

Review

Verena Kunig, Marco Potowski, Anne Gohla and Andreas Brunschweiger*

DNA-encoded libraries – an efficient small molecule discovery technology for the biomedical sciences

<https://doi.org/10.1515/hsz-2018-0119>

Received January 15, 2018; accepted April 12, 2018

Abstract: DNA-encoded compound libraries are a highly attractive technology for the discovery of small molecule protein ligands. These compound collections consist of small molecules covalently connected to individual DNA sequences carrying readable information about the compound structure. DNA-tagging allows for efficient synthesis, handling and interrogation of vast numbers of chemically synthesized, drug-like compounds. They are screened on proteins by an efficient, generic assay based on Darwinian principles of selection. To date, selection of DNA-encoded libraries allowed for the identification of numerous bioactive compounds. Some of these compounds uncovered hitherto unknown allosteric binding sites on target proteins; several compounds proved their value as chemical biology probes unraveling complex biology; and the first examples of clinical candidates that trace their ancestry to a DNA-encoded library were reported. Thus, DNA-encoded libraries proved their value for the biomedical sciences as a generic technology for the identification of bioactive drug-like molecules numerous times. However, large scale experiments showed that even the selection of billions of compounds failed to deliver bioactive compounds for the majority of proteins in an unbiased panel of target proteins. This raises the question of compound library design.

Keywords: bioactive small molecule; DNA-encoded chemistry; DNA-encoded library; drug identification; molecular evolution; screening.

Introduction

Biomedical research nowadays benefits a lot from rapid progress in ‘omics’-technologies (Rigden et al., 2016). DNA sequencing that evolved to enable sequencing of a whole genome in a single experiment at affordable cost and in short time is one example of such a technology (Lander et al., 2001; Venter et al., 2001; Köser et al., 2012). It allowed for rapid identification of mutations in diseased tissues, most notably in malignant diseases (Alioto et al., 2015). RNA sequencing gives insight in protein expression levels in diseased versus healthy tissues (Ziegenhain et al., 2017). The ready availability of these technologies for basic research generated, and continues to generate a wealth of hypotheses on disease mechanisms, pinpointing molecular targets for validation, and eventually therapeutic intervention (Brunschweiger and Hall, 2012). Important compounds to translate this knowledge into a therapy are bioactive small molecules. Such compounds are often cellularly available, they bind to the molecular target, mostly a protein, in a reversible or irreversible manner depending on the design of the compound and modulate its function. In addition to serving as starting points for drug development programs, small bioactive molecules have value for basic science as ‘chemical probes’ that can aid in understanding biological systems (Arrowsmith et al., 2015, see also <http://www.chemicalprobes.org>; Garbaccio and Parmee, 2016; Ellermann et al., 2017). The crucial first step towards development of drugs or probes is the identification of a compound that binds to a biological target. Technologies for identification of such compounds can be broadly divided into rational approaches, i.e. design of compounds (Shoichet, et al., 2016), and screening-based approaches, i.e. *de novo* ligand generation (Macarron et al., 2011; Erlanson et al., 2016). The former requires structural information on the target and/or on molecules binding to the target, while the latter are brute-force strategies resting solely on serendipity. Depending on the technology, screening allows for testing of up to a few million molecules organized as

*Corresponding author: Andreas Brunschweiger, Department of Chemistry and Chemical Biology, TU Dortmund University, Otto-Hahn-Str. 6, D-44227 Dortmund, Germany, e-mail: andreas.brunschweiger@tu-dortmund.de, <http://orcid.org/0000-0002-4401-1495>

Verena Kunig, Marco Potowski and Anne Gohla: Department of Chemistry and Chemical Biology, TU Dortmund University, Otto-Hahn-Str. 6, D-44227 Dortmund, Germany

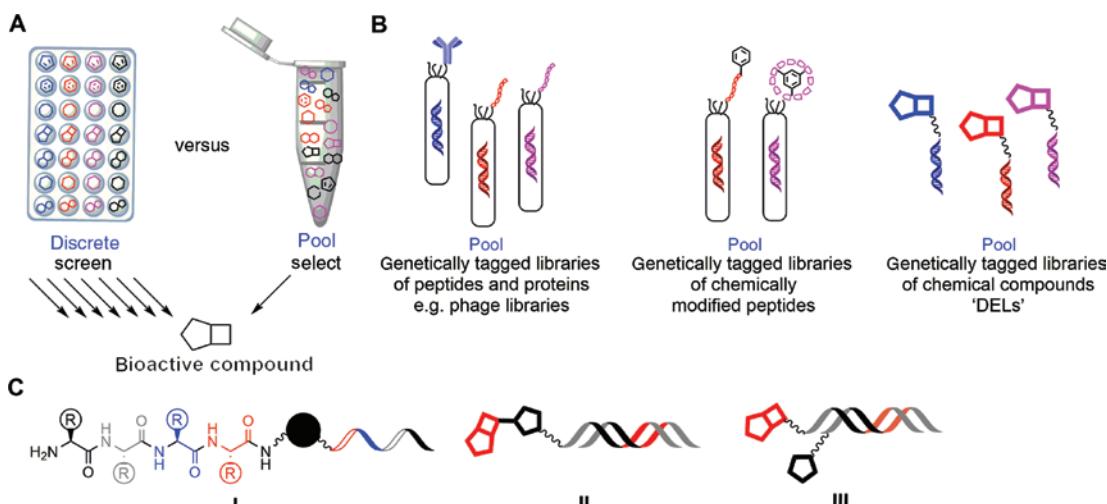


Figure 1: Identification of bioactive compounds by screening versus selection.

(A) Discrete compound libraries necessitate large-scale experimentation for testing all members for bioactive compounds ('screening') whereas pools of compounds can be tested for target ligands in single experiments ('selection'). (B) From genetically tagged libraries of biomolecules via libraries of chemically modified biomolecules to libraries of genetically tagged chemically synthesized compounds. (C) Different formats of encoded chemical libraries: I, encoded peptides on beads; II, DNA-encoded compounds; III, dual display of DNA-tagged fragments.

libraries of discrete entities (Figure 1A) to identify starting points, so-called 'hits', for drug development. However, the infrastructure required to set up and maintain compound library screening is usually only available for large research organizations.

Screening campaigns are often costly and time-consuming endeavors from the initial conception of an assay to its adaption to a format compatible with high throughput, the actual large-scale experimentation, and the evaluation of the results. They are usually only undertaken when the therapeutic value of modulating a target is already reasonably well validated. Yet, even very large discrete compound collections fail in roughly half of the screening campaigns to deliver viable compounds for further development (Macarron et al., 2011). Thus, there is an urgent need both by the drug industry and by the basic biomedical sciences to get access to alternative, more efficient technologies for the experimental identification of starting points for drug development.

Molecular evolution-based technologies represent appealing approaches to the identification of ligands for target structures (Davis et al., 2017). In contrast to the aforementioned screening technologies that test collections of discrete compounds in large-scale experiments, they enable ligand discovery from vast mixtures of molecules in a single experiment (Figure 1A), in which the ligands are selected by Darwinian principles, i.e. the compounds with the longest residence time on a given target structure are usually identified (see below and Figure 4 for a detailed description of the assay). Handling and

testing of mixtures of molecules is enabled by tagging the individual members of a compound collection with information that can be used as a read-out of the experiment. DNA is an attractive compound identifier as it displays an exceptionally large data density, and can be read out and quantitated by massive parallel sequencing. Molecular evolution-based technologies were routinely used by researchers in the life sciences for decades to identify large biological molecules that bind to a target structure, usually a protein, by DNA sequencing. These biomolecules are antibodies and peptides (Figure 1B) that can be isolated from phage- and other display libraries (Smith, 1985; Bradbury et al., 2011), and DNA or RNA oligonucleotides, called aptamers, that are identified by the SELEX technology (systematic evolution of ligands by exponential enrichment, Ellington and Szostak, 1990; Famulok and Mayer, 2014; Gotrik et al., 2016). As DNA is an amplifiable biomolecule, only minute amounts of both the genetically tagged library and the target structure are required for the assay, making it highly efficient. Endeavours, e.g. to expand the spectrum of target structures that can be addressed by these modalities, to improve their chemical and/or metabolic stability, or to modify their physico-chemical properties led to the development of elaborate strategies to chemically modify them (Figure 1B), leading to hybrid structures consisting of natural and artificial building blocks (Heinis et al., 2009; Bashiruddin and Suga, 2015; Tolle et al. 2015; Tjhung et al., 2016). At the far end of this conceptual march from genetically tagged libraries of biological structures via chemically modified

biomolecules to artificial structures are encoded libraries that are no longer synthesized by enzymes, but are prepared by – and therefore benefit from the freedom of – preparative organic chemistry (Figure 1B). Genetically tagged libraries made up of chemically synthesized compounds are known as DNA-encoded libraries or DNA-encoded chemical libraries, often abbreviated 'DELs'. This review is written to give researchers who are not familiar with the technology an introduction to encoded libraries of chemically synthesized compounds. We (a) briefly introduce the different formats of DNA-encoded libraries; (b) describe synthesis strategies and chemical methods focusing on DNA recorded libraries; (c) describe selection methods that have been developed for these libraries; (d) describe approaches to the validation of selection results; (e) assess the utility of DELs as a technology for identification of bioactive compounds.

DNA-encoded libraries

The concept of molecular chimeras consisting of a genetic tag covalently connected to, and thereby informing a chemically synthesized molecule capable of binding to a target structure, was conceived for the first time in a seminal publication by Lerner and Brenner in 1992 (Brenner and Lerner, 1992). In the same manner as oligonucleotide or phage libraries, genetically tagged collections of purely artificial molecules can be generated, handled and screened for target (protein) binders as complex mixtures obviating the need to set up and maintain costly infrastructure for compound screening. In the early days of encoded chemistry, libraries were synthesized on beads that contained two linkers, one allowed for chemical coupling of amino acid building blocks, the other for chemical coupling of DNA nucleotides (Needels et al., 1993; Nielsen et al., 1993). Coupling of amino acid building blocks and of DNA nucleotides were performed in an iterative, combinatorial manner so that each amino acid building block of the growing peptide was chemically encoded by the growing DNA strand (**I**, Figure 1C). A combinatorial synthesis strategy gave efficient access to large numbers of compounds. This encoded chemistry approach required full compatibility of compound synthesis with chemical DNA synthesis. An account published in 1995 that described the combinatorial enzymatic ligation of DNA codes was conceptually a huge step forward, as the reaction vessels in which the codes were assembled were physically separated from the vessels where compound synthesis took place. Thus, the process of compound synthesis was no longer required to be compatible with the

process of gene assembly, and the products of this process were no longer connected to a solid support but represented soluble chimeric structures consisting of a DNA part and chemically synthesized part (**II**, Figure 1C; Kinoshita and Nishigaki, 1995). The process of synthesizing DELs through cycles of alternated synthesis and encoding steps that rests on this approach is commonly known as DNA recorded chemistry (Figure 2). It is today heavily used to synthesize DELs either in solution (Mannocci et al., 2008; Clark et al., 2009) or on the solid phase (MacConnell et al., 2015). Alternative compound encoding strategies are based on the principle of DNA-programming. These are DNA-routed chemistry, DNA-templated chemistry and the 'yoctoreactor'. DNA-routed chemistry uses DNA-template strands that are directed by partially complementary DNA sequences to reaction vessels for compound synthesis (Halpin and Harbury, 2004). Likewise, DNA-templated chemistry relies on libraries of template DNA strands with defined coding regions. The DNA strands that were each coupled with the initial synthetic building block for DEL synthesis recruited small molecule conjugates of complementary (anti-codons) DNA strands, thus forcing building blocks into proximity for a chemical coupling reaction. Following compound synthesis, the anti-codons were removed from the nascent small molecule by a cleavage reaction. DNA-templated chemistry was used to synthesize libraries of peptidic macrocycles and of small molecules (Gartner et al., 2004; Cao et al., 2014). Conceptually related to the template chemistry is a technique called the 'yoctoreactor'. Unlike the DNA-templated chemistry strategy, the DNA-oligonucleotides forming the yoctoreactor were ligated following the chemical reaction to assemble the gene encoding the compound (Hansen et al., 2009). A radically different approach was followed with DNA-encoded fragment libraries **III** (Figure 1C, Melkko et al., 2004; Wichert et al., 2015). These were assembled by hybridizing a sub-library of 5'-modified DNA-fragment conjugates with a sub-library of partially complementary 3'-conjugated DNA-fragment conjugates. Selection of such libraries identifies pairs of building blocks that display cooperative binding to a target protein. Following identification of such pairs, these need to be connected by suitable linker moieties. Dynamic combinatorial chemistry enables the identification of compounds binding to target structures with high affinity by connection of smaller fragments in the binding site (Mondal and Hirsch, 2015). Tagging of fragments with short complementary DNA sequences allows for continuous shuffling of such tagged fragments until pairs of fragments are binding to a target structure. The fragment pairs are then identified by covalent linking of their DNA codes by photo crosslinking

