Published in final edited form as:
Drug Discov Today. 2016 November ; 21(11): 1828–1834. doi:10.1016/j.drudis.2016.07.013.

DNA-encoded chemical libraries: foundations and applications in lead discovery

Gunther Zimmermann and **Dario Neri**
Department of Chemistry and Applied Biosciences, Swiss Federal Institute of Technology (ETH Zürich), Vladimir Prelog Weg 1-5/10, CH-8093 Zürich, Switzerland

Abstract

DNA-encoded chemical libraries have emerged as a powerful means for hit identification in the pharmaceutical industry and in academia. Similar to biological display techniques (such as phage display technology), DNA-encoded chemical libraries contain a link between the displayed chemical building block and an amplifiable genetic “barcode” on DNA. Using routine procedures, libraries containing millions to billions of compounds can be easily produced within a few weeks. The resulting compound libraries are screened in a single test tube against proteins of pharmaceutical interest and hits can be identified by PCR amplification of DNA-barcodes and subsequent high-throughput sequencing. Various types of DNA-encoded chemical libraries can be considered for practical applications. For example, researchers may choose between libraries containing two or three sets of building blocks. The latter approach enables the synthesis of larger libraries, but typically yields hits, which are bigger than 500 Dalton. Moreover, it is possible to display pairs of molecules on complementary DNA strands, thus enabling the combinatorial assembly of library members.

Keywords

DNA-encoded chemical libraries; hit identification; affinity selections

Introduction

The identification of binding molecules to protein targets of pharmaceutical interest is a central problem in drug discovery. In most cases, initial hits are discovered by high-throughput screening of collections of chemical compounds (“chemical libraries). Because of limitations in library size and in logistics, only few companies worldwide can afford to screen more than 1 million compounds, in campaigns, which may last several weeks. Primary hits typically need to be optimized by medicinal chemistry efforts, before they can be considered for industrial applications. Importantly, conventional screening procedures often fail to provide hits against “difficult” target proteins, such as those, which lack hydrophobic cavities or those involved in protein-protein interactions. On the other hand,

Corresponding author: Dario Neri, dario.neri@pharma.ethz.ch.

Conflicts of interest
Dario Neri is a co-founder and shareholder of Philochem AG (Otelfingen, Switzerland)

Europe PMC Funders Author Manuscripts

biological selection methods (such as phage display technology) facilitate the construction and screening of libraries containing billions of polypeptides, as individual library members capable of protein binding ("phenotype") are directly connected to the corresponding genetic information ("genotype). Remarkably, high affinity antibodies can be directly isolated from large combinatorial libraries and used for pharmaceutical applications.[1,2] Phage display technology and similar methodologies are applicable only for the selection of their biosynthetic products (i.e., polypeptides) and cannot be used for the discovery of small organic ligands.

DNA-encoded chemical libraries are collections of organic molecules, individually coupled to distinctive DNA fragments. The DNA fragments do not drive the biosynthesis of the corresponding organic molecules, but merely serve as amplifiable barcodes, allowing the identification and relative quantification of individual molecules in complex mixtures. DNA-encoded chemical library synthesis can be conducted with numerous, structurally diverse chemical building blocks. A range of chemical reactions can be used for library construction, including reductive amination, amide bond formation, Suzuki reactions and nucleophilic aromatic substitutions.[3–5] The concept of DNA-encoded chemical libraries was first introduced by Brenner and Lerner, who postulated that it should be possible to simultaneously synthesize peptides and DNA-tags on beads, thus generating collections of encoded compounds.[6] Subsequently, it was shown that encoded chemical libraries could be conveniently synthesized without beads, which leads to larger library sizes and facilitates screening procedures.[7–9]

Encoding strategies for the construction of DNA-encoded chemical libraries

At present, most encoded chemical libraries are assembled by DNA-recorded synthesis formats in solution phase. In this approach, DNA fragments are solely employed to "record" the identity of individual building blocks, used for synthetic transformations [Figure 1]. Molecules can be assembled from a limited number of building blocks, using alternating steps of chemical synthesis and of DNA encoding, according to "split-and-pool" procedures. Two different enzymatic reactions are frequently used to record the synthetic history of library construction: ligation[5] and polymerase-catalyzed fill-in reactions [Figure 1].[10,11] Recently, also non-enzymatic encoding reactions, such as those based on click chemistry assembly of oligonucleotides, have been described. [12,13]

DNA-templated chemical library synthesis represents an alternative approach for the preparation of encoded compound collections.[7] In a typical setup, oligonucleotides are used to transfer chemical building blocks to a nascent molecule, attached at the extremity of a complementary DNA strand [Figure 1]. The spatial proximity of the two oligonucleotides results in an increased effective molarity and thus facilitates the execution of bimolecular chemical reactions.[14] However, DNA-templated chemistry requires the preparation of numerous DNA conjugates, which incorporate the building blocks to be used for library synthesis. Moreover, the fidelity of encoding could be compromised by a mismatch in the DNA annealing steps.[15]

Using DNA-templated library synthesis, Liu and coworkers have generated compound collections containing thousands of macrocycles and identified binders against a number of target proteins including a macrocyclic binder of Src kinase with an affinity of 680 nM.[16]

