Scuola Universitaria Superiore IUSS Pavia

Development and optimization of versatile screening methodologies in the context of DNA-encoded chemical libraries

A Thesis Submitted in Partial Fulfilment of the Requirements
for the Degree of Doctor of Philosophy in

BIOMOLECULAR SCIENCES AND BIOTECHNOLOGY

by

Alessandro Sannino

November 2020

Scuola Universitaria Superiore IUSS Pavia

Development and optimization of versatile screening methodologies in the context of DNA-encoded chemical libraries

A Thesis Submitted in Partial Fulfilment of the Requirements

for the Degree of Doctor of Philosophy in

BIOMOLECULAR SCIENCES AND BIOTECHNOLOGY

by

Alessandro Sannino

Supervisor: Prof. Dr. Federico Forneris, Department of Biology and Biotechnology "L. Spallanzani", University of Pavia

Co-supervisors: Dr. Florent Samain, Philochem AG;
Prof. Dr. Dario Neri, Institute of Pharmaceutical Sciencies, ETH Zürich / Philochem AG

November 2020

SUMMARY

For many years, the discovery of novel small molecule drugs has been performed by screening large libraries of compounds one-by-one against a validated target of pharmaceutical interest. Given that conventional methodologies based on high-throughput screening (HTS) are highly demanding in terms of time and logistics, novel screening platforms have been developed and implemented in modern drug discovery. In this context, DNA-encoded chemical libraries (DECLs) have emerged as a powerful and versatile tool for lead identification.

In a classical DECL setting, small organic molecules are individually coupled to distinctive DNA oligonucleotide tags which serve as amplifiable identification barcodes and facilitate the construction of combinatorial libraries of unprecedented size. Typically, the identification of binding molecules from DECLs relies on affinity selection procedures against a solid-supported target, followed by PCR amplification and DNA sequencing.

However, the performance of affinity selection procedures has been narrowly investigated so far. This thesis presents a strategy for the quantitative analysis of affinity selection experiments using three sulfonamide-functionalized ligands with different affinity to carbonic anhydrase IX (CAIX) – a well-known validated tumor associated antigen – as model compounds. Firstly, quantitative PCR procedures (qPCR) were implemented in order to evaluate the recovery and selectivity for affinity selection procedures performed using conventional solid supports. Secondly, the recovery of individual CAIX ligands was determined in the context of a single pharmacophore library, containing 360000 DNA-encoded compounds. Collectively, the results show that conventional affinity capture procedures are efficient for the identification of high-affinity ligands but may be suboptimal for the recovery of binders with a K_d in the micromolar range.

Indeed, micromolar binders are characterized by limited kinetic stability of their complex with the cognate protein target, leading to loss of recovery upon implementation of stringent washing procedures. In this context, the use of photocrosslinking reactions may help to "lock the equilibrium" and prevent the disruption of the interaction between putative binders and their target proteins.

Thus, we aimed at developing and performing a systematic evaluation of a novel photocrosslinking screening methodology, featuring a library displayed on single-stranded DNA, which could be hybridized to a complementary oligonucleotide carrying a diazirine photoreactive group as a terminal crosslinker. Investigation of model selection experiments

against CAIX revealed a recovery of individual binders up to 10%, which was mainly limited due to the high reactivity of intermediate carbene species, generated during the photocrosslinking reaction. In the first set of experiments, model sulfonamide ligands featuring acetazolamide and p-phenysulfonamide were recovered more efficiently compared to their counterparts based on 3-sulfamoyl benzoic acid, which were characterized by a lower binding affinity towards the target. Systematic optimization of experimental parameters revealed suitable conditions for the implementation in a real DECL library, featuring 669'240 combinations of two sets of building blocks. Compared to conventional affinity capture procedures, the photocrosslinking methodology provided better discrimination towards the identification of low-affinity CAIX ligands over the background noise.

The results show that DECL libraries in the single-stranded format are versatile discovery tools that can be used for the construction of encoded self-assembling chemical libraries (ESAC) and for the implementation of photocrosslinking methodologies. The present results support both methodologies (affinity capture and photocrosslinking strategies) for the identification of new binders, which could be lost within the screening process.

RIASSUNTO

Tradizionalmente, la scoperta di nuovi farmaci inizia con una fase di screening di vaste librerie di composti chimici mirati contro uno specifico target di interesse. Dato che le procedure convenzionali basate su metodi "high-throughput" possono essere subottimali in termini di tempo e costi di ricerca, nel corso degli anni sono state sviluppate ed implementate nuove tecnologie per l'utilizzo nella moderna ricerca farmaceutica. In questo contesto, l'utilizzo di librerie chimiche codificate dal DNA è stato determinante per la scoperta di nuove molecole farmacologicamente attive.

La costruzione di una libreria prevede la coniugazione di ciascuna molecola organica a uno specifico frammento di DNA che codifica inequivocabilmente la sua natura e facilita la sintesi di vaste librerie combinatoriali di composti chimici.

Tipicamente, la ricerca di ligandi in grado di legare determinati targets proteici si basa su procedure di screening di affinità contro una proteina immobilizzata su una determinata fase solida, seguita da amplificazione in PCR e tecniche di sequenziamento del DNA.

Tuttavia, il rendimento delle procedure moderne utilizzate per lo screening di librerie codificate dal DNA non è stato ancora sufficientemente valutato. In questa tesi è presentata una strategia per l'analisi quantitativa dell'efficienza di metodologie di selezione, utilizzando tre ligandi modello basati su solfonammidi, in grado di legare l'anidrasi carbonica IX (CAIX) con affinità diversa. Inizialmente, il lavoro ha previsto l'implementazione di procedure di amplificazione in PCR quantitativa per l'analisi quantitativa della resa di recupero e della selettività dei metodi di selezione utilizzando supporti solidi comunemente utilizzati in cromatografia di affinità. Successivamente, la quantificazione della resa di ciascun ligando è stata determinata utilizzando una libreria a singolo farmacoforo, contenente 360'000 molecole organiche. I risultati hanno evidenziato che le procedure di selezione convenzionali sono efficienti per l'identificazione di ligandi ad alta affinità, ma possono essere subottimali nel riconoscimento di molecole caratterizzate da una costante di dissociazione micromolare.

I complessi tra ligandi micromolari e i rispettivi targets sono infatti caratterizzati da una bassa stabilità cinetica, che puó portare a una rottura del legame e a un abbassamento della resa di recupero, a seguito della fase di lavaggio per rimuovere i composti non-leganti. Per questo motivo, si possono utilizzare determinate procedure di "photocrosslinking", per poter stabilizzare l'interazione tra ligandi e targets e bloccare l'equilibrio irreversibilmente.

In questo lavoro di tesi è stata sviluppata ed eseguita una valutazione sistematica di una procedura di selezione basata su "photocrosslinking", utilizzando una libreria costruita su un singolo filamento di DNA e ibridato a un oligonucleotide coniugato a una diazirina fotoreattiva. L'analisi di procedure di screening utilizzando composti modello e anidrasi carbonica IX ha permesso la determinazione di rese di recupero fino al 10%, limitate dall'elevata reattività delle specie carbeniche intermedie generate durante la reazione di crosslinking. Inizialmente, i ligandi solfonammidici contenenti acetazolamide e *p*-fenilsolfonammide sono stati recuperati con un'efficienza maggiore, al contrario di quelli basati su *m*-fenilsolfonammidi, caratterizzati da una costante di affinità meno favorevole. Lo studio sistematico delle variabili sperimentali ha permesso l'ottimizzazione della procedura e l'implementazione per lo screening di una libreria, contenente 669'240 molecole organiche. In confronto alle procedure di selezione tradizionali basate su cromatografia di affinità, la metodologia di "photocrosslinking" ha permesso una migliore discriminazione dei ligandi micromolari dal resto della libreria.

I risultati hanno mostrato che le librerie codificate da codici di DNA a singolo filamento sono versatili e possono essere utilizzate non solo per la costruzione di librerie chimiche "DNA-encoded" ad auto-assemblaggio (ESAC), ma anche per l'utilizzo di procedure di "photocrosslinking". In conclusione, i due metodi di screening – su fase solida e basati su "photocrosslinking" – possono essere utilizzati in maniera complementare e rispettivamente per il riconoscimento di ligandi ad alta e bassa affinità.

TABLE OF CONTENTS

1.INTRODUCTION

1.1 SCREENING PLATFORMS USED IN DRUG DISCOVERY

The discovery of specific molecules (e.g., small molecules or biologics), which exhibit activity against biological targets remains a challenging task for the chemistry, biology and biomedical science field and thus requires a multidisciplinary approach. [1]

In pharmacology, a small molecule drug (SMD) is a molecularly defined chemical entity characterized by low molecular weight (MW < 1000 Da) with a diameter in the order of 10^{-9} m. These physical parameters allow small molecules to rapidly diffuse across cell membranes and reach cytoplasmatic intracellular targets. [2] Moreover, small molecule drugs often work rapidly, reversibly and their activity depends on their concentration. Their pharmacokinetic properties – ADME (Adsorption, Distribution, Metabolism and Excretion) are also peculiar, and they can be administrated by a variety of routes, including orally. [3]

It has been proposed that small organic molecules must comply with the Lipinski's Rule of Five (Ro5) in order to be considered "drug-like". [4] The rule was formulated by Christopher Lipinski, based on the experimental observation of physicochemical parameters of many oral drugs available in the market. The rule can be summarized as follows: 1) hydrogen bond donors ≤ 5; 2) hydrogen bond acceptors ≤ 10; 3) molecular weight ≤ 500; 4) clogP ≤ 5.

Biologics are products derived from living organisms (e.g., human, animal or microorganisms) and thus can contain components of living organisms. [5] Typical biologic drugs include monoclonal antibodies, nucleic acids, cells and recombinant proteins. The popularity of these medicines has been increasing over the last decade, thanks to the advances in biotechnology and bioanalytical methods. In terms of specificity and biocompatibility, these drugs are often preferred over small molecule preparations since they may contain proteins which interact in specific cellular processes (e.g., cancer, autoimmune diseases). Side effects of biologics largely depend on the method of introduction into the body, which is mainly by intravenous injection. Moreover, the high costs of production and evaluation of biopharmaceutical drugs reflect their high price of sales in the market.

Recent approval trends and the current pipeline show that small molecules are still preferred on the market. In 2019, the FDA approved 48 novel drugs, 73% of which were small molecules. (Figure 1.1). [6] This trend is partially justified by the high costs of production and evaluation of biopharmaceutical drugs, which is reflected in their price of sales in the market. However, the discovery of novel drugs remains a costly process that takes in average more than a decade from discovery to approval [7]. In the past, the discovery of new medicines mainly relied on the optimization of novel natural product derivatives. [8] With the advent of novel screening technologies (e.g., high-throughput drug screening and combinatorial chemical libraries), drug discovery shifted the focus from the traditional drug discovery process (e.g., use of medicinal plants) to small molecule synthesis. Further development in the production and design of novel methodologies used for the synthesis and screening of compound libraries will likely improve the time and costs of drug discovery programs and the risk of late-stage failure. [9]

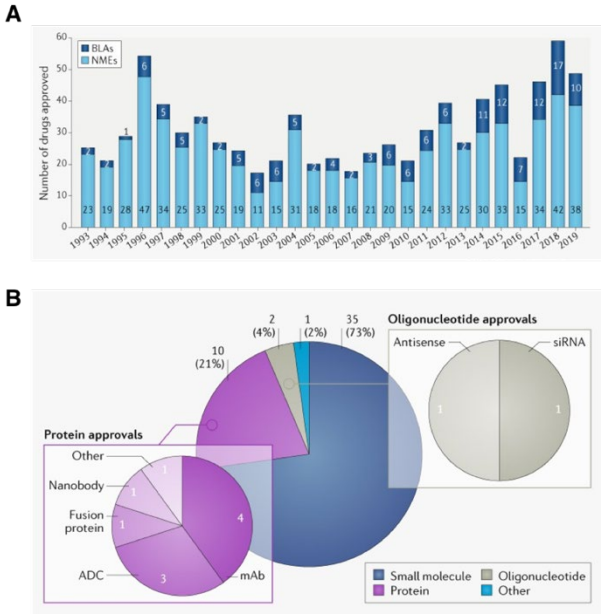


Figure 1.1. A. Annual numbers of FDA's Center for Drug Evaluation and Research (CDER) approvals of new molecular entities (NMEs) and biologics licence applications (BLAs) from 1993 to 2019. B. CDER approvals by modality. NMEs include small molecules and oligonucleotides. Protein-based candidates - nanobodies, fusion proteins, antibody-drug conjugates (ADC) and monoclonal antibodies (mAb) are approved as BLAs. Image adapted from: "2019 FDA drug approvals – Nature Reviews drug discovery". [6]

1.1.1 Screening Platforms Used in Drug Discovery

Nowadays, early drug discovery programs start with the screen of hundreds or thousands of molecules against a validated target of interest. The target can be defined as a molecular entity (e.g., protein, gene, enzyme, receptor) which is involved in a specific biological process. [10] The identification of a target is followed by characterization of the molecular mechanisms in which the target is involved. Defining the "druggability" [11] – the likelihood of being able to modulate the activity of a target using a small molecule drug – is crucial in determining the success of a drug discovery program. A "druggable" target is accessible to a small molecule or a large biologic, and upon binding, elicits a measurable biological response within *in vitro* and *in vivo* biophysical experiments.

Following the validation of a target, compound screening assays are performed during the hit identification and lead discovery phase. In principle, a variety of screening platforms can be used to identify "hit molecules", including high-throughput screening (HTS), fragment screening, ligand-based and structure-based computational methods and combinatorial libraries. These techniques will be introduced in the following paragraphs.

1.1.1.1 *High Throughput Screening (HTS)*

Large libraries of compounds (e.g., chemical libraries comprising 10^3–10^6 compounds) can be screened against a validated drug target using high-throughput screening (HTS). Secondary hit validation assays are generally performed in a second phase to confirm the mechanism of binding and position. [12, 13]

HTS assays are typically performed in microtiter plates (e.g., 96-, 384- or 1536-well formats) and utilize liquid-dispensing and plate-handling robotics for automation. These screens have the potential to analyze between 10^4–10^5 samples per day. Screenings are performed at a compound concentration which is typically between 1–10 μM, but this parameter can be varied to identify compounds with lower or higher activity. [14, 15] In the past two decades, pharmaceutical companies have invested billions of dollars in building multimillion compound libraries, which typically deliver drug leads against many classes of targets.

However, Chen *et al.* [16] reported a success rate as low as 0.0001% in the identification of inhibitors against JIP-JNK. This limitation arises when traditional compound libraries are screened for protein-protein interactions (PPIs) [17], which represent a class of challenging targets due to the highly dynamic and large interfacial areas. These findings are supported by Santos *et al.* [18], which states that small molecule drugs address a limited range of targets (e.g., Rhodopsin-like G protein-coupled receptors, ion channels, kinases, nuclear receptors). Further improvements in library design and fine-tuning of physicochemical parameters are needed to enable a systematic validation of the chemical space.

Another limitation of HTS is the use of complex laboratory automation (e.g., liquid handling devices, sensitive detectors, robotics), which may be demanding in terms of management, costs and logistics. [19]

1.1.1.2 *Fragment-Based Drug Discovery (FBDD)*

Fragment-based drug discovery (FBDD) is a target-based method used to identify small organic starting points ("fragments") which can be further optimized to drug-like leads and clinical candidates by chemical growing and/or linking (Figure 1.2). [20–22]

The procedure relies on the screening of a small number of low molecular weight fragments against a target of interest to identify compounds which give meaningful interactions with the protein. In analogy with Lipinski's Rule of Five, it has been proposed that these fragments should follow the "rule of three" (Ro3) (molecular weight ≤ 3; number of hydrogen bond donors and acceptors ≤ 3; clogP ≤ 3). [23]

Hits arising from an FBDD screen typically display a dissociation constant (K_d) in the range of 0.1–10 mM and above. Therefore, sensitive detection methods are required for the assessment of binding (e.g., X-ray crystallography, NMR, surface plasmon resonance, HTS screens at high concentration) [24]. Calorimetry methods (e.g., isothermal titration calorimetry) are suited for FBDD because they are sensitive to weakly binding ligands and can provide an accurate read-out of functional activity.

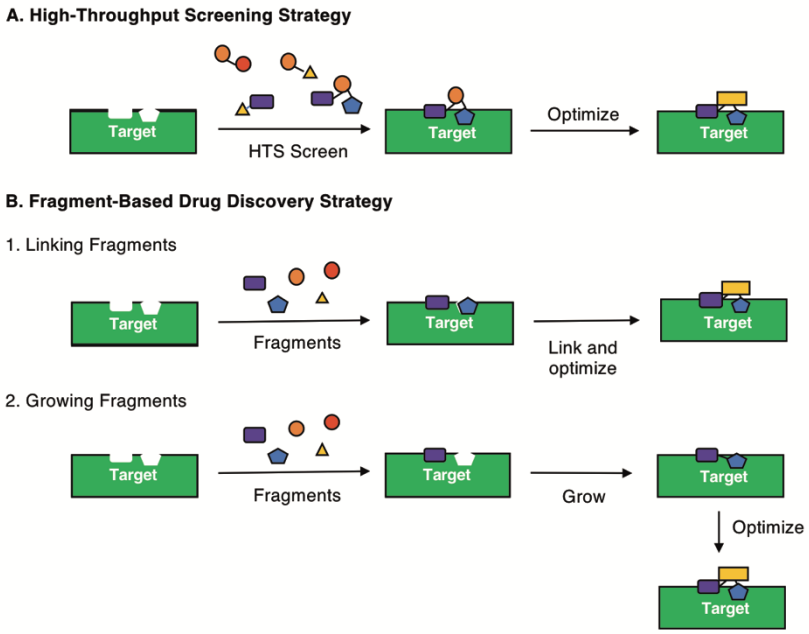


Figure 1.2. Comparison between high-throughput screening (HTS) and fragment-based drug discovery. (FBDD). Using HTS (A), large libraries of compounds are screened against a target of interest to identify potential binders, which can be subsequently optimized. In FBDD (B), the identification of a lead molecule starts with the identification of binding fragments, which can be linked (B.1) or grown (B.2) to generate more potent molecules.

It has been demonstrated that simple organic fragments have a higher probability of binding to complex active sites. [25] High molecular weight molecules can be sterically excluded from these sites. Binding fragments generate an entropy loss of the biological system through the formation of a high-quality interaction. [26, 27] It has been argued that a molecule must sufficiently interact with the target in order to overcome an entropy loss of 15 – 20 kJ mol^{-1} [28], that is due to a reduction in the motion of the molecule from its solution state to bound state. This is not an issue for larger molecules (MW > 500 Da) that can make multiple interactions with a protein, whereas small fragments can overcome this energy barrier only by making particularly strong interactions with a protein. [29] Since the energy penalty has already been "paid" by the single fragment to the protein, additional interactions after optimization of the lead can only improve the binding affinity of the molecule towards the target. [21]

FBDD has been extensively used for challenging targets as binding sites in protein-protein interactions (PPIs). [30] Fesik and co-workers [31] reported the discovery of single-digit nanomolar myeloid cell leukemia-1 (Mcl-1) inhibitor using an NMR-based screen of an extensive fragment library. In 2016, Davies *et al.* [32] described the identification of a potent

nanomolar small molecule inhibitor of Kelch-like ECH-associated protein-1 (KEAP1) using fragment-based drug discovery.

1.1.1.3 *Computer-Aided Drug Design (CADD) methods*

Computer-aided drug design (CADD) methods represent a class of computational techniques that can be used to identify chemical structures which are most likely to bind a target of interest. [33–35] The purpose of the method is generally to screen large compound libraries (> 10^{12} molecules) and identify "smaller clusters" of predicted active compounds. Further optimization of lead compounds is then performed through the improvement of biological properties (e.g., metabolism and pharmacokinetics properties (DMPK), ADMET). [36]

Virtual screenings (VS) can be classified in two main groups: target-structure dependent and ligand-based. [37] Structure-based methods (SBVS) require an in-depth knowledge of the target structure, which makes possible the use of molecular docking experiments of ligands into protein binding sites [38, 39]. Ligand-based virtual screenings (LBVS) utilize the information derived from available small molecules which are biologically active (e.g., chemical similarity criteria) and from predictive quantitative-structure-activity relationship (QSAR) models. [40]

Small molecules are typically screened using algorithms that make use of structural representations of the compounds (2D or 3D) and molecular descriptors. These mathematical parameters describe a specific molecular property, which might be experimentally determined (e.g., logP octanol/water) or theoretically calculated from the chemical structure (e.g., Gibbs free energy). [41] Generally, structure-based methodologies are preferred when soluble proteins can be crystallized, whereas ligand-based is suited for compounds with high affinity to the target. [42]

1.1.2 The Role of Combinatorial Chemistry

Combinatorial chemistry has provided novel synthetic methodologies for the synthesis of focused or diverse chemical libraries with a wide range of chemical moieties (e.g., small molecules, complex natural products, peptides, macrocycles, peptidomimetics). [43]

The concept of combinatorial chemistry was born with the development of solid-phase peptide synthesis methodologies, implemented by Bruce Merrifield *et al.* in 1963. [44] The authors reported the synthesis of a tetrapeptide (H-Leu-Ala-Gly-Val-OH) on polystyrene resin beads using Cbz-protected amino acids. In the following years, other research groups

reported the synthesis of peptides [45], oligonucleotides [46] and oligosaccharides [47], which provided the experimental basis of modern combinatorial chemistry.

In 1992, Ellman and Bunin [48] reported the synthesis of substituted benzodiazepines using solid-phase synthesis. This work represented the first application of combinatorial chemistry for the synthesis of drug-like small molecules and extended the type of chemical reactions that could be applied on solid-phase (e.g., amidation, carbonyl-amino condensations, *N*-alkylations).

The first example of combinatorial "split-and-pool" approach was described by Furka *et al.* [49] and applied to solid-phase peptide synthesis. A combinatorial pool of organic molecules can be synthesized by iterations of "splitting" and "pool" steps of solid-phase intermediates for separate reactions with different chemical reagents ("building blocks"). The number of generated products depends on the number of building blocks and "split-and-pool" iterations (Figure 1.3).

Since that time, combinatorial chemistry approaches have evolved and have been automatized to enhance productivity and efficiency in drug discovery. The "marriage" between combinatorial chemistry and high-throughput screening techniques to hit identification seemed to be convenient to generate large libraries of compounds which could be screened against various targets of interest.

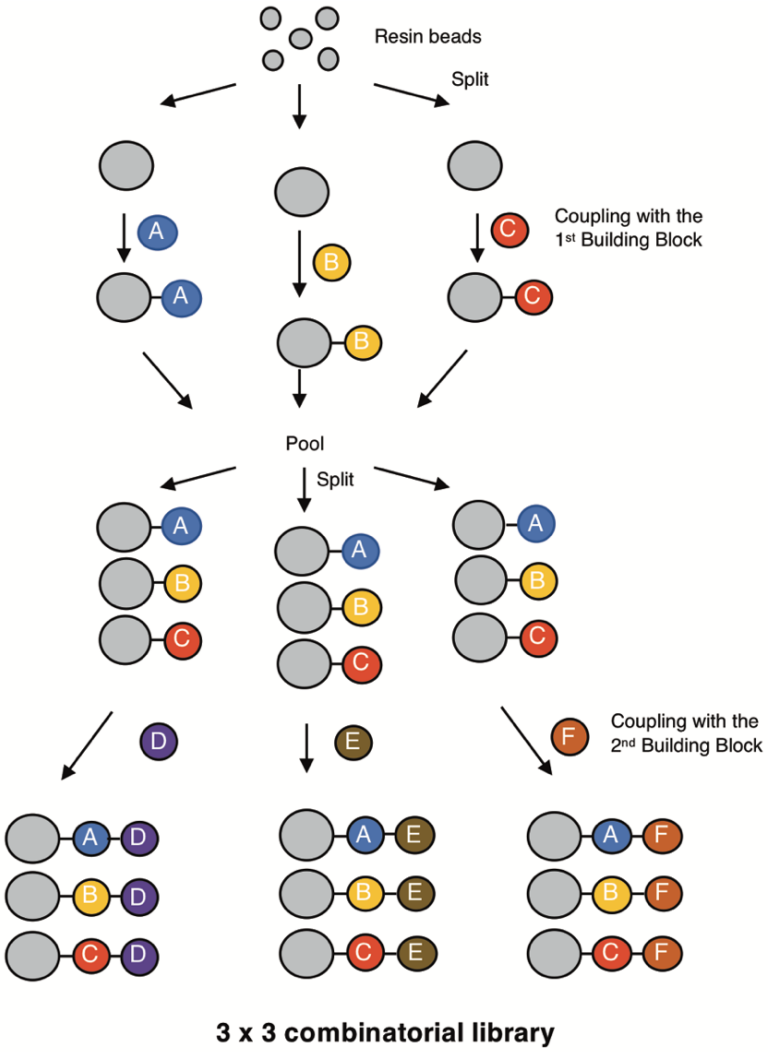


Figure 1.3. Representative synthetic scheme for the generation of a 3 x 3 combinatorial chemical library using combinatorial "split-and-pool" synthesis. For the first coupling reaction, resin beads are distributed into three wells and reacted with various building blocks (A, B, C). After coupling, the products are pooled, and the resulting mixture is split and reacted with the second set of building blocks (D, E, F). The size of the library is calculated from the number of building blocks used in each step (3 x 3 = 9).

The next problem to address was the identification of active compounds after the library screening. Due to the low concentration of each molecule in the library mixture, "tagging" (referred as "encoding") strategies suitable for chemical compounds became a relevant issue.

Several encoding strategies have been proposed and implemented for small molecule combinatorial chemical libraries. Examples include encoding by positional or spatial tagging such as Chiron Mimotipes [50], "teabags" [51], radio-frequency tags [52], chemical encoding by molecular mass [53] and electrophoric molecular tags [54].

In parallel with the implementation of novel encoding strategies, the development of biochemical display technologies (e.g., phage display libraries) provided the idea of using DNA oligonucleotides as a means of encoding. The implementation led to the generation of DNA-encoded chemical libraries, which will be discussed in detail in the next paragraph.

1.2 DNA-ENCODED CHEMICAL LIBRARIES

The concept of DNA-encoded chemical libraries (DECLs) has been derived from the research work on encoded combinatorial libraries of polypeptides, commonly referred as "phage display libraries", which have played a significant role for protein engineering and several research fields. [55– 62]

The technology was first described by George P. Smith in 1985, and it encompasses the display of polypeptides on filamentous *E. coli* phage M13 by fusing the peptide of interest to the gene III of a filamentous phage. [63] Phage libraries can be used for several applications, including B- and T- cell epitope mapping, [64, 65] selection of target-binding peptide motifs [66] and development of peptide-mediated drug delivery systems. [67, 68]

The potential binding epitope ("phenotype") is displayed on a single particle ("phage"), which contains the encoding genetic information ("genotype"). Upon binding to a protein target through affinity capture experiments ("biopanning"), binders are identified by amplification of the encoding genotype. A typical affinity capture experiment involves the panning of the library on a solid-supported antigen, washing out of the non-binders and recovery of putative binders by elution. Generally, three to five rounds of biopanning are required in order to obtain targets that bind with high affinity. [69]

In 1990, McCafferty and Winter [70] suggested using phage display for the discovery and production of recombinant antibodies with desired specificities.

Several biological drugs (>40) have been discovered from antibody phage display technology. [62] Adalimumab (Humira®; AbbVie Inc., formerly Abbott Laboratories, North Chicago, IL) is a fully-human antibody which was discovered thanks to antibody phage display technology with a "guided selection" method involving a mouse mAb against tumor necrosis factor (TNF). [71] The human IgGk antibody is approved for the treatment of several forms of arthritis [72], Crohn's disease [73] and ulcerative colitis. [74]

The power of phage display stems from the physical linkage between phenotype and genotype and the ability to build libraries that range from 10^6 to 10^{11} distinct drug candidates. [75] The physical connection between the displayed polypeptide and the encoding gene allows facile recognition of the binding protein after affinity selection. Moreover, phage display technologies have provided inspiration for DECLs, a powerful technology which can be applied to the discovery of small molecule organic ligands (Figure 1.4).

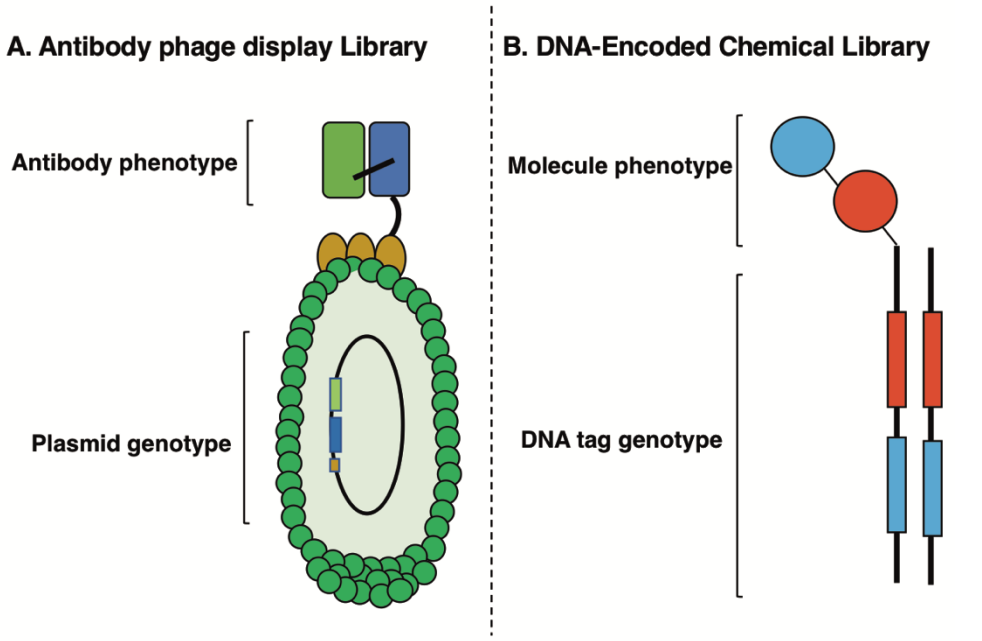


Figure 1.4. Schematic representation of antibody phage display libraries (A) and DNA-encoded chemical libraries (B). In antibody phage display libraries, the phenotype (an scFv antibody fragment) is displayed on the surface of a filamentous phage, and it is physically linked with the genotype, which contains the genetic information. Similarly, in DNA-encoded chemical libraries, each molecule is encoded by a different set of DNA oligonucleotides, which serve as amplifiable identification barcodes.

1.2.1 Historical Overview of DNA-Encoded Chemical Libraries

In 1992, Sidney Brenner and Richard Lerner described in a theoretical article the possibility of coupling solid-phase peptide and oligonucleotide synthesis using orthogonal chemistry and "split-and-pool" methodologies. [76] Using this procedure, a final library of polypeptides can be obtained after cleavage from a pore glass synthesis matrix ("CPG"), and each molecule is encoded by a DNA oligonucleotide (Figure 1.5). The synthetic assembly of the oligonucleotide can be performed through stepwise ligation of DNA fragments. The encoding oligonucleotide merely serves as "identifier" for the corresponding peptide structure after cloning and sequencing, since it is not a biologic genotype. The resulting DECL library can be screened using affinity selection methodologies, in full analogy with phage display techniques.

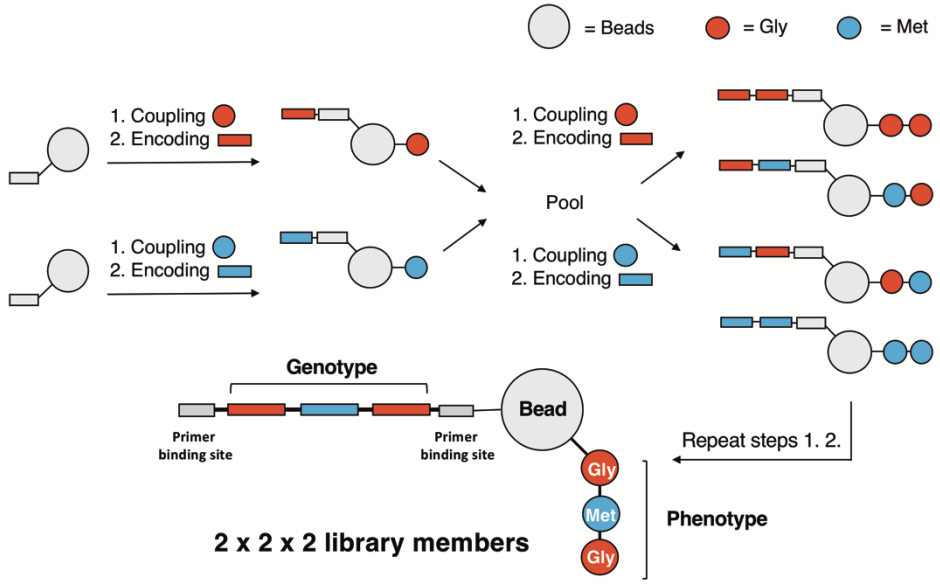


Figure 1.5. Schematic representation of the library synthesis proposed by Brenner and Lerner. In the first step, a solid-phase surface (e.g., beads, represented in gray) is functionalized with a general DNA oligonucleotide which serves as PCR primer binding site. The resin is split in two aliquots and one amino acid (e.g., Gly, Met) is coupled on the bead. The oligonucleotide sequence is orthogonally extended with a 6-mer DNA oligonucleotide, which univocally identifies a set of building blocks. The repetition of split-and-pool cycles leads to the generation of a 2 x 2 x 2 tripeptide library. Gly: glycine; Met: methionine.

The first implementation of this approach was studied by S. Brenner and K. Janda [77] and the group of M. A. Gallop. [78] Brenner and Janda proposed the synthesis of an encoded library of pentapeptides by alternating parallel combinatorial synthesis of the peptide and the oligonucleotide sequence on the same bead, in a split-and-pool fashion. Gallop *et al.* exemplified the Brenner and Lerner approach by synthesizing a library of 800000 heptapeptides, performing seven alternating of "split-and-pool" steps using *D*- and *L*-amino acids. The library was subjected to on-bead screening using fluorescent-activated cell sorter (FACS)-based selections and delivered a peptidic binder against the fluorescent monoclonal antibody D32.39.

In 1995, Kinoshita and Nishigaki [79] demonstrated the feasibility of enzymatic ligation procedures as a means for tagging chemical entities generated in a combinatorial fashion. Beads are not required for the chemical synthesis, but the procedure must be compatible with the presence of oligonucleotide tags and subsequent procedures (e.g., enzymatic ligation, PCR amplification).

These crucial publications outlined key concepts at the basis of DECL technologies, including combinatorial chemical synthesis, polymerase-chain reaction amplification, DNA sequencing and hit validation of drug candidates.

At the beginning of 2000s, DECL libraries were developed by directly attaching small molecule compounds to DNA oligonucleotides, without the use of beads. [80, 81] Several groups in academia and industry started to synthesize libraries of unprecedented size, which provided an economic and feasible platform to explore larger chemical spaces.

Nowadays, DECLs can present >100 billion compounds for selection by a target of interest. For example, companies as *Hitgen*, *X-Chem* and *Nuevolution* have assembled libraries containing 400 billion, 200 million and 40 trillion of compounds, respectively. [82] As a comparison, the size of a standard HTS library is restricted to a few million of compounds due to the costs of synthesis and managing such extensive collections. It has been estimated that the synthesis and screening of an HTS library would cost between $0.4 and $2 billion (ca. $1100 per compound), while screening a DECL library of 800 thousand compounds, would cost around $150000 ($0.0002 per compound). [83] Once prepared, libraries can be stored in a "single tube" and used for thousands of selections in a multiplex form. Since the compound are assayed at once, the costs of screening are negligible compared to one-by-one assay formats.

1.2.2 Classification of DNA-Encoded Chemical Libraries

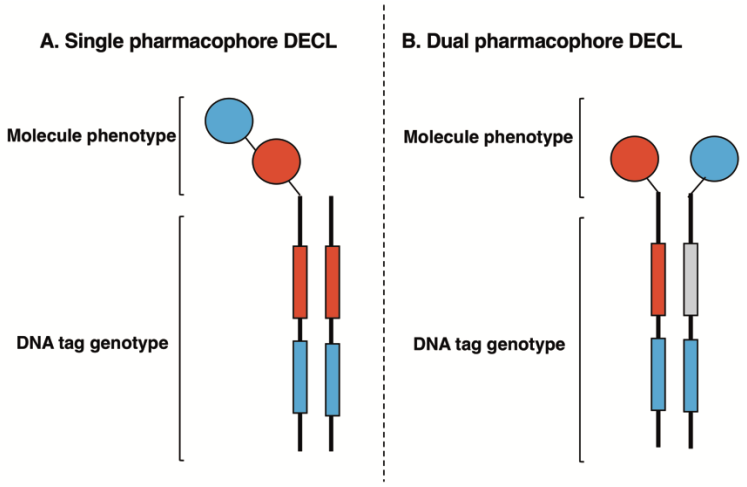


Figure 1.6. Representative examples of single pharmacophore (A) and dual pharmacophore (B) DECLs designs. Building blocks are represented as circles, and the respective encoding regions are represented with colored bars. The DNA oligonucleotide strands are represented in bold lines.

From a synthetic point of view, DNA-encoded chemical libraries can be grouped into "single pharmacophore" libraries and "dual pharmacophore" libraries. In contrast to the first format, where only one chemical moiety is displayed on a DNA strand, "dual pharmacophore" libraries feature two independent small organic molecules in close proximity (Figure 1.6). Various synthetic strategies can be implemented for library synthesis, which will be further described in the next paragraphs.

1.2.2.1 *Single Pharmacophore DNA-Encoded Chemical Libraries*

The traditional approach used for the synthesis of single pharmacophore DNA-encoded chemical libraries relies on a procedure named "DNA-recorded" (also named "DNA-encoded") synthesis. [84] Using this method, the chemical synthesis is "recorded" by a series of short DNA oligonucleotides which can be ligated in a stepwise manner.

In principle, DNA-recorded synthesis can be carried out using short double-stranded (dsDNA) or single-stranded DNA fragments (ssDNA). The two approaches are different and lead to the generation of "single-stranded" or "double-stranded" DECLs.

The use of dsDNA has been described by several groups both in academia and industry. [85–88]. In principle, individual fragments can be covalently linked by enzymatic or chemical ligation. [89] Clark *et al.* [86] described the construction of a DECL library using a combination of enzymatic and chemical synthesis in split-and-pool and using a double-stranded DNA encoding. (Figure 1.7).

The precursor used for the library synthesis was a short DNA duplex ("headpiece") which was further elongated after each step of the chemical synthesis using short dsDNA coding oligos. The authors claimed that the presence of a duplex structure would prevent denaturation reactions on the heterocyclic bases within the DNA helix during the library synthesis. In fact, although DNA is considered a chemically stable macromolecule, it can undergo several side reactions (e.g., loss of DNA bases, depurination, deamination, cleavage of phosphodiester bonds) that could lead to loss of information and failure in the PCR amplification. [90–92]. Moreover, thanks to the heterodimeric structure, the DNA strand interferes less likely with the binding to a putative target (single-stranded DNA can bind to single-stranded DNA-binding proteins as protection means against mutations). [86]

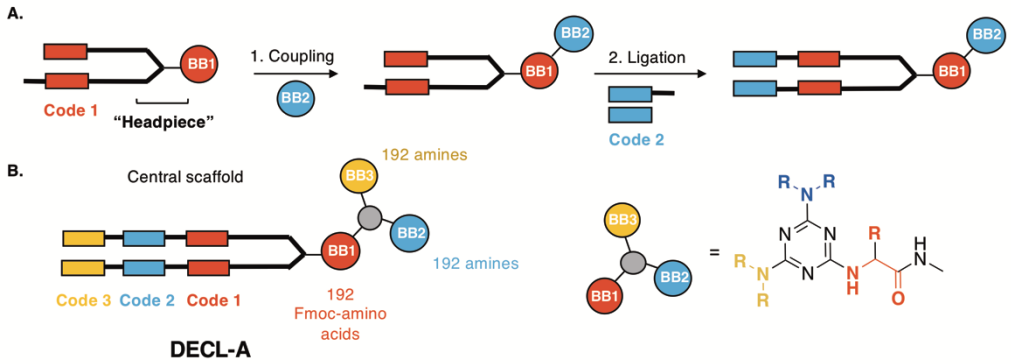


Figure 1.7. A. Stepwise ligation strategy used by Clark *et al.* for the construction of a DNA-encoded chemical library ("DECL-A"). [86] The complementary DNA oligonucleotides are linked by a short oligonucleotide sequence ("headpiece") which supports the synthesis of the library. The coupling of the first set of building blocks is followed by an enzymatic ligation, exploiting a 3' two-base overhang. The encoding scheme is represented for the construction of a two-building block library. B. Schematic representation of DECL-A. The library is based on the presence of a triazine scaffold, which is substituted stepwise with 192 Fmoc amino acids and two sets of 192 amines. BB: building block.

Neri *et al.* described the encoding of DNA-encoded chemical libraries by using ssDNA oligonucleotides. [85, 93– 99]. In this setting, the first building block is attached to one extremity at the 5' end of an amine-functionalized oligonucleotide, which contains a short portion (six to eight bases) designed to be distinctive for each building block ("barcode"). The identity of the second building block is encoded by ligation using a partially complementary oligonucleotide ("DNA adapter"), which can be washed away upon denaturing washes and HPLC purification (Figure 1.8). As an alternative, a Klenow fill-in polymerization procedure can be exploited by using a reverse complementary oligonucleotide, which is partially annealed to the library strand. [95, 98] The oligonucleotide can be introduced as a final barcode element, for the encoding of a subsequent synthetic step, or as a library identifier (Figure 1.8). The main difference between the two approaches is that the latter leads to the generation of a "double-stranded DECL", where the two complementary oligonucleotide fragments are not chemically ligated. These procedures have been used to synthesize libraries up to three sets of building blocks. [93–99].

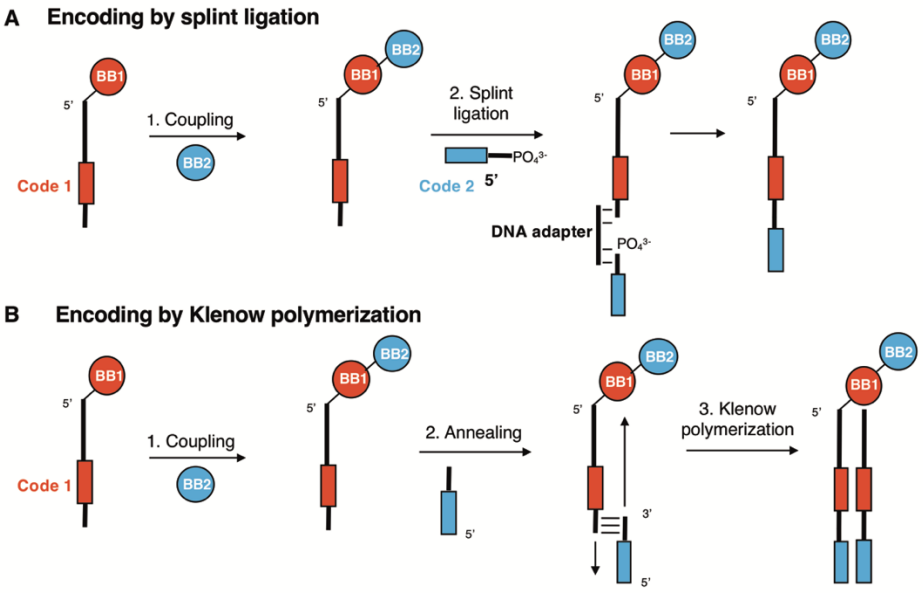


Figure 1.8. Encoding schemes implemented by the Neri group. A. Encoding by enzymatic ligation. B. Encoding by Klenow polymerization. BB: building block.

An alternative methodology for the synthesis of DECLs has been pioneered by the group of David Liu and referred to as "DNA-templated" synthesis. [100] The methodology relies on DNA-directed chemical reactions to promote the coupling of diverse sets of building blocks. The interaction of two nucleobases through hydrogen bonds is known to accelerate bimolecular reactions and to increase the local concentration of the reactants in solution (Figure 1.9A). [101]

The authors described the implementation of DNA-templated synthesis for the construction of a 65-members library of macrocycles, which can be particularly challenging to synthesize using conventional ways. The authors described the use of a 48-base DNA-linked lysine derivative used as "DNA template" to direct three steps of DNA-directed amine acylation reactions with building blocks conjugated to 10-mer or 12-mer biotinylated oligonucleotides. These DNA strands were complementary with three unique coding sequences, displayed on the template. [100] After each coupling, suitable cleavage steps are required to allow the next coupling step to proceed, and the reagent oligonucleotides are removed by affinity capture with streptavidin-linked magnetic beads. However, a limitation of this approach is that the template sequences must be carefully designed and validated, to avoid base mismatch regions and ensure encoding fidelity for subsequent applications (e.g., polymerase chain-reaction amplification).

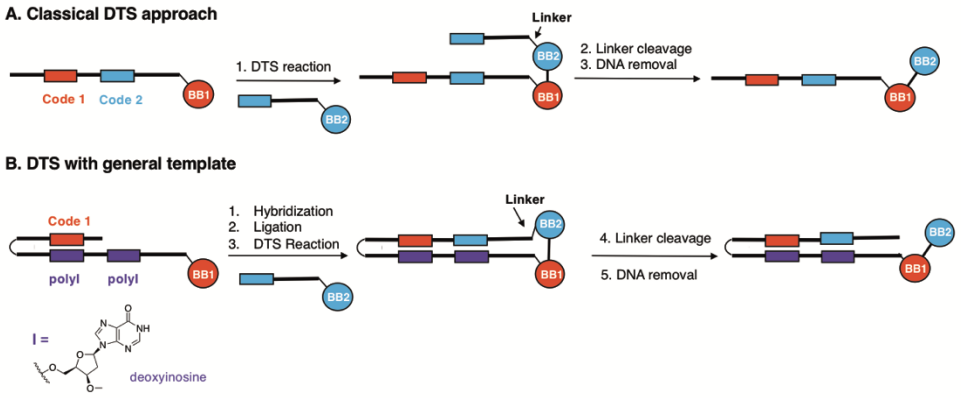


Figure 1.9. Schematic representation of DNA-templated synthesis (DTS) applied for the construction of DNA-encoded chemical libraries. A. Classical DTS approach used for the generation of a single pharmacophore library comprising of two building blocks. After annealing of the oligonucleotides, the first building block (BB1) reacts covalently with a small molecule-DNA conjugate (BB2). At the end of the coupling step, the connection between BB2 and the corresponding oligonucleotide is cleaved. The resulting DNA fragment is removed. B. A universal DNA strand containing polydeoxyinosine regions (polyI) is used as a general template to direct the synthesis of the chemical library after sequential hybridizations and coupling steps.

Using an alternative approach, Li *et al.* reported the use of a single "universal template code", which is capable of directing chemical reactions with multiple small molecule-DNA conjugates displaying various encoding sequences. [102] The universal DNA oligonucleotide contains multiple regions of polydeoxyinosines ("polyI") as anticodons, which can form "omega" structures upon unspecific hybridizations with the four nucleotide bases. This feature is known to improve DTS reaction yields (Figure 1.9B). [100]

A tridimensional evolution of the classical linear DNA-templated encoding strategy was implemented by *Vipergen* and referred as "Yoctoreactor®". This methodology relies on the annealing and subsequent enzymatic ligation of three-way DNA-hairpin-looped junctions that display chemical moieties. In the first step, two single-stranded DECLs are mixed with a third oligonucleotide, which assists the self-assembling of the libraries. Upon the formation of a three-way junction construct, two sets of building blocks are then coupled by DTS reaction, and the resulting library is purified by polyacrylamide gel electrophoresis. After cleavage of the linker and DNA ligation, the procedure is repeated to generate a library displaying three sets of building blocks. The final step involves a PCR primer extension to generate the complementary strand. After hybridization, the building blocks are transferred onto a core acceptor site, and the final library is obtained in double-stranded format (Figure 1.10A). [103, 104]

Another procedure named as "DNA-routing" was described by Harbury group. [105, 106] The procedure features sequential sequence-specific immobilization reaction and purification/elution steps on resin containing complementary DNA codons (Figure 1.10B).

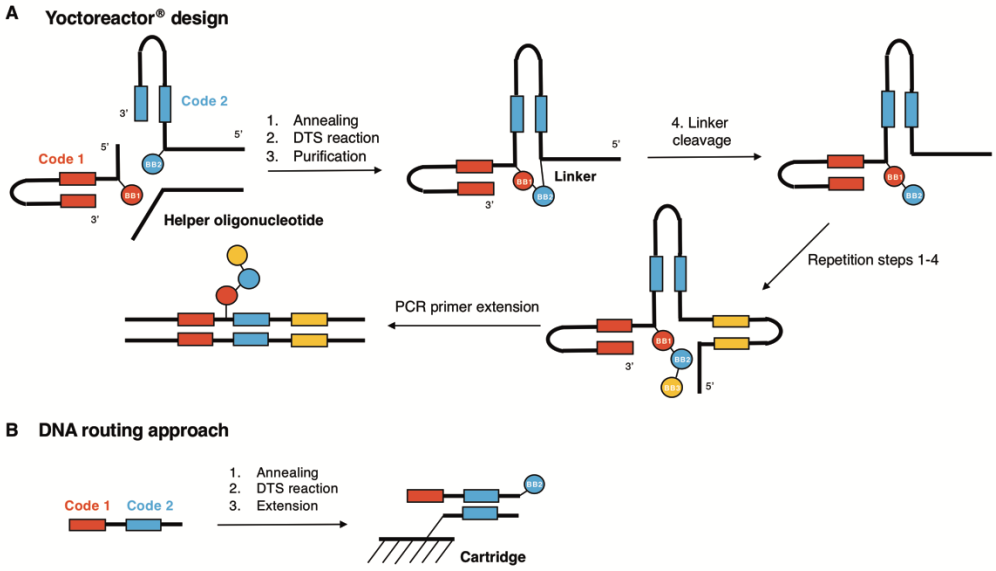


Figure 1.10. A. Schematic representation of the Yoctoreactor® design for the construction of single pharmacophore DECL. B. DNA routing approach as described by Harbury *et al.* Individual DNA oligonucleotides are split ("routed") into sequential cartridges, each containing a complementary DNA sequence. The DNA fragments allow the DTS reaction between oligonucleotides and building blocks and serve as PCR amplifiable identification tags.

1.2.2.2 *Dual Pharmacophore DNA-Encoded Chemical Libraries*

The Neri group has pioneered an alternative strategy for the construction of large and high-purity DNA-encoded chemical libraries. This approach relies on the combinatorial assembly of complementary encoded sublibraries to form a dual pharmacophore library (encoded self-assembling chemical libraries – ESAC). [107–109]

The encoding strategy used for the construction of ESAC libraries has been described in detail by Wichert *et al.* [110] One sublibrary ("sublibrary A") is prepared by coupling sets of individual building blocks to the end of 5'-amino-modified DNA oligonucleotides. The other set of building blocks is coupled to a single oligonucleotide which contains an abasic region ("d-spacer"). The resulting DNA strands are ligated to distinctive coding oligos to yield the sublibrary B. The resulting sublibraries can be annealed, and a code-transfer strategy (Klenow fill-in DNA polymerization step) allows the transfer of the code B onto the complementary strand (Figure 1.11).

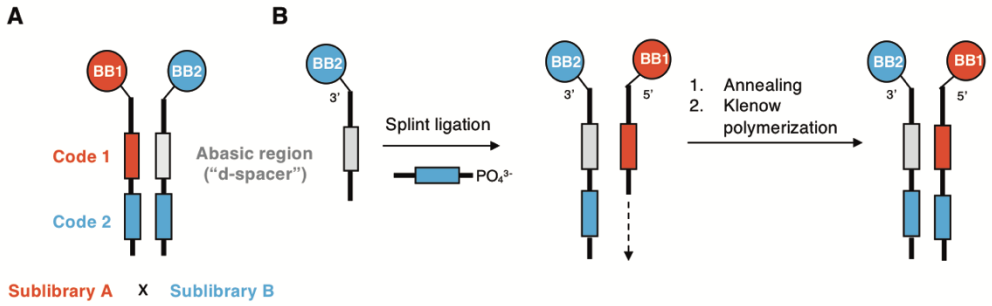


Figure 1.11. A. Schematic representation of the ESAC library design. B. Library construction as described by Wichert *et al.* [110] BB = Building block.