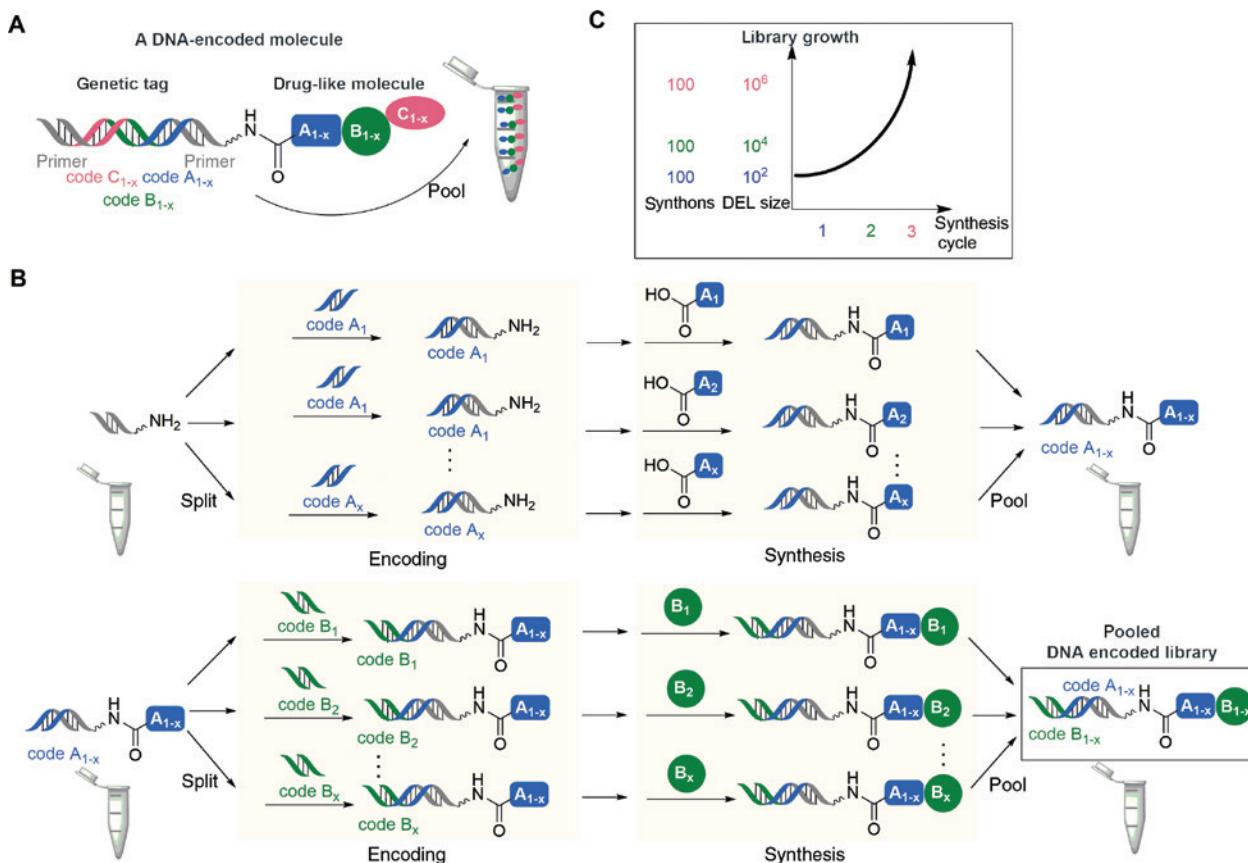


Figure 2: Encoded compound library synthesis.

(A) Structure of a DNA-tagged drug-like small molecule, collections of DNA-tagged compounds are pooled to vast mixtures; wavy bond: linker, usually a short polyethylene glycol. (B) Synthesis of DNA-encoded libraries through iterative, combinatorial cycles of alternated DNA-tagging and organic preparative synthesis steps ('split-and-pool synthesis'). (C) The combinatorial workflow leads to exponential growth of DELs.

followed by sequencing (Li et al., 2015). Finally, compound libraries can be encoded with PNA (Zambaldo et al., 2015). PNA as a coding oligomer is chemically much more stable than DNA and therefore enables the use of a much broader spectrum of chemical methods for library preparation (Chouikhi et al., 2012), but PNA-encoded libraries do not benefit from the efficiency of enzymatic encoding, and the code is also not amplifiable. Taken together, the many formats of encoded chemical libraries reflect the increasing interest in the further development of this technology and its use for the identification of small bioactive compounds (Clark, 2010; Kleiner et al., 2011; Franzini et al., 2014a; Salamon et al., 2016; Goodnow et al., 2017).

Synthesis of DNA-encoded libraries

DNA-encoded libraries are collections of small molecules covalently linked to short, chemically synthesized single- or double-stranded DNA oligonucleotides (Figure 2A).

The DNA sequence consists of terminal primer regions for polymerase chain reaction (PCR) amplification, and internal coding regions that serve as barcodes identifying the individual small molecules. The linker connecting DNA and small molecule serves as a spacer between the two parts of the compound. The most common format for DEL synthesis is the combinatorial split-and-pool approach (Figure 2B; Clark et al., 2009). In the first step, a short, linker-modified DNA is split and a first set of DNA codes identifying the first set of chemical building blocks are introduced by enzymatic (Clark et al., 2009) or chemical ligation techniques (Keefe et al., 2015; Litovchick et al., 2015). Primer extension with Klenow polymerase is an alternative encoding technique (Mannocci et al., 2008). Subsequently, the first set of building blocks are coupled to the DNA. Then, all products are pooled into a single vessel, and split for the next encoding step followed by the coupling of the second set of chemical building blocks. If necessary, the libraries can be purified after a cycle of encoding and synthesis by ion pair reverse

phase high-performance liquid chromatography (HPLC) or by simple ethanol precipitation to remove the excess of reactants and reagents. This iterative combinatorial strategy exploits the power of exponential growth to enable the preparation of huge encoded small molecule libraries depending on the number of synthesis cycles, diversity points and building blocks (Franzini and Randolph, 2016). For instance, the combinatorial connection of $100 \times 100 \times 100$ chemical building blocks through three cycles of synthesis and encoding yields a one million-membered encoded library (Figure 2C). Libraries that are synthesized through two cycles usually contain tens of thousands of compounds, whereas three-cycle libraries easily reach the millions of compounds range. Impressive examples are the generation of billion-membered compound libraries using the split-and-pool approach with four cycles of encoding and building block coupling (Clark et al., 2009; Deng et al., 2012).

Limiting factors for the synthesis of DELs are the availability of chemical building blocks, and the paucity of preparative organic synthesis methods linking these (Goodnow et al., 2017). To be suitable for DEL synthesis, a reaction must be compatible with the DNA to reduce the loss of the genetic information to a minimum (Malone and Paegel, 2016). Strongly acidic reaction conditions, oxidants, and many transition-metal ions compromise the integrity of DNA oligonucleotides, for example, by depurination. Extreme reaction conditions such as prolonged reaction at very high temperatures are not compatible with DNA as well. In addition, to be suitable for the DEL synthesis any reaction must fulfill further requirements. As DNA is insoluble in many organic solvents the reaction should tolerate water or aqueous solvent mixtures. The reactions have to provide defined products, high yields, should show a broad reactant scope, and low by-product formation to ensure a homogeneous representation of library members (Franzini and Randolph, 2016). Additional purification steps like the ‘Cap-and-Catch’ approach might be used to remove unreacted DNA-conjugates (Franzini et al., 2015a). The development of synthetic methodologies for DEL preparation requires systematic optimization and extended test reactions for the different building blocks (Franzini et al., 2014b). As reactions to be used for DEL synthesis need to meet these many requirements, only a very tiny part of the broad spectrum of known organic reactions was applied to DEL synthesis so far.

Currently, most DELs are synthesized by coupling reactions (Figure 3A–F). Amide bond formation belongs to the most important reactions in DEL synthesis. It was applied for the preparation of DELs using Fmoc protected natural and unnatural amino acids as building blocks

resulting in libraries with peptoid character, i.e. compounds with diversity elements that are arranged in a linear fashion (Figure 3A; Wrenn et al., 2007; Mannocci et al., 2008; Leimbacher et al., 2012). In order to increase the structural diversity of DELs, the synthesis of densely functionalized scaffolds as starting points for library synthesis is one important strategy for DEL synthesis (Klika Škopić et al., 2016; Estévez et al., 2017). For instance, scaffolds that are substituted with two orthogonally protected amine groups are heavily used for DEL synthesis (Figure 3B; Encinas et al., 2014; Franzini et al., 2015b). In this context, the chemistry of different protecting groups for amines (and also for carboxylic acids) was shown to be DNA-compatible (Satz et al., 2015). Nucleophilic aromatic substitution of (hetero)aromatic halides is another example for a frequently applied conjugation in DEL synthesis. The stepwise substitution of cyanuric chloride enabled the preparation of a large triaminotriazine library that provided hits in selection experiments (Figure 3C; Clark et al., 2009; Deng et al., 2012). Recently, a 34.7 million-member DEL was generated by a split-and-pool synthesis using trifunctional scaffolds containing a Fmoc-protected amine, a carboxylic acid and an aryl iodide. Palladium catalysis was used to react the aryl iodide of the trifunctional scaffold with boronic acids to furnish biaryls (Figure 3D; Deng et al., 2015).

Significantly less explored is the synthesis of (hetero)cyclic scaffold structures from simple starting materials during the DEL synthesis. One successful example is the reaction towards a benzimidazole scaffold on DNA that was applied for many DELs and yielded several bioactive compounds to date (Figure 3E; Satz et al., 2015). The Diels-Alder reaction was successfully adapted for DNA-encoded library synthesis. Libraries synthesized by this reaction allowed for the identification of a number of protein binders (Figure 3F; Buller et al., 2008, 2011). The accessibility of further heterocyclic scaffolds like imidazolidinone, quinazolinone or imidazopyridine for DELs was demonstrated in proof-of-concept experiments (Satz et al., 2015). However, in contrast to benzimidazole-based DELs (Lewis et al., 2015; Wood et al., 2015) no application in DEL synthesis or identification of bioactive compounds was reported for those reactions to date. Recently, a zirconium(IV)-catalyzed epoxide opening on DNA was established and used for the preparation of a β -amino alcohol DNA-encoded library (Fan et al., 2017). Also, a ring-closing metathesis promoted by ruthenium has been demonstrated (Lu et al. 2017). This reaction can be used to generate both small molecules and macrocycles. To synthesize more complex, spirocyclic structures, the so-called T-reaction (tertiary amino effect reaction) was

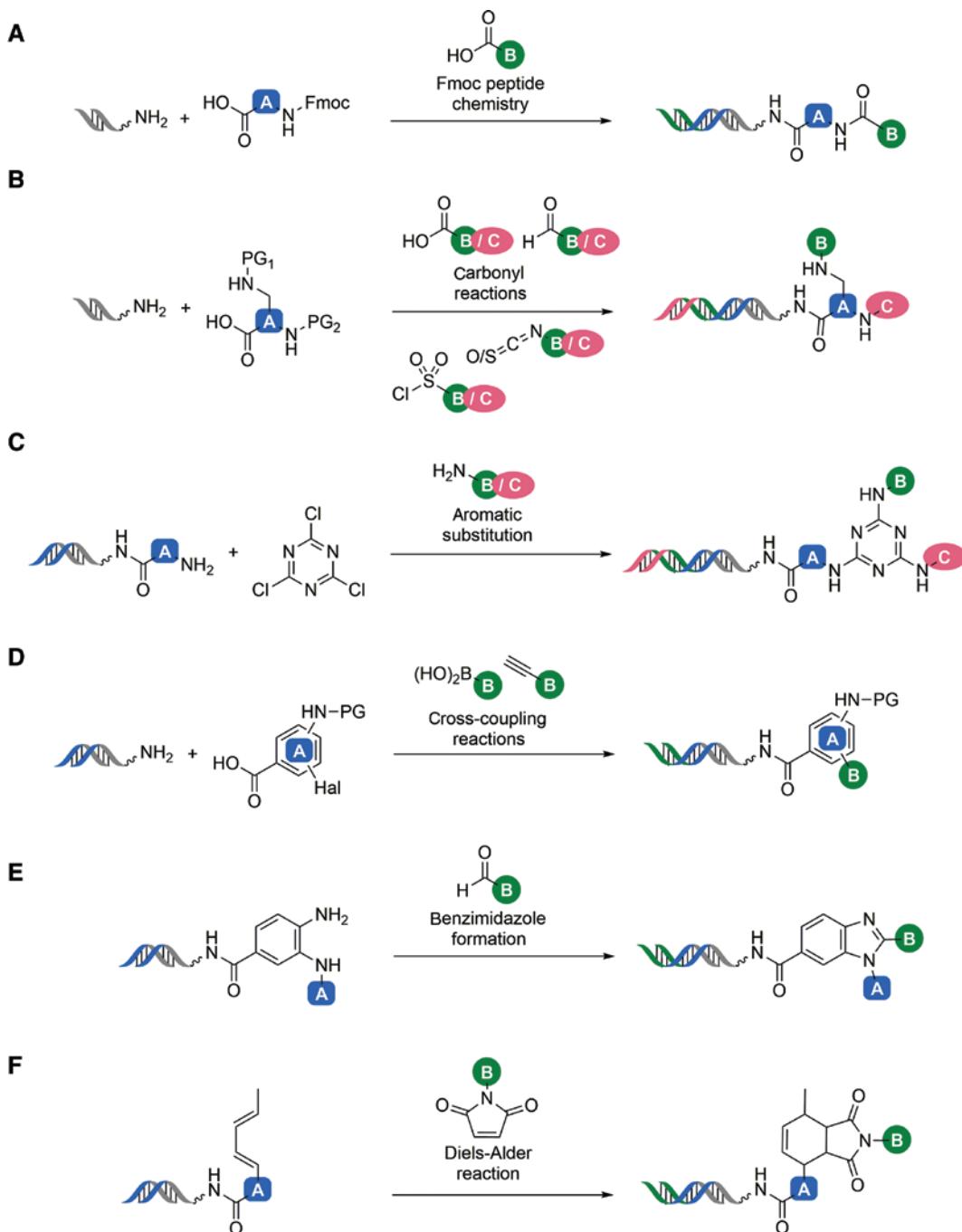


Figure 3: Representative chemical reactions applied to DNA-encoded libraries synthesis.
PG, protective group; Hal, halide; coloured shapes denote diversity elements.