In a different but conceptually related approach, Hansen and coworkers employed self-assembling DNA junctions to generate a “yoctoliter” reactor for templated chemical transformations.[17]

Display formats of DNA-encoded chemical libraries

Two different types of DNA-encoded chemical display formats can be considered. Single pharmacophore libraries result from the coupling of one DNA fragment to a chemical building block, whereas dual pharmacophore libraries have pairs of chemical building blocks attached to the extremities of complementary DNA strands [Figure 2]. [18]

Single pharmacophore DNA-encoded chemical libraries

Using the encoding strategies presented in Figure 1, scientists from different groups have generated single pharmacophore libraries ranging from thousands to billions of compounds. Success stories have even been reported for carefully designed, smaller libraries containing only thousands of compounds. Using such focused libraries, our group has reported the *de novo* discovery of micromolar binders against Bcl-xL,[19] Interleukin-2 [20] and TNF. [21] In 2015, a two building block library, containing 103200 compounds, allowed the isolation of binders against a number of proteins, including prostate-specific membrane antigen and tankyrase-1.[10] Scientists at GlaxoSmithKline have described a number of very large single pharmacophore DNA-encoded libraries using 3 or 4 sets of building blocks. Successful screening campaigns were launched against a number of relevant drug targets and resulted in nanomolar binders against Aurora B kinase, p38MAPK,[5] ADAMTS-4/5,[22,23] SIRT1/2/3, [24] PAD4[25] and RIP1[26]. Scientists at X-Chem have reported the generation of a 334-million single pharmacophore library that provided a low nanomolar binder against sEH.[12]

Dual pharmacophore DNA-encoded chemical libraries

The first dual display chemical library was presented by our group in 2004 and utilizes the display of fragment pairs on each strand of the DNA duplex. These libraries are constructed by the hybridization of two sets of complementary oligonucleotides, which display individual chemical moieties at their extremity. For this reason, the technology is also referred to as “encoded self-assembling chemical libraries” (ESAC).[8] The first libraries were screened using microarrays for decoding purposes and facilitated the identification of novel binders against trypsin [27] and MMP3.[28] Recently, we have reported a novel encoding strategy that allows library decoding by high-throughput sequencing.[29] In this methodology, a single oligonucleotide, containing an abasic site, is coupled to various chemical building blocks and these reactions are subsequently encoded by ligation with suitable oligonucleotide “barcodes” [Figure 3, left]

The resulting sublibrary 1 can then be hybridized to an encoded sublibrary 2, followed by a Klenow polymerization step, which transfers the code information to the complementary DNA strands, making the system suitable for high-throughput sequencing decoding.

In a conceptually similar dual pharmacophore display approach, Winssinger and coworkers employed peptide nucleic acid (PNA) as the coding entity. This procedure led to the identification of hits against HSP70/90[30] and PTP1B.[31] The authors also showed that small, dual pharmacophore libraries of covalent inhibitors may represent an excellent starting point for the discovery of protein binders. [32,33]

Screening and hit validation

DNA-encoded libraries can be screened against target proteins and preferential binders can be identified according to a general procedure, depicted in Figure 4. In a typical screening experiment, a protein of interest is immobilized on a solid support and subsequently incubated with the DNA-encoded library. After several washing steps, binding molecules attached to DNA will remain bound to the target protein, while non-binders should be washed away. Preferential binders can now be eluted from the solid phase by heat-induced protein denaturation and identified by PCR amplification of the DNA-barcode, followed by high-throughput DNA sequencing. Due to the PCR amplification process, even very low copy numbers of the bound DNA-small molecule conjugates can be efficiently detected. As a consequence, DNA-encoding can be used for the synthesis and screening of combinatorial libraries, containing very large numbers of compounds, individually present at very low concentrations. Furthermore, libraries can be stored as frozen aliquots and can be interrogated against numerous protein targets, using robotic procedures.

After decoding, hit compounds are validated by re-synthesis procedures and affinity measurements. It is particularly convenient to resynthesize binders as fluorophore conjugates, thus enabling affinity measurements in solution, using fluorescence polarization (FP) as readout. Conveniently, the fluorophore can be replaced by short fluorescently labeled DNA-tags, which may facilitate the resynthesis procedure.[29]

Selection procedures

The most common procedure for affinity-based selections involves the use of protein immobilized on magnetic beads.[34] Typically, the protein is immobilized prior to the selection process and the incubation and washing steps can be run in an automated fashion. [34] Similarly, affinity resins may also be employed as capture reagents in selection procedures.[12,35] Conveniently, prefabricated pipette tips with different resin bed sizes are commercially available [Figure 5].

