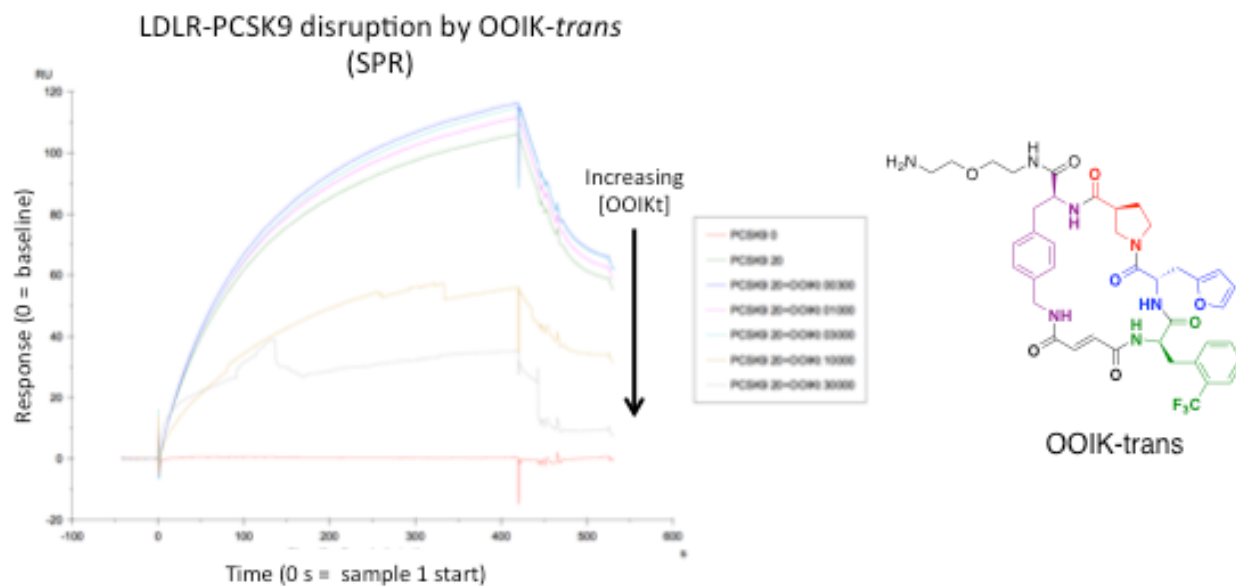


Figure 2.7. Testing the OOIK-trans macrocycle in SPR assays for disruption of the PCSK9 + LDLR protein-protein interaction. A lower SPR response indicates that compound addition interferes with protein-protein binding.

2.4.2 Selections on the BAF complex

Many proteins are commercially available in high-quality (properly folded and enzymatically active) forms, which is convenient for performing a large number of *in vitro* selections without deep biological expertise on every potential target. However, there are a number of protein targets that are not accessible in a recombinant, biologically relevant state due to the requirement for native cell-like environments, cofactors, or binding partners. These targets nonetheless can be very valuable to screen for small molecule modulators. One such target is BAF (SWI/SNF), a ~2 megadalton protein complex that is mutated in ~20% of human cancers [0]. This nucleosome remodeling complex is implicated as the single driver of synovial sarcomas [0] and malignant rhabdoid tumors [0]. Together with Prof. Cigall Kadoch's lab I have been searching for small molecule binders to

this important oncoprotein complex. Such compounds would provide useful tools for the study of the complex's roles in disease progression.

BAF is intractable to standard target-based screening procedures due to the limited amounts of complex that can be purified from tissue culture, as recombinant systems are insufficient for delivering biologically relevant forms of the complex. Due to its heterogeneity and size (even the subunits alone are difficult to purify), target-based screening would be difficult and not necessarily yield binders to native forms of the complex. Cell-based phenotypic assays are possible but would not select for compounds that directly engage the complex. *In vitro* selections using a DNA-encoded library, however, should be able to overcome these challenges, as such selections can evaluate thousands of library members while requiring only micrograms of target protein. We performed selections on entire BAF complexes isolated from cell lysate that had been isolated using an HA-tagged BAF45D subunit. These selections were compared to a pulldown on HA beads of naïve lysate (from cells not transfected with HA-tagged BAF45D) to give us a cleaner baseline for the selection (Figure 2.8). The T*RUU family of barcodes enriched strongly, and some of these macrocycles were synthesized and are awaiting further testing.

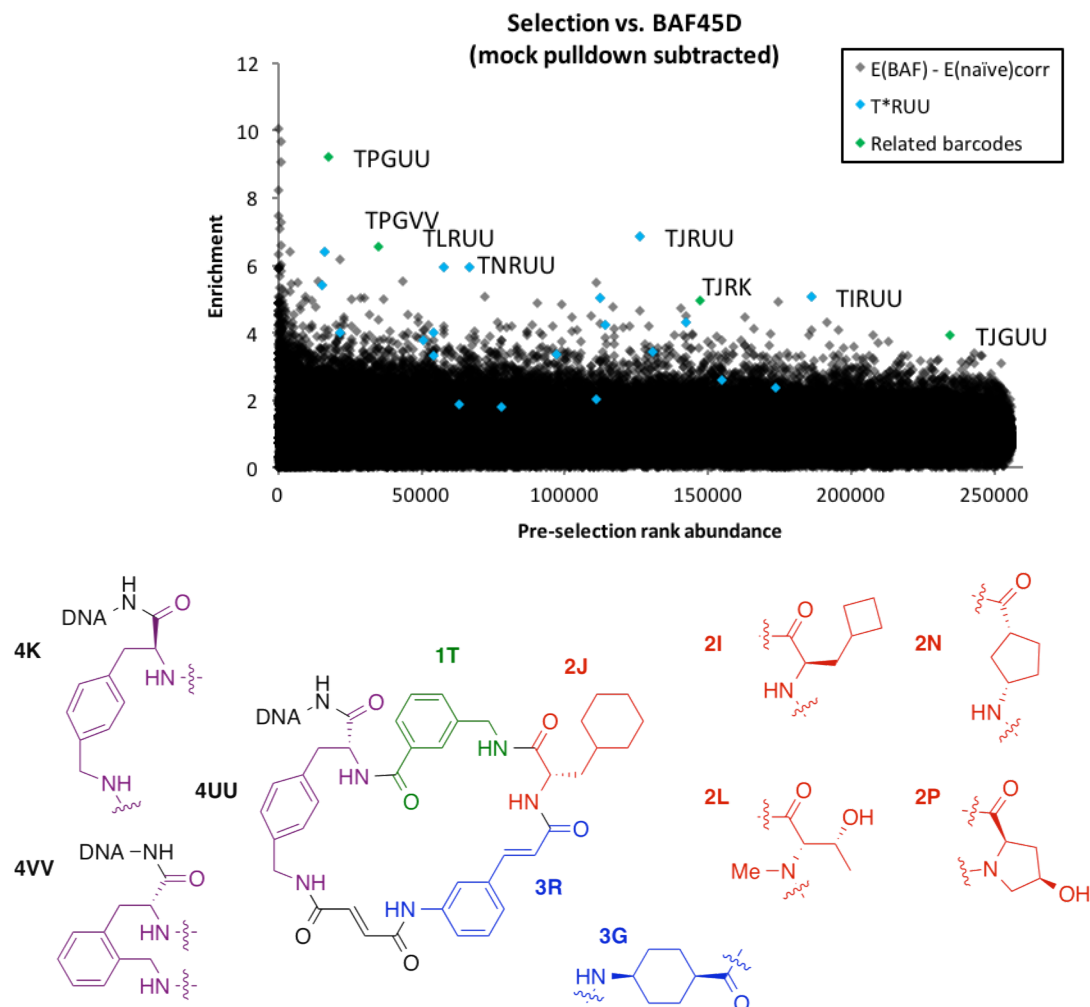


Figure 2.8. Selection of the 256,000-member macrocycle library vs. the BAF complex. Two parallel selections were run against cells transfected with HA-tagged BAF45D and naïve cells. The enrichment values from the naïve selection was subtracted from the values from the real selection to yield the plot above. Representative structures of compounds corresponding to enriched barcodes are also shown.

2.5 Discussion and outlook

Over the course of my graduate work I performed *in vitro* affinity selections on over over a hundred protein targets with our two DNA-templated macrocycle libraries. From these experiments we learned many lessons. We validated that the 2nd-generation 256,000 member macrocycle library can be used as a source of new bioactive compounds, as shown through the synthesis and assessment of IDE inhibitors. I showed that it is possible to perform selections on targets that would be difficult to interrogate in more standard target-based or phenotypic screening methods, such as proteins with native DNA affinity or complexes that are intractable to recombinant expression. Additional selections, such as on PCSK9 or SSX, have shown enriched species that warrant further study to see if these barcodes correspond to bona fide inhibitors. With regard to pursuing additional protein targets, progress has been modest, but not due to a lack of effort. Beverly Mok and Alex Peterson, two younger graduate students in the Liu group, continue to follow up leads from *in vitro* selections on this library that I performed on proteins such as BE3 and CypD. In addition, experiments are currently underway to synthesize a 3rd-generation, 640,000-member macrocycle library that we hope will be a rich source of bioactive compounds in the years to come.

2.6 General protocol for affinity selections

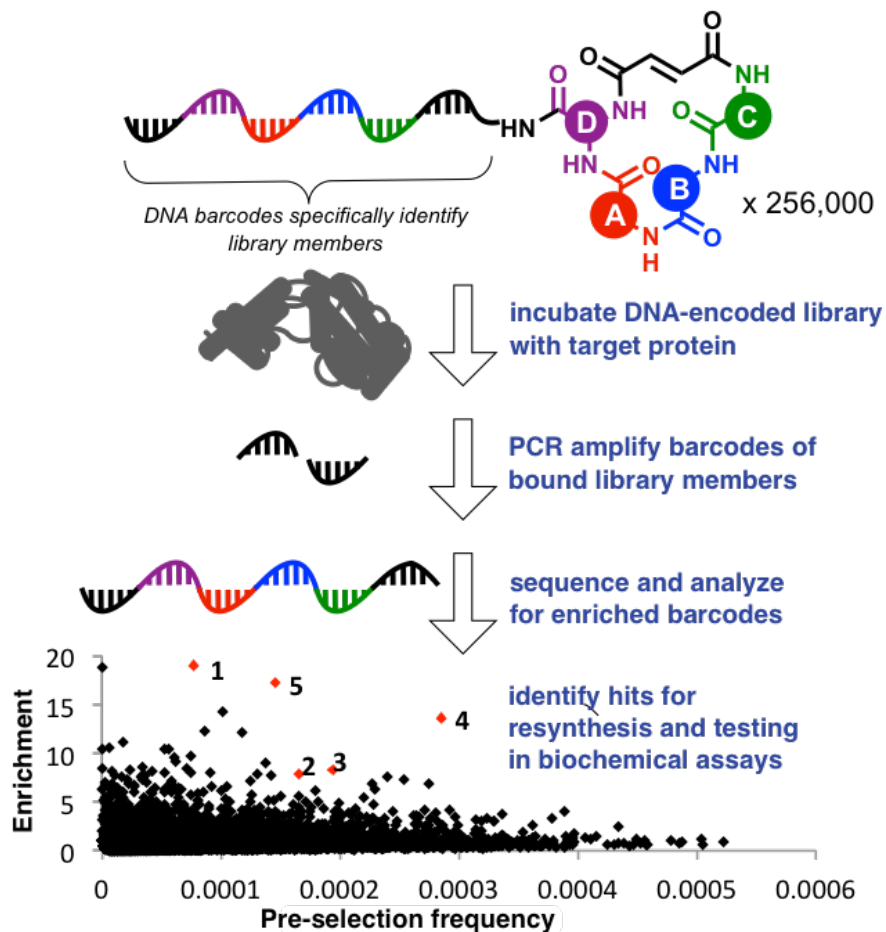


Figure 2.9 Overall workflow of affinity selections on DNA-templated macrocycle libraries for binders to immobilized proteins.

General Notes:

Eppendorf LoBind microcentrifuge tubes (1.5 mL) and MagJet magnetic rack (ThermoFisher) were used for all the operations with magnetic beads. All solutions were chilled on ice. All incubations were conducted via sideways rotation on a tiltable tube rotator, such that the top of the microcentrifuge tube never touches the solution). For the bead washing/elution steps, after each removal of the supernatant on the magnetic rack the beads were resuspended in the next portion of washing/eluting solution and transferred to a new microcentrifuge tube to minimize contamination.

Protocol:

For a His-tagged protein, 25 μ L of Dynabeads (His-Tag Isolation and Pulldown, 10103D) were washed with 2x300 μ L PBST (50 mM sodium phosphate pH 8.0, 300 mM NaCl, 0.01% Tween-20, \pm 5 mM DTT depending on whether the target required reducing buffer). 5-40 μ g of the protein was diluted into 300 μ L PBST and incubated with the beads at 4 $^{\circ}$ C for 30 min. The protein flowthrough fraction was immediately frozen at -78 $^{\circ}$ C. The beads were washed with 2x200 μ L TBST (50 mM Tris-HCl pH 8, 150 mM NaCl, 0.05% Tween-20, \pm 5 mM DTT) followed by a 15-minute incubation with the blocking solution at 4 $^{\circ}$ C (100 μ L TBST, 0.6 mg/mL yeast total RNA). The required amount of the DNA-encoded library (1 pmol for the Tse library, 20 pmol for the Usanov library) was then incubated with the beads in 50 μ L blocking solution for 60 min at 4 $^{\circ}$ C. The flow-throughs from this point on are saved for the library regeneration. The beads were washed with 3x200 TBST. Elution was accomplished by exposure of the beads to 50 μ L of PBST containing 300 mM imidazole (5 min). *Note: whereas BSA was previously used as a blocking agent in addition to yeast RNA, we found that conducting selections without BSA gives cleaner results. For targets prone to covalently bind macrocycles, much shorter incubation with the library (5 min) can be recommended.*

The eluate is directly used for qPCR with adaptor primers for HTS barcoding in order to find the maximum number of cycles within the exponential amplification range. For selections that require denaturing conditions (e.g. 0.1% SDS) for elution, detergents that interfere with PCR can be first removed using size exclusion columns (e.g Centri-Sep Spin Columns, Princeton Separations). Preparative PCR is then run with the identified number of cycles without addition of SYBR Green. The final PCR product was run on a 10%

acrylamide TBE gel (Bio-Rad) and the product band excised, crushed, and gel extracted into 1x TBE buffer. The final purified product was quantified using PicoGreen (Invitrogen), QuBit(Invitrogen) or qPCR (KAPA Biosystems). Sequencing was performed on Illumina MiSeq, NextSeq, or HiSeq systems according to manufacturer's protocols.

Sequencing and Data Analysis

I wrote custom Python scripts to analyze the fastq files from Illumina sequencing after selections. The first takes the raw sequencing data and converts DNA sequences into the shorthand notation (4 or 5 letters) for each sequencing read. It then calculated the frequency of each library member in the post-selection population (Figure 2.10). The second script (Figure 2.11) takes this output and calculates the enrichment of each barcode by comparing the post-selection frequency to the pre-selection frequency of the library.


```

1  #!/usr/bin/python
2
3  '''
4  ***REVISED CODE 07/03/2017 for Usanov library***
5
6  a few notes about this code:
7  * assumes your files fastq files are all unzipped (not fastq.gz's) and in the same folder
8  * set your filepath and directory of this folder below
9  * assumes filenames follow the form 'proteinName_S#.R1.fastq' (NextSeq output format).
10 * the 'S#' indices (e.g. S1) should be mapped to whatever variable bases
11   are in the primer in a separate .csv (barcode keys.csv), saved in the same folder as
12   your fastq data (in the form ['S1' , 'TCACT']). these should be the first bases in each
13   line of the sequence.
14 * don't put underscores in the sample filename, it messes up the regex to do the offset
15   mapping (e.g. use input_library instead of input_library)
16
17 '''
18
19 import collections      #for Counter
20 import glob            #useful for getting all files in a directory
21 import os              # for filepaths
22 import re              # for regex
23 import csv             # for converting to/from csv
24
25
26 '''***CHANGE THE FILEPATH***'''
27
28 #first, find the sequence files
29 filepath = '/Volumes/broad_liulabdata/Alix Chan/LIU DTS/Alix/20170927 NextSeq/'
30 os.chdir(filepath)    # sets the directory to look in
31
32 #dictionaries of anticodons
33 scaffolds = {"TGGG" : "A", "CAAC" : "B", "TTAA" : "C", "ACAA" : "D",
34              "TGAG" : "E", "TTCC" : "F", "TATA" : "G", "AAAT" : "H",
35              "CTAC" : "I", "TCTA" : "J", "AAAC" : "K", "AAAA" : "L",
36              "CAAA" : "M", "ACCT" : "N", "TCCT" : "O", "TTAC" : "P",
37              "TAAT" : "Q", "TAAC" : "R", "AATC" : "S", "CTAT" : "T",
38              "TGAT" : "U", "TTTT" : "V", "CTTT" : "W", "AATT" : "X",
39              "TATC" : "Y", "AACC" : "Z", "TCAC" : "UU", "CACA" : "VV",
40              "CATT" : "WW", "ACTT" : "XX", "TATT" : "YY", "TCTT" : "ZZ"}
41
42 codons_1 = {"AAAGCC" : "A", "AAGCCT" : "B", "TTTGGC" : "C", "GTTCTC" : "D",
43            "CATACG" : "E", "CTCATG" : "F", "TGTCTC" : "G", "CTACAG" : "H",
44            "CAGCTA" : "I", "CTGAGA" : "J", "AGCTCT" : "K", "TGTTGC" : "L",
45            "AAGAGC" : "M", "AGCAGA" : "N", "GATCGA" : "O", "TCAGTC" : "P",
46            "TACTGC" : "Q", "ATACGC" : "R", "GATTCC" : "S", "TGAAGC" : "T"}
47 codons_2 = {"TTCAGC" : "A", "ATCGAC" : "B", "GCAATC" : "C", "AAGTCC" : "D",
48            "ATCCGT" : "E", "ACTCGA" : "F", "TCTTGC" : "G", "CACAAG" : "H",
49            "TTAGCC" : "I", "AGTCCT" : "J", "GCATGA" : "K", "CAGACT" : "L",
50            "TTCCAG" : "M", "GGCAAT" : "N", "TCGAGA" : "O", "CTAAGG" : "P",
51            "AGGCTA" : "Q", "TCACTG" : "R", "TTGCTC" : "S", "AGCTTC" : "T"}
52 codons_3 = {"TCCGAT" : "A", "TGCACA" : "B", "GAGTCT" : "C", "CTGAAG" : "D",
53            "TCGACT" : "E", "CGTCAT" : "F", "AGGTTG" : "G", "TACGGA" : "H",

```

Figure 2.10. Representative Python code used to convert sequencing data (fastq files) to barcodes used in the Usanov library of DNA-templated macrocycles and calculate the frequency of each library member after selection.

