Click-SELEX

- A versatile approach towards nucleobase-modified aptamers -

Dissertation

zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Fabian Tolle

aus
Belo Horizonte

Bonn 2016

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Günter Mayer
2. Gutachter: Prof. Dr. Christoph Thiele

Tag der Promotion: 20.09.2016

Erscheinungsjahr: 2016

Parts of this thesis have been published in:

- Tolle, F., Rosenthal, M., Pfeiffer, F. & Mayer, G. Click Reaction on Solid Phase Enables High Fidelity Synthesis of Nucleobase-Modified DNA. *Bioconjugate chemistry* **27**, 500–503 (2016)
- Tolle, F. & Mayer, G. Preparation of SELEX Samples for Next-Generation Sequencing. *Methods Mol. Biol.* **1380**, 77–84 (2015)
- Tolle, F., Brändle, G. M., Matzner, D. & Mayer, G. A Versatile Approach Towards Nucleobase-Modified Aptamers. *Angew Chem Int Ed Engl* **54**, 10971–10974 (2015)
- Tolle, F., Wilke, J., Wengel, J. & Mayer, G. By-Product Formation in Repetitive PCR Amplification of DNA Libraries during SELEX. *PLoS ONE* **9**, e114693 (2014)

Abstract

Aptamers are short nucleic acids that can be selected to bind specifically and with high affinity to a variety of targets, including small organic molecules and proteins. However, for a multitude of targets, traditional SELEX methods still fail to yield suitable ligands. One possible cause is the limited chemical diversity found in the mere four different building blocks from which nucleic acids are constituted. Therefore, the addition of naturally not represented functional groups, e.g. the ones found in the side chains of proteinogenic amino acids, can increase the repertoire of possible interactions of nucleic acids and thereby expand the addressable target spectrum, enabling the selection of novel high-affinity binders.

Click-SELEX is a versatile method to facilitate the modular introduction of chemical entities during the SELEX process. In this approach chemical groups are introduced into the DNA library before the selection step (*in situ*) via click chemistry and are subsequently removed during the amplification step, thereby avoiding enzymatic incompatibility problems associated with larger nucleobase modifications. Hence, this system allows for an easy implementation of a multitude of different chemical functionalities adapted to the imposed requirements.

The development of all associated methods has been validated by the selection of a highly specific aptamer ("clickmer") interacting with the Green Fluorescent Protein (GFP). A 3-(2-azidoethyl)-indole functionalized DNA library was enriched for C3-GFP immobilized via a His-tag on magnetic beads. Traditional and next-generation sequencing of the enriched library revealed two aptamer families. C12, the most abundant monoclonal sequence, was shown to interact with high affinity with C3-GFP immobilized on beads and in solution. C12's binding affinity towards C3-GFP was demonstrated to be critically dependent on the correct functionalization state of three positions, as omission or substitution of the indole functionalization at these nucleotides led to a complete loss of binding affinity. The C12 clickmer also displayed a high degree of specificity, being capable of differentiating between two members of the GFP family.

Having pioneered the modular introduction of nucleobase functionalization the methods developed in this proof of concept study can now be applied to a variety of other functionalizations, selection strategies, and targets. The vast applicability of the click-SELEX method will rapidly advance the application of *in vitro* selection approaches beyond what was previously feasible, allowing, for example, the generation of aptamers to yet non-targetable molecules.

Zusammenfassung

Aptamere sind kurze Nukleinsäuren, welche selektiert werden können um Zielstrukturen, zum Beispiel kleine organische Moleküle oder Proteine, mit hoher Affinität und Spezifität zu binden. Allerdings können eine Vielzahl der gewünschten Zielstrukturen mit traditionellen SELEX Methoden nicht adressiert werden. Eine mögliche Erklärung für dieses Phänomen liegt in der limitierten chemischen Diversität natürlicher Nukleinsäuren. Die Einführung von zusätzlichen funktionellen Gruppen, wie sie etwa in den Seitenketten der proteinogenen Aminosäuren vorkommen, kann das Repertoire der möglichen Interaktionen deutlich erhöhen. Dadurch kann das adressierbare Spektrum an Zielstrukturen beträchtlich erweitert werden.

Click-SELEX ist eine vielseitige Methode für die modulare Einbringung von chemischen Gruppen während des SELEX Prozesses. Bei diesem Ansatz werden die chemischen Modifikationen vor dem Selektionsschritt (*in situ*) in die DNA Bibliotheken via "Click-Chemie" eingeführt und anschließend im Amplifikationsschritt wieder entfernt. Dabei werden Probleme mit enzymatischer Inkompatibilität, welche häufig mit sperrigen Funktionalisierungen assoziiert sind, vermieden. Dadurch erlaubt dieses System die einfache Einführung einer Vielzahl verschiedener, an die jeweiligen Anforderungen angepasster Funktionalitäten.

Die Entwicklung aller benötigten Methoden wurde durch die Selektion eines hoch spezifischen Aptamers ("Clickmers") für das grün fluoreszierende Protein (GFP) validiert. Eine Indol-funktionalisierte DNA Bibliothek wurde für Bindung an C3-GFP angereichert, welches zuvor mittels eines His-Tag auf magnetischen Partikeln immobilisiert wurde. Traditionelle und Hochdurchsatz-Sequenzierungen der angereicherten Bibliothek erlaubten die Identifizierung von zwei Aptamerfamilien. Für C12, der am häufigsten vertretenen mono-klonalen Sequenz, konnte eine hohe Affinität gegenüber C3-GFP, sowohl immobilisiert als auch in Lösung, nachgewiesen werden. Die Bindungseigenschaften von C12 waren dabei stark von der korrekten Funktionalisierung von drei Positionen abhängig; das Auslassen oder eine Substitution der Indol-Gruppen an diesen Nucleotiden führte zu einem vollständigen Verlust der Bindungsaffinität. C12 demonstrierte auch ein hohes Maß an Spezifität, da es zwischen zwei sehr ähnlichen GFP Varianten zu differenzieren vermochte. Nachdem die Methode nun etabliert wurde, kann click-SELEX auf eine Vielzahl neuer Funktionalisierungen, Selektionsstrategien und Zielmoleküle übertragen werden. Die breite Anwendbarkeit wird die Verwendung von *in vitro* Selektionsmethoden in neuartigen Anwendugsbereichen ermöglichen und damit die Generierung von Aptameren gegen derzeit schwer adressierbare Ziele vorantreiben.

Contents

1. Introduction

1.1. Aptamers, tools for chemical biology and beyond

Fundamentally, life and all its associated biological processes operate at a molecular level. With most molecules of interest being smaller than the wavelength of visible light, one key challenge at this size scale is the ability to specifically identify and interact with the desired molecules.[5] "Chemical Biology" uses chemical techniques and compounds to study and manipulate biological systems at this sub-microscopic level, thereby increasing our general understanding of biology.[6]

Aptamers are one class of molecular tool used in chemical biology. Similar to antibodies they are biopolymers that fold into distinct three-dimensional structures. Thereby, they create a unique interface, capable of specific interaction with their target molecule.[7] In contrast to traditional, small organic molecule based tools, aptamers are usually identified via *in vitro* selection from highly diverse libraries of up to 10^{15} different molecules, representing an enormous molecular repertoire. Aptamers can be selected against a variety of targets including small organic molecules[8], proteins[9], and even complex and unknown targets such as whole cell surfaces[10].

Being nucleic acids, aptamers have several unique features that set them apart from other affinity reagents. In contrast to proteinaceous affinity reagents, they have the ability to be self-complementary via specific base-pairing interactions. As a consequence, aptamers contain in their primary sequence all the information needed for their amplification (genotype). In nucleic acids, this genotype is intrinsically coupled to the three-dimensional structure (phenotype). This enables aptamers to be selected from random libraries, as opposed to the screening process needed for the identification of "small molecule" based affinity reagents. This makes them relatively rapid and easy to obtain, reducing the costs for identifying novel aptamers.[11]

Beside self-complementary, being nucleic acids has additional advantages. With aptamers being bio-molecules, many natural "tools" for their amplification and manipulation, such as polymerases and nucleases, are available. As nucleic acids play a significant role in biology, many methods were already developed for the detection, quantification and interaction with nucleic acids, e.g. polymerase chain reaction (PCR), microarrays, cloning techniques, and various sequencing methods.[12]

Short oligonucleotides can conveniently be synthesized by automated solid-phase chemistry[13], therefore, aptamers can also be regarded as chemicals. That leads to several advantages such as cheap and fast production with low batch-to-batch variation.[14] With solid-phase synthesis being

well established, the only information needed for the manufacture of a functional aptamer is its primary sequence. This allows for a decentralized production of the aptamer anywhere around the globe. The chemical production also facilitates the introduction of chemical modifications such as fluorophores and affinity tags. [15]

Based on these unique properties, in the last 25 years, aptamers were productively employed in a variety of applications spanning a multitude of research areas. These include, but are not limited to, the following fields of research:

Therapeutics

In analogy to the use of small organic molecules and antibodies, the use of aptamers as therapeutics has been extensively studied. [14,16,17] Being the first FDA approved aptamer-based drug, pegaptanib (Macugen) is probably the best-known example of a pharmaceutically relevant application of aptamers. [18] This heavily modified RNA aptamer is directed against the vascular endothelial growth factor (VEGF)-165, the VEGF isoform primarily responsible for pathological ocular neovascularization and vascular permeability. To overcome the problem of rapid renal clearance of aptamers, polyethylene glycol groups (PEG) were added to improve pharmacokinetics and the biodistribution. [19] To increase the stability of aptamers against nuclease digestion, artificial backbone modifications such as 2' fluoro (2'F) and 2' methoxy (2' OMe) groups were introduced increasing the possible spectrum of applications. In December 2004, Macugen was approved by the FDA for the treatment of age-related macular degeneration.

Research tools

Due to the comparably easy and fast identification of aptamers, they represent an important class of molecular research tools. One interesting class of research tools are aptamers that function inside cells, so-called "intramers". [20] Based on their ability to be transcribed inside cells, RNA aptamers modulating protein function inside cells could be generated. [21,22] In a recent example, an RNA aptamer that selectively inhibits the mitogen-activated kinase pathway in neurons was used under intracellular conditions, by application of the aptamer through a patch-clamp pipette. The aptamer was shown to efficiently inhibit mitogen-activated kinase-dependent synaptic plasticity, allowing for selective control of interneuronal signaling. [23]

Sensors

Aptamers comprise special properties that make them predestined to be used as molecular sensors. [24] As aptamers can be specifically amplified by PCR, they can be utilized for the very sensitive detection and quantification of analytes. AptaPCR employs the dual function of aptamers acting both as a selective ligand for target molecules and as a template for qPCR. [25] Aptamers also have the ability to act as bio-molecular "switches". They can be developed to sense environmental

chemical cues and, by undergoing specific, binding-induced conformational changes, transduce this recognition into signal outputs.[26] Based on this electrochemical property aptamer-based DNA sensors have been developed that can operate in real time even in living animals.[27]

With all of these special properties, aptamers represent an important class of affinity reagents, constituting a valuable addition to our repertoire of molecular tools. However, one major obstacle remains, namely the fact that aptamers can not always be selected for every desired target structure. Therefore, further improvements in the identification methods are needed.[28]

1.2. Systematic evolution of ligands by exponential enrichment (SELEX)

In this section, a detailed description of the SELEX process is given. Due to the vast amount of techniques and methods being developed for SELEX, a comprehensive overview of all methods is way beyond the scope of this thesis. Please refer to the following review articles for additional information on available SELEX methods.[12,29–32] For the sake of clarity, all examples refer to DNA-based SELEX. However, most ideas and methods also apply for SELEX with RNA.

1.2.1. Introduction to SELEX

The *in vitro* selection of nucleic acids was independently described by three research groups in 1990. In March of 1990 Debra Robertson and Gerald Joyce published the *in vitro* selection of an RNA enzyme that specifically cleaves single-stranded DNA. By an iterative cycle of mutation, selection and amplification they selected for a mutated form of the *Tetrahymena* ribozyme which could cleave DNA more efficiently than the wild-type enzyme.[33] In August of 1990, Craig Tuerk and Larry Gold described a method henceforth referred to as "Systematic Evolution of Ligands by EXponential Enrichment" (SELEX). Using this process they selected an eight nucleobase long region of RNA binding to T4 DNA polymerase with a binding constant equal to the wild-type ligand.[34] In the same month, Andrew Ellington and Jack Szostak published the selection of RNA molecules that bind specifically to a variety of organic dyes.[35] The group of Jack Szostak later coined the term "aptamer" for these nucleic acid-based ligands, from the Latin *aptus* "to fit", and the Greek *meros* "part".

The *in vitro* selection process described in these three groundbreaking publications is in several aspects similar to Darwinian evolution. The process starts with an enormously diverse library containing a huge variety of different oligonucleotide sequences. This library is submitted to a selection pressure, usually the ability to bind to a target molecule. A small fraction of sequences with the desired properties is separated from the bulk remainder of the library, resulting in an overall enrichment of sequences with the desired properties. This process is similar to the concept known as "survival of the fittest," such as during the selection step only the best adapted ("fittest")

sequences survive, thereby granting that their genotype is transferred to their ”offspring.” Usually, one round of selection is not enough to sufficiently enrich the library for the desired properties. Therefore, in an iterative process, the library is amplified and resubjected to the selection pressure. After several rounds of selection, the library is increasingly enriched, until only the very best adapted sequences prevail. A schematic representation of the SELEX process with DNA libraries is depicted in **figure 1.1**.

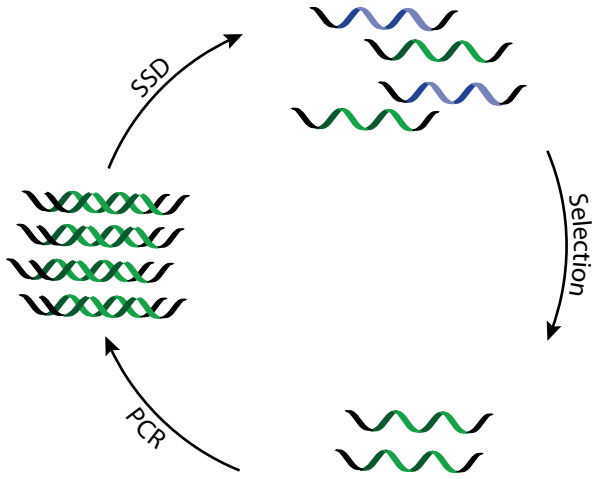

Figure 1.1. – Schematic representation of a generic DNA SELEX process
The DNA library is incubated with the target molecule and non-bound sequences are removed (Selection). The bound sequences are recovered and amplified by PCR. After PCR a single strand displacement (SSD) is undergone in order to obtain single stranded DNA. The enriched library is then subjected to the next selection cycle.

It is important to notice that selection strategies are in many ways superior to screening strategies. In a screening, each molecule is tested separately, which is very time and resource consuming. Although today, due to massive automatization and miniaturization, several million molecules can be screened in a reasonable time scale, these numbers are dwarfed by the up to 10^{15} different molecules that can be tested simultaneously in one selection experiment. The enormously high number of sequences that can be analyzed in one selection experiment increases the probability to find a specific sequence with the desired properties, despite the relatively low chemical diversity found in natural nucleic acids.

1.2.2. Nucleic acid libraries

Nucleic acid libraries for SELEX usually consist of a 20-100 nt long ”random region” which is flanked by two approximately 20 nt long primer binding sites. [36] Allowing for a random distribution of all four building blocks at each position of the random region, a theoretical diversity of 4^n, where n equals the length of the random region, is achieved. As a consequence, with a random region length of 25 nt upwards, only a fraction of the theoretical sequence space can be sampled

per SELEX experiment. However, today a random region length of around 40 nt is routinely used to ensure that the random region is long enough to accommodate stably folded structures.[37] A detailed description of the libraries used in this study is given in chapter 3.

1.2.3. Selection methods

In the last 25 years since the first description, a huge variety of selection methods has been developed. In this context, only two methods will be discussed. However, an excellent overview of the available methods is given in the following review articles.[12,31,32,38,39]

1.2.3.1. Magnetic-bead based selection

Today, one of the most used selection strategies involves the immobilization of the target on magnet particles (**figure 1.2 a**).[31] Many different immobilization chemistries are commercially available including streptavidin coated beads for the immobilization of biotinylated targets, cobalt functionalized beads for the immobilization of His-tagged proteins, and a variety of chemical functionalizations for the covalent attachment of target molecules.[31,40]

By incubating the library with target loaded magnetic beads (**figure 1.2 b**), the best binding sequences will bind to the beads and can conveniently be purified from the supernatant (**figure 1.2 c**). After several washing steps, binding sequences can be eluted from the particles for subsequent PCR amplification. To reduce the risk of enriching sequences that bind to the bead matrix instead of the immobilized target molecule, a negative selection step can be introduced during the SELEX. By incubating the library with the non-functionalized magnetic beads prior to the selection step, the library can be precleared from matrix-binding sequences. Thereby the risk of selecting matrix-binding aptamers is strongly reduced.[41]

Magnetic-bead based selection has the advantage that it is easy and fast to perform and that no expensive instrumentation is needed. Also, the whole process can be automated, facilitating the high-throughput selection of aptamers.[42] However, besides extensive washing of the beads, usually high degrees of unspecific background binding are observed. Unspecific binding sequences reduce the selection pressure and lead to an increase in the number of rounds required for successful selection.[43] Recently combinations of magnetic-bead based selection with washing in microfluidic channels have been developed to reduce the unspecific background binding, reducing the number of necessary selection rounds.[44–46] However, low stringency due to unspecific matrix binding remains a major obstacle for the efficient selection of high-affinity aptamers.[45]

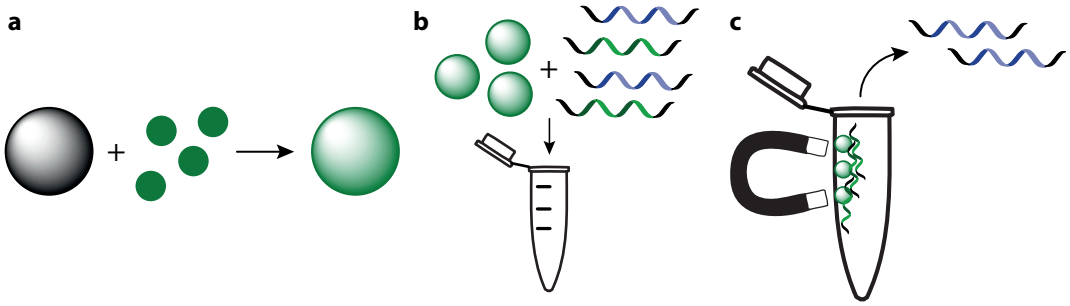

Figure 1.2. – Schematic representation of magnetic-bead based SELEX
a) The target (green circles) is immobilized on magnetic particles. Today, many different immobilization chemistries are commercially available, including streptavidin coated beads for the immobilization of biotinylated targets and cobalt functionalized beads for the immobilization of His-tagged proteins. b) The library is incubated with the target loaded magnetic beads. c) Using magnetic separation, the sequences with the desired properties are separated from the bulk library. After elution from the magnetic beads a sub-library enriched for the desired property is obtained.