This feature allows the identification and recognition of adjacent binding fragments, which can be exploited in fragment-based drug discovery (FBDD) programs, where single moieties can be engaged in two distinct non-overlapping binding events. The overall increase in binding affinity is due to "chelate effect". [111] In analogy with FBDD, the two moieties have to be arranged into a single organic molecule. For this purpose, various bifunctional linkers need to be tested and screened to yield binding molecules devoid of DNA. [112] The identification of the best linker-building block combination may be facilitated by hit validation methodologies using locked nucleic acid (LNA) derivatives. [113, 114]

Using self-assembly, ESAC libraries are versatile and have a higher degree of flexibility in the arrangement of building blocks compared to classic single pharmacophore DECLs. Moreover, small sublibraries can potentially yield large DECLs, which are characterized by a high degree of purity, as each DNA conjugate is distinctively HPLC-purified, analyzed by LC-MS, and mixed in equimolar amounts. The two sublibraries are then mixed together, yielding a library which retains the initial high quality of the two parental derivatives. Library purity is a key parameter in DECL chemistry, since split-and-pool-based libraries may suffer from insufficient homogeneity due to low reaction yields. [115, 116]

Reddavide *et al.* [117] reported the application of "dynamic recombination" for the synthesis of "ESAC-type" libraries. The principle is based on "dynamic combinatorial chemistry" (DCC), which employs reversible covalent chemistry [118] (e.g., disulfide bond, Schiff-base formation) to create dynamic systems of transient small molecule adducts in thermodynamic equilibrium. In comparison with ESAC, the two sublibraries have a short complementary DNA strand which enables the formation of a heat-induced DNA-encoded dynamic combinatorial chemical library (hi-EDCCL). The two sublibraries can hybridize to form unstable dsDNA interactions due to thermodynamic instability. Upon addition of the target protein, the thermodynamic equilibrium is shifted towards the generation of potent bidentate ligands. Therefore, the protein can be considered as a template for the in-situ generation of potent binders through chemical stabilization of unstable adducts. After a round of selection, the non-binding pairs can be "shuffled" by heating and used for a second round of selection to identify the best combinations.

As an alternative, Zhou *et al.* [119] reported the use of a photocrosslinking reaction to "freeze" the shifted thermodynamic equilibrium prior to hit decoding. The authors described the construction of an EDCCL library starting from two sublibraries containing p-stilbazole moieties in the encoding region. Upon mixing with the target protein, the dynamic exchange of each binding pair is frozen by UV-irradiation. The resulting duplexes are then isolated for hit identification and decoded by DNA sequencing.

More recently, a second generation EDCCL has been reported, which utilizes a Y-shaped DNA construct [120] for the dynamic enrichment of potent binding pairs (Figure 1.12). The procedure is based on the synthesis of a third DNA strand at the end of the selection procedure by enzymatic ligation as means for recording the relationships between binders in close proximity.

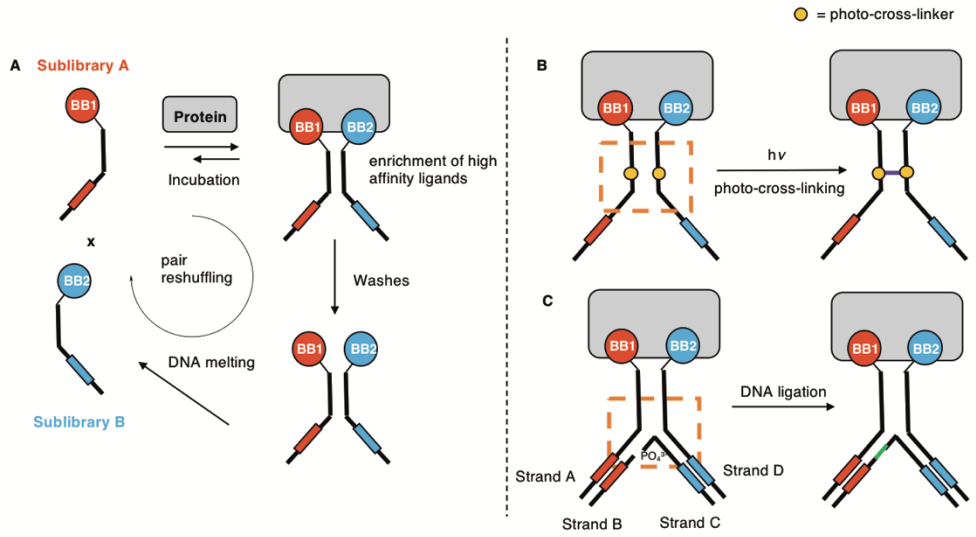


Figure 1.12. A. Schematic representation of hi-EDCCL approach as described by Reddavide *et al.* [117] B. Library design described by Zhou *et al.* [119]. The resulting protein-ligand complexes are irreversibly locked by using a photocrosslinking reaction. C. Second generation of EDCCL using a Y-shaped library design. Upon binding to the target protein, the phosphorylated strand C can be ligated to the strand B to allow the read-out of the binding pair combination. BB = Building block.

Dual pharmacophore chemical libraries can also be constructed using peptide nucleic acids (PNA). Winssinger *et al.* [121–123] proposed the use of DNA templates and PNAs for the generation of heterodimeric and multimeric self-assembling structures (Figure 1.13). PNA-based encoded libraries may be more compatible with reaction conditions than the corresponding DECLs (PNAs are more stable than DNA). [124] The libraries can be generated by stepwise ligation of PNA fragments, but unlike DNA, PNAs cannot be amplified by a suitable amplification procedure (e.g., PCR). For this purpose, the use of a DNA strand, which directs the assembly of PNAs through the formation of a chimeric PNA/DNA assembly, can be exploited as a template for a subsequent PCR amplification after screening against target proteins.

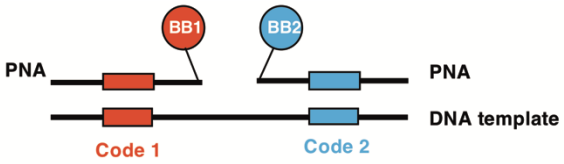


Figure 1.13. Encoding strategy using PNA/DNA strands. Chemically modified PNAs are hybridized to a DNA scaffold which contains the complementary coding sequences. PNA = Peptide nucleic acid. BB = Building block.

1.2.3 Success stories in DNA-Encoded Chemistry

Several examples of hits discovered using different implementations of the DECL technology have been reported in literature. [115]

A micromolar ligand against IL-2 (interleukin-2) has been discovered by Leimbacher *et al.*, [98] by screening a two-building block library featuring 100 different amino acids and 300 carboxylic acids. (Kd = 2.5 μM by fluorescence polarization). By comparative analysis and inspection of 100 selections, the 2-methylindole emerged as important moiety for the recognition of the antigen.

A ligand that inhibits tankyrase-1 (IC_{50} = 290 nM, as measured by enzymatic PARylation assay) was derived from a three-building block library of 103200 members, based on a central diamine scaffold featuring two sets of carboxylic acids (240 x 230). [125]

GlaxoSmithKline (GSK) identified a 2 nM antagonist for neurokinin 3 through screening a 41 million members single pharmacophore library, synthesized using three steps of split-and-pool (64 hydroxy- and amino acids, 854 and 758 amines respectively. [126] Each set of building blocks was displayed on a central 1,3,5- triazine scaffold. Noticeably, triazine-based libraries have also yielded high-affinity ligands to a disintegrin and metalloproteinase with thrombospondin motifs (ADAMTS4), [127] Aurora A kinase and p38 mitogen-activated protein kinase (MAPK), indicating the versatility of this chemical structure for library design. [86]

DECL chemistry has also been exploited to generate libraries based on macrocyclic peptide scaffolds which can be useful to address challenging targets (e.g., protein-protein interactions). [128, 129] David Liu and collaborators reported the identification of a 40 nM inhibitor of insulin-degrading enzyme (IDE) by using a library of 256000 macrocycles synthesized using a DNA-templated synthesis approach. [130] Similarly, scientists at Ensemble isolated XIAP inhibitors (X-linked inhibitor of apoptosis protein) with IC_{50} = 140 nM from a library of 16000 cyclic peptides. [131]

Li *et al.* [132] reported the synthesis of a single pharmacophore library based on a fixed macrocyclic scaffold with antiparallel *β*-sheets, which displayed three sets of building blocks. Specific binders could be isolated against a variety of proteins (e.g., carbonic anhydrase IX, horseradish peroxidase, tankyrase I, human serum albumin, alpha-1-acid glycoprotein, calmodulin, prostate-specific antigen and tumor necrosis factor).

Scientists at *Vipergen* have described the synthesis of a linear library of >12 million compounds by using Yoctoreactor® technology and involving three series of acylation, reductive amination and urea formation steps. The library delivered a MAPK14 inhibitor in the single-digit nanomolar K_d range, as assessed in cellular assays. [133]

Dual pharmacophore libraries based on self-assembling technology have also been very productive. ESAC library allowed the identification of a 190 nM binder against alfa-1-acid glycoprotein (AGP), thanks to the synergistic action of two building blocks. [110]. It is worth to notice that the individual fragments did not exhibit any activity when singularly assayed by fluorescence polarization and isothermal titration calorimetry (ITC), suggesting the importance of chelate-binding effects for molecular recognition.

1.3 SCREENING METHODOLOGIES IN DNA-ENCODED CHEMISTRY

Various screening methodologies have been implemented to select and identify binders to a specific protein of interest using DNA-encoded chemical libraries. [134]. In analogy with phase display technologies, large collections of encoded compounds can be selected using affinity selection procedures (Figure 1.14).

In antibody phage display selections, a library of antibodies is incubated with a target protein of interest, which is immobilized on a solid-phase. Selective binders are captured on the affinity support, whereas a large fraction of non-binders is washed away. The selected phage particles are amplified by bacterial infection and subsequent amplification to generate a secondary library, which can be screened in the next panning. As an alternative, infected bacteria can be directly plated and screened (Figure 1.14A). [69]

DECL libraries can be screened using a similar procedure based on affinity capture procedures. After elution of putative binders, the resulting DNA barcodes can be amplified by PCR, followed by high-throughput DNA sequencing (Figure 1.14B).

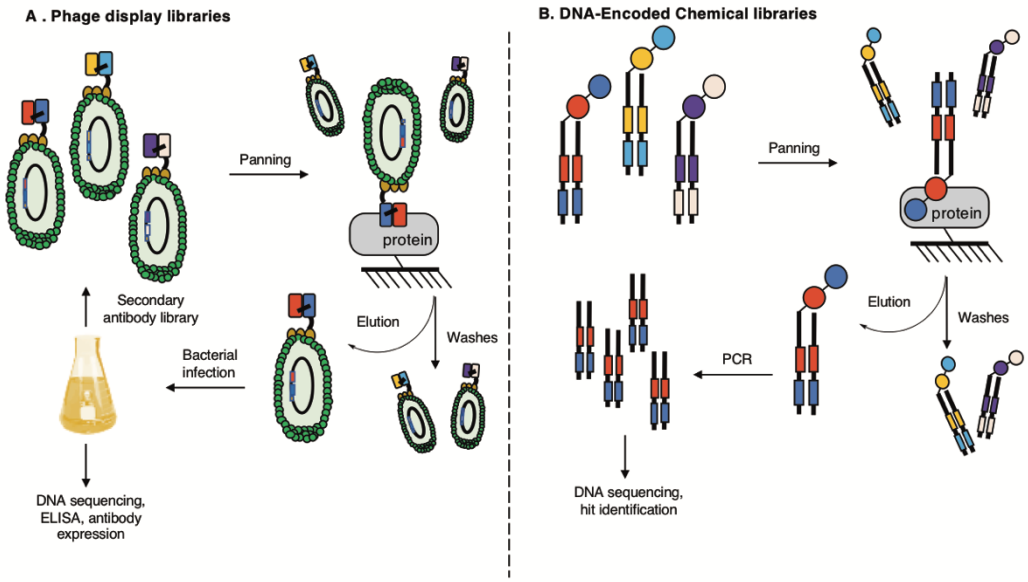


Figure 1.14. Conventional affinity selection strategies using phage display (A) and DNA-encoded chemical libraries (B). PCR = Polymerase chain reaction. ELISA = enzyme-linked immunosorbent assay.

Affinity based selection methodologies are generally performed using purified and immobilized targets of interest. However, various selection methodologies have been tested and implemented both in academia and industry setting, which can also occur in solution

(e.g., photocrosslinking, interaction-dependent PCR). The main advantage of the latter procedures is that the binding event is recorded in the same solution and does not involve the use of a third phase (e.g., solid), which makes mass transport events negligible.

1.3.1 Solid-Phase Affinity Selections

The most common selection methodology relies on the immobilization of a purified protein target of interest onto a solid support (e.g., referred also as "matrix" or "resin"). This process is also defined as "affinity chromatography". The protein can be immobilized by covalent/non-covalent binding onto the resin, and then incubated with the DECL.

Non-covalent binding usually requires solid supports that can be functionalized with specific functionalities (e.g., metal ions, proteins) to allow the capture of the target by using a "molecular tag". [135] The tag is typically expressed on the surface of a recombinant protein. Many affinity tags suitable for affinity chromatography have been recently developed (e.g., Streptavidin/Biotin-based tags, His-Tag). [136]

Biotinylated proteins can be purified by chromatography on streptavidin-functionalized solid supports. The interaction between streptavidin and biotin has an affinity constant of $K_d = 10^{-15}$ M and thus requires strong denaturing conditions to elute the protein (e.g., heat, urea, chaotropic agents). Proteins can be selectively biotinylated by incorporating a specific 15 amino acid peptide sequence ("Avi-Tag") which allows single-site biotinylation in correspondence to a specific lysine residue within the sequence. This prevents overbiotinylation of the protein, which can compromise the activity of the target. [137,138]

The polyhistidine affinity tag ("His-Tag" or His_6) consists of six histidine residues which are consecutively assembled and attached to either the C or N terminus of a recombinant protein. [139] Histidine is an amino acid which can form coordination bonds with various bivalent transition metals (e.g., Co^{2+}, Ni^{2+}, Zn^{2+}, Fe^{3+}). In a typical setting, the metal ions are associated with a specific stationary-phase matrix (e.g., nitriloacetic acid, iminodiacetic agarose), which coordinates the metal ion. [140] As the affinity matrix does not involve the use of a protein, ion-metal affinity chromatography (IMAC) procedures are amenable to denaturing agents (e.g., ionic detergents, weak reducing agents). The elution step is typically carried out by using imidazole solution, which competes with the His-Tag for binding to the metal-charged resin.

Covalent binding involves the direct linkage of the target to the resin by using reactive functional groups on the protein (e.g., primary amines, sulfhydryles, aldehydes, carboxylic acids) and activated solid-phase matrixes (e.g., alkyl halides). It is typically an irreversible covalent interaction (Figure 1.15). [141, 142]

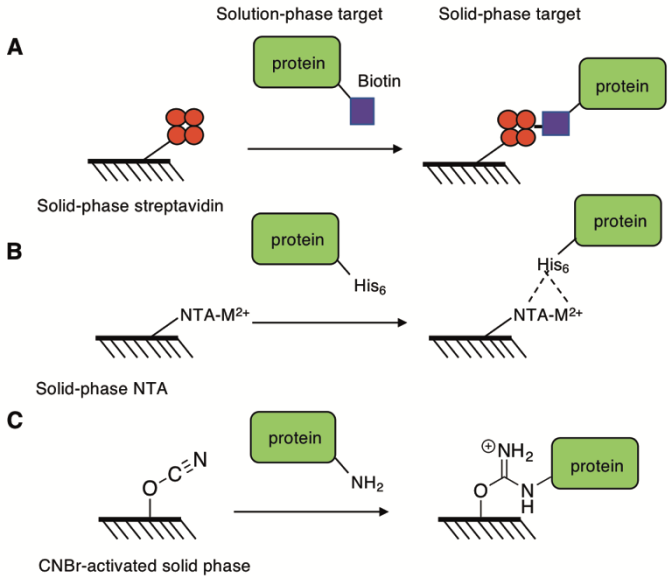


Figure 1.15. Example of supports used in solid-phase affinity selections. A. Solution-phase biotinylated targets can be conveniently transferred onto a solid support by association with coated streptavidin. B. Immobilized metal ions (e.g., Co^{2+}, Ni^{2+})- nitriloacetic acid (NTA) chelates allow the capture of polyhistidine (His_6)-tagged protein by complex formation. C. Solid supports activated with CNBr allow the covalent capture of proteins in solution by formation of isourea derivatives.

Various solid supports have been tested for affinity chromatography (e.g., sepharose beads, resin columns, magnetic beads, chromatography resin tips). In 2006, Dumelin *et al.* described a screening procedure based on the use of sepharose beads which were functionalized with streptavidin, for the screening of biotinylated antigens. [143]. Target proteins can be also covalently functionalized with CnBr-activated sepharose beads for covalent functionalization [94, 95].

Over the past decade, screening protocols have changed, and sepharose beads were replaced by magnetic beads. [144] Magnetic beads are produced as superparamagnetic iron oxide particles that are coated with inert silane derivatives. [145] They have a diameter which can range between 1-4 μm and are non-porous (Figure 1.16A). For this reason, they often exhibit less non-specific binding compared to porous supports. Moreover, due to their magnetic properties, the beads can be conveniently separated by using a magnetic rack, which facilitates the pipetting and decanting of samples. The use of magnetic beads as solid supports enabled the transition from manual to automated selections, which were first described by Decurtins *et al.* [144]

As an alternative approach, resin tips can be used for the screening of DNA-encoded chemical libraries. The biotech company *Phynexus, Inc* offers products for automated-high throughput protein purification (PhyTip® columns) which rely on the principle of Dual Flow Chromatography (DFC) (Figure 1.16B) [146].

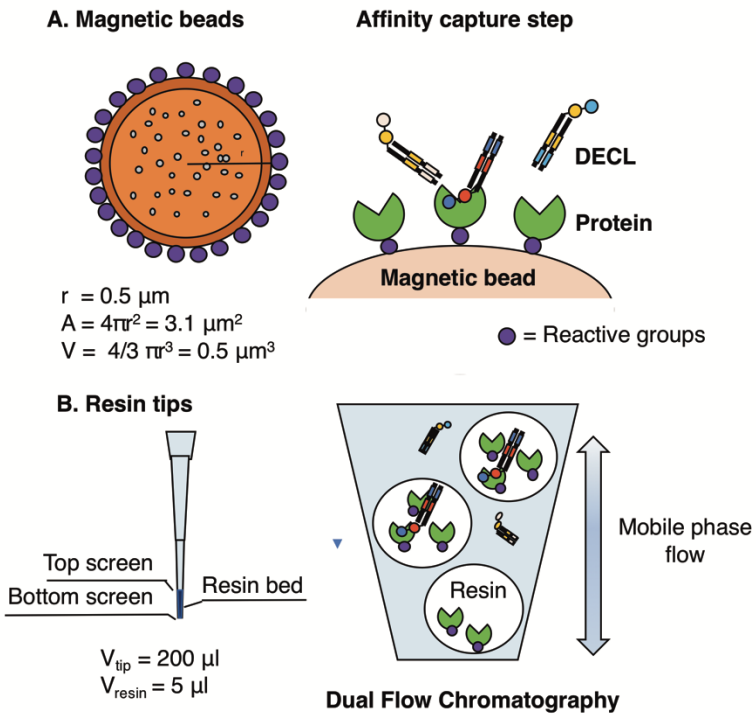


Figure 1.16. A. Graphic representation of the magnetic beads ("Dynabeads") used in solid-phase affinity selection experiments. The beads are composed by a porous polystyrene matrix (orange circle), which is filled with magnetic iron oxide particles (gray dots). The external polymeric coating (depicted in brown) can be functionalized by introducing reactive moieties, used to bind specific biomolecules. The binding events between DECLs and proteins occur on the surface of the bead. B. Resin tips used for affinity chromatography. The bottom of each tip is functionalized with a resin bed, which is contained by a top and bottom screen of woven mesh. In analogy with the magnetic beads, the resin can be functionalized with various reactive groups. The purification of biomacromolecules relies on the dual flow chromatography principle.

Solution-phase biomacromolecules can be purified by using a bi-directional flow (back-and-forth block) of a mobile phase across a stationary phase. The stationary phase is at the end of the pipette tip, where there is a column bed contained by a woven mesh. The mobile

phase is continuously aspirated and ejected in each step (sample loading, column washing and elution). The columns are miniaturized and operate in parallel up to 96 with the use of automated liquid handlers. Thanks to the dual flow system, the number of interactions per time is optimized (there are multiple chances of interaction) and the time is sufficient to drive the equilibrium of sample and column to completion (no effect of slow capture kinetics). For this reason, the interactions between analyte, solid and mobile phase are only dependent on the column and mobile phase (Figure 1.16).

1.3.2 Solution-Phase Affinity Selections

Solution-phase screening methodologies allow the identification of hits from DECLs without the implementation of a solid-phase affinity capture step before library selection. [85] Solid-supported proteins may display altered binding capacity or even the complete loss of activity in comparison with their soluble forms. Steric hindrance can prevent the target protein from being successfully screened, depending on the target orientation and site of binding. Moreover, many complex targets (e.g., targets for modulating protein-protein interactions, GPCRs, membrane proteins and live cells) have to be screened in their native environment. In response, few strategies have been described for the selection of protein directly in the liquid-phase (Figure 1.17).

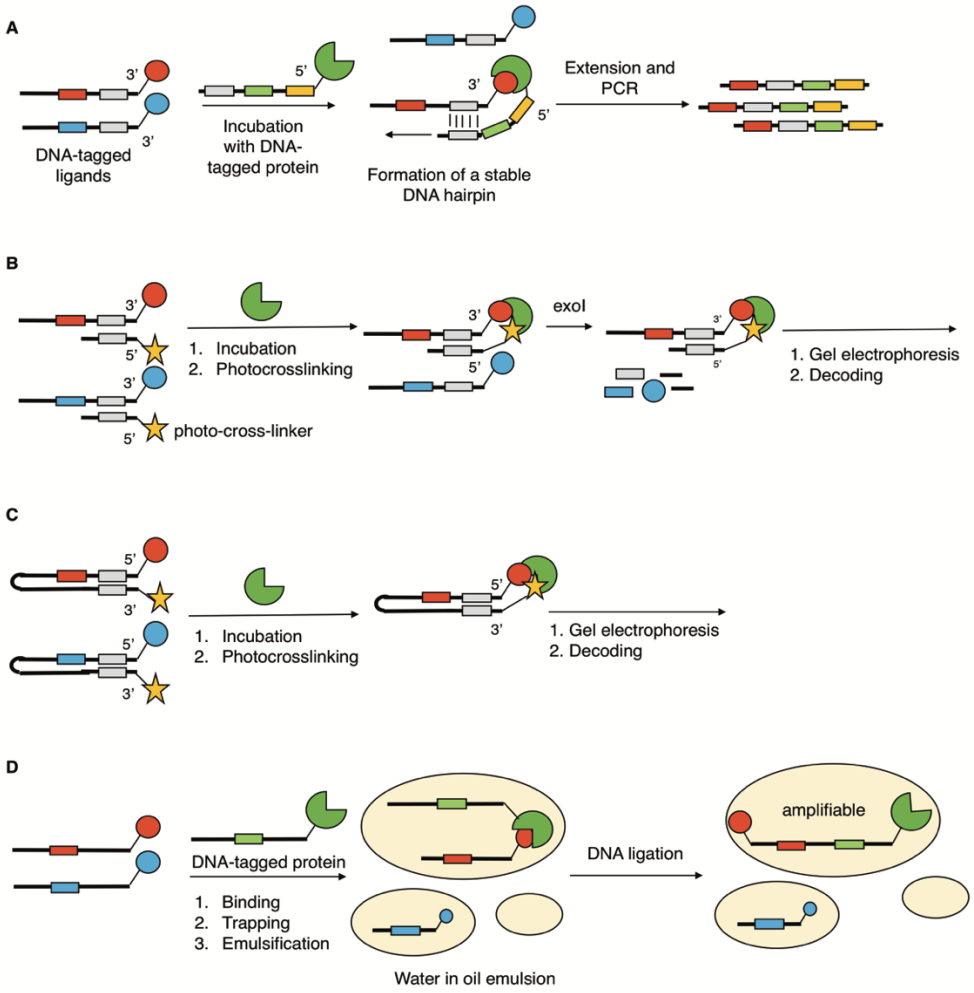


Figure 1.17. Overview of the solution-phase affinity selection methodologies used for DECL screenings. A. Interaction-dependent PCR (IDPCR) approach. The DNA-tagged ligands are mixed with a target protein functionalized with a DNA oligonucleotide. The resulting interaction allows the formation of a stable hairpin, which and allows the transfer of the coding region on the same DNA strand by PCR extension. B. DNA photoaffinity labeling (DPAL) approach. The DNA-tagged ligands are hybridized with a short oligonucleotide, which carries a photoreactive group. Upon UV irradiation, a stable protein-ligand complex is formed. The binding pair protects the corresponding DNA tags from ExoI digestion. C. Ligate-crosslink-purify approach. Single-stranded DECLs are functionalized with a DNA-photocrosslinker at the end distal to the small molecule. Upon irradiation, the resulting complex can be purified by gel electrophoresis and analysed by DNA sequencing. D. Binding Trap Enrichment (BTE) approach implemented by Vipergen. Water oil-emulsion can be used to trap the resulting protein-ligand complexes upon binding events. The DNA-tagged protein and the DNA-tagged ligands are joined by ligation, which irreversibly records the binding information. PCR = Polymerase chain reaction.

David Liu *et al.* [147] proposed and implemented the use of interaction-dependent polymerase chain-reaction (IDPCR) for the simultaneous evaluation of all the ligand-target binding pairs from a single solution containing libraries of ligands and targets. Both ligands and targets are tagged with an encoding DNA oligonucleotide strand. Upon binding, the two strands can hybridize with each other and form a pseudo intramolecular hairpin. The corresponding intermolecular process (DNA duplex formation due to non-specific hybridization) would not occur since the complementary regions are too short to hybridize. The resulting hairpin serves as a starting point for primer extension and subsequent PCR amplification. Thanks to IDPCR, the procedure is highly sensitive, and it can be performed in a few hours. Similarly, the same authors described the determination of DECL ligand/target interactions using unpurified protein cell lysates (IDUP). [148] The protein cell lysate can be functionalized with DNA oligonucleotide tags by using DNA-barcoded target specific antibodies or SNAP-tag moieties. Using this procedure, Liu and co-workers reported the discovery of a novel micromolar inhibitor of MAP2K6 from a focused DNA conjugate library and a cytosolic mixture of 236 SNAP-tagged and DNA-barcoded human kinase proteins. [149]

Alternative solution-phase methodologies rely on the use of crosslinking reactions and hairpin formation to stabilize protein-ligand complexes and facilitate their identification. [150, 151] Zhao and co-workers have developed a novel method for screening non-immobilized and purified proteins, which is based on DNA photoaffinity labeling (DPAL) [152]. The authors described the use of a short (6-8 mer) DNA strand which bears a photoreactive azidophenyl group at 5'-extremity. The strand can hybridize at the primer binding site region of the DNA conjugate strand. Upon binding to protein and photoirradiation, a stable covalent hairpin-like structure is formed, which protects the binding pair from being digested by treatment with ExoI exonuclease at 37 °C. The encoding DNA tags of each non-binder is destroyed, whereas ligand-protein complexes survive and can be decoded or subjected to iterated selections.

Since this method is limited to DECLs with small molecules conjugated at the 3' end of DNA oligonucleotides, Shi *et al.* [153] studied the application of a novel selection method based on "ligate–crosslink–purify". The DNA-encoded library is enzymatically ligated with a terminal DNA-photocrosslinker conjugate (PC-DNA) at the end distal to the small molecule. Since PC-DNA contains a small sequence complementary to the primer binding site, it can form a stable hairpin structure (T_m = 54.5 °C) which allows the crosslinker to be close to the small molecule. Upon irradiation, a stable covalent protein-ligand complex is formed, which can be purified by gel electrophoresis before PCR amplification and hit deconvolution.

Since this method requires further steps for library functionalization (e.g., enzymatic ligation), the same group reported the use of a polymerase-based selection methodology and target-directed photocrosslinking for selecting DECLs against non-immobilized targets. [154]

Other solution-phase methods rely on the use of in vitro compartmentalization (IVC) for the isolation of ligands which bind proteins. IVC enables the generation of cell-like compartments in vitro, which are designed to contain no more than one gene. [155] An emulsion-based technology (Binding Trap Enrichment, BTE®) was developed by *Vipergen*, which allows the formation of cell-like compartments in vitro for the screening of Yoctoreactor® libraries (yR). [103, 104, 133] Water-in-oil emulsions can be used to isolate individual protein-ligand complexes, where proteins and ligands are tagged with a DNA oligonucleotide. The binding information is "recorded" by enzymatic ligation in a droplet. Since binding events can also occur between non-specific binders and proteins (which are randomly trapped within the droplet) high dilution of the selection components is required to reduce the selection noise.

1.3.3 Recent Advances in Screening Methodologies

1.3.3.1 *Selectivity binding assays*

DECLs are characterized by a high rate of "promiscuous" compounds, which can act as binders for diverse proteins belonging to the same family (e.g., kinases) or dissimilar. These small molecules are often referred to as PAINS (Pan-assay Interference Compounds) [156, 157] and can give false-positive results in ligand-binding assay screening formats. Typically, a certain number of functional groups are shared by many PAINS. [158, 159] In a binding screen, these compounds can be washed away by simply "panning" the library against a large excess of diverse untagged proteins or including it in the screening assay. These assessments can also be performed during the hit validation step, through resynthesis of the putative ligand and assessment of its binding properties against other proteins.

Targets and off-targets can be screened in the same assay format by using differential recognition tags (e.g., orthogonal tags in affinity, fluorescence emission). [160] A procedure has been described for OBOC ("one-bead-one-compound") libraries, which relies on the use of a two-color screening strategy for the selective identification of small molecules that can preferentially bind to specific antigen-binding sites of antibodies rich in the serum of patients with active tuberculosis ("target"). [161] After library panning and incubation with the targets, library beads that exhibited a high level of red fluorescence (binding to the target) were sorted using FACS (Figure 1.18).

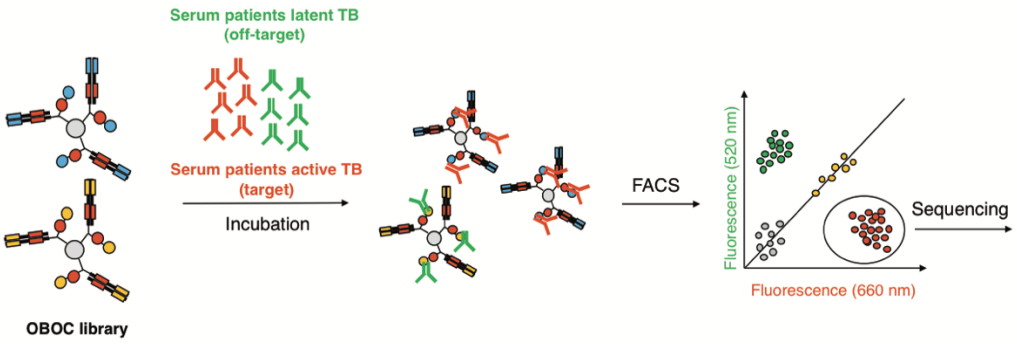


Figure 1.18. Schematic representation of two-color screening strategy used to identify ligands from an OBOC library. The protein target and the off-target are labeled in red and green respectively. The two targets are mixed and incubated with the library. Upon binding, the library beads are sorted by using FACS, depending on the fluorescence color. Beads which exhibit red fluorescence are collected and analysed by sequencing. OBOC = one-bead-one-compound; TB = tuberculosis.

1.3.3.2 *Cell-based screening assays*

Certain protein targets require screenings in their native cell-environment. This applies to proteins that cannot be recombinantly expressed and purified in their active form (e.g., GPCRs, ion channels, receptor tyrosine kinases). Wu *et al.* [162] described selections against immobilized NK3 (neurokinin-3)-overexpressing cells using a DECL of 4.1 x 10^7 members. For this purpose, HEK293-derived cells were transduced with recombinant BacMam viruses encoding the targets of interest. After library incubation at 37 °C, cells were washed and library members eluted by heating the cells at 95 °C. A limitation of this method is the high expression of the protein target, which can be problematic if the protein is toxic at high concentrations (e.g., ion channels) or for the discovery of low-affinity ligands. Bradley and co-workers also described the selection of PNA-encoded libraries against chemokine receptors and integrins directly on live cells. [163, 164]

Cai *et al.* [165] studied the application of covalent crosslinking to enable DECL selections for both membrane and intracellular targets. A cyclic cell-penetrating peptide (cCPP12) developed by Pei and co-workers [166] was attached to the extremity of the library to allow cellular uptake. The authors assessed the amount of amplifiable DNA remaining in cells by qPCR analysis of cell lysates (1% and 11% respectively for 60-bp DNA-cPP12 and 140-bp DNA-cPP12) and the cytosolic delivery of DNA to protein targets by the HaloTag-based chloroalkane penetration assay (CAPA) [167, 168]. Upon internalization of the library, the transient interaction between a cytosolic protein and a DNA-tagged ligand is trapped by covalent cross-linking. After lysis of the cells, the resulting complexes are captured on beads and subsequently identified by DNA sequencing (Figure 1.19).

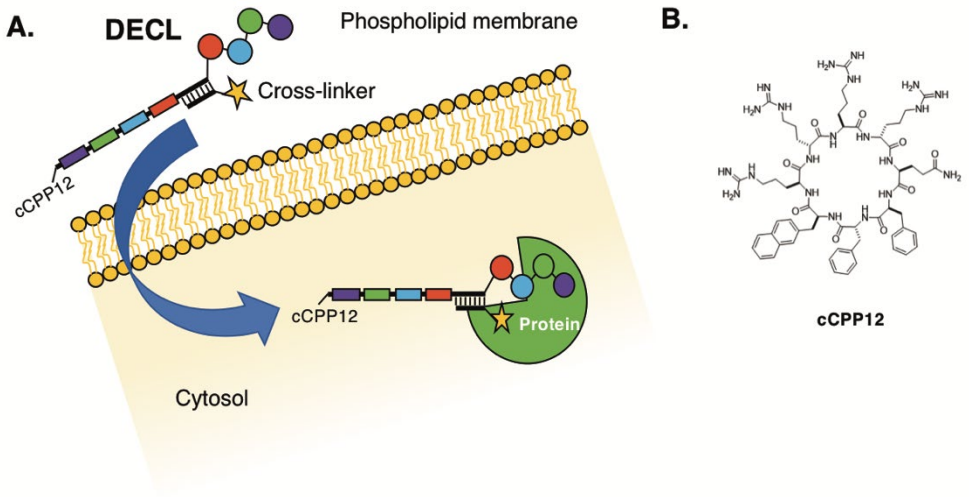


Figure 1.19. A. Graphical representation of the screening methodology on cells as described by *Cai et al.* [165]. This approach relies on the use of cell-penetrating peptides (cCPPs) to allow the delivery of the library across the cell membrane. Upon internalization and binding to a cytosolic ligand, the protein-ligand complex is stabilized by chemical crosslinking. B. Chemical structure of the cell-penetrating peptide used for the experiments (cCPP12). cCPPs generally display positively charged amino acids (e.g., lysine, arginine) or alternating patterns of polar and non-polar groups to allow cell internalization.

1.3.3.3 *Functional screening assays*

In contrast with affinity-based selection methodologies, other assay formats ("functional screening assays"), which can provide complementary information to affinity data (e.g., activity) have been described. Compounds are scored as "hit" depending not only on their binding properties but also on their biological function (assessed in vitro or in cells). Screening conditions can be adjusted to demand a high level of selectivity.

This approach has been applied to the screening of OBOC ("one-bead-one-compound") DECLs using a droplet-based microfluidic system. [169]. Each library bead is incorporated with an enzyme (ATX, phosphodiesterase) and a fluorogenic ATX substrate into a picoliter droplet. By using the microfluidic system, library beads are directed toward a laser, and a certain amount of compound is released upon cleavage of a photochemical linker. Due to the enzymatic reaction, the fluorescence intensity arises if the compound displayed on the single bead is active, thus enabling the sorting of the beads depending on the intensity

threshold. The encoded tags of the beads are then amplified and sequenced to reveal the identity of the putative hits.

Similarly, MacConnell *et al.* described the implementation of a microfluidic circuit that enables functional screening assays using DNA-encoded compounds on beads. The device combines assay reagents and carries out library beads distribution into picoliter-scale droplets. Then, after photochemical cleavage from the beads, the dosed droplets are incubated with the target and sorted based on their fluorescence intensity. [170]

1.3.4 DNA Decoding by Next-Generation Sequencing

The identification of ligands using DECL technology requires the implementation of appropriate decoding strategies after library selection. In the early days of DECL research work, the main aim was to demonstrate that the DNA barcoding was a feasible strategy for the encoding. Since libraries were much smaller, traditional sequencing techniques (e.g., Sanger sequencing) were employed to check the decoding efficiency. [108, 171]

The means for library selection decoding have improved over the last decade. Microarray-based methods [108] were replaced by high-throughput sequencing (NGS – "next-generation sequencing") [95], which had a great impact on decoding efficiency. Modern strategies rely on NGS technologies, such as Roche's 454 Genome Sequencer FLX system (formerly 454 Life Sciences) or Illumina's HiSeq 2500 System (Solexa).

Next-generation sequencing techniques rely on multiple PCR cycles to allow the amplification of the signal. The process can be performed by using emulsion PCR or bridge PCR techniques. In emulsion PCR, individual DNA molecules are isolated as aqueous bubbles within an oil phase along with primer-coated beads. The polymerase chain reaction amplifies the genetic information from each DNA molecule and "coats" the beads with several DNA copies. In bridge PCR, sequencing templates are immobilized on a proprietary flow cell surface which is coated with primers complementary to the DNA library fragments. Upon addition of the polymerase and nucleotides, double-stranded bridged structures are formed, which are subsequently denatured to form clustered single-stranded DNA fragments. Several iterations of the process lead to localized clusters of DNA strands on the solid support (Figure 1.20).

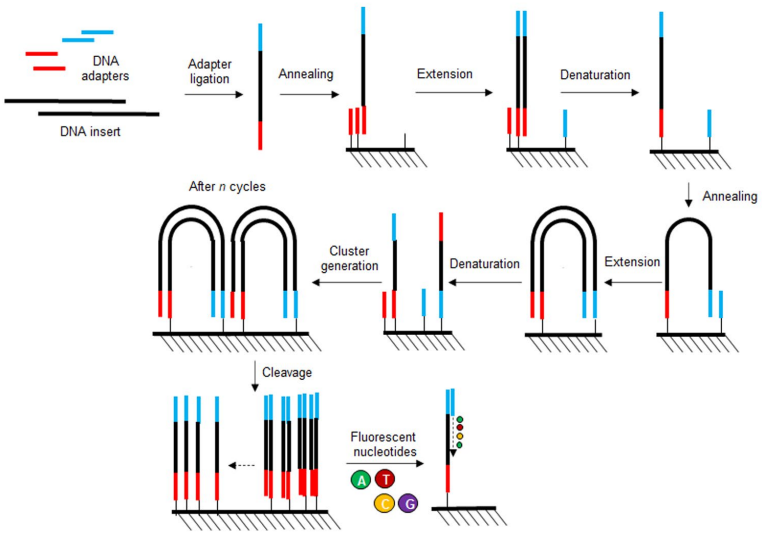


Figure 1.20. Schematic workflow of the Illumina high-throughput sequencing technology. The DNA strands are ligated to single DNA adapters and immobilized on a flow lane by hybridization to solid-supported DNA anticodons. After primer extension and denaturation, the resulting single-stranded DNA oligonucleotides anneal to a second DNA anticodon and form bridge structures, which can be further extended. Iteration of the clonal amplification process generates clusters of bridged structures, which are cleaved and denatured to form single-stranded DNA strands. The DNA sequencing step uses nucleotides bearing cleavable fluorescent groups at the 3'-OH end. Each nucleotide is added one-by-one, and each fluorescence signal is recorded to allow the identification of the original sequence.

The sequencing step typically relies on the "sequencing by synthesis", where a complementary strand is synthesized using a polymerase enzyme and fluorescently tagged deoxynucleotides (dNTPs). Since the fluorescent group acts as a reversible terminator, only one of the four bases (A, T, C, G) can be added per round. After each cycle, the machine identifies the added base by measuring the fluorescence emission. Once the color is recorded, the fluorophore is washed away, and the process is repeated. [172]

In 2008, the Neri group reported the use of NGS techniques for the deconvolution of DECLs using the Roche's 454 Genome Sequencer. [95] The authors described the construction of a single pharmacophore DNA-encoded chemical library containing 4000 compounds (DEL4000). The library was screened against MMP3, streptavidin and IgG and high-throughput sequencing decoding yielded preferential enrichment of binding molecules.

In 2016, Decurtins *et al.* have described a highly detailed and systematic procedure suitable for the analysis of selection outputs using Illumina high-throughput sequencing analysis ("Illumina's HighSeq 2500 system"). [144]

The procedure can be applied to single pharmacophore and dual pharmacophore libraries (ESAC). After affinity selection, the recovered DNA fragments can be amplified by a two-rounded PCR amplification. The first PCR (PCR1) amplifies the selected encoded library members separately for each selection. In order to distinguish individual affinity selections, two sets of primers ("Illumina PCR1a" and "Illumina PCR1b") are introduced, allowing the parallel high-throughput drug screening (HTDS) of different selection experiments on the same Illumina flow lane. The number of selections that can be pooled together and analyzed depends on the library size and the desired counts per selections. Typically, the Illumina HiSeq 2500 system achieves read-lengths of 150 bp and delivers 150 million of readable sequences per flow lane, which is sufficient for the analysis of library sizes between 10^5–10^8 library members. After PCR purification of the PCR reactions (using commercial PCR purification kits) the amplified fragments can be used as templates for the second round of PCR (PCR2). This amplification step introduces two other primers "Illumina PCR2a" and Illumina PCR2b, which are required for the Illumina HiSeq device as recognition adapters. [144]

Nowadays, this technology is used for the decoding of DECL selections by the Neri Group and Philochem.

1.3.5 Data Analysis

HTDS data are delivered as very large raw data files (up to 50–60 Gb per Illumina flow lane), which generally require further data analysis. A C^{++} evaluation program was developed by Neri *et al.* for the generation of output data files that can be and analyzed by using a MATLAB script. [144] The final selection output is a "selection fingerprint" which enables the tridimensional visualization of the "counts" associated to each library member and the determination of the "relative enrichment factor" above the library members. Typically, these fingerprints are cross-compared with the ones obtained with the naïve library (where no compound should be enriched) and after selection on empty beads (useful to determine sticky compounds and off-target ligands which may give a positive signal).

Traditionally, the analysis of DECL selections involves the visualization of a two- or three-dimensional scatter plot ("cubic view") where each point represents a unique library member. The sequencing counts can be visualized on an orthogonal axis. The enrichment of n-synthons can indicate the presence of SARs, binding models or can be associated with the presence of truncated products.

However, there is no general standard of how enrichment fold should be calculated and reported. Generally, enrichment factors are evaluated dividing the number of counts calculated for each library member to the mean counts per library member ("noise"). [144] This enables an estimation of the enrichment per each building block. As an alternative, Buller *et al.* have described the use of a negative binomial distribution to model selection counts and determine p values for enriched compounds. [94, 173, 174]

Satz *et al.* described a simple mathematical method that can be used as a basis for computational simulations of DECL selections. [175, 176] Equilibrium association constants can be derived and assessed from sequencing counts, which take the yield of the chemical synthesis into account. The authors proposed plotting count data of selected library members against various protein concentrations from multiple selection experiments to better estimate affinity constant.

Faver and co-workers [177] proposed the use of a normalized z-score enrichment metric using a binomial distribution model. This metric allows a more quantitative measure of DECL selection data across multiple experiments and allows direct quantitative comparisons of enrichment of n-synthons in DECLs.

1.4 AIMS AND STRUCTURE OF THE THESIS

As previously described in the previous paragraphs, DECLs can be screened using various selection methodologies. In most cases, libraries are screened using solid affinity capture procedures, which have been adapted from antibody phage display technology. Although the performance (in terms of recovery and selectivity) of phage display libraries has been systematically investigated in model systems, the performance of individual steps (e.g., affinity capture, washes, elution) in the selection procedures has been scarcely assessed so far.

The first aim of the research (Paragraph 2.1) was to develop an analytical methodology based on quantitative PCR for the analysis of selection experiments, in order to evaluate the efficiency of classical affinity capture methodologies performed on various solid-phase supports (e.g., magnetic beads, resin tips). Selection experiments were performed using a model target, carbonic anhydrase IX (CAIX) using various DNA-tagged ligands (corresponding to reported sulfonamide-based small molecule ligands against CAIX) and a single pharmacophore DECL.

The second aim of the thesis was to implement an efficient DECL solution-phase screening methodology based on a photocrosslinking strategy to enhance the discrimination of high and medium affinity ligands (e.g., K_d in the micromolar range; Paragraph 2.2) A systematic evaluation of experimental parameters revealed conditions that are suited for library screening and cross-compared with affinity capture procedures.

Affinity selection data rely not only on the affinity selection performance but also on the quality of the library, which can be critical for downstream applications (e.g., PCR amplification). An on-going project is introduced in Paragraph 2.3., where proof-of-concept experiments are performed to demonstrate the feasibility of a solid-phase DECL synthesized on TentaGel® resin beads.

2. RESULTS AND DISCUSSION

2.1 QUANTITATIVE ASSESSMENT OF AFFINITY SELECTION PERFORMANCE BY USING DNA-ENCODED CHEMICAL LIBRARIES

This paragraph is adapted with permission from Sannino, A.; Gabriele, E.; Bigatti, M.; Mulatto, S.; Piazzi, J.; Scheuermann, J.; Neri, D.; Donckele, E. J.; Samain, F. Quantitative Assessment of Affinity Selection Performance by Using DNA-Encoded Chemical Libraries *ChemBioChem* **2019**, *20*, 955–962. [178] Copyright © *2019 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.*

My contribution to this work was first to develop an analytical methodology based on qPCR for the investigation of the efficiency of selection procedures by using DNA-encoded chemical libraries. For this purpose, I have designed and performed all the experiments described in paragraph 2.1.2, which included: 1) synthesis of DNA-tagged sulfonamide ligands; 2) model selection experiments against polyhistidine-tagged and biotinylated carbonic anhydrase IX using magnetic beads and resin tips; 3) analytical method development of qPCR methodologies for the analysis of model selection outputs; 4) qPCR-based quantification, data analysis and investigation of the impact of different selection parameters on the screening procedures. In the second part of the project, I have performed the screening experiments in multiple conditions using a single pharmacophore DECL available in-house. I have developed a novel analytical procedure for the direct analysis of DECL selection outputs by PCR and carried out the quantification experiments, data analysis and evaluation of high-throughput sequencing results using a proprietary MATLAB script (developed at Philogen).

2.1.1 Introduction

As seen in the previous paragraphs, DECLs are large collections of organic compounds, and less expensive than conventional libraries used for HTS screenings. For this reason, DECLs are increasingly being used to discover binders against protein targets of pharmaceutical and biological interest. [Paragraph 1.2, 85, 86, 179–180]

In most cases, DECLs are screened using affinity capture procedures, which have been adapted from similar work, previously performed using antibody phage display libraries (Paragraph 1.3). Most commonly, a protein target of interest is immobilized onto magnetic beads or on resins. Libraries are incubated with the affinity support, and preferential binders are recovered after a series of washing steps. [144]

The performance (in terms of recovery and selectivity) of antibody phage display library selections has been systematically investigated in model systems. [66, 70, 181– 185] Those studies are facilitated by the fact that individual phage particles can be "counted" by infecting *E.coli* bacteria, leading to individual colonies on selective plates. [181] These quantification procedures have a very broad dynamic range (from individual phages up to more than 10^{13} phage particles). Surprisingly, model selection experiments on phage antibodies have revealed that the conventional affinity capture procedures are not very efficient.

For example, Mutuberria and co-workers [181] reported the recovery of as little as 10^5 phage antibodies from an input of 6 x 10^{11} particles (corresponding to a 0.0021 % recovery rate). These possible limitations are often compensated by the use of multiple copies of library members in selection experiments. For example, it is customary to screen libraries containing billions of different phage antibodies, starting from a total population of $>10^{13}$ phage particles (e.g., > 10^4 copies for each antibody clone). [181, 186, 187] Interestingly, the use of library subsets leads to suboptimal recovery of antibody clones, which typically have a lower affinity to the target. [188]

The efficiency of DECL selection procedures has been scarcely investigated so far. Direct quantification of individual DNA barcodes is possible by qPCR techniques, [153, 189– 193] but this technology has a limited sensitivity, due to the unspecific amplification of oligomeric primer structures. [190] Alternatively, investigators have studied the performance of selection procedures by inspecting the features of library fingerprints from DNA sequencing results, comparing DECLs before and after selections. [95, 98, 134, 175 191, 194].

In this work, we have used ligands specific to carbonic anhydrase IX (CAIX, a tumor-associated antigen) as tools for the quantitative evaluation of selection procedures, which are commonly used in academia and industry. [85, 144] Carbonic anhydrases (CAs) represent an important family of zinc-containing metalloenzymes which catalyse the reversible formation of carbonic acid, starting from carbon dioxide and water ($CO_2 + H_2O \leftrightharpoons H^+ + HCO_3^-$). Their activity allows the regulation of acid-base equilibria in physiological systems. CAIX is a homodimeric accessible membrane protein, which is over-expressed in clear-cell renal cell carcinomas and at sites of hypoxia. [195–196] The most widely studied class of CAIX inhibitors are aromatic sulphonamides. The nitrogen atom of the sulphonamide forms a reversibly complex with the Zn^{2+} ion present in the catalytic pocket

and blocks the active site. Claudiu Supuran and others have described a large number of CAIX ligands, differing in binding affinity and specificity. [197] CAIX ligands have been used for molecular imaging applications, [198–200] for functional protein inhibition [201] and drug delivery. [202]

Ligands of different affinity were used both in model selections and in a library comprising 360000 members. Inspection of recovery rates and selectivity for different ligands at different concentrations has revealed that conventional affinity capture procedures can be very efficient for high affinity (Kd < 1 μM) binders, if multiple copies of individual compounds are used in selection experiments. However, inspection of DECL selections performed with copies of individual compounds ranging from 10^4 to 10^8 revealed a dramatic drop of compound recovery efficiency below 10^5 copies per library member used as input. Collectively, the findings of our study provide a rationale for the efficient execution of DECL selections and may stimulate research on novel screening procedures, which may be required when very large libraries (e.g., those containing billions of compounds) are used.

2.1.2 Results

Model selection experiments with pairs of DNA derivatives

We used human carbonic anhydrase IX (CAIX) as a model protein to evaluate the performance of DECL selections. This homodimeric protein is a target of pharmaceutical interest, for which several inhibitors have been described and characterized in terms of dissociation constants (K_d) and other biochemical parameters. [198, 199, 203, 204] Acetazolamide (AAZ) is a potent CAIX inhibitor with a reported K_d value of 13 nM, [198] whereas benzenesulfonamide-based ligands [*p*-sulfamoylbenzoic acid (SABA); 3-sulfamoylbenzoic acid (*m*-SABA)][205] display a reduced binding affinity (K_d = 216 nM [198] and 19 μM, respectively). (Figure 5.1, Paragraph 5.1)

The three CAIX ligands were chemically coupled to one amino-tagged DNA fragment (96 bp) and mixed with a different DNA fragment in 1:1 ratio and at various concentrations, in order to assess experimental recovery rates. (Figure 2.1) Ligand capture procedures were performed using both biotinylated CAIX on streptavidin-coated supports or polyhistidine-tagged protein preparations on metal-based supports, using suitable washing procedures. [144] After selection, the DNA fragments were eluted and quantified by qPCR (Figure 2.1).

