ANNUAL REVIEWS

Annual Review of Biochemistry

DNA-Encoded Chemical Libraries: A Selection System Based on Endowing Organic Compounds with Amplifiable Information

Dario Neri[1] and Richard A. Lerner[2]

[1]Department of Chemistry and Applied Biosciences, Swiss Federal Institute of Technology (ETH Zürich), 8093 Zürich, Switzerland; email: neri@pharma.ethz.ch

[2]Department of Chemistry, The Scripps Research Institute, La Jolla, California 92037, USA; email: rlerner@scripps.edu

Annu. Rev. Biochem. 2018. 87:479–502

First published as a Review in Advance on January 12, 2018

The *Annual Review of Biochemistry* is online at biochem.annualreviews.org

https://doi.org/10.1146/annurev-biochem-062917-012550

Keywords

DNA-encoded chemical libraries, combinatorial chemistry, drug discovery

Abstract

The discovery of organic ligands that bind specifically to proteins is a central problem in chemistry, biology, and the biomedical sciences. The encoding of individual organic molecules with distinctive DNA tags, serving as amplifiable identification bar codes, allows the construction and screening of combinatorial libraries of unprecedented size, thus facilitating the discovery of ligands to many different protein targets. Fundamentally, one links powers of genetics and chemical synthesis. After the initial description of DNA-encoded chemical libraries in 1992, several experimental embodiments of the technology have been reduced to practice. This review provides a historical account of important milestones in the development of DNA-encoded chemical libraries, a survey of relevant ongoing research activities, and a glimpse into the future.

Contents

LIGAND DISCOVERY: A CENTRAL PROBLEM IN CHEMISTRY, BIOLOGY, AND THE BIOMEDICAL SCIENCES

The discovery of specific ligands that bind to protein targets of interest represents an activity of fundamental importance both for basic research and for industrial applications. Many drug discovery programs start with the search for a molecule that interacts with a validated protein target (1–3). Similarly, the deciphering of complex biochemical processes in basic research often relies on the availability of specific reagents, which bind (or even block) macromolecular structures of interest, thus allowing their visualization, quantification, or functional investigation (4–6).

The advent of monoclonal antibodies by hybridoma technology has revolutionized many areas of scientific research. This innovation (for which César Milstein and Georges Köhler were awarded the Nobel Prize in Physiology or Medicine in 1984) has made it possible to detect individual macromolecular structures with an unprecedented level of precision using affinity reagents of exquisite specificity (7). The formidable advances of the last few decades in the molecular characterization of biochemical and cellular processes would have been unthinkable without the use of monoclonal antibody reagents. It would be highly desirable if, in addition to monoclonal antibodies, small organic ligands to protein targets of interest were readily available.

The value of small organic ligands capable of high-affinity binding to a cognate protein is illustrated by the versatility of the biotin–(strept)avidin system (8). Derivatives of biotin are recognized by avidin or streptavidin with dissociation constants in the sub-picomolar range. These tight-binding complexes have found numerous applications not only for the development of specific reagents in biochemistry and immunology but also in areas as diverse as chemical synthesis (9), nuclear medicine (10), and material sciences (11), to name just a few.

For many years, the discovery of small organic ligands to protein targets has been performed by screening very large sets of organic molecules (termed chemical libraries), one by one (1–3, 12). Large pharmaceutical companies typically construct and assay chemical libraries comprising 1 million organic molecules or more, using high-throughput screening procedures. It has

been estimated that the assembly of such libraries may cost between $0.4 and $2 billion (i.e., approximately $1,000 per compound, multiplied by 1 million compounds) (13). While the value of high-throughput library screening has been demonstrated in various pharmaceutical applications, it is not uncommon that binding molecules of sufficient affinity and specificity be discovered using conventional screening campaigns (14).

In light of these considerations, significant efforts have been devoted and continue to be devoted to the discovery and development of methods that facilitate the identification of specific binding molecules to macromolecular targets and proteins in particular. DNA-encoded chemical library technology enables the construction and screening of compound sets of unprecedented size and, as a consequence, the discovery of small organic ligands. When the size of a library grows, the concentration of individual library members decreases, to an extent that those molecules may no longer be detectable even with the most sophisticated analytical methods. However, DNA tags allow the amplification, identification, and relative quantification of molecules in very large libraries.

FROM ENCODED LIBRARIES OF POLYPEPTIDES TO DNA-ENCODED CHEMICAL LIBRARIES

The advent of encoded combinatorial libraries of polypeptides not only has played an important role for the engineering of proteins with novel properties, with applications in many research fields, but also has been conceptually instrumental for the genesis of DNA-encoded chemical libraries. For this reason, it is convenient to briefly discuss a few milestones in this research area.

In 1985, George P. Smith proposed the use of filamentous phage as tools for the display of polypeptides on the surface of these bacterial viruses (15). In a popular implementation, peptides or proteins would be genetically fused at the N-terminal end of the minor coat protein pIII of filamentous phage. The resulting viral particle features a potential functional property (e.g., a binding phenotype, embodied by the polypeptide on the phage surface), while simultaneously bearing the corresponding genetic information (i.e., genotype) as part of the modified phage genome (**Figure 1**). The central idea is that genotype and phenotype are linked in a single phage package. Combinatorial phage display libraries could be generated by inserting a degenerate DNA sequence upstream of gene III (coding for the minor coat protein pIII, downstream of the leader peptide). This simple cloning procedure enables the generation and interrogation of billions of different peptides. Initially, combinatorial phage display libraries of peptides were used to identify the sequence of linear epitopes, recognized by monoclonal antibodies (16). Soon afterward, however, the groups of Sir Gregory Winter (17) and one of the authors (R.A.L; 18) realized that very large combinatorial libraries of antibodies, rather than of short peptides, could be created and that phage display could be used to amplify and isolate rare binding specificities within those libraries (19, 20). In this process, man-made antibodies could be created by interrogating large combinatorial antibody libraries against virtually any target antigen of interest, immobilized on a solid support. The procedure bears logical and functional similarities to the clonal expansion of antigen-specific B cells in the immune system. Indeed, a B lymphocyte with an immunoglobulin on its surface constitutes a cellular entity, which simultaneously features a binding phenotype and the corresponding genotype, in full analogy to an antibody fragment displayed on the surface of a filamentous phage (21). Thus, genotype and phenotype are linked in B cells, albeit at the cellular level.

The impact of antibody phage technology in basic research and in pharmaceutical sciences has been substantial. For example, the fully human antibody Humira™ [a tumor necrosis factor (TNF) blocker and one of the best-selling pharmaceuticals of all time] was generated thanks to antibody phage technology (22). Antibody phage libraries have also taught us important lessons. It has been formally demonstrated that larger libraries are much more likely to yield numerous

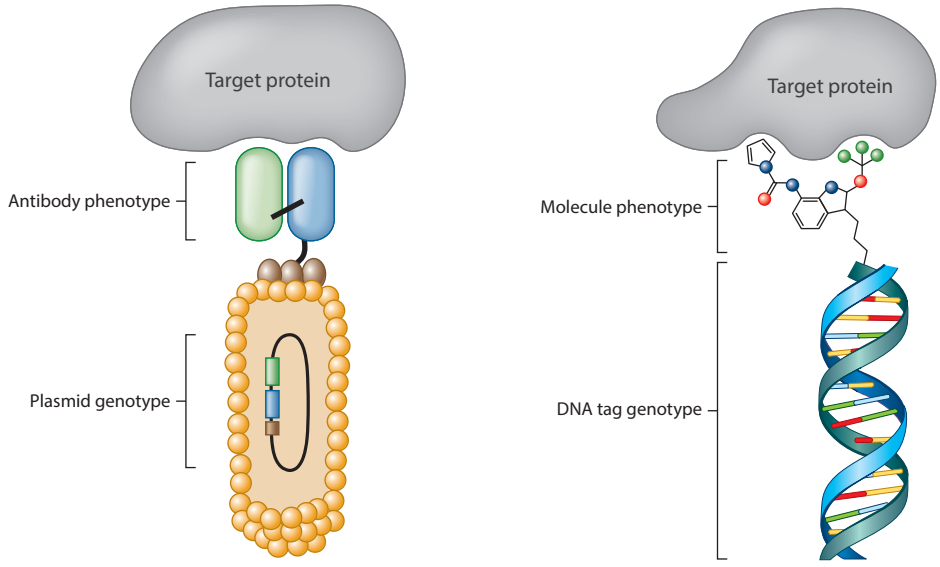


Figure 1

Schematic representation of antibody phage display libraries and of DNA-encoded chemical libraries. In antibody phage display libraries, an antibody fragment (e.g., an scFv fragment) is displayed on the surface of filamentous phage and could potentially bind to a cognate protein target. The genetic information coding for the recombinant antibody is contained in the genome of the bacteriophage. In full analogy, small organic molecules can be attached to DNA fragments, serving as amplifiable identification bar codes.

and high-affinity antibodies, compared with smaller aliquots of the same library. For example, a library comprising 65 billion antibody clones allowed the isolation of numerous antibodies with a dissociation constant in the low nanomolar range, while a portion of the same library, comprising only 10 million antibodies, yielded only a few binders, with Kd values in the 0.8–12 μM range, confirming that not only molecular design but also library size matters (23). Encoded combinatorial libraries of antibodies and of other polypeptides have provided an inspiration for DNA-encoded chemical libraries: a general technology that could be broadly applied to the world of organic molecules. Phage display, yeast display, and ribosome display are some of the most popular selection systems that enable a physical connection between a phenotype and the corresponding genetic information, exploiting the biosynthetic linkage between nucleic acids and proteins (24, 25). This biosynthetic dependence is not needed in DNA-encoded chemical libraries, which use DNA tags only as amplifiable identification bar codes.

In 1992, one of the authors (R.A.L.) together with Sydney Brenner postulated that it should be possible to encode chemistry with DNA (26). The authors envisaged the possibility of simultaneously synthesizing distinctive polypeptide and oligonucleotide sequences on beads, using orthogonal chemistry and split-and-pool procedures. The synthesis of oligonucleotides on beads would not be limited to the stepwise assembly of bases, since other assembly strategies (e.g., the stepwise ligation of DNA fragments) could also be considered. The 1992 landmark article revealed that oligonucleotides could serve as amplifiable bar codes for the corresponding peptide displayed on the same bead. The bar code would only act as identifier for the corresponding peptide structure and would not act as genotype in a biosynthetic sense. Indeed, the original article anticipated that the bead could be replaced by a generic chemical linker and the connection between binding phenotype and the corresponding bar code would be preserved (**Figure 1**).