1.2.3.2. Capillary electrophoresis based selection (CE-SELEX)

An alternative selection strategy, without the need for the immobilization of the target, is capillary electrophoresis (CE). With this method, a very efficient separation of the complex from the unbound library, with very low background binding is achieved. CE-SELEX has been used for aptamer selection against a variety of targets including proteins [47,48] and small molecules [49]. Due to the strong partitioning power of CE, usually, only a few selection cycles are needed. That makes CE-SELEX especially well suited for chemically modified libraries, where the enzymatic amplification is challenging. [50] However, the need for expensive instrumentation and the need for complex optimization of the selection setup obstruct a more widespread use of CE-SELEX.

1.2.4. Amplification of libraries

To generate enough starting material for the next selection round, the enriched library has to be amplified. In the case of DNA-SELEX, the amplification consists of two steps. First the exponential amplification by polymerase chain reaction (PCR) and second, a single strand displacement step to regenerate a single stranded DNA library (ssDNA) from the double stranded PCR product (dsDNA).

1.2.4.1. Polymerase chain reaction (PCR)

An essential step in SELEX is the amplification of enriched libraries to amplify sequences with the desired properties and generate enough material for the next selection round. Albeit being a very well studied technique, PCR of SELEX libraries is not without technical challenges. Due to the random nature of the library sequences, mispriming and, therefore, by-product formation is very common. [51] Several events can lead to the production of undesired by-products. [52] One

source of longer by-products is the priming of the 3' end of one sequence in the random region of a second sequence. This leads to the production of longer sequences, still containing both primer binding sites, therefore still capable of exponential enrichment. [4]

In addition to longer by-products, also, shorter by-product sequences can be formed. Lacking a proper selection pressure, having an advantageous replication behavior, shorter sequences can rapidly dominate an SELEX library. [51,53] Improved PCR procedures such as emulsion PCR or digital-PCR can help to reduce the risk of by-product formation. However more research is needed in this field. [54,55]

The enzymatic amplification, essential for the iterative nature of SELEX, displays a major hurdle for the introduction of unnatural modifications. In SELEX the efficient incorporation of the triphosphates by the polymerase is crucial for the amplification of the library. This explains the widespread use of C5-modified pyrimidines, as modifications at this position are usually well tolerated by polymerases. [56] Nevertheless, even exploiting this privileged position, some modified triphosphates will not be accepted by the polymerase, limiting the use of nucleobase modifications in SELEX. In general, sterically more demanding modifications are less well tolerated by the polymerases, however also other factors such as potential interaction capabilities with the active site influence the acceptance, making the incorporation of modified triphosphates hard to predict. [56]

1.2.4.2. Single strand displacement (SSD)

Many techniques have been developed for single strand displacement during DNA-SELEX. [57–59] In this study, λ-exonuclease digestion [60] has been used exclusively. In this enzymatic technique, the enzyme λ-exonuclease recognizes and specifically digests the 5' phosphorylated strand of a dsDNA duplex. By introducing a 5' phosphate group with the reverse primer in the PCR, the antisense strand can conveniently be removed, yielding the desired ssDNA library for the next selection round. Compared to other more labor-intensive techniques, requiring electrophoretic separation of the desired strand, λ-exonuclease is fast and offers high recovery yields, with the additional benefit of being compatible with high-throughput robotic selection platforms. [57]

1.2.5. Identification of aptamer candidates

After several SELEX rounds, the library should be sufficiently enriched to identify monoclonal aptamer sequences. Aptamer identification is made by sequencing the library and analyzing the frequency of each sequence in the last SELEX round. However, in recent years, next-generation sequencing has been introduced for the analysis of SELEX with an ever-growing role in the analysis and interpretation of SELEX experiments. [31]

1.2.5.1. Sanger sequencing

Traditionally, aptamers were identified by cloning and sequencing of the library of the last selection round. [29] The library is cloned into a plasmid and transformed into competent bacteria. After amplification of the bacteria, the plasmids are purified and sequenced by the chain terminator method developed by Nobel laureate Frederic Sanger (Sanger sequencing). [61] Due to the comparatively low throughput, with this method only up to a few hundred clones can be sequenced practically. Therefore, a high degree of enrichment is necessary to robustly identify sequence families from this small data set. A significant advantage of this technique is that a template plasmid is generated, from which the potential aptamer sequence can be amplified by PCR. Thereby, the material for the validation of potential aptamer sequences can easily be produced, without the need for solid-phase synthesis.

1.2.5.2. Next-generation sequencing (NGS)

The human genome project and other whole genome sequencing projects have created the need for higher throughput than conventional Sanger sequencing could offer. This lead to the development of a new generation of sequencing technologies known as next-generation sequencing (NGS), high throughput sequencing (HTS) or deep sequencing. The high adoption rates of such techniques have lead to a dramatic price drop over the last decade. This degree of affordability has made NGS very attractive to the SELEX community. [62]

With NGS, up to 100 million sequences can be sequenced in one experiment, enabling a much deeper insight into *in vitro* selection. The ability to multiplex several samples in one sequencing run enables deep sequencing, not only of the last round but for each round of the entire SELEX. [63] Capacity to track sequence families and their amplification behavior over several rounds of selection allows much better prediction of interesting sequences, thereby increasing the success chances of *in vitro* selection experiments. [2] Today, an increasing amount of bioinformatic solutions are being developed to facilitate the interpretation and aptamer identification from huge NGS data sets that are becoming available. [38,64–67] Albeit constant improvements in recent years, the processing, visualization and interpretation of NGS data remain challenging and offers an untapped source for the future improvement of the SELEX methodology.

1.2.6. Validation of aptamer candidates

After having identified potential aptamer candidates by sequencing, the candidates have to be validated by interaction analysis. Most methods for protein – protein or protein – small molecule interaction analysis can be used. However, some methods are especially well suited for the mea-

surement of aptamer interactions. The following three methods were used for binding analysis throughout this study. An overview of more available techniques can be found in the following review articles.[64,68,69]

1.2.6.1. Filter retention assay

The filter retention assay is a technique to analyze the interaction of a radioactively labeled aptamer to a protein.[70] It relies on the property of nitrocellulose membranes to interact with proteins, but not with nucleic acids.[71,72] The nucleic acid – protein mixture is filtered through a nitrocellulose membrane, whereby only the protein is retained on the membrane. The nucleic acid is washed away, unless it interacts with the retained protein. The amount of retained nucleic acid is thereby proportional to its interaction with the protein. As the nucleic acid is radioactively labeled, it can be precisely quantified. Radioactive labeling of DNA can easily be performed enzymatically in a kinasation reaction, where the radioactive α-phosphate group of ATP is transferred to the free 5' hydroxyl group of the DNA. Advantages of this method include its high sensitivity and comparably cheap instrumental setup.[73] However, a radionuclide laboratory is required.

1.2.6.2. Microscale thermophoresis (MST)

Microscale thermophoresis (MST) is a comparably new method for interaction analysis.[74] It depends on the measurement of the movement in a microscopic temperature gradient.[75] Interaction of two molecules leads to a reorientation of the hydration shell, thereby affecting the thermophoretic properties.[76] MST offers the advantage to operate in solution, so no immobilization is required. In most cases, aptamers can easily be functionalized with fluorophores at e.g. their 5' position, without affecting their binding abilities. Thereby, no modification of the target molecule is required.[74]

1.2.6.3. Flow cytometry assay

A rather unusual, however very useful way to measure the interaction of aptamers with their target molecule is flow cytometry. Here, the same magnetic particles used for the SELEX are directly analyzed by flow cytometry. In the flow cytometer, the fluorescence intensity of the magnetic beads can be measured. Thereby, binding of fluorescently labeled aptamers to the particles can be detected.[68] This method allows the measurement of enriched libraries or potential aptamer sequences under the exact SELEX conditions and can, therefore, be beneficial for the fast assessment of the enrichment during the SELEX.[3]

1.3. Limitations of current SELEX approaches

The interplay of all methods and concepts introduced in section 1.2 are vital for a successful SELEX experiment. Constant development and improvements in every single technique help to push the SELEX field forward.[37] However, a special role is held by the library.[77] For a successful selection, molecules with the desired properties have to be present in the library. Otherwise, even the most advanced techniques will not be able to identify the desired aptamer. Natural nucleic acid libraries often fail to yield the expected aptamers, most likely due to the limited chemical diversity found in the merely four different building blocks from which nucleic acids are constituted.[78] Therefore, the addition of naturally not represented functional groups, as the ones found in the side chains of amino acids, can increase the repertoire of possible interaction capabilities of nucleic acids, thereby increasing the addressable spectrum of targets and allowing the selection of unprecedented high-affinity binders.[79]

1.4. Chemically modified nucleic acid libraries

Nucleic acid libraries based on the four naturally occurring building blocks can be successfully used for the selection of aptamers with useful functions and many possible applications. However, for some applications, natural nucleic acids have some tremendous drawbacks.[80] Being biomolecules, especially RNA aptamers, are very susceptible to nuclease degradation. To improve certain traits, such as nuclease resistance, unnatural building blocks have been introduced in SELEX libraries.

In addition to stabilizing aptamers against degradation, chemical modifications can also be used to modulate folding and interaction properties.[15] Especially the introduction of unrepresented functional groups can increase the possible spectrum of interactions with the target molecule. Thereby, aptamers for previously non-targetable epitopes can be generated.[78] Chemical modifications usually fall in one of two major categories:

1.4.1. Backbone modifications

To increase nuclease stability, much research has been focused on the 2' position of RNA. Several modifications such as 2'-amino pyrimidines, 2'-fluoro pyrimidines and 2'-methoxy nucleotides have been used successfully and have yielded many high-affinity aptamers with increased nuclease resistance such as the aformentioned pegaptanib (Macugen) (**figure 1.3 a**). Today, especially 2'-fluoro pyrimidines are widely used in SELEX experiments with more than 15 published aptamers generated with this modification.[39]

One special form of 2' modification is found in locked nucleic acids (LNA) (**figure 1.3 b**).[81,82] The incorporation of LNA modifications not only increases the stability against nucleases but also enhances the affinity for complementary DNA and RNA sequences. This property makes LNA modifications especially appealing for the selection and maturation of short, highly stable aptamers.[83–85]

Another very elegant way to stabilize aptamers against nucleases is realized in the "spiegelmer" concept, which exploits the chirality of biomolecules.[86,87] In this approach, a conventional D-aptamer is selected against an unnatural mirror-image target. As a consequence, the synthetic mirror-image L-aptamer is capable of interacting with the natural occurring target, but is not recognized by natural nucleases (**figure 1.3 c**). Today, three spiegelmers, all developed by NOXXON Pharma, are in clinical trials.[88]

Not only sugar-modifications, but also modifications at the phosphate backbone have been employed for aptamer selection. The most prominent class of phosphate modifications are phosphorothioates[89–91] (**figure 1.3 d**) which not only increase the stability against nucleases, but also improve the ability to internalize into cells.[92] Recently, sugar-modifications beyond the 2'-alterations have been realized. Xeno-nucleic acids (XNA) (**figure 1.3 e and f**) are a new class of genetic polymer supporting Darwinian evolution.[93] The development of XNA aptamers was enabled by the engineering of new polymerases capable of DNA-dependant XNA polymerization and XNA dependent DNA polymerization with high fidelity. Using the expanded XNA backbone repertoire several XNA aptamers and XNA catalysts have been selected.[93–95]

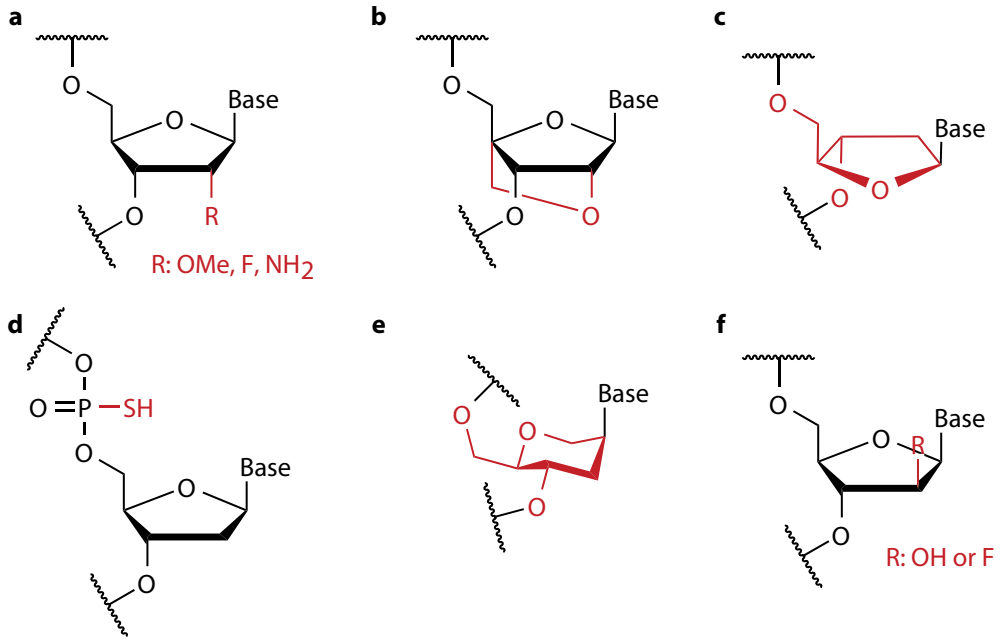

Figure 1.3. – Backbone modifications employed for aptamer selection
a) 2'-amino, 2'-fluoro and 2'-methoxy modifications. b) Structure of locked nucleic acids (LNA). c) Mirror-image D-aptamer backbone found in "spiegelmers". d) Structure of the phosphorothioate backbone. e) Structure of the hexitol nucleic acid (HNA) backbone. f) Structure of the arabinose (ANA) and 2'-fluoro-arabino nucleic acid (FANA) backbone.

1.4.2. Nucleobase modifications

Several nucleobase modifications have been introduced in nucleic acids libraries for aptamer selection.[96] The main objective being, to increase the chemical diversity and, therefore, the available chemical space for selection. The introduction of novel chemical groups allows for additional types of interaction between nucleic acids and their target structures. This, in turn, increases the probability of identifing aptamers with the ability to interact with the target with high affinity and specificity.[97,98]

Modifications have been mainly introduced at the C5-position of pyrimidines, such as deoxyuridine analogs. This position is privileged as modifications are often tolerated by polymerases, retaining the essential compatibility with enzymatic amplification of the library (**figure 1.4 a**). The first example of the use of nucleobase-modified nucleotides in DNA-SELEX was published 1994 by John Latham and coworkers at Gilead Sciences.[99] Using a 5-(1-pentynyl)-2'-deoxyuridine building block (**figure 1.4 b**), instead of the canonical thymidine, they were able to select an aptamer against thrombin. Without showing any sequence similarities, the resulting aptamer displayed a similar dissociation constant for thrombin as described for a previously selected ssDNA aptamer.[9]

Since then several more selections with nucleobase modifications at the C5-position of pyrimidines have been performed [100–104], however arguably the most significant contribution to this field was developed by Bruce Eaton and coworkers at the University of Colorado. With the goal to develop aptamers with slower off-rates (and thus, higher affinities), protein-like side chains have been introduced into DNA libraries. [77,79] This has lead to the development of SOMAmers (Slow Off-rate Modified Aptamers) by SOMAlogic. [78]

A broad range of nucleobase modifications have been used, many inspired by hydrophobic residues similar to the ones found in the amino acids tryptophan or phenylalanine. The structure of some nucleobase modifications employed by SOMAlogic is depicted in **figure 1.4 c**. Having used different libraries for the automated selection against several hundred proteins, a dramatic increase in success rate for the chemically modified libraries has been reported. [78] This proves the hypothesis that indeed an increase in chemical diversity can have a beneficial impact on the interaction capabilities of nucleic acid libraries. These interactions and the deep involvement of the modified nucleotides were confirmed by crystal structures of several SOMAmer-protein complexes. [105–108]

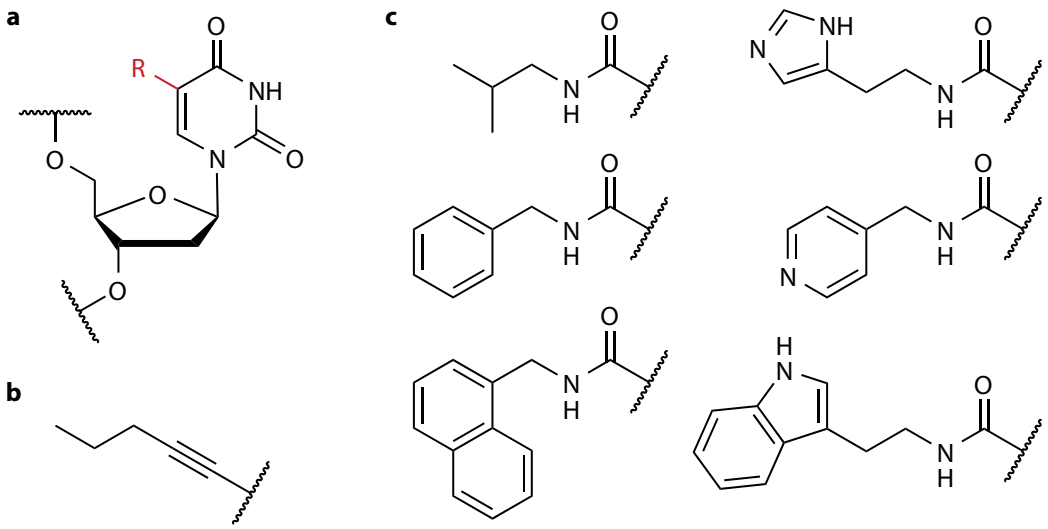

Figure 1.4. – Nucleobase modifications employed for aptamer selection
a) General structure of modified deoxyuridine analogs. b) 5-(1-pentynyl)-modification used by Lantham and coworkers in the first published SELEX with nucleobase-modified nucleotides. c) Overview of some modifications used by SOMAlogic for the selection of SOMAmers.

1.4.3. *In situ* introduction of functionalizations

As chemical modifications are an important way to improve future aptamer selection, a versatile method to facilitate the modular introduction of chemical entities during the SELEX process would be desirable. Such a SELEX concept, based on the *in situ* introduction of the functional-

ization, became possible due to the recent developments in the field of bioconjugation. Although, the general concepts were not new and many techniques had previously been developed [109], the field underwent dramatic acceleration with the discovery of truly bioorthogonal reactions. [110,111]

In 2001, the group of Nobel laureate Barry Sharpless coined the term click chemistry. [112] In their click concept, methods for the quick and reliable formation of chemical structures out of smaller building blocks were described. This idea mimics the way nature generates complex structures by joining small modular units. To be considered a click reaction, a reaction has to fulfill a certain set of criteria such as: be modular and wide in scope, give very high yields and have a strong thermodynamic driving force. Also, it should generate only inoffensive by-products, be stereospecific (but not necessarily enantio-selective), and the product must be stable under physiological conditions. That implies the required process characteristics to include: simple reaction conditions (ideally, the process should be insensitive to oxygen and water), readily available starting materials and reagents, the use of an environmentally friendly solvent such as water, and simple product isolation by non-chromatographic methods, such as crystallization or distillation. [112]

One of the few reactions truly fulfilling most of these requirements is the Cu^{I} catalyzed 1,3-dipolar cycloaddition of alkynes and azides (CuAAC). Although the term click reaction originally described the concept mentioned above, today, through its widespread use, is mostly used synonymously to CuAAC.

