Automated screening for small organic ligands using DNA-encoded chemical libraries

Willy Decurtins[1], Moreno Wichert[1,2], Raphael M Franzini[1,2], Fabian Buller[1,2], Michael A Stravs[1,2], Yixin Zhang[1,2], Dario Neri[1] & Jörg Scheuermann[1]

[1]Department of Chemistry and Applied Biosciences, Institute of Pharmaceutical Sciences, Swiss Federal Institute of Technology (ETH Zurich), Zurich, Switzerland. [2]Present addresses: Roche Pharma Research and Early Development, Roche Innovation Center, Basel, Switzerland (M.W.); Department of Medicinal Chemistry, University of Utah, Salt Lake City, Utah, USA (R.M.F.); Covagen AG, Schlieren, Switzerland (F.B.); Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland, and Institute of Biogeochemistry and Pollutant Dynamics, ETH Zürich, Zürich, Switzerland (M.A.S); B CUBE Center for Molecular Bioengineering, Technische Universität Dresden, Dresden, Germany (Y.Z.). Correspondence should be addressed to D.N. (dario.neri@pharma.ethz.ch) or J.S. (joerg.scheuermann@pharma.ethz.ch).

Published online 17 March 2016; corrected online 1 April 2016 (details online); doi:10.1038/nprot.2016.039

DNA-encoded chemical libraries (DECLs) are collections of organic compounds that are individually linked to different oligonucleotides, serving as amplifiable identification barcodes. As all compounds in the library can be identified by their DNA tags, they can be mixed and used in affinity-capture experiments on target proteins of interest. In this protocol, we describe the screening process that allows the identification of the few binding molecules within the multiplicity of library members. First, the automated affinity selection process physically isolates binding library members. Second, the DNA codes of the isolated binders are PCR-amplified and subjected to high-throughput DNA sequencing. Third, the obtained sequencing data are evaluated using a C++ program and the results are displayed using MATLAB software. The resulting selection fingerprints facilitate the discrimination of binding from nonbinding library members. The described procedures allow the identification of small organic ligands to biological targets from a DECL within 10 d.

INTRODUCTION

Currently available methods for the discovery of novel small-molecule protein binders

Virtually all drugs are characterized by their ability to selectively bind to one or more target proteins. In conventional drug discovery, when a protein has been validated as a drug target, the hit discovery process for small molecules begins with the high-throughput screening (HTS, nomenclature in **Box 1**)[1–3] of individual compounds (one by one) in large collections of molecules, which are generally called chemical libraries. To discover binding molecules, an assay is needed, which is compatible with automation and reveals a desired biomolecular interaction with the target protein of interest. The conventional screening of large chemical libraries by high-throughput methods works well for certain classes of protein targets (e.g., kinases[3,4]), and it represents one of the drug discovery backbones in the pharmaceutical industry.

However, HTS procedures have several limitations: the synthesis, quality control and management of conventional chemical libraries are associated with high costs, lengthy procedures and complex logistics. These constraints restrict *de facto* the practice of HTS with big libraries (i.e., 1 million compounds or more) to large industries or to few large consortia with sufficient economic resources. Furthermore, not all protein targets can be 'drugged' with this procedure or can be produced in sufficient amounts for the individual screening of hundreds of thousands of compounds.

Further developments in drug discovery include fragment-based discovery approaches[5,6] and virtual drug discovery[7–9], in which small libraries of soluble chemical fragments or large virtual collections of molecules, respectively, are screened for binders.

DECLs

The recent development of DECLs[10–12] (**Fig. 1**) has allowed for the creation and screening of libraries of very large size, which can no longer be handled in a 'one well, one compound' manner as in HTS. The main principle of DECLs, as suggested by Brenner and Lerner[13], as well as by Gallop and co-workers[14] in the early 1990s, is to directly link chemical building blocks to oligonucleotides. The identity of the linked building blocks can easily be determined from the DNA coding tags, because the DNA sequence associated with each building block is known. A number of different types of DECLs have been developed, and they can be classified as single-pharmacophore and dual-pharmacophore chemical libraries, depending on whether the displayed compounds are attached to one or two strands of a double-stranded oligonucleotide.

In DNA-encoded single-pharmacophore libraries, small organic molecules are coupled to one DNA strand. In 'DNA-recorded' single-pharmacophore libraries[15–21] (**Fig. 1a**), the oligonucleotide part mainly serves as a DNA barcode, which allows the identification of the individual chemical compounds. The split-and-pool synthesis approach[18,20,22] has been shown to be a versatile tool for the incorporation of various types of chemical moieties into encoded libraries. In this synthetic strategy, sets of building blocks are chemically coupled to form more complex molecular structures. After each synthesis step, suitable oligonucleotides containing a coding sequence are added to the molecular entity, thus 'recording' the identity of the individual compounds. Alternatively, when performing 'DNA-templated' synthesis[23–27], single-pharmacophore libraries are generated on a library of preformed oligonucleotide templates, which contain the DNA codes for the identification of the individual compounds (**Fig. 1b**). At the same time, the hybridization of complementary oligonucleotide derivatives to the DNA template facilitates the transfer of building blocks to a nascent molecular structure. Bringing pairs of building blocks into close spatial proximity, the DNA hybridization step enables chemical reactions that normally do not work efficiently in water[23]. Furthermore, 'DNA-templated' synthesis is potentially compatible with the execution of multiple rounds of library synthesis and selection. The use of a universal

Box 1 | Nomenclature

DECL:	DNA-encoded chemical library. General term for all types of DNA-encoded compound collections.
ESAC:	Encoded self-assembling chemical library. Term for dual-pharmacophore libraries, which are constructed by self-assembly of two complementary sub-libraries.
HTDS:	High-throughput DNA sequencing using a next-generation DNA sequencing technology.
HTS:	High-throughput screening. One-by-one screening of large compound collections (= chemical libraries).
Set of building blocks:	Indicates the spatial position at which building blocks are incorporated (= diversity element[20,21]).
Building block:	A chemical moiety that is chemically coupled to a growing molecule (the displayed compound).
Displayed compound:	A chemical compound (resulting from building block conjugations) covalently linked to an oligonucleotide.
Single-pharmacophore DECL:	A chemical compound displayed at the end of one oligonucleotide strand.
Dual-pharmacophore DECL:	A pair of chemical moieties simultaneously displayed at the end of two assembled oligonucleotide strands.
Recorded DECL:	DECL in which the oligonucleotide part serves the purpose of identifying the attached compound.
Templated DECL:	DECL in which the oligonucleotide assembly also directs the building block in close proximity to the nascent structure, thus allowing its conjugation.

template[28], e.g., by using the ambiguous base-pairing property of deoxyinosine, could enable the generation of larger libraries by DNA-templated synthesis. Alternative ways for constructing single-pharmacophore libraries are the YoctoReactor system[29] and fluidic routing[30].

Dual-pharmacophore DECLs feature pairs of chemical building blocks, attached to adjacent sites on oligonucleotide assemblies, such as the extremities of complementary DNA strands or the junction of two oligonucleotides hybridized to a common template. Individual chemical moieties are typically brought into close spatial proximity with flexible linkers, thereby facilitating their interaction with cognate binding sites on the target protein of interest. Once synergistic building blocks are identified, some synthetic efforts are needed in order to find optimal linkers for the generation of binding molecules in the absence of DNA. Encoded self-assembling chemical (ESAC) libraries[31,32] (**Fig. 1c**) are formed by the combinatorial self-assembly of two complementary sub-libraries, which carry a chemical moiety on both the 5′ and the 3′ ends at the same side of a double-stranded DNA heteroduplex. For some applications, the use of peptide nucleic acids (PNAs)[33,34] rather than DNA offers certain advantages, such as a larger variety of compatible chemical reactions. In dual-pharmacophore PNA libraries (**Fig. 1d**), two PNA sub-libraries, each carrying a coding sequence and a chemical fragment, are hybridized to a complementary DNA template library, containing two specific coding regions[33,35–39].

Advantages of DECLs for drug discovery applications

The use of DECL technology presents distinctive advantages, compared with classical HTS lead discovery:

- A very large library size[18,40] can be obtained, depending on the number of sets of building blocks used[12]. For example, three sets of building blocks with 1,000 members each will yield a library with $1{,}000^3 = 10^9$ displayed compounds. A variety of DNA-compatible synthesis approaches are available[41,42].
- A DECL may be stored in only one small vessel, as the DNA barcodes allow the unambiguous identification of each library member. In contrast, members of conventional chemical libraries must be stored separately, typically in microtiter plates.
- Only a minute amount of a DECL is necessary to perform affinity-based selections against a target protein. One selection is performed in only one reaction vessel. This reduces costs substantially and facilitates the parallel screening of different target proteins.
- No protein-specific assays are needed, as selections with DECLs are solely based on affinity. The principle of affinity selections has been validated extensively for the discovery of target-specific biomacromolecules using a multitude of display technologies—e.g., antibody phage display technology[43–45], mRNA display[46,47], ribosome display[48], yeast display[49], as well as systematic evolution of ligands by exponential enrichment (SELEX) technology[50,51].
- The DNA linkage site of the small-molecule compound serves as a 'modifiable handle' for further medicinal chemistry optimization steps (e.g., introduction of solubility enhancing groups).

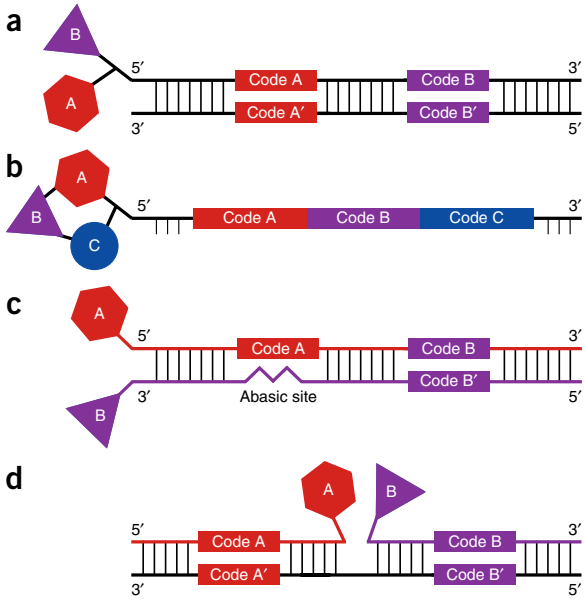


Figure 1 | Comparison of different DECL types. (**a**) Single-pharmacophore two-building-block 'DNA-recorded' library[20]. (**b**) Single-pharmacophore three-building-block 'DNA-templated' library[27]. (**c**) Dual-pharmacophore two-building-block ESAC library[32]. (**d**) Dual-pharmacophore two-building-block PNA/DNA hybrid library[38].