Alternative procedures can be considered, in order to identify DNA-tagged molecules, capable of interacting with target proteins of interest, in large mixtures. The formation of specific complexes can be detected and separated, using native acrylamide gel electrophoresis[36] or capillary electrophoresis[37] for affinity selections [Figure 5]. All these methods involve either a solid support (e.g., magnetic beads, resins) or the use of non-native protein environment (e.g., electrophoresis). Ideally, it would be desirable to carry out

Europe PMC Funders Author Manuscripts

selections in more physiological conditions, such as buffered aqueous solution, cell lysates or even on cells. The groups of Liu and Li have recently reported methods for the selection of DNA-encoded chemical libraries, using template PCR reactions or photo-crosslinking procedures, respectively.[38,39] Researchers at GSK have shown that selections can also be conducted directly against receptors, expressed on the surface of cells.[35]

Conclusions, challenges and future prospects

DNA-encoded chemical libraries have emerged as a powerful tool for the discovery of hits of pharmaceutical interest. Compound collections of unprecedented size can be synthesized with limited efforts and costs, without the need for complex logistics. Libraries can be stored as frozen aliquots for years and multiple simultaneous selections can be performed robotically, thus facilitating comparative analysis of the results.

The synthesis of libraries containing billions of compounds is now routinely possible. However, when 3 or 4 sets of building blocks are used for library assembly, the resulting compounds may not be compatible with Lipinski rules-of-5.[40] We anticipate that various types of libraries may be considered, depending on the specific pharmaceutical requirements (e.g., isolation of small hits with high ligand efficiency vs. identification of binders against large patches on the target protein of interest).

While most libraries described so far made use of a limited set of chemical reactions (most frequently, amide bond formation), the development of DNA-compatible synthetic procedures will certainly expand the scope of the methodology. The availability of thousands of building blocks at reasonably low cost and the introduction of novel library designs (e.g., the use of suitable scaffolds for the directional display of chemical moieties) will facilitate the expansion of the methodology. DNA-encoded chemical libraries facilitate hit finding in different settings including academia, since compound libraries in the million-member range are getting readily affordable. This will not only help academic drug discovery, but also aid in the discovery of novel tool compounds for understanding basic biological processes.

Library purity will continue to represent an important prerequisite for successful selections. The use of high-conversion reactions, the purification of intermediates in library synthesis or the application of cap-and-catch procedures[41] are likely to contribute to the performance of the technology.

Acknowledgements

Financial contributions from ETH Zürich, Swiss National Science Foundation, Deutsche Akademie der Naturforscher Leopoldina and an ERC Advanced Grant (ZAUBERKUGEL) are gratefully acknowledged.

References

1. McCafferty J, et al. Phage antibodies: filamentous phage displaying antibody variable domains. Nature (London). 1990; 348(6301):552–554. [PubMed: 2247164]
2. Heinis C, et al. Phage-encoded combinatorial chemical libraries based on bicyclic peptides. Nat Chem Biol. 2009; 5(7):502–507. [PubMed: 19483697]

3. Franzini RM, et al. Systematic evaluation and optimization of modification reactions of oligonucleotides with amines and carboxylic acids for the synthesis of DNA-encoded chemical libraries. Bioconjug Chem. 2014; 25(8):1453–1461. [PubMed: 25061844]
4. Ding Y, Clark MA. Robust Suzuki-Miyaura cross-coupling on DNA-linked substrates. ACS Comb Sci. 2015; 17(1):1–4. [PubMed: 25459065]
5. Clark MA, et al. Design, synthesis and selection of DNA-encoded small-molecule libraries. Nat Chem Biol. 2009; 5(9):647–654. [PubMed: 19648931]
6. Brenner S, Lerner RA. Encoded combinatorial chemistry. Proc Natl Acad Sci U S A. 1992; 89(12): 5381–5383. [PubMed: 1608946]
7. Gartner ZJ, et al. DNA-Templated Organic Synthesis and Selection of a Library of Macrocycles. Science (Washington, DC, U S). 2004; 305(5690):1601–1605.
8. Melkko S, et al. Encoded self-assembling chemical libraries. Nat Biotechnol. 2004; 22(5):568–574. [PubMed: 15097996]
9. Halpin DR, Harbury PB. DNA display. II. Genetic manipulation of combinatorial chemistry libraries for small-molecule evolution. PLoS Biol. 2004; 2(7):1022–1030.
10. Franzini RM, et al. Identification of structure-activity relationships from screening a structurally compact DNA-encoded chemical library. Angew Chem Int Ed Engl. 2015; 54(13):3927–3931. [PubMed: 25650139]
11. Samain F, et al. Tankyrase 1 Inhibitors with Drug-like Properties Identified by Screening a DNA-Encoded Chemical Library. J Med Chem. 2015; 58(12):5143–5149. [PubMed: 26061013]
12. Litovchick A, et al. Encoded Library Synthesis Using Chemical Ligation and the Discovery of sEH Inhibitors from a 334-Million Member Library. Sci Rep. 2015; 5:10916. [PubMed: 26061191]
13. Keefe AD, et al. Chemical ligation methods for the tagging of DNA-encoded chemical libraries. Curr Opin Chem Biol. 2015; 26:80–88. [PubMed: 25756406]
14. Gartner ZJ, et al. Expanding the reaction scope of DNA-templated synthesis. Angew Chem, Int Ed. 2002; 41(10):1796–1800.
15. Li Y, et al. Multistep DNA-templated synthesis using a universal template. J Am Chem Soc. 2013; 135(47):17727–17730. [PubMed: 24229415]
16. Tse BN, et al. Translation of DNA into a Library of 13 000 Synthetic Small-Molecule Macrocycles Suitable for in Vitro Selection. J Am Chem Soc. 2008; 130(46):15611–15626. [PubMed: 18956864]
17. Hansen MH, et al. A Yoctoliter-Scale DNA Reactor for Small-Molecule Evolution. J Am Chem Soc. 2009; 131(3):1322–1327. [PubMed: 19123795]
18. Franzini RM, Randolph C. Chemical Space of DNA-Encoded Libraries. J Med Chem. 2016
19. Melkko S, et al. Isolation of a Small-Molecule Inhibitor of the Antiapoptotic Protein Bcl-xL from a DNA-Encoded Chemical Library. ChemMedChem. 2010; 5(4):584–590. [PubMed: 20229565]
20. Leimbacher M, et al. Discovery of small-molecule interleukin-2 inhibitors from a DNA-encoded chemical library. Chemistry. 2012; 18(25):7729–7737. [PubMed: 22588840]
21. Buller F, et al. Discovery of TNF inhibitors from a DNA-encoded chemical library based on diels-alder cycloaddition. Chem Biol. 2009; 16(10):1075–1086. [PubMed: 19875081]
22. Deng H, et al. Discovery of Highly Potent and Selective Small Molecule ADAMTS-5 Inhibitors That Inhibit Human Cartilage Degradation via Encoded Library Technology (ELT). J Med Chem. 2012; 55(16):7061–7079. [PubMed: 22891645]
23. Ding Y, et al. Discovery of Potent and Selective Inhibitors for ADAMTS-4 through DNA-Encoded Library Technology (ELT). ACS Med Chem Lett. 2015; 6(8):888–893. [PubMed: 26288689]
24. Disch JS, et al. Discovery of Thieno[3,2-d]pyrimidine-6-carboxamides as Potent Inhibitors of SIRT1, SIRT2, and SIRT3. J Med Chem. 2013; 56(9):3666–3679. [PubMed: 23570514]
25. Lewis HD, et al. Inhibition of PAD4 activity is sufficient to disrupt mouse and human NET formation. Nat Chem Biol. 2015; 11(3):189–191. [PubMed: 25622091]
26. Harris PA, et al. 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. 2016; 59(5):2163–2178. [PubMed: 26854747]