adapted to DNA (Tian et al., 2016). As the T-reaction proceeds over a cascade starting with a Knoevenagel reaction followed by a [1,5]-hydride shift and a subsequent Mannich cyclization without any change in the mass the commonly used MS analysis could not be used. Therefore, the authors described an additional method to follow the reaction using isotopically-labeled substrates and nuclear

magnetic resonance (NMR) analysis. Alternative DNA-tagging strategies that allow for use of a broader scope of catalysts during library synthesis might be exploited to further broaden the chemical space covered by DELs (Klika Škopić et al., 2017). Enzymatically catalyzed reactions are in principle highly attractive to synthesize DELs, as enzymes require aqueous buffer, and work under mild

conditions. Very recently, the first chemo-enzymatic carbohydrate library synthesis on-DNA was published (Thomas et al., 2017). It was shown, that after chemical attachment of a first carbohydrate to a DNA tag, the glycan structure could be elongated with different glycosyltransferases as biocatalysts. Additionally, the sugars were oxidized by GOase (galactose oxidase) to aldehydes that could be further functionalized. From the literature it is clearly visible that the development of chemical methods for DEL synthesis is a major focus of research activities in the field, and several more chemotypes than the few reported might be accessible in an encoded format (Arico-Muendel, 2016).

Selection techniques for DNA-encoded libraries

DNA tags can be enzymatically amplified by PCR and sequenced by massive parallel sequencing to read and relatively quantify them. Thus, technical progress nowadays enables the evaluation of all DNA-tagged small molecules of a very large compound collection for interaction with a target protein in a single selection experiment.

The most common selection method using DELs relies on binding assays with non-covalent, directed immobilization of the target protein on a surface such as magnetic beads or small sepharose columns (Figure 4A). Often,

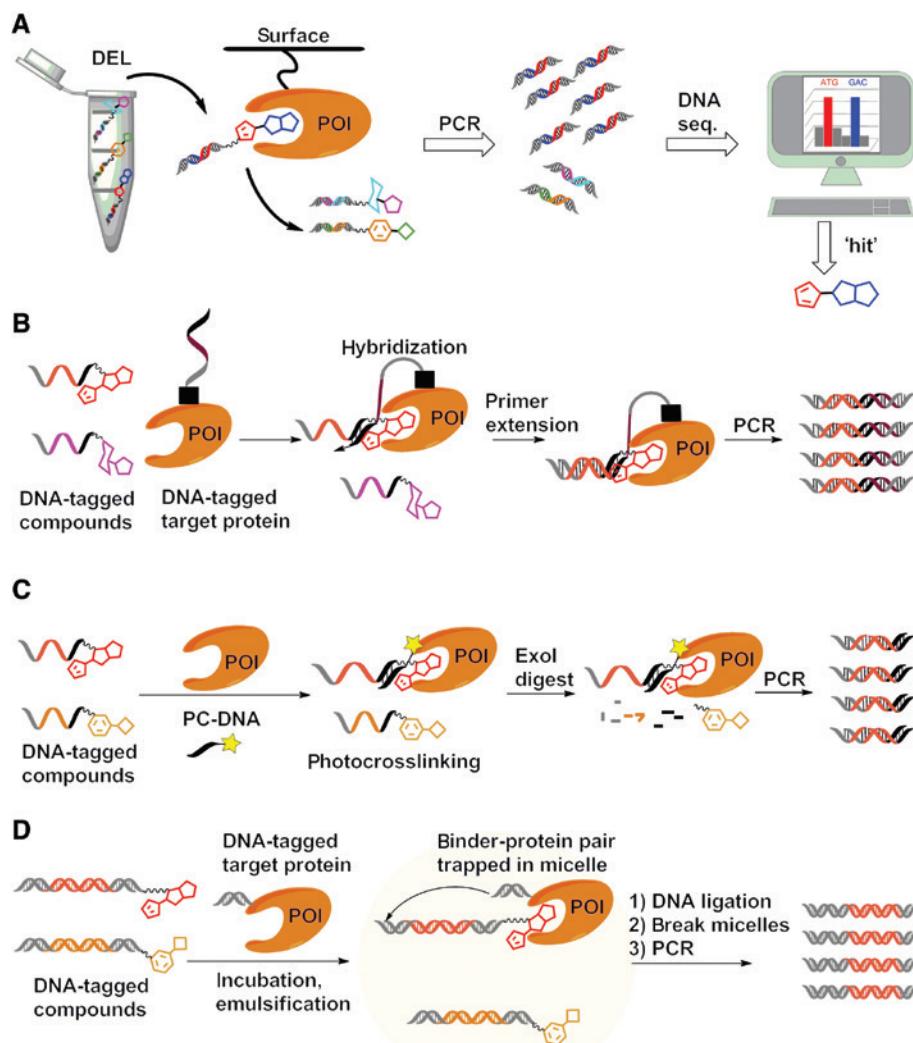


Figure 4: Selection assay formats for identification of protein binders from DELs.

(A) A protein of interest (POI) is immobilized on a surface and incubated with an aliquot of the DEL. Depletion of the non-binding compounds by washing steps is followed by elution of the binding compounds. The DNA-tags of the selected library are PCR amplified and sequenced by massive parallel sequencing. (B) Interaction-dependent PCR/interaction determination using unpurified proteins (IDPCR/IDUP). (C) DNA-programmed affinity labeling (DPAL). (D) Identification of protein binders by ‘binder trap enrichment’ (BTE).

proteins are His-tagged for this purpose and immobilized on an immobilized metal affinity chromatography (IMAC) resin. Alternative protein tags for directed protein immobilization are, e.g. biotin, FLAG and GST. The immobilized target protein is incubated with a pooled DNA-encoded library. Alternatively, protein and compounds are incubated in solution before capturing the complex on solid support (Clark et al., 2009). Washing steps with buffers that include blocking agents (sheared salmon sperm DNA, BSA) that reduce non-specific binding (Clark et al., 2009) are performed to enrich the binding fraction of a library versus a control selection experiment. Winsinger and coworkers developed a selection method using DNA-display and denaturating washing conditions ([phosphate buffered solution (PBS), 1% sodium dodecyl sulfate (SDS)] to discriminate between covalent and high affinity non-covalent ligands (Zambaldo et al., 2016). The washing process to deplete any non-binding compounds can be performed manually or automatically to increase the throughput (Decurtins et al., 2016) so that even dozens of proteins can be screened rapidly as demonstrated by Machutta and colleagues (Machutta et al., 2017). The fraction of the DNA-encoded library enriched with binders is then eluted, e.g. by heat denaturation (10 min at 95°C) of the protein. In order to increase the stringency of the selection process, usually two or three rounds of selection are performed. However, unlike SELEX or phage display, DNA-encoded libraries are not amplified in between the selection rounds, although DNA-programmed compounds could be reproduced from the amplicons. Massive parallel DNA sequencing nowadays yields sufficient data for statistical analysis of selection experiments. Compounds binding to the target protein are identified by comparison of the distribution of DNA sequences of the selection experiment with a negative control selection experiment and/or the non-selected encoded library followed by calculation of enrichment factors. The output of DEL selections is a list composed of small molecule structure/name and enrichment factors which can, e.g. be visualized in a coordinate system (Clark et al., 2009; Satz, 2015). Selection of DELs led to an impressive number of bioactive small molecules a few of which will be described below. However, there are some drawbacks associated with the selection assay. As the amount of encoded library incubated with the target protein cannot be increased beyond a certain total amount of DNA, single library members are diluted with increasing library size. The effect of compound dilution, combined with the fact that only a low percentage of the copies of the binding compounds will survive the washing steps, results in lower enrichment factors complicating data interpretation of very large,

billion-membered DELs (Satz et al., 2017). This might lead either to larger numbers of false positive compounds due to misinterpretation of sequencing data, or false negative compounds due to undersampling. Other drawbacks associated with the selection on immobilized proteins are non-specific binding of compounds to the matrix, multi-valent binding, imprecise control of target concentration on the surface, and the possibility of target denaturation during the selection process. Another disadvantage of this selection method is that the target immobilization can disturb the native conformation of the target or influence the binding of the library members. Beside this, purified target protein has to be used, and very stringent washing may remove low affinity molecules or molecules with low abundance.

To address these drawbacks, the development of selection methods that obviate protein immobilization, and/or provide higher stringency is a highly active field of research (Blakskjaer et al., 2015; Chan et al., 2015). Liu and coworkers developed a solution-phase selection method (Figure 4B) to overcome some of these disadvantages. In interaction-dependent PCR (IDPCR) the binding between complex mixtures of DNA-tagged target proteins and DNA small molecule conjugates can be evaluated. The interaction between DNA-tagged target proteins and small molecules leads to DNA hybridization and formation of a hairpin structure that after primer extension and DNA-amplification encodes the binding molecule as well as the target protein. They showed that IDPCR has the potential to detect ligand-target interactions of varying affinities in a multiplexed format (McGregor et al., 2010). IDPCR was shown to be capable of identifying the binding of small molecules to unpurified proteins in cell lysates (IDUP, interaction determination using unpurified proteins) (McGregor et al., 2010). DNA-programmed affinity labeling (DPAL, Figure 4C) is a solution-phase selection method without the need of modification or immobilization of target proteins, and it is also compatible with targets in cell lysates. In DPAL, a DNA oligonucleotide with a 5'-azidophenyl photocrosslinking moiety (PC-DNA) is hybridized through a complementary region with a library of DNA-encoded small molecules (SM-DNA). After hybridization, binding of the target protein forces the photoreactive chemical group of the PC-DNA into proximity with the target protein. Irradiation triggers the photocrosslinking of the PC-DNA to the target protein. The covalent complex can be isolated by gel electrophoresis to identify the small molecule ligand (Shi et al., 2017), or, in another approach, the small molecule DNA code is protected by the bulky protein against digest by ExoI. Surviving SM-DNAs can be then used for iterated

selection rounds or can be directly decoded by sequencing. A potential advantage of this method is that photo-reactive groups can stabilize weak interactions between the target protein and the small molecule, thus also low-affinity library members can be identified (Zhao et al., 2014). Denton and Krusemark addressed the problem of low recovery of small molecule binders in selection experiments with a photo crosslinking approach. They investigated crosslinking DNA-linked ligands to target proteins by using electrophilic or photoreactive groups in-depth to improve the recovery of small molecule binders. Covalent linking of compounds to proteins indeed allowed for application of stringent washing conditions and improved recovery of the protein binder (Denton and Krusemark, 2016). A selection method to identify binding molecules of a DNA-encoded library using water-in-oil micelles is binder trap enrichment (BTE, Figure 4D). In BTE, binding pairs of DNA-encoded small molecules and DNA-tagged target proteins are trapped in emulsion droplets. This is followed by enzymatic ligation of the target DNA tag and the barcode of the small molecule binder. PCR amplification and massive parallel sequencing enables the identification of the bound molecule (Blakskjaer et al., 2015). Another selection method was based on fluorescence-activated cell sorting (FACS). Libraries of bead-displayed DNA-encoded small molecules (one-bead-one-compound; OBOC) were screened against patient serum (sera from patients with active and infectious tuberculosis) and control serum (from patients with latent, non-infectious tuberculosis) to discover epitope surrogates and to enrich ligands of IgG. The OBOC beads were incubated with sera and washed. Serum IgG-binding hit compound beads were fluorescently labeled with Alexa Fluor 647 antihuman IgG antibody. Performing FACS the authors were able to collect beads with IgG-binding compounds. DNA-amplification and sequencing of the beads identify the compound (Mendes et al., 2016).