1.4.3.1. Azide-alkyne Huisgen cycloaddition (CuAAC)

The 1,3-dipolar cycloaddition is a reaction between a 1,3-dipole (such as an azide) and a dipolarophile (such as an alkyne) forming a five-membered ring (such as a 1,2,3-triazole). The reaction was discovered and investigated by Rolf Huisgen in the 1960s at the LMU in Munich. [113,114] In its non-catalyzed version the reaction proceeds by a pericyclic orbital controlled mechanism, which results in a mixture of the two regio-isomer products. The non-catalyzed 1,3-dipolar cycloaddition is valuable for the production of five-membered heterocyclic rings such as triazoles, but due to its very low regioselectivity and the high temperatures required, for many years, it did not have a big impact on the preparative organic synthesis of biologically relevant compounds.

That changed rapidly when in 2002, the groups of Morten Meldal [115] and Valery Fokin [116] independently discovered that the reaction of an alkyne with an azide can be catalyzed by the addition of Cu^{I} (copper catalyzed azide-alkyne cycloaddition or CuAAC). Not only does the addition of Cu^{I} catalyzes the reaction to work at room temperature, but it also leads to the selective production of the sterically favored 1,4-triazole (**figure 1.5**).

The most convenient source for Cu^{I} is $CuSO_4$ coupled with a reducing agent such as sodium ascorbate for *in situ* generation of the Cu^{I} species. Cu^{I} is thermodynamically very unstable which results in an easy oxidation to Cu^{II} in the presence of water or oxygen. The combination of copper

and sodium ascorbate has been shown to produce oxygen radicals which can be detrimental to DNA and other biomolecules.[117,118] Therefore, in order to protect Cu^{I} from oxidation, stabilizing ligands have to be applied (**figure 1.7**).[119,120]

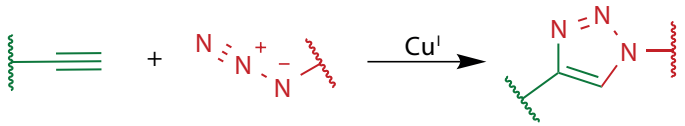

Figure 1.5. – Copper catalyzed azide-alkyne cycloaddition (CuAAc)
An alkyne (green) and an azide (red) react under Cu^{I} catalysis to a 1,4-triazole.

1.4.3.2. CuAAC with nucleic acids

The click-SELEX concept was made possible by novel groundbreaking work in the field of bioconjugation. Although click chemistry has found enormous application within the field of nucleic acid research, within this study only the use for nucleobase modification will be discussed. For additional applications of click chemistry with nucleic acids, please refer to this excellent review article.[121]

The C5-position of pyrimidines is privileged for the introduction of modifications. Unsurprisingly, the incorporation of alkyne groups at this position has been studied extensively.[39] As a result two building blocks with alkyne modifications at the C5-position of deoxyuridine, namely 5-ethynyl-deoxyuridine (EdU) (**figure 1.6 a**) and 5-(Octa-1,7-diynyl)-deoxyuridine (C8-dU) (**figure 1.6 b**), have been well characterized and made commercially available, with higher reaction yields published for C8-dU compared to EdU.[122–124] Based on the idea, that a high degree of functionalization is essential to guarantee a consistent level of modification throughout the SELEX, C8-dU was chosen as a substitution for thymidine in the initial click-SELEX library.

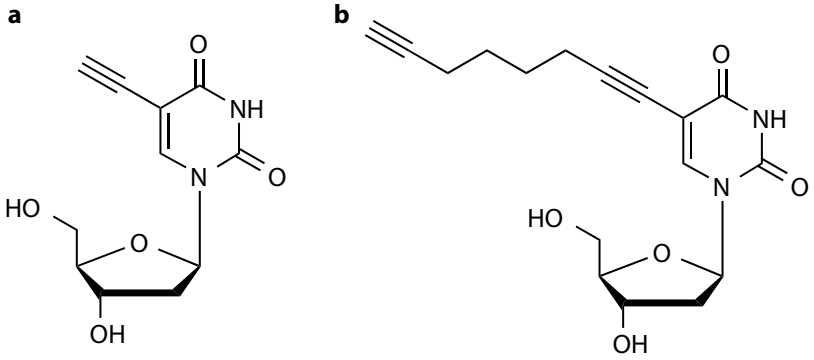

Figure 1.6. – Alkyne functionalized deoxyuridine analogs
a) Structure of the alkyne modified EdU nucleoside) b) Structure of the alkyne modified C8-dU nucleoside.

As mentioned in the previous section, Cu^{I} stabilizing ligands are needed for CuAAC with nucleic acids. Several generations of ligands have been developed. The first generation of ligand was tris(benzyl triazolyl methyl)amine (TBTA) (**figure 1.7 a**).[119] Although many of the first click chemistry studies with nucleic acids were done with this ligand, one big limitation of this ligand is its low solubility in water. Therefore, high amounts of organic co-solvents have to be used, which in turn can be problematic for the solubility of the very polar nucleic acids.

This limitation was addressed with the next generation ligand tris(4-(3-hydroxy-propyl)-[1,2,3]triazol-1-ylmethyl)amine (THPTA) (**figure 1.7 b**).[125] Due to the replacement of the phenyl group found in TBTA with hydroxyl groups, a much better solubility in water was achieved. That makes THPTA the optimal copper stabilizing ligand for aqueous bio-conjugations. Today, many more ligands are described[126–128], however due to the excellent results obtained with THPTA this ligand was used throughout this study.

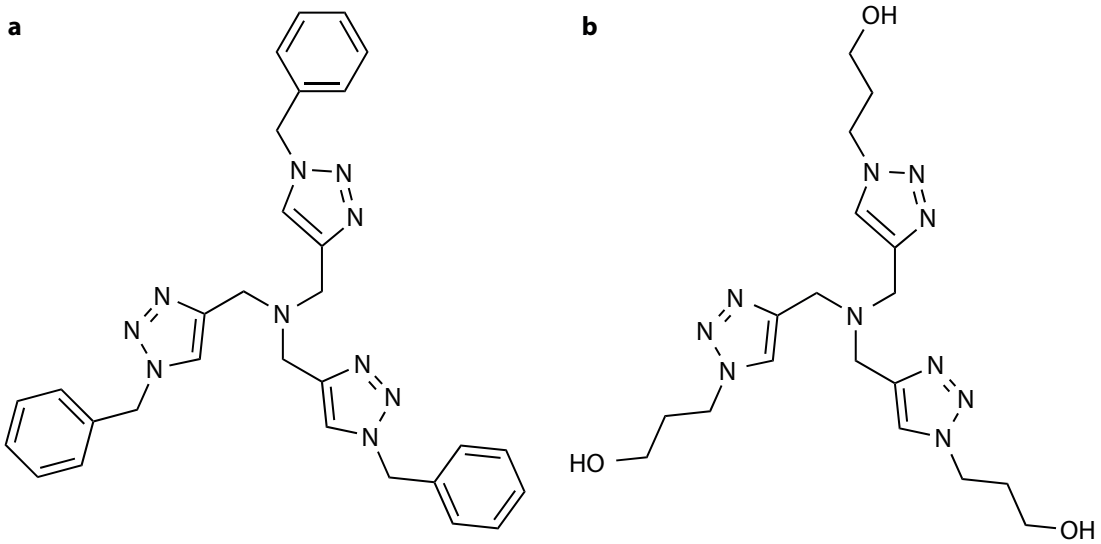

Figure 1.7. – Cu^{I} stabilizing ligands for click chemistry
a) Structure of the first generation copper stabilizing ligand TBTA. b) Structure of the water soluble next generation copper stabilizing ligand THPTA.

2. Aim of this Study

The aim of this study was to develop a modular SELEX method for nucleobase-modified DNA libraries, with the overall goal to further improve the aptamer selection process. In this context, increasing the chemical diversity of SELEX libraries was identified as an important area for innovation. The study was based on the hypothesis that an increase in chemical diversity of the library would allow for more interaction possibilities with the target, which in turn should have a positive impact on the selection of high-affinity aptamers to previously non-targetable molecular structures.

The established practice of introducing nucleobase modification by incorporation of modified dNTPs during PCR has several disadvantages, such as incompatibility of many desirable modifications with enzymatic amplification during PCR and a high synthetic effort needed for the generation of the modified dNTPs. To circumvent these limitations, the introduction of the modification after the enzymatic amplification is required. For such an *in situ* approach, a functionalization method specifically targeting the nucleobases already incorporated in a library is indispensable. A suitable bioconjugation technique known as CuAAC or click chemistry had recently been described, allowing for the modular introduction of a variety of modifications, without the need for extensive organic synthesis.

The concept behind this study was to incorporate CuAAC based nucleobase functionalization into a versatile and modular SELEX protocol. To develop the click-SELEX method, most parts of the SELEX process, such as library design, amplification, functionalization as well as the selection method and aptamer characterization methods had to be adapted and further developed. Finally the concept should be applied to the selection of a nucleobase-modified aptamer (”clickmer”) against a model protein target.

3. Results

3.1. The click-SELEX concept

In the click-SELEX approach, chemical groups are introduced into the DNA library prior to the selection step (*in situ*) via click chemistry (CuAAC) and are subsequently removed during the amplification step. Thereby enzymatic incompatibility problems associated with larger nucleobase modifications are avoided. This system allows for an easy implementation of a multitude of different chemical functionalities, adapted to the imposed requirements.

Figure 3.1 illustrates the general click-SELEX concept. Click-SELEX starts with a synthetic DNA library in which all thymidines are substituted by an alkyne-modified analog. The library is further functionalized with an azide-bearing molecule by CuAAC. After incubation with a target protein immobilized on magnetic particles, unbound molecules are removed and discarded by washing, whereas the target proteins with the bound modified nucleic acids are eluted (selection). The eluted molecules are then subjected to PCR amplification with an alkyne-bearing triphosphate instead of thymidine. This step removes the modification in the elongating strand and reintroduces the alkyne moiety. After amplification, the antisense strand is digested with λ-exonuclease, which selectively recognizes and hydrolyzes the 5'-phosphorylated strand (introduced by the reverse primer) of the double-stranded DNA. The remaining single-stranded alkyne-modified DNA is again reacted with the azide, thereby reintroducing the functionalization and reconstituting the starting library for the next selection cycle.

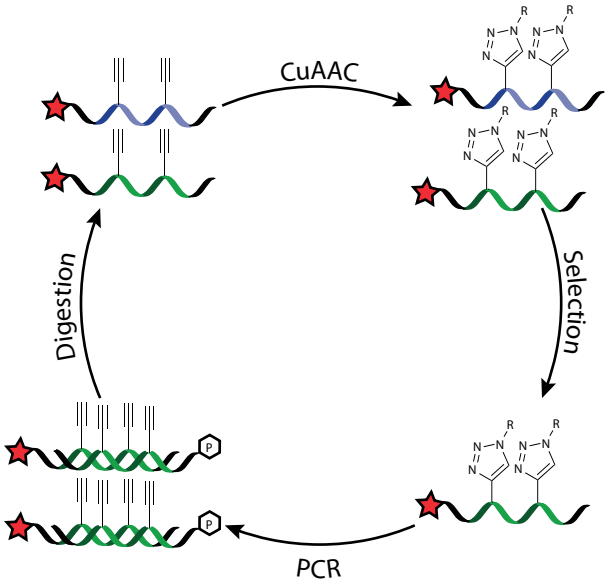

Figure 3.1. – Schematic representation of the click-SELEX process.
Schematic representation of the click-SELEX process. A synthetic alkyne-modified DNA library is further functionalized with an azide-bearing molecule by CuAAC. After incubation with the target molecule and removal of non-bound sequences (selection), the bound sequences are eluted and amplified by PCR using the alkyne-modified triphosphate instead of thymidine. This step removes the modification in the elongating strand and reintroduces the alkyne moiety. Thereby enzymatic incompatibility problems associated with larger nucleobase-modifications are avoided. After PCR, the single stranded DNA is prepared by λ-exonuclease digestion of the 5' phosphorylated antisense strand (digestion). The modification is then reintroduced by click chemistry (CuAAC), and the thereby obtained library is subjected to the next selection cycle.

3.2. Activated Protein C (APC) as a model target protein

For the purpose of developing the click-SELEX method, the focus of this study lied on practical aspects rather than developing a particular aptamer to address a distinct biological question. To facilitate the establishment of the method, a well addressable protein target was desired. Therefore, as a first target, the "Activated Protein C" (APC) was chosen.

Activated protein C (APC) is a serine protease with anticoagulant, anti-inflammatory, and cytoprotective properties. It plays a significant role in the blood clotting process, which consists of a series of enzymatic reactions leading to the final formation of thrombin.[129] APC, which is generated from its zymogen Protein C (PC), is responsible for avoiding over-coagulation, reducing the risk of thrombus formation. In addition to its anticoagulant functions, APC shows cytoprotective effects, including anti-inflammatory and anti-apoptotic activities, and protection of endothelial barrier function.[130] Consequently, substitution with recombinant APC (Drotrecogin alfa (acti-

vated) or "Xigris") was shown to reduces mortality rates in severe sepsis.[131] However, a serious side effect of APC substitution is major bleeding that occurs in 2 % of patients.[132] This severe side effect limits its clinical use and has lead to the withdrawal of Xigris from the market in 2011.

Searching for an APC-specific antidote that would be helpful to manage APC-induced bleeding, Jens Müller and coworkers have previously selected a high-affinity aptamer against APC.[133,134] This aptamer, termed HS02, is known to bind the basic exosite of APC with sub-nanomolar affinity. Knowing to be very well targetable by nucleic acids, and its significant therapeutical importance was thought to render APC an ideal first model target for the development of the click-SELEX method.

3.3. Test SELEX with canonical DNA

The first step towards a functional click-SELEX protocol was to design a suitable library and the adaptation of all required individual techniques. This included the amplification of the DNA library in a PCR reaction, the subsequent single strand displacement (SSD) of the antisense strand via λ-exonuclease digestion and the establishment of a suitable selection method.

To test the library and validate the selection conditions, it was decided to perform a test SELEX with a canonical, non-functionalized DNA library. A traditional DNA library and a magnetic bead-based selection protocol had previously been used for the selection of a high-affinity aptamer against activated protein C (APC) in the laboratory of Prof. Mayer at the LIMES Institute in Bonn[133]. As discussed in section 3.2 (page 20), due to its very good addressability by DNA libraries, APC was chosen as a model target for the establishment of the click-SELEX protocol. For the SELEX conditions, a protocol close to the previously successful selection conditions was developed.

3.3.1. Library design of the FT2-N42 library

The library suitable for click-SELEX was developed based on the most frequently used SELEX library design.[31] Similar to the library used in the previous APC SELEX, it consisted of a 42 nt random region, flanked by two 21 nt primer binding sites. In contrast to a "normal" library, thymidine building blocks should be avoided in the primer binding sites, because those would be substituted by their alkyne-modified counterpart in the amplification step, potentially causing steric problems in the PCR amplification. Following the design of previously successful DNA libraries, an equal amount of all four nucleobases was desired in the random region, statistically resulting in 10.5 thymidine nucleotides or its respective alkyne-functionalized counterpart. A schematic representation of the FT2-N42 library, designed for the test selection, is depicted in **figure 3.2**.

5'-CACGACGCAAGGGACCACAGG---**N42**---CAGCACGACACCGCAGAGGCA-3'

N: dA:dG:dC:T = 1:1:1:1

Figure 3.2. – Schematic representation of the FT2-N42 library
The library is composed of a 42 nt random region (green) flanked by two 21 nt primer binding sites (black). To avoid potential problems due to steric hindrance during the PCR, associated with the alkyne modifications that should later be introduced, the primer binding sites were designed without any thymidine nucleotides. The random region was designed with an equal distribution of all four nucleotides.

3.3.2. Amplification of the FT2-N42 library

The ability to amplify the enriched library for the next selection step is crucial for SELEX. In the case of DNA libraries, the amplification consists of two stages. First, an exponential amplification by PCR, generating multiple copies of double stranded DNA for each template stand. Second, PCR is followed by a single strand displacement (SSD) step, e.g. λ-exonuclease digestion, to generate the desired ssDNA library for the next selection round.

3.3.2.1. PCR amplification of the FT2-N42 library with canonical dNTPs

Before introducing alkyne-modified building blocks, the PCR conditions with the non-modified FT2-N42 library had to be established. It had been described that alkyne-building blocks are well incorporated into DNA by the the *Pwo* polymerase.[124] Therefore, this polymerase was used for all PCR reactions. After careful optimization of the PCR protocol in regard to annealing temperature, primer concentration and magnesium concentration, the following optimized protocol was established. PCR was done in a Mastercycler Personal (Eppendorf). A final concentration of 0.5 µM of both primers and 250 µM of dNTPs were used. *Pwo* DNA polymerase (Genaxxon) and the supplied buffer containing 2 mM Mg^{2+} were used with the following cycling program (2 min 95 °C; 30 s 95 °C, 30 s 62 °C, 1 min 72 °C; hold 4 °C). The samples were always prepared on ice and the thermocycler was pre-heated to 95 °C to reduce the risk of by-product formation. **Figure 3.3** depicts a representative PCR product, proving the successful amplification of the FT2-N42 with canonical nucleotides.

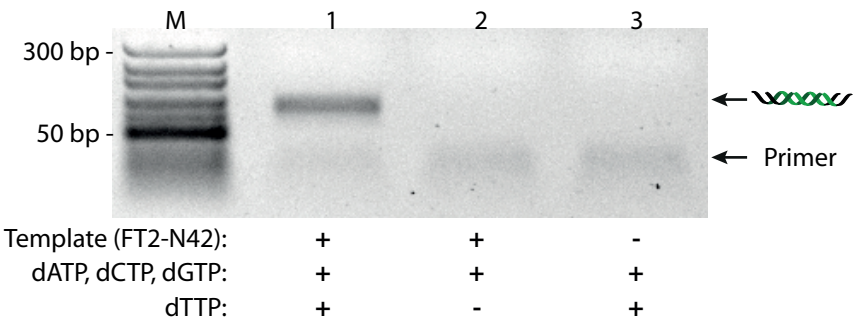

Figure 3.3. – PCR amplification of the FT2-N42 library with canonical dNTPs
Ethidium bromide stained 4% agarose gel. 1) PCR product for the FT2-N42 library amplified using canonical dNTPs. Double stranded DNA of the correct length (84 bp) is produced. 2) Control PCR reaction without the addition of dTTP. No double stranded DNA product is detectable. 3) "No template control" (NTC) PCR reaction without the addition of the template DNA. No double stranded DNA product is detectable.