27. Melkko S, et al. Isolation of high-affinity trypsin inhibitors from a DNA-encoded chemical library. Angew Chem, Int Ed. 2007; 46(25):4671–4674.
28. Scheuermann J, et al. DNA-encoded chemical libraries for the discovery of MMP-3 inhibitors. Bioconjugate Chem. 2008; 19(3):778–785.
29. Wichert M, et al. Dual-display of small molecules enables the discovery of ligand pairs and facilitates affinity maturation. Nat Chem. 2015; 7(3):241–249. [PubMed: 25698334]
30. Daguer JP, et al. DNA display of fragment pairs as a tool for the discovery of novel biologically active small molecules. Chem Sci. 2015; 6(1):739–744.
31. Barluenga S, et al. Novel PTP1B inhibitors identified by DNA display of fragment pairs. Bioorg Med Chem Lett. 2016; 26(3):1080–1085. [PubMed: 26691757]
32. Debaene F, et al. Expanding the scope of PNA-encoded libraries: divergent synthesis of libraries targeting cysteine, serine and metallo-proteases as well as tyrosine phosphatases. Tetrahedron. 2007; 63(28):6577–6586.
33. Daguer J-P. Identification of covalent bromodomain binders through DNA display of small molecules. Angew Chem, Int Ed. 2015; 54(20):6057–6061.
34. Decurtins W, et al. Automated screening for small organic ligands using DNA-encoded chemical libraries. Nat Protoc. 2016; 11(4):764–780. [PubMed: 26985574]
35. Arico-Muendel C, et al. Encoded library technology screening of hepatitis C virus NS4B yields a small-molecule compound series with in vitro replicon activity. Antimicrob Agents Chemother. 2015; 59(6):3450–3459. [PubMed: 25824229]
36. Melkko S, et al. On the Magnitude of the Chelate Effect for the Recognition of Proteins by Pharmacophores Scaffolded by Self-Assembling Oligonucleotides. Chemistry & Biology. 2006; 13(2):225–231. [PubMed: 16492570]
37. Bao J, et al. Prediction of protein-DNA complex mobility in gel-free capillary electrophoresis. Anal Chem. 2015; 87(4):2474–2479. [PubMed: 25582319]
38. McGregor LM, et al. 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. 2010; 132(44):15522–15524. [PubMed: 20949943]
39. Zhao P, et al. Selection of DNA-encoded small molecule libraries against unmodified and non-immobilized protein targets. Angew Chem Int Ed Engl. 2014; 53(38):10056–10059. [PubMed: 25044298]
40. Lipinski CA, et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Delivery Rev. 2001; 46(1-3): 3–26.
41. Franzini RM, et al. “Cap-and-Catch” Purification for Enhancing the Quality of Libraries of DNA Conjugates. ACS Comb Sci. 2015; 17(7):393–398. [PubMed: 26083096]

Teaser: DNA-encoded chemical libraries have emerged as a robust tool for the discovery of novel binders against target proteins since libraries of unprecedented size can be synthesized with limited efforts and costs.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Highlights