- Structure-activity relationships (SARs) can be obtained from the selection results if structurally related compounds are incorporated into the library[20].
- Simultaneously binding fragment pairs can be obtained from dual-pharmacophore fragment-based selections[32].
- The affordable setup makes DECLs an ideal tool for the academic community and for small companies.

Challenges and limitations of DECLs

The limitations of the DECL technology include:

- Sets of building blocks used for library construction are typically limited to two or three, as larger compounds are less 'drug-like'. In libraries based on the combinatorial assembly of multiple sets of building blocks, library size grows exponentially with the number of sets of building blocks. However, the incorporation of multiple building blocks in a molecule leads to properties that are more likely to violate Lipinski's rule of five (ref. 52) and may limit pharmaceutical development possibilities.
- The functional relevance of library size is well illustrated by the following analogy: human monoclonal antibodies are routinely discovered using combinatorial phage display libraries, containing billions of different antibody clones[43,44]. The technology works exceedingly well, and it is virtually always possible to isolate a specific antibody against any target protein of interest. However, if subaliquots of the antibody libraries are used, which contain only a few million antibodies, the process becomes much less efficient and good-quality antibodies are rarely isolated as a result of the process[53].
- In addition to library size, molecular design and library purity also have an important role, thus affecting the functional quality of a DECL. Although it is possible to purify conjugates that contain only one building block to very high purity by HPLC, this procedure cannot be repeated after the introduction of a second building block in a split-and-pool synthesis approach. Although methods for enhancing library purity have been proposed[54], the preparation of high-quality single-pharmacophore libraries remains challenging. In this respect, dual-pharmacophore ESAC libraries[32] offer the advantage of superior purity, as each member of sub-libraries capable of self-assembly can be individually purified by HPLC and subsequently characterized by mass spectrometry.
- The total number and diversity of the used building blocks may be as important as the total number of compounds in a library, for selecting molecules with the desired functional properties.
- Affinity-based selection assays may lead to target binders, which do not exert a functional agonistic or antagonistic effect. In contrast, allosteric binders may be obtained, which might stay undiscovered in a target-specific functional assay.
- For statistically relevant results, analysis of selections by high-throughput DNA sequencing (HTDS)[55,56] should best be performed by oversampling the library—i.e., using sequence counts that exceed the library size[57,58]. Despite the advances in HTDS technology, it may be difficult to achieve oversampling with very large libraries.
- Hit validation of selected binders may be cumbersome if too many hits are obtained from the affinity-based selection. Narrowing down the number of small organic molecules to be synthesized without a DNA handle will speed up hit-to-lead development. Technologies based on locked nucleic acid (LNA) display[32] or hybridization onto DNA slides[31,35] have been proposed.

Selection types for DECLs

Different methodologies for affinity-based selections using DECLs have been developed[29,59]. In this protocol, we describe affinity-based selections on a solid support: A target protein is immobilized on a surface matrix and subsequently incubated with the DECL. Washing steps remove nonbinding library members, whereas binding compounds are eventually removed from the protein and identified. Alternatively, both binding partners are allowed to interact in solution, and the formed complex is then captured on solid support. These solid-phase selection approaches, which are widely used in the field, yielded hit compounds against many classes of target proteins[16,18,20,26,32].

Solution-phase methods rely on the detection of ligand-target interactions without physical removal of nonbinding library members, thereby superseding the application of a solid support. Interaction-dependent PCR (IDPCR)[60] aims to reveal binding events between DNA-linked small molecules and DNA-linked target proteins. Enabled by close proximity of the binding partners, the DNA-tags may anneal and form a hairpin structure, which contains the information of both the ligand and the target. For interaction determination using unpurified proteins (IDUP)[61], an advanced setup of this methodology, the covalent DNA-linkage to the target protein is replaced by antibody-binding or fusion proteins. The indirect target linkage might facilitate the application of this method to unpurified targets in crude cell lysates. Binder trap enrichment (BTE)[29], in analogy to IDPCR, may allow for the selection of DNA-linked small molecules together with a DNA-linked target protein. After reaching equilibrium in solution, the ligand-protein solution is diluted and emulsified. The binding molecules are trapped in water-in-oil droplets, in which the DNA fragments are ligated and eventually identified after breaking the emulsion.

DNA-programmed affinity labeling (DPAL)[62] makes use of irradiation in order to covalently label previously unmodified target proteins. A single DNA-tagged compound, which is hybridized to a shorter capture probe, is incubated with a range of target proteins. Binding of the small molecule brings the capture probe in close proximity to the target protein, thereby facilitating its covalent attachment by irradiation. The ligand-specific target protein may then be identified using the introduced DNA-tag. The DPAL labeling system can be applied for selections[63] when a library of encoded small molecules, hybridized to a photoreactive probe, is incubated with one target protein. Nonbinding library members can be removed by enzymatic digestion, whereas covalently attached library members would be protected from digestion by proximity to the target.

Development and application of the protocol

The screening procedure described in this protocol was first applied in 2004 by Melkko *et al.*[31]. Since then, the protocol was used to screen different DECLs, allowing for the identification of small-molecule binders against alpha-1-acid glycoprotein[32], B cell lymphoma–extra large[15,64], BSA[19], carbonic anhydrase II (ref. 31), carbonic anhydrase IX (refs. 17,32), human serum albumin[15,19,20,22,31], interleukin 2 (ref. 16), matrix

metalloproteinase 3 (ref. 65), prostate-specific membrane antigen[20], rabbit serum albumin[19], streptavidin[66,67], tankyrase 1 (refs. 20,21), trypsin[15,68,69] and tumor necrosis factor alpha[15]. Over the past decade, literally all parts of the protocol were improved substantially: Sepharose beads were replaced by magnetic beads, automation followed manual selections, and selection analysis evolved from microarray-based decoding to more cost-efficient HTDS approaches.

In the initial protocol, Sepharose beads were used for affinity selections[15,17,22,31,57,64–69]. Target proteins were either covalently immobilized on cyanogen bromide (CNBr)-activated Sepharose beads or bound to streptavidin-Sepharose after protein biotinylation. Later, magnetic beads with cobalt-based chemistry[16] for the immobilization of His-tagged proteins or streptavidin-coated beads[19–21,32] (SA beads) for the immobilization of biotinylated proteins were introduced. Magnetic beads enabled the transition from manual to automated selections[19–21,32], using the KingFisher magnetic particle processor.

Means for selection decoding have drastically improved over the past years. Selection decoding provides sequence counts for individual library members before and after selection, thereby enabling the assessment of the enrichment factor achieved with the selection step. Although formerly we used DNA microarrays[31], the advent of HTDS technology[55,56] has greatly affected decoding, as efficient DNA sequencing procedures are crucially important for the success of DECL technology with large libraries. Our DECL team was the first to report the use of Roche's 454 HTDS technology for the decoding of DECL selections[57], replacing formerly practiced microarray-based methods. Nowadays, the Illumina/Solexa technology is more often used[58], which enables the decoding of larger libraries. For smaller libraries, however, decoding by microarray hybridization continues to be a convenient methodology[38].

Considerations on selection parameters

Many selection parameters may be adjusted to individual needs. For example, the solid support can consist of a streptavidin-coated magnetic bead, binding a biotinylated protein, or an activated Sepharose bead to which the target protein is covalently linked. An alternative to the selection automation using magnetic beads, as described in this protocol, is the application of Sepharose columns that fit on liquid handlers (e.g., PhyTips from Phynexus). Besides covalent protein binding and immobilization via biotinylation, tagging strategies such as His-tag, FLAG-tag, Strep-tag and GST-fusion[70–74] may be used. We prefer the streptavidin-biotin system, as the biotinylation of a target protein can easily and reliably be obtained and the interaction is strong enough to prevent dissociation during the course of the affinity-based selection.

In solid-phase selections, the capacity of the beads eventually determines the amount of tagged protein that can be displayed. We usually perform selections with at least two different protein concentrations: first, at conditions in which the beads are saturated with protein and, second, at conditions with lower protein display. The latter allows for more stringency and the identification of higher-affinity hits from the DECL[75]. Optimally, known protein-ligand systems (e.g., carbonic anhydrases/sulfonamides) should be included as positive control selections.

One of the parameters to be examined is the concentration of the DECL. A higher input generally leads to more DNA recovery from the selection, as monitored by quantitative PCR methods. If a large amount of DNA can be recovered, the selection eluate can be used as starting material for further rounds of affinity-based selections on fresh target beads (so-called pseudo-rounds of selection). In addition, different incubation times and buffer compositions may be tested and stringency may be increased with the number of washing steps[75].

Experimental design

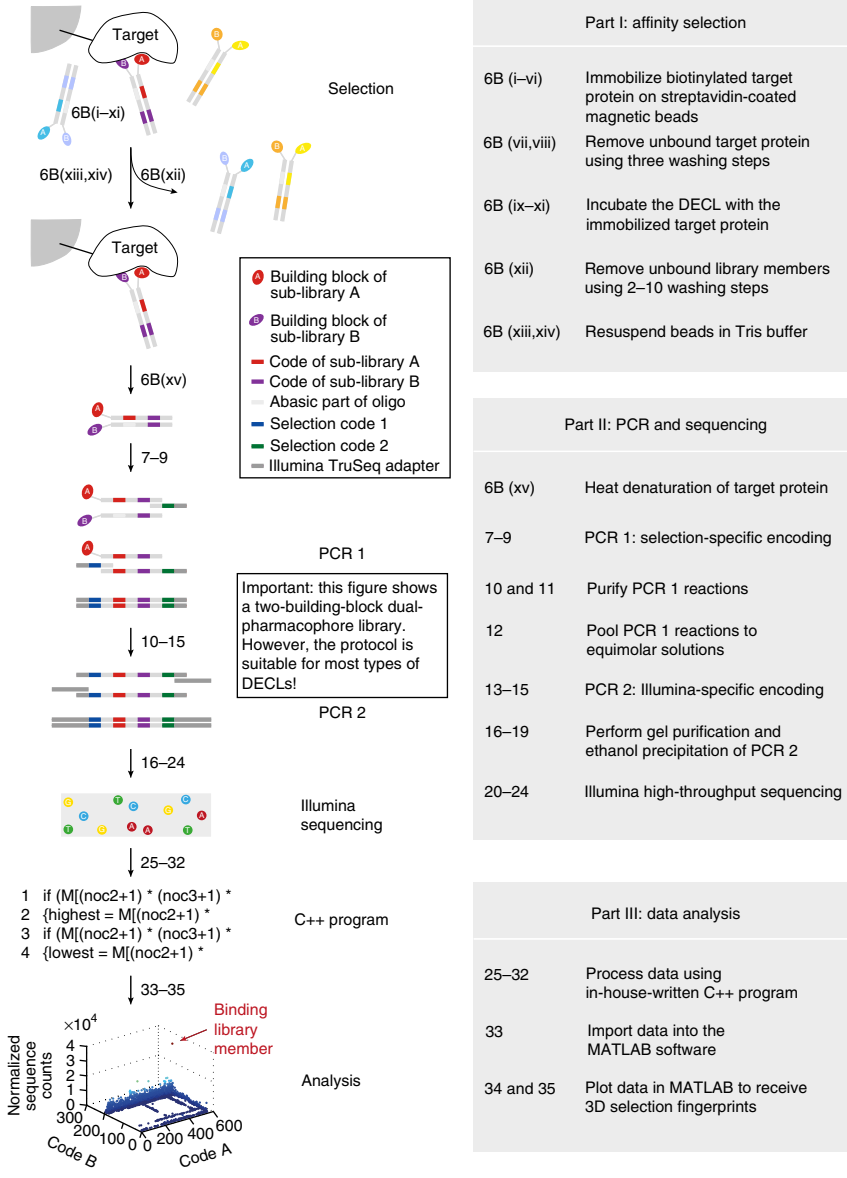
Protocol overview. With this protocol, we aim to provide a generally applicable, inexpensive and fast procedure to obtain small-molecule protein ligands from DECLs for pharmaceutical or chemical biology applications. The protocol is divided into three parts (**Fig. 2**). First, we describe the affinity selection step, which physically separates binding molecules from nonbinding library members. Second, the DNA part of selected binders is amplified in a two-step PCR and subjected to Illumina HTDS. Third, the resulting DNA sequences are processed and analyzed, thereby revealing the relative enrichment of individual library members in relation to the target protein used. The protocol delineates the procedure for libraries comprising two sets of building blocks, yet it can easily be adapted to DECLs comprising three or even four sets of building blocks, as described in the last section.