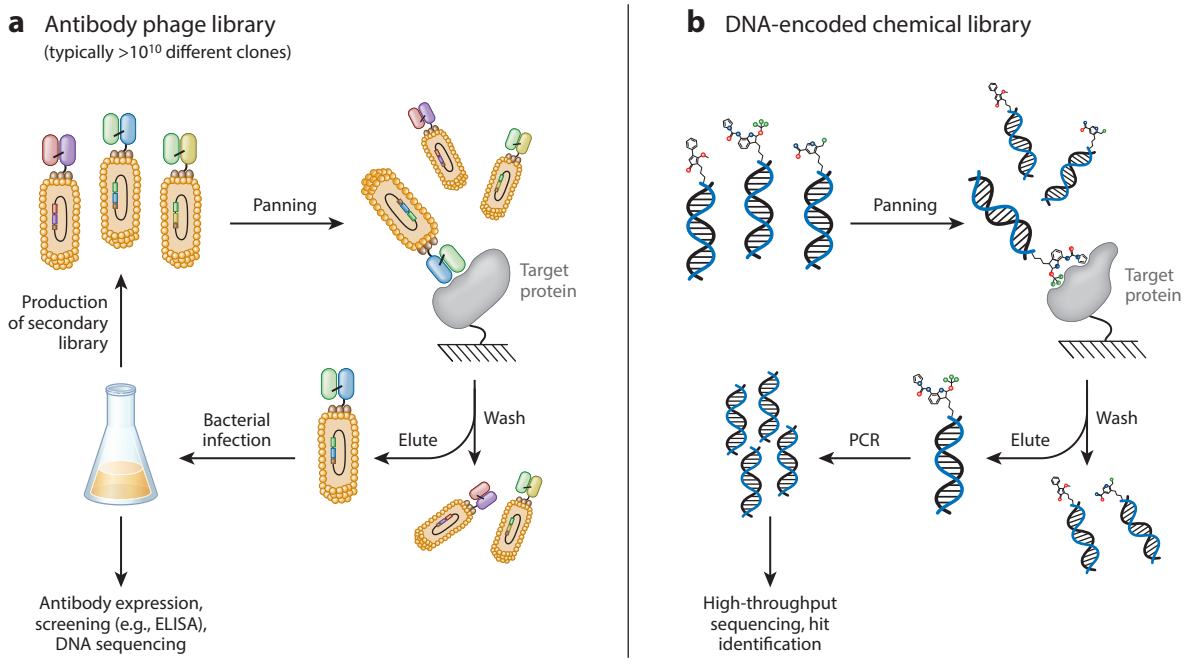


Figure 2

Schematic representation of biopanning procedures with antibody phage display libraries and with DNA-encoded chemical libraries. (*a*) A large library of phage antibodies is incubated with the target protein of interest immobilized on a solid support. Selective binders are captured on the affinity support, while the majority of other phage antibodies (that do not bind to the target) can be washed away. Selected phage particles can be amplified by infecting bacteria, leading to an amplification step and to the generation of more phage particles, which can be submitted to a second round of panning. Alternatively, infected bacteria can be plated onto selective plates and individual colonies correspond to distinct monoclonal antibody clones. (*b*) Similarly, large collections of organic molecules (individually tagged with DNA bar codes) can be interrogated using an affinity capture procedure. Preferentially enriched molecules can be identified by polymerase chain reaction (PCR) amplification of the DNA tags, followed by high-throughput DNA sequencing.

Shortly afterward, Needels and colleagues (27), as well as Janda and colleagues (28), exemplified the Brenner and Lerner concept with the synthesis of peptide libraries, successfully using bead-based libraries to retrieve known antibody epitopes. In 2004, three groups reported on the construction of DNA-encoded combinatorial libraries of organic molecules and on the selection of specific binders (29–31), using libraries devoid of beads and affinity-capture procedures, which are similar to the ones previously used for the panning of phage display libraries (**Figure 2**). Several DNA-encoded chemical libraries were subsequently synthesized and screened, as we see in the following sections.

Many reviews have covered the field of DNA-encoded chemistry, some of which are listed here (32–42). In this review, we wish not only to provide essential information about the origin of the technology but also to discuss strategies that will continue to shape future research activities in the field.

TYPES OF DNA-ENCODED CHEMICAL LIBRARIES

One can distinguish between single-pharmacophore and dual-pharmacophore DNA-encoded chemical libraries (**Figure 3**). In the first case, individual molecules (no matter how complex)

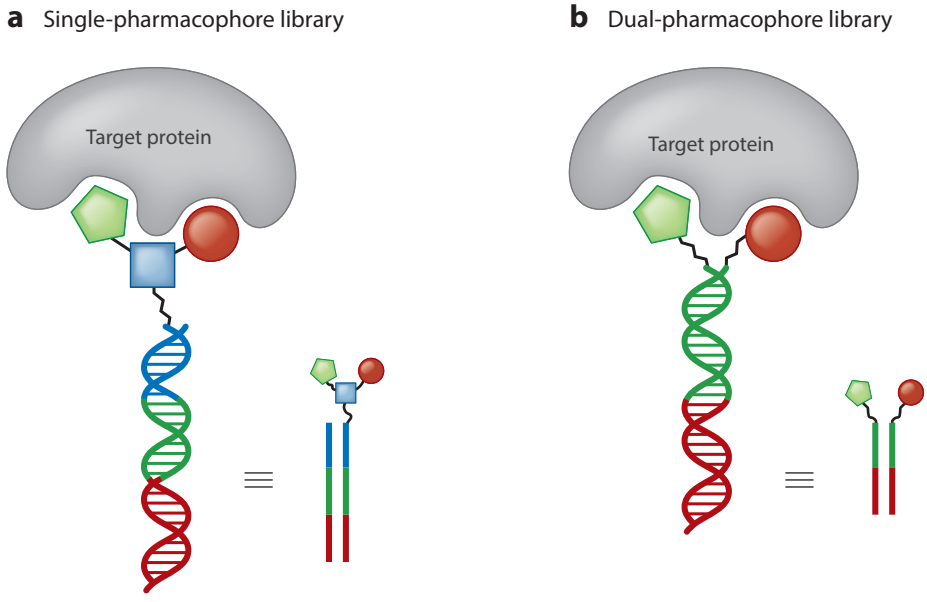


Figure 3

Schematic representation of (*a*) single-pharmacophore and (*b*) dual-pharmacophore DNA-encoded chemical libraries. A linear schematic representation is also displayed next to the double-helix representation of the DNA structure. This graphical representation is used in subsequent figures describing encoding procedures.

are attached to distinctive DNA fragments. In most instances, those molecules are coupled to one of the two complementary DNA strands. However, researchers at Praecis [now GlaxoSmithKline (GSK)] had introduced a technology featuring the use of a chemical linker (43), which would provide a covalent connection between complementary strands and, at the same time, a site for the stepwise growth of organic molecular structures (**Figure 3**).

By contrast, dual-pharmacophore chemical libraries can be produced by coupling pairs of organic molecules at the extremity of complementary DNA strands (**Figure 3**). If two sublibraries are constructed, featuring two sets of partially complementary oligonucleotides (each modified with organic molecules), these two sublibraries may then be reassembled (i.e., hybridized) to create a large combinatorial diversity. This technology is sometimes referred to as encoded self-assembling chemical (ESAC) library technology (30). Details of their construction and specific aspects of the technology are described in more detail in the following section.

ENCODING AND LIBRARY SYNTHESIS STRATEGIES

Over the last two decades, various techniques have been used for the tagging of organic molecules with DNA and, consequently, for the construction of encoded combinatorial libraries. Let us consider the construction of a library comprising 1 billion (10^9) different molecules. Obviously, it would be highly impractical (and prohibitively expensive) to couple 1 billion organic molecules to 1 billion different oligonucleotides (or DNA fragments) one at a time. Luckily, split-and-pool synthesis procedures greatly facilitate the facile and economic assembly of single-pharmacophore chemical libraries. Alternatively, ESAC library technology can also be used to create large molecular repertoires, capitalizing on the combinatorial self-assembly of sublibraries.

Single-Pharmacophore Chemical Libraries

The most commonly used method for the encoding and construction of single-pharmacophore DNA-encoded chemical libraries relies on DNA-recorded (sometimes also referred to as DNA-encoded) synthesis. From a conceptual viewpoint, the simplest way to construct a library relies on the stepwise ligation of double-stranded DNA fragments, which record the identity of individual chemical building blocks (**Figure 4*a***). As an example, let us consider a set of 100 different organic molecules, each of which is coupled to a distinctive double-stranded DNA fragment. The set of 100 organic molecules coupled to the cognate DNA fragments can be pooled, as the DNA sequences allow an unambiguous identification of the corresponding building blocks. The resulting pool can be split into individual reaction vessels (e.g., 100 vessels), each of which is allowed to react with a new chemical moiety. The identity of each building block, which is being added to a nascent molecular structure through a suitable chemical reaction, can then be encoded in a ligation step

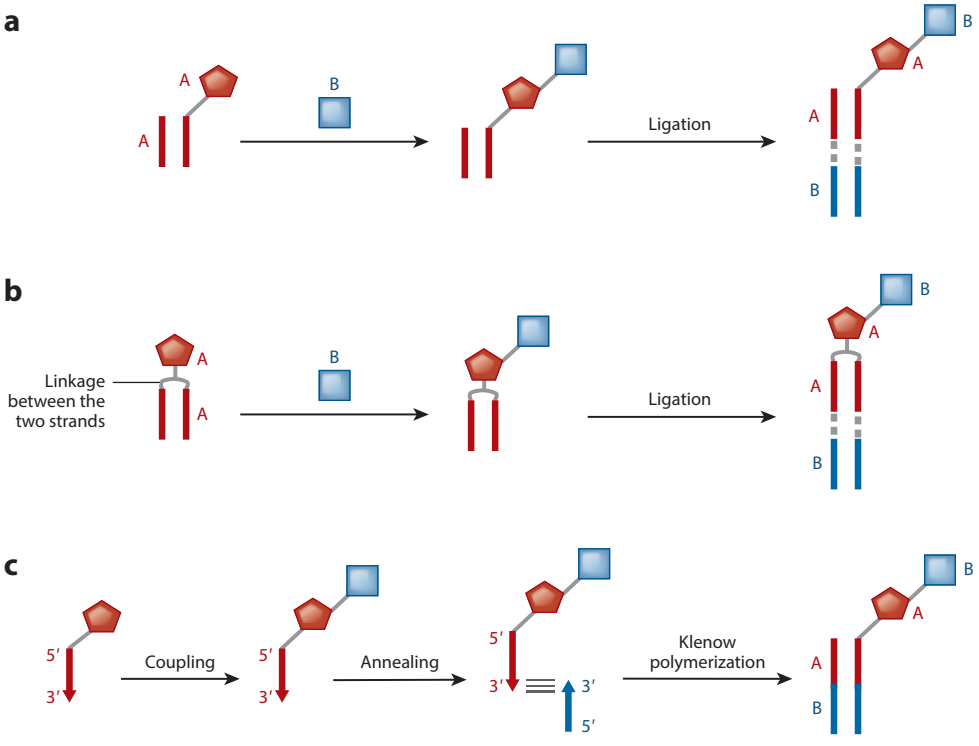


Figure 4

Schematic representation of encoding strategies for DNA-recorded chemical libraries. (*a*) In the simplest implementation of this technology, each building block in the synthesis procedure is encoded (i.e., identified) by a distinctive double-stranded DNA fragment. After each chemical reaction, the identity of the newly introduced building block is provided by an additional DNA fragment, which is ligated to the nascent DNA structure. (*b*) In a variation of the procedure described in panel *a*, the complementary DNA strands are connected by a linker, which also supports the growth of the nascent molecule. (*c*) In a different encoding procedure, a first building block is attached at the 5′ end of an oligonucleotide, which contains a suitable identification code. After a second building block has been added to the nascent molecule by chemical reaction, its identity is encoded by the annealing of a partially complementary oligonucleotide, followed by a Klenow polymerization procedure.

with a double-stranded DNA fragment. That is, the use of 2 × 100 building blocks (i.e., 200 organic molecules) and of 2 × 100 DNA fragments (i.e., 200 DNA fragments) allows the construction of an encoded library, comprising 10,000 organic molecules. These molecules originate from the encoded combination of 100 × 100 building blocks. If the process is reiterated, larger libraries can be constructed (e.g., a 1-million-compound library from 3 sets of 100 building blocks each). A variation on the theme features the use of linkers that connect the two complementary strands and enable the stepwise growth of a molecular structure (**Figure 4*b***) (43). It is worth noting that the molecular mass of the DNA-encoded molecules increases at each round of synthesis (with potentially deleterious effects for pharmaceutical applications). Furthermore, the increase in number of reaction steps may lead to a worsening of library purity.