- DNA-encoded chemical libraries: powerful means for hit identification in pharma
- Link between the chemical building block and an amplifiable genetic “barcode”
- Production of millions of compounds within a few weeks
- Screening of libraries in a single test tube against proteins of pharmaceutical interest
- Numerous protein binders have thus far been identified by means of this technique

Europe PMC Funders Author Manuscripts

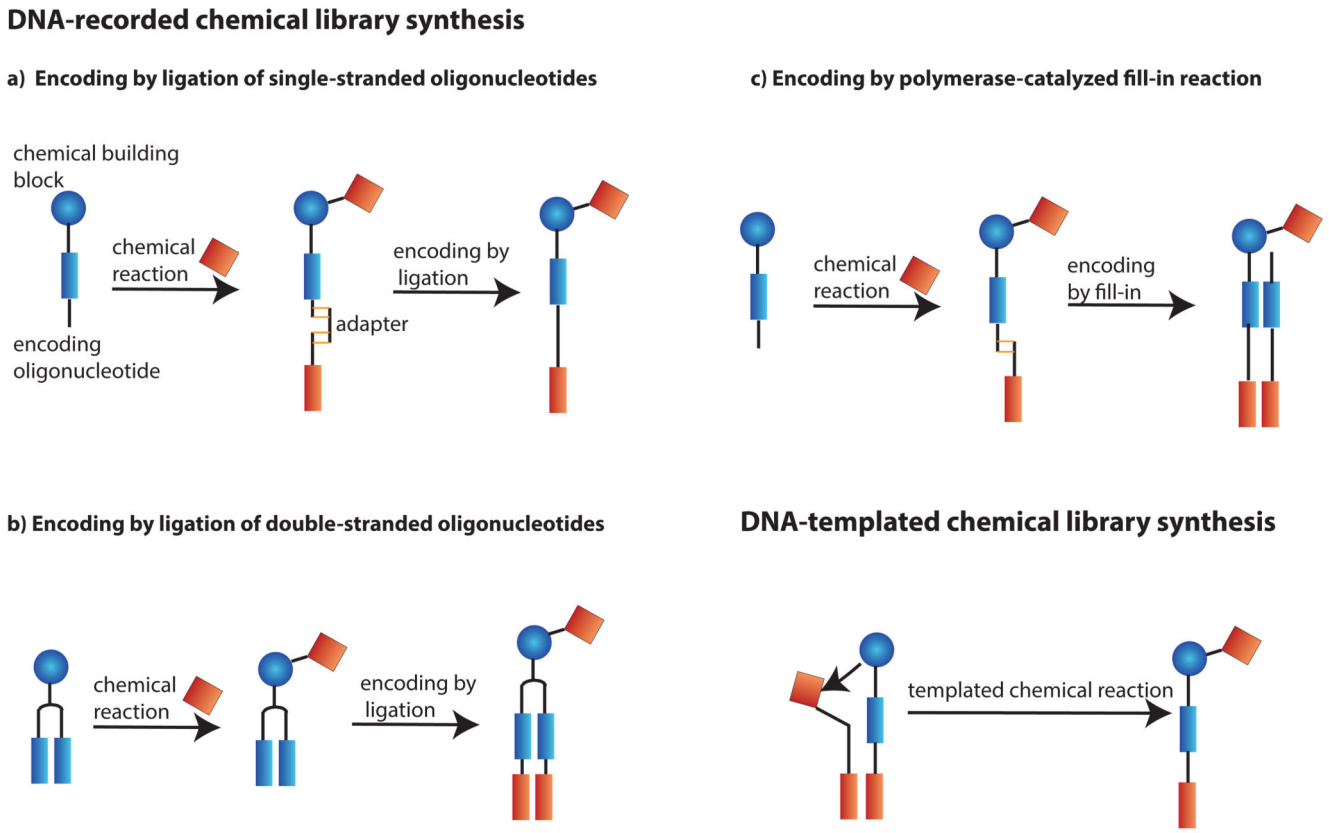

Figure 1.

Encoding strategies for the construction of DNA-encoded chemical libraries. Two synthetic strategies are employed: DNA-recorded chemical library synthesis and DNA-templated library synthesis. In DNA-recorded chemical library synthesis encoding may achieved by ligation using a) single, b) double stranded DNA or c) polymerase-catalyzed fill-in reactions.