Figure 2.10 (continued).

```
54         "GTAAGC" : "I", "CGTAGA" : "J", "TGACAC" : "K", "GTAGTG" : "L",
55         "GTTCAG" : "M", "GACTAG" : "N", "AAACCG" : "O", "AATGGG" : "P",
56         "AGAGAG" : "Q", "CGGTAA" : "R", "ACAGCA" : "S", "ACAAGG" : "T"}
57
58 # list of all codons, make a Counter object with all codons value 1 to compare enrichments
59 allcodons = []
60 for SKey in scaffolds.iterkeys():
61     for key1 in codons_1.iterkeys():
62         for key2 in codons_2.iterkeys():
63             for key3 in codons_3.iterkeys():
64                 allcodons.append(codons_1[key1] + codons_2[key2] + codons_3[key3]
65                                 + scaffolds[SKey])
66 allCodonCount = collections.Counter()
67 for codon in allcodons:
68     allCodonCount[codon]+=1
69 writer = csv.writer(open('allcodons.csv', 'wb'))
70 for barcode, count in allCodonCount.iteritems():
71     writer.writerow([barcode, count])
72
73 #open the file of the selection to be analyzed. extract out the sequences.
74 allFiles = glob.glob('*.*fastq')
75                                     # makes a list of all the fastq files in that directory
76
77 summarycsv = csv.writer(open('sequencing summary.csv', 'wb'))
78 summarycsv.writerow(['Selection', 'Total Reads', 'Valid Reads'])
79
80 with open(filepath + 'barcode keys.csv','rU') as csvfile:
81                                     # rU prevents the universal newline error
82     keyDict = {}
83     keys = csv.reader(csvfile, delimiter = ',')
84     for row in keys:
85         keyDict[row[0]] = row[1]
86                                     # makes a dictionary for the offsets
87
88 for filename in allFiles:
89     file = filepath + filename
90     fasta = open(file)
91     fastaList = fasta.readlines()
92     seqList = fastaList[1::4]
93                                     # opens the fasta file
94                                     # converts the file to a list
95                                     # gets every 4th line, the sequences
96
97     print 'done with getting every 4th line'
98     '''***CHANGE THIS IF FILENAME FORMAT HAS CHANGED***'''
99     findIndices = re.search('(?!<=_)ate*(?!=.R1)',filename)
100     # get the selection indices, which are the first strings between the underscores.
101     indices = findIndices.group(0)
102     # yields S# (whatever the sequencing index is) from the filename
103
104     validseqs = []
105     # offset = ''
106     offset = keyDict[indices]
107                                     # is there an offset in the first 0-5 bases?
108
109     for seq in seqList:
110         if offset == seq[0:len(offset)]:
111             # if the first bases match the offset
112             matchedSeq=re.search('GAGTGGGATG....TAG.....ATCAT.....AACTT.....GTGTACAGGG'
113                                 , seq)
```

Figure 2.10 (continued).

```
108     # changed to be 4Ns for scaffold
109     # regex that finds the library sequence within the sequences nucleotides
110     if matchedSeq is not None: # if a matching sequence is found, add it
111         template = matchedSeq.group(0)
112         # this should only find one result in matchedSeq.group, hence the 0 index
113
114         #match sequences to codons
115         if template[10:14] in scaffolds.keys():
116             #check if all 4 codons are valid
117             if template[17:23] in codons_1.keys():
118                 if template[28:34] in codons_2.keys():
119                     if template[39:45] in codons_3.keys():
120                         scaff = scaffolds[template[10:14]]
121                         #all 4 are valid, so start mapping
122                         A = codons_1[template[17:23]]
123                         B = codons_2[template[28:34]]
124                         C = codons_3[template[39:45]]
125                         decoded = A + B + C + scaff
126                         validseqs.append(decoded + '\n')
127                         #add this sequence to the list
128
129     '''***CHANGE THIS IF FILENAME FORMAT HAS CHANGED***'''
130     findProtein = re.search('.*(?=_S)', filename)
131     #find the protein name, similar logic as for the indices
132     protein = findProtein.group(0)
133
134     print 'Selection ' + indices + ' (' + protein + ') has ' + str(len(validseqs))
135     + ' valid sequences out of ' + str(len(seqList)) + ' reads.'
136     summarycsv.writerow([protein, str(len(seqList)), str(len(validseqs))])
137
138     #now looking through these decoded sequence files and counting them
139     seqCounts = collections.Counter()
140     for seq in validseqs:
141         seqCounts[seq.rstrip()] +=1
142         #rstrip() to get rid of the newline \n to make keys match
143     totCount = seqCounts + allCodonCount
144     #add one to each so that we have all codons represented, and don't divide by 0
145     for count in totCount.iteritems():
146         totCount[count[0]] = float(count[1])/float(len(validseqs))
147         #normalize to that selection
148
149     counterFileName = indices + '_' + protein + '_seqCounts.csv'
150     #export to csv
151     writer = csv.writer(open(counterFileName, 'wb'))
152     writer.writerow([barcode, count, seqCounts[barcode]])
```

```

1  #!/usr/bin/python
2
3  '''
4  this code will calculate enrichments. Put all of the *seqCounts.csv files from the other
5  code into a new folder.
6  Set the input lib and file path
7  '''
8  import glob
9  import csv
10 import os
11 import re
12
13 #first, find the sequence files
14 filepath = '/Volumes/broad_liulabdata/Alix Chan/LIU DTS/Alix/20170927 NextSeq/'
15 os.chdir(filepath) # sets the directory to look in
16 inputLibFile = 'reg_lib_Counts.csv' #set the filepath for the input library
17
18 with open(filepath + inputLibFile,'rU') as csvfile:
19     inputDict = {}
20     keys = csv.reader(csvfile, delimiter = ',')
21     for row in keys:
22         inputDict[row[0]] = [row[1],row[2]] # makes a dictionary for the input enrichments
23
24 '''change input text string if not input'''
25 allFiles = [i for i in glob.glob('*.*csv') if 'seqCounts' in i and 'reg' not in i]
26 # makes a list of all the seq Counts files in that directory except for the input lib
27 #print allFiles
28
29 for file in allFiles:
30     with open(filepath + file,'rU') as csvfile:
31         selectDict = {}
32         keys = csv.reader(csvfile, delimiter = ',')
33         for row in keys:
34             selectDict[row[0]] = [row[1],row[2]]
35             # makes a dictionary for the {barcode: selection enrichments, seq counts}
36
37     enrichCalc = {}
38     for key in selectDict.keys():
39         enrichCalc[key] = float(selectDict[key][0])/(float(inputDict[key][0]))
40
41     findProtein = re.search('(?!=[0-9]).*(?!=_seqCounts)', file)
42         #find the protein name, similar logic as for the indices
43     protein = findProtein.group(0)
44
45     counterFileName = protein + '_enrichment_vs_input.csv'
46     writer = csv.writer(open(counterFileName, 'wb'))
47     writer.writerow(['Sequence', 'Scaffold', 'A Codon', 'B Codon', 'C Codon',
48                     'Raw Selection Freq', 'Raw Selection Count', 'Pre-enrich Freq',
49                     'Pre-enrich Count', 'Enrichment']) #header row
50     for barcode, enrich in enrichCalc.iteritems():
51         writer.writerow([barcode, barcode[0], barcode[1], barcode[2], barcode[3:],
52                         selectDict[barcode][0], selectDict[barcode][1], inputDict[barcode][0],
53                         inputDict[barcode][1], enrich])
54     #barcodes, split by individual codon, and total count

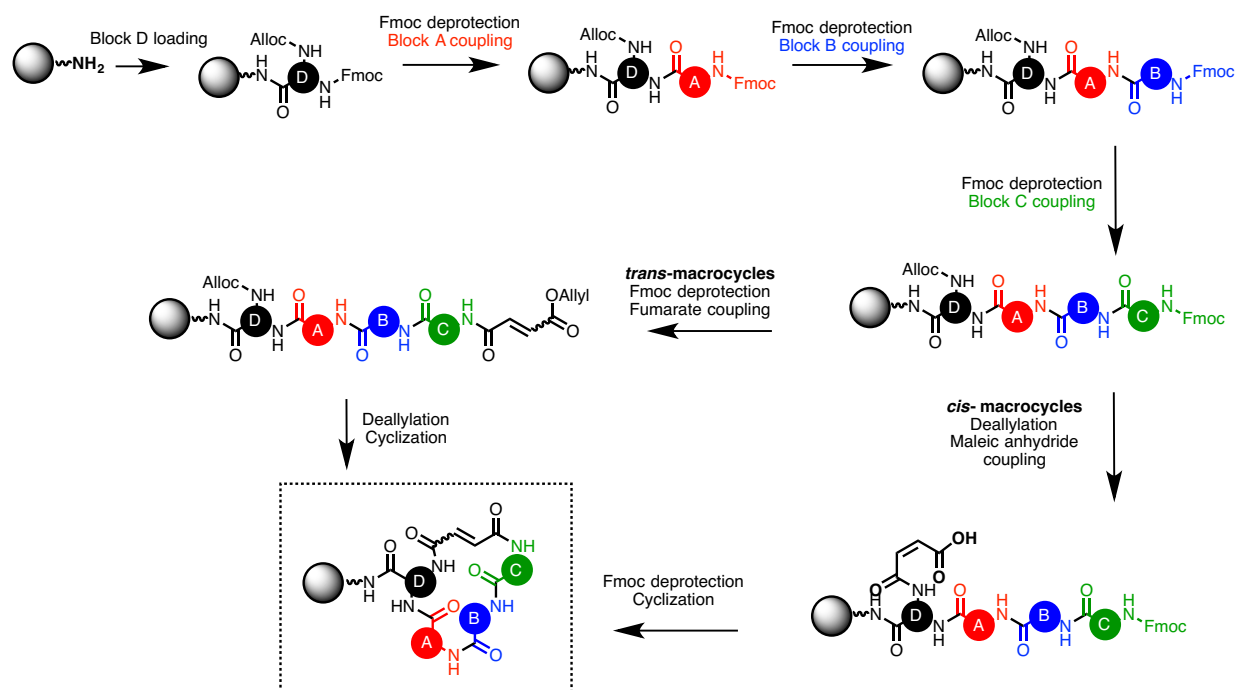
```

Figure 2.11. Representative Python code used calculate post-selection enrichment of every DTS library member. Enrichments were calculated by dividing each post-selection frequency by the corresponding barcode's pre-selection frequency.

After data processing, selections data were plotted (generally enrichment vs. pre-selection rank abundance) and visually inspected for trends in enriched species. Strong enrichments among related barcodes (e.g. sharing three of the four barcode positions) is generally indicative of a family of true binders.

2.7 Follow-up synthesis of macrocycles after DTS library selections.

After sequencing, if we identified families of enriched barcodes, the next step was to synthesize the corresponding macrocycles for testing in biochemical assays. The overall synthetic strategy is shown in Scheme 2.1.



Amino acid couplings: 5 eq. N-Fmoc-amino acid-OH, 4.75 eq HATU, 10 eq DIPEA, DMF, 1 hr rt
 Fmoc deprotections: 3x 20 % piperidine/NMP
 Alloc deprotection: 3x 0.5 eq Pd(PPh₃)₄, 40:2:1 DCM:AcOH:NMM, 3x 1 hr
 Cyclization: 5 eq. pentafluorophenyl diphenylphosphinate, 10 eq DIPEA, DMF, 3 hr rt

Scheme 2.1. General protocol for synthesis of macrocycles on solid support.

Peptide resin was swelled in ~5 mL DMF for 1 hr, agitated either with dry nitrogen bubbling or rocking in a sealed peptide synthesis vessel. Commonly used resins include

Bis-(2-aminoethyl)-ether trityl resin (Novabiochem), Rink amide MBHA resin (Novabiochem), or trityl chloride resin (Novabiochem), depending on the functional group desired at the original exocyclic DNA attachment site. Typical scale for initial selections follow-up was ~ 0.10 mmol scale per macrocycle. Commercial resin loadings typically ranged from 0.2-1 mmol/g; in general a loading of <0.5 mmol/g was optimal for minimizing dimer macrocycle formation in the final synthesis step.

In a separate flask, the diamine scaffold building block (D) (5 equivalents to theoretical resin loading) and 2-(1*H*-7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyl uronium hexafluorophosphate (HATU, 4.75 equiv.) were dissolved in anhydrous dimethylformamide (DMF, ~ 4 mL), then treat with *N,N'*-diisopropylethylamine (DIPEA, 10 equiv.) for 5 min at RT. The solution (usually yellow) was combined with the pre-swollen resin and agitated for 30-60 m. Generally, full coupling for all amino acids was achieved within 30 m, as observed by Kaiser test or microcleavage followed by mass spectrometry [14].

Following amino acid coupling, the reaction vessel was eluted and the resin washed three times with *N*-methyl- 2-pyrrolidone (NMP, ~10 vol.). Following each coupling step, Fmoc deprotection was effected with 20 % piperidine in NMP (~10 vol.) for 5 min, repeated three times, followed by washing three times with NMP (~10 vol.) and twice with DMF (~10 vol.).

The general procedure for amide coupling of building blocks A, B and C was iterative treatment of the resin with solutions of HATU-activated N^{α} -Fmoc amino acids (5 equiv.) for 30-60 minutes in DMF with agitation. The general procedure for HATU-activation was treating a solution of N^{α} -Fmoc amino acid (5 equiv.) and HATU (4.75 equiv.) in anhydrous DMF (10 vol.) with DIPEA (10 equiv.) for 5 min at RT.

For cis-alkene macrocycles, the Fmoc group is not cleaved immediately after C-amino acid coupling. (If both cis- and trans- isomers are synthesized, the resin could be split in two halves at this point.)

2.7.1 Trans-alkene (fumarate) installation:

Following the final Fmoc deprotection procedure, the α -amine of building block C was coupled with allyl fumarate monoester (10 equiv.) using activation conditions as previously described [3] with HATU (9.5 equiv.) and DIPEA (20 equiv.) in anhydrous DMF (~10 vol.). N-hydroxy-succinimide (NHS) (10 eq) may also be added to this coupling. Allyl fumarate coupling was accomplished by 30-60 m agitation, followed by washing five times with NMP (~10 vol.) and three times with CHCl₃ (~10 vol.).

2.7.2 Allyl deprotections for cis and trans macrocycles:

Simultaneous allyl ester and *N*-allyloxycarbonyl group cleavage on solid support were effected with three consecutive treatments with a solution of tetrakis(triphenylphosphine)palladium(0) (0.5 equiv. per Allyl/Alloc group) dissolved in degassed CHCl₃ containing acetic acid and *N*-methylmorpholine (40:2:1 ratio, ~20 vol.), agitated for 1-3 hr. For more acid-sensitive resins such as trityl chloride resins, 0.5 Pd(PPh₃)₄ with 12.5 eq. phenylsilane in DCM was instead used to prevent premature cleavage from the solid support. The resin was then washed twice subsequently with ~20 vol. of 5 % DIPEA in DMF, twice with a 5 % (w/v) solution of sodium diethyldithiocarbamate trihydrate in DMF (~20 vol.), twice with 5 % (w/v) solution of hydroxybenzotriazole monohydrate in DMF, and finally washed with 50 % CH₂Cl₂ in DMF and re-equilibrated with anhydrous DMF (~10 vol.).

2.7.3 Cis-alkene (maleic anhydride) coupling

Maleic anhydride (10 eq) was mixed with DIPEA (20 eq) in DMF and added to the Alloc-protected resin to couple to the newly deprotected amine of the scaffold (D) amino acid. After 1 hour, the resin was washed three times with DMF. The remaining Fmoc protecting group was deprotected using 3x 1 minute treatments with 1% DBU (1,8-Diazabicyclo[5.4.0]undec-7-ene). (It is necessary to use a non-nucleophilic base at this step, to prevent coupling to the free acid.) The resin was subsequently wash three times with a 20% DIPEA/DMF solution to salt exchange at the newly deprotected acid.

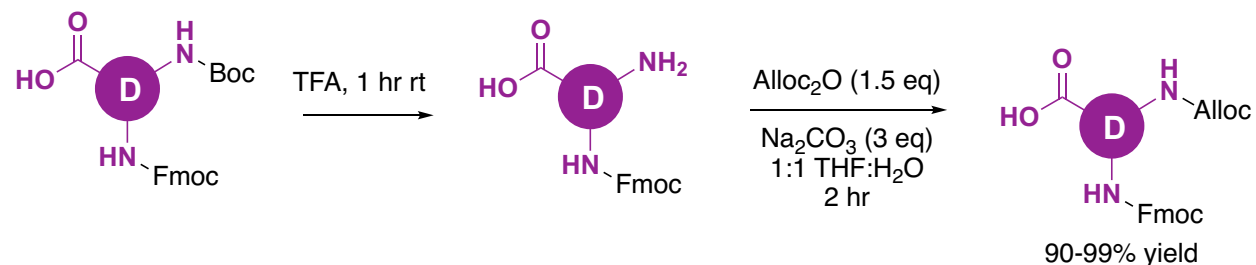
2.7.4 Cyclization

The resin was treated with pentafluorophenyl diphenylphosphinate (5 equiv.) and DIPEA (10 equiv.) in anhydrous DMF (~10 vol.) agitated for 3 hrs – overnight. Repeated treatments were occasionally necessary, especially in the case of hindered C-position amine building blocks in cis-alkene macrocycles. After full conversion to the macrocyclized product (as detected by microcleavage and MS) the resin was washed with NMP (~20 vol.) and CH₂Cl₂ (~20 vol.) and dried.

The macrocyclized product was cleaved from the resin by two 5 min treatments of the macrocycle-bound resin with 95 % TFA containing 2.5 % water and 2.5 % triisopropylsilane (~20 vol.), followed by TFA washes (~5 vol.) until the solvent ran clear (~ 2-4 washes). The TFA solution was dried to a residue under rotatory evaporation, and the peptide was precipitated into cold (-80 C), dry Et₂O. The precipitate was pelleted and the supernatant decanted. The remaining solids were dried and dissolved in a minimum volume of 3:1 DMF/water (~1-2 mL).

2.7.5 General protocol for the synthesis of N-Alloc, N-Fmoc scaffold amino acids

For scaffold ("D" position) amino acids that are not commercially available, the necessary building block can be synthesized from the corresponding N-Boc, N-Fmoc diamine material in two steps. This protocol was adapted from Demmer *et al* [15] and Ahmed *et al* [16].



Scheme 2.2. General synthetic scheme for synthesis of N-Alloc, N-Fmoc diamino scaffold building blocks.

2.5-3.0 g of N-Boc, N-Fmoc amino acid were dissolved in trifluoroacetic acid (15 mL). After stirring for one hour at RT the solvent was removed by rotary evaporation. The resulting product was dissolved in THF and water (1:1, 200 mL) with sodium carbonate (3 eq.) at 0°C. Diallyl dicarbonate (1.5 eq.) was added dropwise and the solution stirred for 2 hr at room temperature. The THF was removed by rotary evaporation. The aqueous solution was washed with diethyl ether (100 mL), then acidified (10% HCl, ~ 15mL) and extracted with ethyl acetate (3 x 100 mL). The combined organic layers were extracted with brine, dried with anhydrous sodium sulfate, and concentrated by rotary evaporation to yield the N-Alloc, N-Fmoc product as either a white solid or pale viscous oil, in 90-99% yield.

2.7.6 High-resolution mass spectrometry data for macrocycles from IDE selection.

compound	calculated	observed
DJPM-cis	836.4341	836.4372
DJPM-trans	836.4341	836.4372
DJPM-amide	749.3657	749.3664
DJPR-cis	760.4028	760.4054
DJPR-trans	760.4028	760.4067
DJLysM	806.4236	806.4258
DJQR-cis	772.4028	772.4033
DJQR-trans	772.4028	772.4058
DJIR-cis	904.4215	904.4231
DJIR-trans	904.4215	904.4254
CODVV-cis	785.3981	785.3973
CODVV-trans	785.3981	785.4003
DJPI-cis	836.4341	836.4368
DJPI-trans	836.4341	836.4356

2.8 References

1. Kleiner, R. E., Dumelin, C. E., Tiu, G. C., Sakurai, K. & Liu, D. R. *In vitro* selection of a DNA-templated small-molecule library reveals a class of macrocyclic kinase inhibitors. *J. Am. Chem. Soc.* **132**, 11779–11791 (2010).
2. Georghiou, G., Kleiner, R. E., Pulkoski-Gross, M., Liu, D. R. & Seeliger, M. A. Highly specific, bisubstrate-competitive Src inhibitors from DNA-templated macrocycles. *Nat. Chem. Biol.* **8**, 366–374 (2012).
3. Maianti, J. P. *et al.* Anti-diabetic activity of insulin-degrading enzyme inhibitors mediated by multiple hormones. *Nature* **511**, 94–98 (2014).
4. Tse, B. N., Snyder, T. M., Shen, Y. & Liu, D. R. Translation of DNA into a library of 13 000 synthetic small-molecule macrocycles suitable for *in vitro* selection. *J. Am. Chem. Soc.* **130**, 15611–15626 (2008).
5. Usanov, D.L.; Chan, A.I.; Maianti, J.P.; Liu, D.R. “Second-generation DNA-templated macrocycle libraries for the discovery of bioactive small molecules.” *Nat. Chem.* In press (2018).
6. Doak, B. C., Over, B., Giordanetto, F. & Kihlberg, J. Oral druggable space beyond the Rule of 5: insights from drugs and clinical candidates. *Chem. Biol.* **21**, 1115–1142 (2014).
7. Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–1251 (2013).
8. Mali, P., Esvelt, K.M. & Church, G.M. Cas9 as a versatile tool for engineering biology. *Nature Methods* **10**, 957–963 (2013)
9. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
10. Lambert, G., Sjouke, B., Choque, B., Kastelein, J. J. P. & Hovingh, G. K. The PCSK9 decade. *J. Lipid Res.* **53**, 2515–2524 (2012).
11. Kadoch, C. & Crabtree, G. R. Reversible disruption of mSWI/SNF (BAF) complexes by the SS18-SSX oncogenic fusion in synovial sarcoma. *Cell* **153**, 71–85 (2013).
12. Kadoch, C. *et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* **45**, 592–601 (2013).
13. Kia, S. K., Gorski, M. M., Giannakopoulos, S. & Verrijzer, C. P. SWI/SNF mediates

polycomb eviction and epigenetic reprogramming of the INK4b-ARF-INK4a locus. *Mol. Cell. Biol.* **28**, 3457–3464 (2008).

14. Coin, I., Beyermann, M. & Bienert, M. Monitoring solid phase peptide synthesis. *Nature Protocol Exchange* (2007). doi:10.1038/nprot.2007.461. Accessed Feb. 23, 2018.
15. Demmer, O., Dijkgraaf, I., Schottelius, M., Wester, H. J. & Kessler, H. Introduction of functional groups into peptides via N-alkylation. *Org. Lett.* **10**, 2015–2018 (2008).
16. Ahmed, S., Beleid, R., Sprules, T. & Kaur, K. Solid-phase synthesis and CD spectroscopic investigations of novel beta-peptides from L-aspartic acid and beta-amino-L-alanine. *Org. Lett.* **9**, 25–28 (2007).

Chapter 3: Discovery of a covalent kinase inhibitor from a DNA-encoded small-molecule library x protein library selection

Alix I. Chan, Lynn M. McGregor, Tara Jain, and David R. Liu.

Lynn McGregor planned and executed experiments to synthesize the protein/small molecule libraries and perform the initial IDUP experiment, with assistance from Tara Jain. Dr. Sunia Trauger provided invaluable assistance with mass spectrometry. I analyzed data from the IDUP selection and performed all follow-up experiments.

The majority of this work was published as: Chan, A. I.; McGregor, L.M.; Jain, T.; Liu, D.R. "Discovery of a covalent kinase inhibitor from a DNA-encoded small-molecule library x protein library selection" *J. Am. Chem. Soc.* 139, 10192-10195 (2017).

3.1 Interaction determination using unpurified proteins (IDUP)

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 [1-3], and they are typically evaluated for binding to individual protein targets by affinity enrichment using immobilized, purified protein targets [4-6]. The effectiveness of these methods is limited by artefactual enrichment of library members that bind the solid support or non-physiologically 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 [7], parallel selections under varied conditions [6], or the use of photocrosslinking probes to perform selections on unmodified proteins [8-9], 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 [10]. 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 3.1). 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 [11] in cell lysate [10] 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 [10]. In this study, we conducted a discovery-oriented IDUP experiment using libraries of DNA-barcoded proteins and small molecules to identify novel binding pairs.

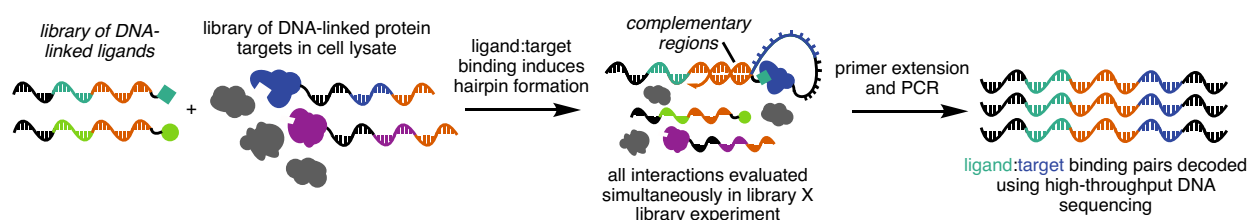


Figure 3.1. Overview of IDUP. DNA-barcoded small molecules and proteins are combined in cell lysate. Primer extension, PCR and DNA sequencing reveal the identity of protein:ligand pairs.