The procedure described above allows the construction of libraries with relatively short DNA bar codes, which are easy to sequence using modern high-throughput procedures. However, the encoding of *n* building blocks through ligation of DNA fragments requires 2*n* oligonucleotides, owing to the heterodimeric nature of DNA. The group of one author (D.N.) has described an encoding procedure that makes use of a set of only *n* oligonucleotides for the encoding of *n* building blocks (**Figure 4*c***). Let us consider a set of 100 building blocks, each of which is coupled at the 5′ end of a corresponding oligonucleotide (e.g., through the functionalization of an appended linker terminating with a primary amine). The oligonucleotides are designed to have an identical sequence, except for a short central portion, which is distinctive for each building block and acts as bar code. The 100-oligonucleotide conjugates can be purified [e.g., by high-performance liquid chromatography (HPLC)] and can be pooled, as each building block can be identified by the bar code to which it is associated. At this stage, splitting of the mixture in individual reaction vessels (e.g., 100 vessels) allows the execution of reactions with a second set of building blocks (e.g., 100 building blocks). The identity of each reaction can be recorded by annealing with a partially complementary oligonucleotide and followed by a Klenow polymerization step (44–49). The procedure can be expanded to three or more sets of building blocks by means of splint oligonucleotides and ligation procedures (50, 51).

An alternative way of encoding single-pharmacophore chemical libraries has been pioneered by the group of David Liu and is often referred to as DNA-templated (or DNA-directed) synthesis (**Figure 5*a***). In this approach, long preformed oligonucleotides are used to mediate a stepwise series of annealing steps with shorter complementary oligonucleotides, carrying chemical moieties that can react with organic structures displayed on the longer oligonucleotide templates (29). These reaction steps may benefit from the high effective molarity advantage caused by the heterodimerization of the complementary oligonucleotides. Indeed, the Liu group (52) has demonstrated the usefulness of DNA-templated methods for the execution of bimolecular chemical reactions, which would otherwise be inefficient in water solution and in the presence of DNA. The transfer of building blocks onto a nascent molecular structure needs to be followed by a suitable cleavage step, to allow for the next round of coupling to proceed. The procedure appears to be particularly suited for the synthesis of complex macrocyclic structures (53). In principle, the requirement for high-fidelity annealing between the long oligonucleotide template and the shorter complementary oligonucleotides could represent a limitation for the DNA-directed synthesis methodology. However, Xiaoyu Li and coworkers (54) have described an ingenious optimization of the encoding method featuring the use of an oligonucleotide containing a nascent organic molecule and a polyinosine segment that serves as promiscuous hybridization stretch for short oligonucleotides, and carries transferrable chemical moieties (**Figure 5*b***).

Two further variations of the DNA-templated synthesis method have been described. The Danish company Vipergen has developed an elegant methodology, termed YoctoReactor™

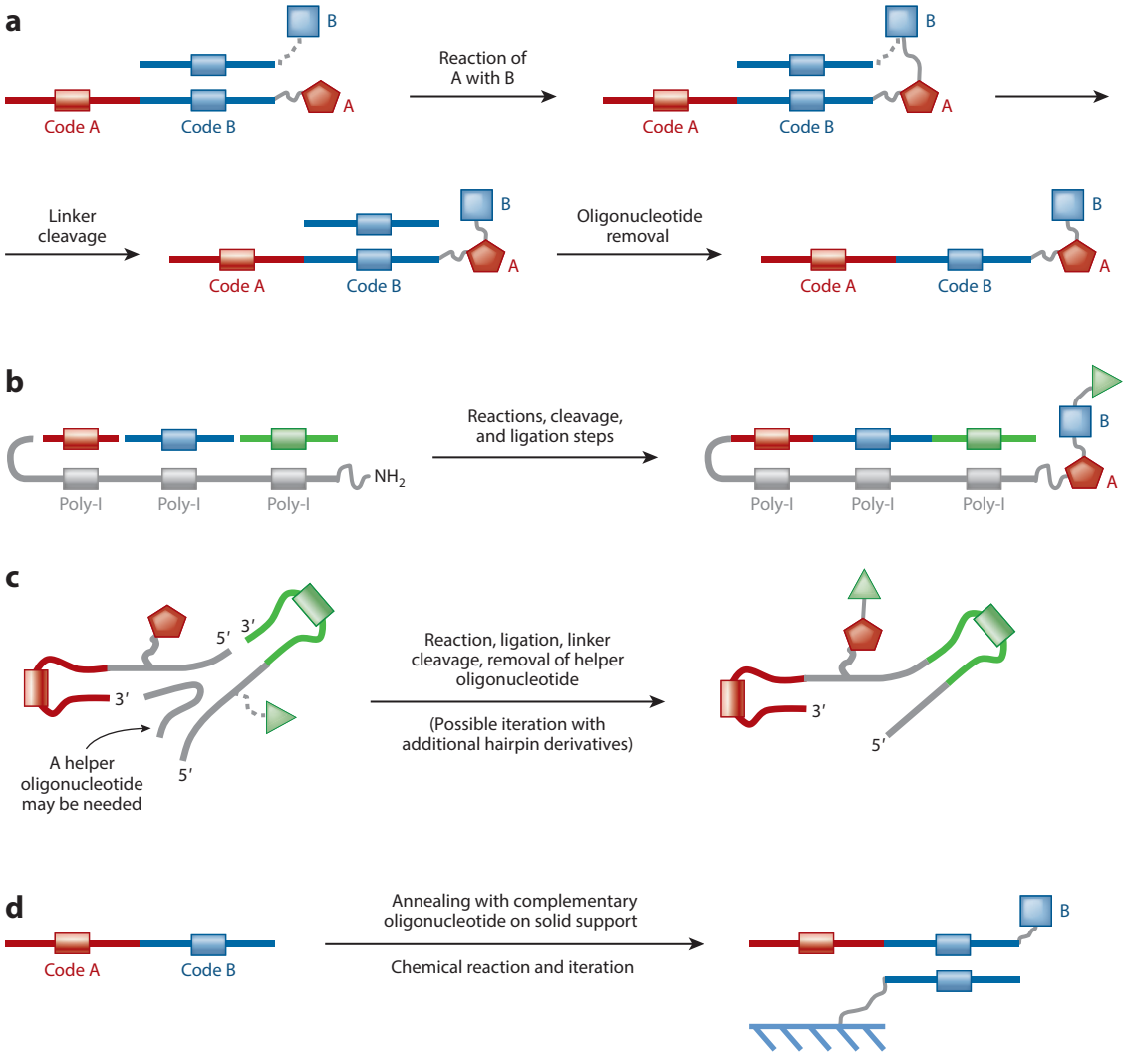


Figure 5

Schematic representation of encoding strategies for DNA-templated synthesis of chemical libraries. (*a*) An oligonucleotide template (here depicted for the construction of a molecule based on two building blocks) is annealed with a chemically modified oligonucleotide, which transfers a chemical moiety to the nascent molecule. At the end of the coupling step, a cleavable connection is made between the building block B and the corresponding oligonucleotide (*dotted line*). (*b*) In a modified procedure, a general template (containing poly-I stretches for annealing with various coding segments) is sequentially hybridized with oligonucleotide derivatives, which transfer building blocks onto a nascent molecule. (*c*) In YoctoReactor™ technology, hairpin structures (containing a coding sequence and a building block) are used to mediate successive cycles of DNA ligation and of chemical reactions. Cleavable linkers are indicated with a dotted line. (*d*) In DNA routing strategies, long oligonucleotides (containing multiple coding sequences) are sequentially hybridized to columns carrying complementary oligonucleotides. Individual column hybridization steps dictate which chemical reaction (i.e., which building block addition) is performed on a given nascent molecule.

technology, that features the annealing and subsequent ligation of hairpin loops, which carry suitable reactive moieties (55). The end result of a series of hybridization, building block transfer, and cleavage events is a set of double-stranded DNA products containing the sequence information needed to unambiguously identify the molecular entities, assembled through proximity-enhanced chemical reactions (**Figure 5*c***). A second technology [described as early as 2004 by the Harbury group (31) and sometimes referred to as DNA routing] makes use of long preformed oligonucleotides, which are sequentially annealed to complementary oligonucleotides on solid supports. These hybridization steps allow the transfer of different building blocks in correspondence to given oligonucleotide hybridization events, thus providing a direct linkage between the DNA sequence of the oligonucleotide templates and the resulting chemical structure of the stepwise synthesized molecules (**Figure 5*d***). The procedure depends on the fidelity of the hybridization steps and may be complex to perform in practice if serial (rather than parallel) affinity-based routing steps are performed to direct chemical synthesis. The authors have reported the successful synthesis of a peptide library, from which a known ligand could be retrieved (31, 56–58).

Dual-Pharmacophore Chemical Libraries

From the beginning, dual-pharmacophore chemical libraries were designed to be compatible with either microarray-based or DNA sequencing–based decoding procedures (30, 51, 59). In the simplest implementation, two sets of oligonucleotides would be used for library construction, featuring a conserved portion (essential for the self-assembly of the library) and a variable portion (distinctive for each chemical compound) (30). The resulting ESAC libraries can be decoded using microarrays coated with sets of complementary oligonucleotides, but this procedure does not provide information about the enrichment factors of individual pairs of building blocks, since the bar codes are present in different DNA strands (**Figure 6*a***).

A more elaborate procedure allows the unambiguous identification of pairs of building blocks in ESAC libraries (described in the section titled Decoding Strategies) and in a library that can be decoded by high-throughput DNA sequencing (**Figure 6*b***). In this setting, a single oligonucleotide containing an abasic site is coupled to various chemical building blocks at the 3′ end, and these reactions are subsequently encoded by ligation with suitable oligonucleotide bar codes (51). In parallel, complementary oligonucleotides can be modified at the 5′ end, generating a second sublibrary that can be used for self-assembly procedures. Conveniently, the base segment that identifies molecular entities at the 5′ end of one sublibrary corresponds to the abasic site in the sublibrary modified at the 3′ end. A Klenow polymerization step allows the incorporation of pairs of coding segments into a single oligonucleotide, thus permitting the unambiguous identification of all building block combinations within the library (**Figure 6*c***).

Heterodimeric and multimeric self-assembling structures can also be generated using DNA templates and peptide nucleic acids (PNAs) (**Figure 6*d***). Winssinger and colleagues (37, 60) have extensively used this technology to isolate binders to various molecular targets. Importantly, PNA-encoded chemical libraries can also be encoded by the stepwise chemical coupling of PNA fragments (61). In most experimental conditions, PNAs are more stable than the corresponding DNA-based oligonucleotides and may therefore be compatible with reaction conditions that are not possible in the presence of DNA (37). However, unlike DNA, PNA structures cannot be amplified by a suitable polymerase procedure [e.g., polymerase chain reaction (PCR)]. The DNA moiety that directs the assembly of chemically functionalized PNAs allows a PCR amplification step after selection on target proteins, thus facilitating subsequent decoding procedures.