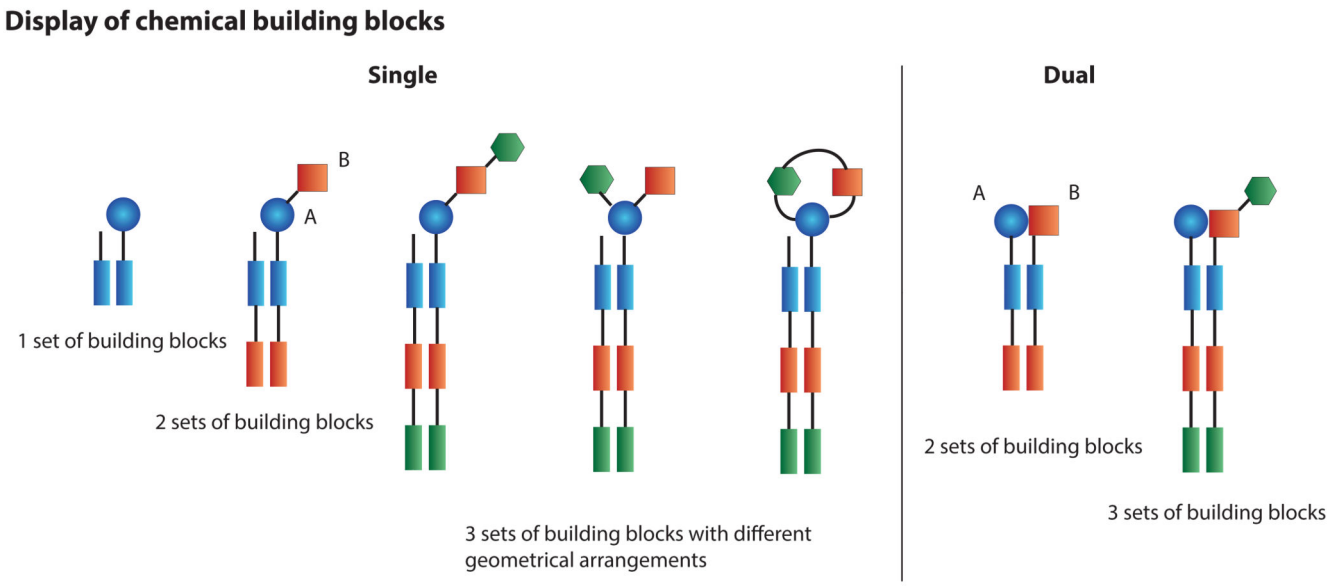

Figure 2.

Different library formats employed for DNA-encoded compound collections. So-called single-pharmacophore libraries (left) employ only one DNA-strand for presenting chemical building blocks. Single pharmacophore libraries are synthesized in a split-and-pool approach resulting in combinatorial libraries of A x B (in case of a two building block library). Dual pharmacophore libraries (right) are assembled by hybridization of individually synthesized sublibraries A and B (in case of a two building block library).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