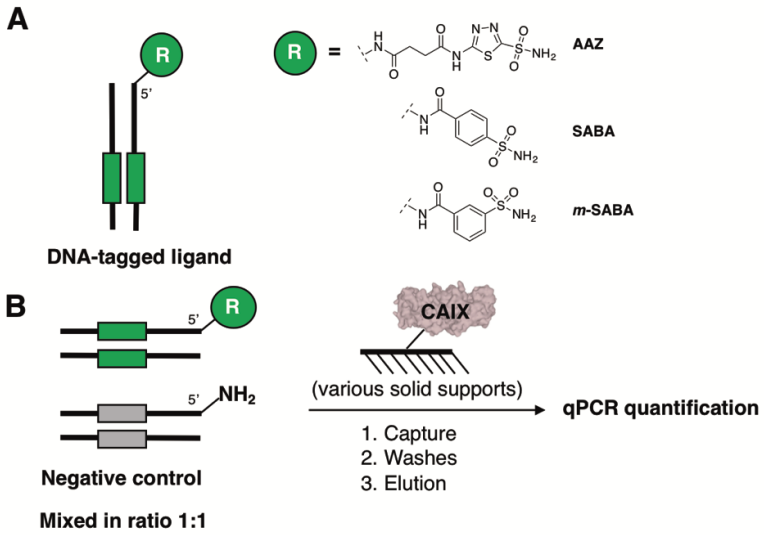


Figure 2.1. A. Schematic representation of DNA-tagged ligands directed against CAIX. R: AAZ, SABA, *m*-SABA. B. Setup of model selection procedures.

Figure 2.2 shows the results of six selection experiments (P) performed against biotinylated CAIX, immobilized on C1 streptavidin-coated magnetic beads or Phynexus tips as chromatographic supports. Both procedures have been often used in the past for DECL selections. Negative control experiments (NP) were performed on "empty" beads and resin tips, which had not been coated with CAIX. All experiments were carried out using identical amounts of DNA input, corresponding to 10^{11} copies of each molecule before selection (I). The DNA fragments were quantified by qPCR methodologies, with calibration curves that were established for each experiment (Figure 5.2, Paragraph 5.1).

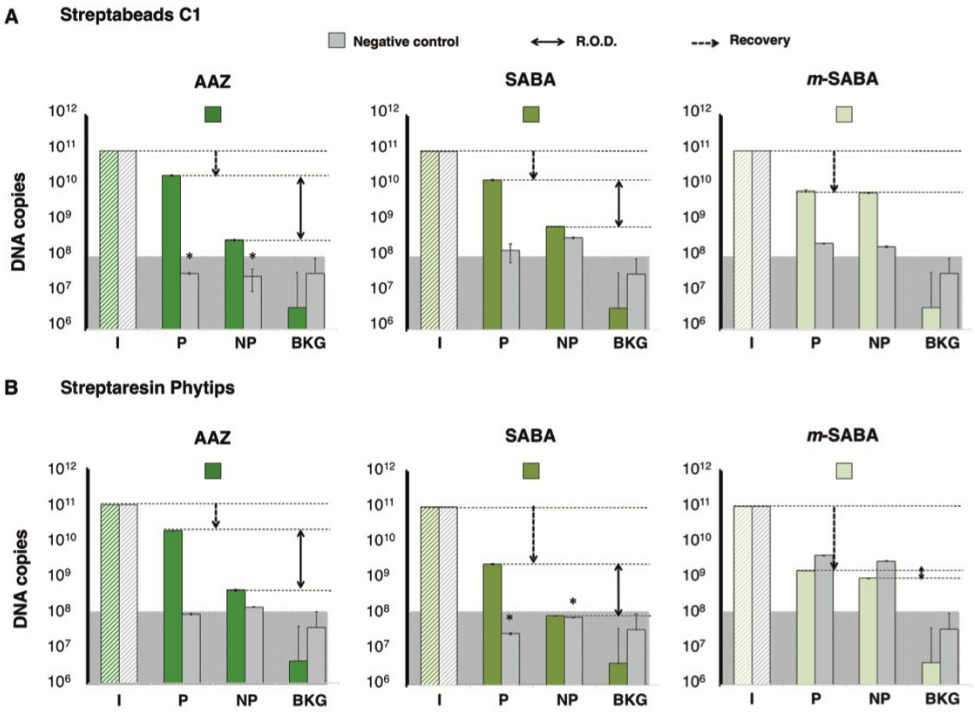
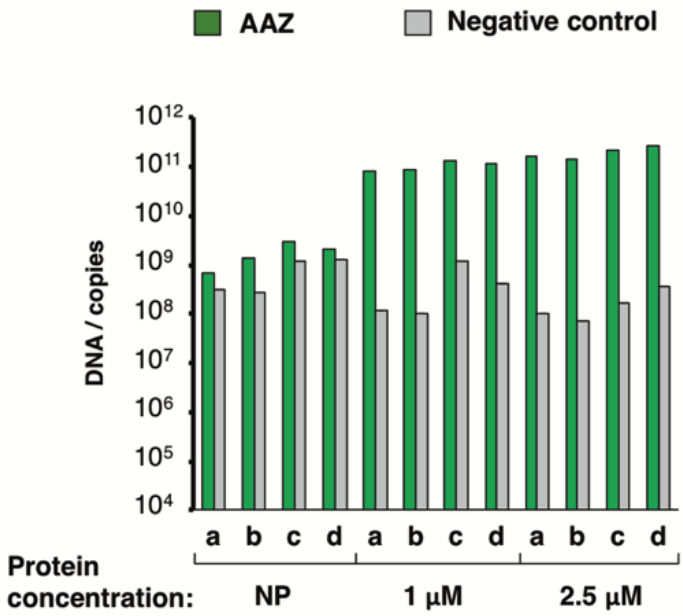


Figure 2.2. qPCR-based quantification of DNA-tagged CAIX ligands after model selections plotted on a logarithmic scale. For each capture experiment, 10^{11} copies of DNA were used (input, I). The results obtained for DNA fragments coupled to CAIX ligands are depicted in green, whereas negative control results associated with unmodified DNA structures were depicted in gray. A range of detection (R.O.D.) can be derived from the comparison of output DNA quantities in selections performed with (protein, P) or without (no protein, NP) immobilized CAIX. The lowest limit of detection, which was determined by the onset of a PCR signal associated with primer dimers formation, is also indicated (Background, BKG). To facilitate visualization of the quantification range, a gray area is plotted in each graph, which corresponds to the upper standard deviation bar for the BKG signal.

The lowest limit of quantification in our qPCR procedures was determined by the onset of a background noise (BKG), caused by the formation of primer dimers. An experimental range of detection (R.O.D.) was derived by the comparison of DNA output values in selections performed in the presence (P) or the absence (NP) of CAIX. [190]

The DNA derivatives of AAZ (e.g., the CAIX ligand with highest binding affinity) were eluted with a ~10% recovery rate, on both magnetic beads and Phynexus tips. In the experimental conditions used, the recovery was ~100-times more efficient, compared to negative control experiments, performed in the absence of immobilized CAIX (Figure 2.2). Selections with SABA derivatives were also efficient on magnetic beads, but the recovery rate dropped when Phynexus tips were used (Figure 2.2). Surprisingly, *m*-SABA (a micromolar CAIX binder) exhibited a less efficient recovery and no detectable enrichment compared to uncoated NP controls for both capture methods (Figure 2.2).

Similar findings were observed when polyhistidine-tagged CAIX was used as the target antigen. We first evaluated the recovery of DNA derivatives of AAZ in various experimental settings. (Figure 2.3) High recovery rates were always observed with AAZ on magnetic beads, and the results were confirmed by comparison with resin tips supports. (Figure 2.4). Interestingly, SABA derivatives were captured more efficiently on Phynexus tips.



Entry	Tween %	Elution buffer
a	0.01%	50 mM Na_3PO_4, 300 mM NaCl, 300 mM imidazole, 0.01% Tween-20, pH = 8
b	0.1%	50 mM Na_3PO_4, 300 mM NaCl, 300 mM imidazole, 0.01% Tween-20, pH = 8
c	0.01%	TRIS Buffer pH = 8.5. Heat at 95°C for 10 min
d	0.1%	TRIS Buffer pH = 8.5. Heat at 95°C for 10 min

Figure 2.3. qPCR quantification of the DNA-tagged AAZ (green) and negative control (gray) recovery after model selections against polyhistidine-tagged CAIX. The results are plotted in logarithmic scale. The experiments were performed changing different experimental conditions (e.g., protein concentration, quantity of Tween-20 in the wash solutions and elution method). NP: no protein selection.

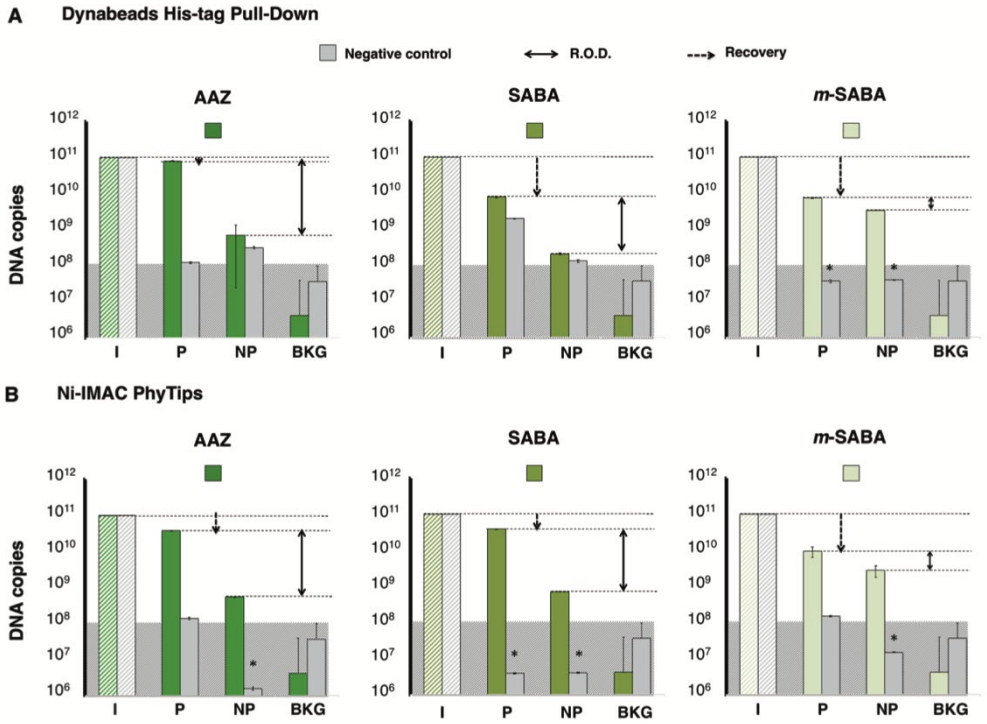


Figure 2.4. qPCR-based quantification of DNA-tagged CAIX ligands after model selections against polyhistidine-tagged CAIX.

These results indicate that high-affinity ligands are efficiently recovered in various experiments, while binders with moderate affinity may benefit from the execution of screening procedures in multiple conditions.

Individual fractions (flow-through, washes and elution) in the selection experiments with biotinylated CAIX were evaluated, in order to assess any loss of DNA. For all three CAIX ligands, most of the DNA (>90%) was lost in the flow-through or in the first wash (almost 10%), while a smaller portion of the input (~0.1%) was recovered in the following washing steps (Figure 2.5). These results may indicate a suboptimal contact between ligands and affinity capture support, in analogy to what had previously been reported for antibody phage selections. [181]

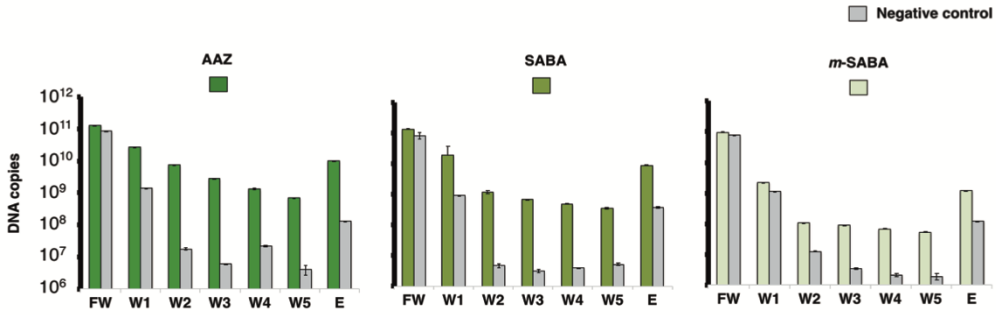


Figure 2.5. qPCR-based quantification of DNA-tagged CAIX ligands in individual fractions (flow-through, FW; washes, W1–W5; elution, E) of the selection experiments with biotinylated CAIX. Negative controls results are depicted in gray.

Selection experiments with DNA-encoded chemical libraries

The selection experiments described in the previous section provided information about the relative recovery and specificity for CAIX ligands of different affinity when the DNA derivatives were used with a large number of copies (10^{11}). In real libraries, the number of copies for individual compounds is typically much lower, and those molecules are diluted in a molar excess of other chemical structures. In order to gain insight on the efficiency of DECL selections in various experimental conditions, we screened a library, comprising 359652 members, which was based on a synthetic strategy previously described by our group (Figure 2.6). [98]

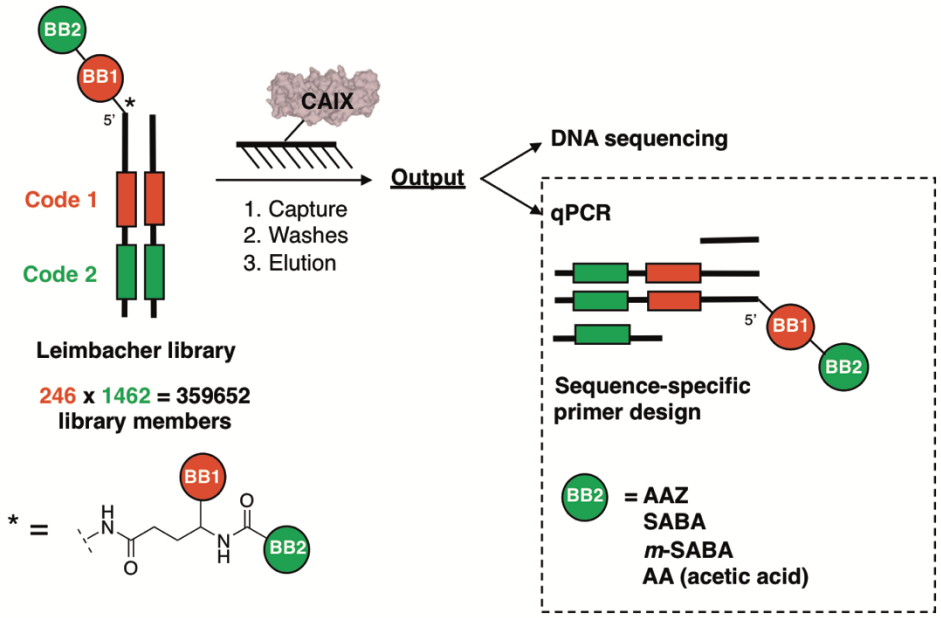


Figure 2.6. Schematic representation of a DECL, comprising 359652 members. Building blocks (BBs) AAZ, SABA, and *m*-SABA were incorporated in the second reaction step, thus facilitating the quantification of CAIX selections by both DNA sequencing and qPCR. One of the many building blocks in position two that could be considered as a negative control for the selection experiments was acetic acid (AA).

The library consisted of 246 amino acids which had subsequently been capped by reaction with carboxylic acids or other reagents (e.g., isothiocyanates, sulphonyl chlorides, etc.). AAZ, SABA and *m*-SABA were among the 1462 building blocks used in the second reaction step, thus facilitating the quantification of ligand recovery both by qPCR (Figure 5.3, Paragraph 5.1) and by the analysis of sequence-derived DECL fingerprints. [190] We performed selections using various amounts of input DNA, capture methods and washing procedures. The results are summarized in Table 2.1.

Table 2.1 Selections performed against CAIX in a real library setting. Experiments were performed using an input of library set at 2.46 x 10^{10} copies for each CAIX ligand (AAZ, SABA, *m*-SABA) displayed as a second building block. Evaluated experimental conditions: type of protein variant (e.g., polyhistidine-tagged and biotinylated-tagged), protein concentration, amount of Tween-20 in the washes, method used). Method A: Affinity selection against biotinylated CAIX using streptavidin-coated beads; Method B: Affinity selection against hisCAIX using Dynabeads His-tag Pull-down. The recovery of each portion of the library containing a specific CAIX ligand as second building block was calculated by using the following formula: (DNA output/DNA input) * 100.

Entry	Ligand	Protein	Protein Conc. (μM)	Tween %	Method	DNA Output	Recovery %
1a	AAZ	hisCAIX	1	0.01%	B	1.25 x 10^{8}	0.5074
1b	AAZ	hisCAIX	2.5	0.01%	B	7.82 x 10^{7}	0.3180
1c	AAZ	hisCAIX	1	0.10%	B	1.28 x 10^{8}	0.5195
1d	AAZ	hisCAIX	2.5	0.10%	B	1.46 x 10^{8}	0.5938
1e	AAZ	bCAIX	1	0.05%	A	1.05 x 10^{8}	0.4256
1f	AAZ	bCAIX	2	0.05%	A	1.02 x 10^{8}	0.4154
1g	AAZ	bCAIX	1	0.10%	A	1.12 x 10^{8}	0.4566
1h	AAZ	bCAIX	2	0.10%	A	1.04 x 10^{8}	0.4241
1i	AAZ	no protein	no protein	0.01%	B	3.91 x 10^{4}	0.0002
1j	AAZ	no protein	no protein	0.10%	B	3.51 x 10^{4}	0.0001
1k	AAZ	no protein	no protein	0.05%	A	1.20 x 10^{5}	0.0005
1l	AAZ	no protein	no protein	0.10%	A	1.72 x 10^{5}	0.0007
2a	SABA	hisCAIX	1	0.01%	B	2.22 x 10^{7}	0.0902
2b	SABA	hisCAIX	2.5	0.01%	B	2.15 x 10^{7}	0.0875
2c	SABA	hisCAIX	1	0.10%	B	2.29 x 10^{7}	0.0930
2d	SABA	hisCAIX	2.5	0.10%	B	3.54 x 10^{7}	0.1439
2e	SABA	bCAIX	1	0.05%	A	4.33 x 10^{7}	0.1762
2f	SABA	bCAIX	2	0.05%	A	4.58 x 10^{7}	0.1862
2g	SABA	bCAIX	1	0.10%	A	4.55 x 10^{7}	0.1851
2h	SABA	bCAIX	2	0.10%	A	4.81 x 10^{7}	0.1956
2i	SABA	no protein	no protein	0.01%	B	5.39 x 10^{5}	0.0022
2j	SABA	no protein	no protein	0.10%	B	6.91 x 10^{5}	0.0028
2k	SABA	no protein	no protein	0.05%	A	1.44 x 10^{5}	0.0006
2l	SABA	no protein	no protein	0.10%	A	1.16 x 10^{5}	0.0005
3a	*m*-SABA	hisCAIX	1	0.01%	B	4.66 x 10^{6}	0.0190
3b	*m*-SABA	hisCAIX	2.5	0.01%	B	4.55 x 10^{6}	0.0185
3c	*m*-SABA	hisCAIX	1	0.10%	B	3.73 x 10^{6}	0.0152
3d	*m*-SABA	hisCAIX	2.5	0.10%	B	5.42 x 10^{6}	0.0220
3e	*m*-SABA	bCAIX	1	0.05%	A	5.46 x 10^{6}	0.0222
3f	*m*-SABA	bCAIX	2	0.05%	A	6.27 x 10^{6}	0.0255
3g	*m*-SABA	bCAIX	1	0.10%	A	6.31 x 10^{6}	0.0256
3h	*m*-SABA	bCAIX	2	0.10%	A	6.42 x 10^{6}	0.0261
3i	*m*-SABA	no protein	no protein	0.01%	B	1.52 x 10^{6}	0.0062
3j	*m*-SABA	no protein	no protein	0.10%	B	1.74 x 10^{6}	0.0071
3k	*m*-SABA	no protein	no protein	0.05%	A	2.56 x 10^{5}	0.0010
3l	*m*-SABA	no protein	no protein	0.10%	A	3.60 x 10^{5}	0.0015
4a	AA	hisCAIX	1	0.01%	B	7.91 x 10^{5}	0.0032
4b	AA	hisCAIX	2.5	0.01%	B	1.19 x 10^{6}	0.0048
4c	AA	hisCAIX	1	0.10%	B	5.80 x 10^{5}	0.0024
4d	AA	hisCAIX	2.5	0.10%	B	1.08 x 10^{6}	0.0044
4e	AA	bCAIX	1	0.05%	A	3.63 x 10^{6}	0.0147
4f	AA	bCAIX	2	0.05%	A	4.61 x 10^{6}	0.0187
4g	AA	bCAIX	1	0.10%	A	3.45 x 10^{6}	0.0140
4h	AA	bCAIX	2	0.10%	A	4.22 x 10^{6}	0.0172
4i	AA	no protein	no protein	0.01%	B	2.38 x 10^{5}	0.0010
4j	AA	no protein	no protein	0.10%	B	8.78 x 10^{4}	0.0004

For illustrative purposes, a representative subset of data is displayed in Figure 2.7. A total amount of 3.6 x 10^{13} DNA-encoded molecules was used as input, which corresponds to 2.46 x 10^{10} molecules (I) containing each of the DNA-tagged CAIX ligands (AAZ, SABA, *m*-SABA). After an affinity capture step on CAIX-coated magnetic beads, AAZ was recovered with the highest efficiency (nearly 0.5% of the input), while the recovery of SABA and *m*-SABA derivatives was 0.1% and 0.01%, respectively. The selectivity of the capture procedures can be assessed by inspection of the results obtained in the absence of protein and of the recovery of negative-control DNA conjugates (e.g., those carrying a terminal acetyl moiety instead of a CAIX ligand as BB2) (Figure 2.7).

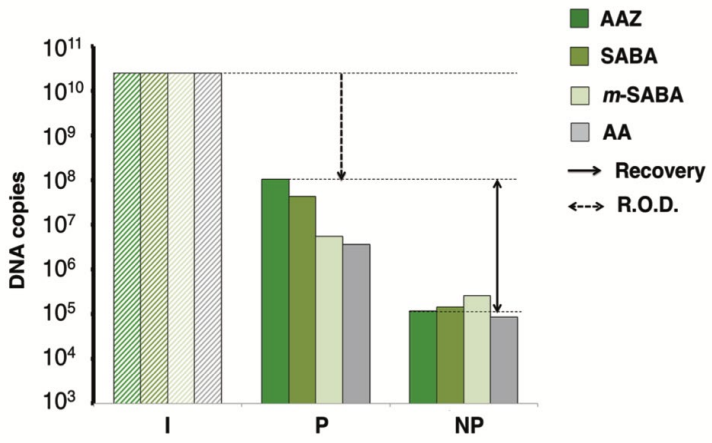


Figure 2.7. qPCR results of CAIX selection experiments performed with a library based on two sets of building blocks. The portion of the library corresponding to the use of AA as a capping reagent for the second reactions step was used as a negative control (AA). Selections were also performed on affinity supports, without immobilized CAIX. I: Input; P: protein selections; NP: no protein selections.

We performed selections at different library input conditions (ranging from 10^4 to 10^8 copies of each library member). In order to obtain the number of library members featuring AAZ, SABA or *m*-SABA moieties, a multiplication times 246 would be needed (corresponding to the 246 building blocks in position 1, which were capped with each of the CAIX ligands). The DNA amount of the elution fraction was analysed by qPCR (Figure 2.8) and the results were correlated with the quantitative aspects of the fingerprints, derived from the sequence analysis of the library before and after affinity capture (Figure 2.9). As shown in Figure 2.8, below a library input of 10^4 copies per library member (2.46 x 10^6 copies for each CAIX ligand), the three CAIX ligands are not discriminated in terms of their enrichment factor, revealing that the procedure may be suboptimal in these conditions.

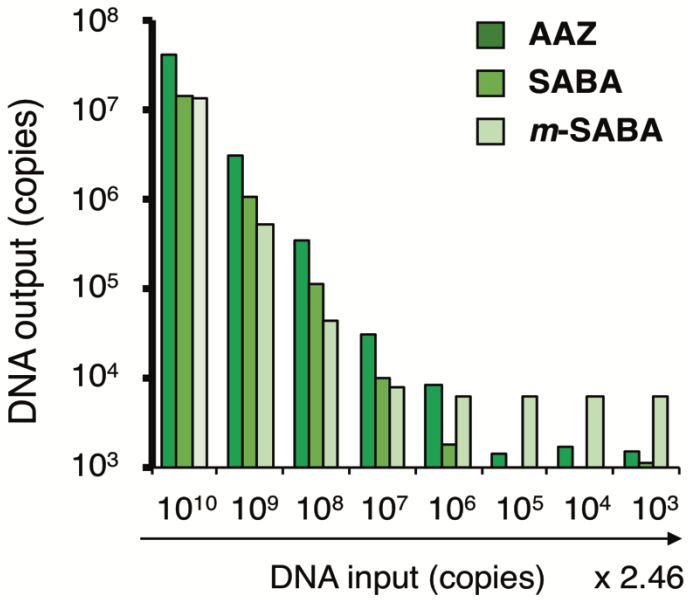


Figure 2.8. qPCR quantification of the recovery of CAIX ligands (AAZ, SABA, *m*-SABA) after selections against biotinylated CAIX, using different inputs of the library (ranging between 2.46 x 10^{10} and 2.46 x 10^3 copies for each CAIX ligand). Each CAIX ligand is displayed as a second building block in a DECL library. The results are plotted in logarithmic scale.

Figure 2.9 shows representative fingerprints in two different graphical representations. The top panels correspond to standard two-dimensional representations of selection results for libraries based on two sets of building blocks. [85] The bottom panels correspond to "cumulative plots", which were obtained by summing up the sequence counts of all library members, featuring a given building block in position 2. Inspection of the selection fingerprints revealed a clear enrichment of AAZ and SABA derivatives, compared to most other building blocks in position 2 for all input conditions, except for the lowest amounts (e.g., 10^4 copies of each library member, 2.46 x 10^6 copies of sulfonamide derivatives). The calculation of the total enrichment factors quantified for each sulfonamide derivative is shown in Figure 2.10.

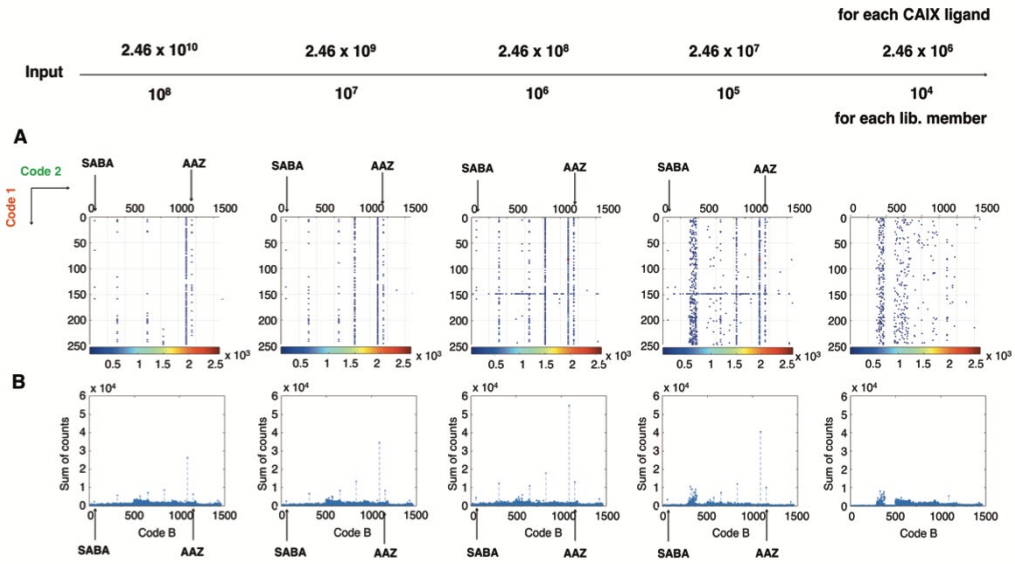


Figure 2.9. Sequencing results of library selections against biotinylated CAIX. Selections were performed with different inputs of the library, ranging from 10^4 to 10^8 copies for each library member (2.46×10^{10} to 2.46×10^6 library members capped with each CAIX ligand). A. Selection fingerprints. The individual library members are univocally identified by their codes 1 (ranging from 1 to 246) and codes 2 (ranging from 1 to 1462). The number of counts is displayed as dots of different colors and ranges from 0 to 2500. The cutoff threshold was set at 60 counts. CAIX ligands (AAZ, SABA, *m*-SABA) are identified by different codes 2. B. Cumulative plots obtained by summing up the sequence counts of all library members containing each building block 2 (code 2).

Interestingly, other novel fragments were found to bind to CAIX in an avid and selection manner. The binding properties of these derivatives will be published elsewhere. The cumulative plot results for *m*-SABA were only slightly higher than the background, indicating potential problems associated with the recovery of binders with single-digit micromolar potency.

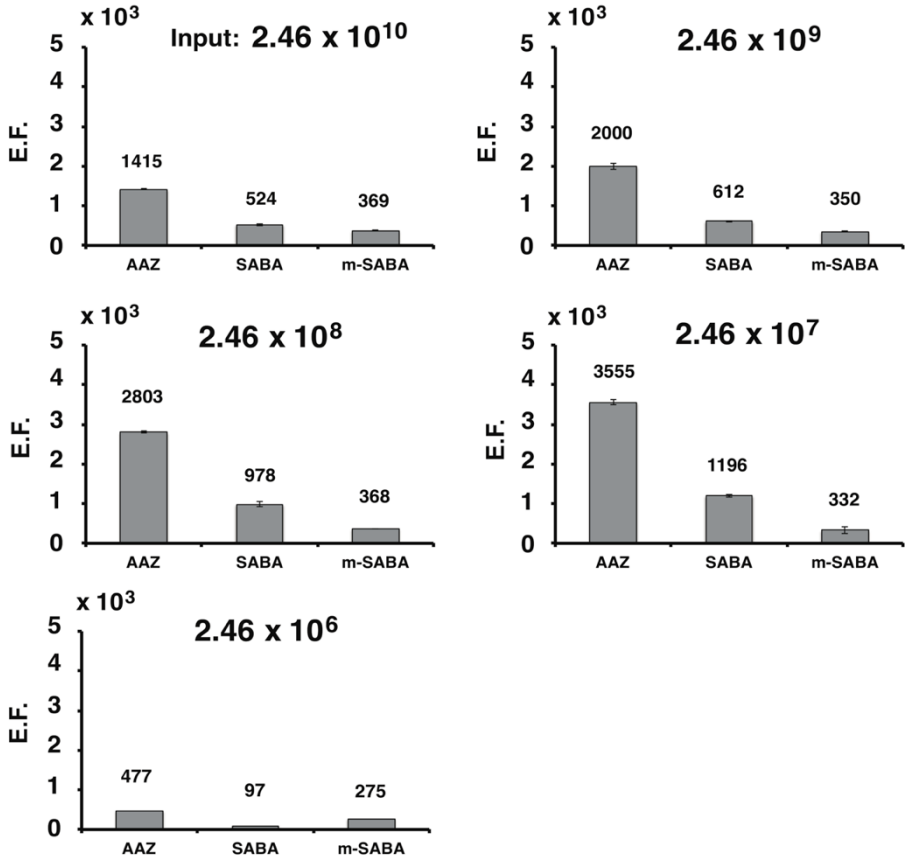


Figure 2.10. Total enrichment factors calculated for each CAIX ligand (AAZ, SABA and *m*-SABA). Selections were performed using different inputs of the library (ranging between 2.46 x 10^{10} and 2.46 x 10^6 copies for each CAIX ligand). Enrichment Factors were calculated with the following formula:

$$E.F. = \frac{cumulative\ counts\ (per\ BB)\ x\ size\ library}{total\ number\ of\ counts}$$

2.1.3 Discussion and Conclusion

We have used CAIX ligands of different affinity and DNA-encoded chemical libraries in order to quantitatively assess the recovery and selectivity of binders in affinity capture procedures. The capture methods used in the study (e.g., streptavidin- or metal-based supports, magnetic beads vs resin) were the ones which are most commonly utilized both in the industrial and academic setting. [85] Some of the main findings can be summarized as follows.

High-affinity binders (e.g., AAZ with a K_d = 13 nM) are always efficiently recovered, irrespective of the affinity capture methodology that was used. The typical recoveries ranged between 10 and 80% per capture step, starting from relatively high input quantities of DNA-encoded compound.

A decrease in binding affinity for CAIX ligands led to a decrease in capture efficiency. Surprisingly, *m*-SABA (a micromolar binder) exhibited poor recoveries and selectivity in most experimental conditions, casting doubts about the potential of reliably identifying micromolar hits using conventional selection methodologies.

Individual capture methods may show advantages and disadvantages, which are ligand- and protein-related. Thus, an empirical optimization procedure (e.g., testing various target immobilization and capture conditions) is likely to be important for future screening campaigns. [98] The results obtained by qPCR assessment of capture efficiency were confirmed by the sequence-based evaluation of selections, performed with a DNA-encoded chemical library, containing both established and novel building blocks, capable of CAIX recognition. Interestingly, when the amounts of individual library members dropped below a certain threshold (approx. 10^5 copies of individual library members) selection efficiency dropped dramatically (Figure 2.9). Collectively, these findings indicate that libraries should be synthesized in sufficiently large amounts, allowing for the use of at least 10^5 (and preferably 10^6) input copies of DNA-encoded compounds.

Some groups have reported the use of two consecutive affinity capture steps, prior to DNA sequencing. [206, 207] This methodology appears to be reasonable with hits comparable to SABA or AAZ are found. However, hits of lower potency are most probably lost if two consecutive screening procedures are performed. Our observations are largely in keeping with the results of Krusemark and colleagues, who studied the efficiency of affinity capture selections using DNA-encoded peptides, capable of binding to CBX. [194] The authors found that lower enrichments and recoveries of ligands are obtained for lower ligand affinity to both CBX7 and CBX8 ChDs. If the findings of both studies are confirmed for other proteins, there may be a need for the development of more efficient selection methods, particularly for targets for which only hits with micromolar potency can be expected. Possible directions include the use of cross-linking after affinity capture (Paragraph 1.3.2), use of chromatographic procedures with rebinding potential, capillary

electrophoresis separations [208, 209]. Potentially, tethering approaches [210– 214] with reversible covalent capture steps [215] could stabilize the interaction of weak binders with target proteins of interest.

The use of qPCR methods and DNA sequencing to quantitatively evaluate DNA-encoded chemical library selections revealed that conventionally used procedures are efficient for the capture of high-affinity ligands but may fall short for the recovery of binders with intermediate affinity. The findings of this study provide a rationale and a motivation for the development of novel selection methodologies, that may facilitate the identification of specific ligands against "difficult-to-drug" protein targets.

2.2 EVALUATION OF PHOTOCROSSLINKING PARAMETERS FOR THE IMPLEMENTATION OF EFFICIENT DNA-ENCODED CHEMICAL LIBRARY SELECTIONS

This paragraph is adapted with permission from Sannino, A.; Gironda-Martínez, A.; Gorre, É. D.; Prati, L.; Piazzi, J.; Scheuermann, J.; Neri, D.; Donckele, E. J.; Samain, F. Critical Evaluation of Photo-cross-linking Parameters for Efficient DNA-Encoded Chemical Library Selections. *ACS Comb. Sci.* **2020**, *22*, 204−212. Copyright © (2020) American Chemical Society. [216]

My contribution to this work was first to develop a photocrosslinking selection methodology using DNA-encoded chemical libraries. I have designed and performed all the experiments reported in paragraph 2.2.2., which included: 1) design of the structure of the photoreactive DNA-ligand assemblies; 2) synthesis and characterization of model photoreactive DNA-conjugates; 3) design of a screening procedure based on photocrosslinking; 4) selection experiments against CAIX and analysis of the outputs by qPCR, gel electrophoresis and ELISA experiments; 5) study of different selection parameters and optimization of the procedure. In the second part of the project, I have implemented the studied methodology for the screening of a single pharmacophore DECL available in-house. For this purpose, I have designed and synthesized a photoreactive DECL analogue, and performed the selection experiments against carbonic anhydrase IX, using both the photocrosslinking methodology and solid-phase affinity selection procedures. I have carried out the qPCR quantification of the selection outputs and the analysis of the sequencing results. I have contributed to the off-DNA synthesis of the ligands identified from DECL screenings and hit validation by fluorescence polarization.

2.2.1 Introduction

Conventionally, DECLs are screened using affinity-capture procedures, in which the target protein of interest is immobilized on solid supports, [144] in full analogy to what has previously been described for other classes of encoded combinatorial libraries (e.g., phage display antibody libraries). As discussed in Paragraph 2.1, we have recently evaluated the performance of affinity capture methodologies for DECLs and shown that high-affinity binders can efficiently be recovered, while the identification of medium-affinity ligands (e.g., those with a K_d value in the micromolar range) may be more problematic. Indeed, micromolar binders are often characterized by a rather limited kinetic stability of their complex with the cognate protein target, leading to loss of recovery upon the implementation of stringent washing procedures. [178] In this context, the use of photocrosslinking reactions may stabilize the interaction between putative binders and their target protein. [119, 150–154, 165, 192]

Since covalent crosslinking was first introduced by Westheimer in the early 1960s [217], chemical crosslinking has been increasingly being used as a tool for the discovery of protein-protein, [218, 219] protein-DNA, [220, 221] and protein-ligand interactions. [222] Despite the significance, few crosslinkers are available including benzophenone, [223] arylazide, [224, 225] sulfonyl fluoride, [165] and diazirine moiety. [152, 226–229] In respect of photocrosslinking, diazirine is utilized as a photoreactive group, which upon irradiation generates a carbene that covalently binds to the protein of interest. The most significant feature of carbenes is their ability to rapidly form a covalent bond with the nearest molecule (e.g., amino acids of the protein) through C–C, C–H, O–H and X–H (X = heteroatom) insertions. [152, 226–229] We decided to use an alkyl diazirine group in our study due to its stability at room temperature, its absorption at longer-wavelengths (350–365 nm), reducing damage to the targeted biological system and its short lifetime upon irradiation and subsequent high reactivity. [230] In general, photocrosslinking efficiency is reported to be low since carbenes can rapidly be quenched in aqueous media. [231, 232] However, this feature is actually an advantage since it minimizes unspecific crosslinking. Only those ligand molecules that have a certain affinity will react covalently to the protein, whereas unbound ligand molecules will react with water before being able to undergo non-specific reactions with the protein of interest. [228, 229] This approach offers a number of conveniences for assembly of a DNA-encoded small molecule with a DNA-linked diazirine group and therefore expands the capabilities of DECL selections. By "locking" the transient interaction of DECL ligands to the target, crosslinking can give improved enrichment of ligands over non-ligands, particularly for low-affinity binders. [150–154, 165, 192]

Most libraries described so far were based on double-stranded DNA. It appears that ssDNA libraries may be more flexible, allowing the construction of encoded self-assembling chemical libraries (ESAC), [107–109, 117, 123] innovative screening techniques (e.g., interaction-dependent PCR) [147] and in this study the hybridization with a complementary oligonucleotide-photocrosslinker conjugate. The implementation of a photocrosslinker at the 5' end of an oligonucleotide primer may be particularly convenient, as it would allow a templated DNA polymerization step after successful ligand binding and crosslinking reaction.

The use of photocrosslinking reactions for the screening of various types of DECLs has previously been reported. [150–154, 165, 192] Zhao and coworkers developed a DNA-based protein labeling method, named DNA-programmed affinity labeling (DPAL), in which a DNA-linked small molecule guides the capture and the identification of its target through photocrosslinking. In this work, the authors reported a methodology based on ligand-directed photocrosslinking, followed by ExoI digestion of non-binders and hit decoding.

Krusemark *et al.* [192] developed an approach for applying photocrosslinking selections, employing a tethered ssDNA construct, which allows for a reactive group to be appended to DNA-encoded ligands after equilibration with protein targets. The authors tested several

reactive groups (e.g., electrophiles and photoreactive groups) and compared crosslinking efficiency in model selection setting, reporting photocrosslinking efficiencies between 12 and 36%.

Here we describe the implementation and systematic evaluation of a photocrosslinking screening procedure, featuring encoded compounds on the 5' end of a single-stranded DECL, which can be hybridized to a complementary strand, used to display a diazirine group as photoreactive moiety on the 3' end. This approach shares the same design of encoding self-assembling chemical libraries (ESAC) [107–109, 117, 123] and it was adapted to be compatible with single pharmacophore libraries based on splint ligation. [85, 93–99] We demonstrate the ability to enrich DNA-encoded ligands after photocrosslinking selection both in model selections and real library setting.

2.2.2 Results

Model Selections with characterized CAIX sulfonamide inhibitors

We used carbonic anhydrase IX (CAIX) as target protein in both model selections and in library screenings. [195–197] Figure 2.11 shows a schematic representation of model screening procedures based on photocrosslinking. Individual small molecule ligands (depicted as green circles) were chemically attached on distinct 48-mer DNA tagged oligonucleotides, carrying a unique 6-base pairs sequence codon used as a bar code. The DNA-ligand derivatives were subsequently paired with a partially complementary oligonucleotide carrying a diazirine at the 3' end and featuring an abasic region in correspondence of the DNA barcode on the cognate strand (Figure 5.8, Paragraph 5.1). [107–109] We and others have previously described how to synthesize single-stranded DECLs, using splint ligation procedures. [85, 93–99] At first, it was convenient to equip the complementary strand, carrying the photoreactive moiety, also with a fluorescent amide (FAM) at the 5' end, thus allowing easy detection of the fluorescent conjugates in gel electrophoresis experiments.

Indeed, individual encoded compounds on DNA or large DECLs can be hybridized to a universal photoreactive DNA strand, allowing complex formation with target proteins of interest, followed by affinity capture in denaturing experimental conditions. The efficiency of screening procedures can be quantitatively assessed using qPCR, ELISA, electrophoretic mobility shift assay (EMSA) or by DNA sequencing and sequence counts. [178] The formation of a stable complex of a DNA-encoded ligand with the target protein of interest, initially driven by the compound displayed on DNA, can be covalently "locked" by irradiation and covalent bond formation.

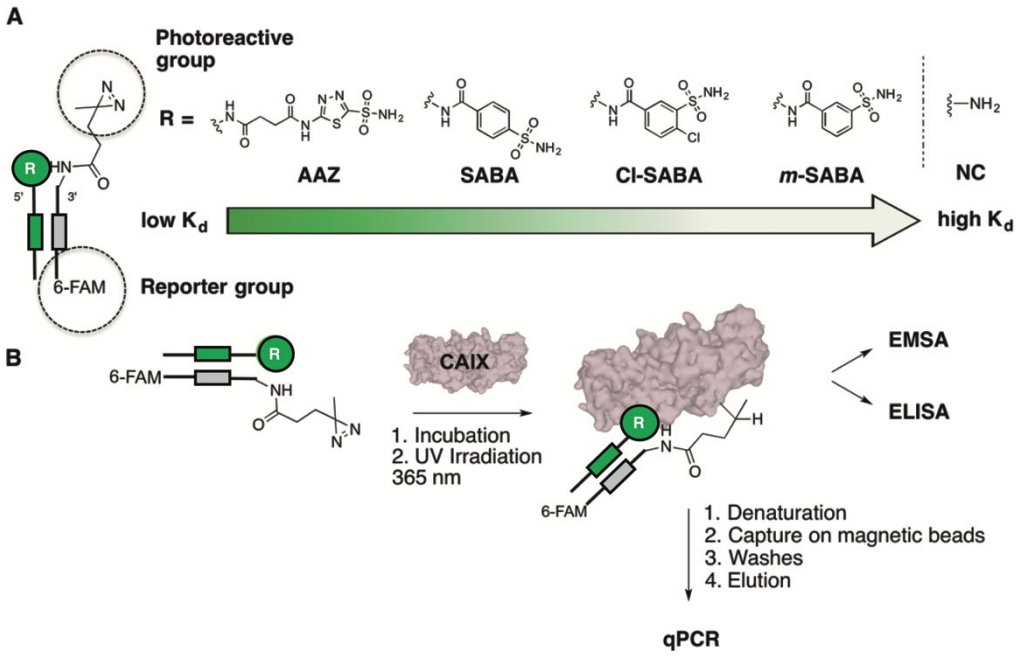


Figure 2.11. A. Schematic representation of DNA-tagged ligands paired with a complementary fluorophore-labeled DNA-tagged photoreactive group. R: AAZ, SABA, m-SABA, Cl-SABA, NC. B. Setup of model selections based on the photocrosslinking methodology.

As depicted in Figure 2.11, the photocrosslinking approach was implemented in model selections by using polyhistidine-tagged CAIX (his-CAIX), and four reported [178, 197, 198] and characterized CAIX sulfonamide inhibitors of various affinities. (Paragraph 2.1.2) Experimental dissociation constants of the fluorescein-labeled acetazolamide (AAZ), p-sulfamoyl benzoic acid (SABA), 4-chloro-3-sulfamoylbenzoic acid (Cl-SABA), and 3-sulfamoyl benzoic acid (*m*-SABA) and derivatives were determined by off-DNA fluorescence polarization yielding Kd values of 8.7, 192, 500 and 1030 nM, respectively. (Figure 5.11). Negative control experiments were performed using free amino-DNA oligonucleotide (FAM-DNA-NC) as input.

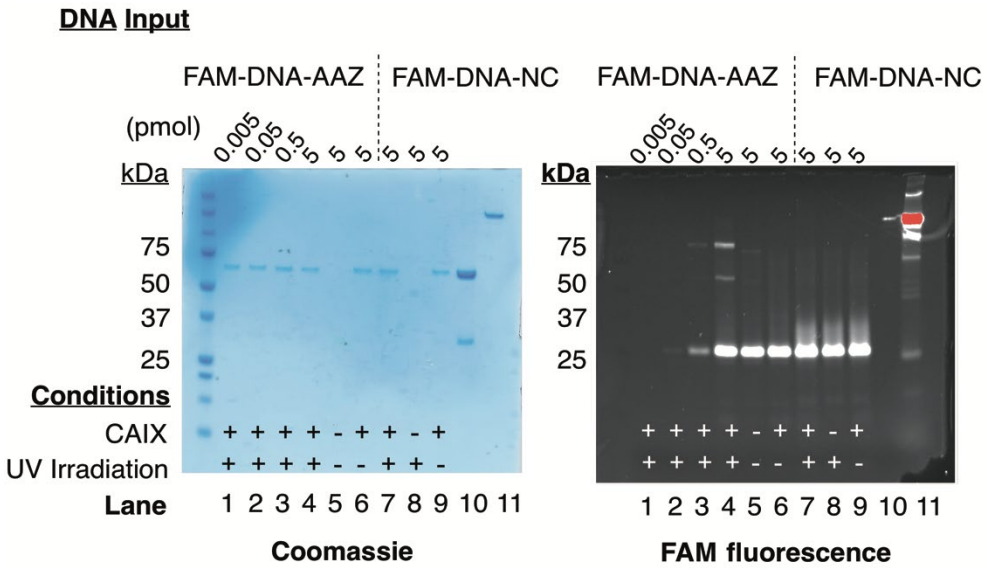


Figure 2.12. SDS-PAGE analysis from FAM-DNA-AAZ and FAM-DNA-NC crosslinking with CAIX, after Coomassie staining (left panel) and FAM fluorescence imaging (right panel). Lanes: 1–4: CAIX-AAZ-FAM complex after UV-irradiation (Input FAM-DNA-AAZ: 0.005 – 0.05 – 0.5 – 5 pmol). Protein concentration: 2 μM. Lane 5: only FAM-DNA-AAZ, no protein control. Lane 6: CAIX-AAZ-FAM, no UV-irradiation control. Lane 7: CAIX-NC-FAM after UV-irradiation (Input FAM-DNA-NC: 5 pmol). Protein concentration: 2 μM. Lane 8: only FAM-DNA-NC. Lane 9: CAIX-NC-FAM, no UV-irradiation control. Lane 10: CAIX control. Lane 11: F8 antibody FITC-conjugated was used as fluorescence control.

Figure 2.12 shows the results of gel electrophoresis analysis for various quantities of photoreactive FAM-DNA-AAZ derivatives, ranging from 0.005 to 5 pmol (lanes 1–4). The FAM-DNA-CAIX ligand was incubated with an excess quantity (100 pmol) of non-immobilized his-CAIX in solution to pre-form the protein-ligand complex, which was subsequently UV-irradiated at 365 nm. Crude reaction mixtures were then analyzed by SDS-PAGE. The polyacrylamide gel was stained with Coomassie Blue or imaged with a fluorescence detector, revealing the covalent formation of a photocrosslinking product between CAIX and the photoreactive ligand with a yield of 10% (lane 4). It has previously been reported that quantitative photocrosslinking yields may be prohibitive, because of the instability of carbene intermediates. [231, 232] Surprisingly, no crosslinking product was detected when FAM-DNA-NC was used as input (lanes 7–9). The formation of covalent adduct was also qualitatively monitored by ELISA (Figure 5.12, Paragraph 5.1) to assess the signal to noise with and without irradiation.

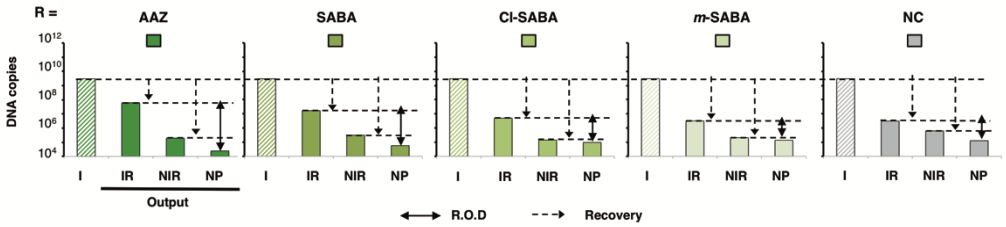


Figure 2.13. qPCR quantification of FAM-DNA-CAIX ligands after model selections against polyhistidine-tagged CAIX plotted on a logarithmic scale. I: Input; IR: Protein selection with UV-irradiation; NIR: Protein selection without UV-irradiation; NP: Selection on empty beads in the absence of the protein. The results obtained for each FAM-DNA-CAIX ligand are depicted in green. A range of detection (R.O.D.) can be derived from the comparison of output DNA quantities in selections performed in the presence (IR) or in the absence (NP) of CAIX.

Fifteen selection experiments were performed against his-CAIX, as displayed in Figure 2.13. The FAM-DNA-CAIX ligands were incubated with the non-immobilized protein in solution. The protein-ligand complexes were further subjected to UV-irradiation (IR) at 365 nm and subsequently captured on Dynabeads magnetic cobalt-based beads. Selection control experiments were also carried out without UV-irradiation (NIR) of the protein-ligand complexes. Control experiments were performed on empty beads using the irradiation conditions in the absence of the protein (NP). Identical amounts of FAM-DNA-CAIX ligands, corresponding to 3 x 10^9 DNA copies of each molecule before selection, were used as input (I).

The captured protein-ligand complexes were then subjected to denaturing stringent washes. Eluted FAM-DNA ligands were analyzed and quantified by qPCR, with standard calibration curves established for each experiment (Figure 5.13, Paragraph 5.1). An experimental range of detection (R.O.D.) was determined by comparison of DNA output selections carried out in the presence (IR) or in the absence of protein (NP). [178] The recovery of FAM-DNA-AAZ ligands varied from 5% yield in the normal conditions to 0.1% and 0.01% for the (NIR) and (NP) control experiments respectively. Similar findings were observed for the FAM-DNA derivatives of SABA, Cl-SABA and *m*-SABA, despite lower recovery efficiencies (0.55% - 0.16% - 0.11%, respectively). In these experimental conditions, the FAM-DNA-NC derivative was also recovered with an efficiency of 0.1%, thus suggesting that the methodology may require further optimization to improve the discrimination of low- affinity ligands.