Part I: affinity selection. In affinity selection procedures, the target protein of interest is typically immobilized on magnetic beads and subsequently incubated with the encoded chemical library. If fitting small organic ligands happen to be present in the DECL, they will be retained on the magnetic beads because of a binding interaction with the immobilized target protein. Nonbinding library members are removed by several washing steps, whereas binding library members remain bound to the solid support and they are eventually dissolved in Tris buffer during the last step of the affinity selection. The use of Tris buffer is compatible with freezing of selection material and with the subsequent PCR amplification procedure.

We perform selections using biotinylated proteins. The chemical biotinylation procedure with *N*-hydroxysuccinimide–biotin reagents is mild and reliable. Our biotinylation protocol, adapted from the supplier's instructions, is provided as **Supplementary Method 1**. After the biotinylation reaction, the biotinylated protein needs to be purified from the free biotin by size-exclusion purification, and it should be quality-controlled on an SDS-PAGE gel. The success of the biotinylation reaction can be assessed using a band-shift assay, or, for some proteins, mass spectrometry (MS) analysis.

The selection buffer used for the affinity selection depends on the target protein. Most proteins can readily be stored and screened in PBS. Standard PBS types, as used for the majority of selections, are detailed in the reagent setup section. If PBS may not be used as selection buffer, e.g., for the screening of phosphatases, alternative buffers such as HEPES may be used. For simplicity, all steps of the affinity selection protocol are described for PBS.

At the beginning of the selection procedure, the biotinylated target protein is immobilized on magnetic beads in PBS, followed by two washing steps with PBS-Tween-biotin (PBST-biotin). The biotin blocks free binding sites on the SA beads in order



Part I: affinity selection	
6B (i–vi)	Immobilize biotinylated target protein on streptavidin-coated magnetic beads
6B (vii,viii)	Remove unbound target protein using three washing steps
6B (ix–xi)	Incubate the DECL with the immobilized target protein
6B (xii)	Remove unbound library members using 2–10 washing steps
6B (xiii,xiv)	Resuspend beads in Tris buffer

Part II: PCR and sequencing	
6B (xv)	Heat denaturation of target protein
7–9	PCR 1: selection-specific encoding
10 and 11	Purify PCR 1 reactions
12	Pool PCR 1 reactions to equimolar solutions
13–15	PCR 2: Illumina-specific encoding
16–19	Perform gel purification and ethanol precipitation of PCR 2
20–24	Illumina high-throughput sequencing

Part III: data analysis	
25–32	Process data using in-house-written C++ program
33	Import data into the MATLAB software
34 and 35	Plot data in MATLAB to receive 3D selection fingerprints

Figure 2 | Overview of the screening process. In part I, binding library members are identified from the multiplicity of library members. Part II describes the amplification of the eluted library members' DNA tags, followed by Illumina HTDS. The sequencing results are analyzed in part III using a C++ program and displayed in MATLAB.

to reduce selection of streptavidin binders from the library. Tween 20 is added to the selection solution in order to prevent the coagulation of magnetic beads and the sticking of beads to the plastic plate and tip comb. Tween 20 may reduce background from false-positive hits, which bind to the target protein by unspecific hydrophobic interactions.

Different types of magnetic beads can be considered for DECL selections: SA beads[19–21,32], beads with cobalt-based chemistry[16] for the immobilization of His-tagged proteins, as well as beads for covalent protein immobilization. We have made good experience with the use of SA beads, and these procedures are detailed in this protocol. Four types of SA beads are commercially available from Thermo Fisher Scientific: Dynabeads M-270 SA, M-280 SA, MyOne SA C1 and MyOne SA T1, which vary in diameter and in terms of whether the beads are pre-blocked with BSA. Depending on the DECL and target protein in use, a different type of SA beads may be optimal. We mostly use Dynabeads MyOne SA T1 (refs. 20,32) and Dynabeads M-270 SA (refs. 19–21). For each selection, 1 mg of magnetic beads[19–21,32] are used. The binding capacity varies with the type of beads. In the case of Dynabeads MyOne SA T1, 0.1 mg of beads has a binding capacity of 40 pmol of biotinylated peptide. However, also lower and higher amounts for protein immobilization may be considered.

The number of washing steps may be adjusted for the individual selection. If a high-affinity ligand can be expected, the stringency (e.g., number and duration) of the washing steps can be increased. By contrast, if a lower binding affinity is expected, fewer washing steps of shorter duration should be considered. A good compromise was found in the use of five washing steps with a duration of 30 s each[19–21,32].

Before a selection experiment, the DECL working solution needs to be prepared and the magnetic beads equilibrated to PBS. We store DECLs in water as concentrated stock solutions, which are diluted to 5 nM working solutions, defined as the concentration of individual library members multiplied by the number of library members (preparation of the DECL working solution, Steps 1 and 2). Herring sperm DNA is added to the DECL as a blocking agent in order to prevent nonspecific binding of the library-oligonucleotides to the target protein, streptavidin, or the beads themselves. The magnetic beads are washed and resuspended in the appropriate amount of PBS (washing the magnetic beads, Steps 3–5).

Originally, affinity selection assays using magnetic beads were performed in a manual manner with the use of a magnetic rack (manual affinity selection, Step 6B). Manual handling possibly introduces some operator bias, especially on the washing steps and when handling many selections in parallel. If short washing times are required, only a few (12 or fewer) samples may be handled in parallel. With the adaptation of the protocol to the use on the Thermo Fisher Scientific KingFisher magnetic particle processor, the quality, reproducibility and throughput of the selection assays were improved substantially (automated affinity selection, Step 6A). The KingFisher device can process up to 24 samples per run, which typically takes 135 min. A standard affinity-selection protocol on the KingFisher device is depicted in **Figure 3**, whereas the corresponding program parameters are summarized in **Table 1**. The complete program, which is ready for import into the KingFisher BindIt software, is provided as **Supplementary Software 1**, together with an example protocol status report as **Supplementary Data 1**.

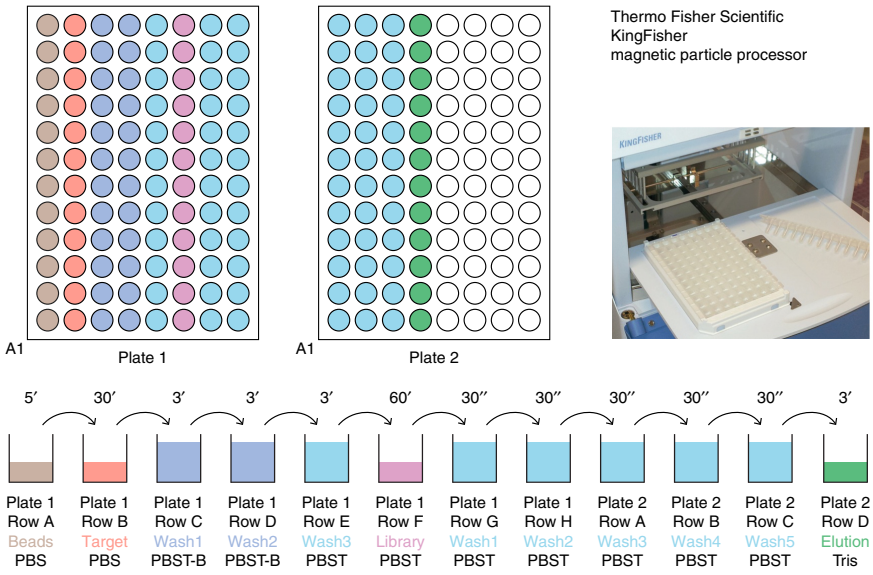


Figure 3 | Plate loading scheme. Top, KingFisher 200-µl plates shown from above. As the magnetic particle processor transfers all magnetic beads contained in a row (e.g., A1 to A12) during each step, the wells are loaded in a row-wise manner (e.g., target proteins in row B of plate 1). The respective solutions are filled into the wells as depicted at the bottom; 200 µl per well for all washing steps and 100 µl of washed beads, target protein, library and Tris buffer. The numbers above the arrows indicate the incubation time of the beads at each step. In this setup, each column allows the performance of an independent selection. Although the handling by the magnetic particle processor is identical for all plate columns, individual selection parameters may be varied in terms of target protein, DECL type and general buffer composition.

Part II: PCR and high-throughput sequencing. The DNA part of the eluted binding library members is amplified in a two-step PCR procedure (PCR amplification of oligonucleotide tags, Steps 7–19), and the resulting DNA amplicons are subjected to Illumina HTDS (Illumina high-throughput sequencing, Steps 20–24). The first PCR (**Fig. 4**, nucleotide sequences are provided in the **Supplementary Note**) amplifies the selected encoded library members for each selection separately. The DNA fragments, which are linked to compounds that have been recovered on beads at the end of the selection procedure (row D of plate 2 in **Fig. 3**), serve as template for these PCRs. Primers 'IlluminaPCR1a' and 'IlluminaPCR1b' introduce two additional DNA codes that are suitable for the identification of individual affinity selections, thus allowing the parallel HTDS of different selection experiments on the same Illumina flow lane. These PCR 1 reactions are purified using a PCR purification kit, and they are pooled to equimolar concentration and used as template for the next amplification step. The number of selections that may be pooled and analyzed in parallel on the same Illumina flow lane depends on both library size and the desired sequence count per selection. The second PCR (**Fig. 4**, nucleotide sequences provided in **Supplementary Note**), using primers 'IlluminaPCR2a' and 'IlluminaPCR2b', eventually introduces the TruSeq adapter sequence[56], which is required for HTDS on the Illumina HiSeq devices.