3.3.2.2. λ-Exonuclease digestion of the FT2-N42 library

After being able to successfully amplify the library by PCR, the next step was to generate single stranded DNA as needed for the next selection round. As discussed in section 1.2.4.2 (page 7), several methods are available. However, λ-exonuclease digestion is one of the fastest and most convenient methods for the single strand displacement (SSD).[60] The enzyme recognizes and specifically digests the 5' phosphorylated strand of a dsDNA duplex. By introducing a 5' phosphate group with the reverse primer in the PCR, the antisense strand can conveniently be removed, yielding the desired ssDNA library for the next selection round. Based on the manufacturer's recommendation the following protocol was developed. 175 µl DNA solution was mixed with 20 µl λ-exonuclease buffer (10 x) and 5 µl λ-exonuclease (5000 U/ml). The sample was incubated on a thermomixer for 1 h at 37 °C and 800 rpm.

Figure 3.4 shows the results for the λ-exonuclease digestion of the non-modified FT2-N42 library. After the enzymatic single strand displacement, no residual dsDNA could be detected on the agarose gel, making this method a good option for the single strand displacement during SELEX.

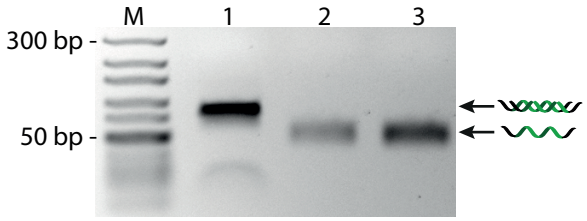

Figure 3.4. – λ-exonuclease digestion of the FT2-N42 library
Ethidium bromide stained 4% agarose gel. 1) Double stranded FT2-N42 library before treatment with λ-exonuclease. 2) FT2-N42 library after treatment with λ-exonuclease. After the enzymatic single strand displacement, no residual dsDNA could be detected. 3) 3 pmol of a solid-phase synthesized ssDNA library (FT2-N42) loaded as a reference.

3.3.3. Adaptation of the selection method

The first selection attempt was based on a well established bead-based selection strategy, which had yielded high-affinity aptamers for APC in the past.[133] Besides the already discussed unusual primer binding site design of the library, additional small changes were made to the published protocol. To reduce non-specific binding of the library, a buffer system for the selection of base-modified DNA aptamers had been previously described.[78] Based on this system, the following SELEX buffer was chosen: Dulbecco's phosphate buffered saline[135] (D-PBS, section 6.1.3, page 99) enriched with 1 mg/ml bovine serum albumin (BSA), 0.1 % TWEEN-20 and 0.1 mg/ml salmon sperm DNA.

Before the selection could be started, the target protein (APC) had to be immobilized on the streptavidin-coated magnetic beads (Dynabeads M-280 Steptavidin). Therefore, the APC had to be biotinylated at random lysine residues with an amine-reactive N-hydroxysuccinimide (NHS) biotinylation reagent. In contrast to the published method, due to its increased solubility in water, a polyethylene glycol based biotinylation reagent was chosen (NHS-PEG4-Biotin by Life Technologies).

3.3.3.1. Preparation of APC-coupled magnetic beads

APC (Haematologic Technologies) was biotinylated with NHS-PEG4-Biotin (Life Technologies) according to the manufacturer's recommendation. Briefly, 200 pmol APC was incubated with 20 equivalents of NHS-PEG4-Biotin, freshly prepared in DMSO to a final volume of 50 µl. The solution was incubated for 1 h on ice and 15 min at 25 °C and subsequently purified with Amicon Ultra centrifugal filters (10 kDa cutoff) according to the manufacturer's recommendation.

The success of the biotinylation reaction was verified with a dot-blot experiment in which conjugated and non-conjugated protein was spotted on a nitrocellulose membrane and treated with a fluorescently labeled avidin conjugate, specifically staining the biotinylated protein.

Figure 3.5 depicts the result of a dot-blot experiment with 1 µg APC before and after biotinylation, stained with Neutravidin-800 (Life Technologies) scanned at 800 nm in a Odyssey imaging system (Li-Cor).

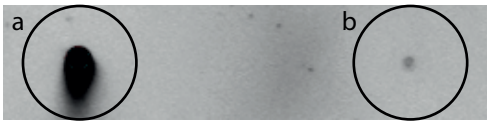

Figure 3.5. – Dot-blot of biotinylated APC
Nitrocellulose membrane treated with a fluorescently labeled avidin conjugate (Neutravidin-800 by Life Technologies) scanned in a Odyssey imaging system at 800 nm. a) After biotinylation a strong signal could be detected for the APC protein (1 µg). b) Before biotinylation of the APC protein (1 µg) no signal could be detected.

After having biotinylated the APC, 200 pmol biotinylated protein were immobilized on 5 mg streptavidin-coated magnetic beads (Dynabeads M-280 Steptavidin) in a total volume of 500 µl SELEX buffer, according to the manufacturer's recommendation. The loaded and washed beads were stored at 4 °C until further usage.

3.3.4. Test SELEX against APC with the canonical DNA library

Having established all necessary techniques and materials, the first SELEX with the FT2-N42 library could be performed. The selection was carried out in SELEX buffer, based on Dulbecco's phosphate buffered saline (D-PBS, section 6.1.3, page 99) enriched with 1 mg/ml bovine serum albumin (BSA), 0.1 % TWEEN-20 and 0.1 mg/ml salmon sperm DNA. As the target, activated protein C (APC) immobilized on magnetic beads was used.

500 pmol of the start library was incubated with 500 µg of non-conjugated streptavidin beads to remove the unwanted matrix binding sequences. The supernatant was incubated for 30 min at 37 °C and 800 rpm with 500 µg APC loaded beads in a total volume of 200 µl. The beads were thoroughly washed two times (30 s and 5 min) with 200 µl SELEX buffer to remove non-specific binding DNA and the enriched library was eluted with 85 µl water from the beads by heat denaturation (80 °C for 3 min). The eluted DNA was PCR amplified in a total volume of 800 µl and purified via spin-columns (section 5.1.3, page 87). PCR product formation (18-22 PCR cycles) was carefully monitored by gel electrophoresis and stopped as soon as the desired product band could be detected on ethidium bromide stained 4 % agarose gels. Single strand displacement of the double-stranded PCR product was carried out by λ-exonuclease digestion as described in section 5.1.5 (page 88) and the resulting single strand DNA was purified via spin-columns (section 5.1.3, page 87). The purified single strand DNA (approx. 10 pmol) was applied in the next SELEX round.

For each selection cycle, the PCR product and the purified single strand DNA were analyzed by gel electrophoresis on a 4 % agarose gel. To increase the selection pressure, the total amount of washing steps was gradually increased up to a maximum of 8 washing steps during the selection. **Figure 3.6** depicts the PCR products obtained for all 11 selection rounds that were performed during the test selection.

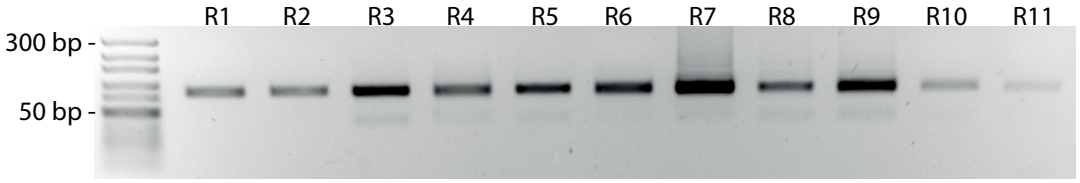

Figure 3.6. – PCR products of the test SELEX against APC
Ethidium bromide stained 4% agarose gel depicting the PCR products from all 11 SELEX rounds. For every selection round a band corresponding to the PCR product of the correct length could be obtained. Although the PCR amplification was stopped as soon as the desired product band could be detected on ethidium bromide stained agarose gels, the precise required number of PCR cycles was difficult to predict, resulting in varying product intensities throughout the SELEX. Despite the formation of PCR by-product could be detected from R3 onwards the selection was carried out for 11 rounds.

The first (R1), 7th (R7), and 11th round (R11) were analyzed for APC binding in a filter retention assay. The samples were labeled with a radioactive ^{32}P phosphate group at the 5' end and measured as described in section 5.5.1 (page 95).

Figure 3.7 summarizes the results of the filter retention assay. As expected, the library after 1 round of selection (gray squares) did not show binding to APC. After 7 rounds of SELEX (rose diamonds), a binding curve could be obtained. For the 11th round of selection (red dots), a further increase in binding affinity could be observed. Due to this sharp rise in binding affinity, it was decided to sequence the 11th round library to verify the enrichment on a monoclonal level.

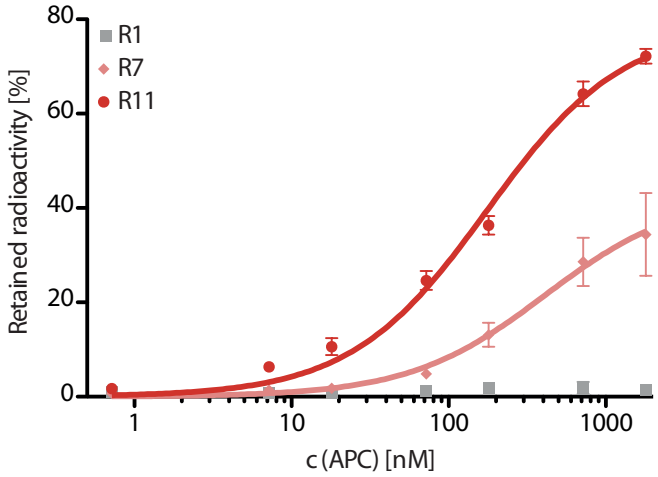

Figure 3.7. – Filter-retention analysis of the test SELEX against APC with canonical nucleobases After 1 round of selection (R1, grey squares) no binding to APC could be observed. The library of R7 (rose diamonds) showed binding to APC. After 4 additional rounds of selection, the R11 library (red dots) bound to APC with an increased binding affinity.

3.3.5. Identification of aptamer candidates from the test SELEX

3.3.5.1. Cloning and Sanger sequencing of the test SELEX

The enriched library of the 11th round was cloned into a vector and transformed into competent cells with the TOPO-TA Cloning kit (Life Technologies) according to the manufacturer's recommendation. Before cloning the sample was PCR amplified with *Taq* polymerase (instead of *Pwo* polymerase) as the addition of a 3' overhang deoxyadenosine is crucial for the cloning strategy. For the TOPO-TA cloning protocol, it is essential that unmodified primers, without phosphate group at the 5' end, are used. After plasmid preparation of successfully transformed colonies, 50 µl plasmid DNA (30 ng/µl) was sent to GATC-Biotech for Sanger sequencing, using the M13 reverse primer.

Figure 3.8 shows the results of the Sanger sequencing. For the full list of all sequencing results, please refer to **figure A.1** (page 103). Sanger sequencing revealed a strong enrichment of the library after 11 rounds of selection. Two sequences (F1 and F2) showed a very high prevalence in the library with 9 (F1) and 7 (F2) identical copies out of 20 sequences. In addition to those two dominating sequences, four orphan sequences (F3-F6) were found. Interestingly, all sequenced clones contained a common consensus motif. This 13 nt long motif (TATCCCGTATGGG) is identical to the previously described consensus motif identified in an earlier aptamer selection experiment against APC[133].

The strong enrichment of two sequence, both containing a consensus motif that was previously described in APC binding DNA aptamers, and the binding data of the 11th round library, together validated the selection strategy to be successful in selecting unmodified DNA aptamers.

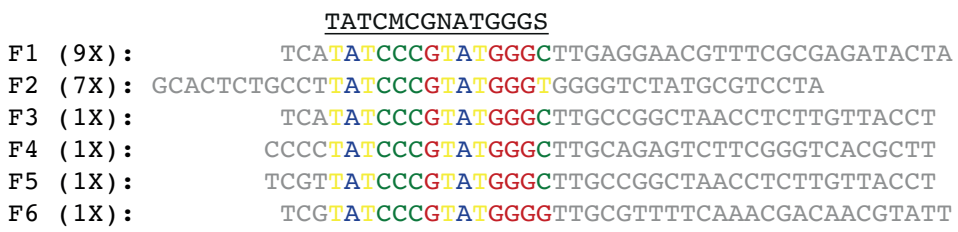

Figure 3.8. – Sanger sequencing results for the test SELEX against APC
Two dominant sequences (F1 and F2) could be identified. All sequenced clones contain a common consensus motif (TATCCCGTATGGG) which is in full accordance to the previously described APC binding motif (TATCMCGNATGGS) [133].

3.3.5.2. Next-generation sequencing (NGS) of the test SELEX

To get a better insight into the SELEX process, an NGS analysis of the successful test SELEX was performed using Illumina's sequencing by synthesis (SBS) technology.[136] With NGS, instead of only analyzing the last round of the selection, all rounds can be analyzed allowing for a better understanding of the enrichment process during SELEX.[63] The test SELEX was chosen as a reference for a selection with a very high degree of enrichment.

To multiplex several samples in one sequencing run, index sequences were added to the ends of the SELEX samples. Therefore, a set of 24 primers with 12 different index sequences at the 5' end was ordered. Each SELEX round was PCR amplified with a different set of primers allowing for multiplexing of 12 different SELEX rounds on one lane of the sequencer. After performing a PCR adding the index primers to each sample, the 12 samples were mixed, and Illumina adapters were ligated. For adapter ligation with some adaptations, the TruSeq DNA PCR-Free (LT) sample preparation kit was used. For further details, please refer to the following publication.[2]

Sequencing was performed on an Illumina HiSeq 1500 instrument in collaboration with the group of Prof. Joachim Schultze (LIMES, Bonn). 100 bp single end reads were performed. After sequencing, the data analysis of the raw sequencing data was done by Dr. Carsten Gröber at AptaIT (Munich) using the COMPAS software.[137] The COMPAS software generates a variety of outputs including basic statistics, but also analyzes the clustering of similar sequences to create so-called "pattern".

Unique sequences

A very basic form of analysis to quickly assess the success of a selection experiment is based on monitoring the amount of unique sequences found in each selection round over the course of the SELEX experiment. The reduction of unique sequences in the library can be used as a first indicator to assess the quality of enrichment and to define a good stopping point for the selection.

In **figure 3.9** the relative amount of unique sequences is plotted over the selection rounds. In the first five rounds of the SELEX, the majority of sequences are only found as one single copy. Over the next two selection rounds a very strong enrichment of sequences could be observed, with only around 20% of unique sequences left in the library. This strong enrichment is continued so that after eleven rounds of selection no unique sequences could be detected anymore.

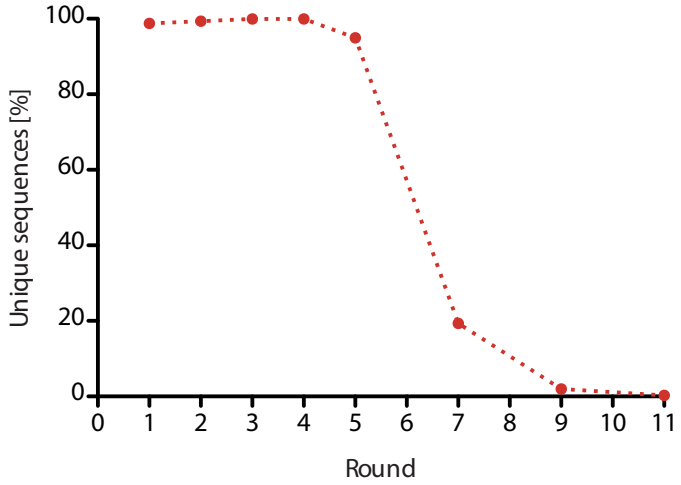

Figure 3.9. – Analysis of unique sequences of the test SELEX against APC
In the first five rounds, little enrichment of sequences could be observed. From round five onwards, a very strong enrichment was obtained. After eleven rounds of selection, no unique sequences could be detected. Rounds six, eight and ten were not sequenced.

Character statistics

A more detailed understanding of the selection process can be obtained by analyzing the frequency of each nucleobase at every position in the random region (character statistics). This analysis is also very valuable to assess the quality and the overall length-distribution of the start library.

The character statistics for the unmodified FT2-N42 start library, compared to the 11^{th} round library are depicted in **figure 3.10**. The starting library showed a relatively equal distribution of all four nucleobases over all positions of the random region. After eleven rounds of selection, a completely different nucleobase distribution was observed. For each position one dominant nucleobase had emerged, indicating a strong decrease in diversity of the library. The analysis of the character statistics is in good accordance with the analysis of the unique sequences, showing a low level of enrichment until round five and a strong enrichment from round seven onwards (**figure A.2**, page 104).

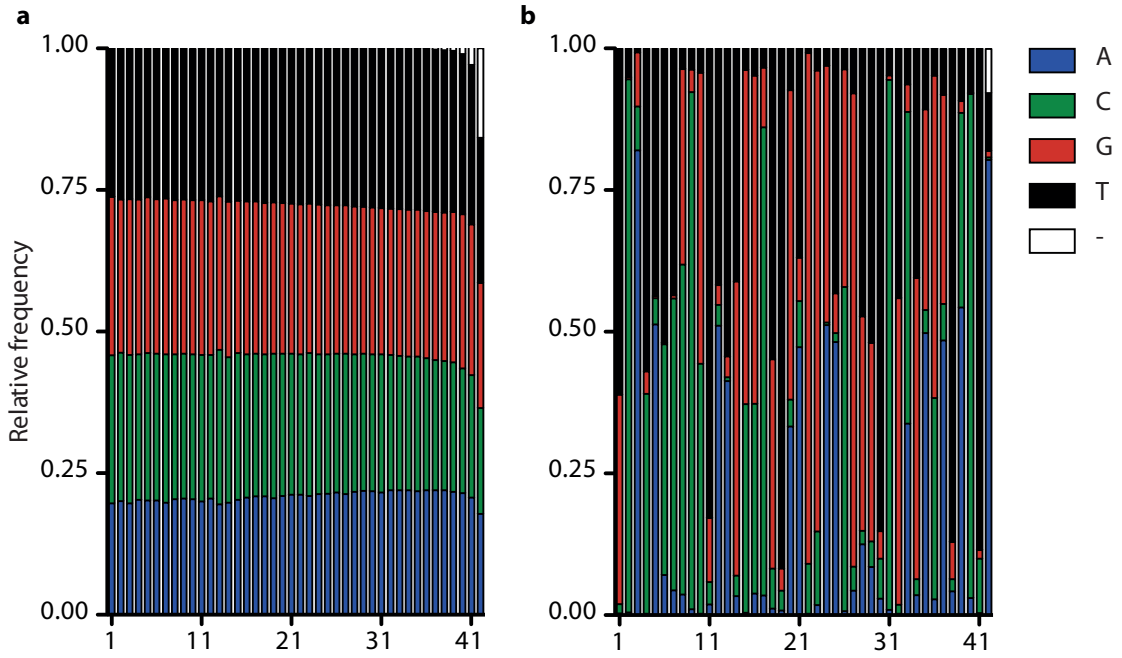

Figure 3.10. – Character statistics for the start library and the 11th round of the test SELEX
a) Statistical base distribution in the FT2-N42 start library, showing an equal distribution of all four nucleobases over all 42 positions of the random region. Approximately 20 % of the sequences display a shortened random region (white). b) Statistical base distribution over all 42 positions of the random region after eleven rounds of selection. For each position one dominant nucleobase has emerged, indicating a strong decrease in diversity of the library.