Validation of compounds identified by DEL selection

Selection of large libraries of small molecules against specific biological targets can result in hit compounds. These have to be confirmed and then validated through orthogonal assays to remove false positive compounds and to elect the best hits for optimizing drug-like properties. As conventional orthogonal assays, biochemical (e.g. enzymatic assays) or biophysical techniques [e.g. fluorescence polarization (FP), isothermal titration calorimetry

(ITC), differential scanning fluorimetry (DSF), surface plasmon resonance (SPR), NMR] can be used.

In the case of DNA-encoded libraries, hit compounds (on-DNA) are typically resynthesized without the DNA tag and tested as small molecules (off-DNA) to confirm their activities in conventional assay methods. Additionally, enriched hits can be resynthesized as fluorescent conjugates to enable fluorescence-based assay formats (Zimmermann and Neri, 2017). However, DNA-encoded libraries may result in hundreds of hit compounds which is why conventional validation assays are often laborious (Buller et al., 2010). Establishing synthesis procedures of enriched compounds without the DNA tags can be very challenging and is currently viewed as a bottleneck of DNA-encoded library technology (Zhang, 2014). Thus, methods that streamline the validation of large numbers of compounds are badly needed. Compounds identified from DEL selections may efficiently be resynthesized as DNA conjugates. Synthesis routines to such conjugates are already established for DEL synthesis and the DNA tag facilitates the purification of compounds by, e.g. DNA precipitation or HPLC purification (Buller et al., 2010). Scheuermann and coworkers described techniques for analyzing binding properties of compounds attached to oligonucleotide tags. For instance, nucleic acid-compound conjugates can be hybridized with fluorescently labeled complementary strands (Figure 5A, Zimmermann et al., 2017). This technique was tested by them with acetazolamide, a binder to carbonic anhydrase IX (CAIX) (Wichert et al., 2015) by using FP, AlphaScreen and microscale thermophoresis technology. Scheuermann and coworkers demonstrated that the usage of ligand-oligonucleotide-conjugates reduces synthetic effort, increases ligand solubility and generates products that are compatible both with K_d and k_{off} measurements and yields comparable values as the non DNA-ligand. Another method that addresses the limitation of the need of resynthesis of enriched compounds without the DNA tag was developed by Zhang and coworkers and is based on regenerable chips that allows a nearly completely automated high-throughput assay for the characterization of the kinetic interactions (K_d , k_{on} , and k_{off}) between DNA-conjugated compounds and target proteins (Figure 5B, Lin et al., 2015). In this method, DNA-conjugated compounds are non-covalently captured on a biosensor support by complementary DNA sequences. After affinity measurements with the target protein at different concentrations, the chip can be regenerated by washing off the DNA-attached compound and reused in the next cycles of measurement with another DNA-conjugated small molecule against the same or a different target protein. As the double helix structure is stable

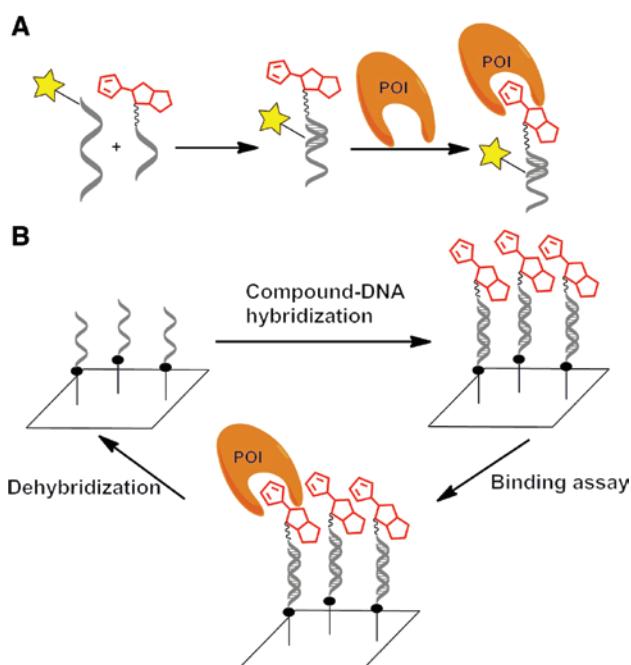


Figure 5: Validation of compounds identified from DEL selections. (A) A DNA-conjugated compound is hybridized to a complementary DNA strand, labeled with a fluorescent tag which is then tested as a complex against specific target proteins in fluorescence-based techniques. (B) Schematic representation of using regenerable chips to characterize DNA-conjugated compounds. A complementary DNA-compound conjugate is hybridized to oligonucleotides on a biosensor chip. After annealing, the target protein is injected at different concentrations for measuring the binding affinity. Afterwards the DNA-tagged compound can be washed off under dehybridization/regeneration conditions. Then another DNA-compound conjugate can be immobilized on the regenerated chip and its binding affinity is measured against the same or a different target protein.

enough through many rounds of binding, dissociation and chip-cleaning processes, this method can also be used to evaluate the binding affinity of a DNA-conjugated ligand towards different target proteins. This technique was successfully implemented in the characterization of cyclosporin A (CsA) derivatives on two different types of biosensors (interferometer and quartz crystal microbalance) against different target proteins (CypA, CypB, Cyp40, and RhTC).

Small molecule identification from DNA-encoded libraries

Over the years, the different formats of DNA-encoded libraries have proven their value as a technology for the identification of compounds binding to proteins from

different families and perturbing their function. Several enzyme inhibitors, among them compounds that show alternative, i.e. not purely competitive, modes of enzyme inhibition will be highlighted in the following lines (Figure 6). A large proportion of research activities was dedicated to kinases as these enzymes play pivotal roles in cell signaling. Often, aberrant kinase function is associated with disease, and most members of the kinase family can be inhibited with drug-like small molecules (Cohen, 2002; Hopkins and Groom, 2002; Santos et al., 2017). Usually, inhibitors of these enzymes compete with ATP. However, as the ATP binding site is well conserved throughout this large family of more than 500 members, not counting the many mutated kinases, a major challenge in the field is the development of isoenzyme-selective inhibitors. Finding compounds with alternative, e.g. allosteric binding modes or compounds that induce large shifts in the binding site can show exceptional isoenzyme selectivity. For instance, compound **1** (VPC00628) identified from an unbiased DEL by the binder trap enrichment technology, inhibits p38 α MAP kinase in the nanomolar range with high selectivity versus the kinome. The high selectivity of the compound could be explained by its induced fit binding mode which was determined by X-ray crystallography (Petersen et al., 2016). Another example of a kinase inhibitor with an alternative binding mode is the phosphoinositide 3-kinase α (PI3K α) inhibitor **2** which binds to the ATP binding site, and an additional pocket (Yang et al., 2015). A covalent c-Jun N-terminal kinase 1 (JNK1) inhibitor was identified from an encoded fragment library that displayed a set of thiol-reactive structures (Zimmermann et al., 2017). The isolated fragments were linked by a flexible polyethylene-linker yielding compound **3**. Covalent binding of **3** to its target protein was proven by mass spectrometry. The kinase inhibitor proved to be selective versus the closely related family members BTK and GAK. Selection of a very different class of encoded compounds, namely peptidic macrocycles, yielded the allosteric Src kinase inhibitor **4** (Kleiner et al., 2010; Georghiou et al., 2012). Finally, a highly selective receptor interacting protein 1 (RIP1) kinase inhibitor **5** that originated from a DNA-encoded library screen was developed towards a clinical candidate currently evaluated in phase 2a clinical trials in psoriasis, rheumatoid arthritis and ulcerative colitis patients (Harris et al., 2017). The high kinase selectivity of the compound was attributed to its binding partly to an allosteric site which is not present in other kinases.

DNA-encoded libraries were not only used to find potent and highly selective kinase inhibitors with novel modes of action (Cuozzo et al., 2017) but proved to be successful

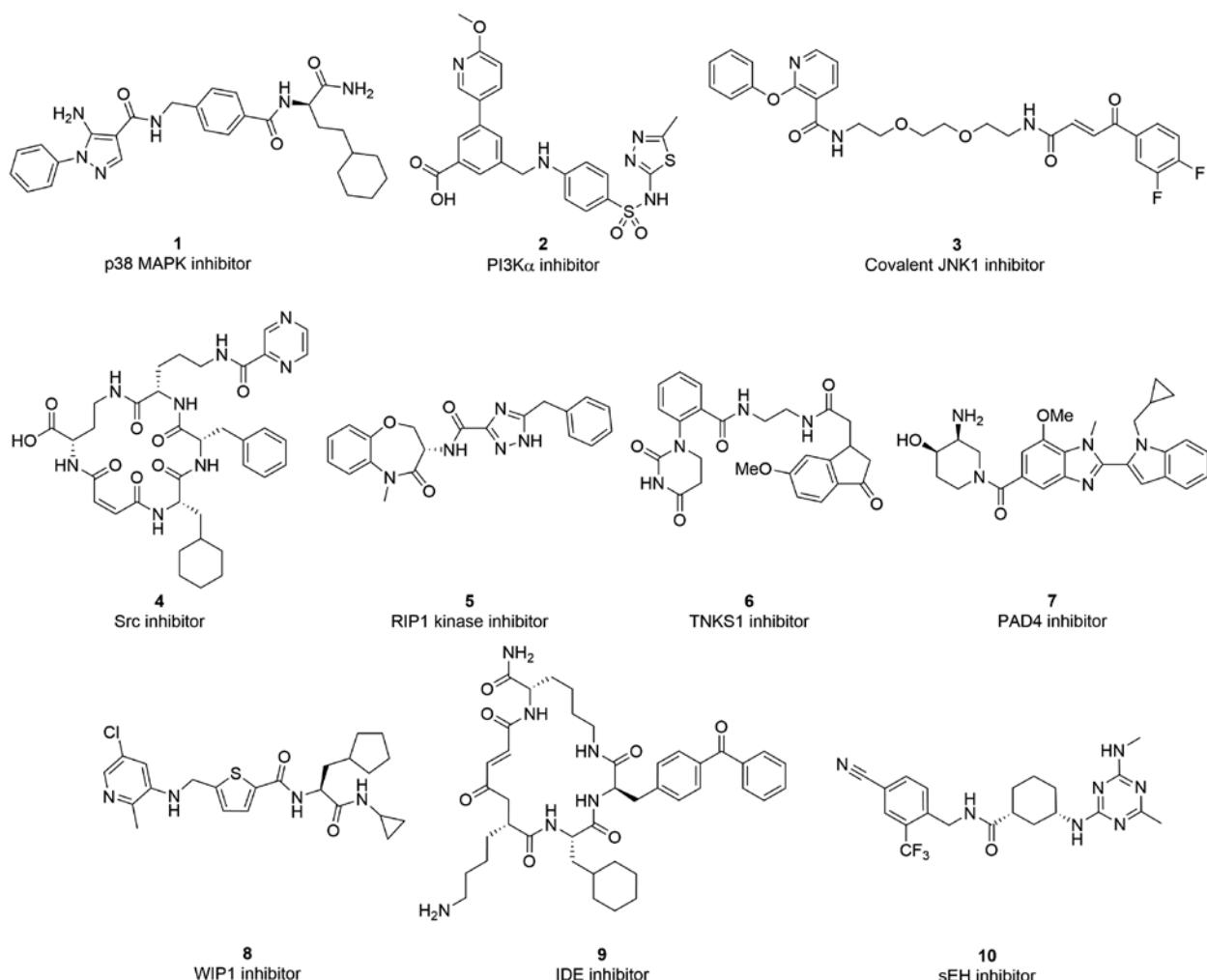


Figure 6: Enzyme inhibitors from DNA-encoded libraries.

across other enzyme families. For instance, inhibitors of the poly(ADP-ribose) polymerase tankyrase 1 (TNKS1) with nanomolar potency such as **6** were identified from a DEL (Franzini et al., 2015a). Several enzyme inhibitors that were initially discovered by selection of DNA-encoded libraries and subsequently optimized for potency, selectivity, and cellular bioavailability are today available as tools for chemical biology studies. These are the protein arginine deiminase 4 (PAD4) inhibitor **7**; the wild-type p53-induced phosphatase 1 (Wip1, PP2C δ , PPM1D) inhibitor **8**; and the insulin degrading enzyme (IDE) inhibitor **9**. All these compounds show unique modes of action for inhibition of their target. The isoenzyme-selective PAD4 inhibitor **7** inhibited the protein arginine deiminase with a mixed mode of inhibition explained by induced-fit-type mechanism (Lewis et al., 2015). It was used in cellular and animal studies to validate the role of PAD4 catalytic activity in the formation of neutrophil extracellular traps (NETs). NETs are networks

of extracellular fibers, primarily composed of DNA from neutrophils that occur in diseases caused by dysregulation of the immune system. The Wip1 inhibitor **8** targets a binding site outside the catalytic center, thereby inhibiting the phosphatase with a non-competitive mode of action. This explains why this phosphatase inhibitor is devoid of the structural features commonly seen on orthosteric phosphatase inhibitors, such as carboxylic or phosphonic acid moieties that are associated with impaired cellular availability and low isoenzyme selectivity. Compound **8** validated Wip1 as an oncogenic target in cellular and mouse xenograft models (Gilmartin et al., 2014). The IDE inhibitor **9** was identified from the same encoded library of macrocyclic structures as the Src inhibitor **4** (Maiani et al., 2014). It inhibited the metallohydrolase potently and selectively versus several related enzymes. As in the aforementioned examples, the IDE inhibitor binds to an allosteric binding site. Thus, the macrocycle does not require a metal

ion-binding moiety for enzyme inhibition which could potentially lead to off-target activity. As the macrocycle was orally active, it could be used to elucidate the role of IDE in the metabolism of insulin, glucagon and amylin in animal studies, giving the first hints towards IDE inhibition as a potential treatment option for diabetes. Finally, the soluble epoxide hydrolase (sEH) inhibitor **10** is another example of a compound originating a DEL selection that entered clinical trials (Belyanskaya et al., 2017).