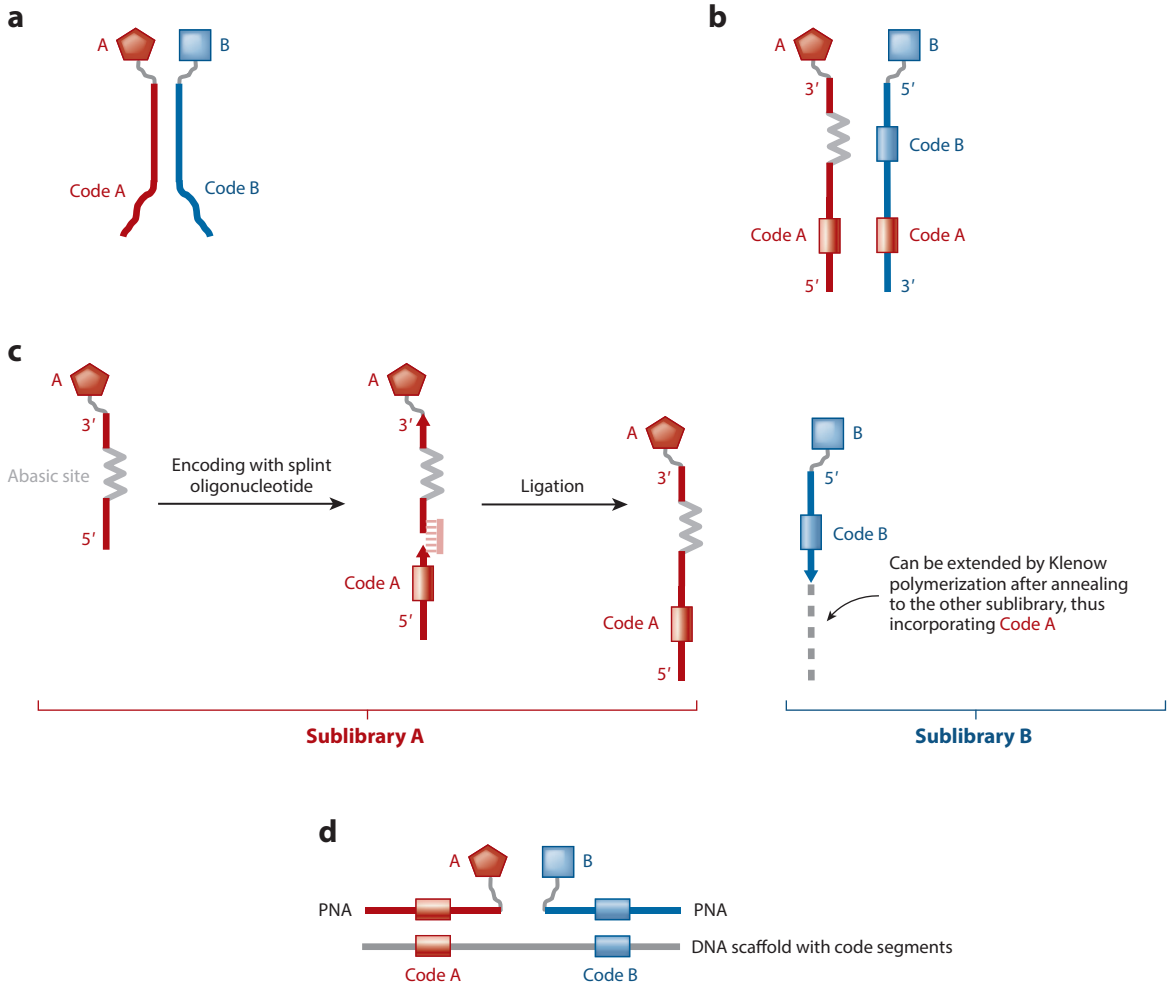


Figure 6

Encoding strategies for dual-pharmacophore encoded chemical libraries. (*a*) General structure of encoded self-assembling chemical (ESAC) library members, which are decoded by hybridization onto oligonucleotide microarrays. (*b*) In a more powerful procedure, ESAC library members are encoded with a methodology that leads to the simultaneous presence of two codes (identifying building blocks A and B) onto one of the two complementary strands. (*c*) This encoding strategy is compatible with high-throughput sequencing decoding procedures. Members of the red sublibrary are constructed by the coupling of building blocks to a general oligonucleotide carrying an abasic site, followed by encoding by ligation assisted by a splint degradable oligonucleotide. By contrast, members of the blue sublibrary are created by coupling building blocks to the 5′ extremity of suitable oligonucleotides, carrying an identification code. The two sublibraries can then be annealed followed by a Klenow polymerization step. (*d*) In one of the possible implementations of the technology, chemically modified peptide nucleic acids (PNAs) are hybridized to DNA templates, which carry suitable complementary coding sequences.

SELECTION METHODOLOGIES AND THE IMPACT OF SELECTION CONDITIONS

Figure 7 schematically summarizes methods that can be used for the screening of DNA-encoded chemical libraries. In full analogy to phage display libraries (17–20, 23) or ribosome display libraries (25), specific binders can be identified from DNA-encoded chemical libraries by performing an

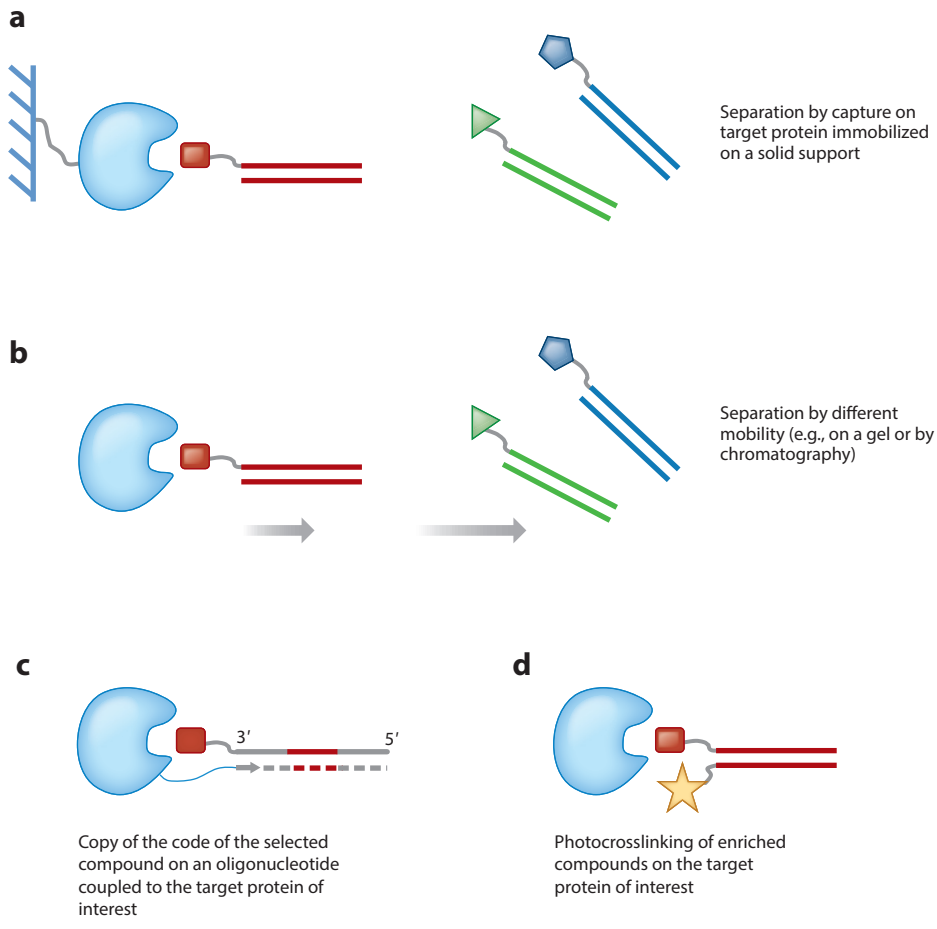


Figure 7

Selection methods for DNA-encoded chemical libraries. (*a*) Library members are captured on a target protein that is immobilized on a solid support. (*b*) DNA derivatives, capable of binding to a target protein with sufficient stability in given experimental conditions, are separated from nonbinding library members by chromatography or by electrophoresis. (*c*) Library members encoded by single-stranded DNA molecules, capable of binding to a target protein of interest equipped with a suitable oligonucleotide primer, can be identified by a DNA polymerization step followed by polymerase chain reaction amplification. (*d*) Capture of preferentially binding library members by photocrosslinking.

affinity capture step on the target protein of interest, immobilized on a solid support (**Figure 7*a***). The use of magnetic beads (46, 47), streptavidin-coated tips (43), and cyanogen bromide (CNBr)-activated sepharose (after a suitable modification step with the target protein; 49, 62) have been described and reviewed (63). Certain experimental parameters (e.g., washing conditions, concentration of detergents, protein coating procedures) have been shown to have a profound impact both on the absolute recovery and on the selectivity of selection fingerprints (47, 63). The use of quantitative PCR methodologies has recently been introduced as a tool for the experimental determination of selection performance, especially when using a mixture of DNA-tagged ligands of known biochemical properties and compounds of irrelevant specificity used as negative controls (64, 65).

Capillary electrophoresis techniques can also be considered for the separation of protein-bound DNA derivatives and the rest of the library (66) (**Figure 7*b***). Alternatively, a selective PCR amplification of the DNA tag associated with preferential binders can be used, if the protein target of interest is equipped with a suitable oligonucleotide primer (**Figure 7*c***) (67, 68). This methodology has been implemented in the context of water-in-oil emulsions, allowing a preferential productive PCR amplification step of those coding sequences associated with protein binders (42).

The use of photocrosslinking methods (**Figure 7*d***) for the screening of DNA-encoded libraries has also been proposed, with encouraging experimental results (69–70). These approaches can be performed in solution and may avoid certain artifacts that could potentially arise from the immobilization of target proteins on solid supports.

DECODING STRATEGIES

The composition of a DNA-encoded chemical library can be characterized by DNA sequencing before and after protein selections, thus providing an experimental determination of relative enrichment factors for all compounds in the library (45, 46, 48, 53, 71). We assume that the frequency at which a certain DNA sequence is found corresponds to the relative abundance of the corresponding DNA fragment (and hence the associated compound) in the library. Since each base can have four different possibilities (A, C, G, and T), a stretch of 10 bases can encode $4^{10} = 2^{20} = 1{,}048{,}576$ different events (41). Considering the iterative nature of most library construction steps, the coding sequences that identify the individual molecules are often found in variable blocks flanked by constant sequences rather than in a single continuous stretch (differing across library members). The use of constant sequence segments facilitates PCR amplification procedures as well as the implementation of certain encoding methods (**Figures 4–6**).

The advent of high-throughput DNA sequencing technologies (e.g., 454 technology or Illumina sequencing; 45, 46, 72) has revolutionized the field of DNA-encoded chemistry and permitted the screening of very large compound collections. The relative abundance of compounds in a library consisting of two sets of building blocks can be depicted as a cube, in which two dimensions are used to determine the identity of pairs of building blocks and the third dimension corresponds to the relative compound frequency (i.e., to the number of times individual sequence tags are read in a high-throughput sequencing experiment) (**Figure 8*a***). A library based on three sets of building blocks needs four dimensions to display the identity of all compounds and their relative frequencies. In most instances, spheres of different radius or color are used to graphically represent this pseudo–four-dimensional space (**Figure 8*b***). The use of suitable cut-off filters in frequency counts allows the display of the most common molecules in a collection (e.g., the most enriched compounds after a selection procedure). Lines and planes of enriched compounds in libraries based on three sets of building blocks correspond to a set of molecules, for which two or one building blocks (respectively) dominate the selection procedure. It is indeed possible, as we see in the next section, that not all building blocks equally contribute to binding affinity.

LIBRARY DESIGN, CHEMICAL REACTIONS, AND HIT COMPOUNDS

Many different chemical strategies can be considered when constructing an encoded library using multistep synthesis procedures. Some approaches have been inspired by previous combinatorial chemistry approaches, making use of DNA-friendly transformations. Compilations of chemical reactions that can be performed on DNA and are potentially compatible with library construction have been described (52, 73). However, not all building blocks are equally reactive even for the most straightforward transformations (e.g., amide bond formation) (74), and investigating the reaction conditions for individual compounds prior to library synthesis is recommended.