Pattern and tracing

Besides the basic statistics, one main feature of the COMPAS software is the ability to cluster similar sequences into families (”pattern”). Thereby, very similar families, which vary in only a few point mutations can be analyzed collectively. The COMPAS software also allows for tracking the frequency of sequences and patterns over several rounds of selection. With this so-called ”tracing” a frequency and an amplification fold profile for each aptamer candidate can be obtained. This profile can be very beneficial to screen for the most promising aptamer candidates, potentially reducing the amount of binding experiments required to identify aptamer sequences with the desired properties. The tracing profile can give valuable insights into the SELEX process and can thereby help to understand and optimize the parameters leading to a successful selection.

In **figure 3.11**, the tracing of the 15 most abundant patterns is depicted. In agreement with Sanger sequencing, the same two major sequence families (F1 and F2) could be identified as the dominant patterns. For the full list of all 15 patterns see **figure A.3** (page 104). As observed in the ”unique sequences” data, from the 7th round onwards a strong enrichment of patterns was revealed. In the 11th round library, the patterns corresponding to F1 and F2 made up roughly 70 % of the total sequences.

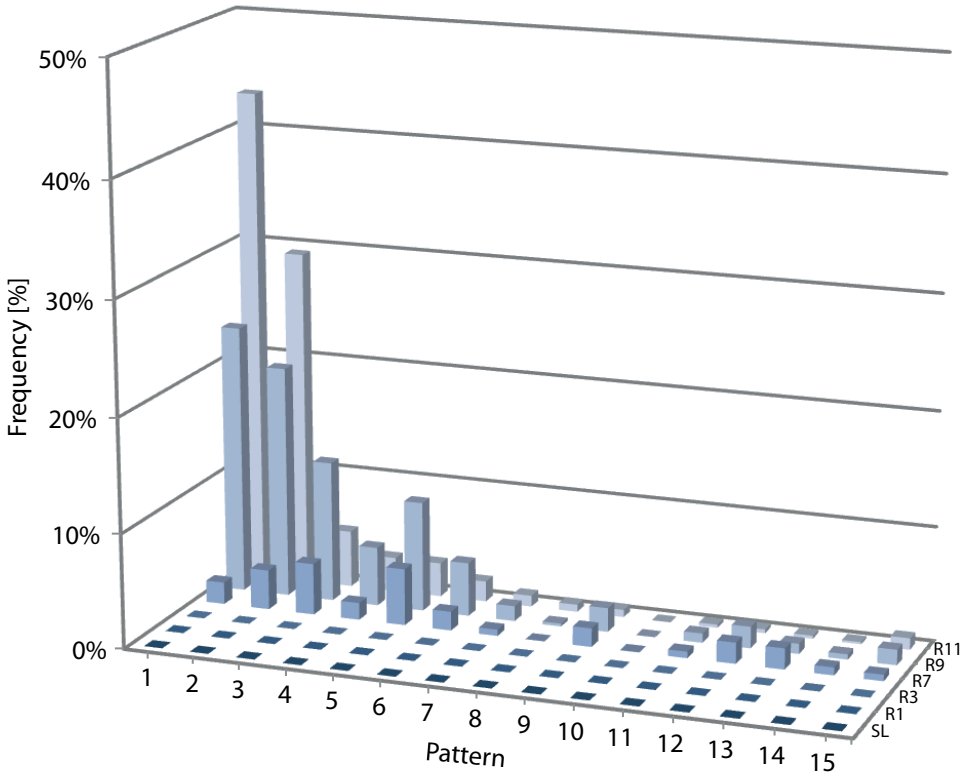

Figure 3.11. – Tracing of the 15 most abundant patterns of the APC test SELEX
Pattern 1 and 2 correspond to the same two dominant sequence families (F1 and F2) previously identified by Sanger sequencing. Together, those two families made up roughly 70 % of the total 11^{th} round library, illustrating the very high degree of enrichment achieved during the selection.

In agreement with the Sanger sequencing and binding data, the NGS data showed a strong enrichment of the FT2-N42 library towards APC. Two dominant aptamer sequences (F1 and F2) could be identified, both containing the previously described APC binding motif. Having demonstrated the capability of the newly developed library and methods to generate a canonical DNA aptamer for APC, the next step, namely the introduction of nucleobase-functionalizations into the SELEX process, could be approached.

3.4. SELEX with C8-dU

As the first test-selection against APC was successful, validating the chosen SELEX conditions, a SELEX with modified nucleotides could be attempted. In this case, a similar DNA library was used, with the only difference being that the canonical thymidine residues were replaced by the alkyne modified C8-dU nucleotides.

3.4.1. Library design of the FT2-N42-C8-dU library

As the non-modified FT2-N42 library was designed with the future application in mind, except for the substitution of all thymidines with the respective alkyne-modified C8-dU nucleotide[123], no further modifications to the library design were necessary. A schematic representation of the FT2-N42-C8-dU library is depicted in **figure 3.12**.

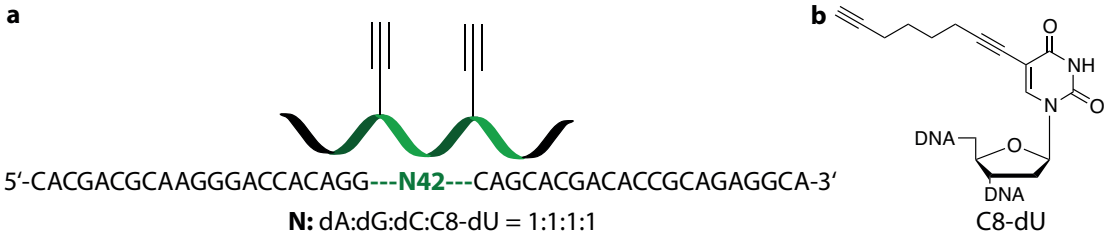

Figure 3.12. – Schematic representation of the FT2-N42-C8-dU library
a) The library is composed of a 42 nt random region (green) flanked by two 21 nt primer binding sites (black). To avoid potential problems in the PCR, associated with the alkyne modification, the primer binding sites were designed without any C8-dU nucleotides. In the random region an equal distribution of all four nucleotides was desired yielding a theoretical statistical distribution of 10.5 functionalization sites per strand. b) Structure of alkyne-modified C8-dU nucleotide.

3.4.2. Amplification of the FT2-N42-C8-dU library

The first step towards the adaptation of the alkyne-library for SELEX was to verify the successful incorporation of the C8-dU nucleotide by PCR and the efficient single strand displacement of the modified antisense strand by λ-exonuclease.

3.4.2.1. Polymerase chain reaction with C8-dU

It had been shown that the C8-dU triphosphate could be incorporated into defined template strands by *Pwo* polymerase[124], but as of yet no PCR with random libraries had been described. However, using the same PCR conditions described in section 3.3.2.1 (page 22) equal product formation with canonical thymidine triphosphate and with C8-dU-triphosphate could be observed (**figure 3.13**).

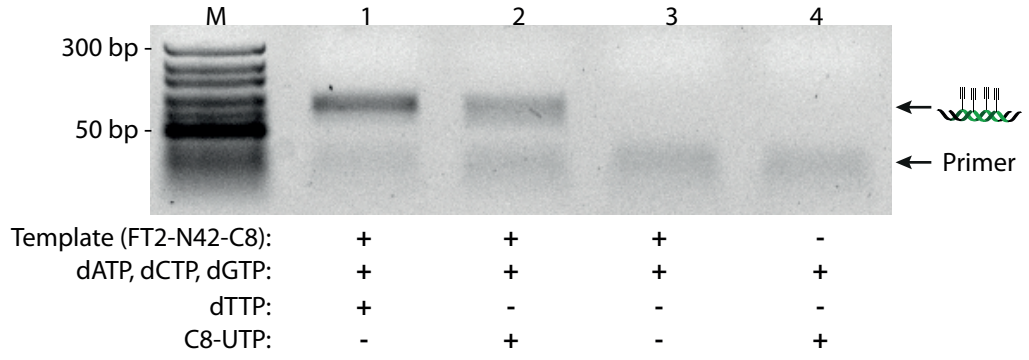

Figure 3.13. – PCR amplification of the FT-N42-C8-dU library with C8-dU containing dNTPs
Ethidium bromide stained 4% agarose gel. 1) PCR products for the FT2-N42 library with canonical nucleotides . 2) PCR products for the FT2-N42-C8-dU library with the C8-dU nucleotide. 3) Control PCR reaction for the FT2-N42-C8-dU library without the C8-dU nucleotide. No double stranded DNA product is detectable. 4) "No template control" (NTC) PCR reaction without the addition of the template DNA. No double stranded DNA product is detectable.

3.4.2.2. λ-Exonuclease digestion of C8-alkyne-DNA

Being able to amplify the C8-dU containing library by PCR, the next step was to generate single stranded DNA needed for the next selection round. As expected, an efficient digestion by λ-exonuclease was achieved.

Figure 3.14 shows the results for the λ-exonuclease digestion of the C8-alkyne containing library in comparison to the non-functionalized library. Both libraries could be efficiently digested by the enzyme enabling the use of λ-exonuclease digestion as a method for the single strand displacement of alkyne-modified libraries during click-SELEX.

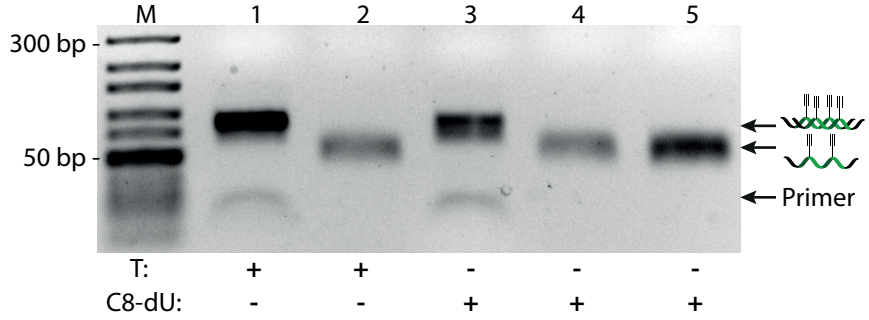

Figure 3.14. – λ-exonuclease digestion of the FT2-N42-C8-dU library
Ethidium bromide stained 4% agarose gel. 1) Unmodified FT2-N42 library before the treatment with λ-exonuclease. 2) Unmodified FT2-N42 library after the treatment with λ-exonuclease. 3) FT2-N42-C8-dU library before the treatment with λ-exonuclease. 4) FT2-N42-C8-dU library after treatment with λ-exonuclease. 5) 3 pmol of solid-phase synthesized ssDNA library (FT2-N42-C8-dU) loaded as a reference.

3.4.3. Functionalization of the FT2-N42-C8-dU library by click chemistry

A key step towards a functional click-SELEX protocol is the capability to functionalize a DNA library reliably before the selection step. For the establishment and validation of an efficient click-reaction protocol, a 16 nt long test-oligonucleotide (Test-C8-dU: 5'-GCACTGT-C8-dU-CATTCGCG-3'), similar to a previously described test-sequence, was used.[123] As a starting point for the optimization recently described click-conditions[120] were used, utilizing the water-soluble THPTA ligand instead of TBTA. For further details please refer to section 1.4.3.2 (page 15).

To establish the click-SELEX protocol, the focus had to be put on one proof of concept modification. From the examples published by SOMAlogic, it was known that empirically the functional groups that produce the best nucleic acid ligands typically have hydrophobic aromatic character.[78] These types of side-chains are also commonly utilized by antibodies in their recognition of protein antigens.[138,139] The indole group, also found in the side-chain of the amino acid tryptophan, was described to increase the success rate of aptamer identification against proteins strongly. Based on this observation, 3-(2-azidoethyl)indole (hereinafter referred to as **1**) (**figure 3.15**) was chosen as first functionalization for click-SELEX.

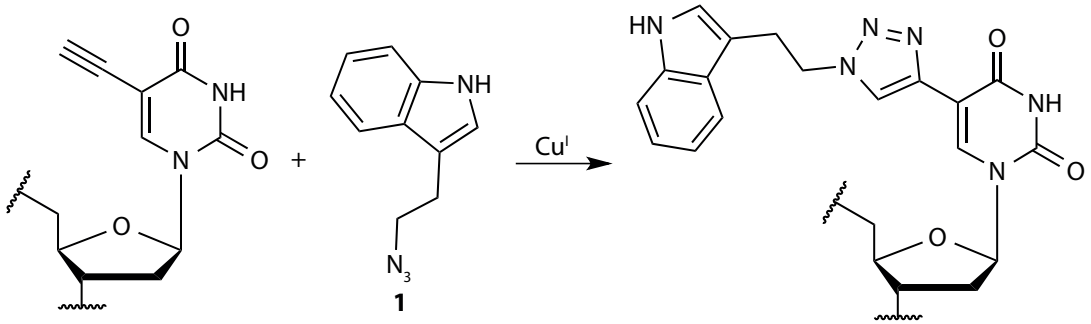

Figure 3.15. – Indole functionalization of alkyne-DNA
Alkyne-modified DNA is reacted with 3-(2-azidoethyl)indole (**1**) to form the functionalized library for click-SELEX.

Using the following optimized click-conditions a full functionalization of the Test-C8-dU oligonucleotide with indole azide **1** could be observed by high-performance liquid chromatography coupled to mass spectrometry (HPLC-MS).

A 1 mM Cu^I catalyst solution was freshly prepared by addition of 25 µl freshly prepared 100 mM sodium ascorbate solution to 75 µl 1:4 $CuSO_4$/THPTA solution (1 mM and 4 mM final concentration respectively). The click reactions were performed in 100 µl final volume. Therefore, 10 µl of a 10 mM azide solution in DMSO was added to 80 µl of DNA solution in 100 mM phosphate buffer (pH 7.0). The click reaction was started by addition of 10 µl catalyst solution, incubating the reaction mixture for 60 min at 37 °C and 800 rpm.

After incubation the samples were purified and buffer exchanged to water with Amicon Ultra centrifugal filters (3 kDa cutoff) according to the manufacturer's recommendation. Samples of the Test-C8-dU oligonucleotide before and after functionalization were analyzed by HPLC-MS as described in section 5.3.1 (page 91). **Figure 3.16** summarizes the results for the functionalization of the test-oligonucleotide with indole azide **1**.

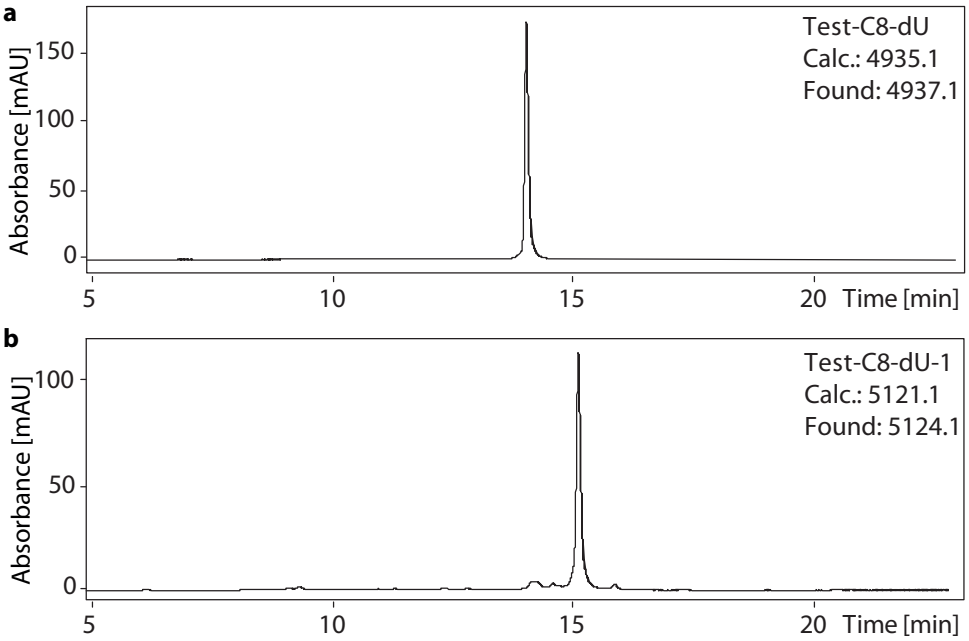

Figure 3.16. – HPLC-MS analysis of the click-reaction with Test-C8-dU
a) Prior functionalization the test-oligo (Test-C8-dU) eluted after 14.0 min. A mono-isotopic mass of 4937.1 could be detected. b) After functionalization of the test-oligonucleotide with indole azide **1**, the DNA (Test-C8-dU-1) eluted after 15.1 min. Near quantitative conversion to the desired product with a mono-isotopic mass of 5124.1 could be detected.

After having established an efficient click-reaction protocol, capable of full functionalization of the Test-C8-dU oligonucleotide, the next step was to validate the protocol with libraries containing multiple modifications per strand. Using the same click-conditions (section 5.2.1, page 90) the FT2-N42-C8-dU DNA library with statistically 10.5 alkynes per strand could be fully functionalized. As HPLC-MS can not directly analyze complex DNA mixtures, the library was enzymatically digested to the individual nucleosides as described in section 5.2.4.1 (page 91).

Briefly, 3 µl of 10 x S1 nuclease reaction buffer (Life Technologies) was added to 200 pmol DNA in 27 µl water. After addition of 0.5 µl S1 nuclease (100 U/µl), the samples were incubated for 60 min at 37 °C and 800 rpm. 3.5 µl 10 x alkaline phosphatase buffer (Promega) and 0.5 µl alkaline phosphatase (CIAP) (1 U/µl), 0.5 µl snake venom phosphodiesterase I (5 U/µl) and 0.5 µl Benzonase nuclease (250 U/µl) were added. The samples were incubated for 120 min at 37 °C and

800 rpm. After digestion, the samples were heated to 95 °C for 3 min and centrifuged for 3 min at 12000 rcf. Samples of the FT2-N42-C8-dU library before and after functionalization were analyzed by HPLC-MS as described in section 5.3.2 (page 92).

The results for analytical HPLC-MS runs with the digested FT2-N42-C8-dU library, before and after functionalization are depicted in **figure 3.17**.