3.2 Construction of a DNA-barcoded protein library

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. Expression of each SNAP-kinase fusion protein was assessed Western blot (Figure 3.2). 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 (Table 3.1). 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.

Figure 3.2. Western blots of SNAP-kinase fusion proteins. Proteins are numbered as according to Table 3.1. Molecular weight markers are in the lanes marked M with weights in kDa shown for the upper left blot.

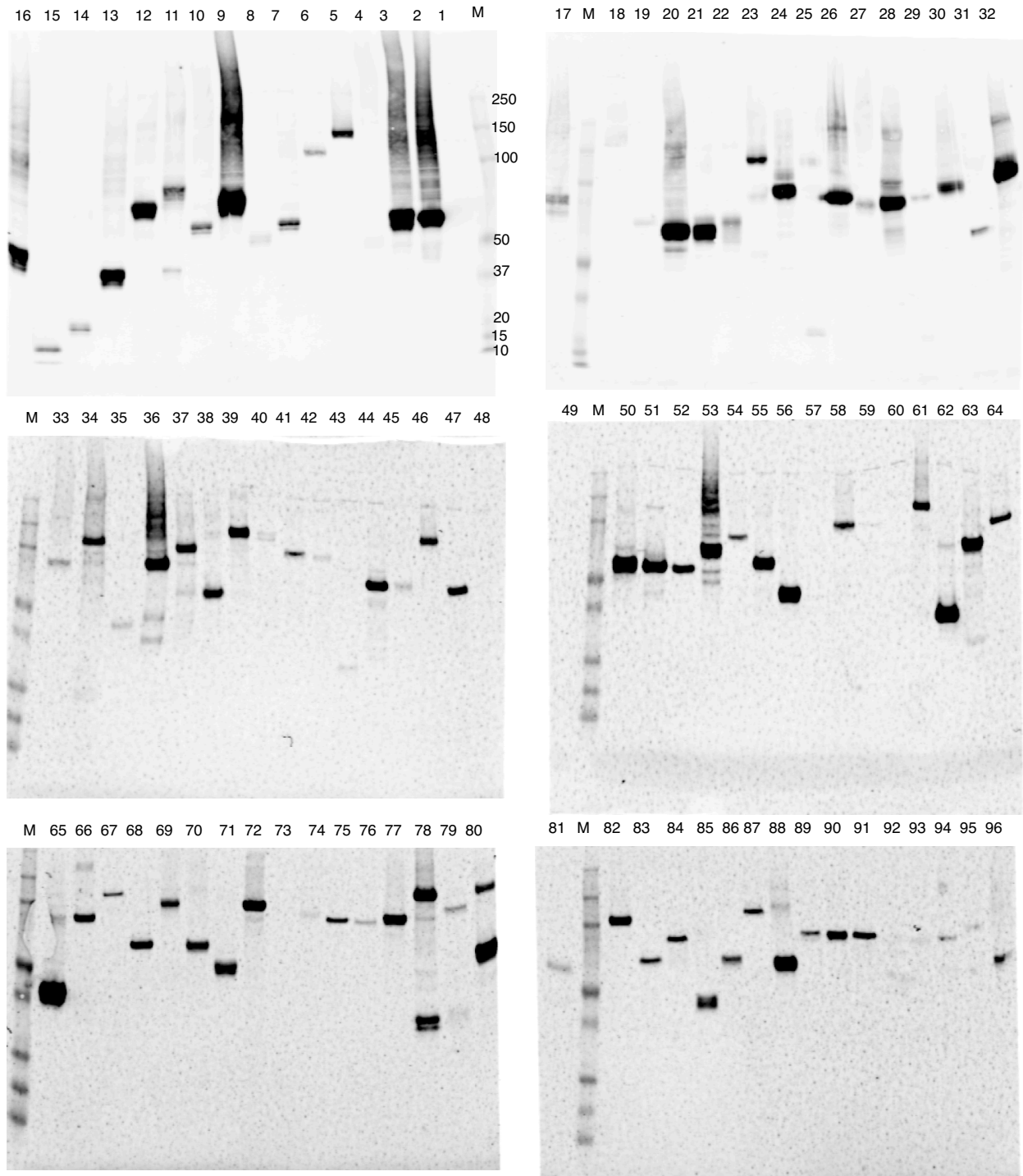


Figure 3.2 (continued).

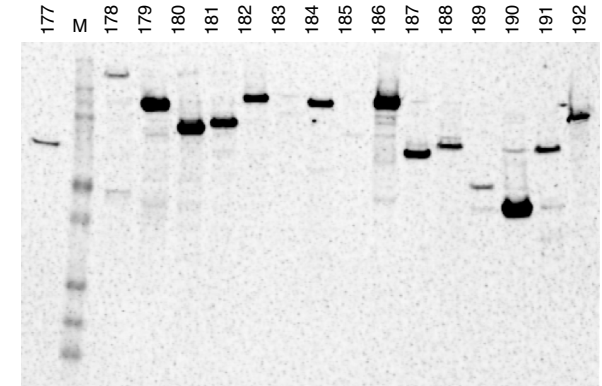
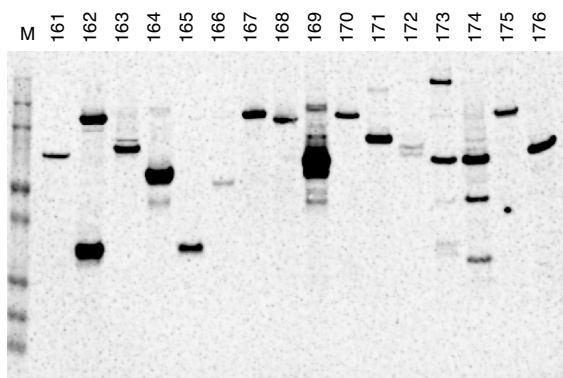
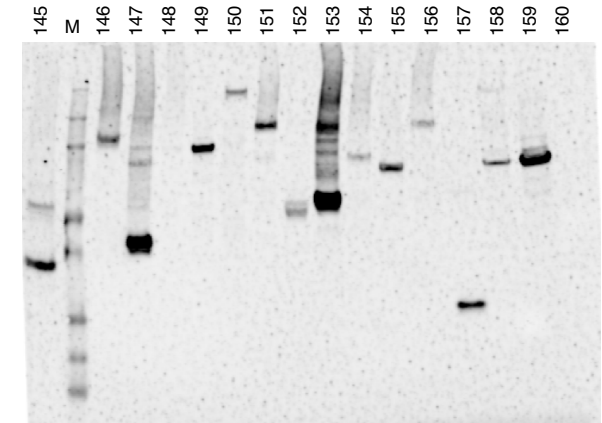
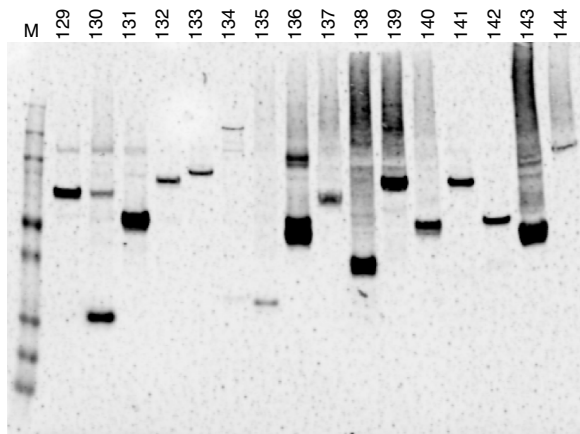
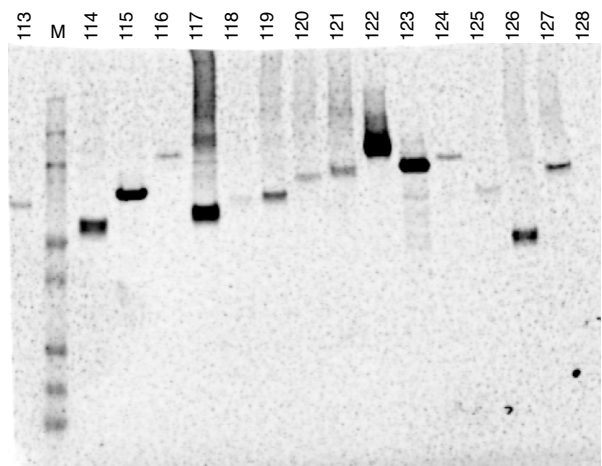
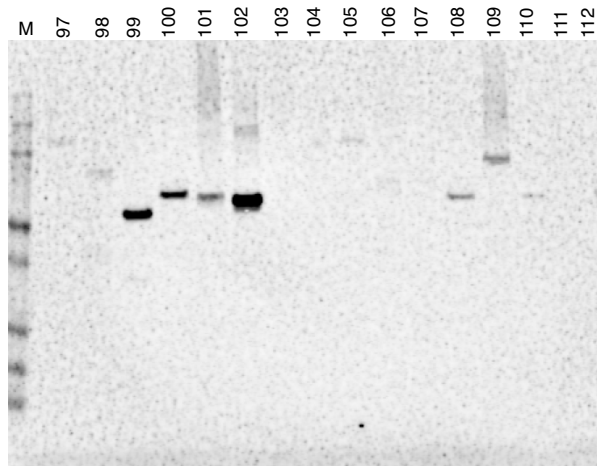


Figure 3.2 (continued).

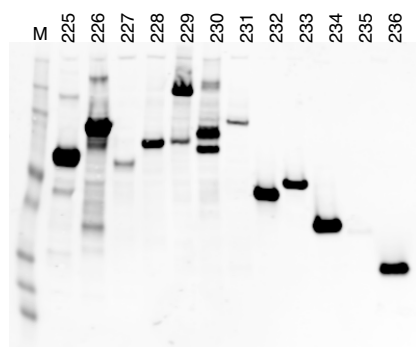
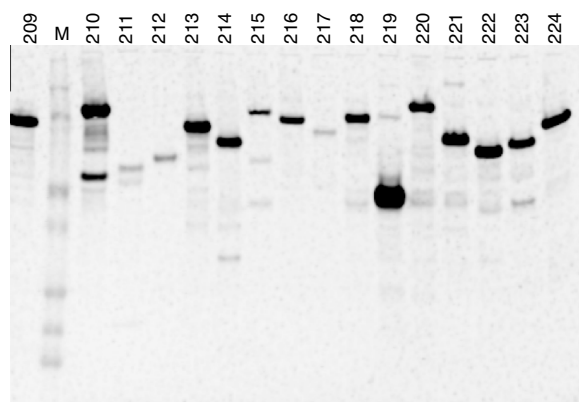
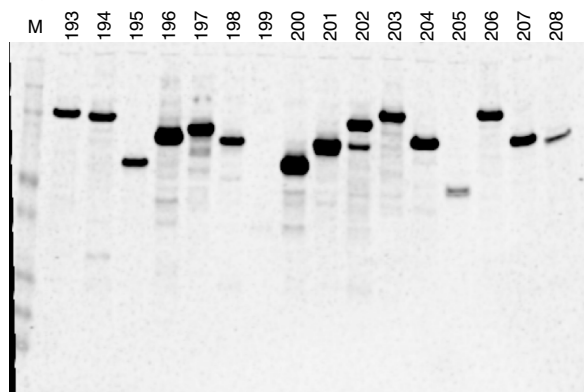


Table 3.1 (continued).

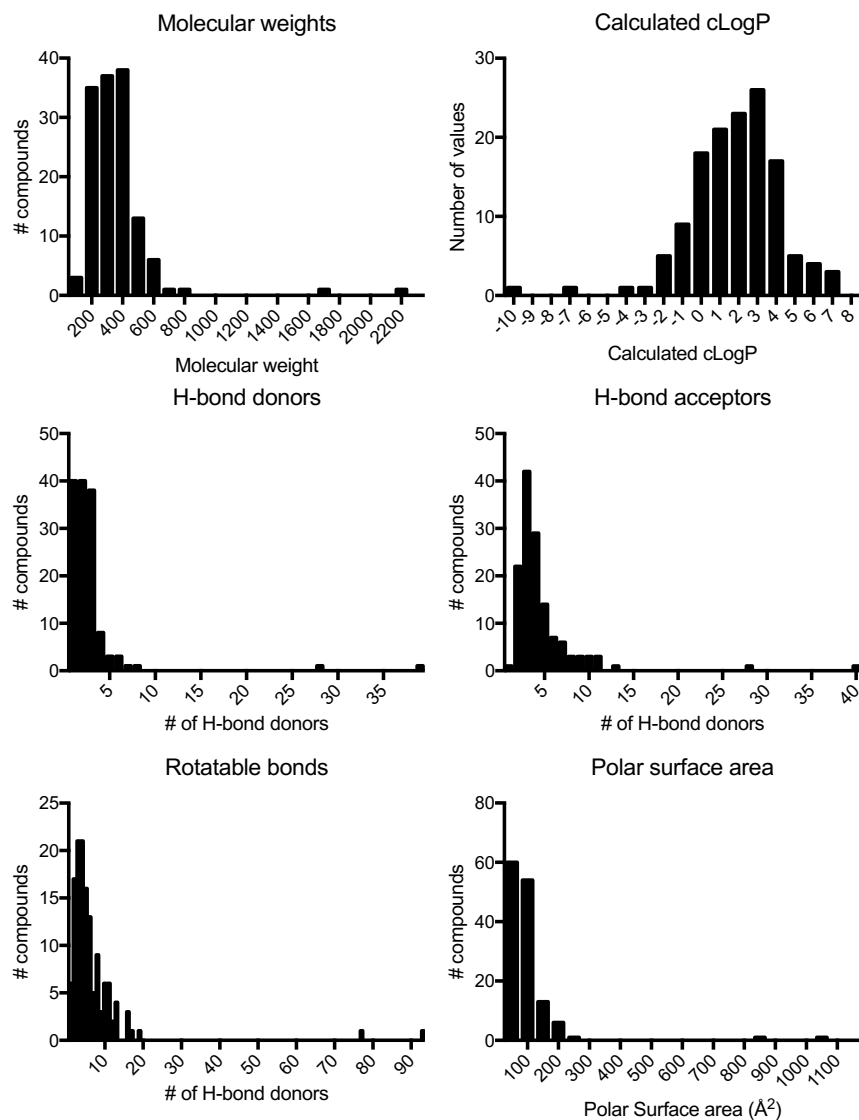
No.	Barcode	Full Sequence	Target Protein
173	PS-173 TCCGTC	ACCTGTGAGAGCTAGTCCGCTCTTGAGTGAG	ATM
174	PS-174 CGAGTC	ACCTGTGAGAGCTAGTCGAGTCTTGAGTGAG	PRKX
175	PS-175 ATTCTC	ACCTGTGAGAGCTAGTATTCTCTTGAGTGAG	MYLK3
176	PS-176 TATCTC	ACCTGTGAGAGCTAGTTATCTCTTGAGTGAG	STK32B
177	PS-177 TCGCTC	ACCTGTGAGAGCTAGTTCGCTCTTGAGTGAG	KRR1
178	PS-178 TGCCTC	ACCTGTGAGAGCTAGTTGCCTCTTGAGTGAG	MINK1
179	PS-179 TTAAGC	ACCTGTGAGAGCTAGTTACTCTTGAGTGAG	SRPK2
180	PS-180 AAAGTC	ACCTGTGAGAGCTAGTAAAGCTCTTGAGTGAG	CAMKK2
181	PS-181 TCTATC	ACCTGTGAGAGCTAGTCTATCTTGAGTGAG	TAF9
182	PS-182 CCGATC	ACCTGTGAGAGCTAGTCGGATCTTGAGTGAG	C9orf95
183	PS-183 TTTATC	ACCTGTGAGAGCTAGTTTATCTCTTGAGTGAG	GSK3A
184	PS-184 AACATC	ACCTGTGAGAGCTAGTAAACATCTTGAGTGAG	RPS8KA1
185	PS-185 ACAATC	ACCTGTGAGAGCTAGTACAATCTTGAGTGAG	DCLK2
186	PS-186 GGTTCG	ACCTGTGAGAGCTAGTGGTTGCTTGAGTGAG	TESK1
187	PS-187 GTGTGC	ACCTGTGAGAGCTAGTGTGTGCTTGAGTGAG	CDK7
188	PS-188 CAGTGC	ACCTGTGAGAGCTAGTCAGTGCCTTGAGTGAG	PFKFB4
189	PS-189 TCTGTC	ACCTGTGAGAGCTAGTTCCTGCTTGAGTGAG	PAPSS2
190	PS-190 CGATGC	ACCTGTGAGAGCTAGTCGATGCTTGAGTGAG	FUK
191	PS-191 GTTGCC	ACCTGTGAGAGCTAGTGTGGCTTGAGTGAG	PLK4
192	PS-192 CATGGC	ACCTGTGAGAGCTAGTCATGGCTTGAGTGAG	TLK1
193	PS-193 CCGGGC	ACCTGTGAGAGCTAGTCGGGCTTGAGTGAG	PINK1
194	PS-194 CGCGGC	ACCTGTGAGAGCTAGTCGGGCTTGAGTGAG	HK2
195	PS-195 CTAGGC	ACCTGTGAGAGCTAGTCTAGGCTTGAGTGAG	GALK2
196	PS-196 GAAGGC	ACCTGTGAGAGCTAGTGAAGGCTTGAGTGAG	CDK18
197	PS-197 TCTCGC	ACCTGTGAGAGCTAGTTCTCGCTTGAGTGAG	STK3
198	PS-198 CCGCGC	ACCTGTGAGAGCTAGTCGGGCTTGAGTGAG	STK40
199	PS-199 TTCCGC	ACCTGTGAGAGCTAGTTTCCGCTTGAGTGAG	ADK
200	PS-200 AACCGC	ACCTGTGAGAGCTAGTAACCGCTTGAGTGAG	MAP2K6
201	PS-201 ACACGC	ACCTGTGAGAGCTAGTACACGCTTGAGTGAG	MAPK9
202	PS-202 CGTAGC	ACCTGTGAGAGCTAGTCGTAGCTTGAGTGAG	PAK4
203	PS-203 CTGAGC	ACCTGTGAGAGCTAGTCTGAGCTTGAGTGAG	AK7
204	PS-204 GAGAGC	ACCTGTGAGAGCTAGTGAGAGCTTGAGTGAG	GK
205	PS-205 ACCAGC	ACCTGTGAGAGCTAGTACCAGCTTGAGTGAG	CDK10
206	PS-206 GGAAGC	ACCTGTGAGAGCTAGTGAAGCTTGAGTGAG	STK38
207	PS-207 ATTTCG	ACCTGTGAGAGCTAGTATTCTCTTGAGTGAG	MLH1
208	PS-208 TATTCC	ACCTGTGAGAGCTAGTTATTCTCTTGAGTGAG	MAP2K7
209	PS-209 TCGTCC	ACCTGTGAGAGCTAGTTCGCTCTTGAGTGAG	PAPSS1
210	PS-210 TGCTCC	ACCTGTGAGAGCTAGTTGCCTCTTGAGTGAG	SCYL3
211	PS-211 TTATCC	ACCTGTGAGAGCTAGTTTATCTCTTGAGTGAG	SGK494
212	PS-212 AAATCC	ACCTGTGAGAGCTAGTAAATCTCTTGAGTGAG	CDKL1
213	PS-213 TCTGCC	ACCTGTGAGAGCTAGTTCGCTCTTGAGTGAG	STK4
214	PS-214 CGGGCC	ACCTGTGAGAGCTAGTCGGGCTTGAGTGAG	IHPK3
215	PS-215 TTCGCC	ACCTGTGAGAGCTAGTTTCGCTCTTGAGTGAG	NEK8
216	PS-216 AACGCC	ACCTGTGAGAGCTAGTAAACGCTTGAGTGAG	MPP4
217	PS-217 ACAGCC	ACCTGTGAGAGCTAGTACAGCCTTGAGTGAG	STK31
218	PS-218 TGTCCC	ACCTGTGAGAGCTAGTTTGCCTTGAGTGAG	RIOK1
219	PS-219 TTGCC	ACCTGTGAGAGCTAGTTTGCCTTGAGTGAG	TK1
220	PS-220 AAGCCC	ACCTGTGAGAGCTAGTAAAGCCTTGAGTGAG	NUAK1
221	PS-221 GCCCCC	ACCTGTGAGAGCTAGTGCCCCCTTGAGTGAG	IHPK1- (2)
222	PS-222 AGACCC	ACCTGTGAGAGCTAGTAGACCCCTTGAGTGAG	MAP2K4
223	PS-223 TTTACC	ACCTGTGAGAGCTAGTTTACTCTTGAGTGAG	CHKA
224	PS-224 AATACC	ACCTGTGAGAGCTAGTAATACCTTGAGTGAG	HCK
225	PS-225 ACGACC	ACCTGTGAGAGCTAGTACGACCTTGAGTGAG	PRPSAP2
226	PS-226 AGCACC	ACCTGTGAGAGCTAGTAGACCCCTTGAGTGAG	CARM1
227	PS-227 ATAACC	ACCTGTGAGAGCTAGTATAACCTTGAGTGAG	GATA1
228	PS-228 TAAACC	ACCTGTGAGAGCTAGTTAAACCTTGAGTGAG	PRMT2
229	PS-229 TCTTAC	ACCTGTGAGAGCTAGTTCTTACTTGAGTGAG	TACC3
230	PS-230 CGGTAC	ACCTGTGAGAGCTAGTCGGTACTTGAGTGAG	TRAF2
231	PS-231 TTCTAC	ACCTGTGAGAGCTAGTTTCTACTTGAGTGAG	BIRC2
232	PS-232 AACTAC	ACCTGTGAGAGCTAGTAACTACTTGAGTGAG	CAII
233	PS-233 ACATAC	ACCTGTGAGAGCTAGTACATACTTGAGTGAG	BeIxL
234	PS-234 CGTGAC	ACCTGTGAGAGCTAGTCGTGACTTGAGTGAG	FKBP
235	PS-235 CTGGAC	ACCTGTGAGAGCTAGTCTGGACTTGAGTGAG	FRB
236	PS-236 GAGGAC	ACCTGTGAGAGCTAGTGAGGACTTGAGTGAG	SNAP

3.3 Synthesis of a DNA-encoded bioactive compound library

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 DMTMM*Cl 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

crosslinker 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) [12] in the IDUP format [10]. The final library contained an equimolar mixture of each of the 136 DNA-linked compounds (Table 3.3). The molecules span a range of chemical properties, including molecular weight (123 to 2,222 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 3.3).