Finding small molecules that bind to protein surfaces, and inhibit the binding of another protein is often a challenging endeavor (Milroy et al., 2014), yet modulation of protein-protein interactions can give insight into biological processes and may in some cases even hold promise as therapeutic options (Wells and McClelland, 2007; Arkin and Whitty, 2009; Laraia et al., 2015). DELs have delivered a number of compounds capable of inhibiting protein-protein interactions (Figure 7), among them are the B-cell lymphoma-extra large (Bcl-xL) antagonist **11** (Melkko et al., 2010), the interleukin 2 (IL-2) inhibitor **12** (Leimbacher et al., 2012), and the lymphocyte function-associated antigen-1 (LFA-1) antagonist **13** (Kollmann et al., 2014).

Understanding epigenetic mechanisms regulating gene expression is a highly active research area in biology. Several proteins modulate gene expression through reversible posttranslational modification of lysine side chains in histones, whereas others read these modifications,

acting, e.g. as cofactors for transcription factors. The ATPase family AAA-domain containing protein 2 (ATAD2, ANCCA), a protein associated with malignant diseases, recognizes acetylated lysine side chains in histones via a bromodomain motif. Selection of 65 billion DNA-encoded compounds versus this protein returned a small molecule binder which was subsequently optimized towards a potent, selective and cell permeable ATAD2-inhibitor **14** (Fernández-Montalván et al., 2017). BAY-850 acts by an intriguing mode of action not often observed by small molecules. It induces dimerization of ATAD2, thus reducing its affinity to acetylated lysine residues in histone-mimicking peptides. However, the compound inhibited cell growth only at very high concentrations, and gene expression studies indicated a weak impact of bromodomain blockade by **14** on target gene expression levels.

Many transmembrane proteins, e.g. G protein-coupled receptors (GPCRs), are promising drug targets, yet isolating them in their native conformation as needed for selection of DNA-encoded libraries is often technically very demanding. Overexpressing a target protein and performing selection experiments with protein-overexpressing cell lines is one strategy to circumvent the often observed lack of stability of isolated transmembrane proteins and led to the isolation of several potent tachykinin receptor 3 (NK3) antagonists (Wu et al., 2015), among them compound **15** (Figure 8). Another strategy to employ transmembrane receptors in selection experiments is to stabilize them by strategic mutations.

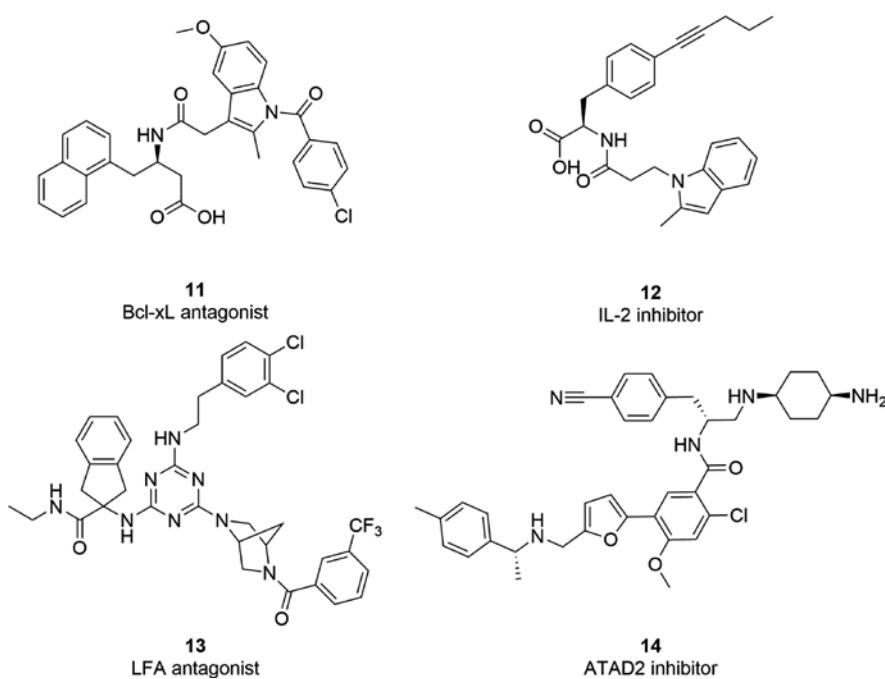


Figure 7: Inhibitors of protein-protein interactions from DNA-encoded libraries.

The protease activated receptor 2 (PAR2) was for a long time an elusive target for small molecules. Selection of encoded libraries on a stabilized PAR2 receptor identified compound **16** (Cheng et al., 2017). Crystal structures of the receptor with compound **16** revealed a previously unknown allosteric small molecule binding site for the inhibition of PAR2 activation. The utility of DELs for the identification of previously unknown binding sites for small molecules on GPCRs

was also demonstrated by compound **17**. A 190 million-membered DNA-encoded library was selected on the native β_2 adrenoceptor embedded in the detergent n-dodecyl- β -D-maltoside. Compound **17** isolated from this library bound to a distinct intracellular pocket close to the G-protein binding site of the β_2 adrenoceptor, inhibiting receptor function allosterically, therefore it was named an ‘allosteric beta-blocker’ (Ahn et al., 2017).

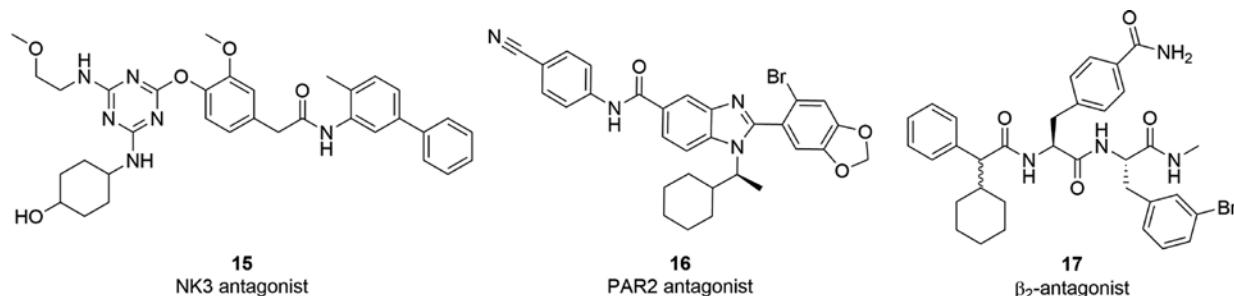


Figure 8: Antagonists of G-protein-coupled receptors from DNA-encoded libraries.

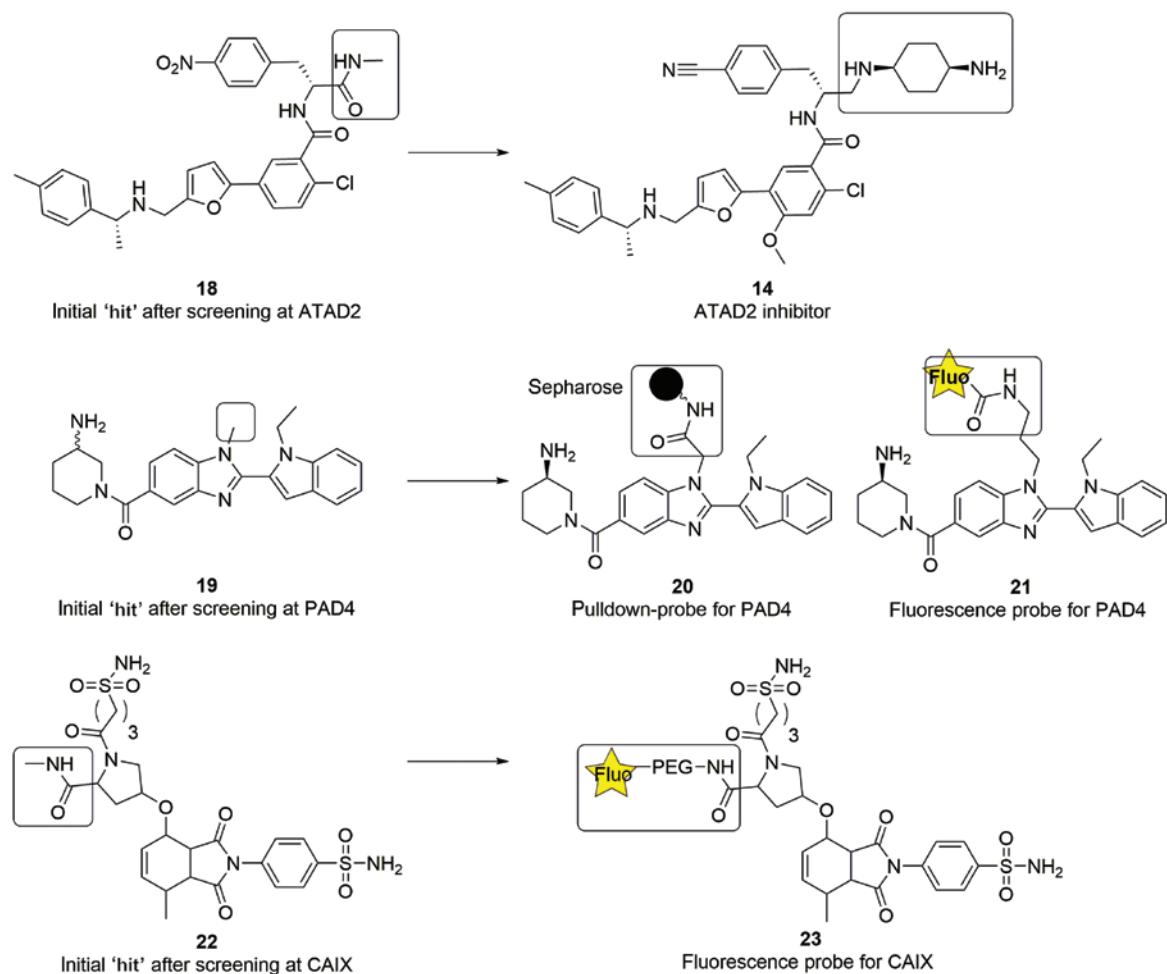


Figure 9: Examples of a compound optimized for cellular studies (**14**) and functional chemical probes (**20**, **21**, **23**) evolved from DEL hits.

Any compound isolated from a DNA-encoded library displays at its former DNA linkage site a position readily available for the chemical modification of choice (Figure 9). Fernández-Montalván derivatized this position in their primary ATAD2 binding hit **18**, and could show improvement of the compounds potency to obtain the probe **14** useful for cellular studies (Fernández-Montalván et al., 2017). The DNA can also readily be replaced with an affinity handle for chemoproteomic studies, e.g. to assess the selectivity of the compound, or with a fluorescence tag for binding studies. This was shown for the PAD4 inhibitor **19**, which was evolved into the chemical probe **7** (shown in Figure 6), and modified with a carboxylic acid **20** for covalent attachment to sepharose for proteomics studies, or with fluorescein **21** for fluorescence polarization assays. The carboanhydrase IX inhibitor **22** was converted to a fluorescence labeled probe **23** to demonstrate effective tissue targeting of carboanhydrase IX-expressing human colorectal adenocarcinoma xenografts in mice *in vivo* (Buller et al., 2011).