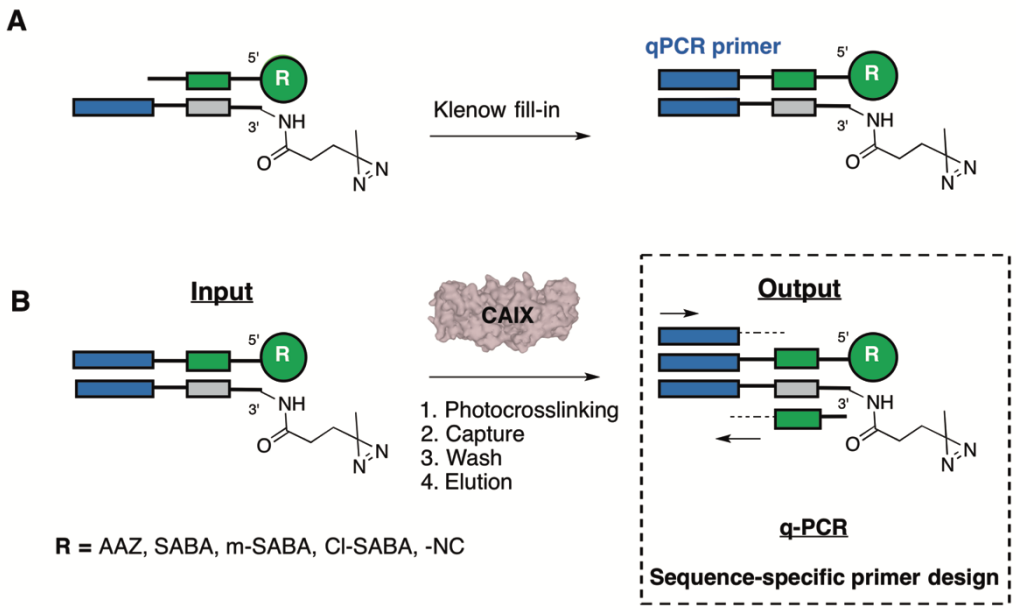


Figure 2.14. A. Preparation of DNA-CAIX ligands for selection experiments. B. The distinct DNA-CAIX binders were pooled in 1:1 ratio and used in the selection experiments as input. After the screening, the amount of the eluted DNA molecules was quantified by qPCR using sequence-specific primer sets.

To gain an insight into the efficiency of the photocrosslinking screening methodology, we then assessed the selectivity of the procedure by performing selection experiments with a pool of DNA-CAIX ligands comprising a distinct PCR primer sequence for further qPCR quantifications (Figure 2.14). The 48-mer DNA-CAIX ligands, AAZ, SABA, Cl-SABA, and *m*-SABA were annealed with a distinct 80-mer complementary DNA-diazirine conjugate comprising a unique qPCR primer sequence (Figure 5.9, Paragraph 5.1). After Klenow-fill-in polymerization, which allows the primer region to be transferred onto the complementary strand, the distinct double-stranded DNA-CAIX ligands were pooled in equimolar ratio. (Figure 2.14A).

Selection experiments were carried out in similar conditions as described above and quantified by qPCR (Figure 5.14. Paragraph 5.1). The orthogonal primer design was evaluated by standard curve calibration experiments (Figure 5.15, Paragraph 5.1). We performed selections by using various amounts of protein (Figure 2.15) quenching methods and buffer formulations (Figure 2.16) and analyzed the eluted DNA-CAIX ligands by qPCR.

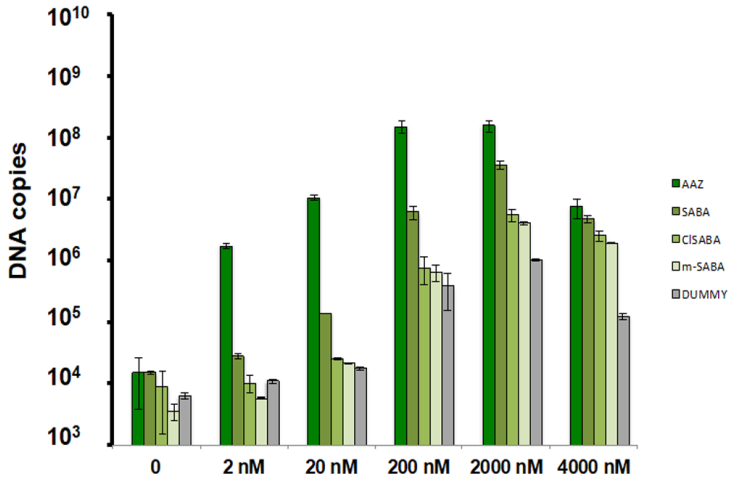
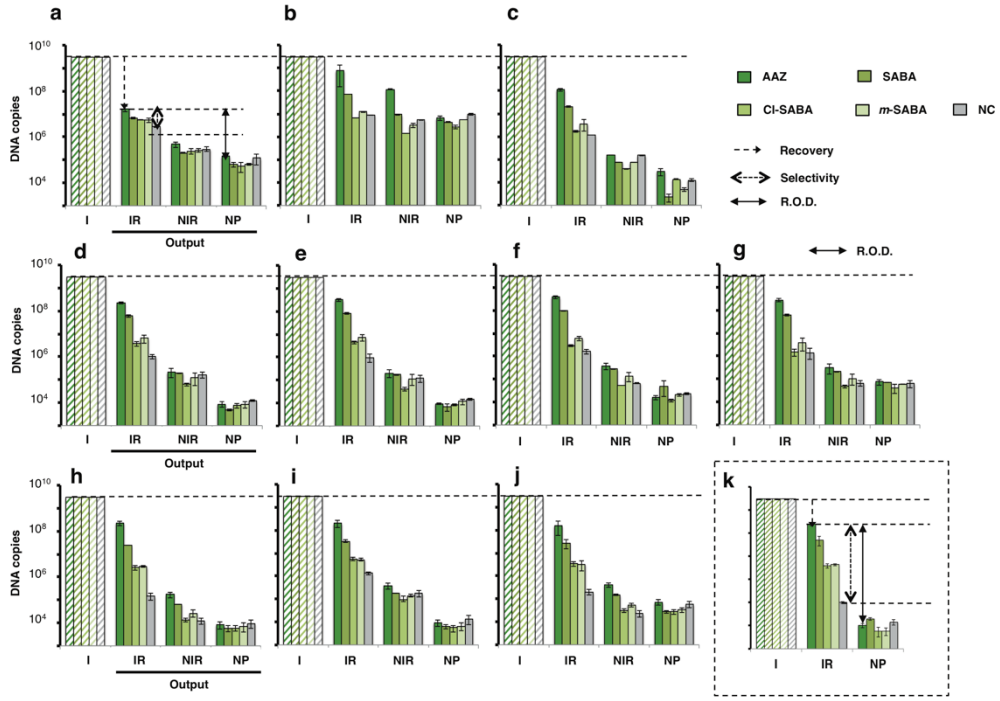


Figure 2.15. Evaluation of the recovery yield and selectivity after model selections against his-CAIX, used at different protein concentrations. (2 nM – 20 nM – 200 nM – 2000 nM – 4000 nM). A pool of the photoreactive DNA-CAIX ligands (3 x 10^9 copies for each conjugate) was used as input. The pool of conjugates was incubated with different concentration of protein before UV irradiation. After denaturation of the protein, capture on magnetic beads/washes/elution, the number of DNA molecules in the output solution was analyzed by qPCR. The quantity of the DNA after selection is plotted in logarithmic scale (green bars). Selection against empty beads (no protein, 0) was included as a negative control. We observed a correlation between recovery and binding constants of the ligands. The best recovery yields were obtained with a protein concentration of 2 μM. At 4 μM protein concentration, we observed a decrease of the recovery yield, probably determined by oversaturation of the beads (hook effect).

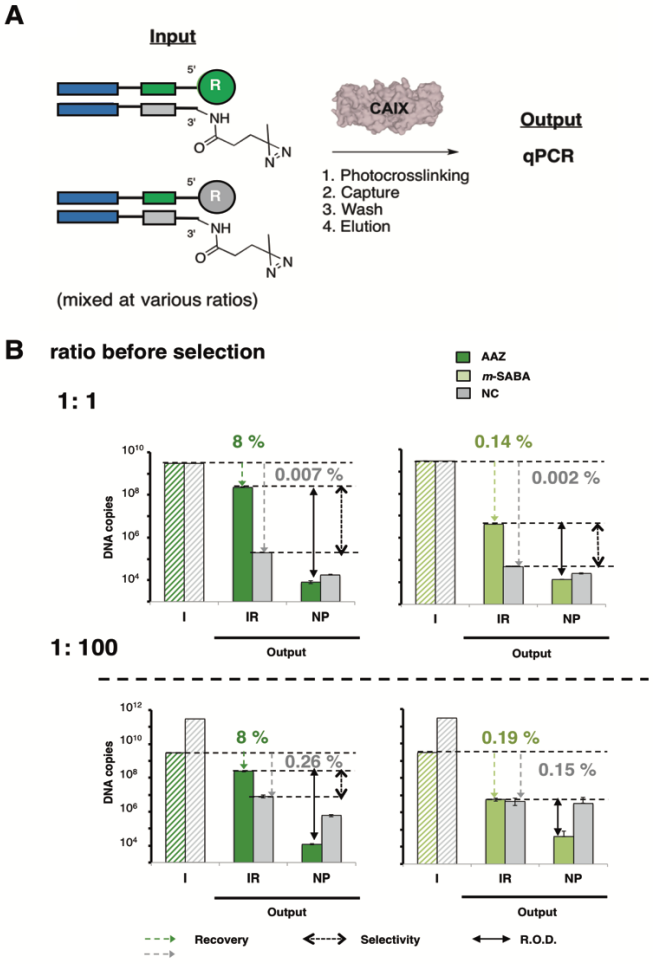


Condition	Quenching	Incubation Buffer	Ionic strength	Blocking additives
a	YES	H_2O		
b	NO	PBS Gibco 1x, pH 7.4		
c	YES	PBS Gibco 1x, pH 7.4		
d	YES	PBS Gibco 1x, pH 7.4	0.05 M NaCl	
e	YES	PBS Gibco 1x, pH 7.4	0.1 M NaCl	
f	YES	PBS Gibco 1x, pH 7.4	0.5 M NaCl	
g	YES	PBS Gibco 1x, pH 7.4	1 M NaCl	
h	YES	PBS Gibco 1x, pH 7.4		0.3 mg/ml HS
i	YES	PBS Gibco 1x, pH 7.4		0.1% Tween-20
j	YES	PBS Gibco 1x, pH 7.4		4% non-fat dry milk
k	YES	PBS Gibco 1x, pH 7.4	0.1 M NaCl	0.3 mg/ml HS

Figure 2.16. A) qPCR quantification of model selection experiments performed using a pool of DNA-CAIX ligands as input. I: Input, IR: Protein Selection with UV-irradiation; NIR: Protein Selection no UV-irradiation; NP: Selection in the absence of protein. Different experimental conditions were evaluated to optimize the recovery of DNA-CAIX ligands, the selectivity of the screening procedure and the R.O.D. (range of detection). Results are plot in logarithmic scale. Conditions b-c. Effect of denaturing quenching conditions adding 5% SDS in the incubation buffer after UV-irradiation. Conditions d-g. Effect of ionic strength, adding various quantities of NaCl (aq) in the incubation buffer. Conditions h-j. Effect of blocking additives in the incubation buffer (HS: Herring Sperm) k. Optimized conditions for the screening methodology.

A total amount of 1.5 x 10^{10} of DNA-encoded molecules, corresponding to 3 x 10^9 copies of each DNA-tagged CAIX ligand was used as input (I), After incubation with polyhistidine-tagged CAIX, UV-irradiation, affinity capture of the protein-ligand complexes and stringent denaturing washes, the eluted DNA-CAIX ligands were analyzed by qPCR (IR). The selectivity of the screening procedure was defined as the ratio between the recovery of the DNA-CAIX ligands (green bars) and the recovery of the free amino-DNA oligonucleotide (DNA-NC). In optimized conditions (entry k; Figure 2.16) DNA-AAZ ligands were recovered with the highest efficiency (8% of recovery), whereas the recovery yields of the medium and low-affinity ligands (SABA, Cl-SABA, *m*-SABA) were less efficient (1.7%, 0.13%, 0.15% each). In these experimental conditions, the recovery of AAZ and *m*-SABA were 2500 and 43 times more efficient than the DNA-NC selection, suggesting that the methodology can potentially provide good discrimination of the putative binders over the background signal.

We also investigated the effect of the background on the screening procedures in model selections, which represents a critical variable when performing DECL screenings (Figure 6). DNA-CAIX ligands (AAZ or *m*-SABA) were mixed with DNA-NC at various input ratios (1:1 and 1:100). The quantity of DNA-CAIX ligands was fixed at 3 x 10^9 copies each. The pools were panned against his-CAIX following qPCR analysis of the outputs. (Figure 6A). The DNA-CAIX ligand / DNA-NC ratio before and after selection was assessed from the recovery ratios of each species, after qPCR quantification. At an initial DNA-AAZ/DNA-NC and DNA-*m*-SABA/DNA-NC 1:1 ratio, the DNA-CAIX ligands became the major species in the mixture (ratio after selection: 1136 and 81, respectively) after one round of selection. Surprisingly, both the highest (DNA-AAZ) and the lowest affinity derivative (DNA-*m*-SABA) were enriched from 0.01 to 31 and 1.28 (corresponding to an enrichment of 3100 and 128, respectively) after one round of selection, when using an initial 1:100 ratio of DNA-CAIX ligand / DNA-NC derivative (Figure 2.17).



C

Ratio of DNA-CAIX ligand to DNA-NC	DNA-AAZ		DNA-*m*-SABA	
	1:1	1:100	1:1	1:100
before selection	1	0.01	1	0.01
after selection	1136	31	81	1.28

Figure 2.17. A. DNA-AAZ or DNA-m-SABA mixed with various quantities of DNA-NC (1:1 and 1:100 ratios, respectively). The DNA quantity of DNA-AAZ or DNA-*m*-SABA was fixed at 3 x 10^9 DNA copies. The pooled DNA ligands were used as input in model selection experiments. B. Amount of DNA-AAZ, DNA-*m*-SABA and DNA-NC before (I) and after selection (IR). The recovery of each DNA-tagged ligand was also assessed after selection on empty beads (NP). Recoveries of DNA amounts are reported in percentage. C. Ratios of DNA-AAZ/DNA-NC and DNA-m-SABA/DNA-NC quantities before and after selection.

To correlate the qPCR findings with the sequence count features stemmed from a photocrosslinking screening experiment, we screened a library, comprising 669240 members, based on a previously described strategy. [98, 178] AAZ, SABA, Cl-SABA, and *m*-SABA were among the second set of 1430 building blocks chemically attached at the 5' extremity of a single-stranded DNA library based on a splint ligation procedure. (Figure 2.18A). This library was hybridized to a single partially complementary strand, featuring the diazirine photoreactive moiety at the 3' end and two abasic regions in correspondence of the DNA barcodes (Figure 5.10, Paragraph 5.1). The Klenow-fill-in polymerization allowed the library identifier (sequence codon C) to be transferred onto the complementary strand. Photocrosslinking selections against CAIX were performed using an input of 10^6 copies per library member. The eluted DNA was then analyzed by DNA sequencing. (Figure 2.18B) shows representative sequencing fingerprints after selection in UV-irradiation conditions (IR), no UV-irradiation (NIR) and on empty beads in the absence of protein (NP).

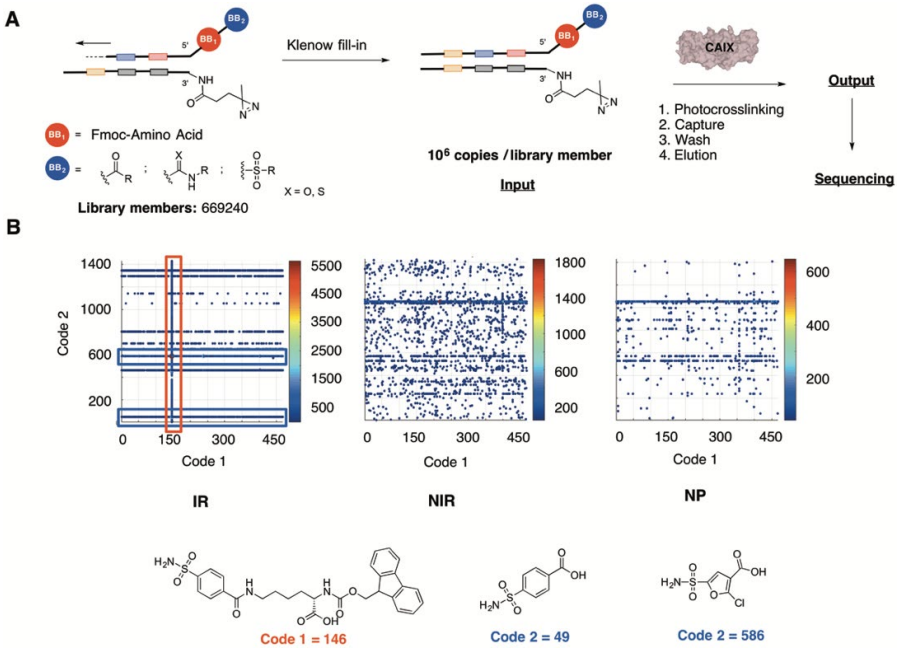


Figure 2.18. A. Schematic representation of a single pharmacophore DECL library, comprising 669240 members, displayed at the 5' end. Code 1: red; Code 2: blue; Code 3 (library identifier): orange. The individual library members are univocally identified by their codes 1 (BB1, ranging from 1 to 468) and 2 (BB2, ranging from 1 to 1430). The library was used as input in photocrosslinking screenings against polyhistidine-tagged CAIX. The selection output was evaluated by DNA sequencing. B. Selection fingerprints. The number of counts is correlated with the color of dots. Cutoff threshold set for the images: 40 counts. IR = UV-irradiation control; NIR = No UV-irradiation control; NP = No Protein control.

We did not observe a significant enrichment of CAIX ligands in the NIR (No UV-irradiation) control and after selection on empty beads (NP). These findings were consistent with the results obtained in the model selection experiments described above. By inspection of the UV-Irradiation control fingerprint (IR), we observed the selective enrichment of two combinations of building blocks (BB1/BB2 = 146/586; BB1/BB2 = 146/49), which corresponded to three sulfonamide derivatives. Those combinations exhibited an enrichment factor (E.F.) of 2803 ± 121 (corresponding to the highest E.F.) and 869 ± 75, respectively (Figure 2.18B). To validate the methodology, we performed fluorescence polarization measurements on the off-DNA small-molecule fluorescent derivatives and determined an experimental dissociation constant of 105 ± 4 nM (BB1/BB2 = 146/586) and 116 ± 6 nM (BB1/BB2 = 146/49 (Figure 5.17, Paragraph 5.1).

The photocrosslinking screening output was contrasted with the sequence count profile obtained with solid-phase affinity capture procedures, which are often used for DECL screening campaigns (Figure 2.19) (Paragraph 1.3.1).

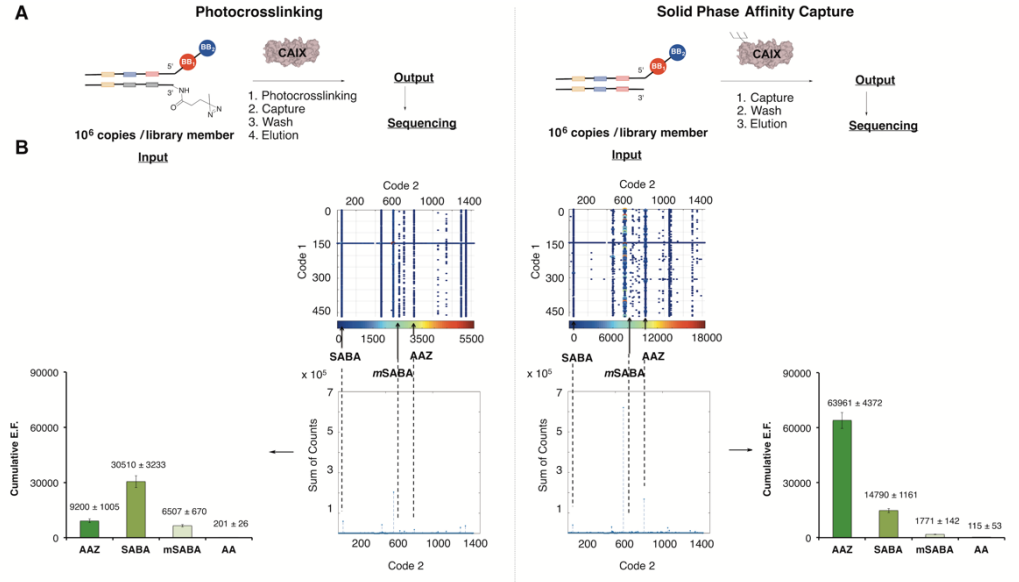


Figure 2.19. A. Schematic representation of DECL selections performed using photocrosslinking (left panel) and solid-phase affinity capture (right panel). B. Top panel: Sequencing fingerprints of library selections against his-CAIX. The cutoff threshold was set at 80 counts. CAIX ligands (AAZ, SABA and m-SABA) are identified by different codes 2. Bottom panel: Cumulative plots obtained with the sum of total sequence counts of the library members featuring CAIX ligands (AAZ, SABA, m-SABA). Cumulative enrichment factors (E.F.) were calculated dividing the cumulative counts by the average counts of single library members in each selection.

The single-stranded single pharmacophore library described in the previous section was annealed with a 34 mer-oligonucleotide comprising a distinct library-identifier codon (Figure 2.19A). After Klenow fill-in polymerization, the sequence codon was transferred onto the complementary strand to obtain a double-stranded DECL. This library was used as input for solid-phase affinity capture selections against the immobilized purified his-CAIX. After incubation of the library with the solid-supported protein, non-denaturing washes, the eluted DNA molecules were analyzed by DNA sequencing. Figure 2.19B features the "cumulative plots" defined as the cumulative counts of the sum of library members, featuring known CAIX-binding fragments (AAZ, SABA, *m*-SABA), compared to all other library members. The background signal of the procedure can be assessed from the cumulative counts of negative control-DNA conjugates carrying a terminal acetyl moiety (AA). The affinity capture method provided better discrimination between high-. medium- and low- affinity binders towards CAIX whereas the photocrosslinking method allowed better discrimination between low-affinity binders (featuring the *m*-SABA moiety) and background (AA) and provided higher enrichment factors of the medium and low-affinity ligands. The methodology may thus be used not only for the discovery of high-affinity ligands but also for single-digit micromolar binders, which are often undetectable with conventional affinity capture techniques.

2.2.3 Discussion and Conclusion

We have used CAIX ligands of various affinities ranging from nanomolar to micromolar, to examine the recovery and selectivity of selection procedures by qPCR and high throughput DNA sequencing. Conventional affinity capture procedures were compared with a screening strategy, based on photocrosslinking. Our study indicates that both affinity capture procedures and photocrosslinking methods can efficiently recover binders out of very large DEL libraries. The photocrosslinking methodology may be better suited for the identification of medium affinity binders (e.g., single-digit micromolar ligands). Photocrosslinking may stabilize the ligand-protein complex by locking the equilibrium irreversibly. [150–154, 165, 192]

The photocrosslinking method suffers from certain limitations. It appears that an upper limit for photoconversion may be reached, which cannot be improved by additional irradiation. [233–236] It is not known, at present, whether photocrosslinking may be further enhanced by the choice of different photoreactive groups (e.g., trifluoromethyl phenyl diazirines, benzophenone or aryl azide) or by the spatial proximity between protein and DNA-encoded ligand. [150, 151] Crosslinking efficiency might be improved by the use of benzophenone, as a photoreactive group, which has the discrete property of repeated photoactivation to form a triplet ketyl diradicals (not found in the case of diazirine or aryl azide) by prolonged UV-irradiation. [223]

The performance of the sequential use of photocrosslinking and affinity capture may ultimately be limited if the protein-ligand complex is not quantitatively recovered. [192] As an alternative, Li and coworkers developed a polymerase-extension based screening procedure against non-immobilized proteins. Indeed, it is worth noting that their current scheme is limited to single-stranded DNA-encoded chemical libraries with encoded compounds at the 3' end. [154]

The construction of DNA-encoded libraries featuring DNA in single-stranded format opens various opportunities (e.g., the use of complementary strands with photocrosslinkers, the conversion into the double-stranded format or the self-assembly with complementary libraries, yielding dual pharmacophore repertoires) [107–109, 117, 123] without any apparent disadvantage. We anticipate that DNA barcodes in single-stranded format may be broadly applicable for experimental goals, which have, so far, mainly been tackled with double-stranded DNA tags.

Encoded chemical libraries in single-stranded DNA format are versatile discovery tools that can be used for the construction of standard double-stranded libraries or for the implementation of photocrosslinking-based selection methodologies. Our study indicates that both affinity capture selection methods and photocrosslinking procedures are complementary. The latter methodology may be better suited for the identification of low-affinity binders (single-digit micromolar), that could be lost during the washing steps in conventional affinity capture procedures.

2.3 DEVELOPMENT OF A ONE-BEAD-ONE-COMPOUND DNA-ENCODED CHEMICAL LIBRARY (OBOC DECL) BASED ON SINGLE-STRANDED DNA.

2.3.1 Introduction

Combinatorial synthesis and "split-and-pool" strategies have allowed the production of large libraries of molecules, characterized by high molecular diversity. The solid-phase combinatorial synthesis [43–47] has enabled the generation of molecular collections of organic molecules known as "one-bead-one-compound" (OBOC) libraries. [237] Such collections display several copies of the same chemical compound on a single bead and have a great potential for the identification of binding molecules against targets of pharmaceutical interest. [238, 239] However, practical aspects such as automation, solvent compatibility, and chemical analysis may restrict its applicability in drug discovery. [240, 241] Typically, the elucidation of hit structures from library repertoires occurs by mass spectrometric (MS/MS) fragmentation analysis, which may be unfeasible for the assessment of stereochemical and regiochemical diversity. [242, 243]. Thus, various synthetic encoding tags (Paragraph 1.1.2) have been explored to circumvent such limitations, of which nucleic acids as encoding molecules have become particularly attractive. For this purpose, DNA-encoded solid-phase synthesis (DESPS) has emerged as a powerful tool for the generation of DNA-encoded OBOC libraries. [76, 170]. The use of DNA-encoding tags may facilitate the identification of hit compounds by PCR amplification and next-generation sequencing. [85]

In the early days of DESPS, libraries were primarily assembled on solid-phase using inorganic supports (e.g., controlled pore glass, CPG) for the synthesis of polypeptide chains. [76, 77]. The use of solid-phase matrixes allows the implementation of various reaction conditions (e.g., organic solvents), which are not suitable for the synthesis of DECL in solution. Moreover, such synthetic methodologies are in principle compatible with a much broader selection of chemical building blocks, which are compatible for the generation of more complex molecules (e.g., macrocyclic scaffolds, cyclic peptidomimetics) [244, 245]. The introduction of more advanced activity-based screening methodologies has sparked a renewed interest in DNA-encoded OBOC libraries, (e.g. DESPS), which have been collectively investigated by academia and industry. [246– 250].

MacConnell *et al.* [247] described a DESPS synthetic procedure for the preparation of libraries on TentaGel® Rink amide resins, comprising steps of reactions in organic solvents and enzymatic ligation of encoding dsDNA oligonucleotides. The synthesis requires the use of a bifunctional linker ("scaffold", e.g., Fmoc-protected alkyne), which once bound on beads is used for the orthogonal synthesis of organic molecules and DNA-encoding in "split-and-pool" fashion. The final DESPS product is a covalently resin-bound oligomeric product, carrying a DNA oligonucleotide tag.

A similar strategy was also employed by Paciaroni *et al.* [248] for the construction of an OBOC library based on an aldehyde scaffold, which can be converted into structurally diverse products (e.g., amines, carboxylic acids, alkynes) by using known synthetic reactions. The implementation of these chemical transformations requires minimum optimization, since it is virtually possible to use any organic solvent on beads. In contrast, solution-phase DECL synthesis requires the use of aqueous solvents, resulting in a limited scope of chemical reactions.

Brunschweiger *et al.* [249] developed a strategy for the synthesis of DNA-coupled isoquinolones and pyrrolidines on solid-phase by using ytterbium- and silver-mediated imine chemistry. The authors reported the use of a protected single-stranded DNA synthesized by standard phosphoramidite chemistry and bound to a controlled pore glass (CPG) matrix to initiate the construction of a DECL. After library synthesis in parallel fashion, the DNA conjugates are cleaved from the solid support and ligated to DNA hairpins, each containing encoding sequences for each heterocyclic structure. Recently, the same group reported a similar strategy for the implementation of Ugi-four-component reactions (Ugi-4CR) on solid-phase coupled DNA. [250]

As seen in the previous examples, OBOC libraries are typically encoded by sequential enzymatic ligation steps promoted by base-paired fragments. [247–250] However, the use of single-stranded DNA oligonucleotides (ssDNA) for the encoding of OBOC libraries has been narrowly investigated so far.

In this study, we have explored the feasibility of the single-stranded DECL encoding on a solid-phase. The generation of a solid-supported library may be advantageous compared to the "in-solution" library synthesis strategy [85, 95–99], which simplify the purification steps (e.g., omission of HPLC purifications) after iterated steps. Moreover, single-stranded (ssDNA) DECL libraries may be more flexible, allowing the construction of encoded self-assembling chemical libraries (ESAC), [107–108, 109, 117, 123] and innovative screening techniques, including interaction-dependent PCR [147] and photocrosslinking methodologies.

2.3.2 Results

We have first investigated a single-stranded DNA-encoding setting suitable for the synthesis of a representative OBOC DECL library (Figure 2.20) The 10 μm Monosized amino-functionalized TentaGel® microspheres were used as solid support. TentaGel®-based resins are composed by grafted copolymers consisting of low-crosslinked polystyrene matrix and polyethylene glycol (PEG), and their use has been described by several groups both in academia and industry. [251, 252] The hydrophobic and hydrophilic properties of the resin became attractive for the construction of DECL libraries.

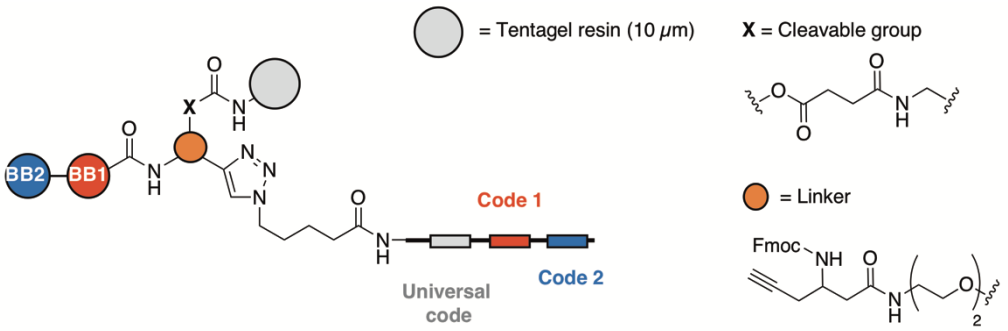


Figure 2.20. Design of a two-building block OBOC-DECL library. BB1: building block 1; BB2: building block 2.

The use of a trifunctional Fmoc-amino acid alkyne scaffold can be particularly convenient for the coupling on the solid support, chemical library synthesis (after cleavage of the Fmoc group) and DNA encoding. [247] A substoichiometric amount of the organic molecules displayed on the solid matrix is covalently linked to a 45-mer single-stranded DNA oligonucleotide ("universal code") through copper-catalyzed azide-alkyne cycloaddition (CuAAC) click chemistry. The universal DNA strand presents complementary bases for the subsequent splint ligation of phosphorylated sequences (e.g., code 1, code 2). The final product is a solid-phase-bound organic molecule tethered to an encoding DNA oligonucleotide, which can be released in solution after cleavage of the ester moiety ("cleavable group").

Such a design requires extensive optimization of critical steps (i.e. coupling on the solid support, splint ligation of ssDNA oligonucleotides performed on TentaGel®) in order to yield DECLs of high purity.

Therefore, we have first evaluated the efficiency of the coupling reaction on solid-phase and the enzymatic ligation step, which is required for the encoding of sets of building blocks displayed on DECLs. To initiate our studies, we adapted a synthetic procedure described by Paciaroni *et al.* [248], which was originally used for the synthesis of an OBOC DECL library based on dsDNA encoding. The synthetic procedure involves alternating coupling and ligation steps on TentaGel® in split-and-pool fashion.

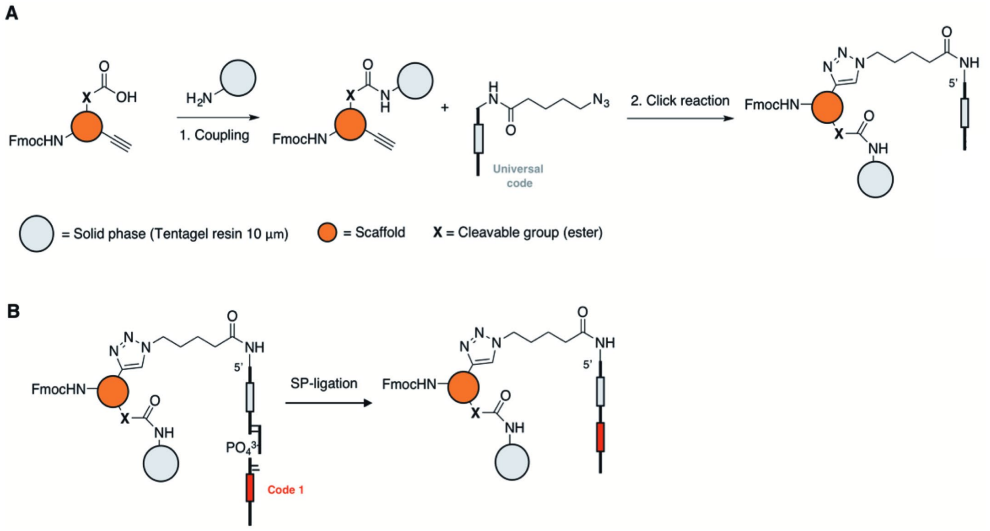


Figure 2.21. A. Synthetic sequence adopted for the synthesis of the DNA conjugate on solid-phase. B. Ligation procedure on solid-phase.

Figure 2.21 represents the trifunctionalized scaffold covalently immobilized on TentaGel® beads through amide bond formation using diisopropylcarbodiimide / ethyl cyano(hydroxyamino)acetate (DIC/OXYMA) as coupling reagent. The resulting solid-phase-bound alkyne derivative was then subjected to click reaction in CuAAC conditions with a substoichiometric amount of a 45-mer ssDNA azidopentoic derivative, to reach a DNA loading on beads of 1%. As reported in literature, [247] the use of higher loading capacities may be detrimental for the subsequent enzymatic ligation procedures, due to steric hindrance. The corresponding DNA derivative was then subjected to enzymatic splint ligation with a 3'-phosphorylated 29-mer oligonucleotide (code 1), which presents a 6-base pair-encoding region for the first set of building blocks. The implementation of denaturing washes conditions after the reaction is necessary for the complete removal of the DNA adapter oligonucleotide, which could remain bound to the product due to Watson-Crick base pairings.

The quantification of the products after each step was determined by LC-MS analysis and qPCR quantification, after cleavage of the linker with NaOH solution. The amount of DNA molecules amplified by qPCR is correlated to the number of sites accessible for enzymatic ligation, which can be limited by steric hindrance.

In this experimental setting, both the click and the ligation reaction were inefficient (2% and 1% of DNA recovery) as determined by qPCR quantification.

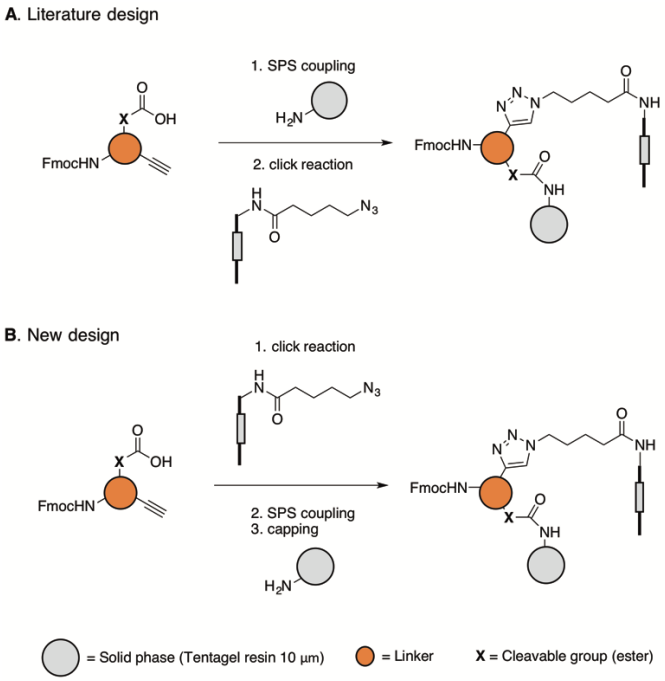
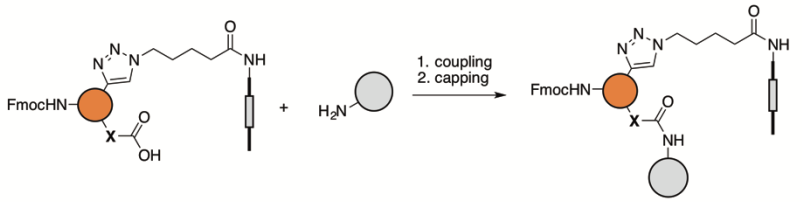


Figure 2.22 Encoding strategies used in literature (A) [248] and developed in this study (B) for the functionalization of TentaGel® resin beads with DNA oligonucleotides.

Therefore, we explored the possibility to perform the click reaction on the scaffold with the DNA-azido derivative to ensure a stoichiometric ratio of DNA encoding tags and small molecule scaffolds. The resulting DNA-scaffold derivative can be conveniently purified by HPLC and supported on TentaGel® beads by solid-phase amide bond formation. An additional step of capping of the unreacted amino groups is required to quench the beads (Figure 2.22).

First, we screened various conditions with the aim to improve the coupling step on solid-phase (Figure 2.23).



Entry	Concentration	Water*/DMF	% Load. Beads	Method	Yield
a	8 μM	1:1 + 0.04% Tween	1%	DMT-MM	2%
b	8 μM	1:1	1%	DMT-MM	5%
c	8 μM	1.5:1	1%	EDC/HOAt/NMM	25%
d	8 μM	1:4	1%	EDC/HOAt/NMM	67%
e	8 μM	4:1	1%	EDC/HOAt/NMM	4%
f	8 μM	1:1	5%	EDC/HOAt/NMM	2%
g	8 μM	1:1	0.25%	EDC/HOAt/NMM	8%
h	0.8 μM	1:1	1%	EDC/HOAt/NMM	1%
i	8 μM	1:1	1%	EDC/HOAt/NMM	18%
j	80 μM	1:1	1%	EDC/HOAt/NMM	2%
k	4 μM	1:10	1%	sNHS/EDC	87%

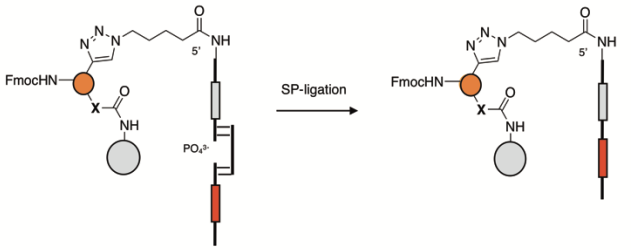
Figure 2.23. Optimization of the amino coupling reaction on TentaGel® beads.

We used various concentrations of reactants, coupling methodologies and loadings of substoichiometric amounts of the DNA-scaffold. As shown in Figure 2.23, the increase of the amount of DMF in the reaction mixture, led to better conversion of the substrate (entry d: 67% yield; entry e: 4%). The high swelling properties of TentaGel® resins allow the creation of enlarged reaction volumes, which can be critical for the chemical synthesis in high purities and yields. [253] The coupling procedure based on the N-hydroxysulfosuccinimide/1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (sNHS/EDC) method provided a better conversion in contrast with EDC/HOAt/MMM (entry k: 87%; entry d: 67%). However, using the former method, the reaction mixture contained more impurities, as assessed by LC-MS. Therefore, entry d was the best compromise in terms of quality and yield.

After the optimization of the coupling procedure, we performed a systematic evaluation of the ligation reaction conditions in order to improve the conversion of the product. We sought to investigate the effect of different experimental variables, including annealing and reaction temperature, catalytic amounts of T4 DNA ligase and quantity of organic solvent

in the reaction mixture. Each test was performed in separated wells, and the amount of ligation product was determined by qPCR analysis after NaOH cleavage. (Figure 2.24).

As shown in Figure 2.24, the best conversions were obtained including a pre-annealing step at 95 °C and increasing the reaction temperature to 30 °C and 37 °C. (entires o–r). Surprisingly, small amounts of organic solvent in the reaction mixture (e.g., DMF, DMA, THF), which are generally used for the swelling of resin beads, can lead to complete inhibition of the enzymatic reaction. Moreover, the quantity of the enzyme was downscaled to 2 U/pmol of substrate.



Entry	Ann. T (°C)	T. reaction (°C)	Org.solv. (%)	T4 Ligase (U/pmol)	Yield
a	25	21	0	2	19 %
b	25	21	0	3	18 %
c	25	21	0	4	19 %
d	95	21	0	2	48 %
e	95	21	0	3	30 %
f	95	21	0	4	29 %
g	60	21	0	2	23 %
h	60	21	0	3	33 %
i	60	21	0	4	33 %
j	95	21	0	2	59 %
k	95	21	15% DMF	2	9 %
l	95	21	50% DMF	2	0 %
m	95	21	15% DMA	2	0 %
n	95	21	15% THF	2	0 %
o	95	30	0	2	99 %
p	95	30	15% DMF	2	6 %
q	95	30	50% DMF	2	0 %
r	95	37	0	2	95 %
s	95	37	15% DMF	2	7 %
t	95	37	50% DMF	2	0 %

Figure 2.24 Optimization of the enzymatic ligation conditions using single-stranded DNA oligonucleotides. Ann. T.: Annealing Temperature.

2.3.3 Discussion and Conclusion

We have implemented a novel synthetic design for the synthesis of OBOC DECL libraries based on ssDNA encoding and supported on TentaGel® resin beads while optimizing coupling and ligation test reactions to assess and optimize the overall conversion of ligated product. Collectively, the results show that DESPS methodologies can be conveniently used for the synthesis of small molecule-DNA oligonucleotide derivatives bound to solid-phase matrixes. DESPS methodologies can be efficient for the production of DECL libraries up to four building blocks [247], which can be chemically coupled in order to synthesize polypeptide chains and macrocyclic structures. In recent years, various synthetic methodologies have been developed for the synthesis of complex heterocyclic and macrocyclic scaffolds (e.g., cyclopeptides, [244, 245] pyrimidines, [254] benzodiazepine-based scaffolds, [255] oxadiazoles [256]). Moreover, the use of high-throughput instrumentation (e.g., filter plates, automated liquid handling systems) enables both synthesis and purification of large libraries in an efficient time and cost manner. [247]

Recent advances in screening methodologies (e.g., Paragraph 1.3.3) have also contributed to the growing interest in solid-phase-bound libraries. Collections of chemical compounds displayed on resin can be conveniently screened on beads by fluorescence sorting and selected in functional screening assays against enzymatic targets. [169, 170] The presence of a cleavable linker between the organic molecules and the solid support may allow the release of the library in solution and therefore, the implementation of classical screening methodologies (e.g., affinity selections). (Paragraph 1.3.1–1.3.2)

The use of single-stranded DECL libraries opens various opportunities (e.g., ESAC libraries, photocrosslinking methodologies) [107–109, 117, 123]. The implementation of DESPS strategies for the synthesis of such collections can be beneficial in terms of time and reaction yields. However, optimal conditions for the enzymatic splint ligation and purification of resin-bound intermediates are required in order to reduce the amount of unreacted DNA oligonucleotides, which can reduce library purity and integrity.

In conclusion, this preliminary study provided a rationale for the construction of single-stranded DECLs on solid supports, which can be versatile synthetic tools for the production of chemical libraries on-beads.

3. CONCLUDING REMARKS AND FUTURE OUTLOOK

During the PhD work, we have implemented analytical and synthetic methodologies useful for the improvement of the DECL platform. In particular, most of the effort was focused on the optimization and development of novel screening procedures suitable for drug discovery using DECL libraries to enhance the quality of hit identification campaigns.

In the first part of the project (Paragraph 2.1), we have quantitatively evaluated the efficiency of screening procedures based on solid-phase affinity capture, which are used both in the industrial and academic setting. The recovery of CAIX ligands of different binding affinity and DECL libraries was assessed by qPCR quantification and analysis of the features of DNA sequencing fingerprints. The data suggested that high-affinity ligands (e.g., AAZ) are almost quantitatively recovered. In constrast, the identification of micromolar binders (e.g. *m*-SABA) may require additional and extensive optimization of selection parameters (e.g., type of target immobilization, capture conditions, library inputs) to minimize recovery losses (e.g., in the washes fraction). Nevertheless, the results indicated that both qPCR analysis and DNA sequencing results could be efficiently used for evaluation studies of the DECL technology.

These findings provided a rationale for the development of a novel DECL screening method, which can be exploited for the identification of micromolar hits. In the second part of the PhD work (Paragraph 2.2), we have developed and optimized a selection procedure using ssDNA DECL libraries and photocrosslinking reactions in solution. Overall, our results demonstrated that the procedure is suitable for the identification of low and medium affinity ligands (e.g., micromolar K_d). Therefore, the photocrosslinking methodology provided complementary information to conventional DECL affinity capture procedures on solid supports. Notably, the results provided evidence that ssDNA DECL libraries can be conveniently hybridized with chemically modified complementary oligonucleotides, which can be useful for many applications (e.g., ESAC libraries, affinity maturation and photocrosslinking selections).

For this reason, in the final part of the project (Paragraph 2.3), we focused our attention on the development of a novel synthetic route for the generation of ssDNA DECL libraries in OBOC format. Such libraries can be conveniently screened using the procedures

described in this thesis (e.g., affinity capture, photocrosslinking) or using activity-based screening assays. The preliminary results have shown that DESPS approaches are versatile for the synthesis of ssDNA DECL libraries in a fast and convenient manner.

Collectively, the results presented in this thesis provided a rationale and motivation for the improvement of the DECL platform, which can be versatile for the discovery of specific ligands against "difficult-to-drug" protein targets.

Future research should consider the potential effect of screening parameters for the identification of high-, medium-, and low-affinity binders, and therefore the use of multiple selection conditions for the confirmation of putative binders to proteins of interest (e.g., quantity of library input, protein concentration, buffer, solid support for affinity capture, wash and elution conditions). As an example, high and low protein amounts in a screening assay may facilitate the identification of weak and strong small molecule binders respectively. Therefore, it may be useful to perform DECL selections at different concentrations. Moreover, each protein target may require orthogonal screening methodologies based on alternative capture procedures (e.g., screening in solution), in order to minimize the risk of selection artefacts (e.g., unspecific binding, presence of PAINS).

DECL library selections based on photocrosslinking represent a promising tool for the identification of low- affinity ligands, thanks to the irreversible stabilization of protein-ligand complexes in solution. In future, alternative crosslinkers may be implemented (e.g., electrophilic crosslinkers, photoreactive groups) to improve the conversion yield, which can be limited by side-reactions of the intermediate carbene species. The design of the photoreactive library described in this thesis can be in principle applied for the encoding of a variety of crosslinkers (e.g., displayed on the 3' end of the photoreactive strand), thus allowing the identification in parallel of the most effective probe in terms of ligand recovery and sequencing counts. Future studies could investigate the use of photocrosslinkers for the screening of difficult protein targets in a native cell environment (e.g., GPCRs). The synthetic approach, described in this thesis, is compatible with the use of cell-penetrating peptides (e.g., cPP12), [165] which can be covalently attached to the 5' end of the photoreactive DNA strand, making the library permeable to the cell membrane.

However, it is worth noting that novel screening methodologies should be developed in parallel with innovative library technologies in order to accelerate drug discovery processes. The application of orthogonal technologies (e.g., self-assembling chemical libraries, photocrosslinking methodologies, DESPS procedures) is likely to be relevant for the discovery of chemical leads of high potency and favourable characteristics.

4. MATERIALS AND METHODS

4.1 CHAPTER 2.1: GENERAL PROCEDURES

4.1.1 List of the oligonucleotides and PCR primers

Custom oligonucleotides were purchased by LGC Biosearch Technologies, Inc and delivered as lyophilized solids. The solids were further purified by EtOH precipitation and redissolved in H_2O. The final concentration was determined by UV absorbance measurement at 260 nm using a NanoDrop instrument.

Oligonucleotide A

5'-GGAGCTTCTGAATTCTGTGTGCTGCGCGCGCGAGTCCCATGGCGC-3'

Modification: 5'-C6-Aminolink

MW: 14056 Da

Oligonucleotide B

5'-CGGATCGACGCGCGCGCGCGTCAGGCAGCTTGAGTGGCGACCGTGCAGAGC-3'

Modification: 5'-Phosphorylated

MW: 15888.2 Da

Oligonucleotide C

5’-GGAGCTTCTGAATTCTGTGTGCTGCGCGCGCGAGTCCCATGGCGCCGGATCGAC
GCGCGCGCGCGTCAGGCAGCTTGAGTGGCGACCGTGCAGAGC-3’

Modification: 5’-C6-Aminolink

MW: 29927.2 Da

Oligonucleotide D

5’-GCTCTGCACGGTCGCCACTCAAGCTGCCTGACGCGCGCGCGCGTCGATCCGGCG
CCATGGGACTCGCGCGCGCAG CACACAGAATTCAGAAGCTCC-3’

MW: 29464 Da

Oligonucleotide E

5’-AACATCCTCAGTCATCACACATCGTATCCACGAGCGTCAGGCAGCGAGCAAGAT
GCGTGCGTGTGCA-3'

Modification: 5’- C6-Aminolink

MW: 20785.4 Da

Oligonucleotide F

5’-TGCACACGCACGCATCTTGCTCGCTGCCTGACGCTCGTGGATACGATGTGTGATG
ACTGAGGA-3’

MW: 19430.6 Da

PCR primers

Primer_DL_1:

5’-GGAGCTTCTGAATTCTGTGTGCT-3’

Primer_DL_2:

5’-CTCTGCACGGTCGCCACTC-3’

Primer_NC_1:

5'-ACATCCTCAGTCATCACACATCG-3'

Primer_NC_2:

5'-GCACACGCACGCATCTTG-3'

Sequences of the forward and the sequence-specific reverse primers (RP) used to quantify AAZ, SABA and *m*-SABA in the library (LB) experiments. The specific sequences are highlighted in bold.

Forward primer LB:

5'-GGAGCTTCTGAATTCTGTGTGCTG-3'

Reverse primer_AAZ

5'-CTGCACGGTCGC**GCGAGAC**-3'

Reverse primer SABA

5'-TGCACGGTCGC**CGCTGCT**-3'

Reverse primer *m*-SABA

5'-CTGCACGGTCGC**CCTTCAT**-3'

Reverse primer_AA

5'-GCTCTGCACGGTCGC**ATACGGA**-3'

Sequences of the forward and the reverse primers used to amplify the library for the total DNA quantification:

Forward primer LB:

5'-GGAGCTTCTGAATTCTGTGTGCTG-3'

Reverse primer LB:

5'-GCTCTGCACGGTCGC-3'

4.1.2 Procedures for on-DNA reactions

4.1.2.1 *Ethanol Precipitation*

To aqueous DNA solutions, 10% (v/v) of 5 M NaCl was added, followed by 2.5–3 volumes of cold EtOH. The colloidal solution was left for 1 h at −20 °C and then centrifuged at 4 °C for 30 min. at 14000 rpm. The resulting supernatant was discarded, and the pellet was rinsed once with cold 80% EtOH. After centrifugation for another 5 min at 4 °C at 14000 rpm, the supernatant was discarded, and the pellet was dried using a SpeedVac. The recovered samples were dissolved in an appropriate buffer for subsequent analysis or experiments. GP1 was generally performed after each chemical reaction.