	Molecular weight	cLogP	H-bond donors	H-bond acceptors	Rotatable bonds	Polar surface area (Å ²)
Min.	123.1	-9.8	1	1	1	35.3
Max.	2222.4	7.4	39	40	93	1045.2
Avg.	357.5	1.8	2.8	4.7	6.9	100.9

Figure 3.3. Summary statistics for DNA-encoded small molecule library. All values reported for compounds not conjugated to DNA barcodes as either the free acid or primary amine. CLogP and polar surface area values calculated using ChemOffice 15 (PerkinElmer Informatics).

3.4 IDUP library x library experiment to detect protein-ligand interactions

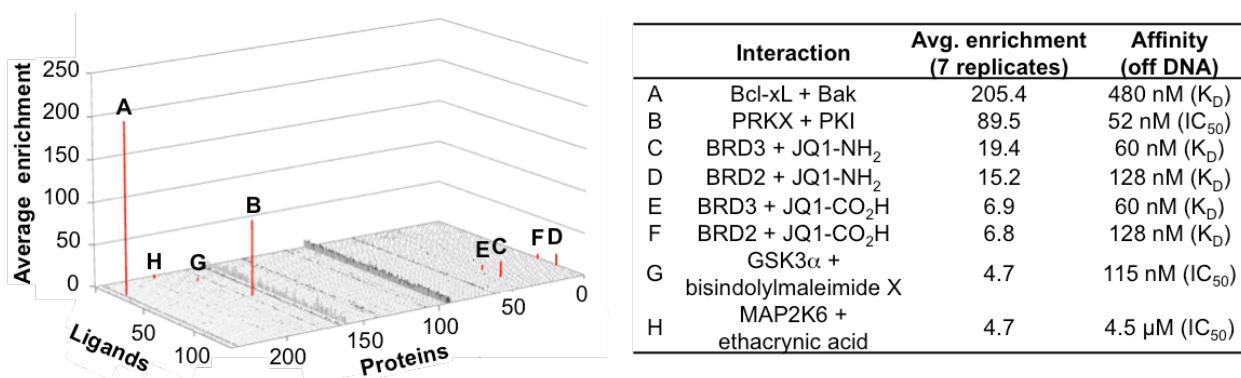


Figure 3.4. Results of seven replicate IDUP experiments of combined protein and small molecule libraries. Eight DNA barcodes corresponding to bona fide interactions (Bak:Bcl-xL, PKI:PRKX, JQ1:BRD2, JQ1:BRD3, bisX:GSK3 α , EA: MAP2K6) enriched (red) out of 32,096 possibilities. K_D and IC_{50} values 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 3.2).

We combined 2 pmol of the DNA-linked small-molecule library with the DNA-tagged protein library and performed IDUP primer extension. 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 3.4). 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 [10]. In addition, we observed high enrichment (89.5-fold) of the barcodes corresponding to PKI peptide (a cAMP-dependent kinase inhibitor [13]) and PRKX (a cAMP-dependent kinase [14]). 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 [15]) and BRD3 (15.2- or 19.4-fold, K_D = 60 nM [15]). 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. I tested 11 interactions encoded by enriched barcodes in either kinase activity or binding assays using the corresponding non-DNA tagged ligands (Table 3.2). Using Z'-LYTE assays (Invitrogen) I measured the inhibition of PRKX by PKI ($IC_{50} = 52$ nM) and GSK3 β by bisindolylmaleimide X (bisX) [16] (4.7-fold IDUP enrichment, $IC_{50} = 115$ nM). Finally, I discovered that ethacrynic acid (EA) inhibits MAP2K6 (4.7-fold IDUP enrichment, Z'-LYTE $IC_{50} = 4.5$ μ M). All six of the compounds found to be bioactive were relatively specific; we did not observe significant enrichment of any of their barcodes in combination with a large fraction of the protein library's barcodes (Figure 3.5).

	Protein target	Ligand	Avg. enrich	IC ₅₀ , nM
Interactions assayed by Z'-LYTE	PRKX	PKI: cAMP-dependent protein kinase inhibitor (5-24)	89.5	52
	SRPK2	bisindolylmaleimide X	5.85	>20000
	SRPK2	Gemifloxacin mesylate	4.98	>6700
	MAP2K6	ethacrynic acid	4.73	4500
	GSK3 α	bisindolylmaleimide X	4.68	115
	SRPK2	Zomepirac	4.61	>20000
	EEF2K	prostaglandin E1	4.50	>20000
	SRPK2	JQ1-COOH	4.16	>20000
	SRPK2	Lithocholic acid	4.12	>20000
Interactions assayed by Lanthascreen	LIMK1	sulindac	4.52	>20000
	LIMK1	desthiobiotin	4.16	>20000

Table 3.2. Interactions from library x library IDUP experiment tested in kinase assays. Interactions were either tested using the Z'-LYTE kinase activity or Lanthascreen kinase binding assay (Invitrogen). Of these, PRKX+PKI, MAP2K6+ethacrynic acid, and GSK3 α +bisX were validated as true binders, indicating that their enrichment level (~4.7) might be taken as the noise limit of the assay. No ligands tested with SRPK2 validated as true binders; this may indicate that the original experiment selected for binding to a different isoform of the kinase than was tested *in vitro*, or that there was nonspecific enrichment of the DNA barcode corresponding to SRPK2.

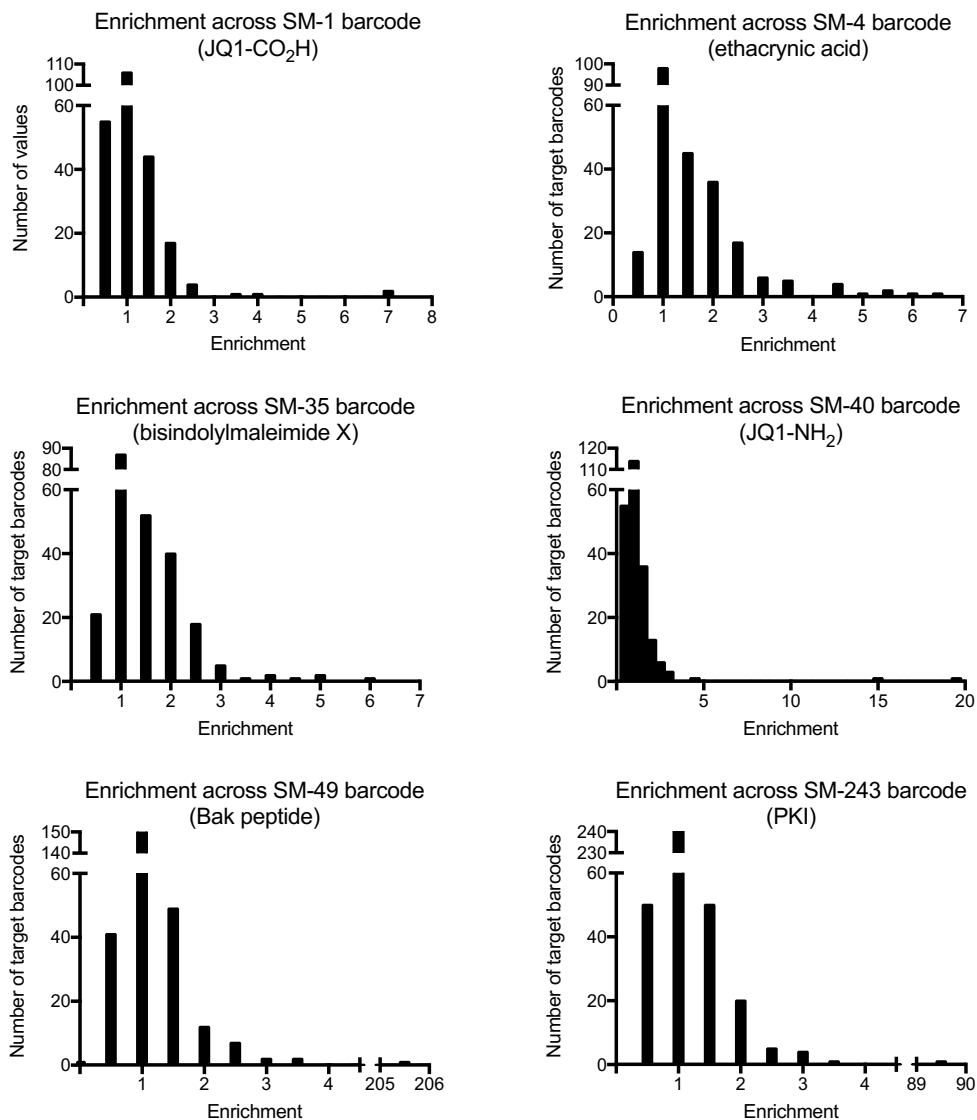


Figure 3.5. Histograms of enrichments of every target DNA barcode with DNA barcodes for ligands which were identified as having true binding partners from the IDUP assay. Target barcodes which show high enrichments across all small molecule barcodes were excluded from analysis (93-96, 162, 166). One potential reason for the signal across these targets' barcodes is related to poor expression of these targets (see, for example, the Western blots for proteins 93-96 and 166 in Figure S2). Protein barcodes 93-96 also amplified poorly in the negative control experiments, possibly due to the common CTG motif in their internal barcodes, which may have artificially inflated the calculated enrichment values (sequence counts in the control experiment are the denominator for the enrichment calculation).

3.5 Identification and biochemical characterization of ethacrynic acid as a covalent inhibitor of MAP2K6

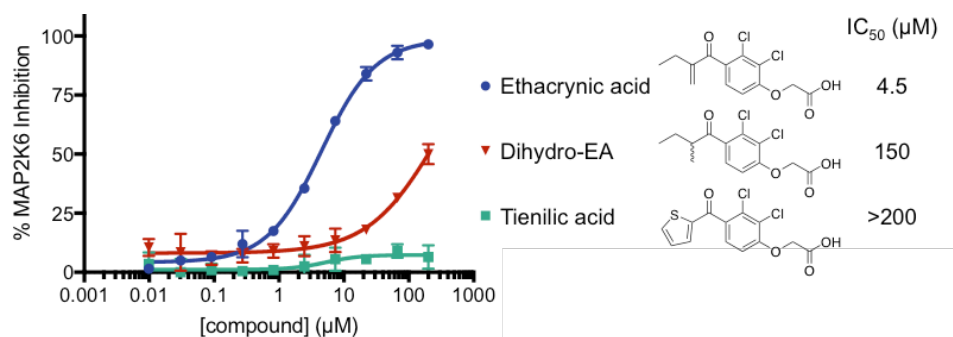


Figure 3.6. Inhibition of MAP2K6 by EA is dependent on the Michael acceptor. The non-electrophilic analogs shown are >30-fold less potent.

EA is an FDA-approved loop diuretic that inhibits the NKCC symporter [17] and has not been previously reported to inhibit any kinases. EA contains a Michael acceptor that reacts with glutathione [17-18] and EA derivatives have been previously used as covalent bromodomain inhibitors [19]. We investigated whether it inhibits MAP2K6 by forming a covalent adduct with the protein. Non-electrophilic analogs of EA (dihydro-ethacrynic acid and tienilic acid) exhibited >30-fold weaker inhibition of MAP2K6 (Figure 3.6). Incubating MAP2K6 with EA yielded a +303 adduct in the intact protein mass spectrum, consistent with covalent modification by EA (Figure 3.7). 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 3.7D), 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 3.8A). This peptide contains a single cysteine residue, and fragmentation of this peptide by tandem mass spectrometry (MALDI-TOF/TOF) confirmed that Cys38 was the site of covalent modification (Figure 3.8C).

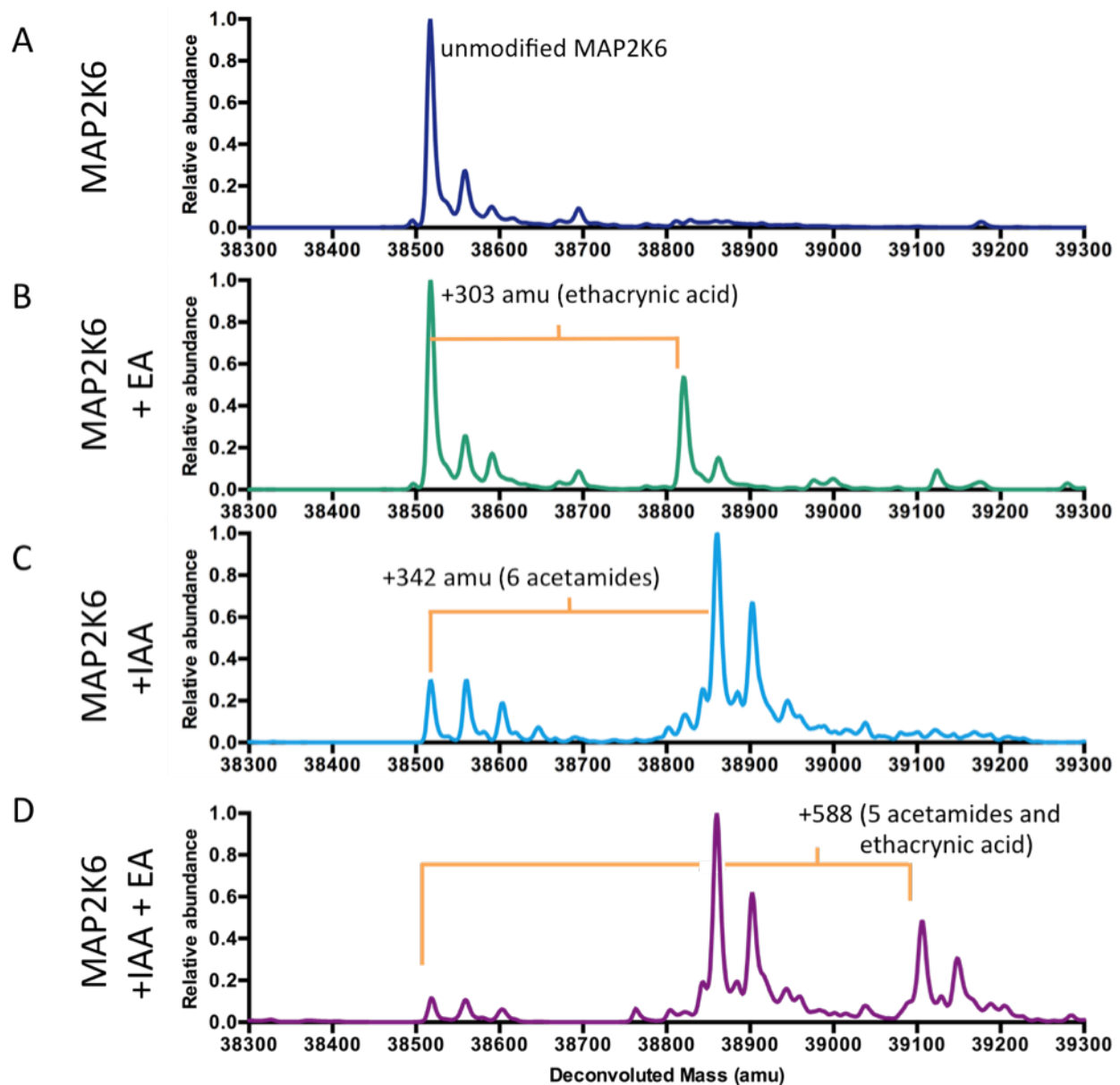


Figure 3.7. Mass spectrometry of MAP2K6 with alkylating agents. These spectra reveal a covalent adduct consistent with alkylation of MAP2K6 by ethacrynic acid (EA) at a cysteine residue. (A) Deconvoluted mass spectrum of unmodified N-His6-MAP2K6. (B) Incubation of MAP2K6 with ethacrynic acid results in a +303 Da mass shift. (C) Incubation of MAP2K6 with iodoacetamide (IAA) nonspecifically alkylates all six cysteines. (D) Incubation with ethacrynic acid and iodoacetamide results in a product that is alkylated with one ethacrynic acid and five iodoacetamides. In (B) and (D) 10 μ M MAP2K6 was incubated with 20 μ M EA for 1 hr at room temperature prior to mass spectrometry. Iodoacetamide was added to a final concentration of 15 mM in (C) and (D) for 15 minutes before mass spectrometry.

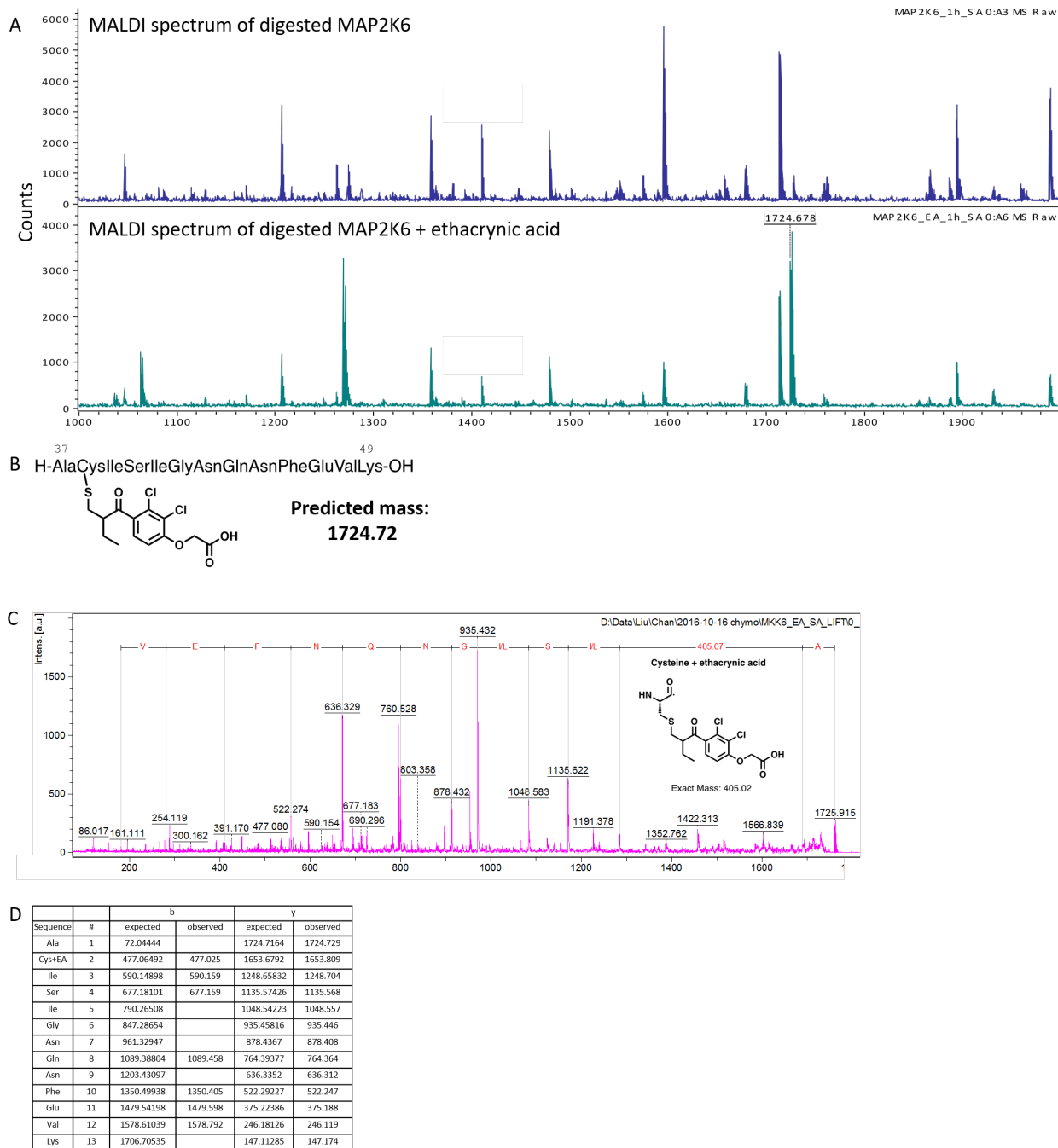


Figure 3.8. Cysteine 38 is the site of ethacrynic acid alkylation of MAP2K6. (A) MALDI-TOF spectrum of trypsin-digested MAP2K6. When MAP2K6 is incubated with ethacrynic acid prior to digestion, a 1724 Da mass fragment is observed. This fragment is absent in the non-ethacrynic acid-treated sample. (B) This mass is consistent with modification of the Ala37-Lys49 fragment of MAP2K6. (C) and (D) Tandem mass spectrometry targeting the 1724 Da peptide. The fragmentation pattern is consistent with the Ala37-Lys49 peptide being modified at Cys38. When inhibition of MAP2K6 is measured biochemically (see Figure 4) the magnitude of the effect of mutating C38 is not as large as one might expect. We suspect that the EA+MAP2K6 adduct is somewhat reversible. For example, after EA

Figure 3.8 (continued) pre-incubation we see evidence of unmodified MAP2K6 in the mass spectrum as well, as shown in Fig. S4 (although ionization is a harsher condition than the biochemical assay). However, the modified Ala37-Lys49 peptide, as shown above in Fig. S5, is still observed after denaturation and trypsin digest, indicating that reversible binding, while possible, is likely to be slow.