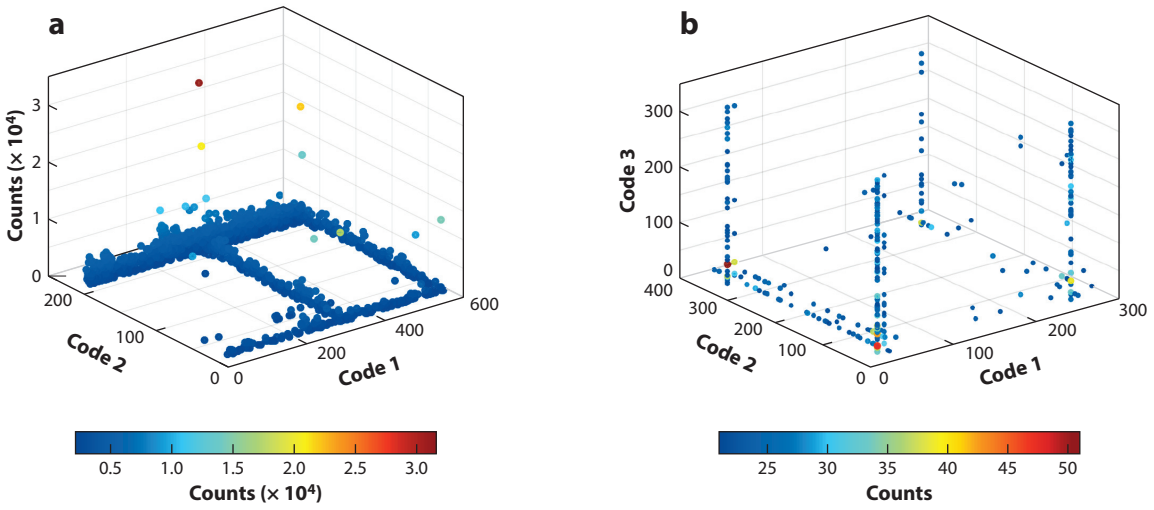


Figure 8

Illustrative fingerprints of selection results, performed with a library consisting of (*a*) two or (*b*) three sets of building blocks.

Members of single-pharmacophore chemical libraries, produced using the DNA-recorded method, are routinely purified after the first step of synthesis. However, all subsequent transformations are performed with pools of compounds and, as a consequence, are not purified and are difficult to characterize using analytical methodologies. Catch and cap technologies have been proposed as an avenue to achieve higher library purity, but the additional chemical steps may lead to loss in DNA quantities (75). Dual-pharmacophore chemical libraries (such as ESAC libraries) are generated by the combinatorial self-assembly of sublibraries of very high quality, whose members are individually purified by HPLC and checked by mass spectrometry.

Recent surveys of selection hits have been published that describe publications with hit compounds isolated from DNA-encoded chemical libraries (32, 40, 41). In this section, we highlight some notable hit discovery procedures, detailing how the corresponding libraries had been constructed and which building blocks mostly contributed to the experimentally observed affinity constants.

A few interesting hits, isolated from DNA-encoded chemical libraries based on two sets of building blocks, have been described. For example, Leimbacher et al. (47) described the synthesis of a combinatorial library featuring a set of amino acids that were subsequently reacted with a set of carboxylic acids, from which specific interleukin-2 inhibitors could be isolated (**Figure 9*a***). In those compounds, a methyl indole derivative played a crucial role in the molecular recognition of the cytokine target (47). In this library and in many other applications, it was convenient to substitute the DNA moiety with a fluorophore to facilitate the determination of dissociation constants in solution using fluorescence polarization methodologies (see also the next section).

A library based on a central scaffold featuring two primary amino groups, which were sequentially modified with two sets of carboxylic acids, yielded high-affinity binders against various target proteins including selective albumin binders (71, 76) and tankyrase-1 inhibitors (71) (**Figure 9*b***). The latter protein could also be drugged using compact structures generated through the acylation of secondary amines (77) (**Figure 9*c***).

ESAC libraries have allowed the isolation of high-affinity α1-acid glycoprotein binders (**Figure 9*d***). Interestingly, the combination of two building blocks yielded binders with a

dissociation constant in the nanomolar range, and the individual building blocks did not exhibit any detectable binding in fluorescence polarization assays or in isothermal titration calorimetry (51). This observation suggests that a chelate-binding effect may be exploited for molecular recognition applications when low-affinity ligands recognizing adjacent pockets on the target protein of interest are joined together. The chemical nature of the linker connecting two building blocks can also strongly contribute to the resulting binding affinity (61, 78).

Libraries based on three or more sets of scaffolds have yielded interesting binders against targets of pharmaceutical interest. For example, scientists at GSK in 2009 described novel libraries, including one which contained >800 million substituted triazines, resulting from the combinatorial assembly of four sets of building blocks (**Figure 9*e***). The authors isolated inhibitors for Aurora A kinase and p38 mitogen-activated protein kinase (MAPK), some of which had potency in the low-nanomolar concentration range (43). Triazine-based libraries have also yielded high-affinity ligands toward other classes of protein targets, including a 2-nM inhibitor of neurokinin 3 (79) and a 10-nM binder to ADAMTS4 (80), indicating the versatility of triazines for library design (**Figure 9*e***). The neurokinin 3 hit discovery activities were remarkable, since encoded library technology was used to find inhibitors of a cell surface target using selection conditions that mimic the cellular environment (79).

The GSK group has reported the isolation of phosphoinositide 3-kinase-α (PI3Kα) inhibitors ($IC_{50} = 10$ nM) using a library of 3.5 million compounds assembled by the sequential reaction of 191 amino acids with an aromatic multifunctional scaffold, which was subsequently reacted with 96 boronates and 196 amines (81) (**Figure 9*f***). Suzuki- and Sonogashira-coupling methodologies are particularly attractive for the formation of carbon–carbon bonds in the presence of DNA (73, 82). Scientists at GSK also reported the successful isolation of potent inhibitors of the receptor-interacting protein 1 (RIP1) kinase from a library of 2.6 billion compounds obtained by the sequential reaction of three sets of compounds (632 amino acids and 6,594 amine capping reagents, such as carboxylic acids) (**Figure 9*g***). Interestingly, the most potent inhibitors (a set of substituted benzodiazepines) relied on the chemical structure of only the second and third building blocks, whereas the first set of building blocks could be removed without loss of inhibitory potency (83). This observation suggests that the same inhibitors could have been isolated from libraries based on two sets of building blocks—for example, having the design of **Figure 9*a***.

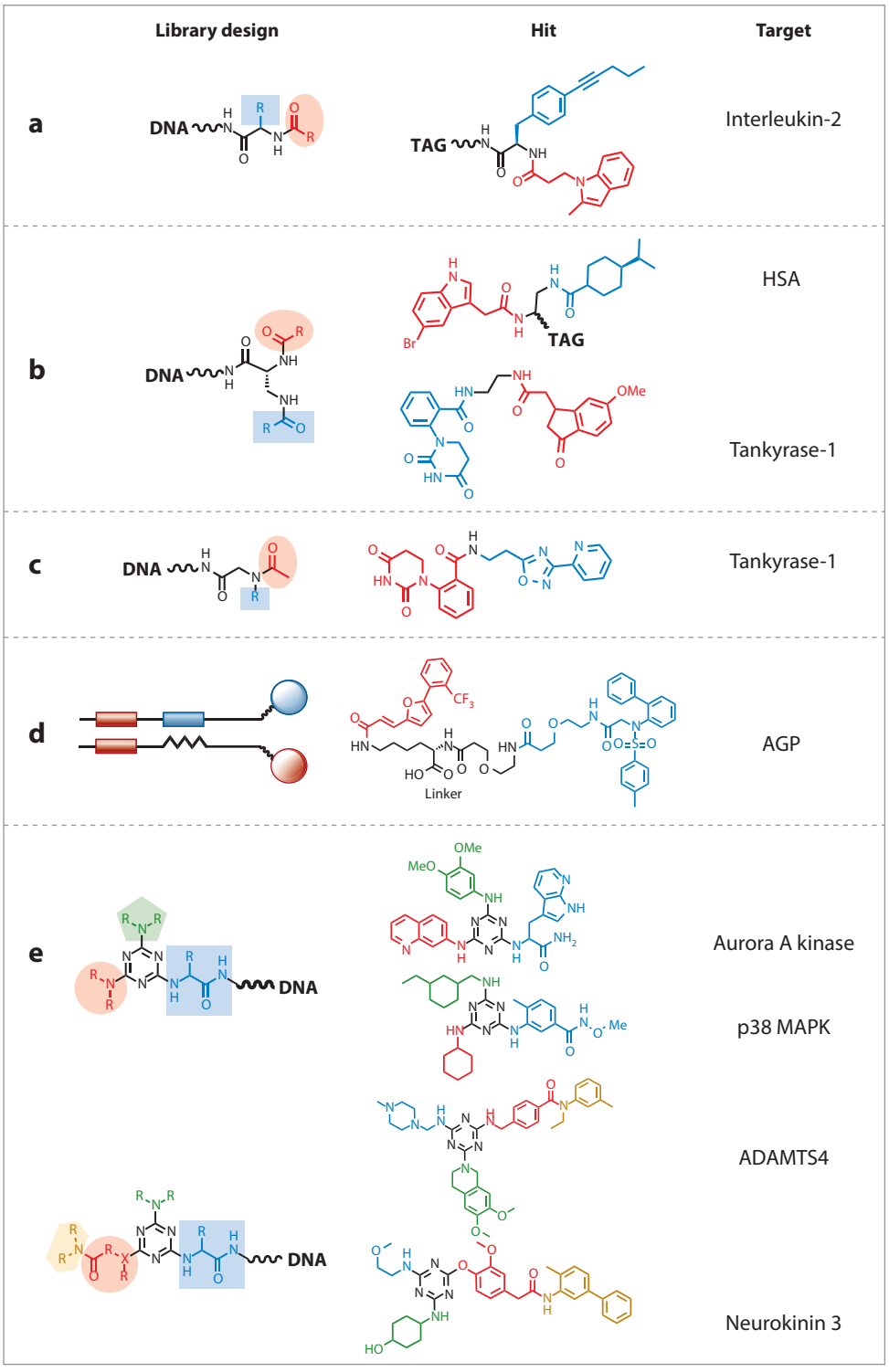
Using YoctoReactor technology, scientists at Vipergen have described the synthesis of a linear library containing >12 million compounds resulting from three series of acylation, reductive amination, and urea formation steps. The library allowed the isolation of an MAPK14 inhibitor with potency in the single-digit nanomolar range in cellular assays (84) (**Figure 9*b***).

Scientists at X-Chem have been very active in the field of DNA-encoded chemical libraries and have reported, among other results, the synthesis of a library containing >300 million compounds resulting from the sequential reaction of 2,259 primary amines, 666 bromoaryl acids, and 667 boronic acids (**Figure 9*i***). Fingerprints of selections performed against soluble epoxide hydrolase enabled the synthesis and confirmation of a hit with 2-nM potency and drug-like properties (molecular weight = 431 Da; cLogP = 3.1) (85).

DNA-encoded chemistry has also been used to generate libraries of peptide macrocycles, from which interesting binders against difficult target proteins have been isolated. For example, David Liu and collaborators (86) reported the isolation of a 50-nM inhibitor of insulin-degrading enzyme (IDE), from a library of 13,824 macrocycles constructed using a DNA-templated synthesis approach (**Figure 9*j***). A similar approach has been used by scientists at Ensemble to isolate XIAP (X-linked inhibitor of apoptosis protein) inhibitors with $IC_{50} = 140$ nM from a library of 160,000 cyclic peptides (87) (**Figure 9*j***).