HTDS technologies[55,56] have greatly improved over the years, whereas costs per sequence have decreased, which makes the technology affordable for academic institutions. Driven by

TABLE 1 | Standard KingFisher program for DECL affinity selections.

	Beginning of step		Mixing/pause		End of step	
	Release time (s)	Release speed	Mixing time (min)	Mixing speed	Collect count	Collect time (s)
p1 A beads	–	–	5	Bottom mix	5	10
p1 B target	30	Medium	30	Medium	5	10
p1 C wash1	30	Medium	3	Medium	5	10
p1 D wash2	30	Medium	3	Medium	5	10
p1 E wash3	30	Medium	3	Medium	5	10
p1 F library	30	Medium	60	Medium	5	10
p1 G wash1	30	Medium	0.5	Medium	5	10
p1 H wash2	30	Medium	0.5	Medium	5	10
p2 A wash3	30	Medium	0.5	Medium	5	10
p2 B wash4	30	Medium	0.5	Medium	5	10
p2 C wash5	30	Medium	0.5	Medium	5	10
p2 D elution	30	Medium	3	Medium	–	–

The options 'Precollect', 'Pause' and 'Postmix' are not used.

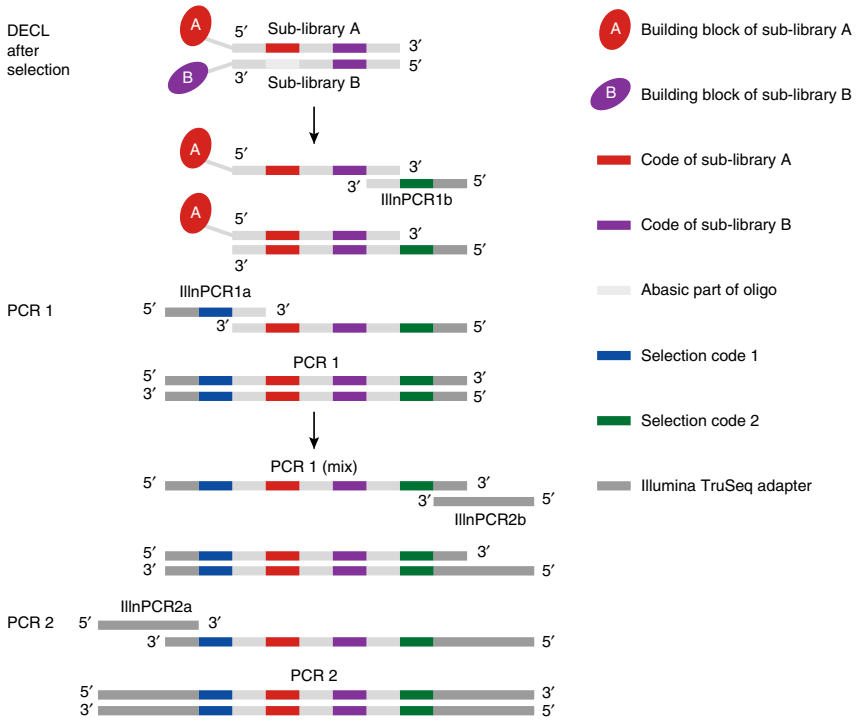


Figure 4 | Layout of the two-step PCR. In the first PCR, selection-specific codes (code 1 and code 2) are added. These reactions are purified, pooled and used as template for the second PCR, which introduces the DNA sequences required for Illumina sequencing.

efforts in genomic research, different and even-more-powerful HTDS technologies have been and continue to be developed, which may serve as an alternative to Illumina sequencing. The choice of a given HTDS technology for DECL selections also depends on the length of the DNA oligonucleotide to be sequenced (varying between 70 and 150 nt) and the desired number of sequence reads, as hundreds of millions of DNA sequences may be required for large libraries.

Illumina sequencing is based on the amplification of 'clustered' DNA strands, which are individually confined to small portions on the surface of the reaction chamber (i.e., an eight-lane flow cell)[56]. Each DNA cluster, which originates from one individual DNA sequence, can be analyzed by the sequential incorporation of fluorescently labeled DNA nucleotides, followed by iterative scanning procedures (DNA sequencing by synthesis). As a good spatial scanning resolution of the clusters cannot be obtained if all clusters are incorporating the same base at a time (as is the case when a constant region of a DECL is sequenced), dummy random genomic DNA (e.g., the 'PhiX Control v3' library) needs to be added to the DNA amplicon to be analyzed in order to obtain optimal sequencing results.

In the past, sequence length restrictions and obtained numbers of sequences posed severe constraints. Nowadays, Illumina HTDS constitutes a suitable platform for the application with DECL selections. By using the fourth version of Illumina HiSeq chemistry, 2×10^9 reads with a length of 125 nt may be obtained per flow cell in single-read mode. As one flow cell comprises eight flow lanes, one flow lane provides 250×10^6 reads. Owing to the addition of 30% dummy PhiX DNA, 175×10^6 reads per flow lane can be used in the best case. In general, it is desirable to oversample the library size (e.g., by a factor 10). This means that, for a library of 10^6 compounds, ~15 selections should best be sequenced on the same flow lane. Undersampling may be necessary for very large libraries (e.g., those with four or more sets of building blocks), but it may still yield valuable hits.

Part III: data analysis. HTDS delivers very large raw data files of up to 50 gigabytes per Illumina flow lane, which need to be analyzed (data analysis, Steps 25–35). An overview about the analysis process is given in **Figure 5**. First, we convert the standard output *.fastq data files into *.fasta files, thereby deleting the two out of four lines per sequence containing the sequence quality information. This information is not needed for our purpose, as, by default, the evaluation program will count only intact sequences—i.e., full length sequences with correct constant parts and codes of proper length.

We developed the C++ evaluation program 'count.cpp', provided as **Supplementary Software 2**, which processes the *.fasta HTDS data on the basis of a definition structure file (**Box 2** and **Supplementary Software 3**). This C++ program can be compiled and used on the platform of choice (PC, Mac, Unix). The structure file contains the path to the *.fasta file (path_to_sequence_file) to be evaluated, as well as a definition of the minimum length of the HTDS sequence (minimum_line_length) to be analyzed. The user may specify whether the applied DECL contains two, three or four sets of building blocks (output_type) and how many errors are allowed per constant region (mismatch_limit, default=0). Four different coding positions (code1-4) are defined, together with the number of coding sequences per code (code1-4_count) and the individual start (code1-4_startpos) and end (code1-4_endpos) positions in the nucleotide sequence (**Supplementary Note**). Separate code lists, which are stored at the path indicated at path_to_code1-4_list, provide the sequences used in the coding positions. An example codelist can be found in **Supplementary Software 4**. Further, the constant regions between the codes are defined (const1-3), their start (const_region_1-3_startpos) and end (const_region_1-3_endpos) positions, as well as the corresponding DNA sequence (const_region_1-3_seq). The example DECL setup provided in **Supplementary Note** and the corresponding structure file (**Box 2** and **Supplementary Software 3**) show a two-building block library[32].

The C++ program checks the millions of sequences obtained from HTDS to determine whether the constant domains match, and it counts the occurrence of the individual code combinations. Three output files are created: (i) 'name_datum_Codecounts.txt', (ii) 'name_datum_eval.txt' and (iii) 'name_datum_evalNorm.txt'. The output file 'name_datum_Codecounts.txt' gives an overview of the evaluation: the number of sequences contained in the *.fasta file, the number of sequences that could be evaluated according to the structure file settings, the codelists used and the

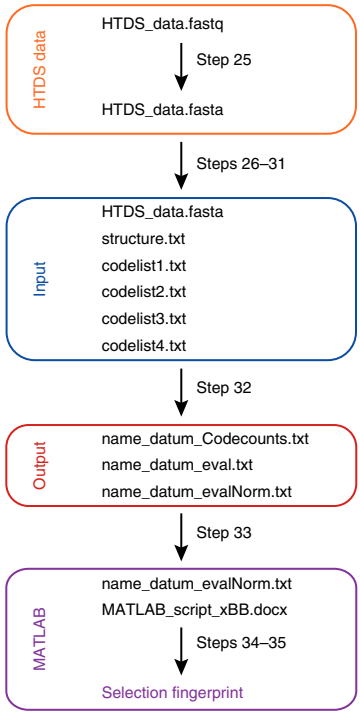


Figure 5 | Data processing workflow. Illumina HTDS raw data are converted from *.fastq to *.fasta. HTDS data, the structure file and the code lists are provided as input for the C++ program, which generates three output files. The normalized output file is imported into MATLAB. The selection fingerprint is obtained using a MATLAB script.

assignment to individual codes within the codelists. A code '0' means that no assignment was possible to a code of the respective codelist. The individual evaluation of a selection, i.e., the counting of code combinations, is given in the evaluation file 'name_datum_eval.txt'. This text file contains comma-separated values and includes a header row with all the code1_code2 combinations (i.e., the individual selections are given as columns), followed by a column with the codeA__codeB (= code3__code4) combinations and two columns with the code A (= code 3) and the code B (= code 4), respectively.