To better assess the mechanism of EA inhibition, we incubated with EA a constitutively active MAP2K6 mutant containing phospho-mimetic 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 3.9). Together, these results suggest that covalent modification of MAP2K6 by EA at Cys38 is partially, but not solely, responsible for kinase inhibition.

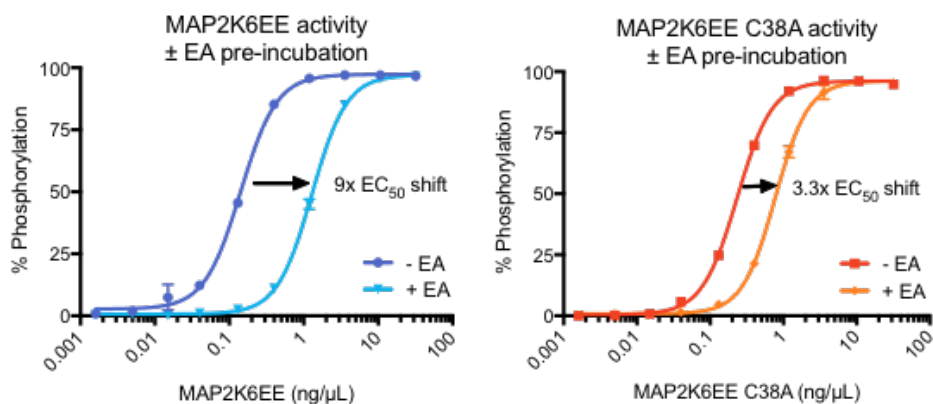


Figure 3.9. Kinases were incubated with 100 μM EA (+EA) or DMSO (-EA), then dialyzed to remove free EA. Activity was assayed as a function of kinase concentration. Ethacrynic acid inhibits MAP2K6 to a greater degree when Cys38 is present (left) than when this residue is mutated to an alanine (right).

3.6 Selectivity of ethacrynic acid for MAP2K6 over other MAP2K proteins

A member of the MAP2K family, MAP2K6 activates p38 MAP kinase in response to environmental stresses [20]. Previous cheminformatic and proteome-wide studies implicated Cys128 (in the Gatekeeper region) or Cys196 (adjacent to the DFG motif) as more accessible or reactive towards small-molecule electrophiles [21,22]. In contrast,

Cys38 is located within a non-active site region with poorly understood function [23] and is not conserved among other MAP2Ks (Figure 3.10). We confirmed that EA has higher affinity for MAP2K6 than other MAP2Ks (Figure 3.11). 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.

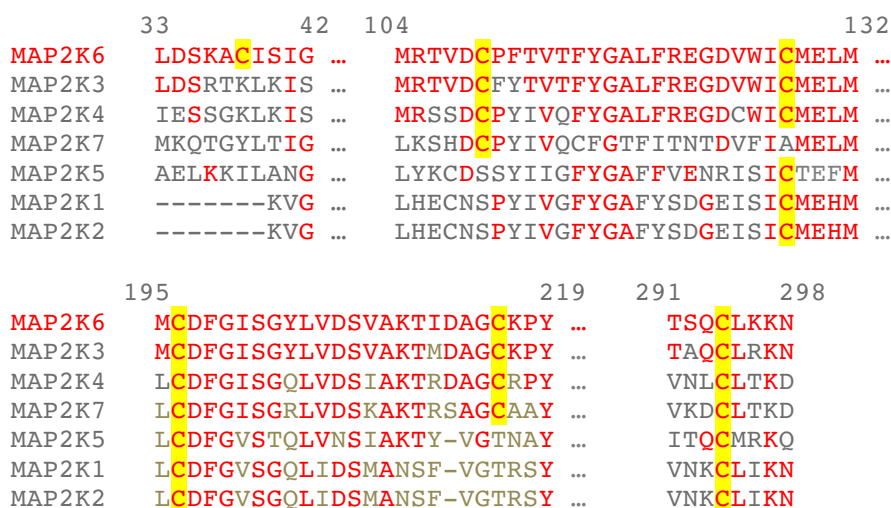


Figure 3.10. Multiple alignment of the cysteines in MAP2K6 with the other 6 members of the MAP2K family. Residues that align to MAP2K6 are in red, and cysteines are highlighted in yellow. Cysteine 38 is uniquely nonconserved among cysteines in these closely related kinases. Alignments were created using COBALT [24].

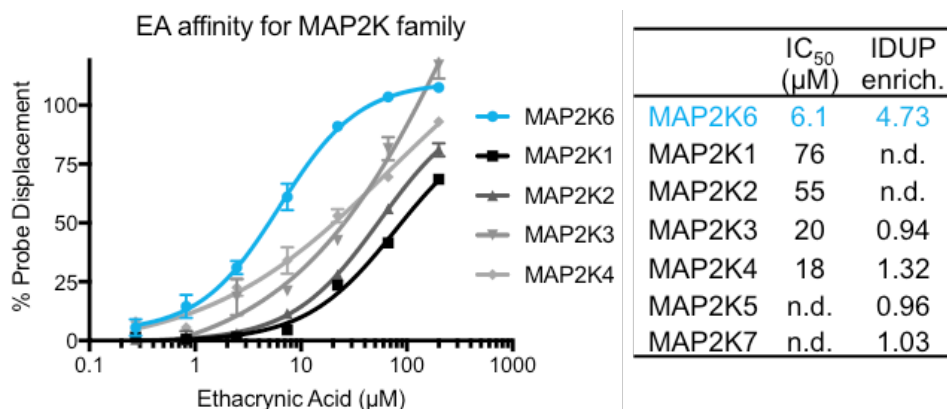


Figure 3.11. EA has higher affinity for MAP2K6 than the related MAP2K family members, as measured in both LanthaScreen Eu assays and our initial IDUP assay. n.d. = not determined.

3.7 Discussion

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 [25], development [26], and cancer [27,28]. EA's inhibition of MAP2K6, a cellular target unrelated to current uses of EA to treat edema or as a probe for GST function [18] or Wnt signaling [29], 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 \square 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 [30], 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 [31] 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 [32]. 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.

3.8 Experimental Methods

3.8.1 General methods

Unless otherwise noted, chemical reagents were purchased from Sigma-Aldrich. Purified water was obtained using a Milli-Q system. Standard DNA oligonucleotides were purchased from Integrated DNA Technologies. Modified oligonucleotides were synthesized using a PerSeptive Biosystems Expedite 8909 DNA synthesizer according to manufacturer's protocols. All reagents for DNA synthesis were purchased from Glen Research. Oligodeoxynucleotides were purified by reverse-phase high-pressure liquid chromatography (HPLC) on an Agilent 1200 system using a C18 stationary phase (Eclipse-XD-B C18, 5 μ M, 9.4 x 200 mm) and an acetonitrile/ 100 mM triethylammonium acetate gradient. Oligonucleotide concentrations were quantitated using a Nanodrop ND1000 spectrophotometer. Non-commercial oligonucleotides were characterized by LC/ESI-MS. Reverse-phase separation was performed on a Waters Acquity ultra-performance LC (UPLC) quadrupole TOF Premier instrument using a UPLC BEH C18 column (1.7 μ M, 2.1 x 50 mm) stationary phase and 6 mM aqueous triethylammonium bicarbonate/MeOH mobile phase.

3.8.2 Assembly of DNA-encoded ligand libraries

Design and validation of DNA sequences

Rather than use DNA sequences from the development of IDUP, we designed new sequences with the goal of avoiding contamination from previous experiments and of relocating the barcode region closer to the primer so that the constant region is long enough for purification by ethanol precipitation. We developed a new set of primer binding sequences for both the target-encoding strand and the small molecule-encoding strand. Keeping the complementary region constant, we first designed a set of potential primer binding sites for both the ligand and target strands. The Oligonucleotide Modeling Platform (OMP, DNASoftware) was used to improve the design of the primer binding sequences, decreasing hybridization between the two primer binding sequences. Finally, OMP was used to confirm that the most common species in a combined solution of the target and ligand strands was a heterodimer of the target and ligand strands via the complementary region. We next generated a text file containing all possible barcoded members of the ligand and target libraries and used UNAFold [33] to individually analyze their secondary structures and ability to hybridize to a single member of the complementary library. We compared these energies to those of the original library, and found that the majority of barcoded sequences exhibited low secondary structure and similar annealing energies, encouragingly suggesting that interactions outside of dimerization via the complementary region were not occurring.

We next performed a negative control library x library experiment to identify (and exclude) barcodes that were subject to overrepresentation or underrepresentation after processing by IDUP. In this experiment, primer extension, PCR and Illumina sequencing were performed on a mixture of all possible DNA barcodes for the small molecule and protein libraries. High throughput sequencing results were analyzed using a custom

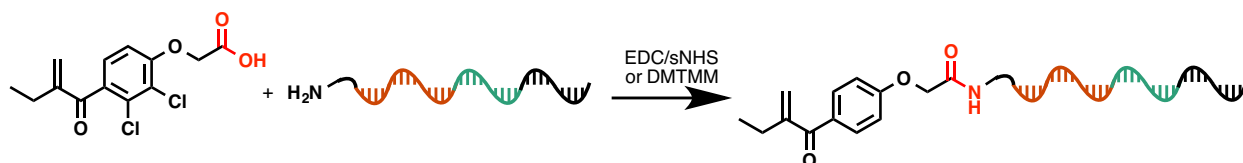
MATLAB script These data were further processed by a MATLAB script designed to remove all sequences that were observed more than 10 or fewer than 1 times. Finally, a MATLAB script was used to select from these sequences a set in which all barcodes are at least two mutations away from all other barcodes, giving a final list of barcodes which were used to purchase oligonucleotides for conjugation to proteins or small molecules.

Splint ligation of DNA strands

To prepare the full-length DNA sequences (5'-ACTATCGTGGCGACTCTAXXXXXXCCGATAGTATCTCACTCA-modifier-3', where XXXXXXX is the 7-nt barcode) for the library of DNA-encoded small molecules, we first synthesized the chemically modified constant sequence. SM1-dithiol or SM1-amine were synthesized using either 3'-Thiol-Modifier C3 S-S CPG or 3'-Amino-Modifier C6 CPG (Glen Research), respectively, with the sequence 5'-GTATCTCACTCA-modifier-3'. SM1-dithiol or SM1-amine (5 nmol in 2.5 μ L water) were phosphorylated by combining with 2 μ L 10X T4 DNA ligase buffer (NEB) and 0.75 μ L T4 PNK (10 U/ μ L) in a total volume of 20 μ L. This mixture was incubated at 37°C for 40 minutes before heat inactivation for 20 minutes at 65°C. To this heat inactivated mixture was added: 10 μ L 10X T4 DNA ligase buffer (NEB), Primer+code 31-mer (5 nmol in 50 μ L water, ACTATCGTGGCGACTCTAXXXXXXCCGATA), splint 12-mer (5 nmol in 5 μ L water, AGATACTATCGG), and 45.3 μ L water. The mixture was heated to 65°C for 3 minutes and cooled to 16°C using a -0.1°C/s ramp. After the mixture reached 16°C, T4 DNA ligase was added (1.25 μ L or 500 U) and the mixture was incubated at 16°C for 16 hours prior to heat inactivation at 65°C for 20 minutes. Amine-linked oligos were recovered by ethanol precipitation, dissolved in 50 μ L 400 mM TEA/HCl pH10 and used directly in reactions with activated carboxylates. Dithiol-linked oligos were deprotected by

addition of 12 μL DTT (1M in water). After 30 min incubation at room temperature, the thiol DNA was recovered by ethanol precipitation and was dissolved in 50 μL water and used in reactions with amino-small molecules.

Conjugation of carboxylate compounds to DNA (Scheme 3.1)



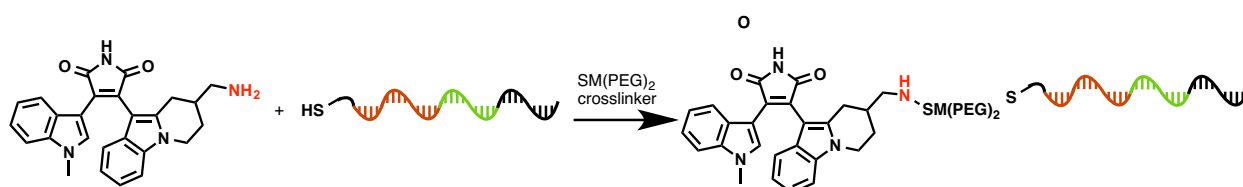
Scheme 3.1. General conditions for coupling carboxylic acid-containing small molecules to amine-functionalized DNA.

Small molecules were dissolved in DMSO to 100 mM. Aliphatic carboxylic acids (1.25 μmol) were activated using 3.3 μmol N-hydroxysulfosuccinimide (sNHS), (333 mM in 2:1 DMSO:water), 1.2 μmol 1-ethyl-3-[3-dimethylaminopropyl]carbodiimide (EDC) (100 mM in anhydrous DMSO) in 215 μL DMSO. This mixture was stirred at room temperature for 30 min before addition of 5-10 nmol 3'-amine-modified DNA, and 50 μL 500 mM TEA/HCl, pH 10. The resulting mixture was stirred 8-16 hours before the DNA was recovered by ethanol precipitation, purified by reverse phase HPLC and characterized by LC/MS.

We found that substituted benzoic acid derivatives, α,β -unsaturated carboxylic acids and β -lactams were not efficiently coupled using those conditions. Instead, the carboxylic acid containing small molecule (900 nmol in 9 μL DMSO) was activated in a mixture containing 162 μL DMSO, 2.5 μmol sNHS (7.5 μL in 2:1 DMSO:water), and DMTMM*Cl (4-(4,6-Dimethoxy-1,3,5,-triazin-2-yl)-4-methylmorpholinium chloride) (10 μmol in 20 μL 1:1 DMSO:500 mM MOPS buffer, pH 7.4). After 20 minutes, DNA (\sim 5 nmol in 50 μL 400 mM TEA/HCl) was added and incubated overnight. DNA was recovered by ethanol precipitation

and purified by reverse phase HPLC and characterized by LC/MS. Certain classes of carboxylic acids could not be efficiently linked to DNA using these protocols and were omitted without further optimization.

Conjugation of amino compounds to DNA (Scheme 3.2)



Scheme 3.2. General method for coupling amine-containing small molecules to thiol-functionalized DNA.

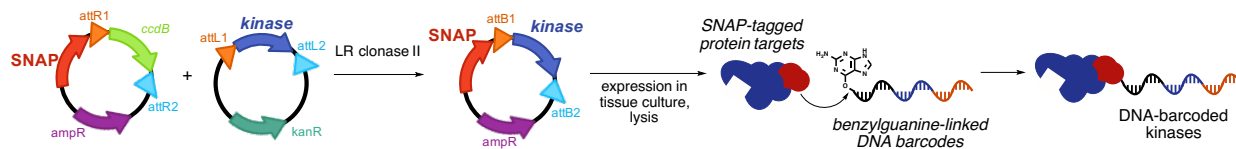
To deprotected thiol-DNA (50 μ L in water) was added 3.75 μ L primary amine-containing small molecule (100 mM in DMSO), 50 μ L DMSO, 14.2 μ L 10X PBS, 0.38 μ L 0.5M tris(carboxyethyl) phosphine (TCEP) (pH 7 in water), 7 μ L EDTA (50 mM in water), and 0.75 μ L SM(PEG)₂ heterobifunctional crosslinker (Pierce) (100 mM in anhydrous DMSO). The resulting mixture was vortexed and incubated overnight at room temperature. DNA was recovered by ethanol precipitation and DNA-small molecule conjugates were purified by reverse phase HPLC and characterized by LC/MS. Certain classes of amines could not be efficiently linked to DNA using these protocols and were omitted without further optimization.

Computational sorting of kinase library

Gene Ontology (GO) terms related to subcellular localization and a complete set of GO annotations for human genes were obtained from The Gene Ontology Consortium (<http://www.geneontology.org>) on February 26, 2014. A MATLAB script was used to generate a file that associates gene names referenced by the DFHCC ORF library with GO terms. The list of gene symbols and GO terms was then sorted into separate files based on

subcellular localization as reported by the corresponding GO term using a MATLAB script (available upon request).

Protein library preparation via Gateway cloning (Scheme 3.3)



Scheme 3.3. Overview of Gateway cloning method for creating DNA-barcoded, SNAP-tagged kinases.

Gateway cloning methods adapted from Yang *et al* [30]. A cassette containing Gateway R1 and R2 recombination sites flanking a chloramphenicol resistance marker and the *ccdB* toxin were inserted into the pSNAPf vector (New England Biolabs) to yield the pDEST-SNAP-ins vector using the following primers to amplify the Gateway cassette from a commercial Gateway destination vector (Invitrogen) and append restriction sites for cloning into the pSNAPf vector: CCTGCAGGACAAGTTTGTACAAAAAAGCTGAACG and CCTCGAGTTATCACCACTTTGTACAAGAAAGCTG. The pDEST-SNAP-ins vector was propagated in *ccdB* resistant cells (Invitrogen).

The Human Kinase Collection of DONR221 vectors was obtained as glycerol stocks from the PlasmID Repository at Harvard Medical School. The glycerol stocks were used to inoculate 1.2 mL cultures in a 96-well deep well plate. The resulting cultures were subjected to plasmid purification using an Epoch 96-well plate with a vacuum apparatus and Qiagen miniprep buffers. Each of the 289 Gateway pDONR221 plasmids corresponding to intracellular kinases were subcloned into the pDEST-SNAP-ins vector. A 5 μ L reaction containing 25 ng of each pDONR plasmid and the pDEST-SNAP-ins was incubated with 0.25

μ L LR clonase II (Invitrogen) overnight in 1X TE buffer. 10 μ l of ccdB-sensitive E. coli DH5 α competent cells (Zymo) were then added to the entire LR reactions for heat-shock transformation. The transformants were selected with 100 μ g/ml carbenicillin on multiwell, gridded, agar plates. We obtained colonies from 98.5% of LR reactions. Single colonies were cultured overnight in LB + carbenicillin in deep well 96-well plates and the plasmids isolated using a 96-well plasmid prep plate (Epoch) with a vacuum apparatus and Qiagen Miniprep kit buffers.

HEK-293T cells (ATCC) were maintained under standard conditions in DMEM supplemented with 5% FBS and Pen/Strep. HEK-293T cells were plated on gelatinized 12-well plates (200,000 cells/well). On the following day, the cells were transfected with Lipofectamine 2000 according to manufacturer's protocols. Cells were harvested 48 hours after transfection by dissociation. Growth medium was removed and the cells were incubated for 5-10 minutes at 37°C in 500 μ L of a sterile-filtered dissociation buffer containing 15 mM sodium citrate and 135 mM KCl. The cells were transferred to a microcentrifuge tube, pelleted at 400g for 4 min, and the supernatant was aspirated. Cells were resuspended in 75 μ L of lysis buffer (10 mM Tris, pH 7.4, 137 mM NaCl, 2 mM EDTA, 1 mM DTT, 2 μ M leupeptin, 1mM PMSF) prior to mechanical lysis using QIASHredder spin columns (Qiagen). The resulting lysate was flash frozen and stored at -80°C prior to use in the library x library experiment. Successful transfection and protein expression were verified by Western blot using the anti-SNAP-tag antibody (New England Biolabs). (Figure 3.2)

Table 3.3. Compounds included in DNA-encoded small molecule library. Supplier name abbreviations: SA = Sigma Aldrich, P = Prestwick, C= Cayman Chemical, SC = Santa Cruz Biotechnology, AK Scientific, I = obtained from departmental inventory, commercial source not specified. Amine and carboxylate derivatives of JQ1 were generously provided from the laboratory of James Bradner (B).