4.1.2.2 *DNA-conjugation of Fmoc-amino acids as building block 1 (DNA-BB1)*

Fmoc-amino acid in DMSO (12.5 µl, 200 mM), *N*-hydroxysulfosuccinimide (s-NHS) (20 µl, 333 mM) in DMSO/H_2O (2:1), followed by 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) in DMSO (24 µl, 100 mM) were added in DMSO (200 µl) and stirred for 30 min at 30 °C. Amino-modified 45-mer oligonucleotide in H_2O (4 µl, 0.15 nmol) in MOPS buffer (50 µl, 100 mM, pH 8.0, 1 M NaCl) was then added. The reaction mixture was stirred for 16 h at 37 °C. The DNA was precipitated following the procedure described in Paragraph 4.1.2.1 and was purified by RP-HPLC on an XTerra® C18 semipreparative using a gradient of eluent A (TEAA 100 mM) and eluent B (TEAA 100 mM in 80% ACN). The fractions containing the product were combined and lyophilized to obtain **DNA-BB1**, as determined by measuring the UV absorbance at 260 nm of a water solution on a Thermofisher Nanodrop 2000.

Procedures for DNA-conjugation of carboxylic acids /sulfonylchlorides /iso(thio)cyanates/ amines as second diversity elements (DNA-BB1-BB2)

4.1.2.3 *Amide bond formation method with DMT-MM*

Carboxylic acid in DMSO (12.5 µl, 200 mM), 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride (DMT-MM, 12 µl, 200 mM) were dissolved in 32.5 µl of DMSO and stirred for 15 min at 25 °C. DNA in H_2O (4 µl, 0.15 nmol) diluted in MOPS buffer (3-(*N*-morpholino)propanesulfonic acid, 35 µl, 100 mM, 1M NaCl, pH = 8) and H_2O (37 µl) was added to the mixture and the reaction was stirred at 25 °C for 16 h. The DNA was precipitated following the procedure described in Paragraph 4.1.2.1 and analyzed by LC-MS.

4.1.2.4 *Amide bond formation method with EDC/HOAt/DIPEA*

Carboxylic acid in DMSO (12.5 µl, 200 mM), EDC (4 µl, 300 mM), 1-Hydroxy-7-azabenzotriazole (HOAt, 4 µl, 60 mM) and DIPEA (4 µl, 300 mM) were dissolved in 32.5 µl of DMSO and stirred for 15 min at 25 °C. DNA in H_2O (4 µl, 0.15 nmol) was diluted in MOPS buffer (35 µl, 100 mM, 1M NaCl, pH = 8) and H_2O (37 µl) was added to the mixture, and the reaction was stirred for 16 h at 37 °C. The reaction was quenched with NH_4OAc (25 µl, 500 mM) and agitated for 30 min at 37 °C. The DNA was precipitated following the procedure described in Paragraph 4.1.2.1 and analyzed by LC-MS.

4.1.2.5 *Amide bond formation method with s-NHS/EDC*

Carboxylic acid in DMSO (12.5 µl, 200 mM), s-NHS (20 µl, 333 mM) in DMSO/H_2O (2:1), followed by EDC in DMSO (24 µl, 100 mM) were added in DMSO (200 µl) and stirred for 30 min at 30 °C. DNA in H_2O (4 µl, 0.15 nmol) diluted in triethylamine hydrochloride buffer (TEA•HCl, 50 µl, 500 mM, pH = 10.0) was then added. The reaction mixture was stirred for 16 h at 37 °C. DNA was precipitated following the procedure described in Paragraph 4.1.2.1 and was analyzed by LC-MS.

4.1.2.6 *Reductive amination reaction*

4-formylbenzoic acid (12.5 µL 200 mM) in DMSO (200 µl), s-NHS (20 µl, 333 mM) in DMSO/H_2O (2:1), followed by EDC in DMSO (24 µl, 100 mM) were added in DMSO (200 µl) and stirred for 30 min at 30 °C. DNA in H_2O (4 µl, 0.15 nmol) diluted in TEA•HCl (50 µl, 500 mM, pH = 10.0) was then added. The reaction mixture was stirred for 16 h at 37 °C. DNA was precipitated following the procedure described in Paragraph 4.1.2.1, and used for the second step without any further purification. The crude pallet was dissolved in H_2O (6 µl, 0.1 nmol) and in phosphate buffer (5 µl, 2 M, pH = 5.5). Amine in DMSO (25 µl) was added to the DNA mixture, and the reaction was heated up to 40 °C for 3h. $NaCNBH_3$ in CH_3CN (3.5 µl, 100 mM) was then added, and the reaction was agitated overnight at 40 °C. DNA was precipitated following the procedure described in Paragraph 4.1.2.1 and was analyzed by LC-MS.

4.1.2.7 *Reverse coupling reaction*

Succinic anhydride in DMSO (15 µl, 200 mM) and DMAP (1.2 µl, 200 mM) were added to a solution of DNA in H_2O (5 µl, 20 nmol,) and TEA•HCl buffer (10 µl, 500 mM, pH = 10.0). The reaction mixture was stirred for 3 h at 60 °C. After completion of the reaction by LC-MS, the DNA conjugate was precipitated using procedure described in Paragraph 4.1.2.1**,** and used for the second step without any further purification. The crude pallet was dissolved in H_2O (37 µl, 0.1 nmol), MOPS buffer (35 µl, 100 mM, 0.5 M NaCl, pH = 8.0), amine in DMSO (15 µl, 100 mM) followed by DMT-MM (12 µl, 208 mM) were added. The reaction mixture was agitated overnight at 37 °C. DNA was precipitated following the procedure described in Paragraph 4.1.2.1 and was analyzed by LC-MS.

4.1.2.8 *Carbamoylation with iso(thio)cyanates*

DNA in H_2O (5 µl, 0.1 nmol), iso(thio)cyanate dissolved in DMSO or CH_3CN (10.5 µl, 100 mM) was diluted in Sodium Borate buffer (26 µl, 500 mM, pH = 9.4). The reaction mixture was stirred at 40 °C overnight. DNA was precipitated following the procedure described in Paragraph 4.1.2.1and was analyzed by LC-MS.

4.1.2.9 *Sulfonylation with sulfonyl chlorides*

DNA (0.1 nmol) was diluted in Sodium Borate buffer (26 µl, 500 mM, pH 9.4) and sulfonyl chloride in CH_3CN (10.5 µl, 100 mM) was added, and the reaction mixture was stirred overnight at 40 °C. DNA was precipitated following the procedure described in Paragraph 4.1.2.1 and was analyzed by LC-MS.

4.1.2.10 *Klenow polymerization reaction*

The pool of 246 DNA-amino acid conjugates (0.1 nmol) was dissolved in water (75.4 µl) and MOPS buffer (10 µl, 100 mM, pH 7.9). The second set of oligonucleotides (100 µM, Code 2-n, 1 < n < 1500, 5'-GCT CTG CAC GGT CGC **XXXXXXX** GCT GCC TGA CGC-3'), which allowed the unambiguous identification of BB2, was added. The solution was stirred at room temperature for 15 min. dNTPs (10 µl, 5 mM) and Polymerase (54 units per µl, 1.5 µl) were then added to the DNA solution. The mixture was left for 1 h at room temperature. All the samples were analyzed by ion-exchange cartridge gel with QiAxcel Advanced Instrument (QIAGEN).

4.1.3 Synthesis of DNA-tagged ligands

4.1.3.1 *Ligation procedure (synthesis of oligonucleotide C)*

Oligonucleotide **A** (1 eq., 21.7 µl, 5 nmol) and oligonucleotide **B** (1.5 eq, 21.4 µl, 7.5 nmol), DNA-RNA adapter in H_2O (1.58 eq, 9.2 µl, 7.9 nmol), 10x ligation reaction buffer (20 µl, New England Biolabs) and H_2O (145 µl) were mixed and heated up to 50 °C for 15 min. The mixture was allowed to cool down to 22 °C and T4 DNA-ligase (5000 U, 12.5 µl, New England Biolabs) was then added. The ligation process was left at room temperature for 16 hours. The reaction was then heated up to 65 °C for 10 min to inactivate the ligase. DNA was precipitated following the procedure described in Paragraph 4.1.2.1. The crude mixture was purified by RP-HPLC on an XTerra® C18 semipreparative using a gradient of eluent A (TEAA 100 mM) and eluent B (TEAA 100 mM in 80% ACN). The fractions containing the product were combined and lyophilized to obtain the desired ligated product, as determined by measuring the UV absorbance at 260 nm of a water solution on a Thermofisher Nanodrop 2000. LC-ESI-MS: 29926 m/z, found: 29926.09 m/z.

4.1.3.2 *Conjugation of 4-sulfamoyl benzoic acid (SABA) and 3-sulfamoyl benzoic acid (m-SABA) with oligonucleotide C (DMT-MM method)*

Reaction with amino-modified 96-mer oligonucleotide was performed using the procedure described in Paragraph 4.1.2.3. The DNA conjugate was purified by 2x EtOH precipitation and purified by DNA-Clean Up spin column (QIAquick PCR Purification Kit, Qiagen) and subsequently analyzed by ESI-LC-MS. ESI-LC-MS (SABA-DNA): 30109.2 m/z, found: 30107.69 m/z. ESI-LC-MS (*m*-SABA-DNA): 30109.2 m/z, found: 30107.57 m/z.

4.1.3.3 *Conjugation of 4-Oxo-4-((5-sulfamoyl-1,3,4-thiadiazol-2-yl)amino)butanoic acid (AAZ) with oligonucleotide C (EDC/S-NHS method)*

Reaction with amino-modified 96-mer oligonucleotide was performed using the procedure described in Paragraph 4.1.2.5. The DNA conjugate was purified by 2x EtOH precipitation and purified by DNA-Clean Up spin column (QIAquick PCR Purification Kit, Qiagen) and subsequently analyzed by ESI-LC-MS. ESI-LC-MS: (AAZ-DNA, 30189.31 m/z, found: 30188.27 m/z).

4.1.3.4 *Thermal hybridization*

Oligonucleotides **C**, **D**, **E**, **F** were first diluted at a concentration of 10^{10} copies of DNA/µl. Equimolar amounts of complementary oligos (oligonucleotide-tagged ligand **C**/**D** and oligonucleotide **E**/**F**, respectively) were mixed and heated up at 95 °C for 10 min. The mixture was allowed to cool down slowly at room temperature. The products were stored at –20°C and directly used for the affinity selections assays.

4.1.4 Affinity selections procedures

4.1.4.1 *Affinity selection against biotinylated CAIX (KingFisher)*

Streptavidin-coated magnetic beads C1-type (10 µl suspensions) were washed with 3 x PBS (Phosphate Buffer Saline, 500 µl, 50 mM sodium phosphate, 100 mM NaCl, pH = 7.4) and resuspended in PBS (100 µl). By using a KingFisher magnetic particle processor, the magnetic beads were transferred to a solution of PBS and biotinylated target (CAIX) at a certain concentration (1 µM, 2 µM) and incubated for 30 min with continuous gentle mixing. The beads were washed (2 x 3 min, 200 µl) with PBST-biotin (50 mM sodium phosphate, 100 mM NaCl, pH 7.4, 0.05% - 0.1% Tween-20, 100 µM biotin) and washed once (1 x 3 min) with PBST (200 µl, 50 mM sodium phosphate, 100 mM NaCl, pH = 7.4, 0.05% - 0.1% Tween-20). The protein-coated beads were transferred to a solution of the DNA conjugate and the negative control (10^{11} copies, ratio 1:1) in PBST (100 µl) containing

0.3 mg/ml herring sperm and incubated for 1 h with continuous gentle mixing. **For the library selections:** The library (10^x copies /lib. member) was diluted in PBST (100 µl) containing 0.3 mg/ml herring sperm and incubated for 1 h with continuous gentle mixing. The beads were removed and washed with PBST (5 x 30 s, 200 µl) and incubated with TRIS buffer (100 µl, 20 mM, pH = 8.4). The DNA conjugates were released by heat denaturation of the proteins (95 °C, 10 min) and the beads were separated. The eluate was analyzed by quantitative PCR and/or amplified by PCR and submitted to the Functional Genomics Center in Zürich for high-throughput DNA sequencing on an Illumina HiSeq 2500.

4.1.4.2 *Affinity selection against polyhistidine-tagged CAIX (KingFisher)*

Dynabeads His-Tag Pull-Down (Invitrogen, 10103D, 12 µl suspension) were washed with 3 x binding buffer (500 µl, 50 mM sodium phosphate, 300 mM NaCl, pH = 8, 0.01%/ 0.1% Tween-20) and resuspended in binding buffer (100 µl). By using the KingFisher magnetic particle processor, the magnetic beads were transferred to a solution of binding buffer and his-tagged target (CAIX) at a certain concentration (1 µM / 2.5 µM) and incubated for 30 min with continuous gentle mixing. The beads were washed (3 x 3 min, 200 µl) with binding buffer and subsequently transferred to a solution of the DNA conjugate and the negative control (10^{11} copies, ratio 1:1) in pull-down buffer (100 µl, 3.25 mM sodium phosphate, 70 mM NaCl, pH = 7.4, 0.01%/ 0.1% Tween-20), containing 0.3 µg/ml herring sperm. **For the library selections**: The library (10^x copies /lib. member) was diluted in in pull-down buffer (100 µl, 3.25 mM sodium phosphate, 70 mM NaCl, pH = 7.4, 0.01%/ 0.1% Tween-20), containing 0.3 µg/ml herring sperm. The beads were incubated for 1 h with continuous gentle mixing. The beads were removed and washed with pull-down buffer (5 x 30 s, 200 µl) and incubated with elution buffer (100 µl, 300 mM imidazole, 50 mM sodium phosphate, 300 mM NaCl, pH = 8, 0.01% Tween-20) for 30 min. After the incubation, the beads were separated. The eluate was analyzed by quantitative PCR and/or amplified by PCR and submitted to the Functional Genomics Center in Zürich for high-throughput DNA sequencing on an Illumina HiSeq 2500.

4.1.4.3 *Affinity selection against biotinylated CAIX (PhyNexus)*

The DNA conjugate and a negative control (10^{11} copies, ratio 1:1) were incubated for 30 min at room temperature with the biotinylated target (CAIX, 50 µl, 2 µM) in B&W selection buffer (50 mM Sodium Phosphate, 100 mM NaCl, pH = 7.4, 0.05% Tween-20) containing 0.2 mg/ml herring sperm. **For the library selections:** Library (10^x copies/ lib. member) was incubated with the biotinylated target (CAIX, 50 µl, 2 µM) in B&W selection buffer (50 mM Sodium Phosphate, 100 mM NaCl, pH = 7.4, 0.05% Tween-20) containing 0.2 mg/ml herring sperm. Streptavidin resin tips columns (5 µl) were pre-equilibrated with 3 x 200 µl Binding & Wash buffer (B&W, 50 mM Sodium Phosphate, 0.05% Tween-20, 100 mM NaCl, pH = 7.4). After pre-equilibration, the tips were dipped into the target solution (6 cycles, 90 µl). After the incubation, tips were washed with 8 x 200 µl B&W buffer (1 cycle, 190 µl) and released into the elution buffer (100 µl, TRIS, 10 mM, pH 8.5). The eluate

was analyzed by quantitative PCR and/or amplified by PCR and submitted to the Functional Genomics Center in Zürich for high-throughput DNA sequencing on an Illumina HiSeq 2500.

4.1.4.4 *Affinity selection against polyhistidine-tagged CAIX (PhyNexus)*

The DNA conjugate and a negative control (10^{11} copies, ratio 1:1) were incubated for 30 min at room temperature with the histidine-tagged target (CAIX, 50 µl, 2 µM) in pull-down buffer (50 mM Sodium Phosphate, 100 mM NaCl, pH = 7.4, 0.01% Tween-20) containing 0.3 mg/ml herring sperm. Streptavidin resin tips columns (5 µl) were pre-equilibrated with 3 x 200 µl Binding & Wash buffer (B&W, 50 mM Sodium Phosphate, 0.01% Tween-20, 100 mM NaCl, pH = 7.4). After pre-equilibration, the tips were dipped into the target solution (6 cycles, 90 µl). **For the library selections**: Library (10^{x} copies/ lib. member) was incubated with the histidine-tagged target (CAIX, 50 µl, 2 µM) in pull-down buffer (50 mM Sodium Phosphate, 100 mM NaCl, pH = 7.4, 0.01% Tween-20) containing 0.3 mg/ml herring sperm. After the incubation, tips were washed with 8 x 200 µl B&W buffer (1 cycle, 190 µl) and released into the elution buffer (TRIS, 100 µl, 10 mM, pH = 8.5). The eluate was analyzed by quantitative PCR and/or amplified by PCR and submitted to the Functional Genomics Center in Zürich for high-throughput DNA sequencing on an Illumina HiSeq 2500.

4.1.5 Quantitative PCR General Procedure

In an Optical 8-Cap Strip (Applied Biosystems) were combined in a total volume of 25 µl: DNA template (2.5 µl) (selection eluate or water for the no-template control "NTC"), 200 nM forward primer (0.5 µl, 10 µM in H_2O), 200 nM reverse primer (0.5 µl, 10 µM in H_2O), PowerUP SYBR Green Master Mix (12.5 µl, Applied Biosystems no. 100031508) and 9 µl H_2O. All qPCR experiments performed with SYBR Green were conducted at 50 °C for 2 min, 95 °C for 2 min, and then 40 cycles of 95 °C for 15 sec (denaturation), 60 °C for 1 min (annealing/extension). The specificity of the reaction was checked by melt curve analysis or by gel electrophoresis. The corresponding C_t values were correlated with the absolute quantity of DNA through standard curve calibration experiments.

4.2 CHAPTER 2.2: GENERAL PROCEDURES

4.2.1 List of the oligonucleotides and PCR primers

Elib2_1

5'-GGAGCTTCTGAATTCTGTGTGCTGAAACGTCGAGTCCCATGGCGCAGC-3'

Modification: 5'-C6-Aminolink

MW: 14994.6 Da

Elib2_2

5'-GGAGCTTCTGAATTCTGTGTGCTGAAACTGCGAGTCCCATGGCGCAGC-3'

Modification: 5'-C6-Aminolink

MW: 14994.6 Da

Elib2_3

5' – GGAGCTTCTGAATTCTGTGTGCTGAAAGCTCGAGTCCCATGGCGCAGC – 3'

Modification: 5'-C6-Aminolink

MW: 15043.7 Da

Elib2_4

5'-GGAGCTTCTGAATTCTGTGTGCTGAAAGAGCGAGTCCCATGGCGCAGC-3'

Modification: 5'-C6-Aminolink

MW: 14994.6 Da

Elib2_5

5'-GGAGCTTCTGAATTCTGTGTGCTGACATCGCGAGTCCCATGGCGCAGC-3'

Modification: 5'-C6-Aminolink

MW: 14970.7 Da

d-spacer

5'-CATGGGACTCGddddddCAGCACACAGAATTCAGAAGCTCC-3'

Modification: 5' - $(PO_4)^{3-}$; 3'-C6-Aminolink

MW: 12058.74 Da

E4 DNA – RNA chimeric adapter

5'-CGA **GUC** CCA TGG C**GC AG**C TGC-3', **bold:** RNA portions

MW: 6520.17 Da

FAM-Elib4

5'-CCTGCATCGAATGGATCCGTGGTCGAATTGCAGCTGCGC-3'

Modification: 5'- FAM

MW: 12497.87 Da (no counter ions); 12537.26 Da (fully protonated)

Elib4_1

5'-TCCTGCATCGAATGGATCGATACTCGCTGGCAGCTGCGC-3'

MW: 11959.78 Da

Elib4_2

5'-ATAACTTCACTGCCGTGTCCATGAACAGAGCAGCTGCGC-3'

MW: 11936.79 Da

Elib4_3

5'-AAGTCTCGTCAATTCACACTGTTGGCGATGCAGCTGCGC-3'

MW: 11958.80 Da

Elib4_4

5'-ACGTACGTCTCATGTGATGCACATGTATCGCAGCTGCGC-3'

MW: 11958.80 Da

Elib4_5

5'-GCTTGGGTGTATTGCACTAGCGTCAAGGCGCAGCTGCGC-3'

MW: 12055.34 Da

Elib6_code

5'-GCTCTGCACGGTCGCCTGAGATGTAGGATCACGCTGCCTGACGCddddddCGTCG ATCCGGCGC-3'

MW: 19024.57 Da (no counter ions); 19089.21 Da (fully protonated)

DNA adapter_Elib6

5'-CGAGTCCCATGGCGCCGGATCGACG-3'

MW: 7669 Da

Library Construction Code 1

5'-GGAGCTTCTGAATTCTGTGTGCTGXXXXXXCGAGTCCCATGGCGC-3'

Library Construction Code 2

5'-CGGATCGACGYYYYYYYGCGTCAGGCAGC-3'

Modification: 5' - $(PO_4)^{3-}$

Library Construction Code 3

5'-GCTCTGCACGGTCGCCTGAGATGCTGCCTGACGC-3'

MW: 10411. 8 Da

DNA Adapter Code1_Code2

5'-CGATCCGGCGCCAT-3'

MW = 4224.78 Da

PCR Primers

LB_FP

5'-GGAGCTTCTGAATTCTGTGTGCTG-3'

Primer_ESAC qPCR

5'-CCTGCATCGAATGGATCCGTG-3'

Primer_AAZ_1

5'-TCTGAATTCTGTGTGCTGAAACGT-3'

Primer_AAZ_2

5'-TCCTGCATCGAATGGATCGATACT-3'

Primer_SABA_1

5'-TCTGAATTCTGTGTGCTGAAACTG-3'

Primer_SABA_2

5'-ATAACTTCACTGCCGTGTCCATGA-3'

PrimerClSABA_1

5'-TCTGAATTCTGTGTGCTGAAAGCT-3'

PrimerClSABA_2

5'-AAGTCTCGTCAATTCACACTGTTG-3'

PrimermSABA_1

5'-TCTGAATTCTGTGTGCTGAAAGAG-3'

PrimermSABA_2

5'-ACGTACGTCTCATGTGATGCACAT-3'

Primer_NC_1

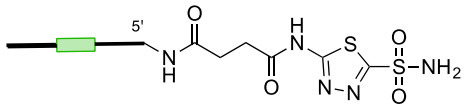
5'-GGAGCTTCTGAATTCTGTGTGCTGACATCG-3'

Primer_NC_2

5'-GCTTGGGTGTATTGCACTAGCGTC-3'

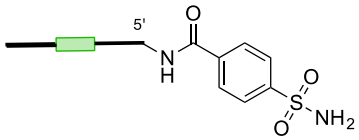
4.2.2 Synthesis of DNA-tagged ligands for model selection experiments

4.2.2.1 *Synthesis of AAZ_Elib2_1*



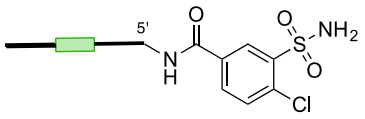
AAZ carboxylic acid derivative (4-Oxo-4-((5-sulfamoyl-1,3,4-thiadiazol-2-yl)amino)butanoic acid) in DMSO (12.5 µl, 200 mM), s-NHS (20 µl, 333 mM) in DMSO/H_2O (2:1), followed by EDC in DMSO (24 µl, 100 mM) were added in DMSO (200 µl) and stirred for 30 min at 30 °C. The oligonucleotide (Elib2_1, 25 µl, 0.66 mM in H_2O) diluted in triethylamine hydrochloride buffer (TEA•HCl, 50 µl, 500 mM, pH = 10.0) was then added. The reaction mixture was stirred for 16 h at 37 °C. The DNA conjugate was purified by EtOH precipitation, followed by RP-HPLC. ESI-LC-MS analysis: theoretical m/z: 15256.87 m/z; found m/z: 15255.98.

4.2.2.2 *Synthesis of SABA_Elib2_2*



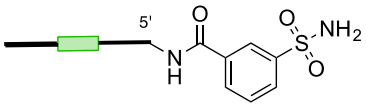
SABA (4-sulfamoyl benzoic acid) in DMSO (12.5 µl, 200 mM), 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride (DMT-MM, 32.5 µl, 208 mM in H_2O) were dissolved in 12 µl of DMSO and stirred for 15 min at 25 °C. The oligonucleotide (Elib2_2, 25 µl, 0.66 mM in H_2O) diluted in MOPS buffer (3-(*N*-morpholino)propanesulfonic acid, 35 µl, 100 mM, 1M NaCl, pH = 8.0) and H_2O (12 µl) was added to the mixture, and the reaction was stirred for 16 h at 25 °C. The DNA was purified by EtOH precipitation, followed by RP-HPLC. ESI-LC-MS analysis: theoretical m/z: 15176.89; found m/z: 15176.99.

4.2.2.3 *Synthesis of Cl-SABA_Elib2_3*



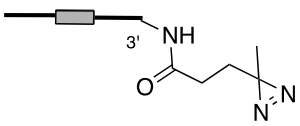
Cl-SABA (4-chloro-3-sulfamoyl benzoic acid) was conjugated to the oligonucleotide Elib2_3, as previously described in Paragraph 4.2.2.2. ESI-LC-MS analysis: theoretical m/z: 15212.24; found m/z: 15211.63.

4.2.2.4 *Synthesis of m-SABA_Elib2_4*



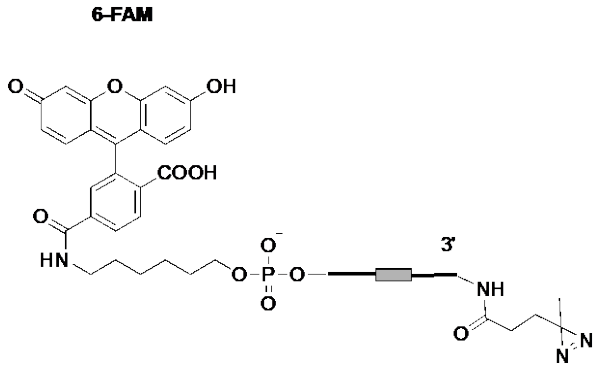
m-SABA (3-sulfamoyl benzoic acid) was conjugated to the oligonucleotide Elib2_4, as previously described in Paragraph 4.2.2.2. ESI-LC-MS analysis: theoretical m/z: 15226.9; found m/z: 15226.25.

4.2.2.5 *Synthesis of DA-d-spacer (DA-d-spacer)*



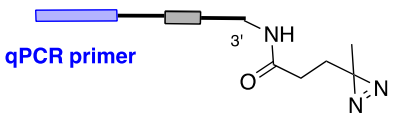
Sulfo-SDA (Sulfo NHS Diazirine, 5 µl, 200 mM in H_2O), d-spacer (65 µl, 304 µM) were dissolved in triethylamine hydrochloride buffer (TEA•HCl, 45 µl, 500 mM, pH = 10.0). The reaction was stirred overnight at 37 °C. The DNA conjugate was purified by EtOH precipitation and RP-HPLC. ESI-LC-MS analysis: theoretical m/z: 12168; found m/z: 12168.20.

4.2.2.6 *Ligation of FAM-Elib4 to DA-d-spacer: synthesis of DA-FAM-Elib4*



DA-d-spacer (25 µl, 80 µM), FAM-Elib4 (26 µl, 100 µM), E4 chimeric DNA/RNA adapter (13.3 µl, 300 µM) were dissolved in H_2O (111 µl) and 10 x T4 DNA Ligase Reaction buffer (B0202S, *New England Biolabs*) was added (20 µl). A pre-hybridization step was performed for 2 min at 90 °C, and the mixture was allowed to cool down at 25 °C. T4 DNA Ligase (5 µl, 400 U/ µl, *New England Biolabs*) was added, and the reaction was left 14 h at 16 °C, before inactivating the ligase for 10 min at 65 °C. The crude ligation product was checked by ESI-LC-MS (theoretical m/z: 24687 found m/z: 24686.58) and directly used for the next step. (Paragraph 4.2.2.8)

4.2.2.7 *Ligation of unique qPCR primer sequences to d-spacer (DA_Elib4_X)*



The ligation procedure described in Paragraph 4.2.2.6 was used to perform the ligations of different PCR-primers. The encoding strategy is displayed in Figure 5.9.

4.2.2.8 *RNAse adapter degradation*

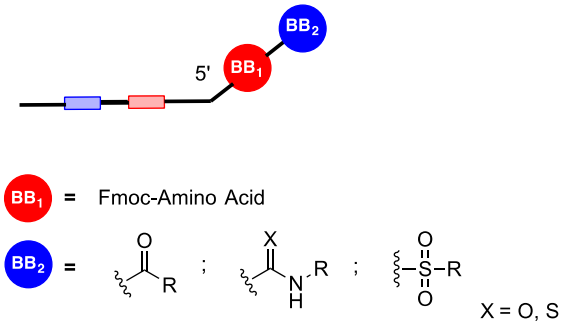
The crude ligation solution (100 µl) was treated with RNAse HII (2 µl) in 10x ThermoPol Reaction Buffer (10 µl, *New England Biolabs*). The degradation of the adapter was confirmed by gel electrophoresis analysis on Novex™ TBE-Urea Gels 15% (*Invitrogen*). The product was purified by nucleotide removal columns (Smartpure DNA Purification Kit, *Eurogenentec*), redissolved in H_2O and the concentration evaluated by Nanodrop measurement.

4.2.2.9 *Klenow hybridization*

FAM-DNA-CAIX Ligand **DNA-CAIX Ligand**

Elib2 conjugates and DA-Elib4 derivatives (with/without FAM) conjugates (reaction scale: 50 pmol) were dissolved in H_2O (final volume: 100 µl) and NEBuffer™ 2 (10 µl, *New England Biolabs*). The code Elib2_5 was used to construct the respective FAM-DNA-NC and DNA-NC derivatives. A pre-hybridization step was performed for 2 min at 90°C, and the mixture was allowed to cool down at 25°C. The DNA Polymerase I, Large (Klenow) Fragment (1 µl, *New England Biolabs)* and deoxynucleotide solution mix (4 µl, 5 mM) were added and incubated for 1 h at 25°C. The reaction was checked by gel electrophoresis analysis on Novex™ TBE-Urea Gels 15% (*Invitrogen*). The product was purified by QIAquick PCR Purification Kit (*Qiagen*) and the concentration evaluated by Nanodrop measurement. The corresponding derivatives (FAM-DNA CAIX ligands and DNA-CAIX ligands) were used as input in model selection experiments. The DNA encoding strategy is displayed in Figure 5.8 and Figure 5.9.

4.2.3 Synthesis of the single pharmacophore DECL

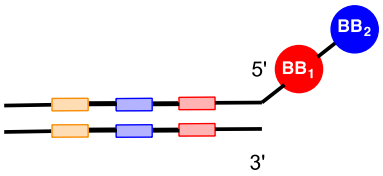


The encoding strategy is displayed in Figure 5.10. The coupling procedures of the first set of building block 1 (BB1) and building block 2 (BB2) were performed as described in Paragraph 4.1.2.

4.2.3.1 *Ligation Procedure for single-stranded DNA library synthesis*

The pool of the code 1-BB1 (1.4 µl, 1 mM H_2O) conjugates, the corresponding code 2 (2.1 µl, 1 mM H_2O) and DNA adapter Code1_Code2 (1.4 µl, 2 mM H_2O), were mixed and incubated at room temperature for 1 hour. In a separate vial, a master mix of the corresponding amount of 10x T4 DNA-ligase buffer (1.7 µl, New England Biolabs), water (0.15 µl) and T4 DNA-ligase (0.2 µl, 400 U/µl, *New England Biolabs*) was prepared and added to the oligo mixture. The ligation process was left for 16 h at 16 °C, before inactivating the ligase for 10 min at 65 °C. The formation of the desired ligated product was confirmed by gel electrophoresis analysis. The mixture was dried over SpeedVac, and the DNA pellet was directly used for the coupling of the second building block (BB2). After the reaction, the pool of the DNA conjugates was purified by ionic exchange chromatography (IE-HPLC).

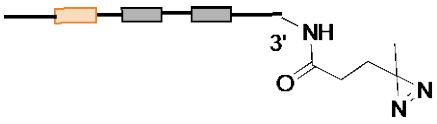
4.2.3.2 *Klenow Polymerization*



The single-stranded single pharmacophore library (31.7 µl, 3.15 µM) and Code 3 (2 µl, 50 µM) were dissolved in H_2O (50.3 µl) and NEB Buffer 10x (10 µl). A pre-hybridization step was performed at 95 °C for 2 min, and the mixture was allowed to cool down at 25 °C. The DNA Polymerase I, Large (Klenow) Fragment (2 µl, *New England Biolabs)* and deoxynucleotide solution mix (4 µl, 5 mM) were added and incubated for 1 h at 25 °C. The reaction was checked by gel electrophoresis analysis on Novex™ TBE-Urea Gels 15% (*Invitrogen*). The library was purified by QIAquick PCR Purification Kit (*Qiagen*), redissolved in 100 µl and the concentration evaluated by Nanodrop measurement.

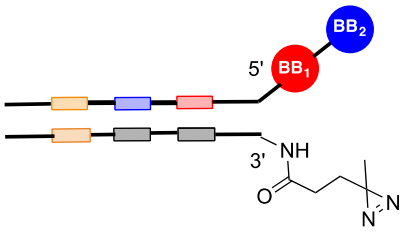
4.2.4 Synthesis of the Photoreactive Library

4.2.4.1 *Ligation procedure: Synthesis of the photoreactive DNA-strand (DA-Elib6)*



DA-d-spacer (25 µl, 80 µM), Elib6_code (5.20 µl, 500 µM), DNA adapter_Elib6 (8 µl, 500 µM) were dissolved in H_2O (49 µl) and 10 x T4 DNA Ligase Reaction buffer (*New England Biolabs*) was added (10 µl). A pre-hybridization step was performed at 90°C for 2 min, before the mixture was allowed to cool down at 25°C. T4 DNA Ligase (2.5 µl, 400 U/ µl, New England Biolabs) was added, and the reaction was left for 14 h at 16 °C, before inactivating the ligase for 10 min at 65 °C. The crude ligation product was checked by LC-MS (theoretical m/z: 31239 found m/z: 31238.42). The product was purified by QIAgen PCR Purification Kit (Qiagen) to remove the excess of code and adapter. The concentration was evaluated by Nanodrop measurement, and the product was directly used for the next step.

4.2.4.2 *Klenow polymerization*



The single-stranded single pharmacophore library, prepared as described in Paragraph 4.2.3.1, (30 µl, 1.31 µM) and the DA-Elib6 (19 µl, 2.11 µM) were dissolved in H_2O (35.5 µl) and NEB Buffer 10x (10 µl). A pre-hybridization step was performed for 2 min at 95 °C, before the mixture was allowed to cool down at 25 °C. The DNA Polymerase I, Large (Klenow) Fragment (1.5 µl, *New England Biolabs)* and deoxynucleotide solution mix (4 µl, 5 mM) were added and incubated for 1 h at 25 °C. The reaction was checked by gel electrophoresis analysis on Novex™ TBE-Urea Gels 15% (*Invitrogen*). The library was

purified by QIAquick PCR Purification Kit (*Qiagen*) and the concentration evaluated by Nanodrop measurement.

4.2.5 Screening Methodologies

4.2.5.1 *Optimized photocrosslinking screening methodology*

The photoreactive DNA-CAIX ligands (0.005 pmol) or the photoreactive library (10^6 copies / lib member) and the 6x-histidine-tagged CAIX (10 µl, 10 µM) were incubated for 1 h in PBS Gibco 1x (final volume: 50 µl, final protein concentration: 2 µM) containing 0.3 mg/ml Herring Sperm (*Thermofisher*) and 0.1 M NaCl. The incubation solution was transferred in a 96-well micro test plate (*Ratiolab*), and UV irradiated over ice for 20 min at 365 nm using a UVP CL-1000 Ultraviolet Crosslinker (*Hoefer*). After the UV reaction, the reaction was quenched with a solution of 10% SDS in H_2O (final concentration: 5% in 100 µl). Dynabeads His-Tag Pull-Down (12 µl / selection, *Invitrogen*) were washed with 3 x PBS containing 0.01% Tween-20 and resuspended in the same buffer (100 µl / selection). By using the KingFisher Magnetic Particle Processor, the magnetic beads were transferred to the target solution and incubated for 1 h with continuous gentle mixing. The beads were washed (5 x 2 min, 200 µl) with washing buffer (PBS Gibco 1x, 0.2% SDS, 0.2% Tween-20) and incubated with elution buffer (100 µl, 10 mM NaH_2PO_4, 0.3 M NaCl and 200 mM Imidazole) for 20 min. After incubation, the beads were separated. Experiments were conducted in duplicate. The eluate was analyzed by qPCR and/or amplified by PCR and submitted to the Functional Genomic Center Zürich for high-throughput DNA sequencing on an Illumina HiSeq 2500 Instrument.

4.2.5.2 *Affinity selection against polyhistidine-tagged CAIX*

Dynabeads His-Tag Pull-Down (12 µl / selection, Invitrogen) were washed with 3 x binding buffer (500 ml; 50 mM sodium phosphate, 300 mM NaCl, pH 8.0, 0.01 %/ Tween-20) and resuspended in binding buffer (100 µl). By using the KingFisher magnetic particle processor, the magnetic beads were transferred to a solution of binding buffer and His-tagged target (CAIX) at a certain concentration (2 µM, 100 µl) and incubated for 30 min with continuous gentle mixing. The beads were washed (3 x 3 min, 200 µl) with binding buffer and subsequently transferred to a solution of the library (10^6 copies / lib. member) in pull-down buffer (100 µl; 3.25 mM sodium phosphate, 70 mM NaCl, pH 7.4, 0.01 % Tween-20), containing herring sperm (0.3 mg/ml, Thermofisher). The beads were incubated for 1 h with continuous gentle mixing. The beads were removed and washed with pull-down buffer (5 x 30 s, 200 µl) and incubated with elution buffer (100 µl, 10 mM

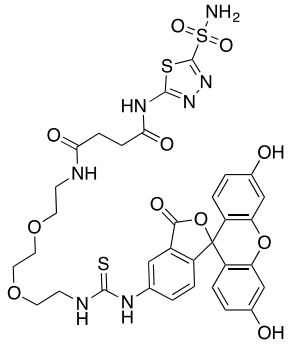
NaH_2PO_4, 300 mM NaCl and 200 mM Imidazole) for 30 min. After incubation, the beads were separated. The eluate was amplified by PCR and submitted to the Functional Genomics Center in Zürich for high-throughput DNA sequencing on an Illumina HiSeq 2500 instrument.

4.2.6 Quantitative PCR: General Procedure

In an Optical 8-Cap Strip (*Applied Biosystems*), in a volume of 25 µl, the following were combined: DNA template (2.5 µl; selection eluate or water for the no-template control), forward primer (0.5 µl, 10 µM in H_2O), reverse primer (0.5 µl, 10 µM in H_2O), SensiFAST™ SYBR Hi-ROX PCR MasterMix (12.5 µl, *Bioline)* and H_2O (9 µl). qPCR program: 2 min at 95°C (hold), 5 s at 95 °C, (denaturation) 10 s at 60°C (annealing), 10 s at 72 °C (extension). The number of denaturation/annealing/extension cycles was set at 40. The corresponding Ct values were correlated with the absolute quantity of DNA through standard curve calibration experiments. The experimental measurements are plotted as absolute quantities of DNA molecules (DNA copies) in 100 µl. The specificity of the reactions was checked by melt curve analysis or gel electrophoresis. Melt curve program: 15 s, 95 °C, 60 °C to 95 °C, 0.3 °C/min.

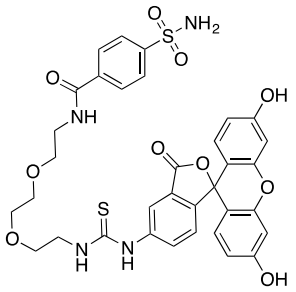
4.2.7 Hit Validation: Synthesis of small-molecule FITC derivatives

4.2.7.1 *Synthesis of AAZ - FITC (compound 1)*



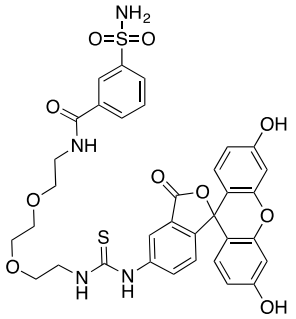
To a solution of 4-Oxo-4-((5-sulfamoyl-1,3,4-thiadiazol-2-yl)amino)butanoic acid (0.1 mmol, 1 eq, 28 mg) in DCM (1 ml) trimethylamine (0.3 mmol, 3 eq, 80 µl) was added. Once dissolved, FITC-$(PEG)_2$ amine (0.1 mmol, 1 eq, 57 mg) and HATU (0.11 mmol, 1.1 eq, 42 mg) were added. The reaction was stirred overnight at 25 °C. The crude was evaporated *in vacuo* and purified by inverse phase HPLC (r.t.: 4.65 min, H_2O/MeCN, from 30% to 60%, 5.0 ml/min, 5 min run), obtaining a yellow powder (23 mg, 28%). ^{1}H NMR (400MHz, DMSO-d_6) δ = 10.03 (br. s., 2 H), 8.39 - 8.24 (m, 3 H), 8.09 (br. s., 1 H), 8.00 (dd, J_1 = 4.6 Hz, J_2 = 4.6 Hz, 1 H), 7.74 (d, J = 8.1 Hz, 1 H), 7.23 - 7.15 (m, 1 H), 6.71 - 6.66 (m, 2 H), 6.64 - 6.53 (m, 4 H), 3.64 - 3.54 (m, 8 H), 3.42 (t, J = 6.0 Hz, 2 H), 3.20 (m, 2 H), 2.292 (m, 2H), 2.78 - 2.70 (m, 2 H). ^{13}C NMR (100MHz, DMSO-d_6) δ = 180.78, 172.02, 171.00, 168.73, 167.61, 164.43, 161.32, 159.72, 152.11, 143.75, 141.55, 129.26, 126.74, 124.30, 112.81, 109.94, 102.45, 87.13, 80.14, 69.76, 69.31, 68.66, 45.98, 43.89, 30.48, 29.59. LCMS (ES^+) m/z 800.0 $(M+H)^+$. TOF-MS (ES^+): $C_{33}H_{34}N_7O_{11}S_3$ $[M+H]^+$ calculated: 800.1473, found: 800.1452.

4.2.7.2 *Synthesis of SABA - FITC (compound 2)*



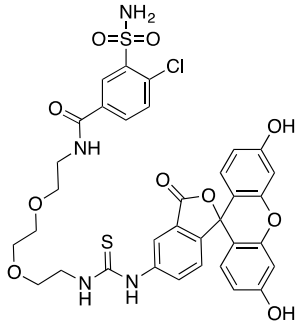
To a solution of 4-sulfamoylbenzoic acid (0.1 mmol, 1 eq, 20 mg) in DCM (1 ml), trimethylamine (0.3 mmol, 3 eq, 80 µl) was added. Once dissolved, FITC-$(PEG)_2$ amine (0.1 mmol, 1 eq, 57 mg) and HATU (0.11 mmol, 1.1 eq, 42 mg) were added. The reaction was stirred overnight at 25 °C. The crude was evaporated *in vacuo* and purified by inverse phase HPLC (r.t.: 4.67 min, H_2O/MeCN, from 30% to 62%, 5.0 ml/min, 6 min run), obtaining a yellow powder (33 mg, 45%). ^{1}H NMR (400MHz, DMSO-d6) δ = 10.12 (br. s., 3 H), 8.73 (t, *J* = 5.6 Hz, 1 H), 8.28 (s, 1 H), 8.17 - 8.05 (m, 1 H), 8.03 - 7.96 (m, 2 H), 7.93 - 7.84 (m, 2 H), 7.73 (m, 1 H), 7.48 (s, 2 H), 7.19 (d, *J* = 8.4 Hz, 1 H), 6.68 (d, *J* = 2.5 Hz, 2 H), 6.64 - 6.51 (m, 4 H), 3.69 (br. s., 2 H), 3.64 - 3.53 (m, 8 H), 3.45 (m, 2 H). ^{13}C NMR (100MHz, DMSO-d_6) δ = 180.77, 168.73, 165.54, 159.67, 152.08, 147.93, 147.42, 146.43, 141.53, 137.49, 132.31, 129.25, 128.06, 126.74, 125.83, 124.26, 116.59, 112.77, 109.91, 102.45, 83.20, 72.00, 69.79, 69.01, 68.65, 43.89. LCMS (ES^+) m/z 721.1 $(M+H)^+$. TOF-MS (ES^+): $C_{34}H_{33}N_4O_{10}S_2$ $[M+H]^+$ calculated: 721.1633, found: 721.1624.

4.2.7.3 *Synthesis m-SABA - FITC (compound 3)*



To a solution of 3-Sulfamoylbenzoic acid (0.1 mmol, 1 eq, 20 mg) in DCM (1 ml), trimethylamine (0.3 mmol, 3 eq, 80 µl) was added. Once dissolved, FITC-$(PEG)_2$ amine (0.1 mmol, 1 eq, 57 mg) and HATU (0.11 mmol, 1.1 eq, 42 mg) were added. The reaction was stirred overnight at 25 °C. The crude was evaporated *in vacuo* and purified by inverse phase HPLC (r.t.: 4.67 min, H_2O/MeCN, from 30% to 62%, 5.0 ml/min, 6 min run), obtaining a yellow powder (30 mg, 41%). ^{1}H NMR (400MHz, DMSO-d_6) δ = 10.32 - 9.95 (m, 3 H), 8.79 (t, *J* = 5.6 Hz, 1 H), 8.32 (m, 1 H), 8.28 (s, 1 H), 8.13 (br. s., 1 H), 8.09 - 8.00 (m, 1 H), 8.00 - 7.92 (m, 1 H), 7.82 - 7.71 (m, 1 H), 7.67 (dd, $J_1 = J_2$ =7.7 Hz, 1 H), 7.44 (br. s., 2 H), 7.21 - 7.14 (m, 1 H), 6.73 - 6.65 (m, 2 H), 6.64 - 6.50 (m, 4 H), 3.68 (br. s., 2 H), 3.64 - 3.54 (m, 8 H), 3.45 (m, 2 H). ^{13}C NMR (100MHz, DMSO-d_6) δ = 180.56, 168.52, 165.15, 159.54, 151.90, 147.58, 147.02, 144.38, 141.33, 135.05, 132.04, 130.13, 129.13, 128.94, 128.11, 124.73, 116.38, 112.63, 109.74, 102.23, 82.26, 69.63, 68.80, 68.43, 43.66. LCMS (ES^+) m/z 721.1 $(M+H)^+$. TOF-MS (ES^+): $C_{34}H_{33}N_4O_{10}S_2$ $[M+H]^+$ calculated: 721.1633, found: 721.1630.

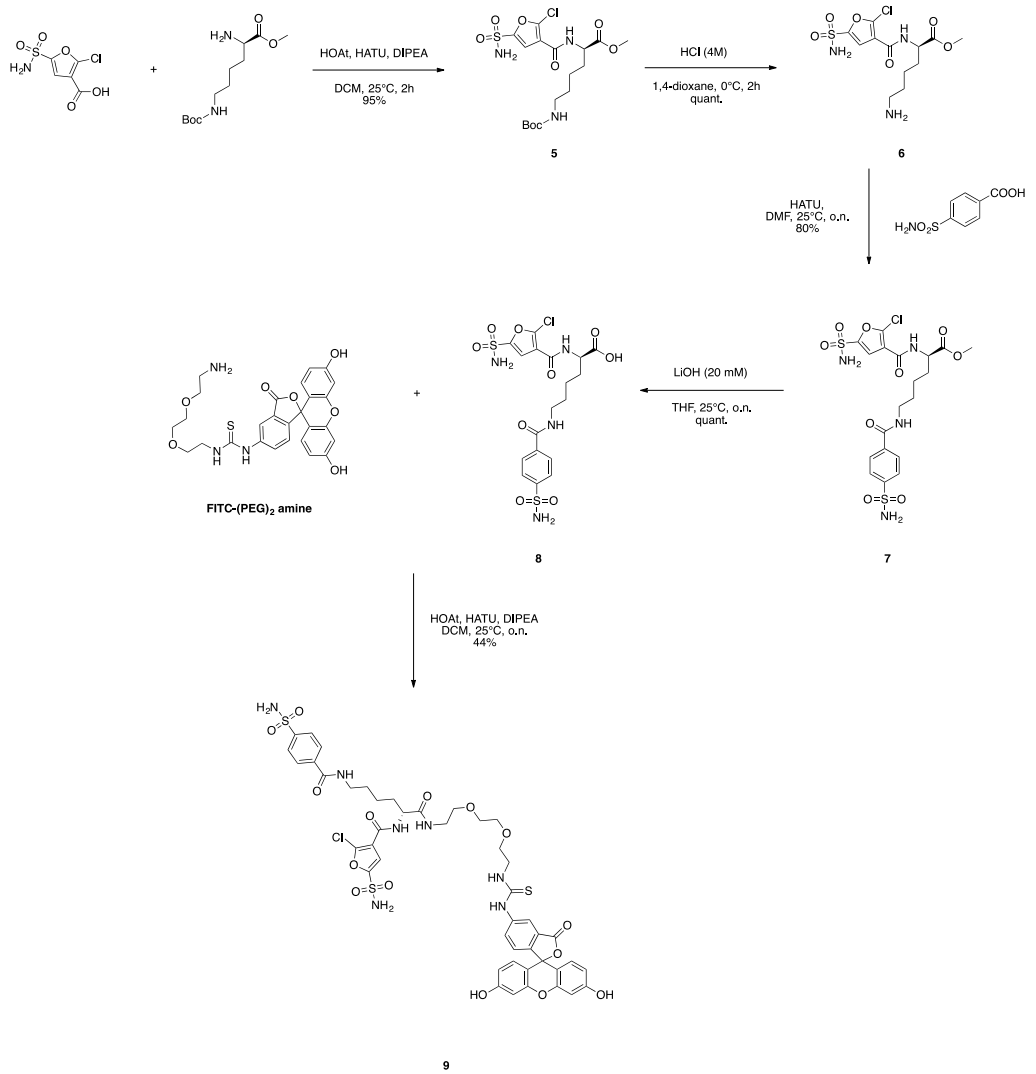
4.2.7.4 *Synthesis Cl-SABA - FITC (compound 4)*



To a solution of 4-chloro-3-sulfamoylbenzoic acid (0.1 mmol, 1 eq, 23 mg) in DCM (1 ml), trimethylamine (0.3 mmol, 3 eq, 80 µl) was added. Once dissolved, FITC-$(PEG)_2$ amine (0.1 mmol, 1 eq, 57 mg) and HATU (0.11 mmol, 1.1 eq, 42 mg) were added. The reaction was stirred overnight at 25 °C. The crude was evaporated *in vacuo* and purified by inverse phase HPLC (r.t.: 4.75 min H_2O/MeCN, from 30% to 62%, 5.0 ml/min, 6 min run), obtaining a yellow powder (35 mg, 46%). ^{1}H NMR (400MHz, DMSO-d_6) δ = 10.02 (br. s., 2 H), 8.85 (t, *J* = 5.4 Hz, 1 H), 8.46 (m, 1 H), 8.28 (s, 1 H), 8.15 - 7.99 (m, 2 H), 7.80 - 7.63 (m, 4 H), 7.19 (d, *J* = 8.4 Hz, 1 H), 6.72 - 6.53 (m, 6 H), 3.72 - 3.65 (m, 2 H), 3.64 - 3.53 (m, 8 H), 3.44 (m, 2 H). ^{13}C NMR (100MHz, DMSO-d_6) δ = 180.56, 176.08, 172.72, 168.51,

164.36, 159.51, 151.89, 141.33, 141.10, 133.32, 133.13, 131.62, 131.44, 129.42, 129.04, 128.13, 126.53, 125.48, 124.08, 116.40, 112.60, 109.72, 102.23, 69.62, 69.57, 68.75, 43.67. LCMS (ES^+) m/z 755.0 $(M+H)^+$. TOF-MS (ES^+): $C_{34}H_{32}N_8ClN_4O_{10}S_2[M+H]^+$ calculated: 755.1243, found: 755.1237.

4.2.7.5 *Synthesis of 146/586 - FITC*



Scheme 4.1. Synthetic route of compound 146/586 FITC (compound 9).