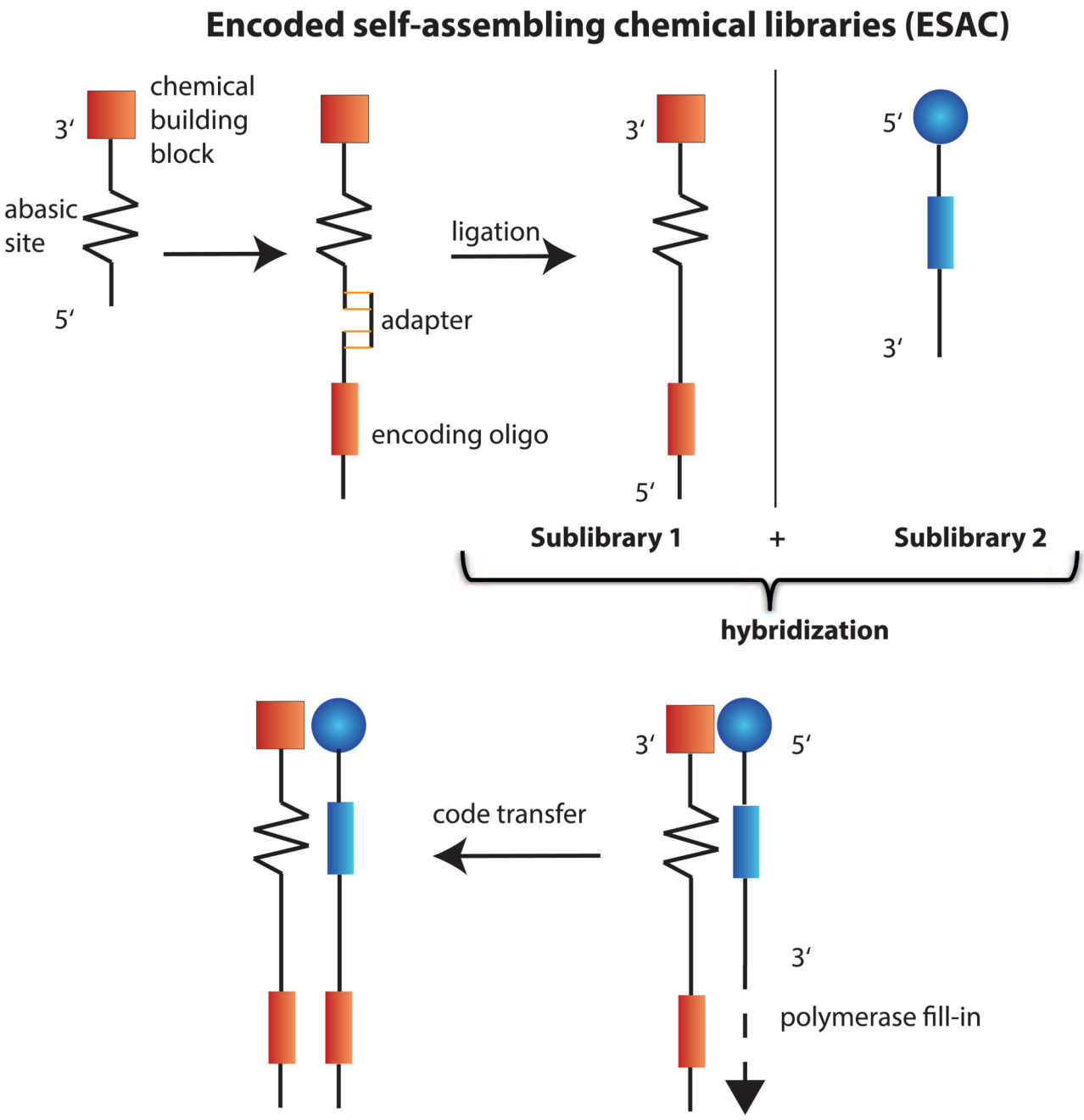

Figure 3.

Dual display, DNA-encoded compound collections employing the “Encoded self-assembling chemical libraries” (ESAC) approach. Sublibrary 1 and 2 are individually synthesized and purified. Sublibrary 1 is generated by coupling a chemical building block to an oligonucleotide which contains an abasic site. Subsequently, enzyme-catalyzed ligation delivers a stable link between the chemical building block and the encoding nucleotide sequence. Sublibrary 2 is synthesized by coupling of chemical building blocks to oligonucleotides bearing a short nucleotide coding sequence. The final combinatorial library

is assembled by hybridization and the coding information is transferred to one strand, making the system suitable for decoding by high-throughput sequencing.

Europe PMC Funders Author Manuscripts

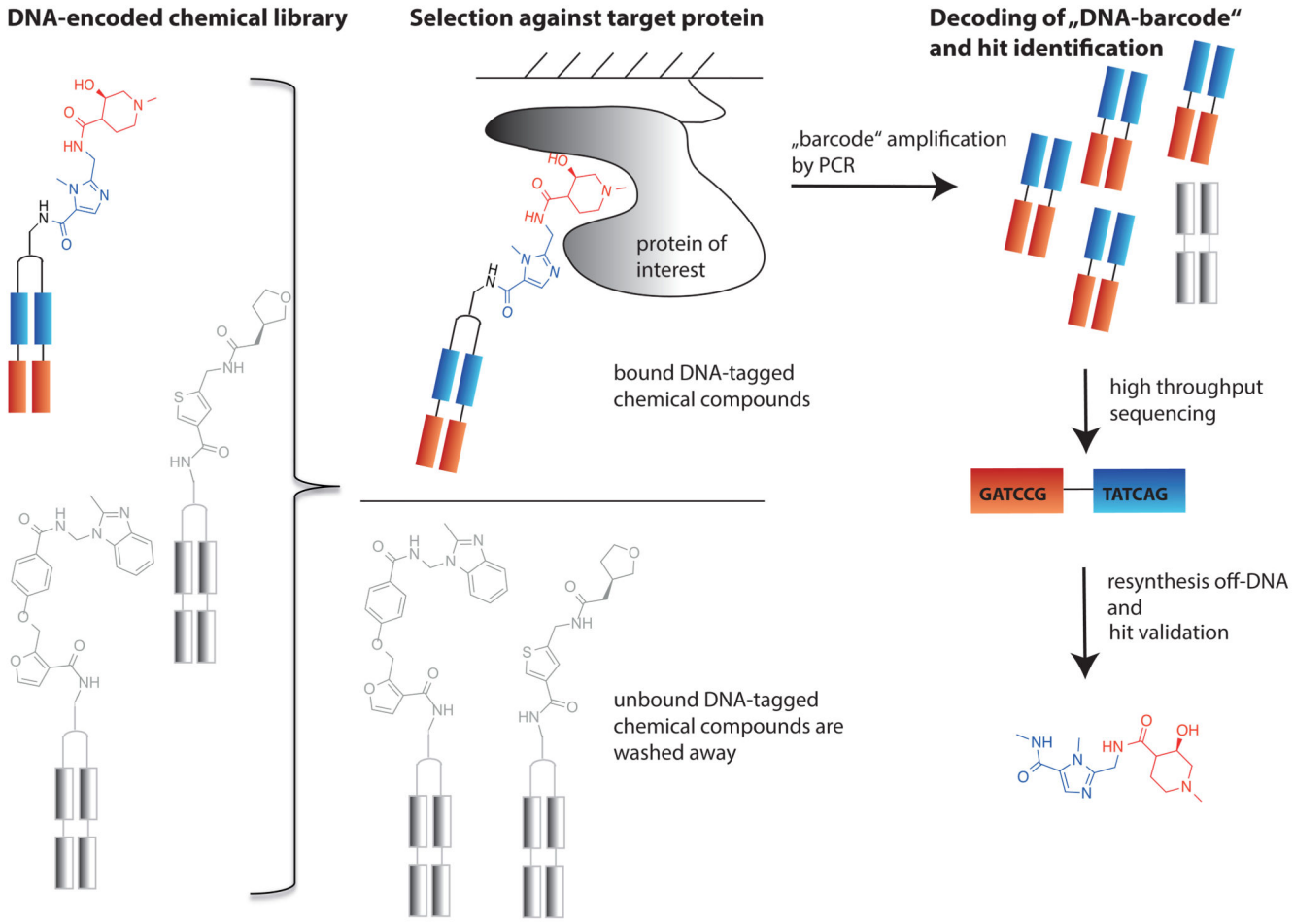

Figure 4.