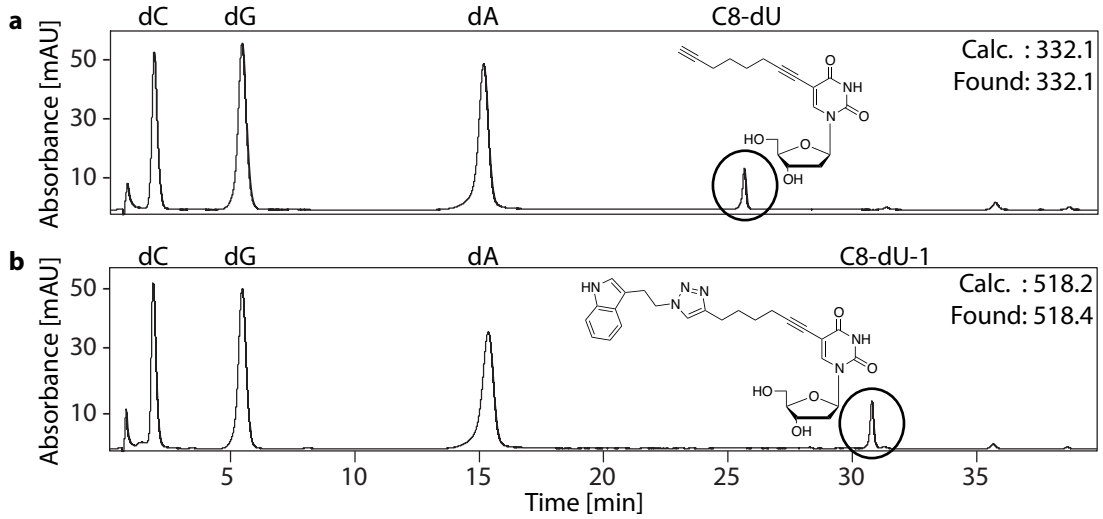

Figure 3.17. – Investigation of the click reaction with the C8-dU containing library
a) HPLC-MS analysis of the C8-dU start library before CuAAC functionalization with indole azide **1**. To assess the yield of functionalized C8-dU the samples were enzymatically digested to single nucleosides. b) After the reaction, full conversion to the product (C8-dU-1) could be observed. Even by mass spectrometry no unreacted C8-dU could be detected.

3.4.4. Adaptation of the magnetic bead-based selection method

In the first selection attempt with the C8-dU modified library, the same selection method and conditions as for the successful test SELEX were used. However, for the functionalization of the library with the desired indole moiety, the previously established additional click-reaction step had to be included (section 3.4.3, page 34).

3.4.5. Magnetic bead-based SELEX against APC

The selection was carried out with the same conditions as the test-selection. 500 pmol of the indole-functionalized start library was incubated with 500 µg of non-conjugated streptavidin beads to remove the unwanted matrix binding sequences. The supernatant was incubated for 30 min at 37 °C with 500 µg APC loaded beads in a total volume of 200 µl. The beads were thoroughly washed two times (30 s and 5 min) with 200 µl SELEX buffer to remove non-specific binding DNA and the enriched library was eluted with 85 µl water from the beads by heat denaturation (80 °C for 3 min).

In the first selection attempt with the indole-functionalized C8-dU library, the enriched library was amplified in a total volume of 800 µl PCR reaction. However, due to the strong dilution of the DNA the PCR failed to produce enough material for the next selection round in an acceptable number of PCR cycles (compared to a control PCR reaction without added template DNA).

To overcome the amplification problems associated with the very small amount of eluted DNA, a two-step PCR protocol was employed. Therefore, all eluted DNA was first amplified in a 100 µl PCR reaction and the resulting PCR product was subsequently further amplified in a second 800 µl PCR reaction. PCR product formation was always carefully monitored by gel electrophoresis and stopped as soon as the desired product band could be detected on ethidium bromide stained 4 % agarose gels.

Single strand displacement of the double-stranded PCR product was carried out by λ-exonuclease digestion, and the resulting single strand DNA was purified via spin-columns. The purified single strand DNA was indole-functionalized as described in section 5.2.1 (page 90) and applied in the next SELEX round. To increase the selection pressure, the stringency of the washing steps was gradually increased up to a maximum of eight washing steps during the selection.

Figure 3.18 summarizes the results obtained after the PCR amplification from each individual SELEX round. From round three onwards by-product formation could be detected, including a shorter second band, possibly representing a shorter PCR by-product or secondary structure. Regardless of this by-products, the SELEX was continued for a total of ten rounds.

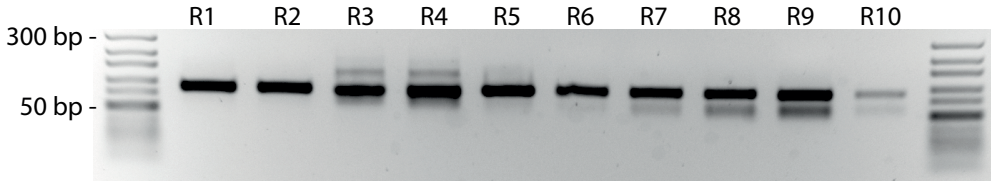

Figure 3.18. – PCR products for bead-based C8-dU-SELEX against APC
Ethidium bromide stained 4% agarose gel depicting the PCR products from all ten SELEX rounds. For every selection round, an intense band corresponding to the PCR product of the correct length could be obtained. From round three onwards by-product formation could be detected, including a shorter second band, possibly representing a shorter PCR by-product or secondary structure. Due to increasing by-product formation the library of round ten was subjected to fewer PCR cycles.

To investigate the success of the selection, the indole-functionalized library after ten rounds of selection was analyzed for APC binding in a filter retention assay. As controls, the functionalized library after one round of selection and the non-functionalized library of the 10^{th} round were prepared. The samples were labeled with a radioactive ^{32}P phosphate group at the 5' end and measured as described in section 5.5.1 (page 95).

Figure 3.19 depicts the results of the filter retention assay. In contrast to the previous test SELEX with canonical DNA, in this selection attempt, no improvement of binding affinity was observable. However, an enhanced maximal binding capacity from 30 % for R1-1 to 60 % for R10-1 was observed.

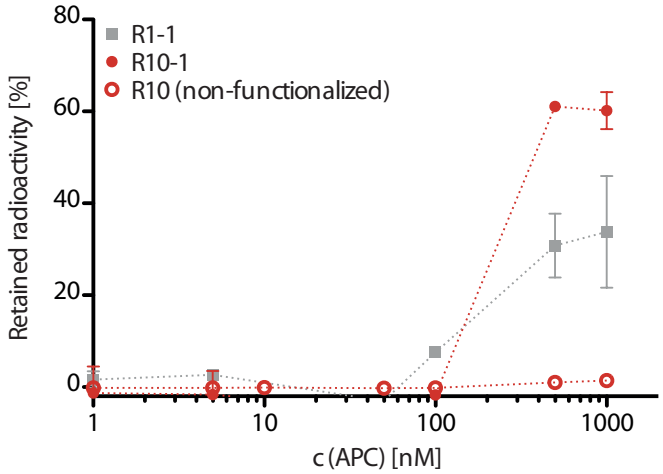

Figure 3.19. – Filter retention assay of the bead-based SELEX against APC
The functionalized library after 10 rounds of selection (R10-1, filled out red circles) did not show a higher binding affinity than library after one round of selection (R1, grey squares). For the non-functionalized 10^{th} round library (R10 non-filled out red circles) no binding to APC could be observed. An enhanced maximal binding capacity from 30 % for R1-1 to 60 % for R10-1 was observed.

3.4.6. Identification of aptamer candidates from the bead-based SELEX

3.4.6.1. Cloning and Sanger sequencing of the bead-based SELEX against APC

To further analyze the outcome of the selection, the library of the 10^{th} round was cloned into a vector and transformed into competent cells with the TOPO-TA Cloning kit (Life Technologies) according to the manufacturer's recommendation. Before cloning the sample was PCR amplified with *Taq* polymerase. After plasmid preparation of successfully transformed colonies, 50 µl plasmid-DNA (30 ng/µl) was sent to GATC-Biotech for Sanger sequencing, amplified with the M13 reverse primer.

In agreement with the shortening of the library observed during the selection, only 8 of the 17 sequenced clones showed a full length 42 nt random region. All of the eight full-length sequences were orphans, so that no promising aptamer candidates could be selected from this data set (**figure 3.20**). A complete list of all sequencing results is depicted in **figure A.4** (page 105).

```
C1:  GCATACCCCGAGCCTAACTGCGAAGCAATGCCAGCGACTGGC
C2:  AAAAGAGGGGGGAGGCGGGAGCGCACGACTTTGGGAACCCGA
C3:  GAGGCGGACACCGCGCGGAGCTGAGGACGCCGAATCGCTTCA
C4:  GCTCGATCGAGGTAAAAAGAGGGAGCCGACGGGAGCCTGGAA
C5:  GGAGACCAAGGGCGGGGAGAGGCCGCACGGCTTTCGGAAGCG
C6:  TAGGCCGCGGGAACTGGGGCCAACTGTCGGCAGCCCAAGTTG
C7:  AGCAGAGCCGGCGAGGCTGAACCCTAGAGACGCGATCATCGG
C8:  GGCGAGTATGGCCATCCGGCGCGCCAAGAGAAGAATACAAAG
```

Figure 3.20. – Sequencing results for the bead-based SELEX against APC
Eight full-length sequences could be obtained. As none of the sequences was detected in multiple copies, no promising aptamer candidates could be selected from this data set.

3.4.6.2. Next-generation sequencing of the bead-based SELEX against APC

Although the filter retention analysis and the Sanger sequencing data of the 10th round library suggested that no enrichment could be obtained, to get a better insight into the SELEX process, a NGS analysis as described in section 3.3.5.2 (page 28) was performed.

Unique sequences

Figure 3.21 shows the analysis of unique sequences over all sequenced selection rounds. In contrast to the previous selection, even after ten rounds of SELEX, only a moderate decrease in the relative frequency of unique sequences down to approximately 70 % could be observed. Together with the lack of increase in binding affinity, this data suggest that not sufficient enrichment of the library could be obtained during the ten rounds of selection.

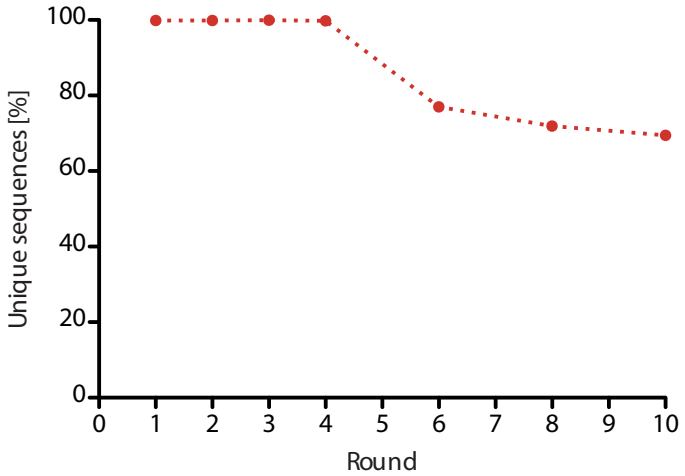

Figure 3.21. – Analysis of unique sequences for the C8-dU-SELEX against APC
After ten rounds of SELEX, only a moderate decrease in the relative frequency of unique sequences could be observed.

Character statistics

The character statistics for the C8-dU start library and the 10th round library are depicted in **figure 3.22**. The start library shows an equal distribution of all four nucleobases over all positions of the random region. However, the relative amount of EdU nucleotides is only around 20 %, resulting in an average amount of 8.5 C8-dU nucleotides per strand. After ten rounds of selection, no position-specific enrichment of nucleobases could be observed. However, a strong decrease in the overall sequence length of the library could be observed, indicating an undesired selection pressure towards shorter sequences. For additional character statistics see **figure A.5**, page 105.

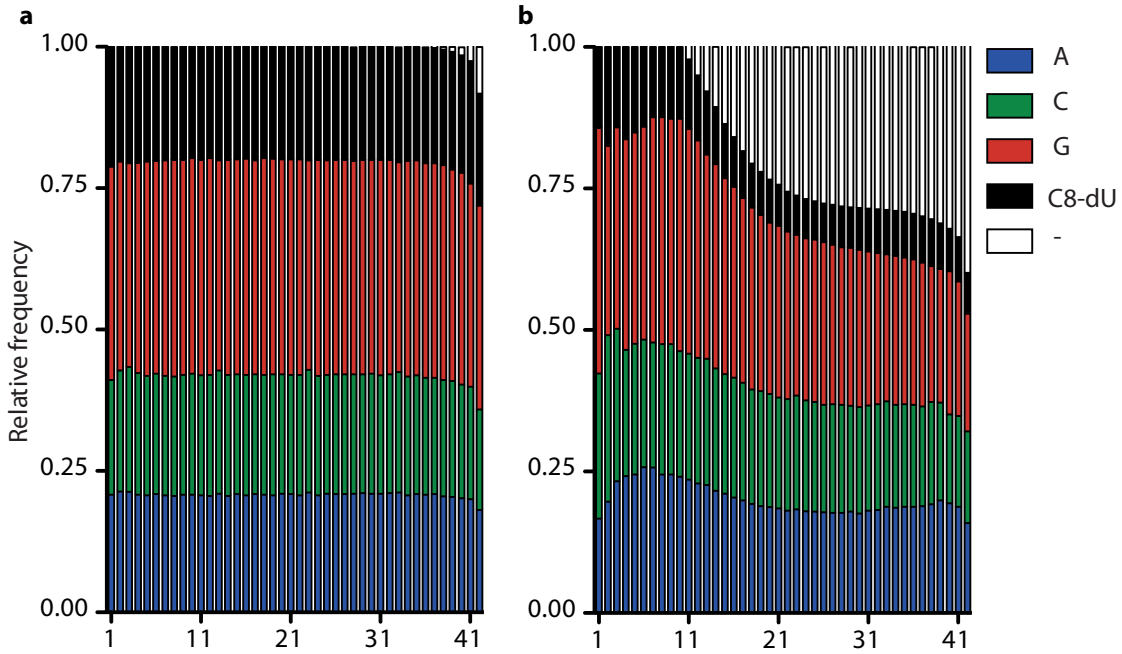

Figure 3.22. – Character statistics for the start library and the 10th round of the APC-SELEX
a) Statistical base distribution in the C8-dU start library, showing an equal distribution of all four nucleobases over all 42 positions of the random region. b) Statistical base distribution over all 42 positions of the random region after ten rounds of selection. Even after ten rounds of selection no position-specific enrichment of nucleobases could be observed. In contrast, a shortening of the random region was observed.

Pattern and tracing

In contrast to the test SELEX, in this case, no full-length pattern could be identified. This lack of potential aptamer candidates is in agreement with the other sequencing results, which all indicate that no successful enrichment could be obtained.

3.4.7. Validation of aptamer candidates from bead-based SELEX

Although sequencing data and the binding data of the 10th SELEX round indicated that this selection attempt did not yield high-affinity aptamer sequences, a filter retention analysis of all full-length sequences obtained by Sanger sequencing was performed.

Figure 3.23 summarizes the results of binding assays with all full length sequences. From all tested sequences only clones C4-1, C3-1 and C5-1 showed signs of a concentration-dependent binding in the range of the tested concentrations. Based on these results additional experiments with the most promising aptamer candidate C4-1 were performed. In contrast to the previous experiment, including only the monoclones, in this experiment C4-1 was compared to the functionalized first round library (Round 1-1) as a control. In comparison to this control, no differences between the library and C4-1 could be identified, showing that no high-affinity aptamer candidate was obtained.

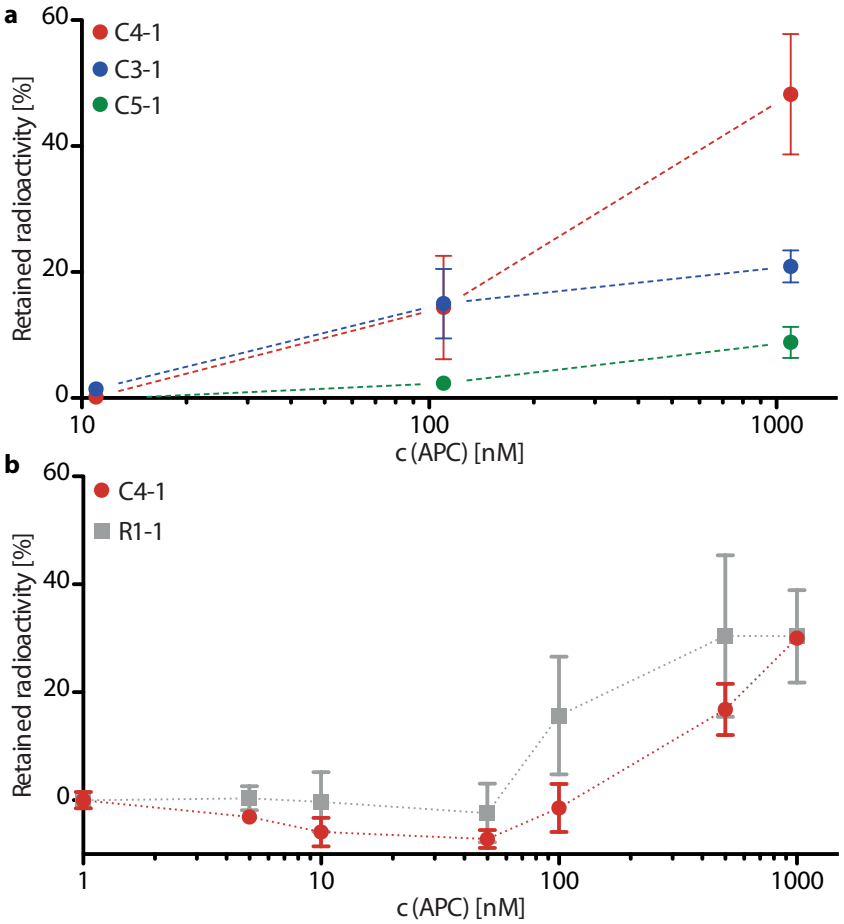

Figure 3.23. – Filter retention assay for monoclones from the bead-based APC SELEX
a) Only C4-1, C3-1, and C5-1 showed a concentration-dependent binding in the range of the tested concentrations. b) In direct comparison with the first round library (R1-1), clone C4-1 did not show any increased binding affinity towards APC.

In summary, no enrichment of the FT2-N42-C8-dU library and, therefore, no identification of any aptamers could be achieved using the magnetic bead-based selection method.

3.4.8. Change of selection method to capillary electrophoresis (CE-SELEX)

As insufficient selection pressure was applied in the previous selection, a method to generate higher selection pressure through better separation of the bound complex species from the non-binding sequences was desired. To achieve this goal, a shift from a bead-based separation to a capillary electrophoresis-based separation was planned.

As described in section 1.2.3.2 (page 6), capillary electrophoresis offers a very high degree of separation, while operating in solution. Thereby, non-specific background binding of the library to the matrix can be reduced. The CE-SELEX was performed in collaboration with Nasim Shahidi-Hamedani and Jens Müller from the Institute of Experimental Haematology and Transfusion Medicine (IHT) in Bonn.

3.4.9. CE-SELEX with FT2-N42-C8-dU library against APC

The indole-functionalized C8-dU starting library (3 µM final concentration) and APC (3 µM final concentration) were incubated in B15 buffer (Tris [25 mM], NaCl [30 mM], KCl [1 mM], $CaCl_2$ [1 mM], $MgCl_2$ [1 mM], pH 8.3). This buffer system was developed by Nasim Shahidi-Hamedani, containing salt concentrations suitable for capillary electrophoresis [58].