Source	Barcode #	Barcode sequence	Structure	CAS #	Common Name	MW	Expected oligo + compound mass (m/z = -6)	Observed mass (m/z = -6)
P	SM-1	GGGTCCT		51-24-1	3,3',5-triiodothyroacetic acid	621.9	2919.4	2919.4
P	SM-2	TTCTCCT		89796-99-6	aceclofenac	354.2	2847.9	2848.1
SA	SM-3	CCATCCT		58-85-5	biotin	244.3	2824.7	2825
SC	SM-4	GTTGCCT		58-54-8	ethacrynic acid	303.1	2850.7	2851.1
AK	SM-6	CGCGCCT		134523-03-8	atorvastatin	558.6	2895.8	2895.1
AK	SM-8	TGTCCCT		41859-67-0	bezafibrate	361.8	2854.4	2854.7
AK	SM-9	CTGCCCT		28395-03-1	bumetanide	364.4	2851.9	2852.4
AK	SM-10	GCCCCCT		62571-86-2	captopril	217.3	2862.9	2788.4
AK	SM-11	ACTACCT		53716-49-7	carprofen	273.7	2835.4	2835
AK	SM-12	CTTTACT		78439-06-2	ceftazidime pentahydrate	546.6	2891.2	2891.7
AK	SM-14	GACTACT		982-57-0	chloramphenicol succinate	422.2	2873.1	2877.8

Table 3.3 (continued)

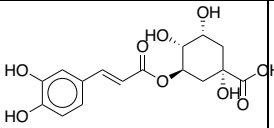
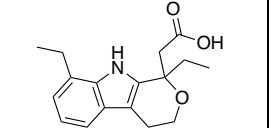
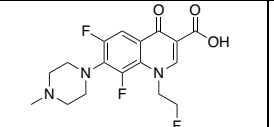
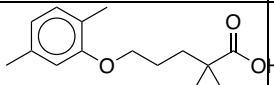
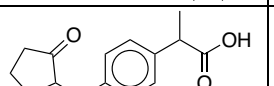
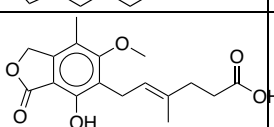
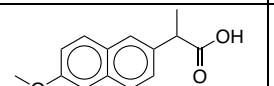
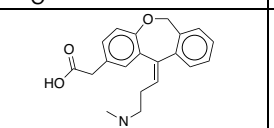
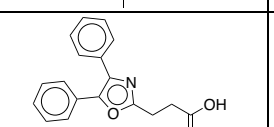
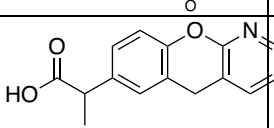
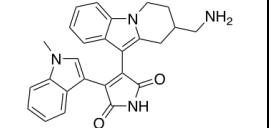
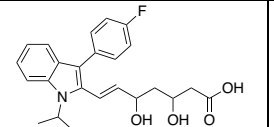
Source	Barcode #	Barcode sequence	Structure	CAS #	Common Name	MW	Expected oligo + compound mass (m/z = -6)	Observed mass (m/z = -6)
AK	SM-15	TGATACT		327-97-9	chlorogenic acid	354.3	2862.5	2863
AK	SM-17	CAGGACT		41340-25-4	etodolac	287.4	2853.1	2851.2
AK	SM-18	GCAGACT		79660-72-3	floxacin	369.3	2867.5	2805.6
AK	SM-20	TATAACT		25812-30-0	gemfibrozil	250.3	2838.5	2792.5
AK	SM-22	CCCAACT		80382-23-6	loxoprofen	245.3	2826.7	2827.1
AK	SM-23	TTTTTAT		24280-93-1	mycophenolic acid	320.3	2851.9	2852.3
AK	SM-24	CGGTAT		26159-34-2	naproxen	229.2	2840.7	2850.9
AK	SM-25	GCCTTAT		14046-2-76-6	olopatadine	337.4	2854.3	2854.7
AK	SM-26	AAATTAT		21256-18-8	oxaprozin	293.3	2851.9	2852.5
AK	SM-28	ATGGTAT		52549-17-4	pranoprofen	255.2	2850.7	2850.9
AK	SM-35	CATATAT		14531-7-11-9	bisindolylmaleimide X	426.0	2873.6	2874.8
AK	SM-36	TCGATAT		93957-55-2	fluvastatin	410.4	2873.7	2874.2

Table 3.3 (continued)

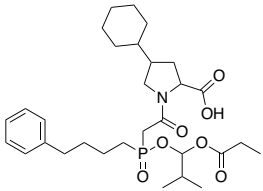
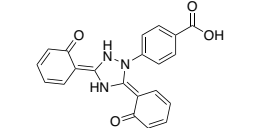
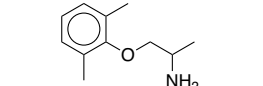
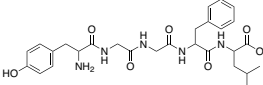
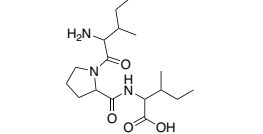
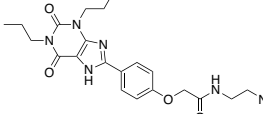
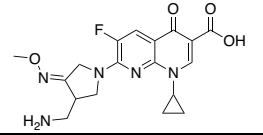
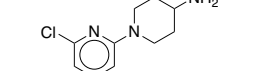
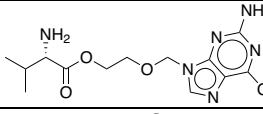
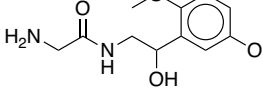
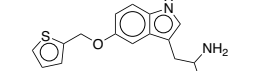
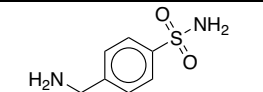
Source	Barcode #	Barcode sequence	Structure	CAS #	Common Name	MW	Expected oligo + compound mass (m/z = -6)	Observed mass (m/z = -6)
AK	SM-40	CCTACTA		88889-14-9	fosinopril	562.6	2915.4	2912.6
AK	SM-46	GGCCGAT		201530-41-8	deferasirox	373.4	2881.1	2909.9
AnaSpec	SM-49	TTCAGAT	H ₂ N-GQVGRQLAIIGDDINR-OH		Bak peptide	1725.0	3142.3	3143.7
AK	SM-53	CCTTGAA		5370-01-4	mexiletene	179.3	2827.9	2828.2
SA	SM-54	CGAAGAT		58822-25-6	leucine enkephalin	555.6	2909.988	2911.1
SA	SM-55	GATTCAT		90614-48-5	diprotin A	341.45	2860.354	2861.5
SA	SM-59	TCTGAAT		96865-92-8	xanthine amine congener	428.5	2882.6	2882.9
AK	SM-60	GCCTGTA		210353-53-0	gemifloxacin mesylate	389.4	2869.9	2870.4
SC	SM-65	CCCTTTG		77145-61-0	SR 57227A	212.8	2844.0	2844.5
SA	SM-66	CGTGTTG		124832-26-4	valacyclovir	307.4	2843.9	2847.7
SA	SM-67	GCGGTTG		133163-28-7	midodrine	254.3	2849.3	2849.8
SA	SM-68	AAAGTTG		160521-72-2	BW 723C86	286.4	2860.7	2861.3
SA	SM-70	TCTCTTG		138-37-4	4-aminomethylbenzene-sulfonamide	187.3	2839.3	2839.6

Table 3.3 (continued)

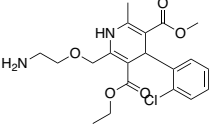
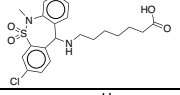
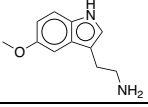
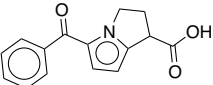
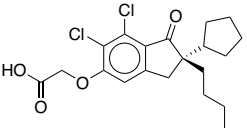
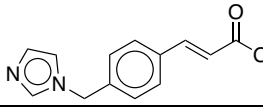
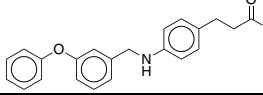
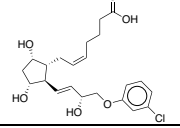
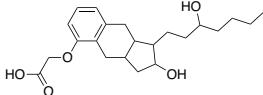
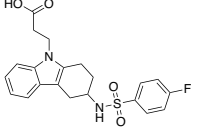
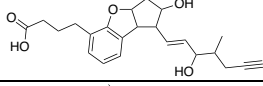
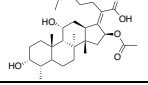
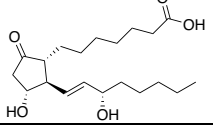
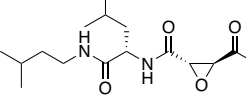
Source	Barcode #	Barcode sequence	Structure	CAS #	Common Name	MW	Expected oligo + compound mass (m/z = -6)	Observed mass (m/z = -6)
P	SM-71	TGCTCAA		88150-42-9	amlodipine	408.9	2867.2	2867.8
AK	SM-74	GAGTGTG		66981-73-5	tianeptine	436.9	2882.2	2807
VWR	SM-75	GAGTGTG		66-83-1	O-methylserotonin hydrochloride	190.24	2845.9	2821.1
AK	SM-76	TTATGTG		74103-07-4	ketorolac	255.3	2848.9	2850.1
SC	SM-77	TATGGTG		81166-47-4	R(+)-DIOA	399.3	2882.7	2882.6
C	SM-81	CTTAGTG		82571-53-7	ozagrel	228.2	2840.5	2816.1
C	SM-82	TCGAGTG		88510-1-89-3	GW9508	347.4	2869.3	2869.8
C	SM-87	GACGCTG		40665-92-7	cloprostenol	424.9	2881.8	2882.1
C	SM-90	TAGCCTG		81846-19-7	treprostinil	390.5	2869.9	2870.4
C	SM-91	GTGACTG		11664-9-85-5	ramatroban	416.5	2883.1	2883.2
C	SM-92	TGAACTG		88430-50-6	beraprost	398.5	2876.3	2876.6
C	SM-93	TACTATG		751-94-0	fusidic acid	516.7	2895.0	2838.1
C	SM-95	ACCGATG		745-65-3	prostaglandin E1	354.5	2864.5	2864.8
C	SM-96	GATCATG		88321-09-9	E-64d	314.4	2868.1	2828.1

Table 3.3 (continued)

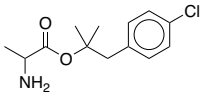
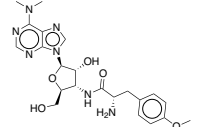
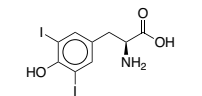
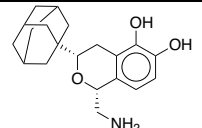
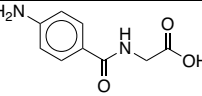
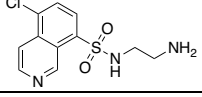
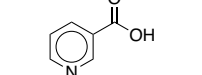
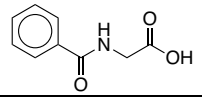
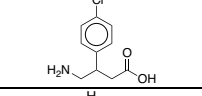
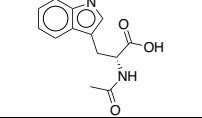
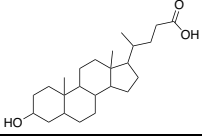
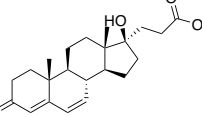
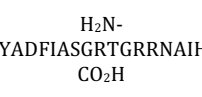
Source	Barcode #	Barcode sequence	Structure	CAS #	Common Name	MW	Expected oligo + compound mass (m/z = -6)	Observed mass (m/z = -6)
P	SM-108	GTCTGGG		60719-83-7	alaproclate	255.7	2854.6	2855.4
SA	SM-111	TGCGGGG		58-58-2	puromycin	471.5	2902.8	2903.7
I	SM-113	CGGCGGG		18835-59-1	3,5-Diiodo-L-tyrosine	433.0	2892.1	2892.8
SA	SM-119	GCTCCGG		14530-7-34-2	3-(1-adamantyl)-1-(aminomethyl)-3,4-dihydro-1H-isochromene-5,6-diol "A-77636"	330.9	2858.6	2959.1
SC	SM-126	ATCAAGG		61-78-9	p-aminohippuric acid	194.2	2837.3	2837.7
SA	SM-127	TATTTCCG		11771-41-67-1	CKI-7	285.7	2846.8	2802
I	SM-128	TATTTCCG		54-86-4	isonicotinic acid	123.1	2819.5	2820
I	SM-130	ATGTTCG		495-69-2	hippuric acid	179.2	2827.7	2828
SC	SM-132	CCACTCG		63701-55-3	arbaclofen	215.12	2824.274	2824.8
I	SM-133	GAATGCA		1218-34-4	N-acetyl-L-tryptophan	246.3	2841.1	2841.7
I	SM-135	TCCTGCG		434-13-9	lithocholic acid	376.6	2862.3	2862.8
I	SM-136	GGCACTA		2181-04-6	canrenoic acid	358.5	2866.7	2866.9
SA	SM-137	AGACGCG	H ₂ N-TTYADFIASGRTGRRNAIHD-CO ₂ H	99534-03-9	PKI [cAMP-dependent protein kinase inhibitor (5-24)]	2222.4	3243.12	2811.5
SA	SM-138	TTGCGCG		138-41-0	carboxybenzene sulfonamide	201.2	2841.9	2842.8

Table 3.3 (continued)

Source	Barcode #	Barcode sequence	Structure	CAS #	Common Name	MW	Expected oligo + compound mass (m/z = -6)	Observed mass (m/z = -6)
I	SM-140	GATAGCG		102-32-9	3,4-dihydroxyphenylacetic acid	168.1	2824.1	2824.4
C	SM-141	AGCTACG		51-48-9	L-thyroxine	776.8	2949.422	2727
I	SM-143	ATAACCG		38194-50-2	sulindac	356.4	2881.1	2881.7
	SM-144	CTCGATA		6385-02-0	meclofenamic acid	296.1	2849.7	2798.2
I	SM-145	GGGGACG		57-66-9	probenecid	285.4	2848.7	2849.3
I	SM-146	TAACACG		53188-07-1	trolox	250.3	2845.3	2845.7
I	SM-147	CTATTAG		13395-35-2	2-iminobiotin	243.3	2855.3	2855.9
SA	SM-148	ACAATAG		533-48-2	desthiobiotin	214.3	2847.5	2847.8
I	SM-149	CAGGGAG		59-05-2	methotrexate	454.4	2881.5	2876.1
C	SM-150	TCTTCAG		30827-99-7	pefabloc	203.2	2828.198	2746.3
I	SM-151	CCCAGAG		1634-82-8	2-(4-hydroxyphenylazo) benzoic acid	242.0	2846.8	2847.5
I	SM-153	AAGTCAG		118-41-2	3,4,5-trimethoxy benzoic acid	212.2	2826.3	2826.6
I	SM-154	CAAACGA		399-76-8	5-fluoroindole-2-carboxylic acid	179.1	2834.5	2834.9
I	SM-158	TTCCAG		306-08-1	homovanillic acid	182.2	2844.7	2845
I	SM-161	GGACCAG		81-23-2	dehydrocholic acid	402.5	2867.1	2867.8

Table 3.3 (continued)

Source	Barcode #	Barcode sequence	Structure	CAS #	Common Name	MW	Expected oligo + compound mass (m/z = -6)	Observed mass (m/z = -6)
I	SM-165	AATTTTC		53-86-1	indomethacin	357.8	2855.4	2856
I	SM-166	ATTTAGA		5728-52-9	felbinac	212.2	2832.3	2833
I	SM-169	TGTATTC		59703-84-3	piperacillin	536.0	2905.2	2905.9
I	SM-171	GAGATCA		12418-2-57-6	CGS-21680	500.9	2892.2	2892.8
SA	SM-174	GGTTGTC		66309-69-1	cefotiam	525.6	2883.2	2892.6
SA	SM-175	CGCGGTC		479-20-9	atranorin	374.3	2865.3	2822.7
SA	SM-176	GCGCGTC		88909-96-0	virstatin	283.3	2838.7	2839.7
SA	SM-177	AAACGTC		66-79-5	oxacillin	401.4	2868.9	2873.3
SA	SM-180	TCGTCTC		79558-09-1	L-165,041	402.4	2859.5	2860.3
SA	SM-182	AGATCTC		32852-81-6	3-phenoxyphenylacetic acid	228.2	2824.5	2825.2
SA	SM-183	TGCCCTC		14277-97-5	domoic acid	311.3	2849.1	2877.9
SA	SM-184	CCTTATC		15087-06-6	estriol 3-(beta-D-glucuronide)	464.5	2887.7	2888.7
SA	SM-186	GTCTATC		33369-31-2	zomepirac	291.7	2839.0	2839.6

Table 3.3 (continued)

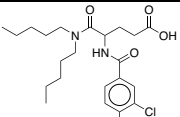
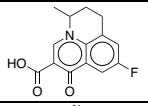
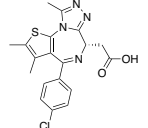
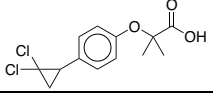
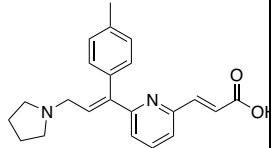
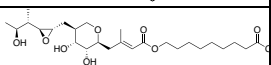
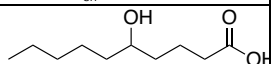
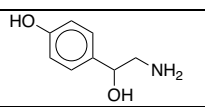
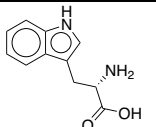
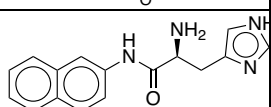
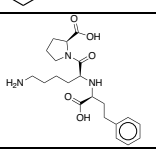
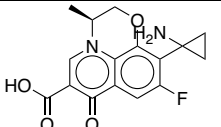
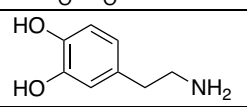
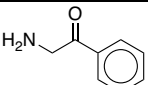
Source	Barcode #	Barcode sequence	Structure	CAS #	Common Name	MW	Expected oligo + compound mass (m/z = -6)	Observed mass (m/z = -6)
SA	SM-187	TTGGATC		97964-56-2	lorglumide	459.4	2877.3	2878.2
SA	SM-189	CTAAATC		42835-25-6	flumequin	261.2	2842.3	2888.9
B	SM-194	TTTCTCA		20259-2-23-2	JQ1-CO ₂ H	399.0	2869.8	2871
SA	SM-196	CGCCTCA		52214-84-3	ciprofibrate	289.2	2852.7	2853.6
SA	SM-197	TTGCTGC		87848-99-5	acrivastine	348.4	2881.7	2887.1
SA	SM-202	TATTGGC		12650-69-0	mupirocin	500.6	2897.0	2891.9
SA	SM-212	AGGTGGC		71186-53-3	5-hydroxydecanoic acid	187.3	2828.1	2829.1
SA	SM-220	AGCCAGC		770-05-8	(±)-octopamine	153.2	2802.1	2803.1
SA	SM-222	TGGCAAC		73-22-3	L-tryptophan	204.2	2836.3	2837.5
SA	SM-225	TCCCTCC		7424-15-9	L-histidine beta-naphthylamide	280.3	2845.3	2846.4
AK	SM-226	TGTGGCC		76547-98-3	lisinopril	405.5	2872.2	2873.1
AK	SM-227	GTACGCC		12704-5-41-4	pazufloxacin	318.3	2869.6	2873.6
I	SM-228	ACGAGCC		645-31-8	3-hydroxy tyramine	153.2	2803.9	2809.8
I	SM-234	GCGTCAA		5468-37-1	2-aminoacetophenone	136.6	2816.4	2817

Table 3.3 (continued)


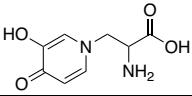
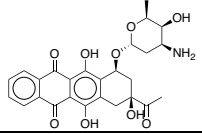
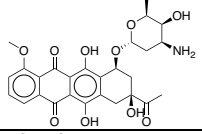
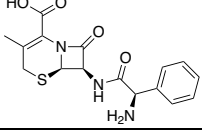
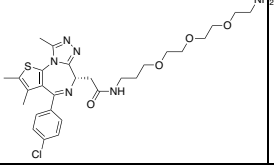
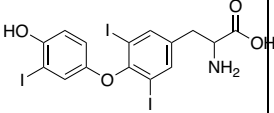

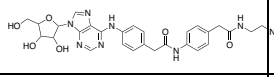
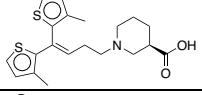
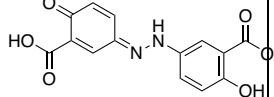
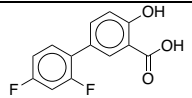

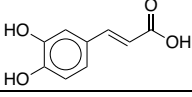
Source	Barcode #	Barcode sequence	Structure	CAS #	Common Name	MW	Expected oligo + compound mass (m/z = -6)	Observed mass (m/z = -6)
I	SM-236	TAAAGAC		61-54-1	tryptamine	160.22	2827.712	2739
I	SM-237	TAATTTA		500-44-7	L-mimosine	198.2	2823.9	2815.6
SA	SM-238	GTATCAC		25316-40-9	daunorubicin	497.5	2894.6	2375
SA	SM-239	CCCCCAC		23214-92-8	doxorubicin	527.5	2879.2	2880.1
I	SM-240	ACGTTAC		15686-71-2	cephalexin monohydrate	347.4	2855.5	2856.8
B	SM-243	TGACGGC			JQ1-NH ₂	589.2	2914.7	2915.6
I	SM-245	TAATTTA		6893-02-3	rathyronine	651.0	2922.1	2820.6
I	SM-249	CGAAGAT		153-98-0	serotonin	176.2	2830.6	2831.5
I	SM-250	CCACTCG		9676-069-9	adenosine amine congener	594.6	2911.2	2908.7
AK	SM-259	ATGCCTA		1458-21-59-6	tigabaine	375.6	2863.8	
AK	SM-260	CCTACTA		6054-98-4	olsalazine sodium	302.2	2841.1	2897.8
I	SM-261	GGCACTA		2249-442-4	diflunisal	250.2	2843.7	2831.4
I	SM-263	TGTGATA		303-38-8	2,3-dihydroxybenzoic acid	154.1	2830.5	2845.7
I	SM-264	TGTGATA		331-39-5	caffeic acid	180.2	2824.7	2881.4

Table 3.3 (continued)

Source	Barcode #	Barcode sequence	Structure	CAS #	Common Name	MW	Expected oligo + compound mass (m/z = -6)	Observed mass (m/z = -6)
AK	SM-265	GAAGATA		9310 6-60- 6	enrofloxacin	359.4	2875.1	2729.3
C	SM-283	CTTGCGA		5026 4-69- 2	lonidamine	321.2	2856.1	2857.3
I	SM-286	GTCCCGA		4800 -94- 6	carbenicillin	378.4	2864.5	2856.4
I	SM-287	CAAACGA		8241 9-36- 1	ofloxacin	361.4	2864.5	2865.3
I	SM-288	ATTTAGA		90- 50- 6	3,4,5- trimethoxycinnami c acid	238.2	2844.1	2845
SA	SM-293	TTTCTCA		3184 2-01- 0	indoprofen	281.3	2838.1	2431.2
SA	SM-294	CGCCTCA		2747 0-51- 5	suxibuzone	438.5	2868.5	2884.7
I	SM-295	GAGATCA		3288 7-01- 7	mecillinam	325.4	2863.5	2867.9
I	SM-296	GAATGCA		599- 79- 1	sulfasalazine	398.4	2878.1	2934.6
AK	SM-301	TAGTCCA		8379 9-24- 0	fexofenadine	501.7	2896.3	2076
SA	SM-302	TCTTACA		5316 4-05- 9	acetaminophen	415.8	2866.8	2462.6

Table 3.3 (continued)

Source	Barcode #	Barcode sequence	Structure	CAS #	Common Name	MW	Expected oligo + compound mass (m/z = -6)	Observed mass (m/z = -6)
AK	SM-305	CCTTGAA		63-45-6	primaquine bisphosphate	265.5	2842.2	2841.9
I	SM-308	GCGTCAA		1132-68-9	p-fluoro-L-phenylalanine	183.2	2830.7	2831.6
AK	SM-309	TGCTCAA		9266-5-29-7	cefprozil	389.4	2866.5	2868.1
C	SM-310	TATGCAA		6171-4-27-0	N-(6-aminohexyl)-5-chloro-1-naphthalenesulfonamide	342.3	2862.3	2863.1
I	SM-316	GCAAAAA		60-19-5	tyramine	137.2	2824.9	2826

3.8.3 Library x library IDUP selection to detect protein-ligand pairs

4.5 μ L of a mixture containing NEBuffer 2 (3.2 μ L) (NEB) and dNTPs (1.3 μ L of a 1 mM stock, NEB) were dispensed into wells of 2.5 96-well PCR plates. 4 μ L of each DNA-BG (Table 3.1) (1 μ M) were added. 17.5 μ L of thawed lysate with SNAP-target library members (prepared as described above) were added. The entire reaction was incubated at 37°C for 30 minutes. After labeling, 2 μ L SNAP-Cell Block (100 μ M in 10% DMSO) (NEB) was added and the reaction incubated for 15 minutes at 37°C. For a negative control sample containing lysates, all lysates were pooled prior to addition of SNAP-Cell Block for 15 min at 37°C followed by addition of pooled PS2-BG sequences (20 min at 37°C). 4 μ L of all labeled lysates were then pooled. 14 μ L of this pooled material was combined with 2 μ L of the pooled small molecule library (1 μ M). This mixture was incubated at 37°C for 15 minutes to allow protein-small molecule binding to occur. Primer extension was performed by adding 4 μ L of a master mixture containing 0.17 μ L T4 DNA Polymerase (3U/ μ L), 0.04 μ L BSA (10

mg/ml), 0.4 µL 10x NEB Buffer 2, and 3.39 µL water (per reaction). Primer extension was performed for 15 minutes at 37°C followed by heat inactivation at 75°C for 20 minutes.