Table 1 summarizes compounds **1–17**, their mode of action, their chemotype and the library size, from which the compound was isolated. Library sizes vary (from 10^4 to $>10^9$ compounds) depending on the chemistry, the DEL synthesis technology, and the institution that synthesized the library. Smaller in size are academic DELs (entries **3, 4, 6, 9, 11, 12**), likely due to a more restricted

access to chemical building blocks, DNA-templated libraries (entries **1, 4, 9**) as they require the synthesis of discrete sets of DNA-conjugates as starting materials for encoded library synthesis, and libraries that encompass the synthesis of a heterocyclic core from simple starting materials (entries **5, 10, 16**). Smaller in size in the field of DNA-encoded chemistry means many thousands of compounds to a few million (!) molecules. Very large, that is multi-millions to a billion compound-numbering DELs were synthesized by mix-and-split routines using step-wise substitution of cyanuric chloride (entries **10, 13, 15**), and carbonyl chemistry such as amide bond formation and reductive alkylation of amines (entries **5, 10, 14, 17**).

The 17 bioactive compounds shown in Table 1, and identified from libraries numbering in total several billion molecules, represent very few chemotypes, here defined as structures that are synthesized by a certain route. These are peptidic macrocycles, acylated amino acid(-derived) structures or acylated diamides, *N*-substituted triaminotriazines, alkylated amines and substituted benzimidazoles, reflecting the currently still limited repertoire of synthesis methods for DELs. This situation is mirrored by two recently published large efforts to find new starting points for the development of antiinfective drugs. One selection campaign involved 84 encoded libraries numbering in total trillions of compounds. It yielded eight chemotypes (with the caveat that not all active compounds

Table 1: Mode of action of DEL-derived compounds, chemotype and size of parent encoded library.

No.	Mode of action	Chemotype	Library size
1	ATP-competitive (type-II) kinase inhibitor, induced fit	Acylated amino acid (diamide)	12.6×10^6
2	ATP-competitive kinase inhibitor, additional interaction with an 'affinity pocket'	Phenyl scaffold, substituted by C-C crosscoupling reaction and reductive amination	3.5×10^6
3	Covalent kinase ligand ^a	Linear diamide	1.5×10^5
4	ATP- and substrate-competitive kinase inhibitor	Peptoid macrocycle	1.3×10^4
5	Allosteric (type-III) kinase inhibitor	Acylated amino acid (carboxylic acid removed from progressed hit)	7.7×10^9
6	Co-substrate competitive TNKS1 inhibitor	Diamide	1.0×10^5
7	Mixed mode of citrullination inhibition, induced fit	Substituted benzimidazole	Not reported
8	Allosteric phosphatase inhibition ^b	Acylated amino acid (diamide)	Not reported
9	Allosteric protease inhibition	Peptoid macrocycle	1.3×10^4
10	Competitive epoxide hydrolase inhibition	Acylated, hetaryl substituted amino acid	Two DELs, 1×10^8 and 8×10^8
11	Competitive inhibition of protein-protein interaction ^a	Acylated amino acid	4×10^3
12	Competitive inhibition of protein-protein interaction ^b	Acylated amino acid	3×10^4
13	Competitive inhibition of protein-protein interaction ^a	Tri-substituted triazine	4.1×10^9
14	Induction and stabilization of target protein dimerization	Alkylated diamide (DEL hit)	Several DELs, in total 65×10^9
15	GPCR antagonism ^a	O-hetaryl substituted phenol carboxylic acid amide	Several DELs, 5×10^5 to 4×10^9
16	Allosteric GPCR antagonism	Substituted benzimidazole	20 DELs, hit from a 7.1×10^6 DEL
17	Allosteric GPCR antagonism	Acylated dipeptide	2×10^8

^aBinding mode not disclosed; ^bbinding mode modeled.

might have been published by the research group) that all except for a substituted pyrimidine fall into the aforementioned classes of compounds (Machutta et al., 2017). The second selection campaign on *Mycobacterium tuberculosis* InhA used 11 encoded libraries of 66 combined billion molecules, and returned ten series of compounds. Eight of these series are acylated, alkylated and/or hetaryl-substituted amino acids, diamides or diamines. These are chemotypes contained in Table 1. Two compound classes represent substituted pyridines not shown in Table 1 (Soutter et al., 2016). These two publications show that while library sizes of corporate compound collections have reached staggering compound numbers, expansion of chemotypes has not kept pace so far. Again, to be fair, one has to take into account that not all compounds might have been published, and publications that demonstrate development of synthesis methods, e.g. heterocycles have appeared rather recently (Satz et al., 2015; Fan and Davie, 2017; Lu et al., 2017).

Intuitively, one would expect that numerically very large compound libraries are required for successful identification of protein binders with novel modes of action. Intuition is supported by the allosteric kinase inhibitor **5**, the allosteric GPCR antagonists **15** and **16** as well as by compound **14** which dimerizes its target protein. Likely, compounds **7** and **8** were isolated from very large encoded compound collections, too. Yet, much smaller encoded libraries yielded protein binders with novel modes of action as well. For instance, macrocycles **4** and **9** that were selected from one of the smallest libraries shown in Table 1, revealed novel modes of action for two protein targets. Thus, compound numbers appear not to be the only parameter for successful selection outcome but also the chemotypes, i.e. the chemistry used to connect chemical building blocks, contained in an encoded library. Another viable strategy to increase the probability of protein ligand identification is the incorporation of molecular sub-structures targeting a protein or a protein class of interest. Such sub-structures can for instance be hinge binding structures as found in the kinase inhibitors **1** and **2**.

This review has described the encoded libraries in a technology-centric manner. A biologist may rather ask about the probability of finding a small molecule ligand on a given target protein using this technology. Two publications cast light on this question. The, by DEL standards, tiny library of 13.000 macrocycles that yielded the kinase inhibitor **4** (and later the IDE inhibitor **9**) was selected on 35 proteins from diverse families (Kleiner et al., 2010). In addition to the Src kinase inhibiting compound **4**, inhibitors for three more kinases were identified. Selection experiments with this small library on the other 31

proteins failed to return any compound. The second publication reported the selection of encoded libraries on a vastly different scale (Machutta et al., 2017). Many billion encoded compounds of a corporate compound collection were selected on 143 target proteins from three pathogens. Roughly two thirds of these selection experiments (93 proteins) yielded enrichment of DNA sequences indicative of successful selection of protein ligands. The research team decided to synthesize and validate small molecule ligands for 45 proteins. They were able to validate compounds for 28 proteins, and were actively pursuing validation of compounds for nine more targets at the time of publication. Thus, the selection of encoded libraries containing many billion compounds on an unbiased panel of 143 proteins yielded a success rate of at best 25%. Is this low success rate due to the coverage of chemical space by the libraries used for small molecule identification?

Summary and outlook

DNA-encoded libraries are nowadays a validated, intensely used technology for the identification of starting points for drug development. For a long time, the technology was viewed by many scientists with scepticism. The DNA tag is several times larger than the small molecule encoded by the tag, and therefore – so it was reasoned – it might prevent binding of the small molecule to a target protein in many cases. The chemical toolbox for synthesis of encoded compounds was, and still is, despite some undeniable advances very restricted. Also, uneven representation of products in complex compound mixtures due to variable synthesis yields, and damage of DNA-codes during the synthesis process were arguments brought forward against encoded libraries. Today, well-established synthesis routines allow for generation of compound collections numbering billions of compounds. These are staggering numbers, especially when comparing them to the database of CAS-registered substances which enumerated around 180 million compounds in early 2018. The feasibility of amplification and sequencing of DNA tags enables the identification of protein binders from vast mixtures of DNA-encoded compounds in a single experiment using very simple instrumentation available to any laboratory: a magnet rack, a PCR cycler, and equipment for DNA purification. DNA sequencing to read the assay is readily available from commercial providers. This contrasts favorably with the large-scale effort required for high-throughput screening of large compound libraries. Unlike other screening methods, a compound identified from a DEL offers a pre-defined position for modification with a label or a payload. It opens intriguing possibilities

for chemical biology studies, but also for tissue targeting (Krall et al., 2013), and for development of novel ubiquitin ligase-recruiting target protein degrading compounds, so-called PROTACS (Neklesa et al., 2017). Several bioactive compounds identified from DELs proved their utility as discovery technology. Some of them show unprecedented modes of action such as induction of protein dimerization (Fernández-Montalván et al., 2017) or uncover novel, allosteric binding sites on proteins (e.g. Gilmartin et al., 2014; Maianti et al., 2014; Lewis et al., 2015; Ahn et al., 2017; Cheng et al., 2017).

It is a common experience in academic and corporate drug development projects that not every small molecule screening experiment is successful (Macarron et al., 2011). Encoded libraries make compound collections that are larger by orders of magnitude than any other screening decks accessible to experimental testing for target protein binding. Yet, these gigantic compound collections also failed to deliver validated bioactive compounds for the majority of proteins in one large-scale selection campaign against an unbiased panel of 143 proteins (Machutta et al., 2017). This raises the question of compound library composition. Should the direction that places emphasis on the synthesis of a large numbers of compounds by only a few chemical reactions be followed further? This strategy necessitates amassing ever larger numbers of costly chemical building blocks. One alternative direction actively followed in DEL research is therefore the development of organic preparative synthesis methods for library synthesis (Satz et al., 2015; Fan and Davie, 2017; Lu et al., 2017). Advances in this direction will increase library diversity and leverage any number of chemical building blocks. But are such libraries more productive? Only further development of the technology and selection experiments will tell. Another important field of research is the development of selection methods that enable selecting for protein(s) (complexes) which cannot be immobilized on a surface. Currently, DELs are predominantly and very successfully used by research groups in the industry. Academic biomedical research would surely benefit from the availability of this technology, e.g. in academic screening centers. Such centers would be able to perform DEL synthesis and selection, and importantly they would also provide compound synthesis and validation for biomedical research groups.

Acknowledgments: This work was supported by the German Federal Ministry of Education and Research (BMBF), Funder Id: 10.13039/501100002347, Grant 1316053, a Boehringer Ingelheim Exploration Grant, Funder Id: 10.13039/501100008454, Grant (to M.P.), and the Mercator Research Center Ruhr Grant Pr-2016-0010 (to V.K.).

References

Ahn, S., Kahsai, A.W., Pani, B., Wang, Q.T., Zhao, S., Wall, A.L., Strachan, R.T., Staus, D.P., Wingler, L.M., Sun, L.D., et al. (2017). Allosteric “beta-blocker” isolated from a DNA-encoded small molecule library. *Proc. Natl. Acad. Sci. USA* **114**, 1708–1713.

Alioto, T.S., Buchhalter, I., Derdak, S., Hutter, B., Eldridge, M.D., Hovig, E., Heisler, L.E., Beck, T.A., Simpson, J.T., Tonon, L., et al. (2015). A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing. *Nat. Commun.* **6**, 10001.

Arico-Muendel, C. (2016). From haystack to needle: finding value with DNA encoded library technology at GSK. *Med. Chem. Commun.* **7**, 1898–1909.

Arkin, M.R. and Whitty, A. (2009). The road less traveled modulating signal transduction enzymes by inhibiting their protein-protein interactions. *Curr. Opin. Chem. Biol.* **13**, 284–290.

Arrowsmith, C.H., Audia, J.E., Austin, C., Baell, J., Bennett, J., Blagg, J., Bountra, C., Brennan, P.E., Brown, P.J., Bunnage, M.E., et al. (2015). The promise and peril of chemical probes. *Nat. Chem. Biol.* **11**, 536–541.

Bashiruddin, N.K. and Suga, H. (2015). Construction and screening of vast libraries of natural product-like macrocyclic peptides using *in vitro* display technologies. *Curr. Opin. Chem. Biol.* **24**, 131–138.

Belyanskaya, S.L., Ding, Y., Callahan, J.F., Lazaar, A.L., and Israel, D.I. (2017). Discovering drugs with DNA-encoded library technology: from concept to clinic with an inhibitor of soluble epoxide hydrolase. *ChemBioChem.* **18**, 837–842.

Blakskjaer, P., Heitner, T., and Hansen, N.J.V. (2015). Fidelity by design: yoctoreactor and binder trap enrichment for small-molecule DNA-encoded libraries and drug discovery. *Curr. Op. Chem. Biol.* **26**, 62–71.

Bradbury, A.R., Sidhu, S., Duebel, S., and McCafferty, J. (2011). Beyond natural antibodies: the power of *in vitro* display technologies. *Nat. Biotechnol.* **29**, 245–254.

Brenner, S. and Lerner, R.A. (1992). Encoded combinatorial chemistry. *Proc. Natl. Acad. Sci. USA* **89**, 5381–5383.

Brunschweiger, A. and Hall, J. (2012). A decade of the human genome sequence – how does the medicinal chemist benefit? *ChemMedChem* **7**, 194–203.

Buller, F., Mannocci, L., Zhang, Y., Dumelin, C.E., Scheuermann, J., and Neri, D. (2008). Design and synthesis of a novel DNA-encoded chemical library using Diels-Alder cycloadditions. *Bioorg. Med. Chem. Lett.* **18**, 5926–5931.