Flowchart for hit identification employing DNA-encoded chemical libraries. Libraries are typically incubated with an immobilized target protein of interest and subsequently washed to select binders of the protein. Bound DNA-tagged molecules may then be eluted by heat-denaturation of the target protein. Coding oligonucleotides are then amplified by PCR and submitted for high-throughput sequencing. Identified hits are resynthesized without the DNA tag and validated by biophysical assays such as fluorescence polarization.

Europe PMC Funders Author Manuscripts

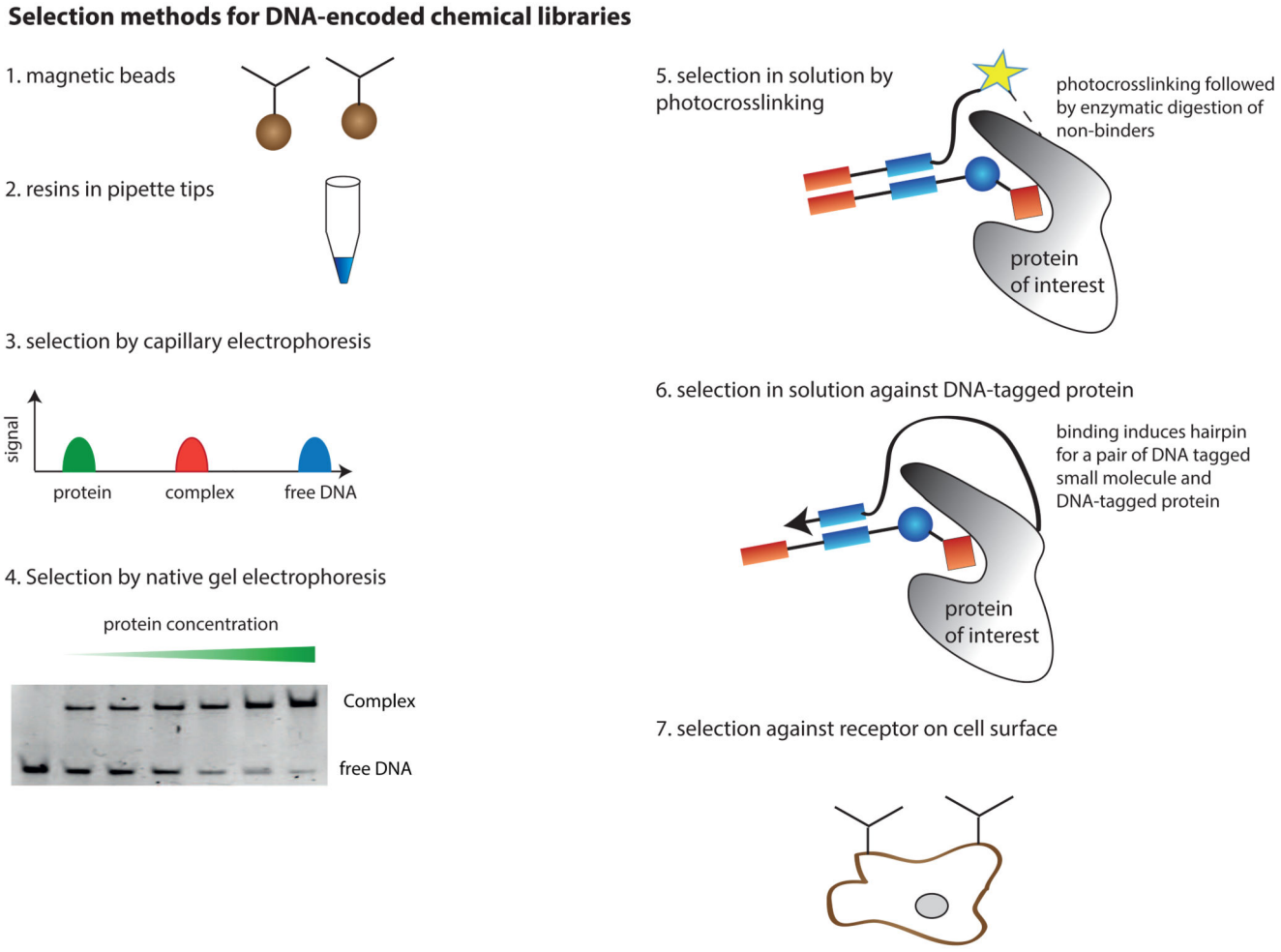

Figure 5.

Different selection methods for DNA-encoded chemical libraries. Magnetic beads and resins in pipette tips (entries 1 and 2) are routinely used for the selection against target proteins. Electrophoretic separation techniques (entries 3 and 4) have only been used for model selections. Similar model affinity enrichments have been demonstrated using in-solution procedures (entries 5 and 6), whereas novel hits have been identified by directly panning DNA-encoded libraries against receptors on the cell surface (entries 7).