Figure 9

Selected examples of binding molecules isolated from DNA-encoded chemical libraries. The individual examples show the general library designs, which were used for the selections. The color code of the building blocks is retained when displaying the chemical structure of the hit molecule, facilitating the detection of synthetic strategies and of modular structures. Abbreviations: ADAMTS4, a disintegrin and metalloproteinase with thrombospondin motif; AGP, α1-acid glycoprotein; BTK, Bruton tyrosine kinase; HSA, human serum albumin; IDE, insulin-degrading enzyme; JNK1, c-Jun N-terminal kinase 1; MAPK, mitogen-activated protein kinase; MEK2, mitogen-activated protein kinase 2; PCAF, P300/CBP-associated factor; PI3Kα, phosphoinositide 3-kinase-α; PNA, peptide nucleic acid; RIP1, receptor-interacting protein 1; sEH, soluble epoxide hydrolase; XIAP, X-linked inhibitor of apoptosis protein.



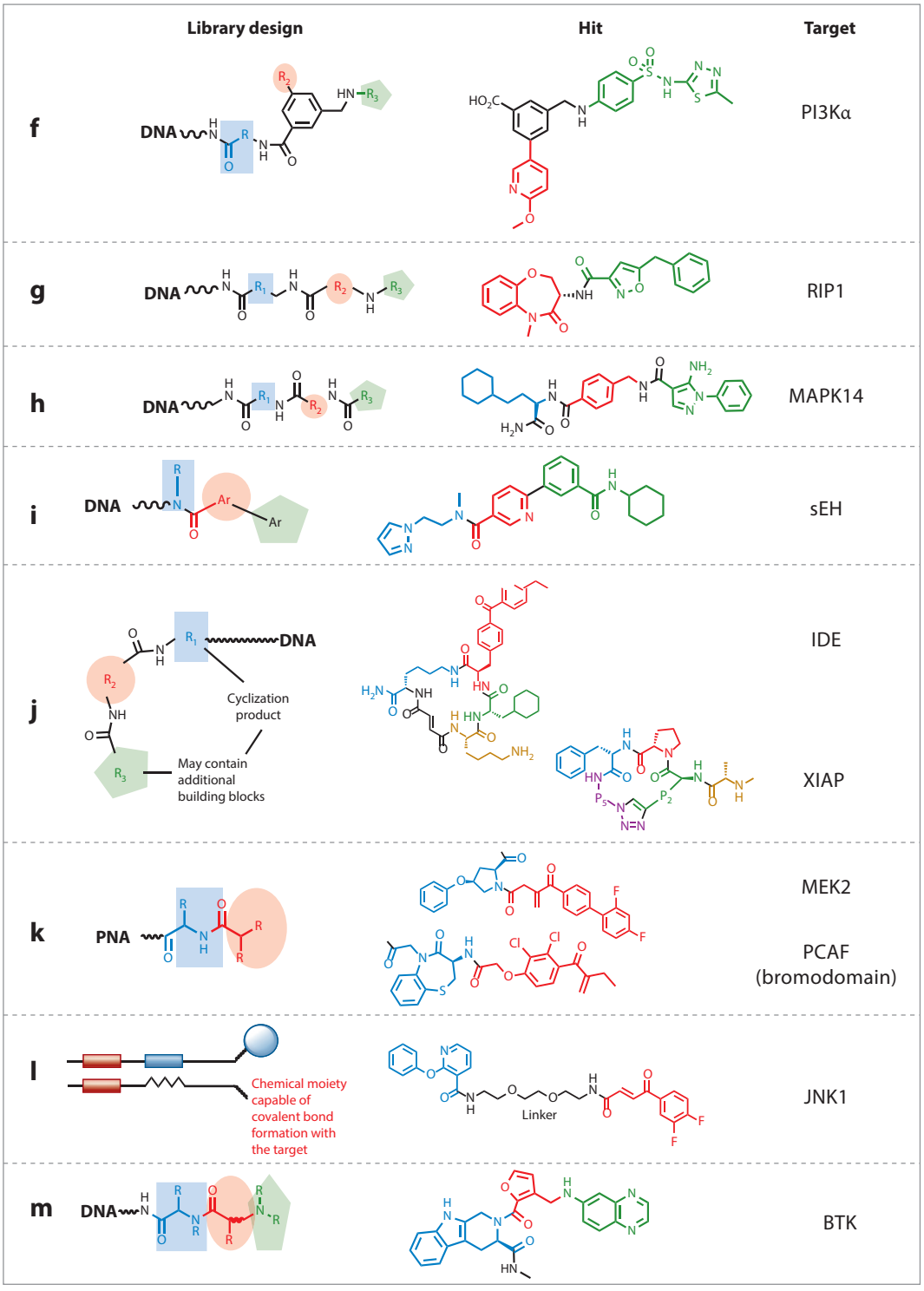


Figure 9

(*Continued*)

DNA-ENCODED CHEMICAL LIBRARIES FOR THE DISCOVERY OF COVALENT INHIBITORS

Most applications of DNA-encoded chemical libraries relate to the isolation of noncovalent inhibitors, but the technology is also ideally suited for the discovery of covalent binders. Winssinger and colleagues (88) used a library of 10,000 compounds, based on chemically modified PNAs and featuring suitable Michael acceptors, for the discovery of irreversible covalent binders of MEK2 (mitogen-activated protein kinase 2) (**Figure 9*k***). The same group reported the use of PNA-encoded chemical libraries for the discovery of two small molecules that form a covalent bond with cysteine residues conserved across the bromodomain family (epigenetic readers that are important for the regulation of transcriptional programs; 89) (**Figure 9*k***).

Zimmermann et al. (90) have recently described the use of ESAC libraries, containing building blocks suitable for the chemical modification of protein targets, to discover a covalent inhibitor of c-Jun N-terminal kinase 1 (JNK1), a protein containing 8 cysteine residues (of which one is in close proximity to the active site). Interestingly, the covalent protein modification was highly selective compared with other kinases with cysteine residues in the active site (90) and proceeded with 1:1 stoichiometry, even in the presence of >1,000-fold excess of glutathione, used to simulate the thiol-rich intracellular-reducing milieu (**Figure 9*l***).

Clark and colleagues (91) reported the construction of an encoded library containing 27 million acrylamide derivatives, from which potent irreversible covalent inhibitors of Bruton tyrosine kinase (BTK; a validated target for the treatment of certain forms of lymphoma with a cysteine residue in the active site) could be isolated (91) (**Figure 9*m***). Liu and coworkers (92) have recently used interaction-dependent screening methodologies with libraries of DNA-encoded bioactive compounds and mixtures of DNA-tagged human kinases to identify ligand–protein binding partners out of 32,096 possible combinations in a single solution–phase panning experiment. The results confirmed known small molecule–protein interactions and also revealed that ethacrynic acid is a novel ligand and inhibitor of MAP2K6 kinase. Ethacrynic acid inhibits MAP2K6, in part, through alkylation of a nonconserved cysteine residue (92).

ON-DNA AND OFF-DNA HIT VALIDATION STRATEGIES

DNA is an invaluable tag for the encoding and decoding of chemical libraries. However, for most applications, it is desirable to use chemical compounds in the absence of DNA. In many cases, the confirmation of hits from library selections proceeds through the resynthesis of the most enriched compounds off-DNA followed by a characterization of their binding properties using biochemical or biophysical assays. However, the synthesis and purification of organic molecules can be labor intensive, depending on the structure of the compounds and the number of hits that need to be measured. Furthermore, when the target is not amenable to biochemical characterization (e.g., using an enzyme inhibition assay), it can be difficult to measure binding affinities in the absence of a suitable tag. Surface plasmon resonance methodologies (e.g., BIAcore) may be complex to perform with proteins that lose activity upon immobilization on a solid support or during microsensor chip regeneration procedures. Isothermal titration calorimetry may allow the determination of dissociation constants in solution, but the procedure poses certain requirements on the availability, buffer composition, and solubility for both binding partners. For these reasons, it is often convenient to resynthesize hit compounds with a fluorophore moiety, at the position originally occupied by the DNA tag, and to perform fluorescence polarization measurements in solution (46). In an initial phase of the hit validation process, it is practical to resynthesize compounds on-DNA (51, 93). Zimmermann and colleagues (93) have recently described a general and

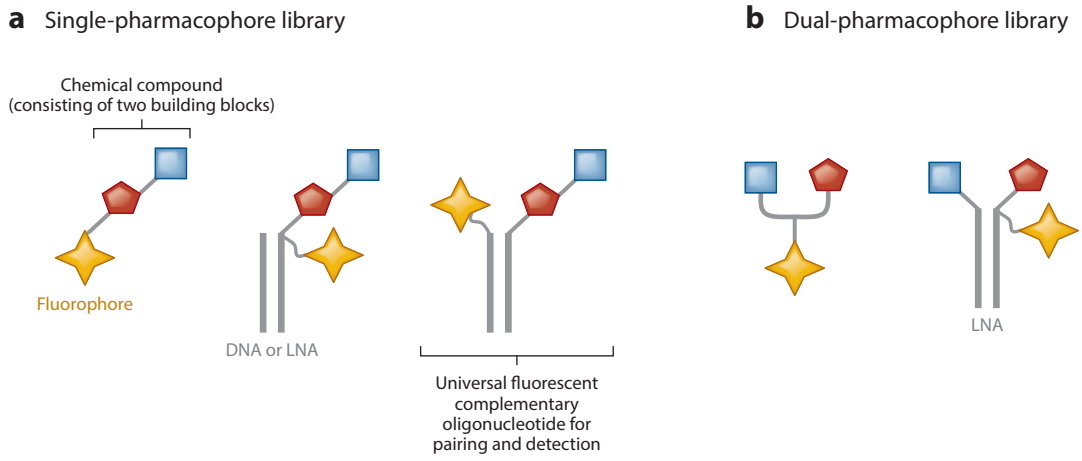


Figure 10

Methods for the fluorescent labeling of library members, enabling fluorescence polarization measurement procedures. For single-pharmacophore library members, hits can be resynthesized as fluorescently labeled compounds or on oligonucleotides [DNA or locked nucleic acid (LNA)], which carry the fluorophore on the same strand or on a complementary strand. Similarly, hits from dual-pharmacophore chemical libraries can be resynthesized as fluorescent molecules (with a suitable linker connecting the two building blocks) or on complementary oligonucleotides (one of which is fluorescently labeled).

versatile methodology based on the hybridization of oligonucleotide conjugates with complementary strands, labeled with a suitable fluorophore and/or additional chemical moieties (**Figure 10**). The technology benefits from the fact that on-DNA synthesis procedures are available from the previously performed library construction steps. Furthermore, oligonucleotides facilitate the purification of the conjugates and their analytical characterization by mass spectrometry while also contributing to the solubilization of lipophilic molecules.