Buller, F., Mannocci, L., Scheuermann, J., and Neri, D. (2010). Drug discovery with DNA-encoded chemical libraries. *Bioconjugate Chem.* **21**, 1571–1580.

Buller, F., Steiner, M., Frey, K., Mircsof, D., Scheuermann, J., Kalisch, M., Bühlmann, P., Supuran, C.T., and Neri, D. (2011). Selection of carbonic anhydrase IX inhibitors from one million DNA-encoded compounds. *ACS Chem. Biol.* **6**, 336–344.

Cao, C., Zhao, P., Li, Z., Chen, Z., Huang, Y., Bai, Y., and Li, X. (2014). A DNA-templated synthesis of encoded small molecules by DNA self-assembly. *Chem. Commun.* **50**, 10997–10999.

Chan, A.I., McGregor, L.M., and Liu, D.R. (2015). Novel selection methods for DNA-encoded chemical libraries. *Curr. Opin. Chem. Biol.* **26**, 55–61.

Cheng, R.K.Y., Fiez-Vandal, C., Schlenker, O., Edman, K., Aggeler, B., Brown, D.G., Brown, G.A., Cooke R.M., Dumelin, C.E., Doré,

A.S., et al. (2017). Structural insight into allosteric modulation of protease-activated receptor 2. *Nature* **545**, 112–115.

Chouikhi, D., Ciobanu, M., Zambaldo, C., Duplan, V., Barluenga, S., and Winssinger N. (2012). Expanding the scope of PNA-encoded synthesis (PES): Mtt-protected PNA fully orthogonal to fmoc chemistry and a broad array of robust diversity-generating reactions. *Chem. Eur. J.* **18**, 12698–12704.

Clark, M.A. (2010). Selecting chemicals: the emerging utility of DNA-encoded libraries. *Curr. Opin. Chem. Biol.* **14**, 396–403.

Clark, M.A., Acharya, R.A., Arico-Muendel, C.C., Belyanskaya, S.L., Benjamin, D.R., Carlson, N.R., Centrella, P.A., Chiu, C.H., Creaser, S.P., Cuozzo, J.W., et al. (2009). Design, synthesis and selection of DNA encoded small-molecule libraries. *Nat. Chem. Biol.* **5**, 647–654.

Cohen, P. (2002). Protein kinases – the major drug targets of the twenty-first century? *Nat. Rev. Drug Discov.* **1**, 309–315.

Cuozzo, J.W., Centrella, P.A., Gikunju, D., Habeshian, S., Hupp, C.D., Keefe, A.D., Sigel, E.A., Soutter, H.H., Thomson, H.A., Zhang, Y., et al. (2017). 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.

Davis, A.M., Plowright, A.T., and Valeur E. (2017). Directing evolution: the next revolution in drug discovery? *Nat. Rev. Drug Disc.* **16**, 681–698.

Decurtins, W., Wichert, M., Franzini, R.M., Buller, F., Stravs, M.A., Zhang, Y., Neri, D., and Scheuermann, J. (2016). Automated screening for small organic ligands using DNA-encoded chemical libraries. *Nat. Protoc.* **11**, 764–781.

Deng, H., O'Keefe, H., Davie, C.P., Lind, K.E., Acharya, R.A., Franklin, G.J., Larkin, J., Matico, R., Neeb, M., Thompson, M.M., et al. (2012). 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.

Deng, H., Zhou, J., Sundarsingh, F.S., Summerfield, J., Somers, D., Messer, J.A., Satz, A. L., Ancellin, N., Arico-Muendel, C.C., Sargent Bedard, K.L., et al. (2015). Discovery, SAR, and X-ray binding mode study of BCATm inhibitors from a novel DNA encoded library. *ACS Med. Chem. Lett.* **6**, 919–924.

Denton, K.E. and Krusemark, C.J. (2016). Crosslinking of DNA-linked ligands to target proteins for enrichment from DNA-encoded libraries. *Med. Chem. Commun.* **7**, 2020–2027.

Eidam, O. and Satz, A.L. (2016). Analysis of the productivity of DNA encoded libraries. *Med. Chem. Commun.* **7**, 1323–1331.

Ellermann, M., Eheim, A., Rahm, F., Viklund, J., Guenther, J., Andersson, M., Ericsson, U., Forsblom, R., Ginman, T., Lindström, J., et al. (2017). Novel class of potent and cellularly active inhibitors devalidates MTH1 as broad-spectrum cancer target. *ACS Chem. Biol.* **12**, 1986–1992.

Ellington, A.D. and Szostak, J.W. (1990). *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822.

Encinas, L., O'Keefe, H., Neu, M., Remuiñán, M.J., Patel, A.M., Guardia, A., Davie, C.P., Pérez-Macías, N., Yang, H., Convery, M.A., et al. (2014). Encoded library technology as a source of hits for the discovery and lead optimization of a potent and selective class of bactericidal direct inhibitors of *Mycobacterium tuberculosis* InhA. *J. Med. Chem.* **57**, 1276–1288.

Erlanson, D.A., Fesik, S.W., Hubbard, R.E., Jahnke, W., and Jhoti, H. (2016). Twenty years on: the impact of fragments on drug discovery. *Nat. Rev. Drug Discov.* **15**, 605–619.

Estévez, A.M., Gruber, F., Satz, A.L., Martin, R.E., and Wessel, H.P. (2017). A carbohydrate-derived trifunctional scaffold for DNA-encoded libraries. *Tetrahedron Asymmetry* **28**, 837–842.

Famulok, M. and Mayer, G. (2014). Aptamers and SELEX in chemistry and biology. *Chem. Biol.* **21**, 1055–1058.

Fan, L. and Davie, C.P. (2017). Zirconium(IV)-catalyzed ring opening of on-DNA epoxides in water. *ChemBioChem.* **18**, 843–847.

Fernández-Montalván, A.E., Berger, M., Kuropka, B., Koo, S.J., Badock, V., Weiske, J., Puetter, V., Holton, S.J., Stöckigt, D., Ter Laak, A., et al. (2017). Isoform-selective ATAD2 chemical probe with novel chemical structure and unusual mode of action. *ACS Chem. Biol.* **12**, 2730–2736.

Franzini, R.M. and Randolph, C. (2016). Chemical space of DNA-encoded libraries. *J. Med. Chem.* **59**, 6629–6644.

Franzini, R.M., Neri, D., and Scheuermann, J. (2014a). DNA encoded chemical libraries: advancing beyond conventional small molecule libraries. *Acc. Chem. Res.* **47**, 1247–1255.

Franzini, R.M., Samain, F., Elrahman, M.A., Mikutis, G., Nauer, A., Zimmermann, M., Scheuermann, J., Hall, J., and Neri, D. (2014b). Systematic evolution 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.

Franzini, R.M., Biendl, S., Mikutis, G., Samain, F., Scheuermann, J., and Neri, D. (2015a). “Cap-and-Catch” purification for enhancing the quality of libraries of DNA conjugates. *ACS Comb. Sci.* **17**, 393–398.

Franzini, R.M., Ekblad, T., Zhong, N., Wichert, M., Decurtins, W., Nauer, A., Zimmermann, M., Samain, F., Scheuermann, J., Brown, P.J., et al. (2015b). Identification of structure-activity relationships from screening a structurally compact DNA-encoded chemical library. *Angew. Chem. Int. Ed.* **54**, 3927–3931.

Garbaccio, R.M. and Parmee, E.R. (2016). The impact of chemical probes in drug discovery: a pharmaceutical industry perspective. *Cell Chem. Biol.* **23**, 10–17.

Gartner, Z.J., Tse, B.N., Grubina, R., Doyon, J.B., Snyder, T.M., and Liu, D.R. (2004). DNA-templated organic synthesis and selection of a library of macrocycles. *Science* **305**, 1601–1605.

Georghiou, G., Kleiner, R.E., Pulkoski-Gross, M., Liu, D.R., and Seeliger, M.A. (2012). Highly specific, bisubstrate-competitive Src inhibitors from DNA-templated macrocycles. *Nat. Chem. Biol.* **8**, 366–374.

Gilmartin, A.G., Faitg, T.H., Richter, M., Groy, A., Seefeld, M.A., Darcy, M.G., Peng, X., Federowicz, K., Yang, J., Zhang, S.Y., et al. (2014). Allosteric Wip1 phosphatase inhibition through flap-subdomain interaction. *Nat. Chem. Biol.* **10**, 181–187.

Goodnow, R.A. Jr., Dumelin, C.E., and Keefe, A.D. (2017). DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nat. Rev. Drug Discov.* **16**, 131–147.

Gotrik, M.R., Feagin, T.A., Csordas, A.T., Nakamoto, M.A., and Soh, H.T. (2016). Advancements in aptamer discovery technologies. *Acc. Chem. Res.* **49**, 1903–1910.

Halpin, D.R. and Harbury, P.B. (2004). DNA Display I. Sequence-encoded routing of DNA populations. *PLoS Biol.* **2**, e173.

Hansen, M.H., Blakskjaer, P., Petersen, L.K., Hansen, T.H., Højfeldt, J.W., Gothelf, K.V., and Hansen, N.J. (2009). A yoctoliter-scale DNA reactor for small-molecule evolution. *J. Am. Chem. Soc.* **131**, 1322–1327.

Harris, P.A., Berger, S.B., Jeong, J.U., Nagilla, R., Bandyopadhyay, D., Campobasso, N., Capriotti, C.A., Cox, J.A., Dare, L., Dong,

X., et al. (2017). Discovery of a first-in-class receptor interacting protein 1 (RIP1) kinase specific clinical candidate (GSK2982772) for the treatment of inflammatory diseases. *J. Med. Chem.* **60**, 1247–1261.

Heinis, C., Rutherford, T., Freund, S., and Winter, G. (2009). Phage-encoded combinatorial chemical libraries based on bicyclic peptides. *Nat. Chem. Biol.* **5**, 502–507.

Hopkins, A.L. and Groom, C.R. (2002). The druggable genome. *Nat. Rev. Drug Discov.* **1**, 727–730.

Keefe, A.D., Clark, M.A., Hupp, C.D., Litovchick, A., and Zhang, Y. (2015) Chemical ligation methods for the tagging of DNA-encoded chemical libraries. *Curr. Opin. Chem. Biol.* **26**, 80–88.

Kinoshita, Y. and Nishigaki, K. (1995). Enzymatic synthesis of code regions for encoded combinatorial chemistry (ECC). *Nucleic Acids Symp.* **34**, 201–202.

Kleiner, R.E., Dumelin, C.E., Tiu, G.C., Sakurai, K., and Liu, D.R. (2010). *In vitro* selection of a DNA-templated small-molecule library reveals a class of macrocyclic kinase inhibitors. *J. Am. Chem. Soc.* **132**, 11779–11791.

Kleiner, R.E., Dumelin, C.E., and Liu, D.R. (2011). Small molecule discovery from DNA-encoded chemical libraries. *Chem. Soc. Rev.* **40**, 5707–5717.

Klika Škopić, M., Bugain, O., Jung, K., Onstein, S., Brandherm, S., Kallioloski, T., and Brunschweiger, A. (2016). Design and synthesis of DNA-encoded libraries based on a benzodiazepine and a pyrazolopyrimidine scaffold. *Med. Chem. Commun.* **7**, 1957–1965.

Klika Škopić, M., Salamon, H., Bugain, O., Jung, K., Gohla, A., Doetsch, L.J., dos Santos, D., Bhat, A., Wagner, B., and Brunschweiger, A. (2017). Acid- and Au(I)-mediated synthesis of hexathymidine-DNA-heterocycle chimeras, an efficient entry to DNA-encoded libraries inspired by drug structures. *Chem. Sci.* **8**, 3356–3361.

Kollmann, C.S., Bai, X., Tsai, C.H., Yang, H., Lind, K.E., Skinner, S.R., Zhu, Z., Israel, D.I., Cuozzo, J.W., Morgan, B.A., et al. (2014). 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. *Bioorg. Med. Chem.* **22**, 2353–2365.

Köser, C.U., Ellington, M.J., Cartwright, E.J.P., Gillespie, S.H., Brown, N.M., Farrington, M., Holden, M.T.G., Dougan, G., Bentley, S.D., Parkhill, J., et al. (2012). Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog.* **8**, e1002824.

Krall, N., Scheuermann, J., and Neri, D. (2013). Small targeted cytotoxins: current state and promises from DNA-encoded chemical libraries. *Angew. Chem. Int. Ed.* **52**, 1384–1402.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.

Laraia, L., McKenzie, G., Spring, D.R., Venkitaraman, A.R., and Huggins, D.J. (2015). Overcoming chemical, biological, and computational challenges in the development of inhibitors targeting protein-protein interactions. *Chem. Biol.* **22**, 689–703.