Eventually, each occurring codeA__codeB combination is counted for a given selection code1_code2. As the number of sequences obtained from HTDS per selection can vary considerably, we work with an internal normalization of a given selection (output file 'name_datum_evalNorm.txt'): The average counts of all compounds (i.e., the codeA__codeB combinations) in a given selection (code1_code2) are determined and the counts for each individual compound codeA__codeB are then divided by this average, multiplied with 100 and rounded to an integer value. This internal normalization, in our opinion, allows for an optimal comparison of selections for the enrichment of individual code combinations. An alternative approach is the external normalization, in which target selections are divided by empty bead selections or the naïve library. Of course, different ways for

Box 2 | Structure file

Template structure file

```
path_to_sequence_file
minimum_line_length
code1_count    code2_count                code3_count                code4_count
output_type
mismatch_limit
code1          code1_startpos             code1_endpos               path_to_code1_list
code2          code2_startpos             code2_endpos               path_to_code2_list
code3          code3_startpos             code3_endpos               path_to_code3_list
code4          code4_startpos             code4_endpos               path_to_code4_list
const1         const_region_1_startpos    const_region_1_endpos      const_region_1_seq
const2         const_region_2_startpos    const_region_2_endpos      const_region_2_seq
const3         const_region_3_startpos    const_region_3_endpos      const_region_3_seq
```

Example structure file

```
sequences/HTDS_data.fasta
50
20     5      575    213
2bb
0
x       1      6     codelists/codelist1.txt
y      87     92     codelists/codelist2.txt
z      31     36     codelists/codelist3.txt
$      58     65     codelists/codelist4.txt
1       7     30     GGAGCTTCTGAATTCTGTGTGCTG
2      37     57     CGAGTCCCATGGCGCAGCTGC
3      66     86     CACGGATCCATTCGATGCAGG
```

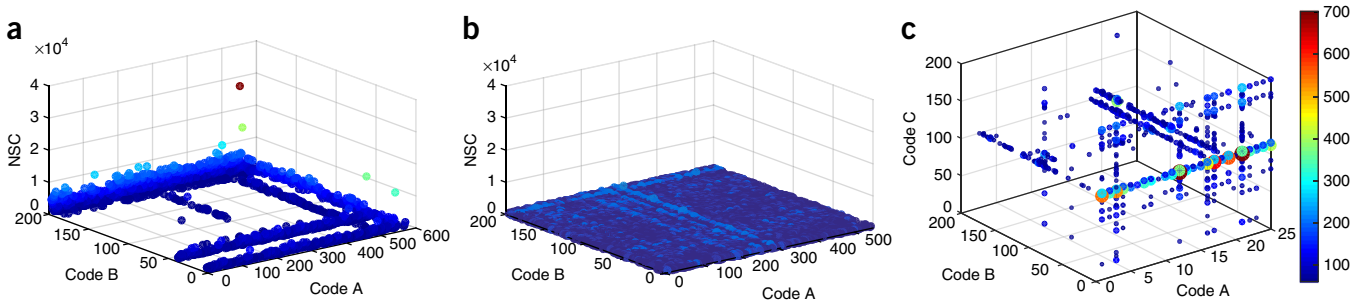


Figure 6 | Selection fingerprints. (**a**) Fingerprint of a two-building-block DECL selection against horseradish peroxidase. NSC, normalized sequence count. Sequence counts normalized to 100; cutoff level = 1,000. Binding building blocks have an elevated NSC value and are visible as lines. The cross-points of binding building blocks feature the highest enrichment, indicating that both building blocks contribute to the binding (chelate effect). (**b**) Unselected, naïve two-building-block DECL library. (normalized to 100, cutoff level = 200). Before the affinity selection, all library members are present in comparable amounts. (**c**) Graphical representation of a three-building-block DECL selection. The axes represent the three sets of building blocks, whereas dot color and size represent the sequence counts of a compound according to the heat scale given at the right.

normalization can be considered and the user may prefer to process the raw evaluation file differently for a desired normalization.

The normalized file can be imported into the MATLAB software for graphical visualization of the selection results in a Cartesian coordinate system plot (**Fig. 6**). Our MATLAB script for two-building block DECL selections is provided as **Supplementary Software 5**, and it needs to be adjusted for input file name, the selection of interest and the desired count cutoff value (highlighted in yellow). Only *z* values exceeding the cutoff are plotted. After affinity selection, certain library members may be enriched, as a result of their preferential interaction with the protein immobilized on the solid support (**Fig. 6a**). This visualization enables the display of structure-activity relationships: binding building blocks are visible as lines for libraries of two sets of building blocks, whereas cross-points indicate contributions to binding from both building blocks. The latter situation is especially remarkable for dual-pharmacophore[32] libraries, profiting from the chelate effect[67]. Apart from counting the codes in selection experiments, it is important to also sequence the unselected, naïve library, as well as selections performed with 'empty' beads, in order to define enrichment factors and to check library homogeneity. In an unselected library (**Fig. 6b**), all compounds should be present at comparable amounts[58].

After analysis of the selection results, the obtained hits need to be validated on- and off-DNA[32,76] in different, target-specific assays and optimized by medicinal chemical activities to become valuable lead structures and eventually drugs. These procedures cannot easily be generalized and are beyond the scope of this protocol.

Adaptation of the protocol to different types of DECLs

The presented protocol has been optimized for selections using dual-pharmacophore DECLs with two sets of building blocks. After affinity-based selection, the primers used for the first PCR step introduce two additional coding regions, which are used to tag the individual selection (nucleotide sequences are provided in the **Supplementary Note**). The same setup can be used for single-pharmacophore libraries containing two sets of building blocks. The encoding of selections with two additional coding domains allows the pooling of 100 or more selections to be conducted with relatively small two-building-block libraries (typical library sizes of 10^5–10^6 library members) for analysis on one Illumina flow lane, making the selection read-out a low-cost effort.

Libraries with three and four sets of building blocks can be treated analogously. As our evaluation software can analyze four coding regions, DECL libraries consisting of three sets of building blocks can be decoded and one additional code for the selection encoding can be introduced in the first PCR step. As three-building-block libraries typically have sizes in the millions range, only a handful of selections should be pooled and analyzed on one Illumina flow lane. Libraries with four sets of building blocks can theoretically be very large (in the billions range), and their selections cannot be barcoded in the first PCR step of our protocol. However, if desired, the individual selections can be tagged at the second PCR step by using barcoded Illumina TruSeq primers.

The structure file for our evaluation program defines whether a library of two, three or four sets of building blocks is evaluated (**Box 2** and **Supplementary Software 3**). For three-building-block libraries, the evaluation file can also be imported into MATLAB. We provide a script for the graphical 3D representation of a three-building-block library in **Supplementary Software 6**. Here, the three sets of building blocks form the *x*, *y* and *z* axes, and the count for each compound is given as a colored dot of different size (**Fig. 6c** shows an example plot). A heat-map scale at the right correlates the color and size of the dot with the normalized sequence counts (NSCs). A graphical representation of selections with four sets of building blocks is out of the scope of this protocol.

MATERIALS

REAGENTS

- Acetic acid (Sigma-Aldrich, cat. no. 33209) **! CAUTION** Concentrated acetic acid is corrosive. Use appropriate personal protective equipment such as nitrile gloves, protective eye goggles, chemical-resistant clothing and shoes. Handle it in a fume hood.
- Agarose (Sigma-Aldrich, cat. no. A9539)
- D-biotin (ChemPep, cat. no. 270201)
- EZ-Link NHS-LC-biotin (Thermo Fisher Scientific, cat. no. 21336)
- Boric acid (Sigma-Aldrich, cat. no. B6768)
- Deoxynucleoside triphosphate set (Roche, cat. no. 11969064001)

- DMSO (Fluka, cat. no. 41641)
- DNA-encoded chemical library (DECL). DECLs have been reported by various academic groups[26,28,32,38,77] and by different companies[18,21,40,78]. Aliquots of DECLs for affinity-based selections may be obtained from ETH Zürich or Philochem AG upon contractual agreement
- Dynabeads MyOne streptavidin C1 (Thermo Fisher Scientific, cat. no. 65001) ▲ **CRITICAL** For this protocol, only these beads have been evaluated. Products of other suppliers may lead to different selection results.
- Dynabeads MyOne streptavidin T1 (Thermo Fisher Scientific, cat. no. 65601) ▲ **CRITICAL** For this protocol, only these beads have been evaluated. Products of other suppliers may lead to different selection results.
- EDTA (Sigma-Aldrich, cat. no. E9884)
- Ethanol, absolute (Fluka, cat. no. 02860)
- GeneRuler Ultra low-range DNA ladder (Thermo Fisher Scientific, cat. no. SM1213)
- HCl (Sigma-Aldrich, cat. no. 30721) ! **CAUTION** Hydrochloric acid is highly corrosive. Use appropriate personal protective equipment such as nitrile gloves, protective eye goggles, chemical-resistant clothing and shoes. Handle it in a fume hood.
- Herring sperm DNA (Thermo Fisher Scientific, cat. no. 15634-017)
- HiSeq SBS kit V4 (Illumina, cat. no. FC-401-4002)
- HiSeq SR cluster kit v4 cBot (Illumina, cat. no. GD-401-4001)
- Isopropanol (Sigma-Aldrich, cat. no. 33539-1L-R)
- NaCl (Merck, cat. no. 1.06404.1000)
- $Na_2HPO_4 \cdot 2H_2O$ (Merck, cat. no. 1.06580.1000)
- $NaH_2PO_4 \cdot 2H_2O$ (Merck, cat. no. 1.06345.1000)
- NaOH (Fisher Chemical, cat. no. 10675692) ! **CAUTION** Sodium hydroxide is highly corrosive. Use appropriate personal protective equipment such as nitrile gloves, protective eye goggles, chemical-resistant clothing and shoes. Handle it in a fume hood.
- NucleoSpin gel and PCR clean-up (Macherey-Nagel, cat. no. 740609.250)
- PD-10 desalting columns (GE Healthcare, cat. no. 17-0851-01)
- PhiX Control v3 (Illumina, cat. no. FC-110-3001)
- Phusion high-fidelity DNA polymerase (New England BioLabs, cat. no. M0530S)
- Sodium acetate (Sigma-Aldrich, cat. no. S2889)
- Target protein(s) of choice, e.g., α-1-acid glycoprotein[32] (Athens Research & Technology, cat. no. 16-16-010700); carbonic anhydrase IX (refs. 32,79), amino acids 120–397; human serum albumin[20] (Sigma-Aldrich, cat. no. A3782); tankyrase 1 (refs. 20,21), amino acids 1106–1325
- Tris (Sigma-Aldrich, Trizma base, cat. no. T6066)
- Tween 20 (Sigma-Aldrich, cat. no. P1379)
- Water, deionized, filtered through a 0.2-µm filter (Millipore, Millipak 40, cat. no. MPGP04001)

EQUIPMENT

- Agarose gel electrophoresis equipment
- High-throughput sequencing service (Illumina, HiSeq 2500)
- KingFisher magnetic particle processor (Thermo Fisher Scientific, cat. no. 5400000)
- KingFisher plate, 200 µl (Thermo Fisher Scientific, cat. no. 97002084)
- KingFisher tip comb (Thermo Fisher Scientific, cat. no. 97002070)
- Magnetic rack, MagRack 6 (GE Healthcare, cat. no. 28-9489-64)
- Microcentrifuge (Eppendorf, Centrifuge 5424)
- NanoDrop spectrophotometer (Thermo Fisher Scientific, NanoDrop 2000)
- pH meter (Mettler-Toledo, FiveEasy Plus, FEP20)
- Reaction tubes, 0.2-ml-capped eight-tube strips (Sarstedt, cat. no. 72.991.002)
- Reaction tubes, 1.5 ml (Sarstedt, cat. no. 72.690.001)
- Reaction tubes, 2.0 ml (Greiner, cat. no. 623 201)
- Rotator (Thermo Fisher Scientific, HulaMixer sample mixer, cat. no. 15920D)
- Thermal cycler (Biometra, T-Gradient Thermoblock)
- UV imaging system (Raytest, Diana 3)
- Vacuum filtration flask, polyethersulfone (PES) membrane, 0.22-µm pore size (TPP, cat. no. 99500)

Software

- BindIt (Thermo Fisher Scientific, provided with the KingFisher instrument)
- Word (Microsoft)
- MATLAB (MathWorks)
- Compiler for C++ program

REAGENT SETUP

▲ **CRITICAL** For optimal purity, filter all PBS buffers, as well as the Tris and NaOAc buffers, using a 0.22-µm PES membrane.