Synthesis of compound **5**

To a solution of 2-chloro-5-sulfamoylfuran-3-carboxylic acid (0.44 mmol, 1 eq, 100 mg) in DCM (1 ml), DIPEA (1.3 mmol, 3 eq, 250 µl) was added. (*R*)-2-Amino-6-tert-butoxycarbonylamino-hexanoic acid methyl ester (0.44 mmol, 1 eq, 114 mg), HOAt (0.22 mmol, 0.5 eq, 30 mg) and HATU (0.53 mmol, 1.2 eq, 200 mg) were then subsequently added. The reaction was stirred for 2 h at 25 °C. The crude was evaporated *in vacuo*, diluted in DCM and washed with NH_4Cl twice. The organic phases were dried and purified by silica gel chromatography (DCM:MeOH 100:0 to DCM:MeOH 93:7) yielding compound **5** as a white powder (194 mg, 95%). ^{1}H NMR (400MHz, CD_3CN) δ = 7.36 (s, 1 H), 7.07 (br. d, *J* = 6.8 Hz, 1 H), 6.10 (br. s., 1 H), 5.29 (br. s., 1 H), 4.52 (td, *J* = 5.1, 7.9 Hz, 1 H), 3.72 (s, 3 H), 3.03 (m, 2 H), 1.93 - 1.71 (m, 2 H), 1.55 - 1.33 (m, 13 H). ^{13}C NMR (100MHz, DMSO-d_6) δ = 171.20, 168.49, 164.65, 164.07, 159.05, 145.67, 144.86, 143.35, 77.38, 52.90, 45.28, 33.75, 28.31, 25.40, 22.89. LCMS (ES^+) m/z 490.1 $(M+Na)^+$. TOF-MS (ES^+): $C_{17}H_{26}ClN_3NaO_8S$ $[M+Na]^+$ calculated: 490.1221, found: 490.1008.

Synthesis of compound **6**

HCl 4 M in 1,4-dioxane (1.0 eq.) previously cooled to 0 °C was added to a 1 M solution of compound **5** in DCM. The reaction was stirred and slowly warmed to room temperature for 2 h. The mixture was evaporated *in vacuo*, Et_2O was added, and the mixture was evaporated several times until a yellowish powder was obtained (compound **6**; 290 mg, quantitative). The crude was used for the following step without further purification.

Synthesis of compound **7**

To a solution of 4-sulfamoylbenzoic acid (0.2 mmol, 1 eq, 45 mg) in DCM (1 ml), DIPEA (0.6 mmol, 3 eq, 104 µl) was added. Compound **6** in 1 mL of DMF (0.2 mmol, 1 eq, 90 mg), and HATU (0.24 mmol, 1.2 eq, 91 mg) were then added. The reaction was stirred overnight at 25 °C. The crude was evaporated *in vacuo*, diluted in DCM and washed with NH_4Cl twice. The organic phases were dried and purified by silica gel chromatography (DCM:MeOH 100:0 to DCM:MeOH 95:5) obtaining compound **7** as a reddish powder (88 mg, 80%). ^{1}H NMR (400MHz, CD_3CN) δ = 8.03 - 7.86 (m, 4 H), 7.41 (s, 1 H), 7.26 (br. t, *J* = 6.1 Hz, 1 H), 7.20 (br. d, *J* = 7.1 Hz, 1 H), 5.96 (br. s., 4 H), 4.52 (td, *J* = 4.8, 10.0 Hz, 1 H), 3.71 (s, 3 H), 3.39 (s, 2 H), 1.86 (s, 2 H), 1.64 (s, 2 H), 1.54 - 1.42 (m, 2 H). ^{13}C NMR (100MHz, DMSO-d_6) δ = 171.62, 169.91, 165.07, 164.49, 159.47, 152.68, 151.86, 146.09, 145.28, 139.24, 129.04, 63.67, 53.32, 45.70, 34.17, 28.73, 23.21. LCMS (ES^+) m/z 551.0 $(M+H)^+$, TOF-MS (ES^+): $C_{19}H_{24}ClN_4O_9S_2$ $[M+H]^+$ calculated: 551.0673, found: 551.0664.

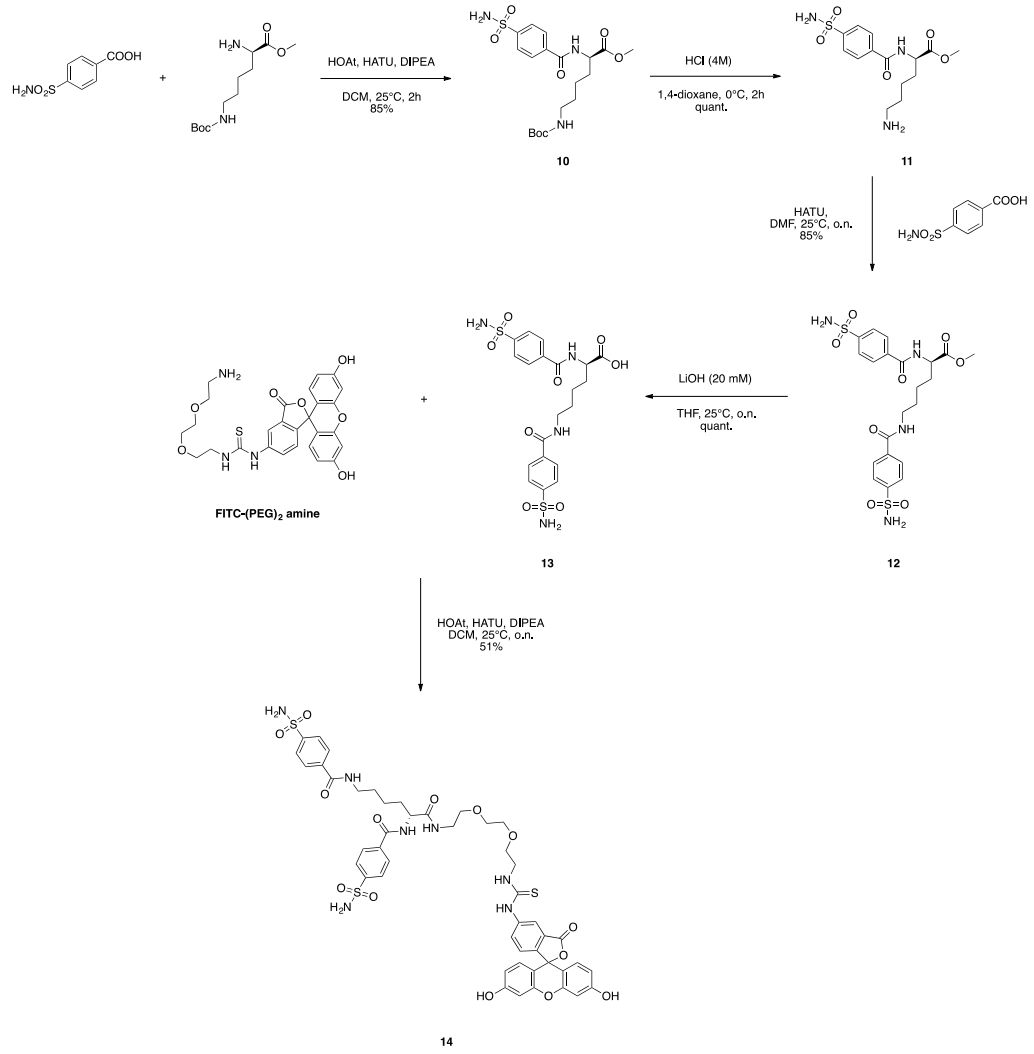
Synthesis of 146/586 (compound **8**)

To a solution of compound **7** (0.054 mmol, 1 eq, 30 mg) in THF (20 mM), a solution of 20 mM LiOH (aq) (0.054 mmol, 1 eq, 2.75 ml) was added. The reaction was stirred overnight at 25 °C. The crude was then neutralized with 1 M HCl(aq) (c.a. 50 µl), evaporated *in vacuo*, ACN was added, and the mixture was evaporated several times until a white powder is obtained. The product (compound **8**) was used for the following step without further purification. LCMS (ES^+) m/z 537.0 $(M+H)^+$.

Synthesis of 146/586 - FITC (compound **9**)

To a solution of 146/586 (compound **8**; 0.019 mmol, 1 eq, 10 mg) in DMF (1 mL), DIPEA (0.058 mmol, 3 eq, 12 µl), FITC-$(PEG)_2$ amine (0.02 mmol, 1.1 eq, 10 mg), HOAt (0.01 mmol, 0.5 eq, 1.5 mg) and HATU (0.02 mmol, 1.1 eq, 7.6 mg) were added. The reaction was stirred overnight at 25 °C. The crude was evaporated *in vacuo* and purified by inverse phase HPLC (r.t.: 4.74 min, H_2O/MeCN, from 30% to 68%, 5.0 ml/min, 7 min run), obtaining compound **9** as a yellow powder (9.0 mg, 44%). ^1H NMR (400MHz ,DMSO-d_6) δ = 10.15 (br. s., 2 H), 8.64 (s, 1 H), 8.40 - 8.27 (m, 2 H), 8.22 - 8.11 (m, 1 H), 8.05 (s, 1 H), 8.01 - 7.94 (m, 1 H), 7.94 - 7.85 (m, 2 H), 7.75 (m, 2 H), 7.59 (s, 1 H), 7.46 (s, 1 H), 7.18 (d, *J* = 8.4 Hz, 1 H), 6.72 - 6.49 (m, 6 H), 4.36 - 4.24 (m, 1 H), 3.75 - 3.64 (m, 2 H), 3.63 - 3.51 (m, 8 H), 3.45 - 3.37 (m, 2 H), 3.29 - 3.17 (m, 3 H), 3.10 (dd, *J* = 4.6, 7.1 Hz, 2 H), 1.86 - 1.26 (m, 6 H). ^{13}C NMR (100MHz, DMSO-d_6) δ = 179.16, 172.10, 169.39, 165.55, 165.22, 164.97, 159.95, 153.16, 152.34, 146.57, 145.76, 144.25, 139.72, 138.02, 135.70, 134.89, 132.43, 129.51, 128.25, 126.05, 125.09, 120.50, 112.14, 110.17, 107.26, 103.03, 102.70, 81.20, 70.09, 69.42, 53.79, 46.17, 44.59, 34.65, 29.21, 23.68. LCMS (ES^+) m/z 1056.0 $(M+H)^+$. TOF-MS (ES^+): $C_{45}H_{47}ClN_7O_{15}S_3$ $[M+H]^+$ calculated: 1056.1981, found: 1056.1987.

4.2.7.6 *Synthesis of 146-49 – FITC (compound 14)*



Scheme 4.2. Synthetic route of compound 146/49 FITC (compound 14).

Synthesis of compound **10**

To a solution of 4-sulfamoylbenzoic acid (1 mmol, 1 eq, 201 mg) in DCM (1 ml), DIPEA was added (3 mmol, 3 eq, 500 µl). (*R*)-2-Amino-6-tert-butoxycarbonylamino-hexanoic acid methyl ester (1 mmol, 1 eq, 260 mg), HOAt (0.5 mmol, 0.5 eq, 68 mg) and HATU (1.2 mmol, 1.2 eq, 456 mg) were added. The reaction was stirred for 2 h at 25 °C. The crude was evaporated *in vacuo*, diluted in DCM and washed with NH_4Cl twice. The organic phases were dried and purified by silica gel chromatography (DCM:MeOH 100:0 to DCM:MeOH 93:7) obtaining compound **10** as a white powder (376 mg, 85%). ^{1}H NMR (400MHz, DMSO-d_6) δ = 8.91 (br. d, *J* = 7.3 Hz, 1 H), 8.07 - 7.99 (m, 2 H), 7.95 - 7.88 (m, 2 H), 7.63 - 7.35 (m, 2 H), 6.79 (br. t, *J* = 5.6 Hz, 1 H), 4.51 - 4.32 (m, 1 H), 3.66 (s, 3 H), 3.00 - 2.83 (m, 2 H), 1.92 - 1.68 (m, 2 H), 1.36 (s, 13 H). ^{13}C NMR (100MHz, DMSO-d_6) δ = 172.22, 165.66, 165.48, 147.58, 146.50, 137.97, 137.69, 77.97, 53.98, 45.13, 31.86, 29.19, 25.50, 23.70. LCMS (ES$^+$) m/z 465.1 $(M+Na)^+$. TOF-MS (ES$^+$): $C_{19}H_{29}N_3O_7SNa$ $[M+Na]^+$ calculated: 466.1618, found: 466.1611.

Synthesis of compound **11**

4 M HCl in 1,4-dioxane (1.0 eq.) previously cooled to 0 °C, was added to a 1 M solution of compound **10** in DCM. The reaction was stirred and slowly warmed to room temperature for 2 h. The mixture was evaporated *in vacuo*, Et_2O was added, and the mixture was evaporated several times until a yellowish powder was obtained (compound **11**; 290 mg, quantitative). The crude was used without further purifications for the next step.

Synthesis of compound **12**

To a solution of 4-sulfamoylbenzoic acid (0.94 mmol, 1.1 eq, 189 mg) in DCM (1 ml), DIPEA (4.3 mmol, 5 eq, 500 µl) was added. Once dissolved, compound **11** in 1 mL of DMF (0.85 mmol, 1 eq, 296 mg), and HATU (1 mmol, 1.2 eq, 390 mg) were added. The reaction was left stirring for 2 hours. The crude was evaporated *in vacuo*, diluted in DCM and washed with NH_4Cl_2 two times. The organic phases were dried and purified by silica gel chromatography (DCM: MeOH 100:0 to DCM: MeOH 85:15) obtaining compound **12** as a white powder (376 mg, 85%). ^{1}H NMR (400MHz, DMSO-d_6) δ = 8.93 (d, *J* = 7.4 Hz, 1 H), 8.65 (t, *J* = 5.6 Hz, 1 H), 8.06 - 8.00 (m, 2 H), 8.00 - 7.95 (m, 2 H), 7.95 - 7.86 (m, 4 H), 7.47 (br. s., 4 H), 4.54 - 4.39 (m, 1 H), 3.65 (s, 3 H), 3.30 - 3.18 (m, 2 H), 1.94 - 1.77 (m, 2 H), 1.65 - 1.51 (m, 2 H), 1.51 - 1.35 (m, 2 H). ^{13}C NMR (100MHz, DMSO-d_6) δ = 171.80, 165.24,165.06, 159.46, 158.86, 163.16, 152.60, 147.08, 144.35, 140.55, 135.27, 133.01, 129.03, 64.16, 53.56, 45.71, 31.44, 28.77, 23.29. LCMS (ES$^+$) m/z 527.8 $(M+H)^+$. TOF-MS (ES$^+$): $C_{21}H_{27}N_4O_8S_2$ $[M+H]^+$ calculated: 527.1265, found: 527.1268.

Synthesis of **146/49** (compound **13**)

To a solution of compound **12** (0.075 mmol, 1 eq, 40 mg) in THF (0.1 M), a solution of 2 M LiOH aq. (0.23 mmol, 3 eq, 115 µl) was added. The reaction was stirred overnight at 25 °C. The crude was then neutralized with 1 M HCl aq. (about 250 µl) and evaporated *in vacuo*. ACN was added, and the mixture was evaporated several times until a white powder was obtained. The product **146/49** (compound **13**) was used for the following step without further purification. LCMS (ES^+) m/z 513.0 $(M+H)^+$

Synthesis of 146/49 FITC (compound **14**)

To a solution of 146/49 (compound **13**; 0.019 mmol, 1 eq, 10 mg) in DMF (1 ml) was added DIPEA (0.058 mmol, 3 eq, 12 µl), FITC-$(PEG)_2$ amine (0.02 mmol, 1.1 eq, 10 mg), HOAt (0.01 mmol, 0.5 eq, 1.5 mg) and HATU (0.02 mmol, 1.1 eq, 7.6 mg) were added. The reaction was stirred overnight at 25 °C. The crude was evaporated *in vacuo* and purified by inverse phase HPLC (r.t.: 4.25 min, H_2O/MeCN, from 30% to 68%, 5.0 ml/min, 7 min run), obtaining compound **14** as yellow powder (10 mg, 51%). ^{1}H NMR (400MHz, DMSO-d_6) δ = 10.29 - 10.01 (m, 3 H), 8.78 - 8.58 (m, 2 H), 8.30 (br. s., 1 H), 8.15 (br. s., 1 H), 8.11 - 7.93 (m, 4 H), 7.93 - 7.85 (m, 3 H), 7.80 - 7.71 (m, 1 H), 7.54 - 7.42 (m, 3 H), 7.19 (d, *J* = 8.4 Hz, 1 H), 6.68 (d, *J* = 2.3 Hz, 2 H), 6.64 - 6.53 (m, 4 H), 4.44 (br. s., 1 H), 3.69 (br. s., 2 H), 3.59 (br. s., 8 H), 3.44 (br. s., 2 H, overestimated due to H_2O), 3.26 (br. s., 3 H), 3.11 (s, 2 H), 2.83 (s, 2 H), 1.76 (br. s., 2 H), 1.64 - 1.50 (m, 2 H), 1.50 - 1.28 (m, 3 H). ^{13}C NMR (100MHz, DMSO-d_6) δ = 181.05, 172.28, 169.00, 165.85, 165.72, 165.54, 159.94, 152.34, 147.64, 146.78, 146.56, 141.83, 138.03, 137.75, 137.49, 129.51, 128.69, 128.32, 128.25, 126.04, 126.94, 124.51, 116.80, 113.04, 110.18, 102.70, 101.13, 100.99, 100.67, 83.45, 70.09, 70.04, 69.45, 69.26, 54.04, 46.19, 44.13, 31.923, 29.25, 23.76. LCMS (ES^+) m/z 1033.0 $(M+H)^+$. TOF-MS (ES^+): $C_{47}H_{50}N_7O_{14}S_3$ $[M+H]^+$ calculated: 1032.2578, found: 1032.2563.

4.2.8 Fluorescence Polarization procedure

FITC small molecule derivatives were dissolved in DMSO obtaining 10 μM stock solutions. Stocks were serially diluted with PBS Gibco 1X pH = 7.4 at a concentration of 20 nM. A serial dilution of the protein in PBS Gibco 1X (starting from 10^{-4} to 10^{-11} M) was prepared in a 384-wells plate. The FITC-labeled small molecule solution was added in each well at a final concentration of 10 nM. After incubation for 30 min, the fluorescence was measured with a 384-well microplate reader. The raw data were exported as anisotropy values and plotted against protein concentration. The data were fitted on Kaleida Graph using the following equation (**1**):

$$A = \alpha \times L_0 + (\beta - \alpha) \times \frac{1}{2} \times \left[(P_0 + L_0 + K_d) - \sqrt{(P_0 + L_0 + K_d)^2 + 4P_0L_0}\right]$$

A = measured anisotropy, [P] = protein concentration, L_0 = initial ligand concentration, α, β = coefficients.

4.3 CHAPTER 2.3: GENERAL PROCEDURES

4.3.1 List of the oligonucleotides and PCR primers

Universal code

5'-GGAGCTTCTGAATTCTGTGTGCTGGCCTCGCGAGTCCCATGGCGC-3'

Modification: 5'–C6-AminoLink

Code 1

5'-CGGATCGACGGTCTCACGCGTCAGGCAGC-3'

Modification: 5' - $(PO_4)^{3-}$

DNA Adapter

5'-CGATCCGGCGCCAT-3'

MW = 4224.78 Da

PCR primers

LB_FP

5'-GGAGCTTCTGAATTCTGTGTGCTG-3'

Code1_RV

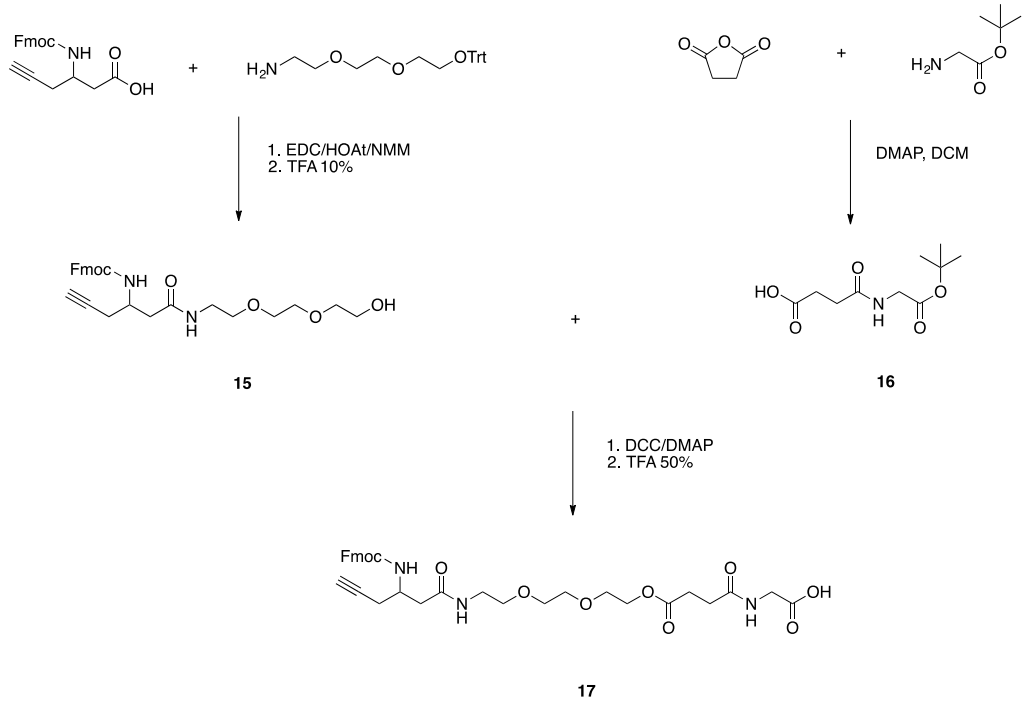
5'-GCGCCATGGGACTCGC-3'

Code2_RV

5'-GCTGCCTGACGCGTGAG-3'

4.3.2 Synthetic procedures

4.3.2.1 *Synthesis of the Fmoc-alkyne trifunctional scaffold (compound 17)*



Synthesis of compound **15**

Fmoc-3-Amino-5-hexynoic acid (334 mg, 1 mmol, 1 eq.) was diluted in DCM (5 mL) then 2-[2-(2-trityloxyethoxy)ethoxy]ethanamine (391 mg, 1 mmol, 1 eq.), 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide HCl (EDC, 170 mg, 1.1 mmol, 1.1 eq.), 1-Hydroxy-7-azabenzotriazole (HOAt, 68 mg, 0.5 mmol, 0.5 eq.) and N-Methylmorpholine (330 µl, 3 mmol, 3eq.) were added. The reaction was left stirring overnight at room temperature. The solvent was removed at reduced pressure, diluted with DCM and washed twice with water (2 x 10 mL). The collected organic phases were dried on Na_2SO_4, deposited on silica and quickly purified by flash column chromatography (DCM/MeOH 100:0 to 97:3) yielding Trt-protected **1** 635 mg (88%). The trityl protected compound **1** (600 mg, 0.83 mmol) was diluted in DCM (7.2 mL) and treated with 800 µl TFA. The reaction was followed by TLC

and after three hours the solvent was removed, and the excess of TFA stripped with 3 consecutive aliquots of toluene. The crude was purified by flash column chromatography (DCM/MeOH 100:0 to 90:10) yielding 302 mg compound **15** (yield: 76%).

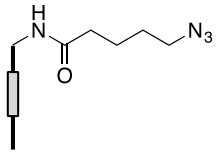
Synthesis of compound **16**

Succinic anhydride (200 mg, 2 mmol, 1eq.) and DMAP (49 mg, 0.4 mmol, 0.2 eq.) were added to a solution of tert-butyl 2-aminoacetate HCl (334 mg, 2 mmol, 2eq.) in DCM (20 mL). The reaction was stirred overnight, then the solvent removed at reduced pressure and the crude diluted in water. The aqueous solution was acidified with HCl 1M (pH=4) and extracted with EtOAc (5 x 20 mL). The organic phases were dried on Na_2SO_4, and the solvent was evaporated *in vacuo*. The obtained compound **16** (240 mg, 51%) was used without further purification for the subsequent reaction.

Synthesis of compound **17**

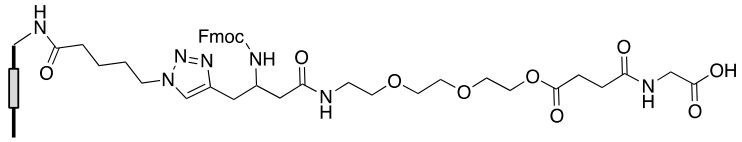
DCC (57 mg, 0.275 mmol, 1.1 eq.), DMAP (33 mg, 0.275 mmol, 1.1 eq.) and compound **15** (116 mg, 0.25 mmol, 1 eq.) were added to a solution of compound **16** (58 mg, 0.25 mmol, 1 eq.) in DCM (20 mL). After 3 hours, a second addition of 1 equivalent of DCC and compound **15** was performed. The reaction was left stirring overnight. After filtration, the solvent was removed, and the crude diluted in water and was extracted with DCM. The organic phases were dried on Na_2SO_4 and purified by flash column chromatography (DCM/MeOH 100:0 to 95:5) to obtain the product with traces of **1**. The crude was then directly deprotected by treatment with TFA 50% in DCM (5 mL) for 2 hours. The solvent was removed *in vacuo,* and the product was purified by chromatography (DCM/MeOH 100:0 to 80:20) to obtain 40 mg compound **17** (25% over two steps).

4.3.2.2 *Synthesis of N_3-DNA conjugate*



A solution of 5-azidopentanoic acid in DMSO (12.5 µl, 200 mM) was diluted with 32.5 µl of DMSO and 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride (DMT-MM; 8.5 µl, 300 mM in H_2O) was added. The solution was stirred for 15 min at 25 °C. The oligonucleotide (universal code; 15 µl, 0.66 mM in H_2O) was dissolved in MOPS buffer (35 µl, 100 mM, 1 M NaCl, pH = 8.0) and H_2O (22 µl), and the solution was added to the mixture. The reaction was stirred for 16 h at 25 °C. The DNA conjugate was purified by EtOH precipitation, followed by RP-HPLC.

4.3.2.3 *Synthesis of the DNA-functionalized scaffold*



The DNA conjugated azido-valerate (8 µl, 500 µM H_2O) was diluted in MOPS (8 µl, 100 mM,1 M NaCl, pH 8). Compound **3** (4 µl, 40 mM DMSO), TBTA (4 µl, 60 mM DMSO), CuSO4•5 H_2O (4 µl, 50 mM H2O) and (+)-Sodium L-ascorbate (4 µl, 70 mM H_2O) were subsequently added, and the reaction was let to proceed for 2 h at 37 °C. The DNA was purified by EtOH precipitation, followed by RP-HPLC.

4.3.2.4 *Coupling of the DNA-functionalized scaffold on TentaGel resin beads*

TentaGel® M NH_2 resin beads (5 mg, 0.23 mmol/g, 10 μm, 11 μmol, *Rapp Polymere*) were added to an Eppendorf and swelled in 0.5 mL DMF at room temperature for 1 h. The resin was centrifuged, and the solvent was removed. Then, the resin was deprotected with 0.5 mL of a 20% piperidine in DMF solution and stirred at room temperature for 2 x 30 min. The beads were centrifuged and washed 5 times with DMF. The resin was diluted in 0.434 mL of DMF, and a premixed solution of EDC/HOAt/NMM 100/20/100 mM (60 μl) was added to the resin. The DNA conjugate **4** (5 nmol, 1% loading) was diluted in 140 μl of MOPS buffer and 24 μl of H_2O and subsequently added to the resin. The resulting mixture was shaken at 37 °C for 16 hours. The resin was then centrifuged and washed with DMF 3 times. An aliquot of the beads (5%) was treated with 100 μl NaOH 0.2 M overnight at 37 °C to cleave the DNA for LC-MS and qPCR analysis (80% yield).

The remaining free amino functionalities on the beads were capped by incubating for 4 hours at 37 °C the resin with a pre-activated solution of benzoic acid (48 μl, 200 mM solution in DMSO) with EDC/HOAt/NMM 100/20/100 mM (48 μl) in 450 μl of DMF. The solid was then centrifuged and washed with 3 x DMF.

4.3.2.5 *Splint ligation procedure*

The resulting functionalized resin (4000 pmol) was equilibrated in 500 μl BTPWB (50 mM NaCl, 0.04% Tween-20, 10 mM bis-tris propane, pH 7.6) in an Eppendorf tube for 1 hour at r.t. The resin was centrifuged and washed with 3 x 500 μl BTPWB and 2 x 500 μl BTPLB (50 mM NaCl, 10 mM $MgCl_2$, 1 mM ATP, 0.02% Tween-20, 10 mM bis-tris propane, pH 7.6). The resin was aliquoted in Eppendorf tube (150 pmol each, 190 μg), Code 2 (4.5 μl, 50 μM, 1.5 eq.), DNA adapter (6 μl, 50 μM, 2 eq), BTPLB (9 μl) and water (67 μl) were added. The resin was heated at 95 °C for 1 minute and let cool down for 30 minutes. Then another aliquot of BTPLB (3 μl) and water (26 μl) was added, and T4 DNA ligase (400 U/μl, 0.75 μl) was mixed. The solution was stirred at 30 °C overnight. The supernatant was removed, and the resin was washed with BTPWB (5 x 100 μl). Then, the denaturing buffer (100 μl; 25 mM bis-tris propane, 6 M urea, 0.02% Tween-20, pH 8) was added to the solid and incubated for 4 h at 37°C. The resin was washed with 2 x 100 μl denaturing buffer and 5 x 100 μl BTPWB. The resin was cleaved with 100 μl NaOH 0.2 M and incubated at 37°C overnight for LC-MS and qPCR analysis.

5. SUPPLEMENTARY FIGURES AND TABLES

5.1 SUPPLEMENTARY FIGURES

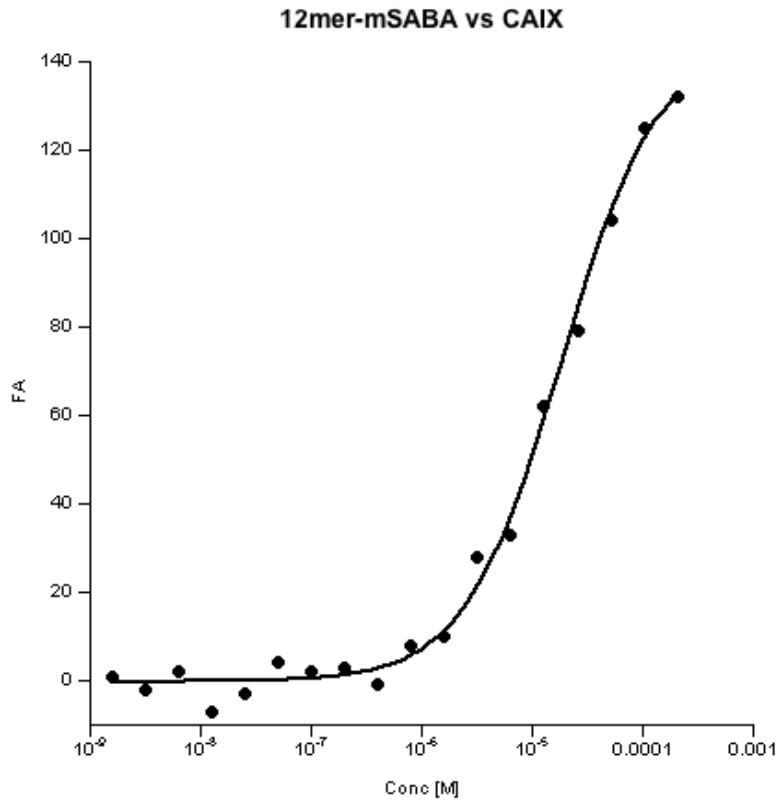


Figure 5.1. Fluorescence polarization assay of *m*-SABA. The assay was performed following the procedure previously described by Bigatti *et al.* [112] A dissociation constant of 19 μM ± 2 μM was observed.

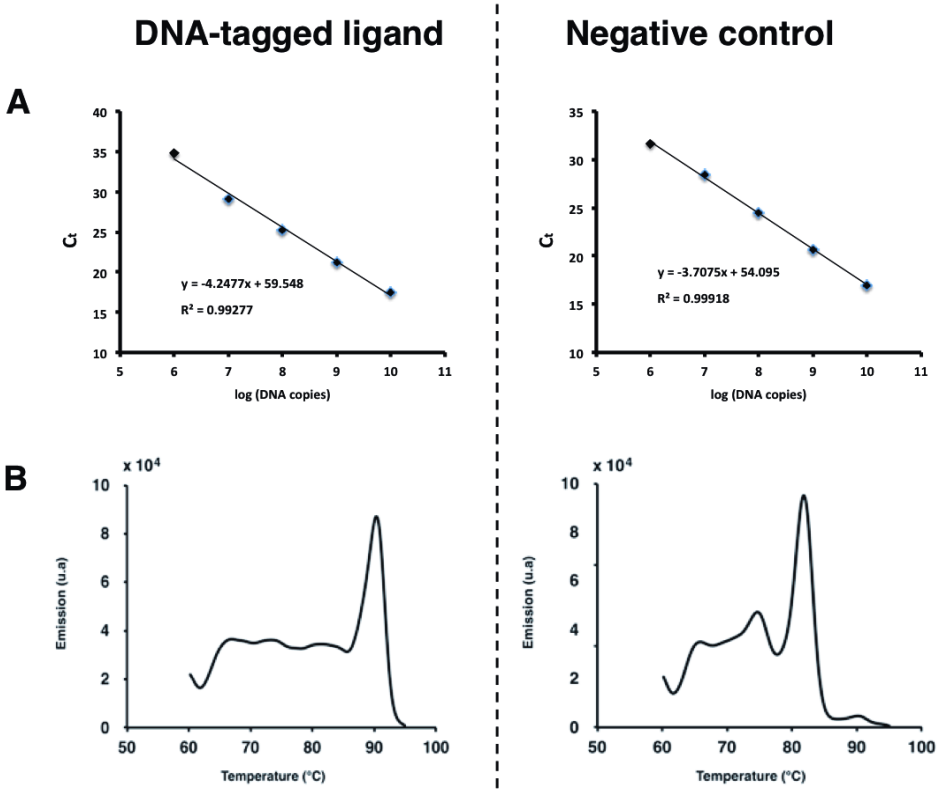


Figure 5.2. Representative calibration curves (A) used for the qPCR quantification of DNA-tagged ligands and negative controls in model selection experiments. The orthogonality of each primer set was checked by melt curve analysis (B) to ensure no-cross amplification reactions during the analyses.

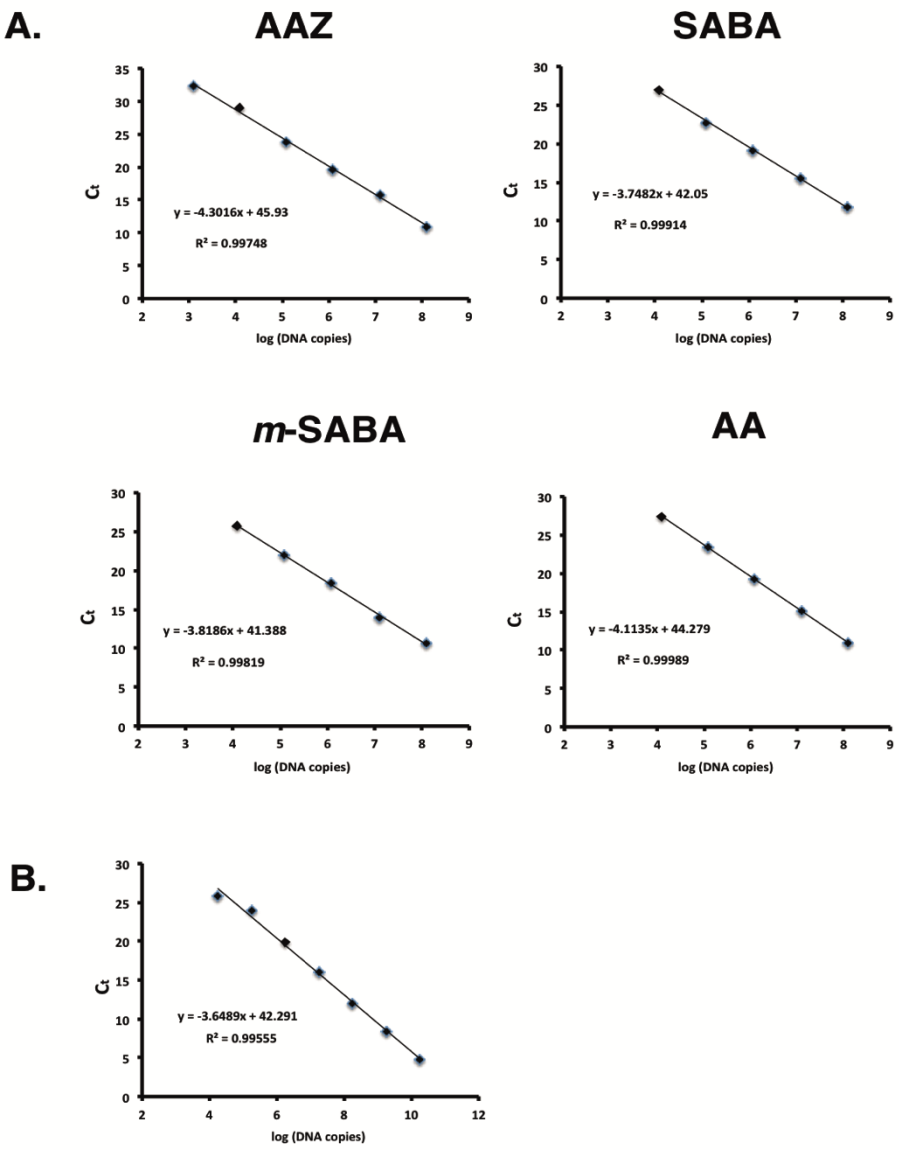


Figure 5.3. Representative calibration curves used for qPCR quantification of DECL library in model selection experiments. A. Calibration curves used in sequence-specific primer design experiments. Each primer set was used for the quantification of ligand recovery, each containing one of the DNA-tagged CAIX ligands (AAZ, SABA, *m*-SABA) and an acetyl group as a terminal moiety. B. Calibration curve used for the quantification of library recovery after selection.

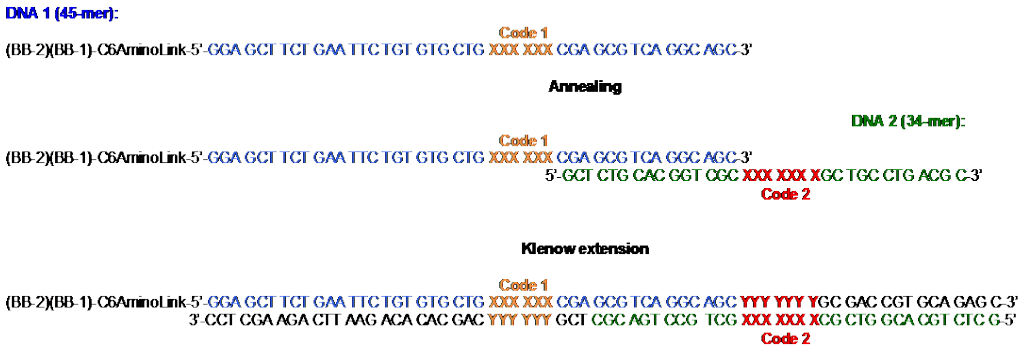


Figure 5.4. Schematic representation of the encoding strategy used for the construction of the Leimbacher library. Building blocks 1 (BB1) are coupled to 45-mer DNA oligonucleotides (DNA 1) through a C6 amino linker (C6-AminoLink). Each DNA strand contains an encoding region (code 1), which is specific for each building block 1. After the coupling reaction between BB1 and building blocks 2 (BB2), a Klenow fill-in polymerization procedure is used to introduce a second barcode element (DNA 2), which encodes the second set of building blocks (BB2).

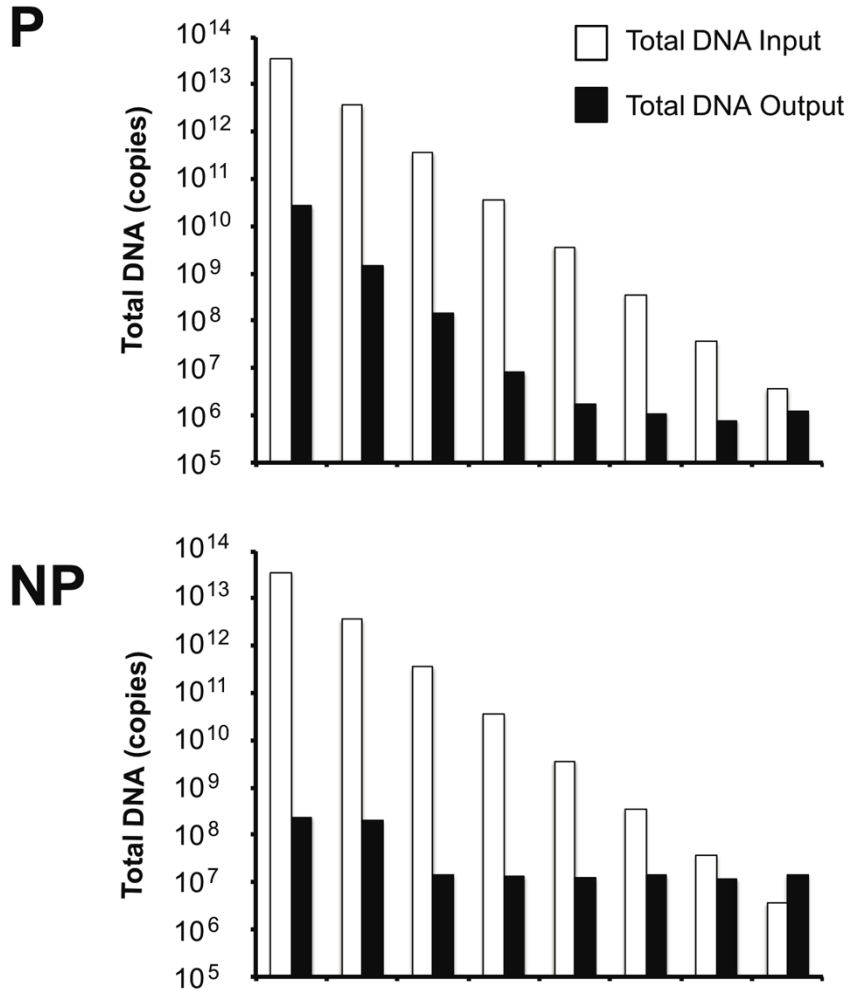


Figure 5.5. qPCR quantification of library recovery after selections against biotinylated CAIX plotted on a logarithmic scale. For each experiment, different inputs of the library (total DNA input, white bar) were used, ranging between 3.6 x 10^{13} and 3.6 x 10^{6} copies of total DNA. The respective outputs (total DNA output, black bar) are depicted in black. P = Protein selections against biotinylated CAIX; NP = No protein selection.

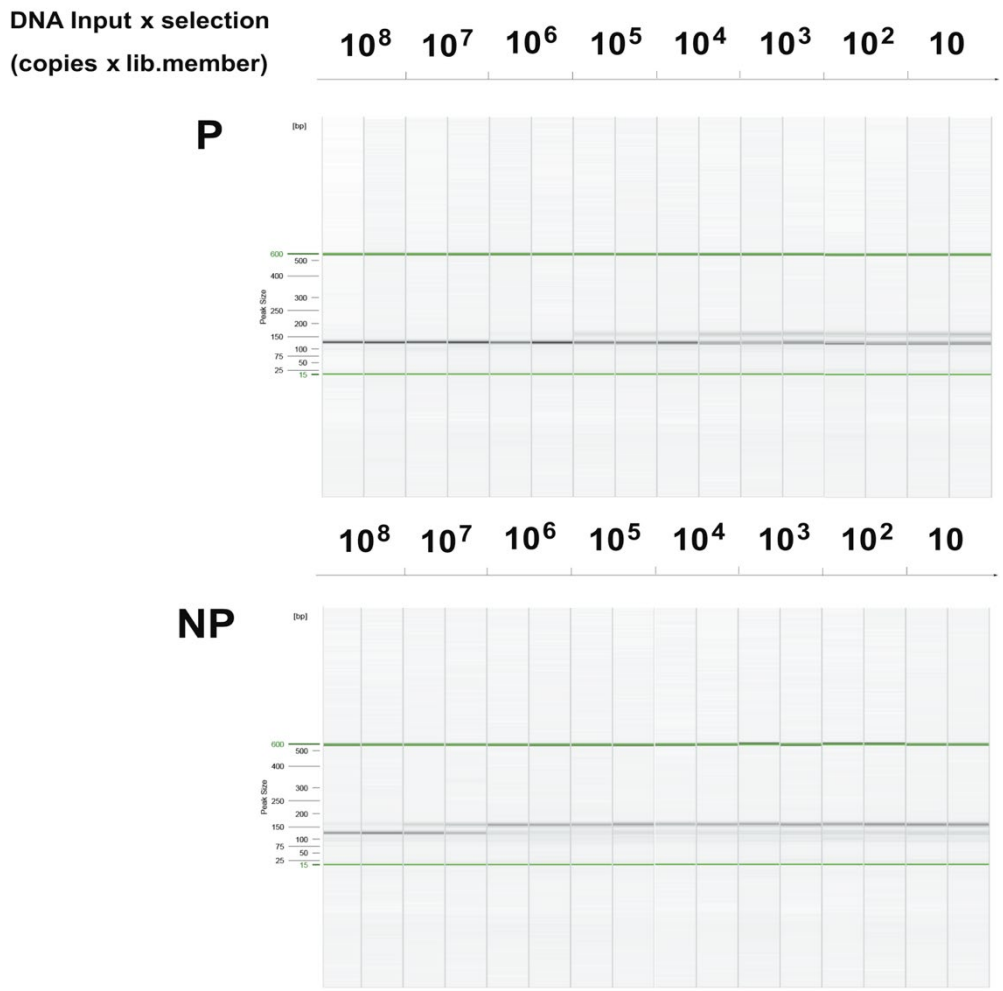


Figure 5.6. Gel electrophoresis analysis after PCR1 both after protein selections (P) and no protein selection (NP). Below an input of 10^4 (P) / 10^6 (NP) copies per library member, primer dimers formation was observed. These results showed a correlation with the quantitative PCR results, where the total amount of DNA in the output was calculated (Figure 5.5).

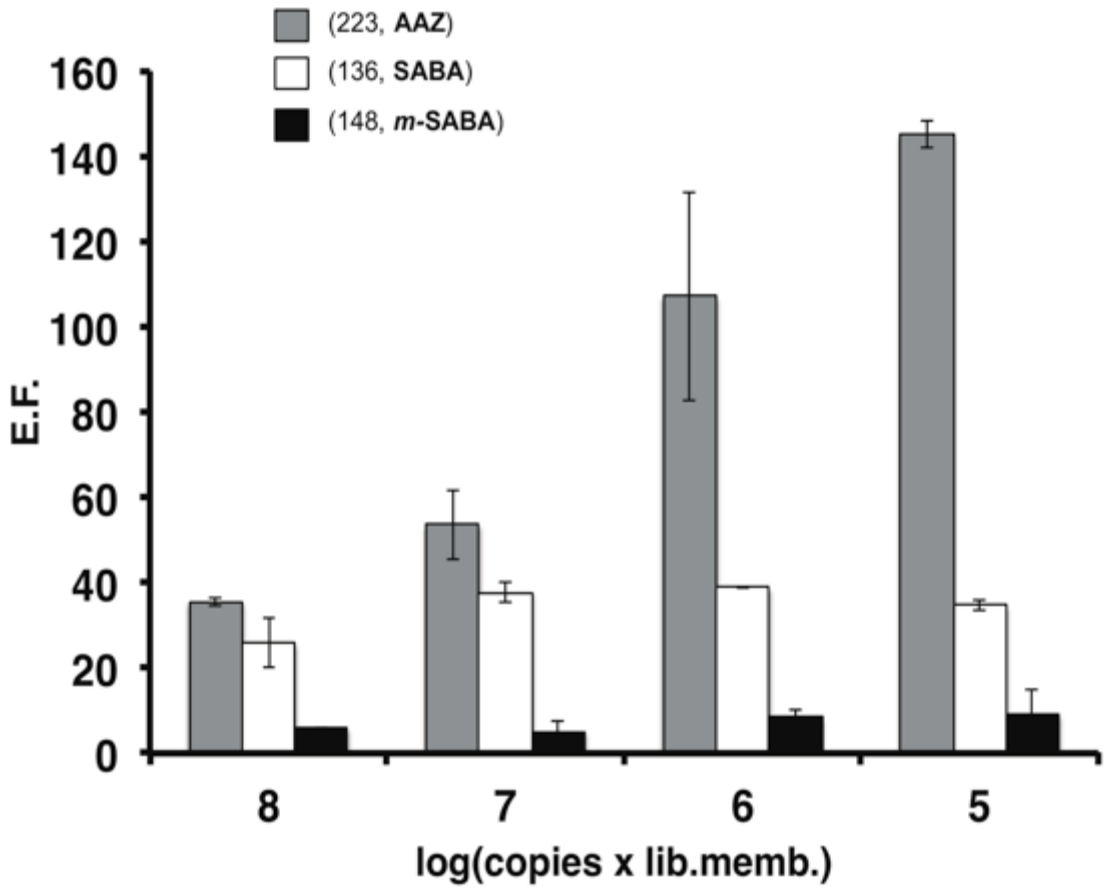


Figure 5.7. Enrichment factors calculated for representative binding pairs, containing the three CAIX ligands, after library selections and sequencing. Selections were performed with different inputs of library, ranging between 10^8 and 10^5 copies per library member. Enrichment factors were calculated using the following formula:

$$E.F. = \frac{counts \ \ x \ size \ library}{total \ number \ of \ counts}$$

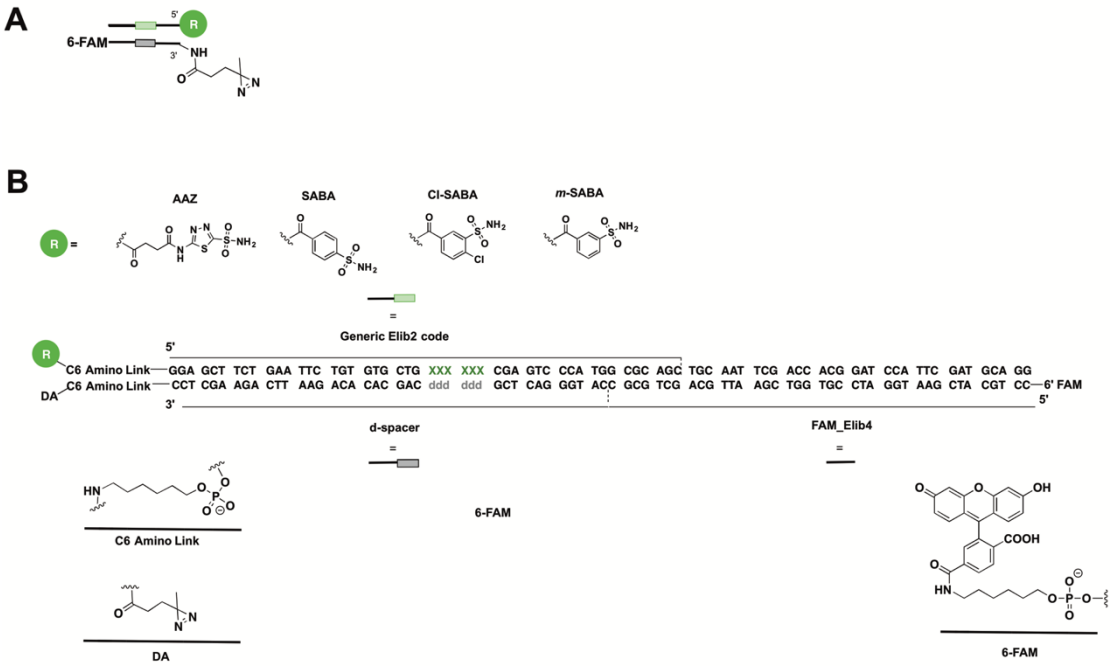


Figure 5.8. Graphical representation (A) and oligonucleotide sequence (B) of the FAM-DNA-CAIX ligands used for the model selections against CAIX using the photocrosslinking procedure.