Three consecutive 20 min separation runs were performed as described in section 5.4.2 (page 93) and collected in 80 µl B15 buffer. The eluted DNA (enriched library) was PCR amplified as described in section 5.1.2 (page 87). The first PCR product was divided into 16 aliquots and further PCR amplified. As a backup of this SELEX round, 10 % of the purified dsDNA was stored at -20 °C. The purified amplified double stranded library was processed to single strand DNA as described in section 5.1.5 (page 88). Before the next SELEX round the ssDNA was functionalized with indole azide **1** as described in section 5.2.1 (page 90).

In total five rounds of SELEX, following the same protocol as in the first round, were performed. However, to increase the selection pressure the DNA and APC concentration were reduced as summarized in table 3.1.

Table 3.1. – Summary of protein and DNA concentrations used for CE-SELEX.

Round	APC [µM]	DNA [µM]
1	3	3
2	0.23	3
3	0.3	0.3
4	0.1	0.06
5	0.3	0.12

Figure 3.24 depicts the PCR products obtained from all five selection rounds that were performed during the CE-selection.

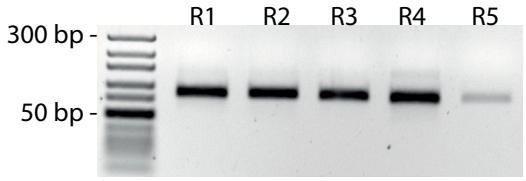

Figure 3.24. – PCR products for CE-SELEX against APC
Ethidium bromide stained 4% agarose gel depicting the PCR products from all 5 SELEX rounds. For every selection round an intense band corresponding to the PCR product of the correct length could be obtained. Due to increasing by-product formation the library of fifth round was subjected to fewer PCR cycles.

Figure 3.25 shows the electropherograms for the first three SELEX rounds. Although the concentration of the DNA and the protein were reduced from round to round, an increased signal with the retention time of the complex (10-12 min) was measured. This increase in signal intensity, indicative for an increase in complex formation, suggested an enrichment of the indole-functionalized FT2-N42-C8-dU library towards APC binding.

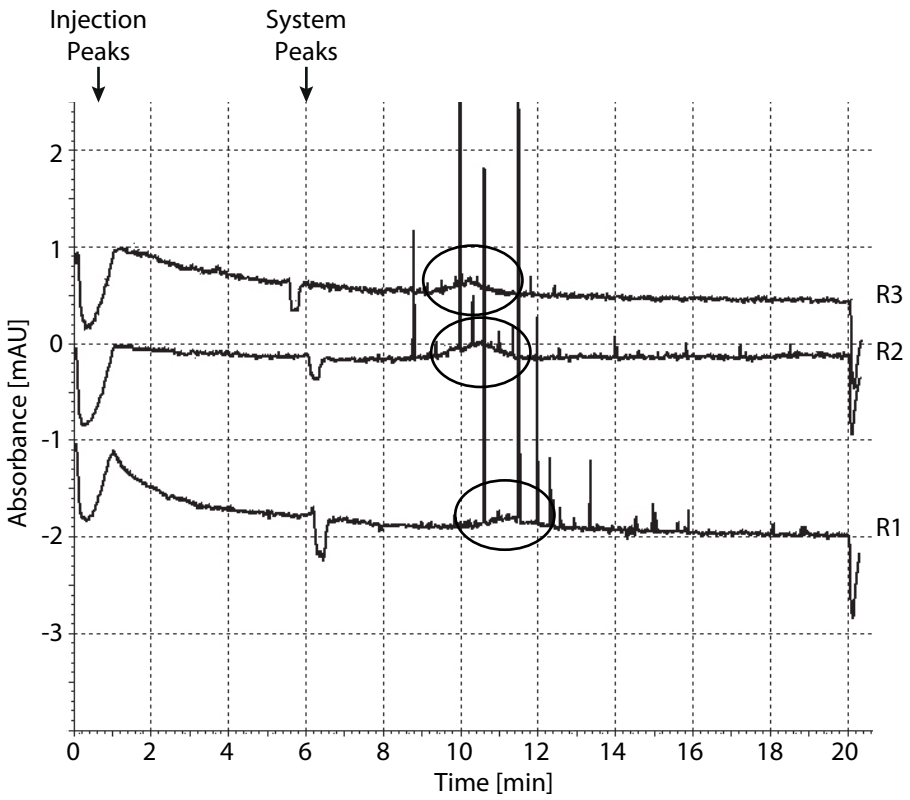

Figure 3.25. – Electropherograms for the first three SELEX rounds of the CE APC-SELEX
Although the concentration of the DNA and the protein were reduced from round to round (table 3.1), an increased signal with the retention time of the complex (10-12 min, black circles) was measured. The spikes in the electropherograms represent measuring artefacts due to e.g. air bubbles.

3.4.10. Identification of aptamer candidates from CE-SELEX

3.4.10.1. Cloning and Sanger sequencing of CE-SELEX

As the electropherograms of the selection indicated a potential enrichment of APC binding sequences, the 5th round library was cloned into plasmids and 30 monoclones were sent to GATC for Sanger sequencing.

The full list of all sequencing results is listed in **figure 3.26** (page 45). Surprisingly, no enrichment of sequence families or motifs could be detected. As the relatively small sample size of 30 sequences was not enough to identify potential aptamer candidates, a larger sample size was needed. Therefore, a NGS analysis of all five selection rounds was prepared.

```
C1:  CGATGGATAGGCTCTATCAGGGGTTGGCTGGTGTACGGTGGA
C2:  CTGTATAAGGCGAATAGGTGCAGTAACCTATCTGGCTGGCTG
C3:  GTTGCATAAAGGGTGAAGGGTTTAGGGGGTTGAAATTGGGAA
C4:  CTACATAGGGGGTTAGTGAGTAGGGGGGTAGCACTTGTGGTC
C5:  CTGTGAACCAAGGTCTAGCTTGCTATCCTGGGACATCCAGGA
C6:  GCGGGGGGCGAATCGATGTGGATGACGTGCCAAGCAGCTAAA
C7:  CCAACTGATGATAAACAGGATACAGCATTGGGGATGCCCAAG
C8:  ACGGCTAATGCAAA
C9:  GCGGCTAATGCAAA
C10: GTTTAAGGGGTGTTAAATGGCTCAGGCTGGATGGCTAATACT
C11: CTGGCCATTCGGAGGGTGTGAACTTCCACAGCTTGATCAAG
C12: GCAGGAGGGGTATGGCTTAACAGGGCGTGGGGGTTGGTTTCG
C13: TGATCATTATGGTAAGACGCCGAATGGGCAGGTTTGGACGGG
C14: CTGATTAAAGTTTGTGCCTCAGGGCGCGTGGTAACAGTGTGG
C15: CTGGATGAGTAGGTGTGGTAGGGTGGAGGGGGGCACTCCAAC
C16: CTACATAGGGGGTTAGTGAGTAGGGGGGTAGCACTTGTGGTC
C17: CAAAGCGGGCCATGAGATGGTAGGTAGCCAAGGCTAAGTAA
C18: GTGACACGTGGCAGAGGGTTGGAGTAGTAAGGGGTAGGTGCA
C19: GTGTGCATTAGACGATTGCACTGGGATTGATAGGGTGTTGGC
C20: TGGCATAACTAGCAGTTATTCAACTTGGGGGTTGGCGAGGGG
C21: GTTAGGACGGAATAG
C22: CTAGACGAAGTTGTATTATGATGAACTTAGGGTGGGGCAGTT
C23: GTGGAAGCGTGTCGACAAGGTAGGCTGGGAGGTTAGAAGCG
C24: TAAAGAAGGGGGTAGCATAACCTAAGTTGGTTCTGGGGCTGG
C25: GCTTAAATGGTGCGTTGCCAGGGTCTGAGGTTTCGCCCCGGG
C26: GTACGATGGGGGCTAATCCAGCTACATGGAGTGTCAAGGGGG
C27: GTAGGACGCGGTACGTGAGCCCTGTGCGTTGCCCCGTGCACG
C28: CTGGCACATGACTACGTTAAAATGGCCTCGCCTTTAAGCTAA
C29: CTGTAGTAGGTCGTCTAAATCAGGGGGTGTCCAACGTGGGGG
C30: CTGAGCAGGGCTAGGGTAGGCGTTGATCAGCAGGCAGCTTCA
```

Figure 3.26. – Sequencing results for the CE-SELEX with C8-dU against APC
The sequence of 27 full-length mono-clones could be identified from the library of the 5th SELEX round.

3.4.10.2. Next-generation sequencing of CE-SELEX

In the hope of obtaining a better understanding of the outcome of the CE-SELEX a NGS analysis, as described in section 3.3.5.2 (page 28), was performed. Each SELEX round was PCR amplified with a different set of primers allowing for multiplexing of 12 different SELEX round on one lane of the sequencer. After indexing PCR, the 12 samples were mixed, and Illumina adapters were ligated. For adapter ligation with some adaptations, the TruSeq DNA PCR-Free (LT) sample preparation kit was used.

Sequencing was performed on an Illumina HiSeq 1500 instrument in collaboration with the group of Prof. Joachim Schultze (LIMES, Bonn). 100 bp single end reads were performed. After sequencing the data analysis of the raw sequencing data was done by Dr. Carsten Gröber at AptaIT (Munich) using the COMPAS software.

Unique sequences

As the first indicator for a successful enrichment, the relative frequency of unique sequences was analyzed (**figure 3.27**). In contrast to the previous test selection with canonical nucleotides, no clear decrease in the relative frequency of unique sequences could be observed. Even after five rounds of CE-selection still almost 80 % of the library were present as a unique sequence, indicating that no enrichment of specific APC binding sequences could be obtained.

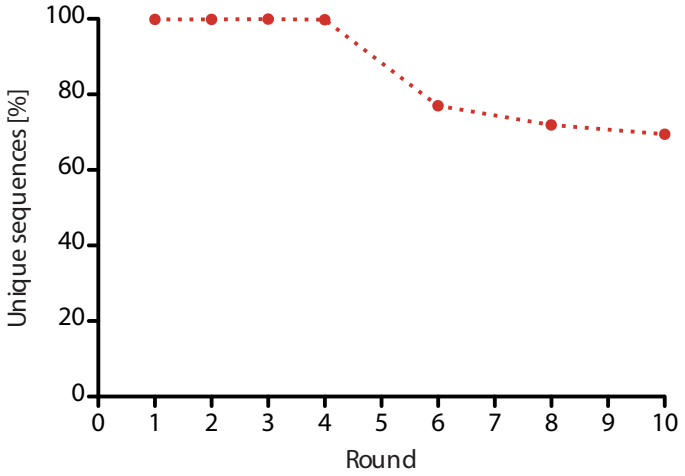

Figure 3.27. – Analysis of unique sequences for the CE-SELEX against APC
No clear trend could be observed. Even after five rounds of CE-selection still almost 80 % of the library were present as a unique sequence, indicating that no enrichment of specific APC binding sequences could be obtained.

Character statistics

The character statistics for the C8-dU start library and the round 5 library are depicted in **figure 3.28**. The start library showed an equal distribution of all four nucleobases over all positions of the random region. However, it is noticed that the relative amount of EdU nucleotides was only around 20 %, resulting in an average amount of 8.5 C8-dU nucleotides per strand. Even after five rounds of CE based selection, no position specific enrichment of individual nucleotides could be observed. For additional character statistics please refer to **figure A.6** (page 106).

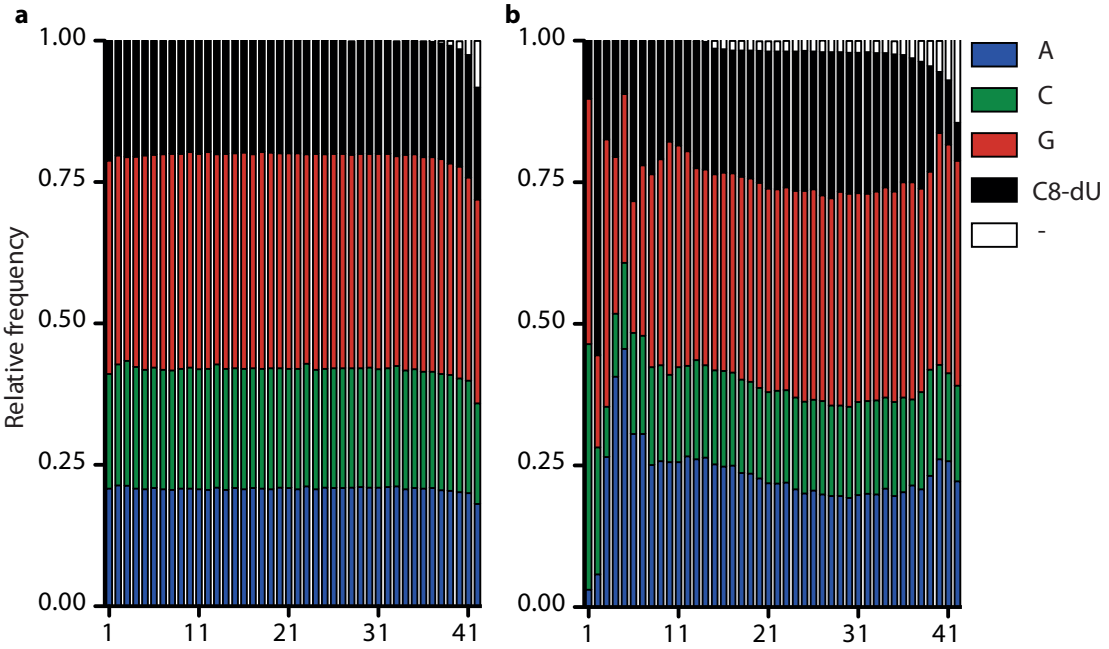

Figure 3.28. – Character statistics for the start library and the 5th round of the CE-APC-SELEX
a) Statistical base distribution in the C8-dU start library, showing an equal distribution of all four nucleobases over all 42 positions of the random region. b) Statistical base distribution over all 42 positions of the random region after five rounds of selection. Even after five rounds of CE based selection, no position specific enrichment of individual nucleotides could be observed.

Pattern and tracing

Even with around 100,000 sequenced DNA strands from the last selection round, no meaningful pattern formation could be performed. Therefore, as with the Sanger sequencing data, no potential aptamer candidates could be identified by NGS.

3.4.11. Validation of aptamer candidates from CE-SELEX

Because the electropherograms of the selection indicated an increased binding ability of the later selection rounds towards APC, additional binding assays with the enriched library were performed. As the selection was performed in solution, a solution based binding assay was desired to reduce the risk that immobilization of the target might have an adverse effect towards the binding affinity of the library. For that reason, microscale thermophoresis (MST) measurements were performed.

MST depends on the measurement of the movement in a microscopic temperature gradient. [75] Interaction of two molecules leads to a reorientation of the hydration shell, thereby affecting the thermophoretic properties. [76]

To establish the MST assay, as a positive control, the HS02-52G aptamer against APC was used. [133] As described in section 5.5.2 (page 96) 16 glass capillaries were prepared by dipping the capillaries in 10 µl of a protein dilution series containing a constant concentration of Cy5 la-

beled DNA (10 nM) in the respective selection buffers. To reduce non-specific interactions such as the adsorption to the capillary walls, the following three components were added to the buffer: BSA (0.1 %), TWEEN-20 (0.1 %) and salmon sperm DNA (0.1 mg/ml).

A summary of the MST results is graphed in **figure 3.29**. For the positive control (HS02-52G, blue squares) a concentration-dependent binding curve could be obtained, from which a K_D of 9.4 ± 0.9 nM could be calculated. For the indole-functionalized start library (SL-1, gray circles) and the library after five rounds of CE-SELEX (5-1, red diamonds) no binding could be measured.

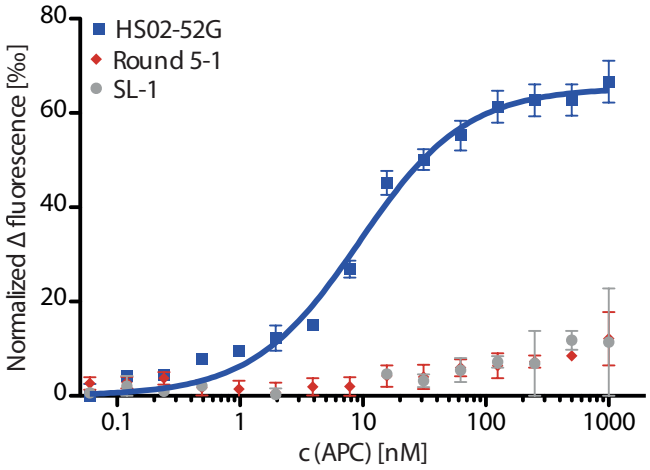

Figure 3.29. – MST binding data against APC
For the positive control (HS02-52G, blue squares) a concentration-dependent binding curve could be obtained, from which a K_D of 9.4 ± 0.9 nM could be inferred. For the indole-functionalized start library (SL-1, gray circles) and the library after five rounds of CE-SELEX (5-1, red diamonds) no binding could be measured.

In conclusion, even after several selection attempts and optimizations, all selections with the C8-dU libraries failed to show any enrichment towards the desired target structures. Therefore, a careful re-evaluation of all selection parameters was necessary. In search of an explanation for the lack of the appropriate selection pressure, the choice of the relatively long and very flexible alkyne linker of the C8-dU was questioned.

3.5. SELEX with EdU

Unsatisfying results for the selection attempts with C8-dU led to the conclusion to change the alkyne modified nucleobase from C8-dU to EdU. This shift in nucleobase choice was driven by the assumption that the relative long aliphatic C8 linker found in C8-dU gives the molecule too much flexibility, preventing a positive contribution of the modification to the binding affinity.

As discussed in section 1.4.3.2 (page 15) alternative alkyne modified nucleobases were described in the literature and were commercially available. Therefore, a selection attempt with the sterically more constricted EdU was conceived.

Because EdU is very similar in structure to the canonical thymine nucleobase, an easy adaptation of all required methods was estimated. The following chapter describes the adaptation of the click-SELEX process towards EdU, leading to a successful selection against the new model-target C3-GFP.

3.5.1. Library design of the FT2-N42-EdU library

The FT2-N42-EdU library was designed with the same principles in mind as the previously successfully employed FT-N42 library. However, this time, all thymidine residues were substituted by EdU instead of C8-dU. Also, to facilitate the detection and quantification of the DNA, a Cy5 fluorophore was introduced at the 5' end. A schematic representation of the FT2-N42-EdU library is depicted in **figure 3.30**

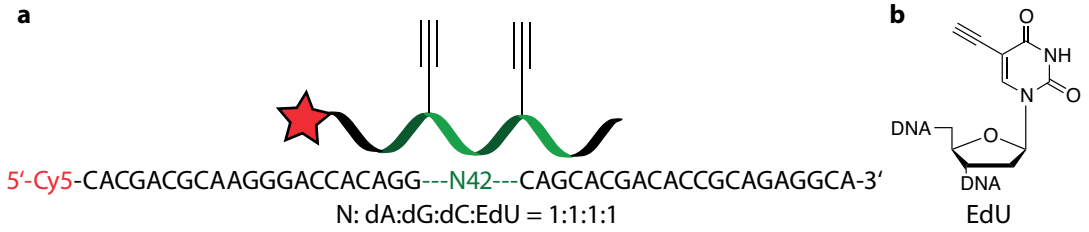

Figure 3.30. – Schematic representation of the N42-EdU library
The FT2-N42-EdU library is composed of a 42 nt random region (green) flanked by two 21 nt primer binding sites (black). To avoid potential problems in the PCR, associated with the alkyne modifications, the primer binding sites were designed without any thymidine nucleotides. In the random region, an equal distribution of all four nucleotides was desired. To facilitate the quantification and interaction analysis a Cy5 fluorophore was introduced at the 5' end.