HCl Prepare a 5 M HCl solution. Dilute the 37% (wt/vol) solution in a fume hood. First, add water into a vessel, and then add the 37% HCl solution to yield a 5 M solution. HCl is stable at room temperature (22 °C) for at least 12 months. ! **CAUTION** Hydrochloric acid is highly corrosive. Use appropriate personal protective equipment such as nitrile gloves, protective eye goggles, chemical-resistant clothing and shoes.

NaOH Prepare a 5 M NaOH solution. Dissolve the white pellets in a fume hood. First, add water into a vessel (place it on ice), and then add the pellets to yield a 5 M solution. Ensure constant stirring. NaOH is stable at room temperature for at least 12 months. ! **CAUTION** Sodium hydroxide is highly corrosive. Use appropriate personal protective equipment such as nitrile gloves, protective eye goggles, chemical-resistant clothing and shoes.

D-Biotin Dissolve D-biotin in DMSO to yield a 200 mM solution. D-Biotin is stable at −20 °C for at least 12 months.

PBS PBS solution is 50 mM sodium phosphate and 100 mM NaCl, pH 7.4. Dissolve sodium phosphate and sodium chloride in water under constant stirring. Adjust the pH with 5 M NaOH. The buffer should be freshly prepared.

PBST PBST is PBS, to which Tween 20 is added under constant stirring, to yield a Tween 20 concentration of 0.05% (vol/vol). The buffer should be freshly prepared.

PBST-biotin PBST-biotin is PBST, to which D-biotin (described above) is added to a final concentration of 100 µM. The buffer should be freshly prepared.

PBST-HS PBST-HS is PBST, containing 0.2 mg/ml herring sperm DNA. Prepare by adding 10mg/ml herring sperm DNA solution to PBST. The buffer should be freshly prepared..

Tris Prepare a 10-mM Tris solution, pH 8.5. Dissolve Tris in water under constant stirring. Adjust the pH with 5 M HCl. The buffer is stable at room temperature for at least 12 months.

NaOAc Prepare a 3 M sodium acetate solution, pH 4.7, dissolving equimolar amounts of sodium acetate and glacial acetic acid (use a fume hood) in water. NaOAc is stable at room temperature for at least 12 months.

! **CAUTION** Concentrated acetic acid is corrosive. Use appropriate personal protective equipment such as nitrile gloves, protective eye goggles, chemical-resistant clothing and shoes.

EQUIPMENT SETUP

KingFisher A personal computer with a Windows operating system is required for running the protocol using the KingFisher magnetic particle processor.

Illumina sequencing Illumina HiSeq 2500 sequencing was conducted by the Functional Genomics Center Zürich.

Hard-disk space The storage and processing of HTDS data requires a large amount of hard-disk space. Make sure to have at least 1 terabyte of free space available.

PROCEDURE

Preparation of the DECL working solution ● TIMING 10 min

▲ **CRITICAL** Library concentration is indicated for the total of all library members.

1| Dilute the DECL stock solution to a concentration of 100 nM using deionized water.

2| Mix the 100 nM DECL solution with PBST-HS and PBST in order to obtain the DECL working solution.

Component	Volume per reaction (µl)	Final concentration
DECL in mQ (100 nM)	5	5 nM
PBST-HS (200 µg/ml HS)	5	10 µg/ml
PBST	up to 100	

Washing the magnetic beads ● TIMING 20 min

▲ **CRITICAL** The magnetic beads stay in suspension for a short time, and then they sink to the bottom of the vessel. Before pipetting a defined amount of beads, ensure that the beads are resuspended thoroughly.

3| Pipette the desired amount of magnetic beads from the manufacturer's vial into a new 2-ml reaction tube.

4| Place the 2-ml reaction tube in the magnetic rack, discard the buffer and resuspend the magnetic beads in 1 ml of PBS. Repeat this step three times in total.

5| Place the 2-ml reaction tube in the magnetic rack, discard the buffer and resuspend the magnetic beads in the final volume of PBS.

Affinity selection

▲ **CRITICAL** In addition to the selections against target proteins, it is vital to also perform selections against uncoated magnetic beads in order to identify potential signal not originating from the target protein.

6| In this step, an automated affinity selection procedure using the KingFisher magnetic particle processor can be used by following option A. If this device is not available, a manual selection procedure can be performed, as described in option B.

(A) Automated affinity selection ● TIMING 3 h (for 24 selections)

(i) Ready the KingFisher magnetic particle processor and load the appropriate program on the connected computer.

(ii) Place a new tip comb into the KingFisher.

(iii) Take two new KingFisher 200-µl plates and label them 'plate 1' and 'plate 2', respectively.

(iv) Pipette the buffers PBST-biotin (200 µl per well), PBST (200 µl per well) and Tris (100 µl per well) into the plates, as described in **Figure 3**.

(v) Add the DECL working solution (100 µl per well), prepared in Step 2, to the appropriate wells of plate 1 (Row F, **Fig. 3**).

(vi) Dilute the biotinylated target protein (prepared using **Supplementary Method 1**) using PBS to the final concentration and add it (100 µl per well) to plate 1 (Row B, **Fig. 3**).

(vii) Resuspend the washed magnetic beads (from Step 5) thoroughly.

(viii) Distribute the washed magnetic beads (100 µl per well) to plate 1 (Row A, **Fig. 3**) and immediately start the KingFisher program.

▲ **CRITICAL STEP** The magnetic beads stay in suspension for a short time, and then they sink to the bottom of the well, where they can be resuspended by manual pipetting but not by the KingFisher device. Thus, it is crucial to thoroughly mix the beads before distribution to plate 1 and to start the KingFisher program immediately after addition of the beads.

(ix) The KingFisher transfers the beads from row A to row H (**Fig. 3**). Once the process in row H is completed, remove plate 1 from the KingFisher and insert plate 2 in the same position. Immediately continue with the program.

▲ **CRITICAL STEP** While the KingFisher waits for the user to change plates, the magnetic beads are being held on the tip comb in the air, outside of the plate. Thus, it is crucial to immediately change the plates and to continue the program, as the protein may degrade if it is held outside of the buffer for a prolonged period of time.

? TROUBLESHOOTING

(x) Upon completion of the program, the KingFisher releases the magnetic beads into the wells containing Tris buffer. Transfer the Tris buffer with the magnetic beads into capped PCR strips.

(xi) Heat the magnetic beads for 10 min at 95 °C to denature the target protein.

■ **PAUSE POINT** The oligonucleotide part of the library members is stable, and it can be stored in Tris buffer for up to 1 year at –20 °C.

(B) Manual affinity selection ● TIMING 6 h (for 24 selections)

(i) Dilute the biotinylated target protein (prepared using **Supplementary Method 1**) using PBS to the final concentration.

(ii) Prepare one 1.5-ml reaction tube per selection.

(iii) Distribute 100 µl of washed magnetic beads (from Step 5) to each 1.5-ml reaction tube.
(iv) Place the 1.5-ml reaction tube in the magnetic rack and discard the buffer.
(v) Resuspend the magnetic beads in 100 µl of protein solution.
(vi) Incubate the beads for 30 min on a rotator.
▲ **CRITICAL STEP** It is important to use a rotator that provides end-over-end rotation. If the incubation step is not done using a rotator (e.g., using a shaker), the beads may sink down to the bottom of the reaction tube, resulting in insufficient protein immobilization.
(vii) Place the 1.5-ml reaction tube in the magnetic rack, discard the buffer and resuspend the beads in 200 µl of PBST-biotin. Repeat this step two times in total.
(viii) Place the 1.5-ml reaction tube in the magnetic rack, discard the buffer and resuspend the beads in 200 µl of PBST.
(ix) Place the 1.5-ml reaction tube in the magnetic rack and discard the buffer.
(x) Resuspend the magnetic beads in 100 µl of DECL working solution.
(xi) Incubate the beads for 1 h on a rotator.
▲ **CRITICAL STEP** It is important to use a rotator that provides end-over-end rotation. If the incubation step is not done using a rotator (e.g., using a shaker), the beads may sink down to the bottom of the reaction tube, resulting in suboptimal selection results.
(xii) Place the 1.5-ml reaction tube in the magnetic rack, discard the buffer and resuspend the beads in 200 µl of PBST. Repeat this step five times in total.
(xiii) Place the 1.5-ml reaction tube in the magnetic rack and discard the buffer.
(xiv) Resuspend the magnetic beads in 100 µl of Tris buffer.
(xv) Heat the magnetic beads for 10 min at 95 °C in order to denature the target protein.
■ **PAUSE POINT** The oligonucleotide part of the library members is stable, and it can be stored in Tris buffer for up to 1 year at –20 °C.

PCR amplification of oligonucleotide tags ● TIMING 2 d (for 24 samples)

▲ **CRITICAL** In addition to performing PCR amplification reactions using your selections against target proteins and magnetic beads as template, also include PCR amplification reactions using the unselected library as template. For this purpose, take an aliquot of the DECL working solution, dilute it 1:10, and process it as additional sample.

7| Perform PCR 1 using the eluted library members as template. It is optional to remove the magnetic beads, as they do not impede the PCR.

▲ **CRITICAL STEP** Use a different primer combination for every selection performed. This way, after HTDS on the same Illumina flow lane, all selections can be evaluated independently.

Component	Volume per reaction (µl)	Final concentration
Eluted library members	5	–
Phusion high-fidelity (HF) buffer (5×)	10	1×
Phusion $MgCl_2$ solution (50 mM)	2	2 mM
dNTPs (5 mM)	2.5	250 µM
Primer IlluminaPCR1a (10 µM)	3	0.6 µM
Primer IlluminaPCR1b (10 µM)	3	0.6 µM
Phusion (2 U/µl)	0.25	0.5 U per 50 µl
Water	up to 50	–

8| Run the PCR 1 program.

Cycle number	Denature	Anneal	Extend
1	98 °C for 3 min		
2–36	98 °C for 45 s	69 °C for 45 s	72 °C for 45 s
37			72 °C for 5 min

■ **PAUSE POINT** Amplified DNA is stable. All intermediate steps can be stored at –20 °C for up to 1 year.

9| Analyze the length and purity of the PCR 1 products by agarose gel (2% (wt/vol)) electrophoresis.
▲ **CRITICAL STEP** Check whether the PCR 1 product shows a band of the expected size (example setup provided in **Supplementary Note**). There should not be any additional bands or smears visible on the gel.
? TROUBLESHOOTING

10| Purify each PCR product using the Macherey-Nagel NucleoSpin gel and PCR clean-up kit, according to the manufacturer's instructions. Elute in 20 µl of buffer NE (included in the kit).
▲ **CRITICAL STEP** Perform two washing steps with buffer NT3 (included in the kit) for highest DNA purity.