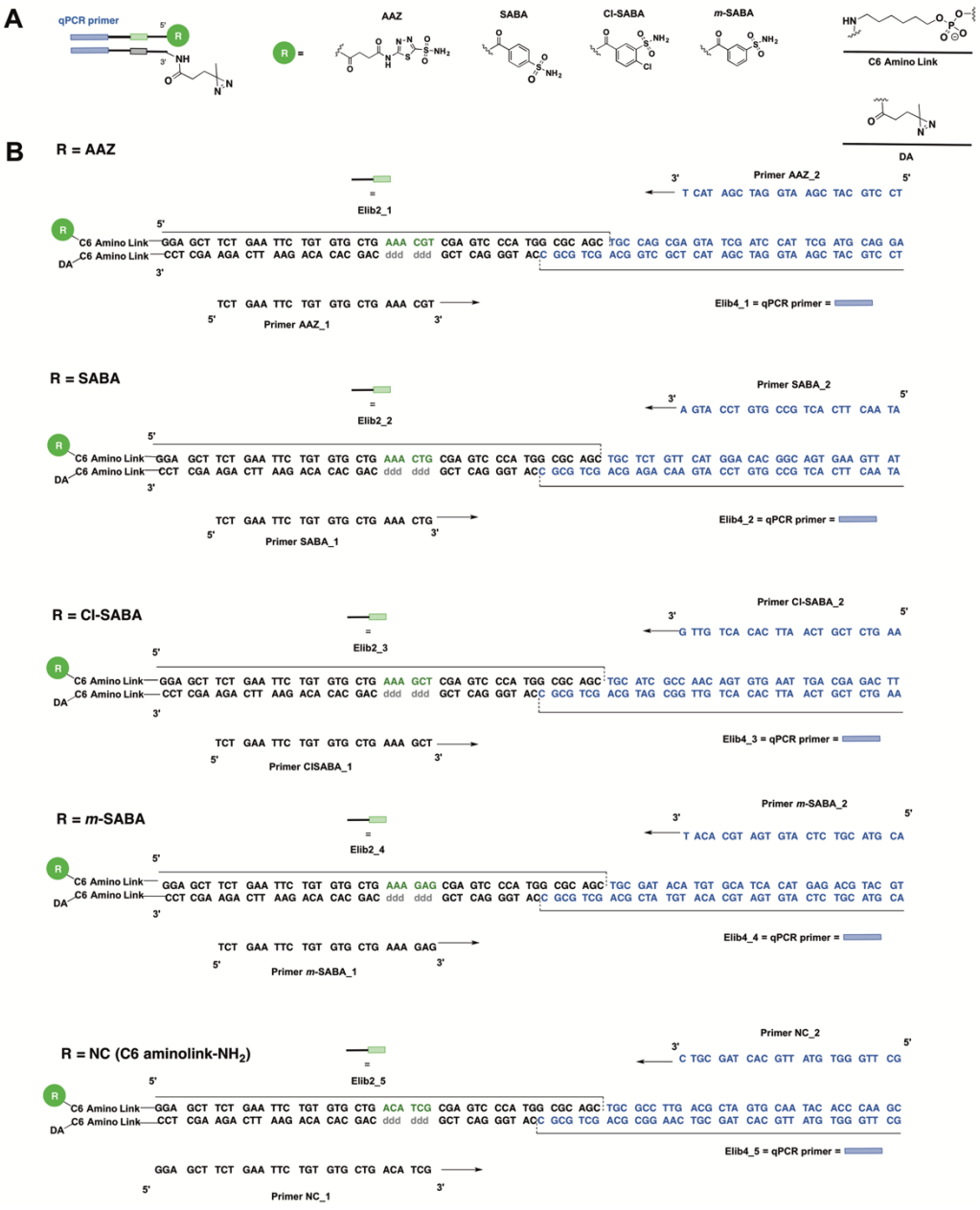


Figure 5.9. Graphical representation (A) and oligonucleotide sequences (B) of the DNA-CAIX ligands used as input in selection experiments. The sequences of the primers used for the qPCR quantification are displayed in (B).

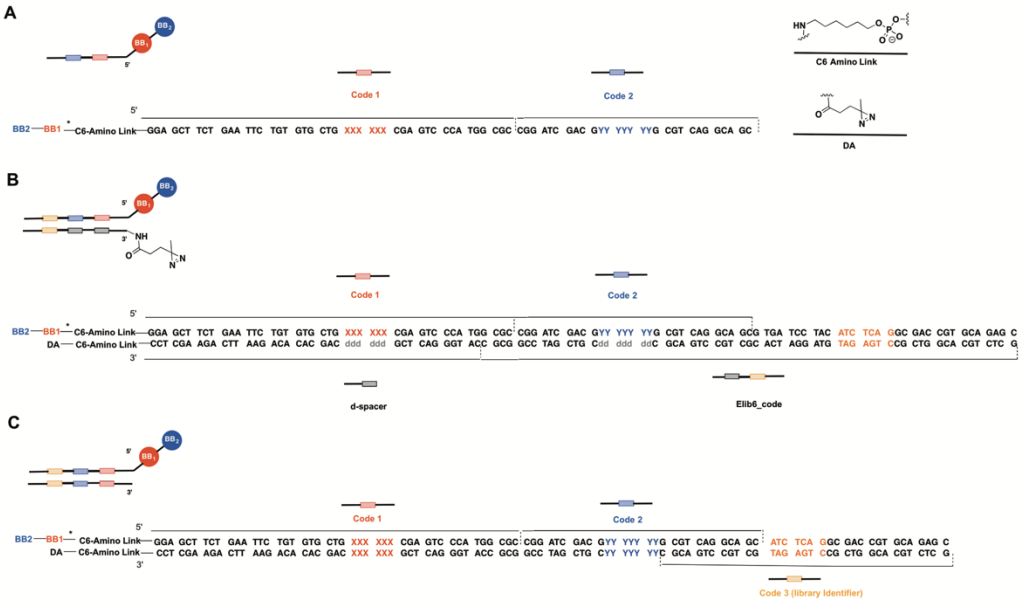


Figure 5.10. A. DNA encoding scheme of the single-stranded single pharmacophore library synthesized using a splint-ligation procedure. DNA-encoded compounds are displayed at the 5' end of a 45-mer oligonucleotide carrying a sequence codon (Code 1, red), which is used to encode the first set of building blocks (BB1, red). After split-and-pool of the first set of conjugates, a 29-mer oligonucleotide carrying the second sequence codon (Code 2, blue) was ligated through splint-ligation, using a 14-mer DNA adapter. B. DNA-encoding scheme of the photoreactive library. C. Double-stranded DNA library analogue of the DECL above mentioned (A).

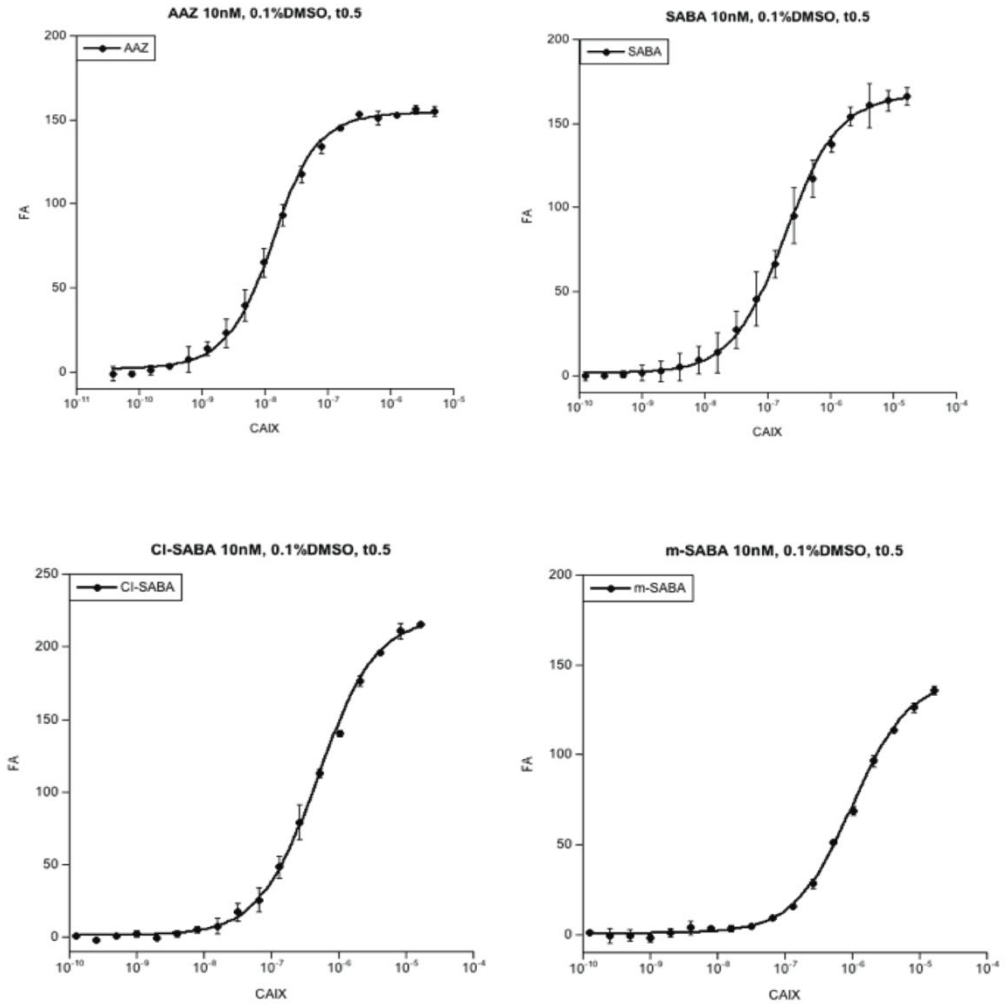


Figure 5.11.Fluorescence polarization measurements of the CAIX-ligands-FITC (AAZ - FITC, SABA - FITC, Cl-SABA - FITC, and m-SABA - FITC) used as positive controls in the photocrosslinking model selection setting. Kd meaurements: AAZ - FITC (Kd = 8.7 ± 0.5 nM; R= 0.99936), SABA - FITC (Kd = 192 ± 7 nM; R = 0.99956), Cl–SABA - FITC (Kd = 503 ± 21 nM; R = 0.99943), and m-SABA - FITC (Kd = 1.029 ± 0.045 μM; R = 0.99942).

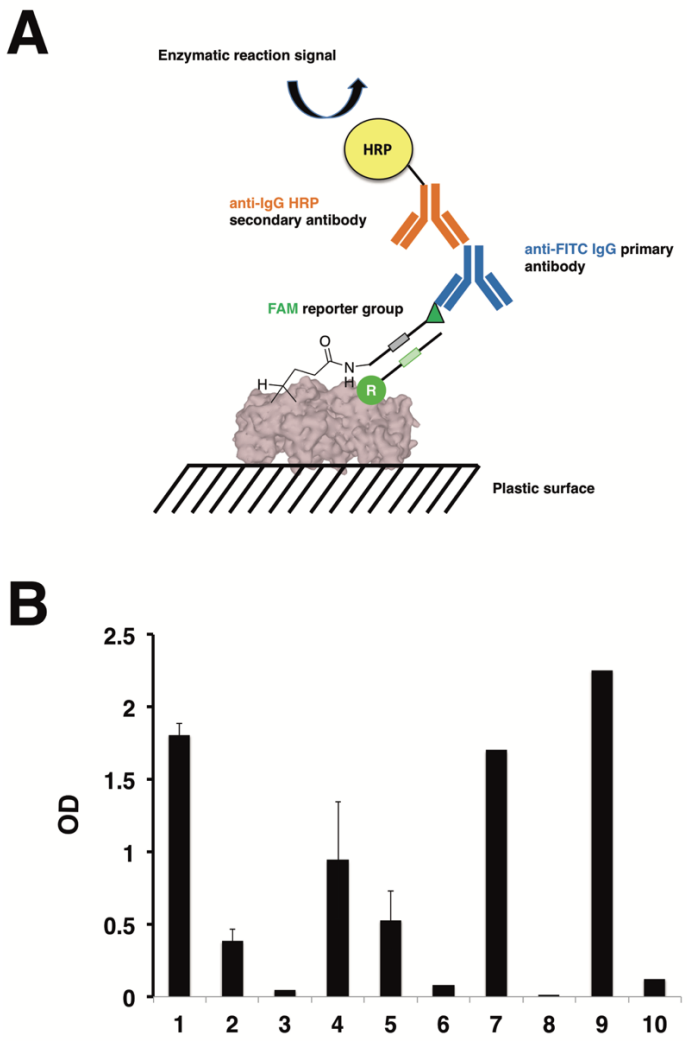


Figure 5.12. The formation and stability of a covalent adduct between his-CAIX and the FAM-tagged oligonucleotide carrying the photoreactive moiety using an ELISA procedure. A monoclonal antibody specific to fluorescein was used as the primary reagent, eventually leading to the concentration of horseradish peroxidase onto the target protein immobilized on a solid support and to the conversion of a chromogenic substrate. Lane 1: CAIX-AAZ-FAM-complex after UV Irradiation at 365 nm (IR) (Input: FAM-DNA-AAZ: 5 pmol) Protein concentration: 2 μM, Lane 2: no UV irradiation control (NIR). Lane 3: no protein control (NP). Lane 4: CAIX-NC-FAM-complex after UV Irradiation at 365 nm (IR) (Input: FAM-DNA-NC: 5 pmol) Protein concentration: 2 μM. Lane 5: NIR control. Lane 6: NP control. Lane 7 – 10: controls. (lane 7: his-CAIX treated with anti-6X His tag (HRP) (coating control); lane 8: his-CAIX treated with anti-FITC IgG/anti-IgG-HRP antibody. Lane 9: F8-FITC treated with anti-FITC HRP/anti-IgG-HRP. Lane 10: blank of the method.

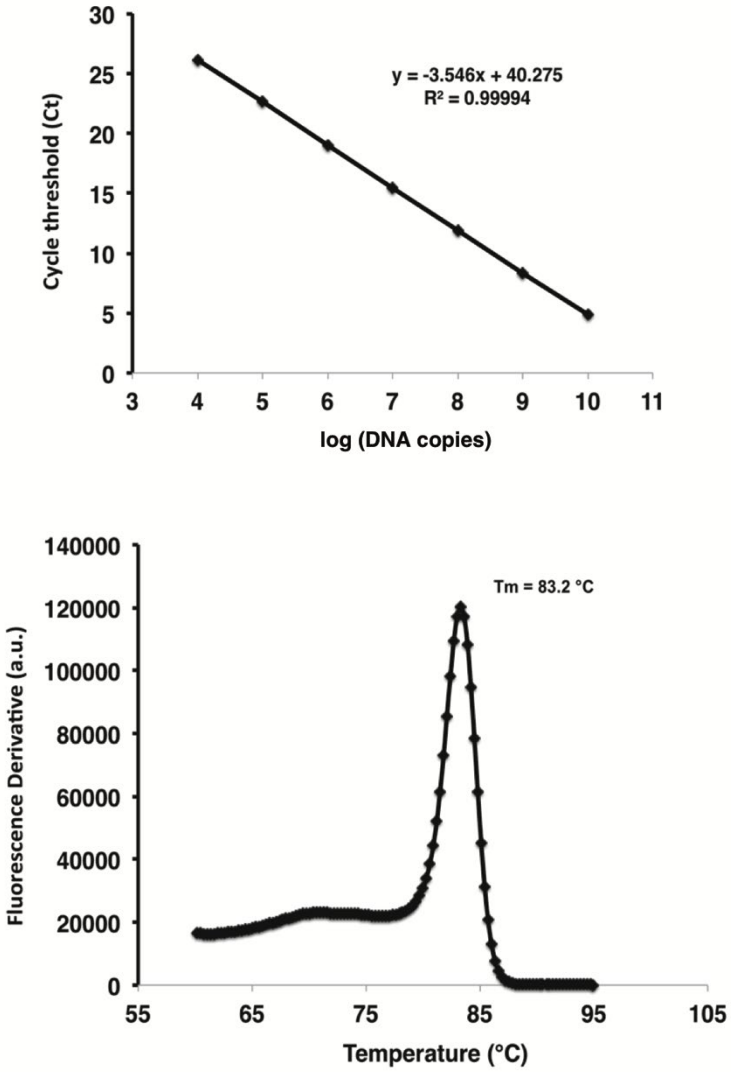


Figure 5.13. Representative calibration curves and melt curve profile used for the qPCR quantification of FAM-DNA CAIX ligand derivatives in model selection experiments. The calibration curve is plotted in a semi-logarithmic scale. Primer sets used for the quantification: FAM_DNA_CAIX ligand Primers: LB_FP; ESAC_q_PCR RP.

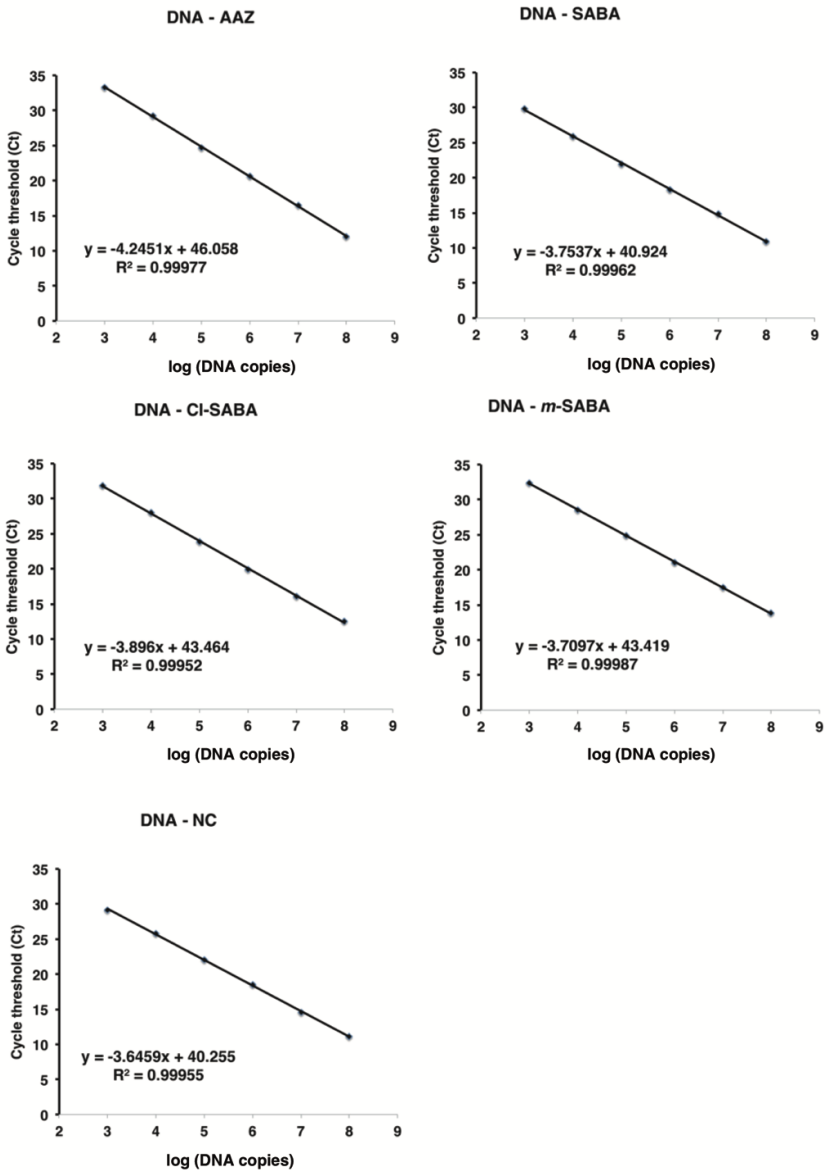


Figure 5.14. Representative calibration curves used for the qPCR quantification of DNA-CAIX ligand derivatives after model selection experiments. Calibration curves are plotted in semi-logarithmic scale. Primers sets used for the quantification: DNA_AAZ (PrimerAAZ_1, PrimerAAZ_2); DNA_SABA (PrimerSABA_1/PrimerSABA_2); DNA_Cl_SABA (PrimerClSABA_1/PrimerClSABA_2); DNA_*m*-SABA (PrimermSABA_1 / PrimermSABA_2); DNA_NC (PrimerNC_1 / PrimerNC_2).

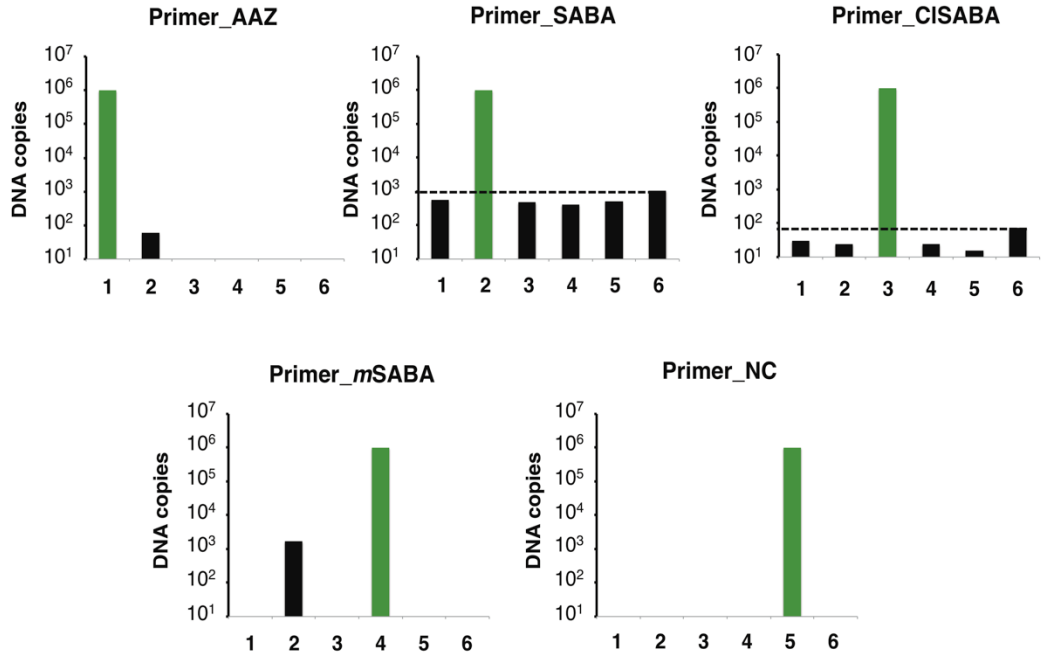


Figure 5.15. Evaluation of the orthogonality of the qPCR primer sets used for the quantification of DNA-CAIX ligands. Each DNA-CAIX ligand (quantity: 10^6 copies of DNA molecules) was quantified with various sets of primers (Primers_AAZ, Primers_SABA, Primers_Cl-SABA, Primers_mSABA, Primers_NC). 1: DNA-AAZ; 2: DNA-SABA, 3: DNA-Cl-SABA, 4: DNA-mSABA, 5: DNA-NC. The lowest limit of detection is also indicated (6, no template control). Results show that each ligand is efficiently amplified only by the corresponding primer set, respectively (green bar). Cross-amplification is not efficient when primer sets are not compatible with a template having non-complementary primer binding sites (black bars).

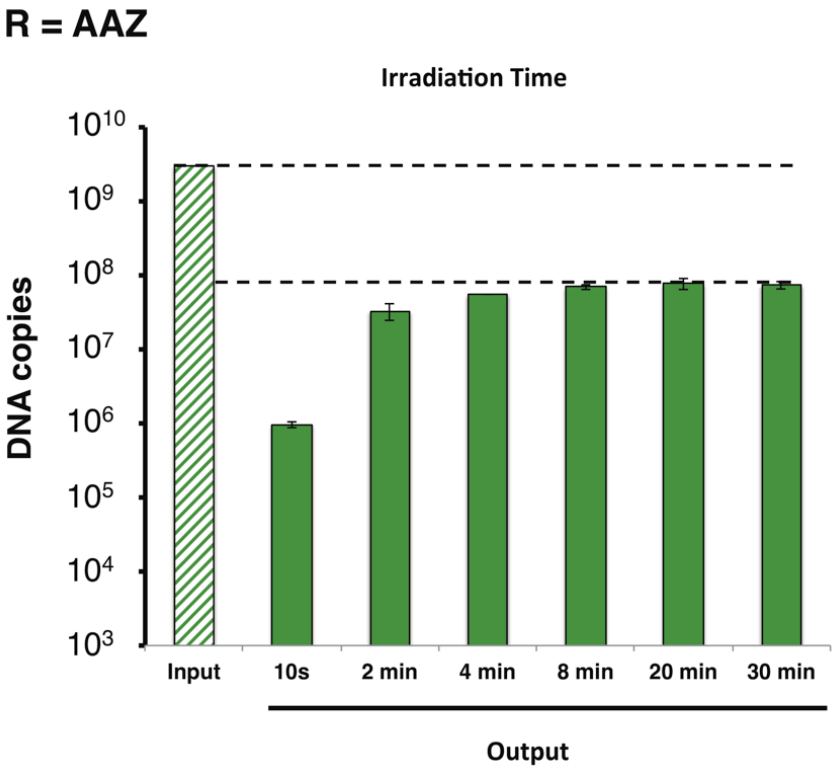


Figure 5.16. Evaluation of the optimal UV irradiation time for the photocrosslinking reaction. Model selections were performed against his-CAIX (2 μM). Photoreactive DNA-AAZ ligand (3 x 10^9 copies of DNA molecules) was used as input. The incubation mixtures of DNA-AAZ and CAIX were irradiated for different time frames. After denaturation of the protein, capture on magnetic beads/washes/elution, the quantity of DNA molecules in the output solution was analyzed by qPCR. The quantity of the DNA after selection is plotted in logarithmic scale (green bars). After 20 minutes, no improvement in the recovery yield was observed.

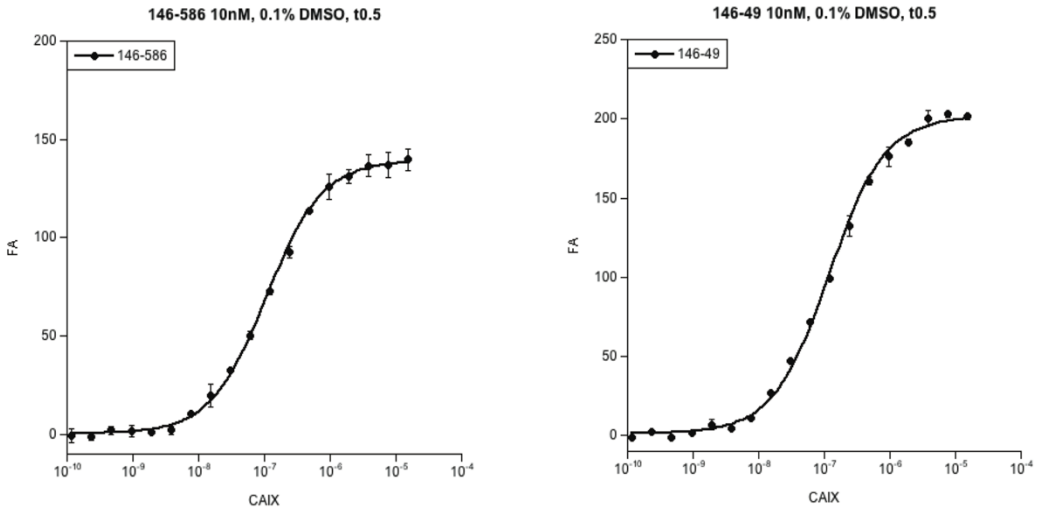


Figure 5.17. Fluorescence polarization measurements of 146_586 FITC and 146_49 FITC derivatives. Kd measurements: 146/586_FITC: Kd = 105 ± 4 nM, R= 0.99951, 146/49_FITC: Kd= 116 ± 6 nM R= 0.99912.

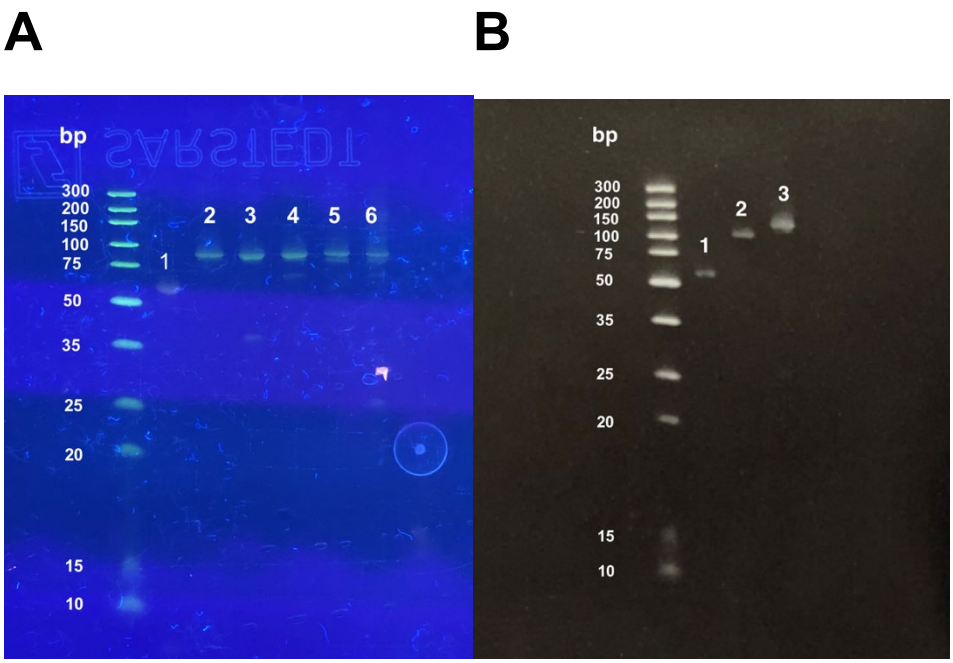


Figure 5.18. Representative 15% TBE-Urea gel electrophoresis of the model DNA-CAIX ligand photoreactive derivatives. Lane 1: DA_Elib4_X before Klenow fill-in. Lanes 2-6: DNA-CAIX ligand derivatives after Klenow polymerization B. Representative 15% TBE-Urea gel electrophoresis analysis of the photocrosslinking library. Lane 1: Single-stranded single pharmacophore library Lane 2: Photoreactive strand (DA_Elib6) Lane 3: Photoreactive library after Klenow fill-in polymerization.

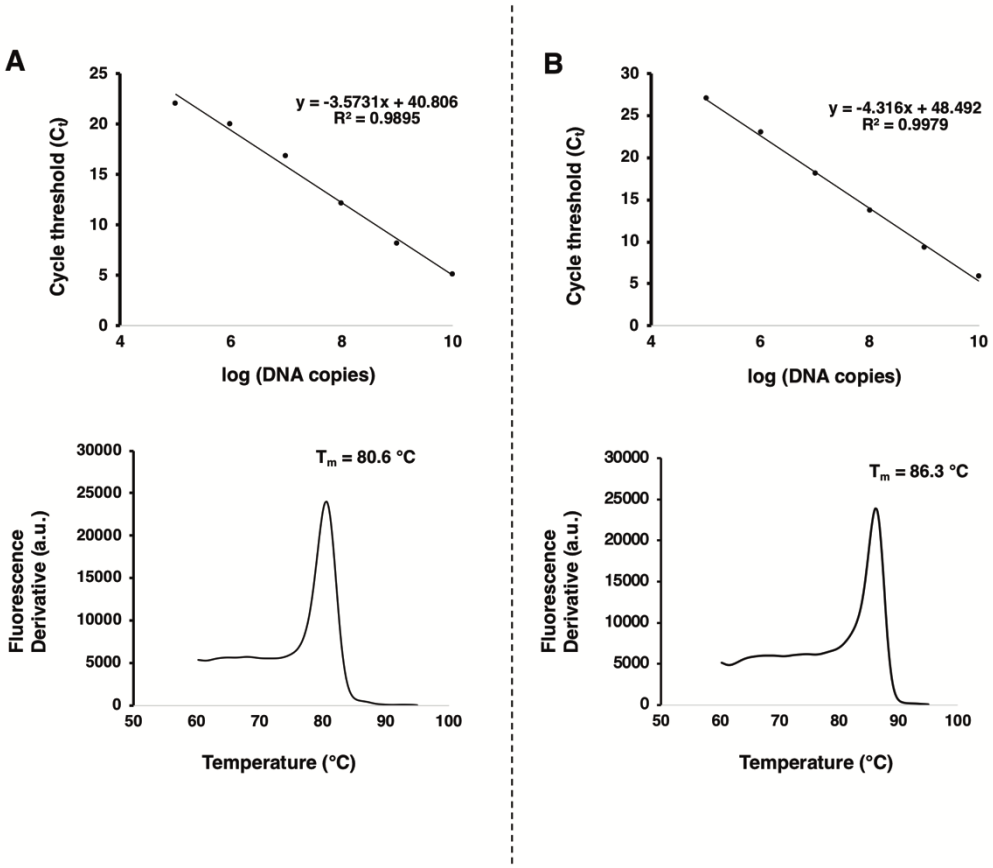


Figure 5.19 Representative calibration curves and melt curves used for the quantification of the universal code (A) and ligation product (B) in the experiments described in Paragraph 2.3.2. Calibration curves are plotted in a semi-logarithmic scale. A. Primers sets used for the quantification: A. LB_FP; CODE1_RV. B. LB_FP: CODE2_RV.

6. ACKNOWLEDGEMENTS

First of all, I would like to thank Philogen and IUSS Pavia for providing me with the opportunity to pursue an industrial PhD in Biotechnology.

I would like to thank Prof. Dr. Dario Neri for giving me the opportunity to work in Philochem laboratories and perform my doctoral thesis in the multidisciplinary field of DNA-encoded chemical libraries. I am very grateful for the many scientific discussions and useful advice, which were fundamental for the completion of the thesis work.

I greatly acknowledge Prof. Dr. Federico Forneris for being my supervisor at IUSS Pavia and for always being available and supportive during my thesis.

I am very thankful to Dr. Florent Samain for his excellent supervision and constant assistance and guidance throughout these years. It was a great pleasure to work in the Chemistry team, and I wish the group all the best for the future.

My special thanks go Dr. Etienne Donckèle, Dr. Luca Prati, Dr. Elena Gabriele and Martina Bigatti for always being very helpful and I really appreciated to work with them on many scientific projects. I wish them all the best for their future careers.

My acknowledgement belongs to PD. Dr. Jörg Scheuermann for his scientific advice and valuable input in DECL field.

Finally, I would like to thank the current and former members of Philochem AG for the pleasant working atmosphere and nice time in (and outside) the lab.

Lastly, many thanks to my family, friends and Chiara for always being very patient with me and for their precious support during these years in Switzerland.

7.ABBREVIATIONS

6-FAM: 6-carboxyfluorescein

AAZ: acetazolamide

ACN: acetonitrile

ADME: absorption, distribution, metabolism and excretion

ADMET: absorption, distribution, metabolism, excretion and toxicity

BB: building block

bp: base pair

BTE: binding trap enrichment

CADD: computer-aided drug design

CAIX: carbonic anhydrase IX

Cbz: benzyloxycarbonyl group

Cl-SABA: 4-chloro-3-sulfamoyl benzoic acid

CPG: controlled pore glass

cPP: cell-penetrating peptide

C_t: cycle threshold

CuAAC: copper(I)-catalyzed azide-alkyne cycloaddition

DCC: dynamic combinatorial chemistry

DECL: DNA-encoded chemical library

DESPS: DNA-encoded solid-phase synthesis

DFC: dual flow chromatography

DIC: 1,3-diisopropylcarbodiimide

DIPEA: N,N-diisopropylethylamine

DMA: dimethylacetamide

DMF: dimethylformamide

DMPK: Drug Metabolism and Pharmacokinetics

DMSO: dimethylsulfoxide

DMT-MM: 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride

DNA: deoxyribonucleic acid

dNTPS: deoxynucleoside triphosphate

DPAL: DNA-programmed affinity labelling

dsDNA: double-stranded DNA

DTS: DNA-templated synthesis

EDC: 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide

EDCCL: DNA-encoded dynamic combinatorial chemical libraries

ESAC: encoded self-assembling chemical library

ESI: electrospray ionization

FACS: fluorescence-activated cell sorting

FDA: Food and Drug administration

FITC: fluorescein isothiocyanate

Fmoc: fluorenylmethyloxycarbonyl

FP: fluorescence polarization

GPCRs: G protein-coupled receptor

HFIP: hexafluoroisopropanol

His: polyhistidine tag

HOAt: 1-hydroxy-7-azabenzotriazole

HPLC: high-performance liquid chromatography

HRP: horseradish peroxidase

HTDS: high throughput drug screening

HTS: high throughput screening

IDPCR: interaction-dependent PCR

IDUP: interaction determination using unpurified proteins

IgG: immunoglobulin G

IL-2: interleukin-2

IMAC: ion metal affinity chromatography

ITC: isothermal titration calorimetry

IVC: in vitro compartmentalization

K_d: dissociation constant

LBVS: ligand-based virtual screening

LC-MS: liquid chromatography-mass spectrometry

LNA: locked nucleic acid

logP: partition coefficient octanol-water

m-SABA: m-sulfamoyl benzoic acid

MeOH: methanol

MOPS: 3-(N-morpholino)propanesulfonic acid

NGS: next-generation sequencing

NMM: N-methylmorpholine

NMR: nuclear magnetic resonance

OBOC: one-bead-one-compound

OXYMA: ethyl cyanohydroxyiminoacetate

PAINS: pan-assay interference compounds

PBS: phosphate-buffered saline

PC: photocrosslinker

PCR: polymerase chain reaction

PEG: polyethylene glycol

PNA: peptide nucleic acid

PPI: protein-protein interaction

qPCR: quantitative PCR

QSAR: quantitative structure-activity relationship

RP-HPLC: reversed-phase HPLC

rpm: revolutions per minute

s-NHS: sulfo N-hydroxysulfosuccinimide

SABA: p-sulfamoyl benzoic acid

SAR: structure-activity relationship

SBVS: structure-based virtual screening

scFv: single-chain variable fragment

SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis

SM: small molecule

SMD: small molecule drug

ssDNA: single-stranded DNA

TEA: triethylamine

TEAA: triethylammonium acetate

TFA: trifluoroacetic acid

THF: tetrahydrofuran

TLC: thin-layer chromatography

UV: ultraviolet

8. REFERENCES

1. Erlanson, D. A.; Fesik, S. W.; Hubbard, R. E.; Jahnke, W.; Jhoti, H. Twenty years on: the impact of fragments on drug discovery. *Nat. Rev. Drug. Discov.* **2016**, *15* (9), 605–619.

2. Trabocchi, A.; Lenci, E. *Small Molecule Drug Discovery: Methods, Molecules and Applications*; 1st Edition; Elsevier: 2019.

3. Ganellin, C. R.; Jefferis, R.; Roberts, S. M. *Introduction to biological and small molecule drug research and development: theory and case studies*; 1st Edition; Elsevier: 2013.

4. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **2001**, *46* (1-3), 3–26.

5. Morrow, T.; Felcone, L. H. Defining the difference: What Makes Biologics Unique. *Biotechnol. Healthc.* **2004**, *1* (4), 24–29.

6. Mullard A. 2019 FDA drug approvals. *Nat. Rev. Drug Discov.* **2020**, *19* (2), 79–84.

7. Berdigaliyev, N.; Aljofan, M. An overview of drug discovery and development. *Future Med. Chem.* **2020**, *12* (10), 939–947.

8. Cragg, G. M.; Newman, D. J. Natural products: a continuing source of novel drug leads. *Biochim. Biophys. Acta.* **2013**, *1830* (6), 3670–3695.

9. Kraljevic, S.; Stambrook, P. J.; Pavelic, K. Accelerating drug discovery. *EMBO Rep.* **2004**, *5* (9), 837–842.

10. Hughes, J. P.; Rees, S.; Kalindjian, S. B.; Philpott, K. L. Principles of early drug discovery. *Br. J. Pharmacol.* **2011**, *162* (6), 1239–1249.

11. Cheng, A. C.; Coleman, R. G.; Smyth, K. T.; Cao, Q.; Soulard, P.; Caffrey, D. R.; Salzberg, A. C.; Huang, E. S. Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.* **2007**, *25* (1), 71–75.

12. Carnero, A. High throughput screening in drug discovery. *Clin. Transl. Oncol.* **2006**, *8* (7), 482–490.

13. Kenny, B. A.; Bushfield, M.; Parry-Smith, D. J.; Fogarty, S.; Treherne, J. M. The application of high-throughput screening to novel lead discovery. *Prog. Drug. Res.* **1998**, *51*, 245–269.

14. Szymański, P.; Markowicz, M.; Mikiciuk-Olasik, E. Adaptation of High-Throughput Screening in Drug Discovery–Toxicological Screening Tests. *Int. J. Mol. Sci.* **2012**, *13* (1), 427–452.

15. Attene-Ramos, M. S.; Austin, C. P.; Xia, M. High Throughput Screening. In: *Encyclopedia of Toxicology* (Third Edition); Wexler, P. Eds.; Oxford: Academic Press. 2014; pp 916–917.

16. Chen, T.; Kablaoui, N.; Little, J.; Timofeevski, S.; Tschantz, W. R.; Chen, P.; Feng, J.; Charlton, M.; Stanton, R.; Bauer, P. Identification of small-molecule inhibitors of the JIP-JNK interaction. *Biochem. J.* **2009**, *420* (2), 283–294.

17. Mabonga, L.; Kappo, A. P. Protein-protein interaction modulators: advances, successes and remaining challenges. *Biophys. Rev.* **2019**, *11* (4), 559–581.

18. Santos, R.; Ursu, O.; Gaulton, A.; Patrícia Bento, A.; Donadi, R. S.; Bologa, C. G.; Karlsson, A.; Al-Lazikani, B.; Hersey, A.; Oprea, T. I.; Overington, J. P. A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **2017**, *16* (1), 19–34.

19. Mayr, L. M.; Bojanic, D. Novel trends in high-throughput screening. *Curr. Opin. Pharmacol.* **2009**, *9* (5), 580–588.

20. Erlanson, D. A.; Davis, B. J.; Jahnke, W. Fragment-Based Drug Discovery: Advancing Fragments in the Absence of Crystal Structures. *Cell Chem. Biol.* **2019**, *26* (1), 9–15.

21. Price, A. J.; Howard, S.; Cons, B. D. Fragment-based drug discovery and its application to challenging drug targets. *Essays Biochem.* **2017**, *61* (5), 475–484.

22. Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R. Fragment-based lead discovery. *Nat. Rev. Drug. Discov.* **2004**, *3* (8), 660–672.

23. Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A 'rule of three' for fragment-based lead discovery? *Drug Discov. Today* **2003**, *8* (19), 876–877.

24. Kirsch, P.; Hartman, A. M.; Hirsch, A. K. H.; Empting, M. Concepts and Core Principles of Fragment-Based Drug Design. *Molecules* **2019**, *24* (23), No. 4309.

25. Hann, M. M.; Leach, A. R.; Harper, G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (3), 856–864.

26. Page, M. I.; Jencks, W. P. Entropic Contributions to Rate Accelerations in Enzymic and Intramolecular Reactions and the Chelate Effect. *Proc. Natl. Acad. Sci. U. S. A.* **1971**, *68* (8), 1678–1683.

27. Finkelstein, A. V.; Janin, J. The price of lost freedom: entropy of bimolecular complex formation. *Protein Eng.* **1989**, 3 (1), 1–3.

28. Williams, G.; Ferenczy, G. G.; Ulander, J.; Keserü, G. M. Binding thermodynamics discriminates fragments from druglike compounds: a thermodynamic description of fragment-based drug discovery. *Drug Discov. Today* **2017**, *22* (4), 681–689.

29. Ferenczy, G. G.; Keserü, G. M. On the enthalpic preference of fragment binding. *Med. Chem. Commun.* **2016**, *7* (2), 332–337.

30. Turnbull, A.; Boyd, S.; Walse, B. Fragment-based drug discovery and protein-protein interactions. *Res. Rep. Biochem.* **2014**, *4*, 13–26.

31. Burke, J. P.; Bian, Z.; Shaw, S. Zhao, B.; Goodwin, C. M.; Belmar, J.; Browning, C. F.; Vigil, D.; Friberg, A.; Camper, D. V.; Rossanese, O. V.; Lee, T.; Olejniczak, E. T.; Fesik, S. W. Discovery of Tricyclic Indoles That Potently Inhibit Mcl-1 Using Fragment-Based Methods and Structure-Based Design. *J. Med. Chem.* **2015**, *58* (9), 3794–3805.

32. Davies, T. G.; Wixted, W. E.; Coyle, J. E.; Griffiths-Jones, C.; Hearn, K.; McMenamin, R.; Norton, D.; Rich, S. J.; Richardson, C.; Saxty, G.; Willems, H. M. G.; Woolford, A. J-A.; Cottom, J. E.; Kou, J-P.; Yonchuk, J. G.; Feldser, H. G.; Sanchez, Y.; Foley, J. P.; Bolognese, B. J.; Logan, G.; Polodin, P. L.; Yan, H.; Callahan, J. F.; Heightman, T. D.; Kerns, J. K. Monoacidic Inhibitors of the Kelch-like ECH-Associated Protein 1: Nuclear Factor Erythroid 2-Related Factor 2 (KEAP1:NRF2) Protein-Protein Interaction with High Cell Potency Identified by Fragment-Based Discovery *J. Med. Chem.* **2016**, *59* (8), 3991–4006.

33. Leelananda, S. P.; Lindert, S. Computation methods in drug discovery. *Beilstein J. Org. Chem.* **2016**, *12*, 2694–2718.

34. Kapetanovic, I. M. Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach. *Chem. Biol. Interact.* **2008**, *171* (2), 165–176.

35. Kalyaanamoorthy, S.; Chen, Y-P., P. Structure-based drug design to augment hit discovery. *Drug. Discov. Today* **2011**, *16* (17–18), 831–839.

36. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W. Jr. Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2014**, *66* (1), 334–395.

37. Lavecchia, A.; Di Giovanni, C. Virtual screening strategies in drug discovery: a critical review. *Curr. Med. Chem.* **2013**, *20* (23), 2839–2860.

38. Cheng, T.; Li, Q.; Zhou, Z.; Wang, Y.; Bryant, S. H. Structure-based virtual screening for drug discovery: a problem-centric review. *AAPS J.* **2012**, *14* (1), 133–141.

39. Lionta, E.; Spyrou, G.; Vassilatis, D. K.; Cournia, Z. Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Curr. Top. Med. Chem.* **2014**, *14* (16), 1923–1938.

40. Ripphausen, P.; Nisius, B.; Bajorath, J. State-of-the-art in ligand-based virtual screening. *Drug. Discov. Today*. **2011**, *16* (9–10); 372–376.

41. Xu, J.; Hagler, A. Chemoinformatics and Drug Discovery. *Molecules* **2002**, *7* (8), 566–600.

42. Baig, M. H.; Ahmad, K.; Roy, S.; Ashraf, J. M.; Adil, M.; Siddiqui, M. H.; Khan, S.; Kamal, M. A.; Provazník, I.; Choi, I. Computer Aided Drug Design: Success and Limitations. *Curr. Pharm. Des.* **2016**, *22* (5), 572–581.

43. Liu, R.; Li, X.; Lam, K. S. Combinatorial chemistry in drug discovery. *Curr. Opin. Chem. Biol.* **2017**, *38*, 117–126.

44. Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85* (14), 2149–2154.

45. Geysen, H. M.; Meloen, R. H.; Barteling, S. J. Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. *Proc. Natl. Acad. Sci. U. S. A.* **1984**, *81* (13), 3998–4002.

46. Letsinger, R. L.; Mahadevan, V. Oligonucleotide synthesis on a polymer support. *J. Am. Chem. Soc.* **1965**, *87*, 3526–3527.

47. Fréchet, J. M.; Schuerch, C. Solid-phase synthesis of oligosaccharides. I. Preparation of the solid support. Poly[p-(1-propen-3-ol-1-yl)styrene]. *J. Am. Chem. Soc.* **1971**, *93* (2), 492–496.

48. Bunin, B. A.; Ellman, J. A. A general and expedient method for the solid-phase synthesis of 1,4-benzodiazepine derivatives. *J. Am. Chem. Soc.* **1992**, *114* (27), 10997–10998.

49. Furka, A.; Sebestyén, F.; Asgedom, M.; Dibó, G. General method for rapid synthesis of multicomponent peptide mixtures. *Int. J. Pept. Protein. Res.* **1991**, *37* (6), 487–493.

50. Hlaváč, J., Soural, M., Krchňák, V. Practical aspects of combinatorial solid-phase synthesis. In: *Solid-Phase Organic Synthesis: Concepts, Strategies and Applications.* Toy, P. H.; Lam, Y., Eds.; John Wiley & Sons, Inc.: Hobokem, N. J., 2012; pp 95–130.

51. Houghten, R. A. General method for the rapid solid-phase synthesis of large numbers of peptides: specificity of antigen-antibody interaction at the level of individual amino acids. *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82* (15), 5131–5135.

52. Nicolaou, K. C.; Xiao, X.-Y.; Parandoosh, Z.; Senyei, A.; Nova, M. P. Radiofrequency encoded combinatorial chemistry. *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 2289–2291.

53. Geysen, H. M.; Wagner, C. D.; Bodnar, W. M.; Markworth, C. J.; Parke, G. J.; Schoenen, F. J.; Wagner, D. S.; Kinder, D. S. Isotope or mass encoding of combinatorial libraries. *Chem. Biol.* **1996**, *3* (8), 679–688.

54. Ohlmeyer. M. H. J.; Swanson, R. N.; Dillard, L. W.; Reader, J. C.; Asouline, G.; Kokayashi, R.; Wigler, M.; Still, W. C. Complex synthetic chemical libraries indexed with molecular tags. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (23), 10922–10926.

55. Winter, G.; Griffiths, D.; Hawkins, R. E.; Hoogenboom, H. R. Making antibodies by phage display technology. *Annu. Rev. Immunol.* **1994**, *12*, 433-455.

56. Binz, H. K.; Plückthun, A. Engineered proteins as specific binding reagents. *Curr. Opin. Biotechnol.* **2005**, *16* (4), 459–469.

57. Heinis, C.; Rutherford, T.; Freund, S.; Winter, G. Phage-encoded combinatorial chemical libraries based on bicyclic peptides. *Nat. Chem. Biol.* **2009**, *5* (7), 502–507.

58. Mattheakis, L. C.; Bhatt, R. R.; Dower, W. J. An in vitro polysome display system for identifying ligands from very large peptide libraries. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91* (19), 9022–9026.

59. Boder, T.; Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **1997**, *15* (6), 553–557.

60. Baggio, R.; Burgstaller, P.; Hale, S. P.; Putney, A. R.; Lane, M.; Lipovsek, D.; Wright, M. C.; Roberts, R. W.; Liu, R.; Szostak, J. W.; Wagner, R. W. Identification of epitope-like consensus motifs using mRNA display. *J. Mol. Recognit.*, **2002**, *15* (3), 126–134.

61. Giovannoni, L.; Viti, F.; Zardi, L.; Neri, D. Isolation of anti-angiogenesis antibodies from a large combinatorial repertoire by colony filter screening. *Nucleic Acids Res.* **2001**, *29* (5), e27.

62. Nixon, A. E.; Sexton, D. J.; Ladner, R. C. Drugs derived from phage display: from candidate identification to clinical practice. *MAbs*. **2014**, *6* (1), 73–85.

63. Smith, G. P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*. **1985**, *228* (4705), 1315–1317.

64. Wang, L-F.; Yu, M. Epitope identification and discovery using phage display libraries: applications in vaccine development and diagnostics. *Curr. Drug Targets*, **2004**, *5* (1), 1–15.

65. Potocnakova, L.; Bhide, M.; Pulzova, L. B. An Introduction to B-Cell Epitope Mapping and In Silico Epitope Prediction. *J. Immunol. Res.* **2016**, No. 6760830.

66. Rebollo, I. R.; Sabisz, M.; Baeriswyl, V.; Heinis, C. Identification of target-binding peptide motifs by high-throughput sequencing of phage-selected peptides. *Nucleic Acids Res.* **2014**, *42* (22), e169.