THE CHEMISTRY OF LARGE NUMBERS

DNA-encoded chemical libraries are often referred to as "the chemistry of large numbers." What is meant by this is that most chemical procedures are about the reaction between molecules but usually one at a time. However, if one has a strategy to deconvolute reactions in a mixture, DNA encoding allows one to carry out vast numbers of reactions at once. Current encoded libraries can be used to find ligands for many protein targets but not (yet) for all (94). This raises the question of how large should these libraries be. A simple answer is that, when one builds diversity systems, the incorporation of more diversity is never a bad idea, so long as one has a robust strategy for the deconvolution of binding events. In our experience, when screening libraries with conventional affinity capture methodologies, it is convenient to ensure that individual molecules are present in a selection experiment with at least 10^5 copies each (64, 65, 95, 96). In practice, this prerequisite may constrain the size of encoded libraries that can be efficiently selected. In most cases, we aim at perturbing the biological world either by constructing a ligand that mimics the physiological binding event or by generating a new binding event such as an allosteric effector. Thus, it is useful to ask what nature uses as components of ligands, because it is these we are trying to duplicate. Actually, nature uses relatively few chemical interactions (i.e., hydrogen bonds, pi stacking, hydrophobic interactions, etc.) but can achieve enormous binding energy by using them in concert in a relatively anhydrous, confined space, such as the active site of an enzyme.

Some of the major reactions to generate DNA-encoded chemical libraries include amide bond formation and reductive amination, which can be used with various molecular scaffolds bearing reactive moieties similar to those used in nature. However, natural ligands are a product of evolution, through which the best chemical entities and their organization in space were selected over vast periods of time. The central concept is that nature took advantage of what was available or could be biologically constructed (such as by incorporating cofactors co-opted from the environment) but used selection over time to optimize the organization of the individual components of a ligand. In the case of DNA-encoded chemical libraries, the large numbers are the equivalent of evolutionary time, because one hopes that somewhere in the large number of molecules is one or a few that are at least as good as the one that was the product of evolution. There are, however, other significant directions being pursued in the generation of DNA-encoded chemical libraries. Importantly, the use of enzymes in the synthesis of organic ligands is only beginning to be explored (97). Additionally, organic chemists are inventing new reactions that are compatible with DNA and water solutions. This allows DNA-encoded chemical libraries to incorporate adducts that were not available to natural selection. These new adducts may yield components or scaffolds that are beyond the chemistry available to nature.

Thus, in the end the numbers question needs to be rephrased as a diversity question. As for numbers, one simply asks how many unique members there are in the system, whereas diversity concerns how different the members are from each other. To achieve both large numbers and maximum diversity, it would seem that one should link simple amide-based chemistries to more sophisticated chemical transformations. In the end, in an amalgamation of genetics and chemical synthesis, new exciting diversity systems will probably result from the sequential use of simple and complex organic transformations exploring a chemical space that may contain a solution to biological problems that cannot be addressed by other means. These diversity systems, and the DNA-encoded chemical libraries described in this review, are new to organic chemistry because they endow organic molecules with information capable of replication, thereby linking phenotype to genotype in single molecules.

DISCLOSURE STATEMENT

D.N. is a cofounder and shareholder of Philogen, a Swiss-Italian biotechnology group that, inter alia, works on DNA-encoded chemistry.

ACKNOWLEDGMENTS

D.N. acknowledges financial support from ETH Zürich, the Swiss National Science Foundation (CRSII2_160699 and 310030B_163479/1), the European Research Council Advanced Grant "Zauberkugel," and Philochem AG. R.A.L. acknowledges financial support from the JPB Foundation. We are grateful to Gerry Joyce, Ian Wilson, Raphael Franzini, and Jörg Scheuermann for reading the manuscript and providing useful insights. We also thank Nicholas Favalli, Gabriele Bassi, and Jörg Scheuermann for help in the preparation of **Figure 9**.

LITERATURE CITED

1. Erlanson DA, Fesik SW, Hubbard RE, Jahnke W, Jhoti H. 2016. Twenty years on: the impact of fragments on drug discovery. *Nat. Rev. Drug Discov.* 15:605–19
2. Folmer RH. 2016. Integrating biophysics with HTS-driven drug discovery projects. *Drug Discov. Today* 21:491–98

3. Nielsen TE, Schreiber SL. 2008. Towards the optimal screening collection: a synthesis strategy. *Angew. Chem. Int. Ed. Engl.* 47:48–56
4. Schreiber SL, Kotz JD, Li M, Aubé J, Austin CP, et al. 2015. Advancing biological understanding and therapeutics discovery with small-molecule probes. *Cell* 161:1252–65
5. Arrowsmith CH, Audia JE, Austin C, Baell J, Bennett J, et al. 2015. The promise and peril of chemical probes. *Nat. Chem. Biol.* 11:536–41
6. Dar AC, Shokat KM. 2011. The evolution of protein kinase inhibitors from antagonists to agonists of cellular signaling. *Annu. Rev. Biochem.* 80:769–95
7. Köhler G, Milstein C. 1975. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 256:495–97
8. Wilchek M, Bayer EA, Livnah O. 2006. Essentials of biorecognition: the (strept)avidin–biotin system as a model for protein–protein and protein–ligand interaction. *Immunol. Lett.* 103:27–32
9. Heinisch T, Ward TR. 2016. Artificial metalloenzymes based on the biotin-streptavidin technology: challenges and opportunities. *Acc. Chem. Res.* 49:1711–21
10. Paganelli G, Pervez S, Siccardi AG, Rowlinson G, Deleide G, et al. 1990. Intraperitoneal radio-localization of tumors pre-targeted by biotinylated monoclonal antibodies. *Int. J. Cancer* 45:1184–89
11. Sapsford KE, Algar WR, Berti L, Gemmill KB, Casey BJ, et al. 2013. Functionalizing nanoparticles with biological molecules: developing chemistries that facilitate nanotechnology. *Chem. Rev.* 113:1904–2074
12. Halai R, Cooper MA. 2012. Using label-free screening technology to improve efficiency in drug discovery. *Expert Opin. Drug Discov.* 7:123–31
13. Goodnow RA Jr. 2014. The changing feasibility and economics of chemical diversity exploration with DNA-encoded combinatorial approaches. In *A Handbook for DNA-Encoded Chemistry: Theory and Applications for Exploring Chemical Space and Drug Discovery*, ed. RA Goodnow Jr., pp. 417–26. Hoboken, NJ: Wiley
14. Stark JL, Powers R. 2012. Application of NMR and molecular docking in structure-based drug discovery. *Top. Curr. Chem.* 326:1–34
15. Smith GP. 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228:1315–17
16. Clackson T, Wells JA. 1994. In vitro selection from protein and peptide libraries. *Trends Biotechnol.* 12:173–84
17. McCafferty J, Griffiths AD, Winter G, Chiswell DJ. 1990. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* 348:552–54
18. Huse WD, Sastry L, Iverson SA, Kang AS, Alting-Mees M, et al. 1989. Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda. *Science* 246:1275–81
19. Marks JD, Hoogenboom HR, Bonnert TP, McCafferty J, Griffiths AD, et al. 1991. By-passing immunization. Human antibodies from V-gene libraries displayed on phage. *J. Mol. Biol.* 222:581–97
20. Kang AS, Barbas CF, Janda KD, Benkovic SJ, Lerner RA. 1991. Linkage of recognition and replication functions by assembling combinatorial antibody Fab libraries along phage surfaces. *PNAS* 88:4363–66
21. Winter G, Milstein C. 1991. Man-made antibodies. *Nature* 349:293–99
22. Jespers LS, Roberts A, Mahler SM, Winter G, Hoogenboom HR. 1994. Guiding the selection of human antibodies from phage display repertoires to a single epitope of an antigen. *Biotechnology* 12:899–903
23. Griffiths AD, Williams SC, Hartley O, Tomlinson IM, Waterhouse P, et al. 1994. Isolation of high affinity human antibodies directly from large synthetic repertoires. *EMBO J.* 13:3245–60
24. Gai SA, Wittrup KD. 2007. Yeast surface display for protein engineering and characterization. *Curr. Opin. Struct. Biol.* 17:467–73
25. Plückthun A. 2012. Ribosome display: a perspective. *Methods Mol. Biol.* 805:3–28
26. Brenner S, Lerner RA. 1992. Encoded combinatorial chemistry. *PNAS* 89:5381–83
27. Needels MC, Jones DG, Tate EH, Heinkel GL, Kochersperger LM, et al. 1993. Generation and screening of an oligonucleotide-encoded synthetic peptide library. *PNAS* 90:10700–4
28. Nielsen J, Brenner S, Janda KD. 1993. Synthetic methods for the implementation of encoded combinatorial chemistry. *J. Am. Chem. Soc.* 115:9812
29. Gartner ZJ, Tse BN, Grubina R, Doyon JB, Snyder TM, et al. 2004. DNA-templated organic synthesis and selection of a library of macrocycles. *Science* 305:1601–5

26. Earliest proposal of using DNA tags to encode chemical libraries.

27. Demonstration that DNA-encoded peptides of known specificity to a cognate antibody can be retrieved from a mixture of compounds.

28. Demonstration that a set of peptides can be coupled to different oligonucleotides, which do not interfere with peptide binding to cognate antibodies; retrieval of encoded peptides upon binding to cognate antibodies; and application of the DNA-encoding methodology to dendrimeric structures.

29. Earliest proposal and implementation of a DNA-templated chemical library.

30. Melkko S, Scheuermann J, Dumelin CE, Neri D. 2004. Encoded self-assembling chemical libraries. *Nat. Biotechnol.* 22:568–74

31. Halpin DR, Harbury PB. 2004. DNA display I. Sequence-encoded routing of DNA populations. *PLOS Biol.* 2:E173

32. Franzini RM, Neri D, Scheuermann J. 2014. DNA-encoded chemical libraries: advancing beyond conventional small-molecule libraries. *Acc. Chem. Res.* 47:1247–55

33. Kleiner RE, Dumelin CE, Liu DR. 2011. Small-molecule discovery from DNA-encoded chemical libraries. *Chem. Soc. Rev.* 40:5707–17

34. Zimmermann G, Neri D. 2016. DNA-encoded chemical libraries: foundations and applications in lead discovery. *Drug Discov. Today* 21:1828–34

35. Franzini RM, Randolph C. 2016. Chemical space of DNA-encoded libraries. *J. Med. Chem.* 59:6629–44

36. Clark MA. 2010. Selecting chemicals: the emerging utility of DNA-encoded libraries. *Curr. Opin. Chem. Biol.* 14:396–403

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

38. Shi B, Zhou Y, Huang Y, Zhang J, Li X. 2017. Recent advances on the encoding and selection methods of DNA-encoded chemical library. *Bioorg. Med. Chem. Lett.* 27:361–69

39. Keefe AD, Clark MA, Hupp CD, Litovchick A, Zhang Y. 2015. Chemical ligation methods for the tagging of DNA-encoded chemical libraries. *Curr. Opin. Chem. Biol.* 26:80–88

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

41. Goodnow RA Jr., Dumelin CE, Keefe AD. 2017. DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nat. Rev. Drug Discov.* 16:131–47

42. Blakskjaer P, Heitner T, Hansen NJ. 2015. Fidelity by design: Yoctoreactor and binder trap enrichment for small-molecule DNA-encoded libraries and drug discovery. *Curr. Opin. Chem. Biol.* 26:62–71

43. Clark MA, Acharya RA, Arico-Muendel CC, Belyanskaya SL, Benjamin DR. 2009. Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nat. Chem. Biol.* 5:647–54

44. Buller F, Mannocci L, Zhang Y, Dumelin CE, Scheuermann J, Neri D. 2008. Design and synthesis of a novel DNA-encoded chemical library using Diels-Alder cycloadditions. *Bioorg. Med. Chem. Lett.* 18:5926–31

45. Buller F, Zhang Y, Scheuermann J, Schäfer J, Bühlmann P, et al. 2009. Discovery of TNF inhibitors from a DNA-encoded chemical library based on Diels-Alder cycloaddition. *Chem. Biol.* 16:1075–86