11| Measure the concentration of the purified PCR 1 products using the NanoDrop spectrophotometer.

12| Pool the PCR 1 amplification products in one 1.5-ml reaction tube to equimolar concentration (250 nM total). Dilute to 10 nM using buffer NE.

13| Perform PCR 2 using the pooled PCR 1 products (10 nM) as template.

Component	Volume per reaction (µl)	Final concentration
Pooled PCR 1 from Step 12	10	1 nM
Phusion HF buffer (5×)	20	1×
Phusion $MgCl_2$ solution (50 mM)	4	2 mM
dNTPs (5 mM)	5	250 µM
Primer IlluminaPCR2a (10 µM)	6	0.6 µM
Primer IlluminaPCR2b (10 µM)	6	0.6 µM
Phusion (2 U/µl)	0.5	1.0 U per 100 µl
Water	up to 100	–

14| Run the PCR 2 program.

Cycle number	Denature	Anneal	Extend
1	98 °C for 3 min		
2–16	98 °C for 45 s	69 °C for 45 s	72 °C for 45 s
17			72 °C for 5 min

15| Analyze the length and purity of the PCR 2 products by agarose gel (2% (wt/vol)) electrophoresis.

16| Perform agarose gel (2% (wt/vol)) purification of the PCR 2 products using the Macherey-Nagel NucleoSpin gel and PCR clean-up kit, according to the manufacturer's instructions. Elute in 40 µl of buffer NE.
▲ **CRITICAL STEP** Perform two washing steps with buffer NT3 for highest DNA purity.

17| Perform an ethanol precipitation of the gel-extracted PCR 2 product. Detailed instructions are available in **Supplementary Method 2**.

18| Measure the concentration of the purified PCR 2 product using the NanoDrop spectrophotometer.

19| Dilute the PCR 2 product with buffer NE to a final concentration of 100 nM.

Illumina high-throughput sequencing ● TIMING 6 d (for one flow cell in single-read mode)

20| Submit 50 µl of 100 nM PCR 2 product to Illumina HTDS.

21| Choose the appropriate read length. In case of the two-building-block ESAC library, the read length needs to be 100 nt or higher.

22| Add PhiX Control v3 to a final concentration of 30%.

23| Run high-throughput sequencing using a Illumina HiSeq 2500 or comparable device in a single-read sequencing assay.

24| Obtain sequencing data. Sequencing results are prefiltered according to the Illumina sequencing primer and arrive in the *.fastq format. The expected file size is up to 50 gigabytes per flow lane. One flow cell comprises eight flow lanes.

Data analysis ● TIMING 1 d (for 24 selections)

▲ **CRITICAL** The following protocol is given for Macintosh/Unix platforms. An overview of the data analysis process is given in **Figure 5**.

25| Convert the *.fastq file to a *.fasta file using an open-source program (e.g., FASTX-Toolkit: http://hannonlab.cshl.edu/fastx_toolkit/).

26| Compile the C++ program 'count.cpp' (code provided in **Supplementary Software 2**) for your platform (Windows, Mac or Unix) using an appropriate compiler (e.g., using 'g++ -o count count.cpp –lpthread'). Place the compiled program into a new folder, e.g., named 'evaluation', and create the empty subfolders 'codelists' and 'sequences'.

? TROUBLESHOOTING

27| Prepare the structure file 'structure.txt' (**Box 2** and **Supplementary Software 3**), and place it in the 'evaluation' folder.

? TROUBLESHOOTING

28| Prepare the codelists as *.txt files (an example codelist is given as **Supplementary Software 4**) and place them in the 'codelists' folder.

29| Place the raw data *.fasta file in the 'sequences' folder.

30| Open a shell/terminal and change the directory to the 'evaluation' folder. Type './count'.

31| Input the desired evaluation filename 'name'.

32| Check whether the following evaluation files have been created: name_datum_Codecounts.txt, name_datum_eval.txt and name_datum_evalNorm.txt.

? TROUBLESHOOTING

33| In MATLAB, import the normalized evaluation file using 'Home -> Import Data'. A matrix window opens. Go to 'Import', make sure to select 'Matrix' in the list and then click on 'Import Selection'. After importing, which may take some time for large files, close the matrix window.

34| The MATLAB processing script for two-building-block libraries is provided in **Supplementary Software 5**; the one for three-building-block libraries is provided in **Supplementary Software 6**. Open the appropriate script in Word and adjust it according to your needs: define the input file, the selection number to be displayed and the desired cutoff value.

35| Copy the script and paste it into the command window of MATLAB. A new window will open and display the 3D plot of the chosen selection. For two-building-block libraries, the *x* and *y* axes represent code A and code B, respectively, whereas the *z* axis represents the NSCs (average count = 100). The 3D plot for three-building-block libraries shows the codes A, B and C as *x*, *y* and *z* axes. Dot color and size represent sequence counts.

? TROUBLESHOOTING

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 2**.

TABLE 2 | Troubleshooting table.

Step	Problem	Possible reason	Solution
6A(ix)	The beads clump and are not carried efficiently	The immobilized protein alters the properties of the beads	Increase the percentage of Tween 20 in the buffers
9	The negative controls show the same band as the samples	Contamination with template	Handle the KingFisher plates with care. Change gloves if appropriate. Use filter tips for pipetting the PCR reactions
26	Compiling problems in Windows	Lack of a pthread library	Make sure that the 'POSIX threading library for Win32' (mingw32-pthreads-w32) is installed
27	The structure file or the code lists are not properly recognized	Line feeding is not according to MS-DOS convention	Make sure that MS-DOS line-feeding is used: save as plain text file *.txt from Word choosing MS-DOS, CR/LF
32	Segmentation fault error occurs	Stack size is too small	It may be necessary to first increase the stack size. In the shell/terminal, type: 'ulimit –s 16384'
35	Not enough hits can be detected	Stringency may be too high	Redo selection with quality-controlled protein at higher concentration and at less stringent conditions (fewer washing steps of shorter duration)
	Too many hits are detected	Stringency may be too low	Use more stringent selection conditions (more washing steps of longer duration)

● TIMING

Steps 1 and 2, preparation of the DECL working solution: 10 min
Steps 3–5, washing of the magnetic beads: 20 min
Step 6A, automated affinity selection (24 selections): 3 h
Step 6B, manual affinity selection (24 selections): 6 h
Steps 7–19, PCR amplification of oligonucleotide tags (24 samples): 2 d
Steps 20–24, Illumina high-throughput sequencing (per run): 6 d
Steps 25–35, data analysis (24 selections): 1 d

ANTICIPATED RESULTS

The first PCR (**Fig. 4**) uses the selected library as template (row D of plate 2 in **Fig. 3**) and adds two selection codes using the primers IlluminaPCR1a (48 nt) and IlluminaPCR1b (46 nt). With the design of the ESAC library[32], as shown in the **Supplementary Note**, the PCR 1 product has a length of 134 nt. The agarose gel electrophoresis of the PCR 1 product ideally shows a clean band and thus requires no gel purification. By using the first PCR amplification product as template, the second PCR (**Fig. 4**) introduces sequences required for HTDS through the primers IlluminaPCR2a (58 nt) and IlluminaPCR2b (63 nt), resulting in a final PCR 2 product of 213-nt length. To guarantee optimal purity for HTDS, the second PCR product is purified by agarose gel extraction, followed by ethanol precipitation.

The final products of this protocol are individual selection fingerprints, which optimally have the properties of the ones shown in **Figure 6**. As a control experiment, we process the unselected two-building block library with PCR and HTDS. The resulting fingerprint in **Figure 6b** demonstrates that all library members are present in comparable amounts. The fingerprint of a two-building block DECL selection against horseradish peroxidase in **Figure 6a** shows enriched building blocks, visible as lines, in both sub-libraries. Importantly, the cross-points feature the highest enrichment, indicating that both building blocks contribute to the binding. **Figure 6c** displays the result of a selection using a three-building-block DECL. The NSC is represented by dot color and dot size.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS This work was supported by ETH Zürich, the Swiss National Science Foundation (SNSF), the Commission for Technology and Innovation (CTI) of the Swiss Confederation, Krebsliga Schweiz/Krebsforschung Schweiz, the European Union's Research and Innovation funding program and Philochem AG. R.M.F. acknowledges a Vizepräsident für Forschung und Wirtschaftsbeziehungen (VPFW)-ETH postdoctoral fellowship endowed by ETH Zürich and Marie-Curie actions. The authors thank S. Melkko, C.E. Dumelin, L. Mannocci, M. Jaggi, M. Leimbacher, M. Steiner, H. Röst, A. Nauer, D. Bajic, M. Valk, S. Biendl and F. Samain for their help with establishing and improving the protocol. The authors thank C. Aquino, L. Poveda and L. Opitz (Functional Genomics Center Zürich) for help with high-throughput DNA sequencing. TruSeq adapter oligonucleotide sequences © 2015 Illumina. All rights reserved. Derivative works created by Illumina customers are authorized for use with Illumina instruments and products only. All other uses are strictly prohibited.

AUTHOR CONTRIBUTIONS W.D. established the selection protocol using magnetic beads and the KingFisher magnetic particle processor. W.D., M.W. and

R.M.F. applied the protocol and optimized affinity-based selections. W.D. and J.S. optimized the PCR encoding system. F.B. adapted the protocol to use with Illumina sequencing and developed the MATLAB scripts. M.A.S. and Y.Z. wrote and optimized the evaluation software. W.D., D.N. and J.S. wrote the manuscript.

COMPETING FINANCIAL INTERESTS The authors declare competing financial interests: details are available in the online version of the paper.