3.8.4 Preparation of sequences for high throughput sequencing

These primer extension products were then amplified by PCR and prepared for high throughput sequencing as previously described [10]. Briefly, adapters compatible with Illumina paired end sequencing were installed in two sequential PCR steps. An analytical qPCR was performed in a 25 µL reaction volume with a final concentration of 1x Q5 buffer, 200 µM each dNTPs, 0.5 µM each primer, 1.25 µL 10X SYBR Green I (Invitrogen), 0.25 U Q5 Hot Start DNA polymerase (NEB), and 1 µL of the IDUP primer extension product (Primers: 5'- TGGAGTTCAGACGTGTGCTCTTCCGATCTACTATCGTGGCGACTCT -3' and 5'- ACACTCTTCCCTACACGACGCTCTTCCGATCTNNNNACCTGTGAGAGCTAGTCT-3'). PCR Conditions: 30 sec at 95°C, followed by 40 cycles of [10 sec at 95°C, 10 sec at 65°C, 20 sec at 72°C]. The samples were prepared in 50 µL PCR reactions, stopping at the C_T value of each sample. Primers were removed using a PCR Cleanup Kit (Qiagen). The resulting samples were diluted 1:100 and 1 µL was used as a template for the 2nd qPCR and PCR. For the fourteen total samples (seven replicates and seven negative controls) we used the following primers in the second PCR combinatorially in order to demultiplex each sample after pooling for sequencing. (Table 3.4, Table 3.5)

Table 3.4 Forward primers used in Illumina sequencing.

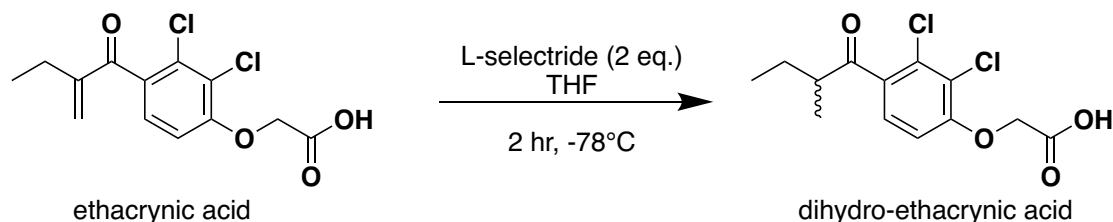
Illadapterfwd-2A	AATGATACGGCGACCACCGAGATCTACAC TTACTCG ACTCTTTCCTA CACGAC
Illadapterfwd-2B	AATGATACGGCGACCACCGAGATCTACACT CCGGAGA AACTCTTTCCTA CACGAC
Illadapterfwd-2C	AATGATACGGCGACCACCGAGATCTACAC CGCTCATT AACTCTTTCCTA CACGAC
Illadapterfwd-2D	AATGATACGGCGACCACCGAGATCTACAC GAGATTCC AACTCTTTCCTA CACGAC

Table 3.5 Reverse primers used in Illumina Sequencing.

PE_REV-18B	CAAGCAGAAGACGGCATAACGAGATGCGTACGTGTGACTGGAGTTCAGACG TGTGCT
PE_REV-23B	CAAGCAGAAGACGGCATAACGAGATCCACTCATGTGACTGGAGTTCAGACGT GTGCT
PE_REV-25B	CAAGCAGAAGACGGCATAACGAGATAGGAATAAGTGACTGGAGTTCAGACG TGTGCT
PE_REV-27B	CAAGCAGAAGACGGCATAACGAGATATCAGTATGTGACTGGAGTTCAGACG TGTGCT

The products of the 2nd PCR were purified by PAGE (5% Criterion TBE (Bio-Rad), 200V, 40 min, stained with SYBR Gold (Invitrogen)). DNA was eluted from excised bands by incubating with 150 μ L 10 mM Tris, pH 7.5 overnight in a 37°C shaker. Eluted DNA was purified with a PCR Cleanup Kit (Qiagen), and quantified using the Quant-iT picoGreen kit (Invitrogen). The pooled samples were further quantified by qPCR using a Library Quantification Kit (Kapa Bioscience). The samples were sequenced on an Illumina MiSeq using a 50-cycle MiSeq Reagent kit v2 (Illumina). Approximately 900,000 reads were obtained per experimental replicate from sequencing. The MiSeq data was processed using two in-house MATLAB scripts (see Lynn McGregor's thesis).

3.8.5 Synthesis of dihydro-ethacrynic acid (Scheme 3.4)



Scheme 3.4. Reduction of ethacrynic acid to dihydro-ethacrynic acid.

The synthetic route was adapted from Nibbs *et al* [34]. Ethacrynic acid (100 mg, Abcam) was added to a flame-dried 4-dram vial equipped with magnetic stirring bar. The vial was purged with N₂ and the solid dissolved in anhydrous THF (1.5 mL). The reaction

was cooled to -78°C and 0.66 mL of a 1 M solution of L-selectride (2 eq) was slowly added via syringe. The reaction mixture was stirred for 2 hours at -78 °C, then quenched slowly with MeOH, followed by 1.0M aqueous HCl. The mixture was extracted three times with ethyl acetate and washed once with brine. The combined organic layers were dried with MgSO₄, filtered, and concentrated by rotary evaporation. The crude material was purified by flash chromatography (SiO₂, 50% EtOAc/1% acetic acid/hexanes) as a white solid (56 mg, 56% yield). ESI-MS calculated for [M-H]⁻: 303.0196. Found: 303.0194. ¹H NMR (CDCl₃, 500 MHz): δ 7.26 (d, 1 H, *J* = 8.5 Hz), 6.79 (d, 1 H, *J* = 8.5 Hz), 4.79 (s, 2 H), 3.17 (m, 1 H), 1.78 (m, 1 H), 1.15 (d, 3H, *J* = 6 Hz), 0.92 (d, 3H, *J* = 6.5 Hz).

3.8.6 Construction of the N-His₆ MAP2K6 plasmid

MAP2K6 was cloned from the MAP2K6 pDONR221 into the pTrcHisA vector (Thermo Scientific Fisher) via USER (uracil-specific excision reaction) cloning. PCR was performed separated on the two vectors using deoxyuracil-containing primers in 200 μ L reactions with 40 μ L VeraSeq Buffer 2, 4 μ L 10 mM dNTPs, 0.5 μ L reverse and forward primers (100 μ M stock), 2 μ L VeraSeq polymerase (Enzymatics) and 5 ng template DNA. The MAP2K6 insert was amplified using the following primers: 5'-ATCATCATCATCATCATATGTCTCAGUCGAAAGGCAAGAA-3' and 5'-AGCCATACCCTAGTCTCCAAGAAUCAGTTTTACAAAAGA-3'. The pTrcHisA plasmid was amplified using the primers 5'-ACTGAGACATATGATGATGATGATGAUGAGAACCCC-3' and 5'-ATTCTTGGAGACTAGGGTATGGCUAGCATGACTGGTG-3'. PCR products were confirmed by agarose gel electrophoresis, purified using the MinElute PCR Purification Kit (Qiagen), and quantified by UV absorbance (NanoDrop). 0.2 pmol of each PCR product were combined in a 10 μ L reaction with 1 μ L NEB Buffer 4, 0.75 μ L DpnI, and 0.75 μ L USER mix

(NEB). The reaction was incubated at 37°C for 45 minutes, then 80°C for 3 minutes, followed by cooling to 30 °C at a rate of 0.2 °C/min. 1 µL of this mixture was directly transformed into NEB Turbo cells (NEB) according to manufacturer's standard protocols and plated onto LB agar with carbenicillin (100 µg/mL). Single colonies were cultured overnight in LB + carbenicillin and the pTrcHisA-MAP2K6 plasmid isolated using the QIAprep Miniprep Kit (Qiagen). The sequence of the final vector was confirmed via Sanger sequencing.

3.8.7 Recombinant MAP2K6 protein expression and purification

The pTrcHisA-MAP2K6 vector was transformed into Rosetta 2 DE3 cells (Novagen) according to manufacturer's protocols. Single colonies were grown to OD 0.6-0.7 in 1 L LB with carbenicillin (100 µg/mL) and chloramphenicol (35 µg/mL) followed by induction with 0.5 mM IPTG for 16 hours at 20 °C. Cells were pelleted by centrifugation (10 min, 10,000 *g* at 4°C) and resuspended in TBS (50 mM Tris-HCl pH 7.5 with 300 mM NaCl, 1 mM dithiothreitol [DTT]) with 1% Triton X-100, 1 mM phenylmethylsulfonyl fluoride (PMSF), and 10 mM imidazole. Cells were lysed by probe sonication for 2 x 3 min at 4 °C and pelleted by centrifugation (10 min, 10,000 *g* at 4°C). The supernatant was incubated with 1 mL HisPur Ni-NTA resin (Thermo Scientific) for 1 hr at 4 °C. The resin was subsequently washed twice with 10 mL of the lysis buffer, then twice with 1 mL of TBS + 50 mM imidazole and twice with 1 mL of TBS + 250 mM imidazole. Protein was eluted with two 1 mL washes of TBS + 500 mM imidazole. N-His₆-MAP2K6 was buffer exchanged using Slide-A-Lyzer Dialysis Cassettes (Thermo Fisher) into storage buffer (50mM Tris pH 7.5, 150mM NaCl, 0.5mM EDTA, 0.05% Triton X-100, 2mM DTT, 20% glycerol), diluted to 1 mg/mL,

flash frozen, and stored at -80 °C. Protein yield was typically between 5-10 mg/L and >90% based on gel electrophoresis analysis (Coomassie stained).

3.8.8 Site-directed mutagenesis of MAP2K6

Site-directed mutagenesis of MAP2K6 was performed using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs) according to manufacturer's standard protocols, using the pTrcHisA-MAP2K6 vector as template and the primers described in Table 5.6.

Plasmids were expressed and purified as described above (Section 3.8.7).

Table 5.6 Primers used for site-directed mutagenesis of MAP2K6.

Primer	Sequence
MAP2K6_T211E_fw	5'-GTTGCTAAAGAAATTGATGCAGGTTG-3'
MAP2K6_S207E_re	5'-CTCGTCCACCAAGTAGCCACT-3'
MAP2K6_C38A_fw	5'-TTAGACTCCAAGGCTGCCATTTCTATTGAAA-3'
MAP2K6_C38A_re	5'-ATCTCGAGGTGGTGTGGAAC-3'

3.8.9 Intact protein LC-MS

N-His₆-MAP2K6 (20 μM) was combined with ethacrynic acid (40 μM) or DMSO in a 500 μL reaction in Kinase buffer A (50 mM HEPES pH 7.5, 10 mM MgCl₂, 1 mM EGTA, and 0.01% Brij-35) (Invitrogen) for 1 hr at room temperature. The samples were analyzed on an Agilent 6220 ESI-TOF mass spectrometer equipped with an Agilent 1260 HPLC. The separation and desalting was performed on an Agilent PLRP-S Column (1000A, 4.6 x 50 mm, 5 μm). Mobile the phase A was 0.1% formic acid in water and mobile phase B was acetonitrile with 0.1% formic acid. A constant flow rate of 0.250 ml /min was used. Ten microliters of the protein solution was injected and washed on the column for the first 3 minutes at 5%B, diverting non-retained materials to waste. The protein was then eluted using a linear gradient from 5%B to 100%B over 7 minutes. The mobile phase composition was maintained at 100%B for 5 minutes and then returned to 5%B over 1 minute. The

column was then re-equilibrated at 5%B for the next 4 minutes. Data was analyzed using Agilent MassHunter Qualitative Analysis software (B.06.00, Build 6.0.633.0 with Bioconfirm). The charge state distribution for the protein produced by electrospray ionization was deconvoluted to neutral charge state using Bioconfirm's implementation of MaxEnt algorithm, giving a measurement of average molecular weight.

3.8.10 Identification of ethacrynic acid-modified peptide by MALDI-TOF

N-His₆-MAP2K6 (20 μM) was combined with ethacrynic acid (40 μM) or DMSO in a 500 μL reaction in Kinase buffer A (Invitrogen) for 1 hr at room temperature. Trypsin digest was carried out following manufacturer's protocols. In short, the protein was buffer exchanged into 50 mM Tris-HCl pH 8 buffer with 8M urea and 5 mM DTT to remove unreacted ethacrynic acid and denatured at 37°C for 1 hr. Iodoacetamide was then added to a final concentration of 15 mM and the reaction incubated for an additional 30 min at rt in the dark. The reaction was buffer exchanged into 50 mM Tris pH 8 and 6.3 μL of 25 μg/mL Trypsin Gold (Promega) was added, for a final ratio of 30:1 MAP2K6:trypsin by weight. Digestion was allowed to proceed for 1 hr at rt. Samples were desalted using C18 ZipTips (Millipore) according to manufacturer's protocols and eluted using 70% acetonitrile/water saturated with sinapic acid as MALDI carrier onto a stainless steel MALDI plate. Data were collected on a ultrafleXtreme MALDI-TOF/TOF Mass Spectrometer (Bruker) in reflector mode.

3.8.11 Z'-LYTE Kinase Activity and LanthaScreen Eu Kinase Binding Assays

Z'-LYTE were performed either by submitting compounds to Invitrogen's SelectScreen Kinase Profiling service or according to manufacturer's protocols. Z'-LYTE assays for MAP2K6 were performed using in the cascade format using the Ser/Thr 03 assay kit.

MAP2K6 protein was either purchased from Invitrogen (wild-type) or recombinantly expressed. MAPK12 (inactive) was purchased from Invitrogen.

For determination of compound IC_{50} 's, MAP2K6 (final concentration 0.5-5 $\mu\text{g}/\text{mL}$) was combined in a 10 microliter reaction with MAPK12 (inactive, 5 $\mu\text{g}/\text{mL}$), 100 nM ATP, and 2 μM Ser/Thr 03 peptide (Invitrogen) and inhibitor in 1x Kinase Buffer A (Invitrogen). Each reaction condition was measured in quadruplicate. MAP2K6 concentration was chosen to yield ~30-40% phosphorylation of the peptide substrate at the assay endpoint in the absence of inhibitor. After 1 hour at room temperature, 5 μL of Development Reagent A (Invitrogen), diluted 1:1024 in Development Buffer B (Invitrogen) was added to each reaction. After another hour at room temperature, 5 μL of Stop Reagent (Invitrogen) were added to the reaction. The ratio of emissions at 520 nm and 445 nm, after excitation at 400 nm, was measured and the percent phosphorylation of the peptide substrate calculated according to manufacturer's protocols.

For assaying the ability of ethacrynic acid to inhibit MAP2K6 after dialysis, we incubated MAP2K6 or MAP2K6 C38A (1 mg/mL) with ethacrynic acid (100 μM) or DMSO (0.5 %) in 200 μL storage buffer. After 1 hour at room temperature, the proteins were buffer exchanged into Kinase Buffer A using two successive Zeba Spin Desalting Columns (ThermoFisher Scientific) according to manufacturer's protocols. Kinase activity was immediately assayed according to manufacturer's protocols for the Z'-LYTE Ser/Thr 03 peptide kinase assay.

LanthaScreen Eu assays were performed at Invitrogen's SelectScreen Kinase profiling service using commercially available proteins.

3.9 References

1. Zimmermann, G. & Neri, D. DNA-encoded chemical libraries: foundations and applications in lead discovery. *Drug Disc. Today* **21**, 1828–1834 (2016).
2. Goodnow, R. A., Dumelin, C. E. & Keefe, A. D. DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nat. Rev. Drug Discov.* **16**, 131–147 (2017).
3. Salamon, H., Klika Škopić, M., Jung, K., Bugain, O. & Brunschweiler, A. Chemical biology probes from advanced DNA-encoded libraries. *ACS Chem. Biol.* **11**, 296–307 (2016).
4. Maianti, J. P. *et al.* Anti-diabetic activity of insulin-degrading enzyme inhibitors mediated by multiple hormones. *Nature* **511**, 94–98 (2014).
5. Wichert, M. *et al.* Dual-display of small molecules enables the discovery of ligand pairs and facilitates affinity maturation. *Nat. Chem.* **7**, 241–249 (2015).
6. Soutter, H. H. *et al.* Discovery of cofactor-specific, bactericidal Mycobacterium tuberculosis InhA inhibitors using DNA-encoded library technology. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E7880–E7889 (2016).
7. Wu, Z. *et al.* Cell-based selection expands the utility of DNA-encoded small-molecule library technology to cell surface drug targets: identification of novel antagonists of the NK3 tachykinin Receptor. *ACS Comb. Sci.* **17**, 722–731 (2015).
8. Zhao, P. *et al.* Selection of DNA-encoded small molecule libraries against unmodified and non-immobilized protein targets. *Angew. Chem. Int. Ed. Engl.* **53**, 10056–10059 (2014).
9. Li, G. *et al.* Design, preparation, and selection of DNA-encoded dynamic libraries. *Chemical Science* **6**, 7097–7104 (2015).
10. McGregor, L. M., Jain, T. & Liu, D. R. Identification of ligand-target pairs from combined libraries of small molecules and unpurified protein targets in cell lysates. *J. Am. Chem. Soc.* **136**, 3264–3270 (2014).
11. Wu, C.-Y. *et al.* Rapid discovery of functional small molecule ligands against proteomic targets through library-against-library screening. *ACS Comb. Sci.* **18**, 320–329 (2016).
12. Petros, A. M. *et al.* Rationale for Bcl-xL/Bad peptide complex formation from structure, mutagenesis, and biophysical studies. *Protein Sci.* **9**, 2528–2534 (2000).
13. Glass, D. B., Cheng, H. C., Mende-Mueller, L., Reed, J. & Walsh, D. A. Primary structural determinants essential for potent inhibition of cAMP-dependent protein kinase by

- inhibitory peptides corresponding to the active portion of the heat-stable inhibitor protein. *J. Biol. Chem.* **264**, 8802–8810 (1989).
14. Li, X. *et al.* PRKX, a phylogenetically and functionally distinct cAMP-dependent protein kinase, activates renal epithelial cell migration and morphogenesis. *P. Nat. Acad. Sci. USA* **99**, 9260–9265 (2002).
 15. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010).
 16. Brehmer, D., Godl, K., Zech, B., Wissing, J. & Daub, H. Proteome-wide identification of cellular targets affected by bisindolylmaleimide-type protein kinase C inhibitors. *Mol. Cell Proteomics* **3**, 490–500 (2004).
 17. Molnar, J. & Somberg, J. C. The clinical pharmacology of ethacrynic acid. *Am. J. Ther.* **16**, 86–92 (2009).
 18. Maron, B. A.; Rocco, T. P., *Goodman & Gilman's: The Pharmacological Basis of Therapeutics, 12e*, Brunton, L. L.; Chabner, B. A.; Knollmann, B. C., Eds. McGraw-Hill Education: New York, NY, 2011.
 19. Daguer, J.-P. *et al.* Identification of covalent bromodomain binders through DNA display of small molecules. *Angew. Chem. Int. Ed. Engl.* **54**, 6057–6061 (2015).
 20. Raman, M., Chen, W. & Cobb, M. H. Differential regulation and properties of MAPKs. *Oncogene* **26**, 3100–3112 (2007).
 21. Liu, Q. *et al.* Developing irreversible inhibitors of the protein kinase cysteinome. *Chem. Biol.* **20**, 146–159 (2013).
 22. Weerapana, E. *et al.* Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **468**, 790–795 (2010).
 23. Matsumoto, T. *et al.* Crystal structure of non-phosphorylated MAP2K6 in a putative auto-inhibition state. *J. Biochem.* **151**, 541–549 (2012).
 24. Papadopoulos, J. S. & Agarwala, R. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **23**, 1073–1079 (2007).
 25. Diao, Y. *et al.* Oxidation-induced intramolecular disulfide bond inactivates mitogen-activated protein kinase kinase 6 by inhibiting ATP binding. *P. Nat. Acad. Sci. USA.* **107**, 20974–20979 (2010).
 26. Warr, N., Siggers, P., Carré, G.-A., Wells, S. & Greenfield, A. Genetic analyses reveal functions for MAP2K3 and MAP2K6 in mouse testis determination. *Biol. Reprod.* **94**, 103 (2016).