46. Mannocci L, Zhang Y, Scheuermann J, Leimbacher M, De Bellis G, et al. 2008. High-throughput sequencing allows the identification of binding molecules isolated from DNA-encoded chemical libraries. *PNAS* 105:17670–75

47. Leimbacher M, Zhang Y, Mannocci L, Stravs M, Geppert T, et al. 2012. Discovery of small-molecule interleukin-2 inhibitors from a DNA-encoded chemical library. *Chemistry* 18:7729–37

48. Buller F, Steiner M, Frey K, Mircsof D, Scheuermann J, et al. 2011. Selection of carbonic anhydrase IX inhibitors from one million DNA-encoded compounds. *ACS Chem. Biol.* 6:336–44

49. Mannocci L, Melkko S, Buller F, Molnàr I, Bianké JP, et al. 2010. Isolation of potent and specific trypsin inhibitors from a DNA-encoded chemical library. *Bioconjug. Chem.* 21:1836–41

50. Bain JD, Switzer C. 1992. Regioselective ligation of oligoribonucleotides using DNA splints. *Nucleic Acids Res.* 20:4372

51. Wichert M, Krall N, Decurtins W, Franzini RM, Pretto F, et al. 2015. Dual-display of small molecules enables the discovery of ligand pairs and facilitates affinity maturation. *Nat. Chem.* 7:241–49

52. Kanan MW, Rozenman MM, Sakurai K, Snyder TM, Liu DR. 2004. Reaction discovery enabled by DNA-templated synthesis and in vitro selection. *Nature* 431:545–49

53. Kleiner RE, Dumelin CE, Tiu GC, Sakurai K, Liu DR. 2010. In vitro selection of a DNA-templated small-molecule library reveals a class of macrocyclic kinase inhibitors. *J. Am. Chem. Soc.* 132:11779–91

54. Li Y, Zhao P, Zhang M, Zhao X, Li X. 2013. Multistep DNA-templated synthesis using a universal template. *J. Am. Chem. Soc.* 135:17727–30

30. Earliest proposal and implementation of dual-pharmacophore DNA-encoded chemical libraries.

31. Earliest proposal of encoding by DNA routing.

43. One of the earliest examples of the synthesis and successful screening of a very large DNA-encoded chemical library.

46. Earliest publication of a DNA-recorded chemical library and the use of high-throughput DNA sequencing for library decoding.

51. Earliest publication of the use of high-throughput DNA sequencing for the decoding of dual-pharmacophore chemical libraries.

55. **Hansen MH, Blakskjaer P, Petersen LK, Hansen TH, Højfeldt JW, et al. 2009. A yoctoliter-scale DNA reactor for small-molecule evolution. *J. Am. Chem. Soc.* 131:1322–27**
56. Halpin DR, Harbury PB. 2004. DNA display II. Genetic manipulation of combinatorial chemistry libraries for small-molecule evolution. *PLOS Biol.* 2:E174
57. Halpin DR, Lee JA, Wrenn SJ, Harbury PB. 2004. DNA display III. Solid-phase organic synthesis on unprotected DNA. *PLOS Biol.* 2:E175
58. Wrenn SJ, Weisinger RM, Halpin DR, Harbury PB. 2007. Synthetic ligands discovered by in vitro selection. *J. Am. Chem. Soc.* 129:13137–43
59. Neri D, Melkko S. 2002. *Encoded self-assembling chemical libraries*. Patent No. WO03/076943
60. Gorska K, Huang KT, Chaloin O, Winssinger N. 2009. DNA-templated homo- and heterodimerization of peptide nucleic acid encoded oligosaccharides that mimick the carbohydrate epitope of HIV. *Angew. Chem. Int. Ed. Engl.* 48:7695–700
61. Dagauer JP, Zambaldo C, Ciobanu M, Morieux P, Barluenga S, et al. 2015. DNA display of fragment pairs as a tool for the discovery of novel biologically active small molecules. *Chem. Sci.* 6:739–44
62. Dumelin CE, Trüssel S, Buller F, Trachsel E, Bootz F, et al. 2008. A portable albumin binder from a DNA-encoded chemical library. *Angew. Chem. Int. Ed. Engl.* 47:3196–201
63. Decurtins W, Wichert M, Franzini RM, Buller F, Stravs MA, et al. 2016. Automated screening for small organic ligands using DNA-encoded chemical libraries. *Nat. Protoc.* 11:764–80
64. Li Y, Zimmermann G, Scheuermann J, Neri D. 2017. Quantitative PCR is a valuable tool to monitor the performance of DNA-encoded chemical library selections. *ChemBioChem* 18:848–52
65. Belyanskaya SL, Ding Y, Callahan JF, Lazaar AL, Israel DI. 2017. Discovering drugs with DNA-encoded library technology: from concept to clinic with an inhibitor of soluble epoxide hydrolase. *ChemBioChem* 18:837–42
66. Bao J, Krylova SM, Cherney LT, Hale RL, Belyanskaya SL, et al. 2015. Predicting electrophoretic mobility of protein-ligand complexes for ligands from DNA-encoded libraries of small molecules. *Anal. Chem.* 88:5498–506
67. McGregor LM, Gorin DJ, Dumelin CE, Liu DR. 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.* 132:15522–24
68. McGregor LM, Jain T, Liu DR. 2014. Identification of ligand-target pairs from combined libraries of small molecules and unpurified protein targets in cell lysates. *J. Am. Chem. Soc.* 136:3264–70
69. Denton KE, Krusemark CJ. 2016. Crosslinking of DNA-linked ligands to target proteins for enrichment from DNA-encoded libraries. *Med. Chem. Comm.* 7:2020–27
70. Shi B, Deng Y, Zhao P, Li X. 2017. Selecting a DNA-encoded chemical library against non-immobilized proteins using a "ligate-cross-link-purify" strategy. *Bioconjug. Chem.* 28:2293
71. Franzini RM, Ekblad T, Zhong N, Wichert M, Decurtins W, et al. 2015. Identification of structure-activity relationships from screening a structurally compact DNA-encoded chemical library. *Angew. Chem. Int. Ed. Engl.* 54:3927–31
72. Buller F, Steiner M, Scheuermann J, Mannocci L, Nissen I, et al. 2010. High-throughput sequencing for the identification of binding molecules from DNA-encoded chemical libraries. *Bioorg. Med. Chem. Lett.* 20:4188–92
73. Satz AL, Cai J, Chen Y, Goodnow R, Gruber F, et al. 2015. DNA compatible multistep synthesis and applications to DNA encoded libraries. *Bioconjug. Chem.* 26:1623–32
74. Li Y, Gabriele E, Samain F, Favalli N, Sladojevich F, et al. 2016. Optimized reaction conditions for amide bond formation in DNA-encoded combinatorial libraries. *ACS Comb. Sci.* 18:438–43
75. Franzini RM, Biendl S, Mikutis G, Samain F, Scheuermann J, et al. 2015. "Cap-and-catch" purification for enhancing the quality of libraries of DNA conjugates. *ACS Comb. Sci.* 17:393–98
76. Franzini RM, Nauer A, Scheuermann J, Neri D. 2015. Interrogating target-specificity by parallel screening of a DNA-encoded chemical library against closely related proteins. *Chem. Commun.* 51:8014–16
77. Samain F, Ekblad T, Mikutis G, Zhong N, Zimmermann M, et al. 2015. Tankyrase 1 inhibitors with drug-like properties identified by screening a DNA-encoded chemical library. *J. Med. Chem.* 58:5143–49
78. Melkko S, Zhang Y, Dumelin CE, Scheuermann J, Neri D. 2007. Isolation of high-affinity trypsin inhibitors from a DNA-encoded chemical library. *Angew. Chem. Int. Ed. Engl.* 46:4671–74

55. Earliest report on YoctoReactor encoding technology.

79. Wu Z, Graybill TL, Zeng X, Platchek M, Zhang J, 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.* 17:722–31
80. Ding Y, O'Keefe H, DeLorey JL, Israel DI, Messer JA, et al. 2015. Discovery of potent and selective inhibitors for ADAMTS-4 through DNA-encoded library technology (ELT). *ACS Med. Chem. Lett.* 6:888–93
81. Yang H, Medeiros PF, Raha K, Elkins P, Lind KE, et al. 2015. Discovery of a potent class of PI3Kα inhibitors with unique binding mode via encoded library technology (ELT). *ACS Med. Chem. Lett.* 6:531–36
82. Malone ML, Paegel BM. 2016. What is a "DNA-compatible" reaction? *ACS Comb. Sci.* 18:182–87
83. Harris PA, King BW, Bandyopadhyay D, Berger SB, Campobasso N, et al. 2016. DNA-encoded library screening identifies benzo[*b*][1,4]oxazepin-4-ones as highly potent and monoselective receptor interacting protein 1 kinase inhibitors. *J. Med. Chem.* 59:2163–78
84. Petersen LK, Blakskjaer P, Chaikuad A, Christensen AB, Dietvorst J, et al. 2016. Novel p38α MAP kinase inhibitors identified from yoctoReactor DNA-encoded small molecule library. *Med. Chem. Comm.* 7:1332–39
85. Litovchick A, Dumelin CE, Habeshian S, Gikunju D, Gujé MA, 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
86. Maianti JP, McFedries A, Foda ZH, Kleiner RE, Du XQ, et al. 2014. Anti-diabetic activity of insulin-degrading enzyme inhibitors mediated by multiple hormones. *Nature* 511:94–98
87. Seigal BA, Connors WH, Fraley A, Borzilleri RM, Carter PH, et al. 2015. 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.* 58:2855–61
88. Zambaldo C, Daguer JP, Saarbach J, Barluenga S, Winssinger N. 2016. Screening for covalent inhibitors using DNA-display of small molecule libraries functionalized with cysteine reactive moieties. *Med. Chem. Comm.* 7:1340–51
89. Daguer JP, Zambaldo C, Abegg D, Barluenga S, Tallant C, et al. 2015. Identification of covalent bromodomain binders through DNA display of small molecules. *Angew. Chem. Int. Ed. Engl.* 54:6057–61
90. Zimmermann G, Rieder U, Bajic D, Vanetti S, Chaikuad A, et al. 2017. A specific and covalent JNK-1 ligand selected from an encoded self-assembling chemical library. *Chemistry* 23:8152–55
91. Cuozzo JW, Centrella PA, Gikunju D, Habeshian S, Hupp CD, 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* 18:864–71
92. Chan AI, McGregor LM, Jain T, Liu DR. 2017. Discovery of a covalent kinase inhibitor from a DNA-encoded small-molecule library x protein selection. *J. Am. Chem. Soc.* 139:10192–95
93. Zimmermann G, Li Y, Rieder U, Mattarella M, Neri D, et al. 2017. Hit-validation methodologies for ligands isolated from DNA-encoded chemical libraries. *ChemBioChem* 18:853–57
94. Machutta CA, Kollmann CS, Lind KE, Bai X, Chan PF, et al. 2017. Prioritizing multiple therapeutic targets in parallel using automated DNA-encoded library screening. *Nat. Commun.* 8:16081
95. Satz AL, Hochstrasser R, Petersen AC. 2017. Analysis of current DNA encoded chemical library screening data indicates higher false negative rates for numerically larger libraries. *ACS Comb. Sci.* 19:234
96. Li Y, De Luca R, Cazzamalli S, Pretto F, Bajic D, et al. 2018. Versatile protein recognition by the encoded display of multiple chemical elements on a constant macrocyclic scaffold. *Nat. Chem.* 10:441–48
97. Thomas B, Lu X, Birmingham WR, Huang K, Both P, et al. 2017. Applications of biocatalysis to on-DNA carbohydrate library synthesis. *ChemBioChem* 18:858