1. Mayr, L.M. & Bojanic, D. Novel trends in high-throughput screening. *Curr. Opin. Pharmacol.* **9**, 580–588 (2009).
2. Wigglesworth, M.J., Murray, D.C., Blackett, C.J., Kossenjans, M. & Nissink, J.W. Increasing the delivery of next generation therapeutics from high-throughput screening libraries. *Curr. Opin. Chem. Biol.* **26**, 104–110 (2015).
3. Macarron, R. *et al.* Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* **10**, 188–195 (2011).
4. Zhang, J., Yang, P.L. & Gray, N.S. Targeting cancer with small-molecule kinase inhibitors. *Nat. Rev. Cancer* **9**, 28–39 (2009).
5. Shuker, S.B., Hajduk, P.J., Meadows, R.P. & Fesik, S.W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**, 1531–1534 (1996).
6. Rees, D.C., Congreve, M., Murray, C.W. & Carr, R. Fragment-based lead discovery. *Nat. Rev. Drug Discov.* **3**, 660–672 (2004).
7. Jorgensen, W.L. The many roles of computation in drug discovery. *Science* **303**, 1813–1818 (2004).
8. Reker, D. *et al.* Revealing the macromolecular targets of complex natural products. *Nat. Chem.* **6**, 1072–1078 (2014).
9. Reker, D. & Schneider, G. Active-learning strategies in computer-assisted drug discovery. *Drug Discov. Today* **20**, 458–465 (2015).
10. Mannocci, L., Leimbacher, M., Wichert, M., Scheuermann, J. & Neri, D. 20 years of DNA-encoded chemical libraries. *Chem. Commun.* **47**, 12747–12753 (2011).
11. Kleiner, R.E., Dumelin, C.E. & Liu, D.R. Small-molecule discovery from DNA-encoded chemical libraries. *Chem. Soc. Rev.* **40**, 5707–5717 (2011).
12. Franzini, R.M., Neri, D. & Scheuermann, J. DNA-encoded chemical libraries: advancing beyond conventional small-molecule libraries. *Acc. Chem. Res.* **47**, 1247–1255 (2014).
13. Brenner, S. & Lerner, R.A. Encoded combinatorial chemistry. *Proc. Natl. Acad. Sci. USA* **89**, 5381–5383 (1992).
14. Needels, M.C. *et al.* Generation and screening of an oligonucleotide-encoded synthetic peptide library. *Proc. Natl. Acad. Sci. USA* **90**, 10700–10704 (1993).
15. 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).
16. Leimbacher, M. *et al.* Discovery of small-molecule interleukin-2 inhibitors from a DNA-encoded chemical library. *Chemistry* **18**, 7729–7737 (2012).
17. Buller, F. *et al.* Selection of carbonic anhydrase IX inhibitors from one million DNA-encoded compounds. *ACS Chem. Biol.* **6**, 336–344 (2011).
18. Clark, M.A. *et al.* Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nat. Chem. Biol.* **5**, 647–654 (2009).
19. Franzini, R.M., Nauer, A., Scheuermann, J. & Neri, D. Interrogating target-specificity by parallel screening of a DNA-encoded chemical library against closely related proteins. *Chem. Commun.* **51**, 8014–8016 (2015).
20. Franzini, R.M. *et al.* Identification of structure-activity relationships from screening a structurally compact DNA-encoded chemical library. *Angew. Chem. Int. Ed. Engl.* **54**, 3927–3931 (2015).
21. Samain, F. *et al.* Tankyrase 1 inhibitors with drug-like properties identified by screening a DNA-encoded chemical library. *J. Med. Chem.* **58**, 5143–5149 (2015).
22. Buller, F. *et al.* Design and synthesis of a novel DNA-encoded chemical library using Diels-Alder cycloadditions. *Bioorg. Med. Chem. Lett.* **18**, 5926–5931 (2008).
23. Gartner, Z.J. *et al.* DNA-templated organic synthesis and selection of a library of macrocycles. *Science* **305**, 1601–1605 (2004).
24. 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).
25. 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).
26. Maianti, J.P. *et al.* Anti-diabetic activity of insulin-degrading enzyme inhibitors mediated by multiple hormones. *Nature* **511**, 94–98 (2014).
27. 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).
28. 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).
29. Blakskjaer, P., Heitner, T. & Hansen, N.J. 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).
30. Weisinger, R.M., Wrenn, S.J. & Harbury, P.B. Highly parallel translation of DNA sequences into small molecules. *PLoS ONE* **7**, e28056 (2012).
31. Melkko, S., Scheuermann, J., Dumelin, C.E. & Neri, D. Encoded self-assembling chemical libraries. *Nat. Biotechnol.* **22**, 568–574 (2004).
32. 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).
33. Zambaldo, C., Barluenga, S. & Winssinger, N. PNA-encoded chemical libraries. *Curr. Opin. Chem. Biol.* **26**, 8–15 (2015).
34. Eberhard, H., Diezmann, F. & Seitz, O. DNA as a molecular ruler: interrogation of a tandem SH2 domain with self-assembled, bivalent DNA-peptide complexes. *Angew. Chem. Int. Ed. Engl.* **50**, 4146–4150 (2011).
35. Winssinger, N. *et al.* PNA-encoded protease substrate microarrays. *Chem. Biol.* **11**, 1351–1360 (2004).
36. Ciobanu, M. *et al.* Selection of a synthetic glycan oligomer from a library of DNA-templated fragments against DC-SIGN and inhibition of HIV gp120 binding to dendritic cells. *Chem. Commun.* **47**, 9321–9323 (2011).
37. Gorska, K., Huang, K.T., Chaloin, O. & Winssinger, N. DNA-templated homo- and heterodimerization of peptide nucleic acid encoded oligosaccharides that mimick the carbohydrate epitope of HIV. *Angew. Chem. Int. Ed. Engl.* **48**, 7695–7700 (2009).
38. Daguer, J.P. *et al.* DNA display of fragment pairs as a tool for the discovery of novel biologically active small molecules. *Chem. Sci.* **6**, 739–744 (2015).
39. 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).
40. Litovchick, A. *et al.* Encoded library synthesis using chemical ligation and the discovery of sEH inhibitors from a 334-million member library. *Sci. Rep.* **5**, 10916 (2015).
41. Franzini, R.M. *et al.* Systematic evaluation and optimization of modification reactions of oligonucleotides with amines and carboxylic acids for the synthesis of DNA-encoded chemical libraries. *Bioconjug. Chem.* **25**, 1453–1461 (2014).
42. Satz, A.L. *et al.* DNA compatible multistep synthesis and applications to DNA encoded libraries. *Bioconjug. Chem.* **26**, 1623–1632 (2015).
43. Smith, G.P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315–1317 (1985).
44. McCafferty, J., Griffiths, A.D., Winter, G. & Chiswell, D.J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**, 552–554 (1990).
45. Kang, A.S., Barbas, C.F., Janda, K.D., Benkovic, S.J. & Lerner, R.A. Linkage of recognition and replication functions by assembling combinatorial antibody Fab libraries along phage surfaces. *Proc. Natl. Acad. Sci. USA* **88**, 4363–4366 (1991).
46. Liang, P. & Pardee, A.B. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257**, 967–971 (1992).
47. Josephson, K., Ricardo, A. & Szostak, J.W. mRNA display: from basic principles to macrocycle drug discovery. *Drug Discov. Today* **19**, 388–399 (2014).
48. Hanes, J. & Pluckthun, A. *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. USA* **94**, 4937–4942 (1997).
49. Boder, E.T. & Wittrup, K.D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15**, 553–557 (1997).
50. Ellington, A.D. & Szostak, J.W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
51. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).
52. Lipinski, C.A., Lombardo, F., Dominy, B.W. & Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46**, 3–26 (2001).

53. Griffiths, A.D. *et al.* Isolation of high-affinity human antibodies directly from large synthetic repertoires. *EMBO J.* **13**, 3245–3260 (1994).
54. Franzini, R.M. *et al.* 'Cap-and-Catch' purification for enhancing the quality of libraries of DNA conjugates. *ACS Comb. Sci.* **17**, 393–398 (2015).
55. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
56. Bentley, D.R. *et al.* Accurate whole-human-genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
57. Mannocci, L. *et al.* High-throughput sequencing allows the identification of binding molecules isolated from DNA-encoded chemical libraries. *Proc. Natl. Acad. Sci. USA* **105**, 17670–17675 (2008).
58. 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).
59. 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).
60. 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).
61. 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).
62. Li, G. *et al.* Photoaffinity labeling of small-molecule-binding proteins by DNA-templated chemistry. *Angew. Chem. Int. Ed. Engl.* **52**, 9544–9549 (2013).
63. 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).
64. Melkko, S. *et al.* Isolation of a small-molecule inhibitor of the antiapoptotic protein Bcl-xL from a DNA-encoded chemical library. *ChemMedChem* **5**, 584–590 (2010).
65. Scheuermann, J. *et al.* DNA-encoded chemical libraries for the discovery of MMP-3 inhibitors. *Bioconjug. Chem.* **19**, 778–785 (2008).
66. Dumelin, C.E., Scheuermann, J., Melkko, S. & Neri, D. Selection of streptavidin binders from a DNA-encoded chemical library. *Bioconjug. Chem.* **17**, 366–370 (2006).
67. Melkko, S., Dumelin, C.E., Scheuermann, J. & Neri, D. On the magnitude of the chelate effect for the recognition of proteins by pharmacophores scaffolded by self-assembling oligonucleotides. *Chem. Biol.* **13**, 225–231 (2006).
68. Melkko, S., Zhang, Y., Dumelin, C.E., Scheuermann, J. & Neri, D. Isolation of high-affinity trypsin inhibitors from a DNA-encoded chemical library. *Angew. Chem. Int. Ed. Engl.* **46**, 4671–4674 (2007).
69. Mannocci, L. *et al.* Isolation of potent and specific trypsin inhibitors from a DNA-encoded chemical library. *Bioconjug. Chem.* **21**, 1836–1841 (2010).
70. Kohn, M. Immobilization strategies for small molecule, peptide and protein microarrays. *J. Pept. Sci.* **15**, 393–397 (2009).
71. Waugh, D.S. Making the most of affinity tags. *Trends Biotechnol.* **23**, 316–320 (2005).
72. Terpe, K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **60**, 523–533 (2003).
73. Smith, D.B. Generating fusions to glutathione *S*-transferase for protein studies. *Methods Enzymol.* **326**, 254–270 (2000).
74. Bornhorst, J.A. & Falke, J.J. Purification of proteins using polyhistidine affinity tags. *Methods Enzymol.* **326**, 245–254 (2000).
75. Satz, A.L. DNA encoded library selections and insights provided by computational simulations. *ACS Chem. Biol.* **10**, 2237–2245 (2015).
76. Lin, W., Reddavide, F.V., Uzunova, V., Gur, F.N. & Zhang, Y. Characterization of DNA-conjugated compounds using a regenerable chip. *Anal. Chem.* **87**, 864–868 (2015).
77. 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).
78. Seigal, B.A. *et al.* The discovery of macrocyclic XIAP antagonists from a DNA-programmed chemistry library, and their optimization to give lead compounds with *in vivo* antitumor activity. *J. Med. Chem.* **58**, 2855–2861 (2015).
79. Ahlskog, J.K. *et al.* Human monoclonal antibodies targeting carbonic anhydrase IX for the molecular imaging of hypoxic regions in solid tumours. *Br. J. Cancer* **101**, 645–657 (2009).

Erratum: Automated screening for small organic ligands using DNA-encoded chemical libraries

Willy Decurtins, Moreno Wichert, Raphael M Franzini, Fabian Buller, Michael A Stravs, Yixin Zhang, Dario Neri & Jörg Scheuermann

Nat. Protoc.; doi:10.1038/nprot.2016.039; corrected 1 April 2016

In the version of this article initially published online, two of the reference citations were to the wrong references. In addition, incorrect units were given for the final concentration of PBST-biotin. The errors have been corrected for all versions of this article.