Leimbacher, M., Zhang, Y., Mannocci, L., Stravs, M., Geppert, T., Scheuermann, J., Schneider, G., and Neri, D. (2012). Discovery of small-molecule interleukin-2 inhibitors from a DNA-encoded chemical library. *Chem. Eur. J.* **18**, 7729–7737.

Lewis, H.D., Liddle, J., Coote, J.E., Atkinson, S.J., Barker, M.D., Bax, B.D., Bicker, K.L., Bingham, R.P., Campbell, M., Chen, Y.H., et al. (2015). Inhibition of PAD4 activity is sufficient to disrupt mouse and human NET formation. *Nat. Chem. Biol.* **11**, 189–191.

Li, G., Zheng, W., Chen, Z., Zhou, Y., Liu, Y., Yang, J., Huang, Y., and Li, X. (2015). Design, preparation, and selection of DNA-encoded dynamic libraries. *Chem. Sci.* **6**, 7097–7104.

Lin, W., Reddavide, F.V., Uzunova, V., Nadi Gür, F., and Zhang, Y. (2015). Characterization of DNA-conjugated compounds using a regenerable chip. *Anal. Chem.* **87**, 864–868.

Litovchick, A., Dumelin, C.E., Habeshian, S., Gikunju, D., Guié, M.A., Centrella, P., Zhang, Y., Sigel, E.A., Cuozzo, J.W., Keefe, A.D., et al. (2015). Encoded library synthesis using chemical ligation and the discovery of sEH inhibitors from a 334-million member library. *Sci. Rep.* **5**, 10916.

Lu, X., Fan, L., Phelps, C.B., Davie, C.P., and Donahue, C.P. (2017). Ruthenium promoted on-DNA ring-closing metathesis and cross-metathesis. *Bioconjug. Chem.* **28**, 1625–1629.

Macarron, R., Banks, M.N., Bojanic, D., Burns, D.J., Cirovic, D.A., Garyantes, T., Green, D.V., Hertzberg, R.P., Janzen, W.P., Paslay, J.W., et al. (2011). Impact of high throughput screening in biomedical research. *Nat. Rev. Drug Discov.* **10**, 188–195.

MacConnell, A.B., McEnaney, P.J., Cavett, V.J., and Paegel, B.M. (2015). DNA-encoded solid-phase synthesis: encoding language design and complex oligomer library synthesis. *ACS Comb. Sci.* **17**, 518–534.

Machutta, C.A., Kollmann, C.S., Lind, K.E., Bai, X., Chan, P.F., Huang, J., Ballell, L., Belyanskaya, S., Besra, G.S., Barros-Aguirre, D., et al. (2017). Prioritizing multiple therapeutic targets in parallel using automated DNA-encoded library screening. *Nat. Commun.* **8**, 16081.

Maianti, J.P., McFedries, A., Foda, Z.H., Kleiner, R.E., Du, X.Q., Leissring, M.A., Tang, W.J., Charron, M.J., Seeliger, M.A., Saghatelian, A., et al. (2014). Anti-diabetic activity of insulin degrading enzyme inhibitors mediated by multiple hormones. *Nature* **511**, 94–98.

Malone, M.L. and Paegel, B.M. (2016). What is a “DNA-compatible” reaction? *ACS Comb. Sci.* **18**, 182–187.

Mannocci, L., Zhang, Y., Scheuermann, J., Leimbacher, M., De Bellis, G., Rizzi, E., Dumelin, C., Melkko, S., and Neri, D. (2008). High-throughput sequencing allows the identification of binding molecules isolated from DNA-encoded chemical libraries. *Proc. Natl. Acad. Sci. USA* **105**, 17670–17675.

McGregor, L.M., Gorin, D.J., Dumelin, C.E., and Liu, D.R. (2010). 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.

Melkko, S., Scheuermann, J., Dumelin, C.E., and Neri, D. (2004). Encoded self-assembling chemical libraries. *Nat. Biotechnol.* **22**, 568–574.

Melkko, S., Mannocci, L., Dumelin, C.E., Villa, A., Sommavilla, R., Zhang, Y., Grütter, M.G., Keller, N., Jermutus, L., Jackson, R.H., et al. (2010). Isolation of a small-molecule inhibitor of the antiapoptotic protein Bcl-xL from a DNA-encoded chemical library. *ChemMedChem.* **5**, 584–590.

Mendes, K.R., Malone, M.L., Ndungu, J.M., Suponitsky-Kroyter, I., Cavett, V.J., McEnaney, P.J., MacConnell, A.B., Doran, T.M., Ronacher, K., Stanley, K., et al. (2016). High-throughput identification of DNA-encoded IgG ligands that distinguish active and latent *Mycobacterium tuberculosis* infections. *ACS Chem. Biol.* **12**, 234–243.

Milroy, L.G., Grossmann, T.N., Hennig, S., Brunsfeld, L., and Ottmann, C. (2014). Modulators of protein-protein interactions. *Chem. Rev.* **114**, 4695–4748.

Mondal, M. and Hirsch, A.K. (2015). Dynamic combinatorial chemistry: a tool to facilitate the identification of inhibitors for protein targets. *Chem. Soc. Rev.* **44**, 2455–2488.

Needels, M.C., Jones, D.G., Tate, E.H., Heinkel, G.L., Kochersperger, L.M., Dower, W.J., Barrett, R.W., and Gallop, M.A. (1993). Generation and screening of an oligonucleotide-encoded synthetic peptide library. *Proc. Natl. Acad. Sci. USA* **90**, 10700–10704.

Neklesa, T.K., Winkler, J.D., and Crews, C.M. (2017). Targeted protein degradation by PROTACs. *Pharmacol. Ther.* **174**, 138–144.

Nero, T.L., Morton, C.J., Holien, J.K., Wielens, J., and Parker, M.W. (2014). Oncogenic protein interfaces: small molecules, big challenges. *Nat. Rev. Cancer* **14**, 248–262.

Nielsen, J., Brenner, S., and Janda, K.D. (1993). Synthetic methods for the implementation of encoded combinatorial chemistry. *J. Am. Chem. Soc.* **115**, 9812–9813.

Petersen, L.K., Blakskjær, P., Chaikuad, A., Christensen, A.B., Dietvorst, J., Holmkvist, J., Knapp, S., Kořínek, M., Larsen, L.K., Pedersen, A.E., et al. (2016). Novel p38 α MAP kinase inhibitors identified from yoctoReactor DNA-encoded small molecule library. *Med. Chem. Commun.* **7**, 1332–1339.

Rigden, D.J., Fernández-Suárez, X.M., and Galperin, M.Y. (2016). The 2016 database issue of Nucleic Acids Research and an updated molecular biology database collection. *Nucleic Acids Res.* **44**, D1–D6.

Salamon, H., Klika Škopić, M., Jung, K., Bugain, O., and Brunschweiger, A. (2016). Chemical biology probes from advanced DNA-encoded libraries. *ACS Chem. Biol.* **11**, 296–307.

Santos, R., Ursu, O., Gaulton, A., Bento, A.P., Donadi, R.S., Bologa, C.G., Karlsson, A., Al-Lazikani, B., Hersey, A., Oprea, T.I., et al. (2017). A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16**, 19–34.

Satz, A.L. (2015). DNA encoded library selections and insights provided by computational simulations. *ACS Chem. Biol.* **10**, 2237–2245.

Satz, A.L., Cai, J., Chen, Y., Goodnow, R., Gruber, F., Kowalczyk, A., Petersen, A., Naderi-Oboodi, G., Orzechowski, L., and Streb, Q. (2015). DNA compatible multistep synthesis and applications to DNA encoded libraries. *Bioconjug. Chem.* **26**, 1623–1632.

Satz, A.L., Hochstrasser, R., and Petersen, A.C. (2017). Analysis of current DNA encoded library screening data indicates higher false negative rates for numerically larger libraries. *ACS Comb. Sci.* **19**, 234–238.

Shi, B., Deng, Y., Zhao, P., and Li, X. (2017). Selecting a DNA-encoded chemical library against non-immobilized proteins using a “ligate-cross-link-purify” strategy. *Bioconjug. Chem.* **28**, 2293–2301.

Shoichet, B.K., Walters, W.P., Jiang, H., and Bajorath, J. (2016). Advances in computational medicinal chemistry: a reflection on the evolution of the field and perspective going forward. *J. Med. Chem.* **59**, 4033–4034.

Smith, G.P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315–1317.

Soutter, H.H., Centrella, P., Clark, M.A., Cuozzo, J.W., Dumelin, C.E., Guie, M.A., Habeshian, S., Keefe, A.D., Kennedy, K.M., Sigel, E.A., et al. (2016). Discovery of cofactor-specific, bactericidal *Mycobacterium tuberculosis* InhA inhibitors using DNA-encoded library technology. *Proc. Natl. Acad. Sci. USA* **113**, E7880–E7889.

Thomas, B., Lu, X., Birmingham, W.R., Huang, K., Both, P., Martinez, J.E.R., Young, R.J., Davie, C.P., and Flitsch, S.L. (2017). Application of biocatalysis to on-DNA carbohydrate library synthesis. *ChemBioChem.* **18**, 858–863.

Tian, X., Basarab, G.S., Selmi, N., Kogej, T., Zhang, Y., Clark, M., and Goodnow, Jr. R.A. (2016). Development and design of tertiary amino effect reaction for DNA-encoded library synthesis. *Med. Chem. Commun.* **7**, 1316–1322.

Tjhung, K.F., Kitov, P.I., Ng, S., Kitova, E.N., Deng, L., Klassen, J.S., and Derda, R. (2016). Silent encoding of chemical post-translational modifications in phage-displayed libraries. *J. Am. Chem. Soc.* **138**, 32–35.

Tolle, F., Brändle, G.M., Matzner, D., and Mayer, G. (2015). A versatile approach towards nucleobase-modified aptamers. *Angew. Chem. Int. Ed.* **54**, 10971–10974.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001). The sequence of the human genome. *Science* **291**, 1304–1351.

Wells, J.A. and McClendon, C.L. (2007). Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **450**, 1001–1009.

Wichert, M., Krall, N., Decurtins, W., Franzini, R.M., Pretto, F., Schneider, P., Neri, D., and Scheuermann, J. (2015). Dual-display of small molecules enables the discovery of ligand pairs and facilitates affinity maturation. *Nat. Chem.* **7**, 241–249.

Wood, E.R., Bledsoe, R., Chai, J., Daka, P., Deng, H., Ding, Y., Harris-Gurley, S., Krynn, L.H., Eldridge Nartey, E., Nichols, J., et al. (2015). The role of phosphodiesterase 12 (PDE12) as a negative regulator of the innate immune response and the discovery of antiviral inhibitors. *J. Biol. Chem.* **290**, 19681–19696.

Wrenn, S.J., Weisinger, R.M., Halpin, D.R., and Harbury, P.B. (2007). Synthetic ligands discovered by *in vitro* selection. *J. Am. Chem. Soc.* **129**, 13137–13143.

Wu, Z., Graybill, T.L., Zeng, X., Platckel, M., Zhang, J., Bodmer, V.Q., Wisnoski, D.D., Deng, J., Coppo, F.T., Yao, G., et al. (2015). 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.

Yang, H., Medeiros, P.F., Raha, K., Elkins, P., Lind, K.E., Lehr, R., Adams, N.D., Burgess, J.L., Schmidt S.J., Knight, S.D., et al. (2015). Discovery of a potent class of PI3K α inhibitors with unique binding mode via encoded library technology (ELT). *ACS Med. Chem. Lett.* **6**, 531–536.

Zambaldo, C., Barluenga, S., and Winssinger, N. (2015). PNA-encoded chemical libraries. *Curr. Opin. Chem. Biol.* **26**, 8–15.

Zambaldo, C., Daguer, J.-P., Saarbach, J., Barluenga, S., and Winssinger, N. (2016). Screening for covalent inhibitors using DNA-display of small molecule libraries functionalized with cysteine reactive moieties. *Med. Chem. Commun.* **7**, 1340–1351.

Zhang, Y. (2014). Hit identification and hit follow-up. In: *A Handbook for DNA-encoded Chemistry: Theory and*

Applications for Exploring Chemical Space and Drug Discovery, R.A. Goodnow, ed. (Hoboken, New Jersey, USA: Wiley & Sons), pp. 357–376.

Zhao, P., Chen, Z., Li, Y., Sun, D., Gao, Y., Huang, Y., and Li, X. (2014). Selection of DNA-encoded small molecule libraries against unmodified and non-immobilized protein targets. *Angew. Chem. Int. Ed.* 53, 10056–10059.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., and Enard, W. (2017). Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* 65, 631–643.

Zimmermann, G. and Neri, D. (2017). DNA-encoded chemical libraries: foundations and applications in lead discovery. *Drug Discov. Today* 21, 1828–1834.

Zimmermann, G., Li, Y., Rieder, U., Mattarella, M., Neri, D., and Scheuermann, J. (2017). Hit-validation methodologies for ligands isolated from DNA-encoded chemical libraries. *ChemBioChem.* 18, 853–857.