67. Majumdar, S.; Siahaan, T. J. Peptide-mediated targeted drug delivery. *Med. Res. Rev.* **2012**, *32* (3), 637–658.

68. Newman, M. R.; Benoit, D. S. W. In Vivo Translation of Peptide-Targeted Drug Delivery Systems Discovered by Phage Display. *Bioconjug. Chem*. **2018**, *29* (7), 2161–2169.

69. Ledsgaard, L.; Kilstrup, M.; Karatt-Vellatt, A.; McCafferty, J.; Laustsen, A. H. Basics of Antibody Phage Display Technology. *Toxins* **2018**, *10* (6), No. 236.

70. McCafferty, J.; Griffiths, A. D.; Winter, G.; Chiswell, D. J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature.* **1990**, *348* (6301), 552–554.

71. Jespers L. S; Roberts A.; Mahler S. M.; Winter G.; Hoogenboom H, R. Guiding the selection of human antibodies from phage display repertoires to a single epitope of an antigen. *Biotechnology (N.Y.)* **1994**, *12* (9), 899–903.

72. Mease, P. J. Adalimumab in the treatment of arthritis. *Ther. Clin. Risk. Manag.* **2007**, *3* (1), 133–148.

73. Asgharpour, A.; Cheng, J.; Bickston, S. J. Adalimumab treatment in Crohn's disease: an overview of long-term efficacy and safety in light of the EXTEND trial. *Clin. Exp. Gastroenterol.* **2013**, *6*, 153–160.

74. Bálint, A.; Farkas, K.; Palatka, K.; Lakner, L.; Miheller, P.; Rácz, I.; Hegede, G.; Vincze, Á.; Horváth, G.; Szabó, A.; Nagy, F.; Szepes, Z.; Gábor, Z.; Zsigmond, F.; Zsóri, Á.; Juhász, M.; Csontos, Á.; Szűcs, M.; Bor, R.; Milassin, Á.; Rutka, M.; Molnár, T. Efficacy and Safety of Adalimumab in Ulcerative Colitis Refractory to Conventional Therapy in Routine Clinical Practice. *J. Crohns Colitis* **2016**, *10* (1), 26–30.

75. Weber, M.; Bujak, E.; Putelli, A.; Villa, A.; Matasci, M.; Gualandi, L.; Hemmerle, T.; Wulhfard, S.; Neri, D. A highly functional synthetic phage display library containing over 40 billion human antibody clones. *PLoS One.* **2014**, 9 (6): e100000.

76. Brenner, S.; Lerner, R. A. Encoded combinatorial chemistry. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (12), 5381–5383.

77. Nielsen, J.; Brenner, S.; Janda, K. D. Synthetic methods for the implementation of encoded combinatorial chemistry. *J. Am. Chem. Soc.* **1993**, *115* (21), 9812–9813.

78. Needels, M. C.; Jones, D. G.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A. Generation and screening of an oligonucleotide-encoded synthetic peptide library. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (22), 10700–10704.

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

80. Scheuermann, J.; Dumelin, C. E.; Melkko, S.; Neri, D. DNA-encoded chemical libraries. *J. Biotechnol.* **2006**, *126* (4), 568–581.

81. Melkko, S.; Dumelin, C. E.; Scheuermann, J.; Neri, D. Lead discovery by DNA-encoded chemical libraries. *Drug Discov. Today.* **2007**, *12* (11–12), 465–471.

82. Nuevolution. Nuevolution Scales Its Chemetics ™ Drug Discovery Platform (press release). **2017.** https://nuevolution.com/nuevolution-technology-progress-nuevolution-scales-itscompound-collection-to-40-trillion-using-its-chemeticsdrug-discovery-platform/

83. Goodnow, R. A. Jr.; Dumelin, C. E; Keefe, A. D. DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nat. Rev. Drug Discov.* **2017**, *16* (2), 131–147.

84. Zimmermann, G.; Neri, D. DNA-encoded chemical libraries: foundations and applications in lead discovery. *Drug Discov. Today.* **2016**, *21* (11), 1828–1834.

85. Neri, D.; Lerner, R. A. DNA-Encoded Chemical Libraries: A Selection System Based on Endowing Organic Compounds with Amplifiable Information. *Annu. Rev. Biochem.* **2018**, *87*, 479–502.

86. Clark, M.A.; Acharya, R.A.; Arico-Muendel, C.C.; Belyanskaya, S.L.; Benjamin, D.R.; Carlson, N.R.; Centrella, P.A.; Chiu, C.H.; Creaser, S.P.; Cuozzo, J.W.; Davie, C.P.; Ding, Y.; Franklin, G.J.; Franzen, K.D.; Gefter, M.L.; Hale, S.P.; Hansen, N.J.; Israel, D.I.; Jiang, J.; Kavarana, M.J.; Kelley, M.S.; Kollmann, C.S.; Li, F.; Lind, K.; Mataruse, S.; Medeiros, P.F.; Messer, J.A.; Myers, P.; O'Keefe, H.; Oliff, M.C.; Rise, C.E.; Satz, A.L.; Skinner, S.R.; Svendsen, J.L.; Tang, L.; van Vloten, K.; Wagner, R.W.; Yao, G.; Zhao, B.; Morgan, B.A. Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nat. Chem. Biol.* **2009**, *5* (9), 647-654.

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

88. Machutta, C. A.; Kollmann, C. S.; Lind, K. E.; Bai, X.; Chan, P. F.; Huang, J.; Ballell, L.; Belyanskaya, S.; Besra, G. S.; Barros-Aguirre, D.; Bates, R. H.; Centrella, P. A.; Chang, S. S.; Chai, J.; Choudhry, A. E.; Coffin, A.; Davie, C. P.; Deng, H.; Deng, J.; Ding, Y.; Dodson, J. W.; Fosbenner, D. T.; Gao, E. N.; Graham, T. L.; Graybill, T. L.; Ingraham, K.; Johnson, W. P.; King, B. W.; Kwiatkowski, C. R.; Lelièvre, J.; Li, Y.; Liu, X.; Lu, Q.; Lehr, R.; Mendoza-Losana, A.; Martin, J.; McCloskey, L.; McCormick, P.; O'Keefe, H. P.; O'Keeffe, T.; Pao, C.; Phelps, C. B.; Qi, H.; Rafferty, K.; Scavello, G. S.; Steiginga, M. S.; Sundersingh, F. S.; Sweitzer, S. M.; Szewczuk, L. M.; Taylor, A.; Toh, M. F.; Wang, J.; Wang, M.; Wilkins, D. J.; Xia, B.; Yao, G.; Zhang, J.; Zhou, J.; Donahue, C. P.; Messer, J. A.; Holmes, D.; Arico-Muendel, C. C.; Pope, A. J.; Gross, J. W.; Evindar, G. Prioritizing multiple therapeutic targets in parallel using automated DNA-encoded library screening. *Nat. Commun.* **2017***, 8*, No. 16081.

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

90. Lindahl, T.; Nyberg, B. Rate of depurination of native deoxyribonucleic acid. *Biochemistry*. **1972**, *11* (19), 3610–3618.

91. Lindahl, T.; Andersson, A. Rate of chain breakage at apurinic sites in double-stranded deoxyribonucleic acid. *Biochemistry*. **1972**, *11* (19), 3618–3623.

92. Frederico, L. A.; Kunkel, T. A.; Shaw, B. R. Cytosine deamination in mismatched base pairs. *Biochemistry*. **1993**, *32* (26), 6523–6530.

93. Buller, F.; Mannocci, L.; Zhang, Y.; Dumelin, C.E.; Scheuermann, J.; Neri, D. Design and synthesis of a novel DNA-encoded chemical library using Diels-Alder cycloadditions. *Bioorg. Med. Chem. Lett.* **2008**, *18* (22), 5926–5931.

94. Buller, F.; Zhang, Y.; Scheuermann, J.; Schäfer, J.; Bühlmann, P.; Neri, D. Discovery of TNF inhibitors from a DNA-encoded chemical library based on Diels-Alder cycloaddition. *Chem. Biol.* **2009**, *16* (10), 1075–1086.

95. Mannocci, L.; Zhang, Y.; Scheuermann, J.; Leimbacher, M.; De Bellis, G.; Rizzi, E.; Dumelin, C.; Melkko, S.; Neri, D. High-throughput sequencing allows the identification of binding molecules isolated from DNA-encoded chemical libraries. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (46), 17670–17675.

96. Buller, F.; Steiner, M.; Frey, K.; Mircsof, D.; Scheuermann, J.; Kalisch, M.; Buhlmann, P.; Supuran, C. T.; Neri, D. Selection of Carbonic Anhydrase IX Inhibitors from One Million DNA-Encoded Compounds. *ACS Chem. Biol.* **2011**, *6* (4), 336–344.

97. Mannocci, L.; Melkko, S.; Buller, F.; Molnàr, I.; Bianké, J. P.; Dumelin, C. E.; Scheuermann, J.; Neri, D. Isolation of potent and specific trypsin inhibitors from a DNA-encoded chemical library. *Bioconjug. Chem.* **2010**, *21* (10), 1836–1841.

98. Leimbacher, M.; Zhang, Y.; Mannocci, L.; Stravs, M.; Geppert, T.; Scheuermann, J.; Schneider, G.; Neri, D. Discovery of small molecule interleukin-2 inhibitors from a DNA-encoded chemical library. *Chemistry* **2012**, *18* (25), 7729–7737.

99. Favalli, N.; Biendl, S.; Hartmann, M.; Piazzi, J.; Sladojevich, F.; Gräslund, S.; Brown, P.J.; Näreoja, K.; Schüler, H.; Scheuermann, J.; Franzini, R.; Neri, D. A DNA-Encoded Library of Chemical Compounds Based on Common Scaffolding Structures Reveals the Impact of Ligand Geometry on Protein Recognition. *ChemMedChem*. **2018**, *13* (13), 1303–1307.

100. Gartner, Z. J.; Tse, B. N.; Grubina, R.; Doyon, J. B.; Snyder, T. M.; Liu, D. R. DNA-templated organic synthesis and selection of a library of macrocycles. *Science* **2004**, *305* (5690), 1601–1605.

101. Calderone, C. T.; Puckett, J. W., Gartner, Z. J.; Liu, D. R. Directing otherwise incompatible reactions in a single solution by using DNA-templated organic synthesis. *Angew. Chem. Int. Ed. Engl.* **2002**, *41* (21), 4104–4108.

102. Li, Y.; Zhao, P.; Zhang, M.; Zhao, X.; Li, X. Multistep DNA-templated synthesis using a universal template. *J. Am. Chem. Soc.* **2013**, *135* (47), 17727–17730.

103. Hansen, M. H.; Blakskjaer, P.; Petersen, L. K.; Hansen, T. H.; Højfeldt, J. W.; Gothelf, K. V.; Hansen, N. J. A yoctoliter-scale DNA reactor for small-molecule evolution. *J. Am. Chem. Soc.* **2009**, *131* (3), 1322–1327.

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

105. Halpin, D. R.; Harbury, P. B. DNA display II. Genetic manipulation of combinatorial chemistry libraries for small-molecule evolution. *PLoS Biol.* **2004**, *2* (7), e174.

106. Halpin, D. R.; Harbury, P. B. DNA display I. Sequence-encoded routing of DNA populations. *PLoS Biol.* **2004**, *2* (7), e173.

107. Neri, D.; Melkko, S. Encoded self-assembling chemical libraries. Patent No. WO03/076943. 2002.

108. Melkko S.; Scheuermann, J.; Dumelin, C. E.; Neri, D. Encoded self-assembling chemical libraries. *Nat. Biotechnol.* **2004**, *22* (5), 568–574.

109. Scheuermann, J.; Dumelin, C. E.; Melkko, S.; Zhang, Y.; Mannocci, L.; Jaggi, M.; Sobek, J.; Neri, D. DNA-encoded chemical libraries for the discovery of MMP-3 inhibitors. *Bioconjug. Chem.* **2008**, *19* (3), 778–785.

110. Wichert, M.; Krall, N.; Decurtins, W.; Franzini, R. M.; Pretto, F.; Schneider, P.; Neri, D.; Scheuermann, J. Dual-display of small molecules enables the discovery of ligand pairs and facilitates affinity maturation. *Nat. Chem.* **2015**, *7* (3), 241–249.

111. Melkko, S.; Dumelin, C. E.; Scheuermann, J.; Neri, D. On the magnitude of the chelate effect for the recognition of proteins by pharmacophores scaffolded by self-assembling oligonucleotides. *Chem. Biol.* **2006**, *13* (2), 225–231.

112. Bigatti, M.; Dal Corso, A.; Vanetti, S.; Cazzamalli, S.; Rieder, U.; Scheuermann, J.; Neri, D.; Sladojevich, F. Impact of a Central Scaffold on the Binding Affinity of Fragment Pairs Isolated from DNA-Encoded Self-Assembling Chemical Libraries. *ChemMedChem* **2017**, *12* (21), 1748–1752.

113. Zimmermann, G.; Li, Y.; Rieder, U.; Mattarella, M.; Neri, D.; Scheuermann, J. Hit-Validation Methodologies for Ligands Isolated from DNA-Encoded Chemical Libraries. *ChemBioChem* **2017**, *18* (9), 853–857.

114. Prati, L.; Bigatti, M.; Donckele, E. J.; Neri, D.; Samain, F. On-DNA hit validation methodologies for ligands identified from DNA-encoded chemical libraries. *Biochem. Biophys. Res. Commun.* **2020**, DOI: 10.1016/j.bbrc.2020.04.030.

115. Favalli, N.; Bassi, G.; Scheuermann, J.; Neri, D. DNA-encoded chemical libraries - achievements and remaining challenges. *FEBS Lett.* **2018**, *592* (12), 2168–2180.

116. Zhao, G.; Huang, Y.; Zhou, Y.; Li, Y.; Li, X. Future challenges with DNA-encoded chemical libraries in the drug discovery domain. *Expert Opin. Drug. Discov.* **2019**, *14* (8), 735–753.

117. Reddavide, F. V.; Lin, W.; Lehnert, S.; Zhang, Y. DNA-Encoded Dynamic Combinatorial Chemical Libraries. *Angew. Chem. Int. Ed. Engl.* **2015**, *54* (27), 7924–7928.

118. Bandyopadhyay, A.; Gao, J. Targeting biomolecules with reversible covalent chemistry. *Curr. Opin. Chem. Biol.* **2016**, *34*, 110–116.

119. Zhou, Y.; Li, C.; Peng, J.; Xie, L.; Meng, L.; Li, Q.; Zhang, J.; Li, X. D.; Li, X.; Huang, X.; Li, X. DNA-Encoded Dynamic Chemical Library and Its Applications in Ligand Discovery. *J. Am. Chem. Soc.* **2018**, *140* (46), 15859–15867.

120. Reddavide, F. V.; Cui, M.; Lin, W.; Fu, N.; Heiden, S.; Andrade, H.; Thompson, M.; Zhang, Y. Second generation DNA-encoded dynamic combinatorial chemical libraries. *Chem. Commun.* **2019**, *55* (26), 3753–3756.

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

122. Gorska, K.; Huang, K.T.; Chaloin, O.; Winssinger, N. DNA-templated homo- and heterodimerization of peptide nucleic acid encoded oligosaccharides that mimic the carbohydrate epitope of HIV. *Angew. Chem. Int. Ed. Engl.* **2009**, *48* (41), 7695–7700.

123. Winssinger, N. Nucleic acid-programmed assemblies: translating instruction into function in chemical biology. *Chimia* **2013**, *67* (5), 340–348.

124. Pellestor, F.; Paulasova, P. The peptide nucleic acids (PNAs), powerful tools for molecular genetics and cytogenetics. *Eur. J. Hum. Genet.* **2004**, *12* (9), 694–700.

125. Samain, F.; Ekblad, T.; Mikutis, G.; Zhong, N.; Zimmermann, M.; Nauer, A.; Bajic, D.; Decurtins, W.; Scheuermann, J.; Brown, P. J.; Hall, J.; Gräslund, S.; Schuler, H.; Neri, D.; Franzini, R. M. Tankyrase I Inhibitors with Drug-like Properties Identified by Screening a DNA-Encoded Chemical Library. *J. Med. Chem.* **2015**, *58* (12), 5143–5149.

126. Wu, Z.; Graybill, T. L.; Zeng, X.; Platchek, M.; Zhang, J.; Bodmer, V. Q.; Wisnoski, D. D.; Deng, J.; Coppo, F. T.; Yao, G.; Tamburino, A.; Scavello, G.; Franklin, G. J.; Mataruse, S.; Bedard, K. L.; Ding, Y.; Chai, J.; Summerfield, J.; Centrella, P. A.; Messer, J. A.; Pope, A. J.; Israel, D. I. 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.* **2015**, *17* (12), 722-731.

127. Ding, Y.; O'Keefe, H.; DeLorey, J. L.; Israel, D. I.; Messer, J. A.; Chiu, C. H.; Skinner, S. R.; Matico, R. E.; Murray-Thompson, M. F.; Li, F.; Clark, M. A.; Cuozzo, J. W.; Arico-Muendel, C.; Morgan, B. A. Discovery of Potent and Selective Inhibitors for ADAMTS-4 through DNA-Encoded Library Technology (ELT). *ACS Med. Chem. Lett.* **2015**, *6* (8), 888–893.

128. Dougherty, P. G.; Qian, Z.; Pei, D. Macrocycles as protein-protein interaction inhibitors. *Biochem. J.* **2017**, *474* (7), 1109–1125.

129. Gao, M.; Cheng, K.; Yin, H. Targeting Protein-Protein Interfaces Using Macrocyclic Peptides. *Biopolymers* **2015**, *104* (4), 310–316.

130. Usanov, D. L.; Chan, A. L.; Maianti, J. P.; Liu, D. R. Second-generation DNA-templated macrocycle libraries for the discovery of bioactive small molecules. *Nat. Chem.* **2018**, *10* (7), 704–714.

131. Seigal, B. A.; Connors, W. H.; Fraley, A.; Borzilleri, R. M.; Carter, P. H.; Emanuel, S. L.; Fargnoli, J.; Kim, K.; Lei, M.; Naglich, J. G.; Pokross, M. E.; Posy, S. L.; Shen, H.; Surti, N.; Talbott, R.; Zhang, Y.; Terrett, N. K. 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.* **2015**, *58* (6), 2855–2861.

132. Li, Y.; De Luca, R.; Cazzamalli, S.; Pretto, F.; Bajic, D.; Scheuermann, J.; Neri, D. Versatile protein recognition by the encoded display of multiple chemical elements on a constant macrocyclic scaffold. *Nat. Chem.* **2018**, *10* (4), 441–448.

133. Petersen, L. K.; Blakskjaer, P.; Chaikuad, A.; Christensen, A. B.; Dietvorst, J.; Holmkvist, J.; Knapp, S.; Kořínek, M.; Larsen, L. K.; Pedersen, A. E.; Röhm, S.; Sløk, F. A.; Hansen, N. J. V. Novel p38α MAP kinase inhibitors identified from yoctoReactor DNA-encoded small molecule library. *MedChemComm* **2016**, *7*, 1332–1339.

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

135. Arnau, J.; Lauritzen, C.; Petersen, G. E.; Pedersen, J. Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expr. Purif.* **2006**, *48* (1), 1–13.

136. Kimple, M. E.; Brill, A. L.; Pasker, R. L. Overview of affinity tags for protein purification. *Curr. Protoc. Protein. Sci.* **2013**, *73*, 9.9.1–9.9.23.

137. Chapman-Smith, A.; Cronan Jr., J. E. The enzymatic biotinylation of proteins: a post-translational modification of exceptional specificity. *Trends Biochem. Sci.* **1999**, *24* (9), 359–363.

138. Cull, M. G.; Schatz, P. J. Biotinylation of proteins in vivo and in vitro using small peptide tags. *Methods Enzymol.* **2000**, *326*, 430–440.

139. Spriestersbach, A.; Kubicek, J.; Schäfer, F.; Block, H.; Maertens, B. Purification of His-Tagged Proteins. *Methods Enzymol.* **2015**, *559*, 1–15.

140. Block, H.; Maertens, B.; Spriestersbach, A.; Brinker, N.; Kubicek, J.; Fabis, R.; Labahn, J.; Schäfer, F. Immobilized-metal affinity chromatography (IMAC): a review. *Methods Enzymol.* **2009**, *463*, 439–473.

141. Wong, L. S.; Khan, F.; Micklefield, J. Selective covalent protein immobilization: strategies and applications. *Chem. Rev.* **2009**, *109* (9), 4025–4053.

142. Rusmini, F.; Zhong, Z.; Feijen, J. Protein immobilization strategies for protein biochips. *Biomacromolecules* **2007**, *8* (6), 1775–1789.

143. Dumelin, C. E.; Scheuermann, J.; Melkko, S.; Neri. D. Selection of streptavidin binders from a DNA-encoded chemical library. *Bioconjug. Chem.* **2006**, *17* (2), 366–370.

144. Decurtins, W.; Wichert, M.; Franzini, R. M.; Buller, F.; Stravs, M. A.; Zhang, Y.; Neri, D.; Scheuermann, J. Automated screening for small organic ligands using DNA-encoded chemical libraries. *Nat. Protoc.* **2016**, *11* (4), 764–780.

145. Neurauter, A. A.; Bonyhadi, M.; Lien, E.; Nøkleby, L.; Ruud, E.; Camacho, S.; Aarvak, T. Cell isolation and expansion using Dynabeads. *Adv. Biochem. Eng. Biotechnol.* **2007**, *106*, 41–73.

146. Gjerde, D. T.; Bonn, G. Dual Flow Chromatography. *LCGC Europe* **2017**, *30* (8), 404–411.

147. McGregor, L. M.; Gorin, D. J.; Dumelin, C. E.; Liu, D. R. Interaction-dependent PCR: identification of ligand-target pairs from libraries of ligands and libraries of targets in a single solution-phase experiment. *J. Am. Chem. Soc.* **2010**, *132* (44), 15522–15524.

148. McGregor, L. M.; Jain, T.; Liu, D. R. Identification of ligand-target pairs from combined libraries of small molecules and unpurified protein targets in cell lysates. *J. Am. Chem. Soc.* **2014**, *136* (8), 3264–3270.

149. Chan, A. I.; McGregor, L. M.; Jain, T.; Liu, D. R. Discovery of a Covalent Kinase Inhibitor from a DNA-Encoded Small-Molecule Library × Protein Library Selection. *J. Am. Chem. Soc.* **2017**, *139* (30), 10192–10195.

150. Zhao, P.; Chen, Z.; Li, Y.; Sun, D.; Gao, Y.; Huang, Y.; Li, X. Selection of DNA-encoded small molecule libraries against unmodified and non-immobilized protein targets. *Angew. Chem. Int. Ed. Engl.* **2014**, *53* (38), 10056–10059.

151. Li, G.; Liu, Y.; Yu, X.; Li, X. Multivalent Photoaffinity Probe for Labeling Small Molecule Binding Protein. *Bioconjug. Chem.* **2014**, *25* (6), 1172–1180.

152. Li, G.; Liu, Y.; Liu, Y.; Chen, L.; Wu, S.; Liu, Y.; Li, X. Photoaffinity labeling of small-molecule-binding proteins by DNA-templated chemistry. *Angew. Chem. Int. Ed. Engl.* **2013**, *52* (36), 9544–9549.

153. Shi, B.; Deng, Y.; Zhao, P.; Li, X. Selecting a DNA-Encoded Chemical Library against Non-immobilized Proteins Using a "Ligate–Cross-Link–Purify" Strategy. *Bioconjug. Chem.* **2017**, *28* (9), 2293–2301.

154. Shi, B.; Deng, Y.; Li, X. Polymerase-Extension-Based Selection Method for DNA-Encoded Chemical Libraries against Nonimmobilized Protein Targets. *ACS Comb. Sci.* **2019**, *21* (5), 345–349.

155. Tawfik, D. S.; Griffiths, A. D. Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.* **1998**, *16* (7), 652–656.

156. Baell, J. B.; Nissink, J. W. M. Seven Year Itch: Pan-Assay Interference Compounds (PAINS) in 2017–Utility and Limitations. *ACS Chem. Biol.* **2018**, *13* (1), 36–44.

157. Jasial, S.; Hu, Y.; Bajorath, J. How Frequently Are Pan-Assay Interference Compounds Active? Large-Scale Analysis of Screening Data Reveals Diverse Activity Profiles, Low Global Hit Frequency, and Many Consistently Inactive Compounds. *J. Med. Chem.* **2017***, 60* (9), 3879–3886.

158. Napper, A. D. Discriminating High-Value Hits from PAINS. *Assay Drug Dev. Technol.* **2016**, *14* (3), 157.

159. Dahlin, J. L.; Walters, M. A. How to Triage PAINS-Full Research. *Assay Drug Dev. Technol.* **2016**, *14* (3), 168–174.

160. Kodakek, T.; Paciaroni, N. G.; Balzarini, M.; Dickson, P. Beyond protein binding: recent advances in screening DNA-encoded libraries. *Chem. Commun.* **2019**, *55* (89), 13330–13341.

161. Mendes, K. R.; Malone, M. L.; Ndungu, J. M.; Suponitsky-Kroyter, I.; Cavett, V. J.; McEnaney, P. J.; MacConnell, A. B.; Doran, T. M.; Ronacher, K.; Stanley, K.; Utset, O.; Walzl, G.; Paegel, B. M.; Kodadek, T. High-throughput Identification of DNA-Encoded IgG Ligands that Distinguish Active and Latent Mycobacterium tuberculosis Infections. *ACS Chem. Biol.* **2017**, *12* (1), 234–243.

162. Wu, Z.; Graybill, T. L.; Zeng, X.; Platchek, M.; Zhang, J.; Bodmer, V. Q.; Wisnoski, D. D.; Deng, J., Coppo, F. T.; Yao, G.; Tamburino, A.; Scavello, G.; Franklin, G. J.; Mataruse, S.; Bedard, K. L.; Ding, Y.; Chai, J.; Summerfield, J.; Centrella, P. A.; Messer, J. A.; Pope, A. J.; Israel, D. I. 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.* **2015**, *17* (12), 722–731.

163. Svensen, N.; Díaz-Mochón, J. J.; Bradley, M. Decoding a PNA encoded peptide library by PCR: the discovery of new cell surface receptor ligands. *Chem. Biol.* **2011**, *18* (10), 1284–1289.

164. Svensen, N.; Díaz-Mochón, J. J.; Bradley, M. Encoded peptide libraries and the discovery of new cell binding ligands. *Chem. Commun.* **2011**, *47* (27), 7638–7640.

165. Cai, B.; Kim, D.; Akhand, S.; Sun, Y.; Cassell, R. J.; Alpsoy, A.; Dykhuizen, E. C.; Van Rijn, R. M.; Wendt, M. K.; Krusemark, C. J. Selection of DNA-Encoded Libraries to Protein Targets within and on Living Cells. *J. Am. Chem. Soc.* **2019**, *141* (43), 17057–17061.

166. Qian, Z.; Martyna, A.; Hard, R. L.; Wang, J.; Appiah-Kubi, G.; Coss, C.; Phelps, M. A.; Rossman, J. S.; Pei, D. Discovery and Mechanism of Highly Efficient Cyclic Cell-Penetrating Peptides. *Biochemistry* **2016**, *55* (18), 2601–2612.

167. Los, G. V.; Encell, L. P.; McDougall, M. G.; Hartzell, D. D.; Karassina, N.; Zimprich, C.; Wood, M. G.; Learish, R.; Ohana, R. F.; Urh, M.; Simpson, D.; Mendez, J.; Zimmerman, K.; Otto, P.; Vidugiris, G.; Zhu, J.; Darzins, A.; Klaubert, D. H.; Bulleit, R. F.; Wood, K. V. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **2008**, *3* (6), 373–382.

168. Peraro, L.; Deprey, K. L.; Moser, M. K.; Zou, Z.; Ball, H. L.; Levine, B.; Kritzer, J. A. Cell Penetration Profiling Using the Chloroalkane Penetration Assay. *J. Am. Chem. Soc.* **2018**, *140* (36), 11360–11369.

169. Cochrane, W. G.; Malone, M. L.; Dang, V. Q.; Cavett, V.; Satz, A. L.; Paegel, B. M. Activity-Based DNA-Encoded Library Screening. *ACS Comb. Sci.* **2019**, *21* (5), 425–435.

170. MacConnell, A. B.; Price, A. K.; Paegel, B. M. An Integrated Microfluidic Processor for DNA-Encoded Combinatorial Library Functional Screening. *ACS Comb. Sci.* **2017**, *19* (3), 181–192.

171. Sanger, F.; Nicklen, S.; Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74* (12), 5463–5467.

172. Canard, B.; Sarfati, R. S. DNA polymerase fluorescent substrates with reversible 3' tags. *Gene* **1994**, *148* (1), 1–6.

173. Buller, F.; Mannocci, L.; Scheuermann, J.; Neri, D. Drug discovery with DNA-encoded chemical libraries. *Bioconjug. Chem.* **2010**, *21* (9), 1571–1580.

174. Buller, F.; Steiner, M.; Scheuermann, J.; Mannocci, L.; Nissen, L. Kohler, M.; Beisel, C.; Neri, D. High-throughput sequencing for the identification of binding molecules from DNA-encoded chemical libraries. *Bioorg. Med. Chem. Lett.* **2010**, *20* (14), 4188–4192.

175. Satz, A. L. DNA Encoded Library Selections and Insights Provided by Computational Simulations. *ACS Chem. Biol.* **2015**, *10* (10), 2237–2245.

176. Satz, A. L. Simulated Screens of DNA Encoded Libraries: The Potential Influence of Chemical Synthesis Fidelity on Interpretation of Structure–Activity Relationships. *ACS Comb. Sci.* **2016**, *18* (7), 415–424.

177. Faver, J. C.; Riehle, K.; Lancia, D. R. Jr.; Milbank, J. B. J.; Kollmann, C. S.; Simmons, N.; Yu, Z.; Matzuk, M. M. Quantitative Comparison of Enrichment from DNA-Encoded Chemical Library Selections. *ACS Comb. Sci.* **2019**, *21* (2), 75–82.

178. Sannino, A.; Gabriele, E.; Bigatti, M.; Mulatto, S.; Piazzi, J.; Scheuermann, J.; Neri, D.; Donckele, E. J.; Samain, F. Quantitative Assessment of Affinity Selection Performance by Using DNA-Encoded Chemical Libraries. *ChemBioChem* **2019**, *20* (7), 955–962.

179. Kleiner, R. E.; Dumelin, C. E.; Liu, D. R. Small-molecule discovery from DNA-encoded chemical libraries. *Chem. Soc. Rev.* **2011**, *40* (12), 5707–5717.

180. Franzini, R. M.; Neri, D.; Scheuermann, J. DNA-Encoded Chemical Libraries: Advancing beyond Conventional Small-Molecule Libraries. *Acc. Chem. Res.* **2014**, *47* (4), 1247–1255.

181. Mutuberria, R.; Hoogenboom, H. R.; van der Linden, E.; de Bruïne, A. P.; Roovers, R. C. Model systems to study the parameters determining the success of phage antibody selections on complex antigens. *J. Immunol. Methods* **1999**, *231* (1–2), 65–81.

182. Fischer, N. Sequencing antibody repertoires: The next generation. *mAbs* **2011**, *3* (1), 17–20.

183. 't Hoen, P. A.; Jirka, S. M.; Ten Broeke, B. R.; Schultes, E. A.; Aguilera, B.; Pang, K. H.; Heemskerk, H.; Aartsma-Rus, A.; van Ommen, G. J.; den Dunnen, J. T. Phage display screening without repetitious selection rounds. *Anal. Biochem.* **2012**, *421* (2), 622–631.

184. Ngubane, N. A.; Gresh, L.; Ioerger, T. R.; Sacchettini, J. C.; Zhang, Y. J.; Rubin, E. J.; Pym, A.; Khati, M. High-throughput sequencing enhanced phage display identifies peptides that bind mycobacteria. *PLoS One.* **2013**, *8* (11), e77844.

185. Christiansen, A.; Kringelum, J. V.; Hansen, C. S.; Bøgh, K. L.; Sullivan, E.; Patel, J.; Rigby, N. M.; Eiwegger, T.; Szépfalusi, Z.; de Masi, F.; Nielsen, M.; Lund, O.; Dufva, M. High-throughput sequencing enhanced phage display enables the identification of patient-specific epitope motifs in serum. *Sci Rep.* **2015**, *5*, No. 12913.

186. Hoogenboom, H. R.; Winter, G. By-passing immunisation. Human antibodies from synthetic repertoires of germline VH gene segments rearranged in vitro. *J. Mol. Biol.* **1992**, *227* (2), 381–388.

187. Marks J. D.; Hoogenboom, H. R.; Bonnert, T. P.; McCafferty, J.; Griffiths, A. D.; Winter, G. By-passing immunization. Human antibodies from V-gene libraries displayed on phage. *J. Mol. Biol.* **1991**, *222* (3), 581–597.

188. Griffiths, A. D.; Williams, S. C.; Hartley, O.; Tomlinson, I. M.; Waterhouse, P.; Crosby, W. L.; Kontermann, R. E.; Jones, P. T.; Low, N. M.; Allison, T. J. Isolation of high affinity human antibodies directly from large synthetic repertoires. *EMBO J.* **1994**, *13* (14), 3245–3260.

189. Belyanskaya, S. L.; Ding, Y.; Callahan, J. F.; Lazaar, A. L.; Israel, D. I. Discovering Drugs with DNA-Encoded Library Technology: From Concept to Clinic with an Inhibitor of Soluble Epoxide Hydrolase. *ChemBioChem* **2017**, *18* (9), 837–842.

190. Li, Y.; Zimmermann, G.; Scheuermann, J.; Neri, D. Quantitative PCR is a Valuable Tool to Monitor the Performance of DNA-Encoded Chemical Library Selections. *ChemBioChem* **2017**, *18* (9), 848–852.

191. Li, G.; Zheng, W.; Chen, Z.; Zhou, Y.; Liu, Y.; Yang, J.; Huang, Y.; Li, X. Design, preparation, and selection of DNA-encoded dynamic libraries. *Chem. Sci.* **2015**, *6* (12), 7097–7104.

192. Denton, K. E.; Krusemark, C. J. Crosslinking of DNA-linked ligands to target proteins for enrichment from DNA-encoded libraries. *MedChemComm* **2016**, *7* (10), 2020–2027.

193. Kuai, L.; O'Keeffe, T.; Arico-Muendel, C. Randomness in DNA Encoded Library Selection Data Can Be Modeled for More Reliable Enrichment Calculation. *SLAS Discov.* **2018**, *23* (5), 405–416.

194. Denton, K. E.; Wang, S.; Gignac, M. C.; Milosevich, N.; Hof, F.; Dykhuizen, E. C.; Krusemark, C. J. Robustness of In Vitro Selection Assays of DNA-Encoded Peptidomimetic Ligands to CBX7 and CBX8. *SLAS Discov.* **2018**, *23* (5), 417–428.

195. McDonald, P. C.; Winum, J-Y.; Supuran, C. T.; Dedhar, S. Recent developments in targeting carbonic anhydrase IX for cancer therapeutics. *Oncotarget* **2012**, *3* (1), 84–97.

196. Pichake, J.; Kharkar, P. S.; Ceruso, M.; Supuran, C. T.; Toraskar, M. P. Carbonic Anhydrase Inhibitors: Design, Synthesis and Biological Evaluation of Novel Sulfonyl Semicarbazide Derivatives. *ACS Med. Chem. Lett.* **2014**, *5* (7), 793–796.

197. Supuran, C. T. Carbonic anhydrases: novel therapeutic applications for inhibitors and activators. *Nat. Rev. Drug. Discov.* **2008**, *7* (2), 168–181.

198. Krall, N.; Pretto, F.; Decurtins, W.; Bernardes, G. J.; Supuran, C. T.; Neri, D. A small-molecule drug conjugate for the treatment of carbonic anhydrase IX expressing tumors. *Angew. Chem. Int. Ed. Engl.* **2014**, *53* (16), 4231–4235.

199. Cazzamalli, S.; Dal Corso, A.; Neri, D. Acetazolamide Serves as Selective Delivery Vehicle for Dipeptide-Linked Drugs to Renal Cell Carcinoma. *Mol. Cancer Ther.* **2016**, *15* (12), 2926–2935.

200. Cazzamalli, S.; Dal Corso, A.; Neri, D. Linker stability influences the anti-tumor activity of acetazolamide-drug conjugates for the therapy of renal cell carcinoma. *J. Control. Release.* **2017**, *246*, 39–45.

201. Dubois, L.; Peeters, S.; Lieuwes, N. G.; Geusens, N.; Thiry, A.; Wigfield, S.; Carta, F.; McIntyre, A.; Scozzafava, A.; Dogné, J. M.; Supuran, C. T.; Harris, A. L.; Masereel, B.; Lambin, P. Specific inhibition of carbonic anhydrase IX activity enhances the in vivo therapeutic effect of tumor irradiation. *Radiother. Oncol.* **2011**, *99* (3), 424–431.

202. Wichert, M.; Krall, N. Targeting carbonic anhydrase IX with small organic ligands. *Curr. Opin. Chem. Biol.* **2015**, *26*, 48–54.

203. Monti, S. M.; Supuran, C. T.; De Simone, G. Anticancer carbonic anhydrase inhibitors: a patent review (2008 - 2013). *Expert Opin. Ther. Pat.* **2013**, *23* (6), 737–749.

204. Supuran, C. T.; Winum, J-Y. Carbonic anhydrase IX inhibitors in cancer therapy: an update. *Future Med. Chem.* **2015**, *7* (11), 1407–1414.

205. Rutkauskas, K.; Zubrienė, A.; Tumosienė, I.; Kantminienė, K.; Kažemėkaitė, M.; Smirnov, A.; Kazokaitė, J.; Morkūnaitė, V.; Čapkauskaitė, E.; Manakova, E.; Gražulis, S.; Beresnevičius, Z. J.; Matulis, D. 4-amino-substituted benzenesulfonamides as inhibitors of human carbonic anhydrases. *Molecules* **2014**, *19* (11), 17356-17380.

206. Kollmann, C. S.; Bai, X.; Tsai, C. H.; Yang, H.; Lind, K. E.; Skinner, S. R.; Zhu, Z.; Israel, D. I.; Cuozzo, J. W.; Morgan, B. A.; Yuki, K.; Xie, C.; Springer, T. A.; Shimaoka, M.; Evindar, G. Application of encoded library technology (ELT) to a protein-protein interaction target: discovery of a potent class of integrin lymphocyte function-associated antigen 1 (LFA-1) antagonists. *Bioorg. Med. Chem.* **2014**, *22* (7), 2353–2365.

207. Zhu, Z.; Shaginian, A.; Grady, L. C.; O'Keeffe, T.; Shi, X. E.; Davie, C. P.; Simpson, G. L.; Messer, J. A.; Evindar, G.; Bream, R. N.; Thansandote, P. P.; Prentice, N. R.; Mason, A. M.; Pal, S. Design

and Application of a DNA-Encoded Macrocyclic Peptide Library. *ACS Chem. Biol.* **2018**, *13* (1), 53–59.

208. Bao, J.; Krylova, S. M.; Cherney, L. T.; Hale, R. L.; Belyanskaya, S. L.; Chiu, C. H.; Arico-Muendel, C. C.; Krylov, S. N. Prediction of protein-DNA complex mobility in gel-free capillary electrophoresis. *Anal Chem.* **2015**, *87* (4), 2474–2479.

209. Bao, J.; Krylova, S. M.; Cherney, L. T.; Hale, R. L.; Belyanskaya, S. L.; Chiu, C. H.; Shaginian, A.; Arico-Muendel, C. C.; Krylov, S. N. Predicting Electrophoretic Mobility of Protein-Ligand Complexes for Ligands from DNA-Encoded Libraries of Small Molecules. *Anal. Chem.* **2016**, *88* (10), 5498–5506.

210. Erlanson, D. A.; McDowell, R. S.; He, M. M.; Randal, M.; Simmons, R. L.; Kung, J.; Waight, A.; Hansen, S. K. Discovery of a new phosphotyrosine mimetic for PTP1B using breakaway tethering. *J. Am. Chem. Soc.* **2003**, *125* (19), 5602–5603.

211. Hyde, J.; Braisted, A. C.; Randal, M.; Arkin, M. R. Discovery and characterization of cooperative ligand binding in the adaptive region of interleukin-2. *Biochemistry* **2003**, *42* (21), 6475–6483.

212. Erlanson, D. A.; Wells, J. A.; Braisted, A. C. Tethering: fragment-based drug discovery. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 199–223.

213. Yang, W.; Fucini, R. V.; Fahr, B. T.; Randal, M.; Lind, K. E.; Lam, M. B.; Lu, W.; Lu, Y.; Cary, D. R.; Romanowski, M. J.; Colussi, D.; Pietrak, B.; Allison, T. J.; Munshi, S. K.; Penny, D. M.; Pham, P.; Sun, J.; Thomas, A. E.; Wilkinson, J. M.; Jacobs, J. W.; McDowell, R. S.; Ballinger, M. D. Fragment-based discovery of nonpeptidic BACE-1 inhibitors using tethering. *Biochemistry* **2009**, *48* (21), 4488–4496.

214. Erlanson, D. A.; Arndt, J. W.; Cancilla, M. T.; Cao, K.; Elling, R. A.; English, N.; Friedman, J.; Hansen, S. K.; Hession, C.; Joseph, I.; Kumaravel, G.; Lee, W. C.; Lind, K. E.; McDowell, R. S.; Miatkowski, K.; Nguyen, C.; Nguyen, T. B.; Park, S.; Pathan, N.; Penny, D. M.; Romanowski, M. J.; Scott, D.; Silvian, L.; Simmons, R. L.; Tangonan, B. T.; Yang, W.; Sun, L. Discovery of a potent and highly selective PDK1 inhibitor via fragment-based drug discovery. *Bioorg. Med. Chem. Lett.* **2011**, *21* (10), 3078–3083.

215. Dal Corso, A.; Catalano, M.; Schmid, A.; Scheuermann, J.; Neri, D. Affinity Enhancement of Protein Ligands by Reversible Covalent Modification of Neighboring Lysine Residues. *Angew. Chem. Int. Ed. Engl.* **2018**, *57* (52), 17178–17182.

216. Sannino, A.; Gironda-Martínez, A.; Gorre, É. D.; Prati, L.; Piazzi, J.; Scheuermann, J.; Neri, D.; Donckele, E. J.; Samain, F. Critical Evaluation of Photo-cross-linking Parameters for Efficient DNA-Encoded Chemical Library Selections. *ACS Comb. Sci.* **2020**, *22* (4), 204−212.

217. Singh, A.; Thornton, E. R.; Westheimer, F. H. The photolysis of diazoacetylchymotrypsin. *J. Biol. Chem.* **1962**, *237*, 3006–3008.

218. Berggård, T.; Linse, S.; James, P. Methods for the detection and analysis of protein-protein interactions. *Proteomics* **2007**, *7* (16), 2833–2842.

219. Petrotchenko, E. V.; Borchers, C. H. Crosslinking combined with mass spectrometry for structural proteomics. *Mass Spectrom. Rev.* **2010**, *29* (6), 862–876.

220. Kuo, M. H.; Allis, C. D. In vivo cross-linking and immunoprecipitation for studying dynamic Protein: DNA associations in a chromatin environment. *Methods* **1999**, *19* (3), 425–433.

221. Qin, H.; Wang, Y. Exploring DNA-binding proteins with in vivo chemical cross-linking and mass spectrometry. *J. Proteome Res.* **2009**, *8* (4), 1983–1991.

222. Sumranjit, J.; Chung, S. J. Recent advances in target characterization and identification by photoaffinity probes. *Molecules* **2013**, *18* (9), 10425–10451.

223. Prestwich, G. D.; Dormán, G.; Elliott, J. T.; Marecak, D. M.; Chaudhary, A. Benzophenone photoprobes for phosphoinositides, peptides and drugs. *Photochem. Photobiol.* **1997**, *65* (2), 222–234.

224. Platz, M. S. Comparison of Phenylcarbene and Phenylnitrene. *Acc. Chem. Res.* **1995**, *28* (12), 487–492.

225. Gritsan, N. P.; Gudmundsdóttir, A. D.; Tigelaar, D.; Zhu, Z.; Karney, W. L.; Hadad, C. M.; Platz, M. S. A laser flash photolysis and quantum chemical study of the fluorinated derivatives of singlet phenylnitrene. *J. Am. Chem. Soc.* **2001**, *123* (9), 1951–1962.

226. Kambe, T.; Correia, B. E.; Niphakis, M. J.; Cravatt, B. F. Mapping the protein interaction landscape for fully functionalized small-molecule probes in human cells. *J. Am. Chem. Soc.* **2014**, *136* (30), 10777–10782.

227. Wang, D. Y.; Cao, Y.; Zheng, L. Y.; Chen, L. D.; Chen, X. F.; Hong, Z. Y.; Zhu, Z. Y.; Li, X.; Chai, Y. F. Target Identification of Kinase Inhibitor Alisertib (MLN8237) by Using DNA-Programmed Affinity Labeling. *Chemistry* **2017**, *23* (45), 10906–10914.

228. Dubinsky, L.; Krom, B. P.; Meijler, M. M. Diazirine based photoaffinity labeling. *Bioorg. Med. Chem.* **2012**, *20* (2), 554–570.

229. Horne, J. E.; Walko, M.; Calabrese, A. N.; Levenstein, M. A.; Brockwell, D. J.; Kapur, N.; Wilson, A. J.; Radford, S. E. Rapid Mapping of Protein Interactions Using Tag-Transfer Photocrosslinkers. *Angew. Chem. Int. Ed. Engl.* **2018**, *57* (51), 16688–16692.

230. Hatanaka, Y.; Sadakane, Y. Photoaffinity labeling in drug discovery and developments: chemical gateway for entering proteomic frontier. *Curr. Top. Med. Chem.* **2002**, *2* (3), 271–288.

231. Wang, J.; Kubicki, J.; Peng, H.; Platz, M. S. Influence of solvent on carbene intersystem crossing rates. *J. Am. Chem. Soc.* **2008**, *130* (20), 6604–6609.

232. Preston, G. W.; Wilson, A. J. Photo-induced covalent cross-linking for the analysis of biomolecular interactions. *Chem. Soc. Rev.* **2013**, *42* (8), 3289–3301.

233. Galardy, R. E.; Craig, L. C.; Jamieson, J. D.; Printz, M. P. Photoaffinity labelling of peptide hormone binding sites. *J. Biol. Chem.* **1974**, *249* (11), 3510–3518.

234. Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of Escherichiacoli. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (17), 11020–11024.

235. Chin, J. W.; Schultz, P. G. In Vivo Photocrosslinking with Unnatural Amino Acid Mutagenesis. *ChemBioChem* **2002**, *3* (11), 1135–1137.

236. Murale, D. P.; Hong, S. C.; Haque, M. M.; Lee, J. S. Photo-affinity labelling (PAL) in chemical proteomics: a handy tool to investigate protein-protein interactions (PPIs). *Proteome Sci.* **2016**, *15*, No: 14.

237. Lam, K. S.; Lebl, M.; Krchnák, V. The "One-Bead-One-Compound" Combinatorial Library Method. *Chem. Rev.* **1997**, *97* (2), 411-448.

238. Wu, C. Y.; Wang, D. H.; Wang, X.; Dixon, S. M.; Meng, L.; Ahadi, S.; Enter, D. H.; Chen, C. Y.; Kato, J.; Leon, L. J.; Ramirez, L. M.; Maeda, Y.; Reis, C. F.; Ribeiro, B.; Weems, B.; Kung, H. J.; Lam, K. S. Rapid Discovery of Functional Small Molecule Ligands against Proteomic Targets through Library-Against-Library Screening. *ACS Comb. Sci.* **2016**, *18* (6), 320–329.

239. Teng, P.; Zhang, X.; Wu, H.; Qiao, Q.; Sebti, S. M.; Cai, J. Identification of novel inhibitors that disrupt STAT3-DNA interaction from a γ-AApeptide OBOC combinatorial library. *Chem. Commun.* **2014**, *50* (63), 8739-8742.

240. Lebl, M.; K Krchňák, V.; Salmon, S. E.; Lam, K. S. Screening of Completely Random One-bead One-Peptide Libraries for Activities in Solution. *Methods* **1994**, *6* (4), 381–387.

241. Pei, D.; Appiah Kubi, G. Developments with bead-based screening for novel drug discovery. *Expert Opin. Drug. Discov.* **2019**, *14* (11), 1097–1102.

242. Vinogradov, A.A.; Gates, Z. P.; Zhang, C.; Quartararo, A. J.; Halloran, K. H.; Pentelute, B. L. Library Design-Facilitated High-Throughput Sequencing of Synthetic Peptide Libraries. *ACS Comb. Sci.* **2017**, *19* (11), 694–701.

243. Martínez-Ceron, M. C.; Giudicessi, S. L.; Saavedra, S. L.; Gurevich-Messina, J. M.; Erra-Balsells, R.; Albericio, F.; Cascone, O.; Camperi, S. A. Latest Advances in OBOC Peptide Libraries. Improvements in Screening Strategies and Enlarging the Family From Linear to Cyclic Libraries. *Curr. Pharm. Biotechnol.* **2016**, *17* (5), 449–457.

244. Prior, A. M.; Hori, T.; Fishman, A.; Sun, D. Recent Reports of Solid-Phase Cyclohexapeptide Synthesis and Applications. *Molecules* **2018**, *23* (6), No. 1475.

245. Itoh, H.; Inoue, M. Full solid-phase total synthesis of macrocyclic natural peptides using four-dimensionally orthogonal protective groups. *Org. Biomol. Chem.* **2019**, *17* (27), 6519–6527.

246. Madsen, D.; Azevedo, C.; Micco, I.; Petersen, L. K.; Hansen, N. J. V. An overview of DNA-encoded libraries: A versatile tool for drug discovery. *Prog. Med. Chem.* **2020**, *59*, 181–249.

247. MacConnell, A. B.; McEnaney, P. J.; Cavett, V. J.; Paegel, B. M. DNA-Encoded Solid-Phase Synthesis: Encoding Language Design and Complex Oligomer Library Synthesis. *ACS Comb. Sci.* **2015**, *17* (9), 518–534.

248. Paciaroni, N. G.; Ndungu, J. M.; Kodadek, T. Solid-phase synthesis of DNA-encoded libraries via an "aldehyde explosion" strategy. *Chem. Commun.* **2020**, *56* (34), 4656–4659.

249. Potowski, M.; Kunig, V. B. K.; Losch, F.; Brunschweiger, A. Synthesis of DNA-coupled isoquinolones and pyrrolidines by solid phase ytterbium- and silver-mediated imine chemistry. *MedChemComm.* **2019**, *10* (7), 1082–1093.

250. Kunig, V. B. K.; Ehrt, C.; Dömling, A.; Brunschweiger, A. Isocyanide Multicomponent Reactions on Solid-Phase-Coupled DNA Oligonucleotides for Encoded Library Synthesis. *Org. Lett.* **2019**, *21* (18), 7238–7243.

251. Quarrell, R.; Claridge, T. D.; Weaver, G. W.; Lowe, G. Structure and properties of TentaGel resin beads: implications for combinatorial library chemistry. *Mol. Divers.* **1996**, *1* (4), 223–232.

252. Doran, T. M.; Dickson, P.; Ndungu, J. M.; Ge, P.; Suponitsky-Kroyter, I. An, H.; Kodadek, T. Synthesis and screening of bead-displayed combinatorial libraries. *Methods Enzymol.* **2019**, *622*, 91–127.

253. Wilson, S. R.; Czarnik, A. W. *Combinatorial Chemistry, Synthesis and Application*; Wiley: New York, 1997.

254. Aparna, E. P.; Devaky, K. S. Advances in the Solid-Phase Synthesis of Pyrimidine Derivatives. *ACS Comb. Sci.* **2019**, *21* (2), 35–68.

255. Kamal, A.; Reddy, K. L.; Devaiah, V.; Shankaraiah, N.; Reddy, D. R. Recent advances in the solid-phase combinatorial synthetic strategies for the benzodiazepine based privileged structures. *Mini Rev. Med. Chem.* **2006**, *6* (1), 53–69.

256. Abdildinova, A.; Gong, Y. D. Current Parallel Solid-Phase Synthesis of Drug-like Oxadiazole and Thiadiazole Derivatives for Combinatorial Chemistry. *ACS Comb. Sci.* **2018**, *20* (6), 309–329.