27. Mitra, A. P. *et al.* The use of genetic programming in the analysis of quantitative gene expression profiles for identification of nodal status in bladder cancer. *BMC Cancer* **6**, 1 (2006).
28. Parray, A. A *et al.* MKK6 is Upregulated in Human Esophageal, Stomach, and Colon Cancers. *Cancer Invest.* **32**, 416-422. (2014).
29. Lu, D. *et al.* Ethacrynic acid exhibits selective toxicity to chronic lymphocytic leukemia cells by inhibition of the Wnt/ β -Catenin pathway. *PLoS ONE* **4**, e8294 (2009).
30. Yang, X *et al.* A public genome-scale lentiviral expression library of human ORFs. *Nat. Methods.* **8**, 659–661 (2011).
31. Zambaldo, C., Daguer, J. P., Saarbach, J., Barluenga, S. & Winssinger, N. Screening for covalent inhibitors using DNA-display of small molecule libraries functionalized with cysteine reactive moieties. *MedChemComm* **7**, 1340–1351 (2016).
32. Backus, K. M. *et al.* Proteome-wide covalent ligand discovery in native biological systems. *Nature* **534**, 570–574 (2016).
33. Markham, N. R. & Zuker, M. UNAFold: software for nucleic acid folding and hybridization. In *Bioinformatics, Volume II. Structure, Function and Applications*, Keith, J. M., Ed. Humana Press: Totowa, NJ; pp 3-31. (2008)
34. Nibbs, A. E., Baize, A.-L., Herter, R. M. & Scheidt, K. A. Catalytic asymmetric alkylation of substituted isoflavanones. *Org. Lett.* **11**, 4010–4013 (2009).

Chapter 4: Future directions for DNA-templated libraries

Alix I. Chan and David R. Liu

I proposed and carried out all experiments described in this section. I am grateful to Phillip Lichtor and Travis Blum for helpful discussions.

4.1 A new DNA-templated reaction: Suzuki-Miyaura couplings

The use of amino acid building blocks in the Tse [1] and Usanov [2] libraries had many advantages, including two reactive chemical groups (amine and carboxylic acid) for iterative assembly and commercially available, properly protected monomers for DNA-templated and solid-supported synthesis. However, the exclusive use of amino acid building blocks comes with several caveats, such as the polarity of the H-bond donor/acceptor in every amide bond formed, posing a limitation in making druglike compounds. In addition, while the side chains of unnatural amino acids span a wide swath of chemical space, it would be fruitful to explore a more diverse array of chemical building blocks that can be coupled through alternative reactions for future DTS libraries. Many different DTS-compatible reactions were explored in the past [3,4]. However, since then a very large number of DNA-compatible reactions [5-7] have been reported and developed. It would immensely enabling to adapt some of these to DNA-templated chemical library synthesis.

One possibility is to adapt the Suzuki-Miyaura coupling-based iterative C-C bond formation strategy reported by Burke and coworkers [8,9] to DNA-templated synthesis. These haloboronic acid building blocks also have two chemical handles for chemical elaboration - a vinylic or aryl halide and an N-methyliminodiacetic acid (MIDA)-protected boronic acid. Iterative, solid-phase synthesis is enabled by using MIDA to selectively (de)protect the boronic acids when desired. Figure 4.1 illustrates the parallels between peptide synthesis and the Burke method.

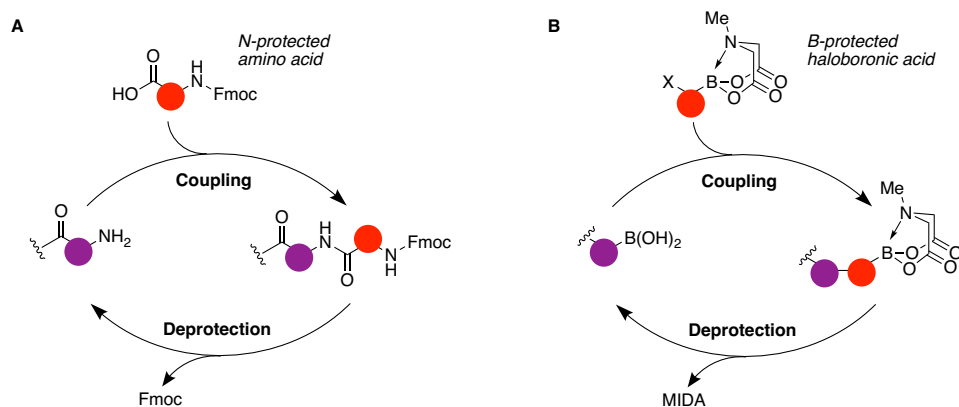


Figure 4.1. The parallels between solid-phase peptide synthesis (A) and iterative MIDA-protected haloboronic acid cross couplings (B). Both strategies rely on alternating steps of monomer coupling and deprotection to create the final synthetic products. Figure adapted from [9].

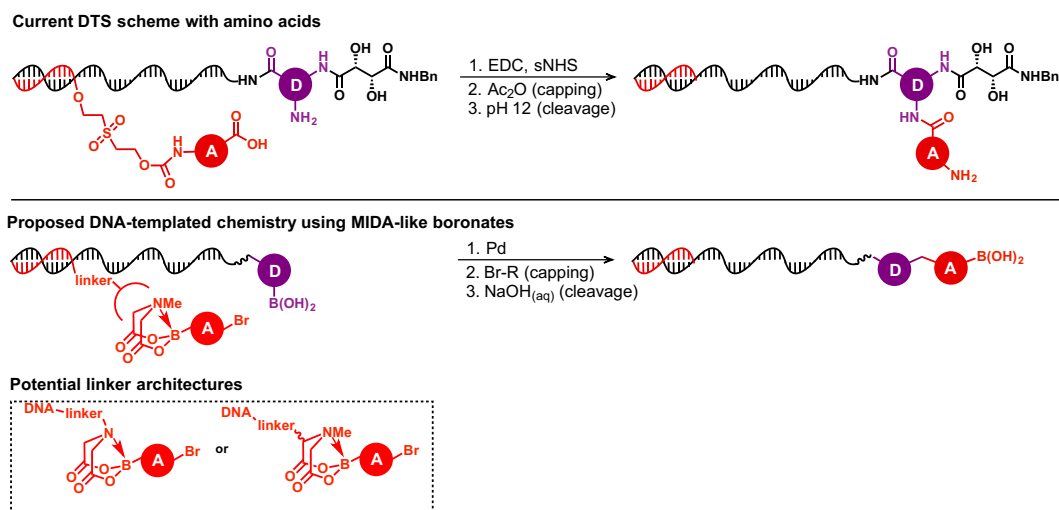


Figure 4.2. Schemes of current DTS strategy using amino acid building blocks (top), and an analogous DNA-templated synthetic scheme using MIDA-like boronates (middle). Potential strategies for linking the MIDA group to DNA are highlighted (bottom).

Given the similarities in these two chemical strategies, iterative Suzuki couplings could theoretically be adapted to our DNA-templated reaction scheme (Figure 4.2). DNA-compatible Suzuki reactions have been reported in applications such as labeling of nucleosides/nucleotide [10] as well as combinatorial library assembly [11]. The main non-obvious step is that the current DTS scheme relies on the BSOCOES (bis(2-(succinimidooxycarbonyloxy)ethyl)sulfone) linker to (a) protect the nucleophile on the

incoming building block, (b) reveal the amine after successful coupling, and (c) link the reagent to DNA. The MIDA group already (a) protects the boronic acid on the incoming reagent and (b) selectively reveals this coupling partner when desired. It should be chemically possible to modify the MIDA group to be a similarly “scarless” linker to DNA; two sites for possible modification (at the methylenes or methyl group in MIDA) are highlighted (Figure 4.2).

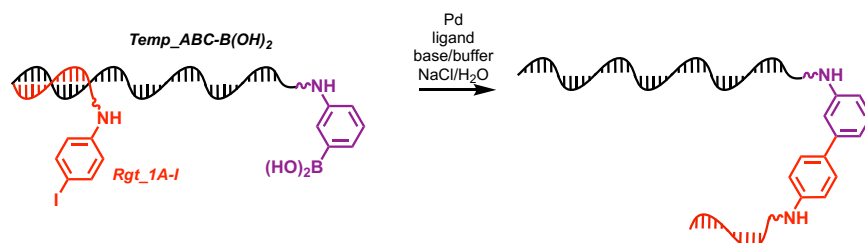


Figure 4.3. General schematic for initial DTS Suzuki test reactions.

To validate the first step of this strategy, I tested to see if a DNA-templated reaction was possible between a phenylboronic acid-bearing DNA template (**Temp_ABC-B(OH)₂**) and an iodobenzene-carrying DNA reagent (**Rgt_1A-I**). I tested a number of Pd(II) salts, water-soluble ligands (see Figure 4.4), and sources of base, among other variables, to see if it was possible to observe a DNA-templated reaction between the template and reagent DNA strands (Table 4.1). Initial conditions were chosen based on literature reports for aqueous, (oligo)nucleotide-compatible reactions that should also allow for the DNA hybridization necessary for DTS (i.e. temperatures below 37 °C) [10, 12-16]. Reactions were monitored by PAGE.

Pd(II) salt	Ligand	Base	Alkali metal salt	Alkaline metal salt concentration	Temp.	Time
Pd(OAc) ₂	TPPTS	Na ₂ CO ₃ (2 eq.)	NaCl	0.1-2.5M	25 °C	18-72 h
Pd(NO ₃) ₂	TXPTS	Tris pH 8	LiCl		30 °C	
Na ₂ PdCl ₄	ADHP	sodium phosphate pH 8	KCl		37 °C	
PdSO ₄		sodium carbonate pH 8	CsCl			

Table 4.1. Initial variables tested in DNA-templated Suzuki reactions. Not all possible combinations of all variables were tested.

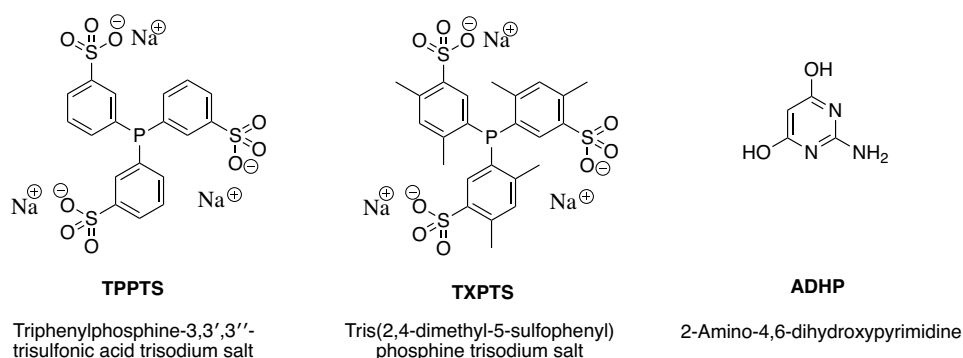


Figure 4.4. Structures of water-soluble Pd ligands tested in DTS Suzuki reactions.

I found that with 50 mM Tris pH 8, 0.5-1.0 M NaCl, and the TPPTS ligand I was able to observe 20-30% conversion from starting template to a larger product that was consistent with a DNA-templated reaction. Multiple Pd(II) salts worked under these conditions to catalyze product formation (Pd(OAc)₂, Pd(NO₃)₂, PdSO₄). Examples of successful reactions are shown in Figure 4.5.

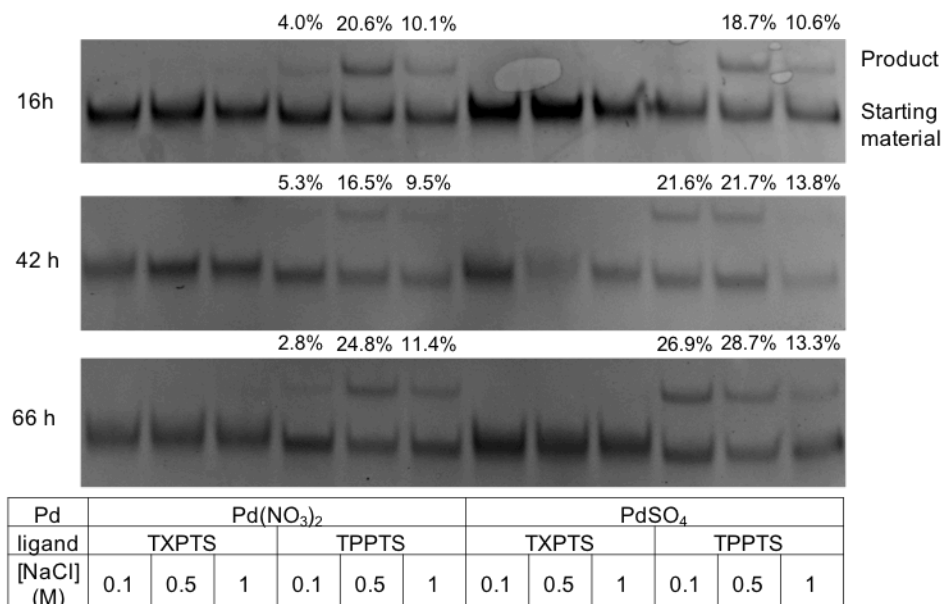


Figure 4.5. Representative gels showing DNA-templated cross-coupling between **Temp_ABC-B(OH)₂** (starting material) and **Rgt_1A-I** (not detected by staining) to form a higher molecular weight product. Percent conversions, as measured by densitometry, are shown where obvious product bands were observed. For all reactions, [Pd] = 1 mM, [ligand] = 2.5 mM, the base was 50 mM Tris pH 8, and reactions were carried out at 30°C. Reactions with Pd(NO₃)₂ were carried out in water, reactions with PdSO₄ contained 5% MeCN in water.

These reactions only work at 30° C (not 25 or 37 °C) and require a long incubation time (up to 3 days). Ideally, reactions would operate on a timescale and yield closer to a few hours and >90%, to make iterative and combinatorial library synthesis feasible. However, this is the first demonstration that a DNA-templated Suzuki reaction is possible and could potentially be used in future library syntheses. An immediate next step would be to validate if similar reaction conditions are able to catalyze DNA-templated cross coupling of other vinylic or aryl boronic acid and halide substrates. For further validation, chemistry efforts are required to develop the as-of-yet theoretical ‘scarless linker’ to reversibly link the boronic acid to DNA via a MIDA-like scaffold.

4.2 Outlook

If haloboronic acid-based DNA-templated C-C couplings can be optimized for library synthesis, we could combinatorially assemble libraries of new classes of compounds based on terpenoids, arenes, or other building blocks that explore chemical space beyond strictly amino acid-based molecules. This synthetic strategy could also be employed in multireaction DTS schemes [17] where both C-C and amide bond formation steps are employed (we could still even make macrocycles using the same ring-closing Wittig reaction). In addition, other strategies for iterative cross coupling, through alternate boron-protecting strategies or using orthogonal chemistries such as Buchwald-Hartwig amination or Sonogashira coupling [18] could be explored. Especially given the recent optimizations to the codon sets and other improvements to the DTS methodology [2], access to alternative iterative chemistry should be especially enabling in our library syntheses and subsequent selections. Taken together, these advances should allow future DTS libraries to explore previously untapped chemical space and provide fruitful starting points for the development of new bioactive molecules.

4.3 Experimental Methods

See section 3.8.1 for general methods.

4.3.1 Synthesis of modified DNAs

Coupling conditions were adapted from [19]. **Temp_ABC-B(OH)₂** was prepared by mixing 10 μL of 5' amine-modified DNA (5 mM in H₂O, /5AmMC6/CCCTGTACACAGACTCAAGTTGTCGATATGATGGCTTTCTACATCCCACTC-3' (IDT)) with 2.5 μL 3-carboxyphenylboronic acid (1 M in DMF, Sigma Aldrich) and 5 μL DMTMM*BF₄ (1 M in DMSO, Sigma Aldrich) in 200 mM borate buffer pH 9.4 (25 μL total reaction volume). The reaction was agitated overnight at room temperature. The reaction was desalted using a Nap5 column (GE Health Sciences) according to manufacturer's protocols and purified via HPLC. The purified product was analyzed by LC/MS. Expected mass 15847.1, found [m-2H₂O] = 15,815.5. The loss of two water molecules in ESI-MS is consistent with other literature reports of boronic acid-modified oligonucleotides [20].

Rgt_1A-I was prepared by mixing 10 μL of 3' amine-modified DNA (5 mM in H₂O, 5'-TAGAAGCCTATAGGG/3AmMO/ (IDT)) with 2.5 μL 4-iodobenzoic acid (1 M in DMF) and 5 μL DMTMM*BF₄ (1 M in DMSO) in 200 mM borate buffer pH 9.4 (25 μL total reaction volume). The reaction was agitated overnight at room temperature. The reaction was desalted using a Nap5 column (GE Life Sciences) according to manufacturer's protocols and purified via HPLC. The purified product was analyzed by LC/MS. Expected mass 5091.2, found mass 5,090.5.

4.3.2 Representative test Suzuki reaction

Temp_ABC_B(OH)₂ (20 pmol) and **Rgt_1A-I** (30 pmol) were mixed with 1 μL PdSO₄ (10 mM in 1:1 water/acetonitrile), 0.5 μL TPPTS (50 mM in water), 5 μL NaCl (1 M in

water), and 0.5 μL Tris pH 8 (1M in water) and water for a total reaction volume of 10 μL . The reaction was incubated at 30 °C for 66 hours and the DNA recovered by ethanol precipitation (using 3 volumes cold ethanol: 0.1 volume 3M sodium acetate). Reaction yield was quantitated by denaturing polyacrylamide gel electrophoresis (10% acrylamide in TBE-Urea) followed by staining with ethidium bromide, UV visualization, and densitometry (using ImageJ) of the product and template starting material bands. Yields were calculated assuming that the templates and products stained with intensity linearly correlated to mass. Calculated conversion: 29%.

4.4 References

1. Tse, B. N., Snyder, T. M., Shen, Y. & Liu, D. R. Translation of DNA into a library of 13 000 synthetic small-molecule macrocycles suitable for *in vitro* selection. *J. Am. Chem. Soc.* **130**, 15611–15626 (2008).
2. Usanov, D.L.; Chan, A.I.; Maianti, J.P.; Liu, D.R. “Second-Generation DNA-Templated Macrocycle Libraries for the Discovery of Bioactive Small Molecules.” *Nat. Chem.* In press (2018).
3. Gartner, Z. J., Kanan, M. W. & Liu, D. R. Expanding the reaction scope of DNA-templated synthesis. *Angew. Chem. Int. Ed. Engl.* **41**, 1796–1800 (2002).
4. Li, X., Gartner, Z. J., Tse, B. N. & Liu, D. R. Translation of DNA into synthetic N-acyloxazolidines. *J. Am. Chem. Soc.* **126**, 5090–5092 (2004).
5. Malone, M. L. & Paegel, B. M. What is a ‘DNA-compatible’ reaction? *ACS Comb. Sci.* **18**, 182–187 (2016).
6. Luk, K.-C. & Satz, A.L. “DNA-compatible chemistry.” In *A Handbook for DNA-Encoded Chemistry: Theory and Applications for Exploring Chemical Space and Drug Discovery, First Edition*. 67-98 (John Wiley & Sons, Inc. 2014).
7. Satz, A. L. *et al.* DNA compatible multistep synthesis and applications to DNA encoded libraries. *Bioconjug. Chem.* **26**, 1623–1632 (2015).
8. Li, J. *et al.* Synthesis of many different types of organic small molecules using one automated process. *Science* **347**, 1221–1226 (2015).
9. Li, J., Grillo, A. S. & Burke, M. D. From synthesis to function via iterative assembly of N-methyliminodiacetic acid boronate building blocks. *Acc. Chem. Res.* **48**, 2297–2307 (2015).
10. Hervé, G., Sartori, G., Enderlin, G., Mackenzie, G. & Len, C. Palladium-catalyzed Suzuki reaction in aqueous solvents applied to unprotected nucleosides and nucleotides. *RSC Adv.* **4**, 18558-18594 (2014).
11. Ding, Y. *et al.* Design and Synthesis of Biaryl DNA-Encoded Libraries. *ACS Comb. Sci.* **18**, 625–629 (2016).
12. Lercher, L., McGouran, J. F., Kessler, B. M., Schofield, C. J. & Davis, B. G. DNA modification under mild conditions by Suzuki-Miyaura cross-coupling for the generation of functional probes. *Angew. Chem. Int. Ed. Engl.* **52**, 10553–10558 (2013).
13. Omumi, A., Beach, D. G., Baker, M., Gabryelski, W. & Manderville, R. A. Postsynthetic guanine arylation of DNA by Suzuki–Miyaura cross-coupling. *J. Am. Chem. Soc.* **133**,

42–50 (2011).

14. Cahová, H. & Jäschke, A. Nucleoside-based diarylethene photoswitches and their facile incorporation into photoswitchable DNA. *Angew. Chem. Int. Ed. Engl.* **52**, 3186–3190 (2013).
15. Western, E. C., Daft, J. R., Johnson, E. M., Gannett, P. M. & Shaughnessy, K. H. Efficient one-step Suzuki arylation of unprotected halonucleosides, using water-soluble palladium catalysts. *J. Org. Chem.* **68**, 6767–6774 (2003).
16. Enderlin, G., Sartori, G., Hervé, G., & Len, C. Synthesis of 6-arylluridines via Suzuki–Miyaura cross-coupling reaction at room temperature under aerobic ligand-free conditions in neat water. *Tet. Lett.* **54**, 3374–3377 (2013).
17. Li, X., Gartner, Z. J., Tse, B. N. & Liu, D. R. Translation of DNA into synthetic N-acyloxazolidines. *J. Am. Chem. Soc.* **126**, 5090–5092 (2004).
18. Wang, C. & Glorius, F. Controlled iterative cross-coupling: on the way to the automation of organic synthesis. *Angew. Chem. Int. Ed. Engl.* **48**, 5240–5244 (2009).
19. Ding, Y. & Clark, M. A. Robust Suzuki–Miyaura cross-coupling on DNA-linked substrates. *ACS Comb Sci* **17**, 1–4 (2015).
20. Lin, N. *et al.* Design and synthesis of boronic-acid-labeled thymidine triphosphate for incorporation into DNA. *Nuc. Acids Res.* **35**, 1222–1229 (2007).