3.5.2. Amplification of the FT2-N42-EdU library

Due to EdU's strong similarity to the canonical thymine nucleobase, no major incompatibility problems were anticipated. However, the compatibility of EdU containing libraries with the enzymatic reactions needed for the selection process had to be verified. As expected, PCR and λ-exonuclease digestion were shown to be fully compatible with the EdU containing library (**figure 3.31**). EdU containing DNA behaved equally to C8-dU containing DNA in all enzymatic reactions and purification properties so that the EdU containing library could be used for SELEX without further adaptation of the required methods.

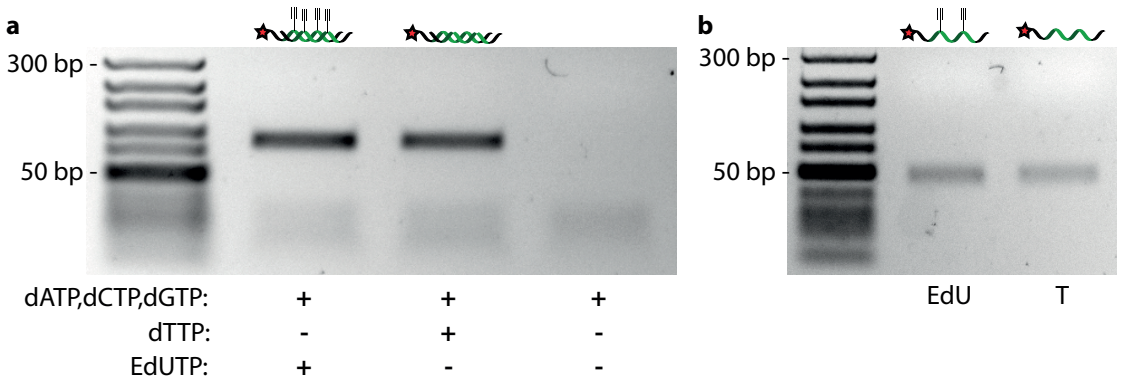

Figure 3.31. – Adaptation of PCR and λ-exonuclease digestion to EdU library
a) Enzymatic incorporation of EdU into the dsDNA library with EdUTP compared to the canonical dTTP. b) λ-exonuclease based single strand displacement of EdU incorporated dsDNA compared to the thymidine-containing equivalent.

3.5.3. Functionalization of EdU by click chemistry

In the initial consideration for which of the alkyne modified nucleobases to work with, C8-dU was preferred over EdU because it was reported for EdU to show inferior reaction yields [122,123]. With the optimized protocol described in section 3.4.3 (page 34) nearly quantitative reaction yields were observed for C8-dU containing DNA. Therefore, the same conditions were applied to test the click-reaction yield of EdU containing DNA, utilizing a short test-oligonucleotide bearing only one alkyne modification per strand (Test-EdU: 5'-GCACTGT-EdU-CATTCGCG-3').

Figure 3.32 shows the results of the HPLC-MS analysis of Test-EdU before the click-reaction, whereas **figure 3.33** depicts the results after the functionalization with indole azide **1**. In agreement with the published results, only approximately 80 % yield could be detected for the EdU containing oligo, whereas previously for the C8-dU containing oligonucleotide quantitative conversion could be observed. This was not only true for indole azide **1**, but also for all other investigated azides. Optimization attempts modifying the click conditions and the use of a "next generation" stabilizing ligand (BTTAA) [128] were not successful in increasing the reaction yield. Nevertheless, functionalization experiments with the EdU containing library (FT2-N42-EdU) were performed.

Interestingly, when the PCR amplified, EdU containing library was analyzed for click-reaction yield, no residual EdU nucleoside could be detected by HPLC-MS (**figure 3.34**). This counter-intuitive result had previously been described, however, at the time no fully conclusive explanation for this observation could be determined. [124,140] Later, new experimental results and a newly published article allowed for the full elucidation of this finding. A detailed analysis of this phenomenon is described in section 3.5.8 (page 71).

At this point, after having shown to be able to fully functionalize EdU containing DNA libraries with indole azide **1**, all methods for SELEX were established, and a new SELEX approach could be started.

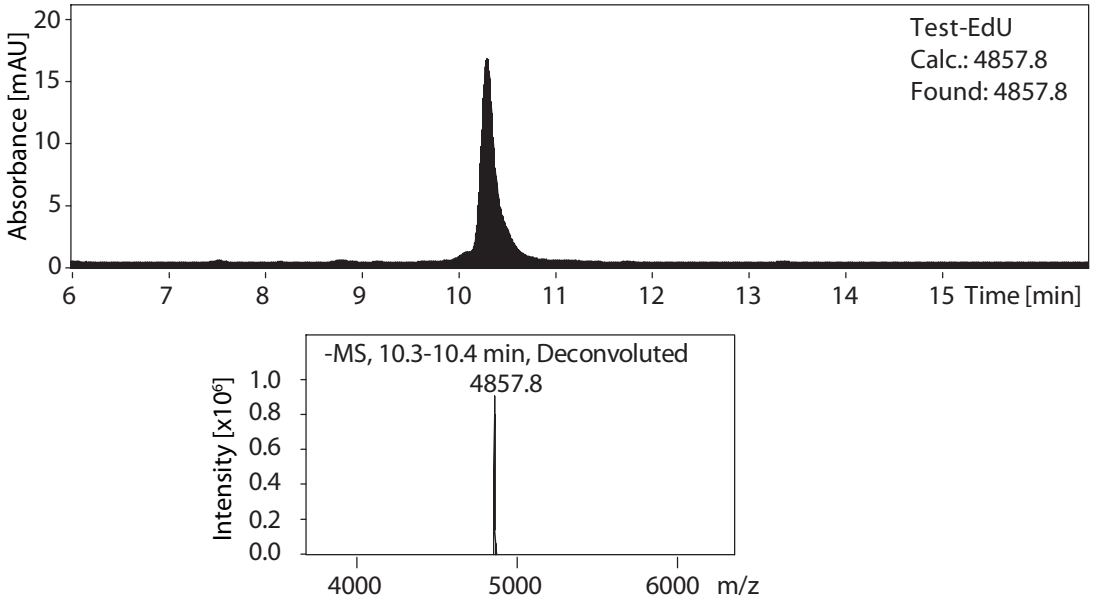

Figure 3.32. – HPLC-MS analysis of the EdU containing test-oligonucleotide
Prior functionalization the test-oligonucleotide eluted after 10.3 min. As calculated a mono-isotopic mass of 4857.8 could be detected.

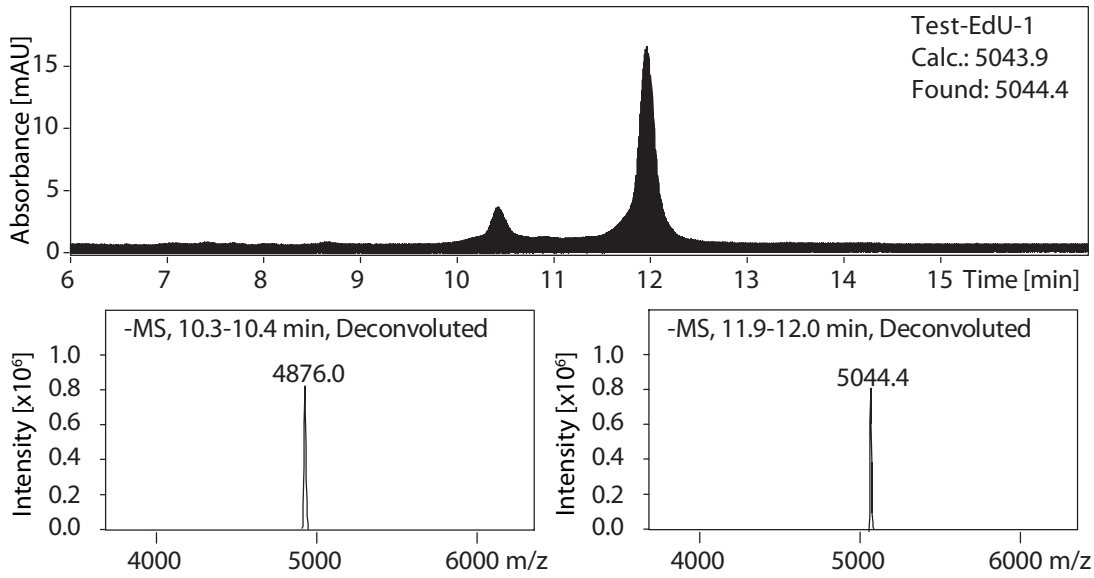

Figure 3.33. – Investigation of the click-reaction with a EdU containing test-oligonucleotide
After functionalization of the test-oligo with indole azide **1** the DNA eluted after 11.9 min. Approximately 80 % of the product with a mono-isotopic mass of 5043.9 could be detected. The remaining 20 % DNA eluted at the same retention time as the Test-EdU oligo, however for this DNA an increased mono-isotopic mass of 4876.0 was measured. For further information please refer to section 3.5.8 (page 71).

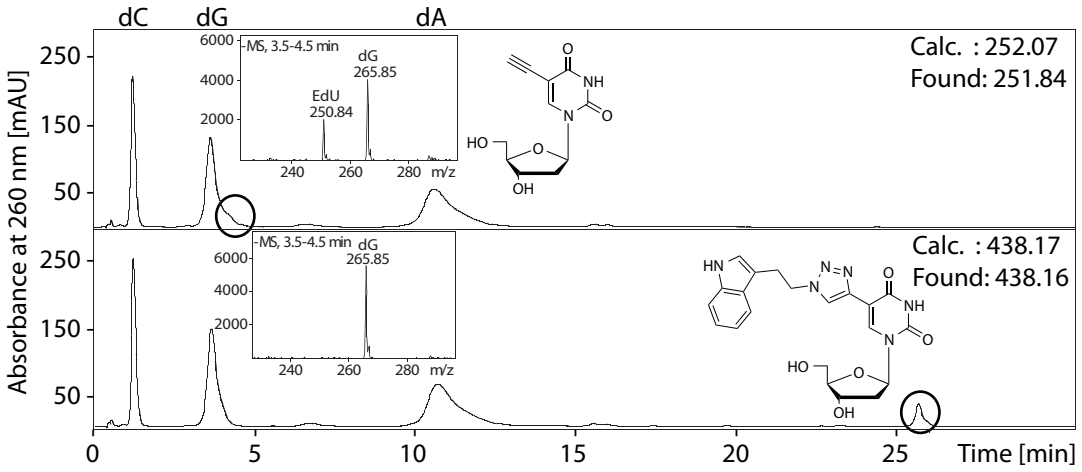

Figure 3.34. – Investigation of the click reaction with the EdU containing library
HPLC-MS analysis of the EdU start library before (upper trace) and after (lower trace) CuAAC functionalization with indole azide **1**. To assess the yield of functionalized EdU the samples were enzymatically digested to single nucleosides. After the reaction no unreacted EdU could be detected by mass spectrometry.

3.5.4. Change of target protein and immobilization strategy

In addition to moving from a C8-dU to an EdU-containing library, the choice of a different model target was considered. Initially, APC was chosen as a model target due to its good addressability by nucleic acid libraries, being considered as "easy" first target to validate the click-SELEX technique. However, as the previous unsuccessful selection attempts suggested, it might not be the optimal target for the selection with hydrophobic modifications. Therefore, to rule out that the problems in enriching a modified library were simply due to an unlucky choice of the model target, a more suitable new target was selected.

GFP was chosen to be a better model target due to several reasons. It is a fairly stable and commercially available protein with widespread use throughout the life-sciences. Its fluorescence properties make it easy to monitor when loading on to the magnetic beads, facilitating the validation of this important step. To our knowledge there were only RNA [141], but no DNA aptamers [142] available for GFP, making a more stable and cheaper to produce DNA aptamer highly desirable.

The green fluorescent protein (GFP) is a protein that exhibits bright green fluorescence when exposed to light in the blue to ultraviolet spectrum. The use of GFP and its derivatives has fundamentally redefined fluorescence microscopy. Due to its enormous importance for modern cell biology, Martin Chalfie, Osamu Shimomura, and Roger Tsien were awarded the 2008 Nobel Prize in Chemistry for their discovery and development of this protein.

GFP has a beta-barrel structure consisting of eleven β-strands, with an alpha helix containing the covalently bonded chromophore 4-(p-hydroxy benzylidene)imidazolidin-5-one (HBI) running through the center. In wt-GFP HBI, which is formed of the tri-peptide Ser/Tyr/Gly, exists

mainly in the unionized phenol form. However, a single point mutation (S65T) dramatically improves the spectral characteristics of GFP by deprotonating HBI to the phenolate form. This results in increased fluorescence, photostability, and a shift of the major excitation peak to 488 nm. The excitation maximum matches better with the spectral characteristics of commonly available filter sets, increasing the practicality of use by the general researcher. [143]

Since this discovery, many mutated forms have been described in the pursue to improve properties of GFP. These optimizations usually fall into two categories: In one line of research, as already described, the fluorescent properties of the protein are altered. The other optimization goal, however, is aimed at the improvement of the folding properties of GFP at 37 °C.

One such optimized variant, which itself was identified in an *in vitro* selection experiment, is "cycle3-GFP" (C3-GFP) used as a target protein in this study. [144] As a control protein, monomeric E-GFP (mE-GFP) was used. Belonging to the E-GFP family this version contains the S65T point mutation in addition to some mutations associated with improved folding properties at 37 °C, reducing its tendency to multimerization and aggregation.

Figure 3.35 shows the structure of wt-GFP with the positions of the point mutations found on the surface of C3-GFP and mE-GFP highlighted in red. A detailed analysis of all mutations found in C3-GFP and mE-GFP is shown in **figure A.7** (page 107).

Together with the change of the target protein, a change in the immobilization matrix on the magnetic beads was executed. Therefore, cobalt-functionalized magnetic beads were chosen. On these beads proteins can be immobilized via a His-tag sequence. Many recombinant proteins are expressed with a His-tag, making this method easily adaptable to many other proteins. The use of His-tagged proteins spares the biotinylation step, facilitating the immobilization process and preserving potential binding epitopes on the surface of the protein, which could potentially be blocked by the biotinylation-agent. With fewer modifications to the protein and its structure, the chance of selecting an aptamer that is capable of recognizing the native, unmodified protein is increased. As an additional advantage over the previously used streptavidin coated beads, no additional potential target protein is presented, reducing the risk of accidentally selecting matrix binding aptamers.

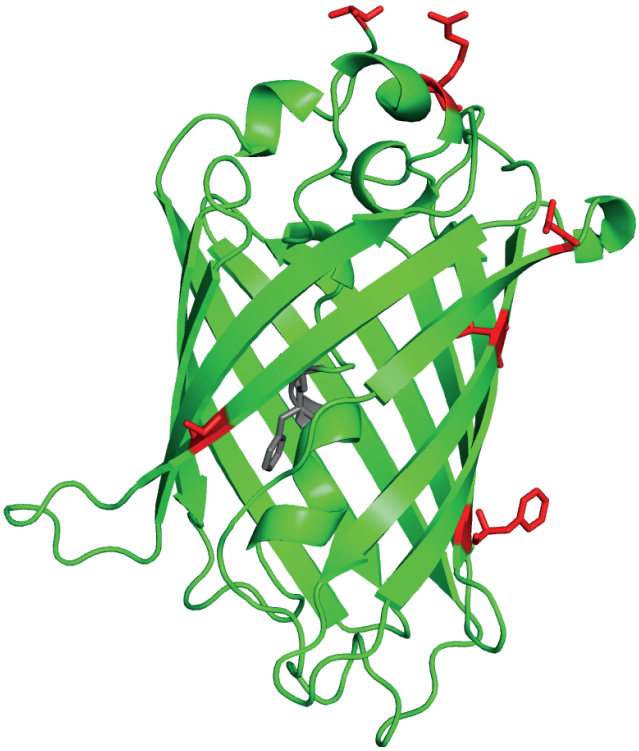

Figure 3.35. – Crystal structure of the green fluorescent protein (GFP)
Crystal structure of wt-GFP (PDB: 1GFL) highlighting mutated amino acids found in C3-GFP or mE-GFP. In red are mutations on the surface of the protein. In gray are mutations inside the barrel structure.

3.5.4.1. Immobilization of GFP on cobalt-functionalized magnetic beads

Commercially available C3-GFP (Life Technologies) was immobilized via its His-tag on cobalt-functionalized magnetic beads as described in section 5.4.3.1 (page 94). Briefly, 25 µl of resuspended cobalt Dynabeads (His-Tag Isolation and Pulldown kit) (40 mg/ml) were washed three times with 500 µl SELEX buffer and resuspended in 500 µl SELEX buffer. 500 µl of a 3.5 µM protein solution in SELEX buffer was added and incubated for 30 min at 25 °C at 800 rpm. The protein-loaded beads were washed three times with 500 µl SELEX buffer and resuspended in 500 µl SELEX buffer.

To check the success of the protein immobilization on the beads, GFP's inherent fluorescence properties were exploited. In addition to the protein immobilization, the elution of the protein with imidazole was analyzed. Therefore, 100 µl of a 300 mM imidazole solution was added to 5 µl of the C3-GFP loaded beads and incubated for 15 min at 37 °C. As a control, unloaded beads treated in the same way as the GFP coupled beads were prepared. Each tube of bead sample was filled up to 200 µl with D-PBS and 50000 events were measured in the FITC channel of a flow cytometer (FACSCanto II from BD Biosciences).

The results are depicted in **figure 3.36**. Evident by the increase in green fluorescence of the beads after addition of GFP, the immobilization was successful. After treatment with imidazole, the green fluorescence is strongly reduced indicating a successful elution of the protein.

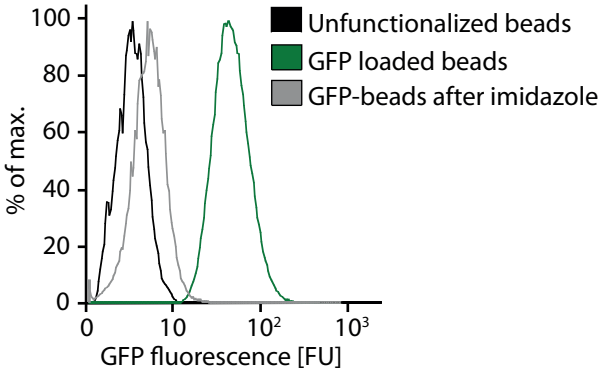

Figure 3.36. – Immobilization and elution of C3-GFP from magnetics beads
Flow cytometry analysis of bead functionalization with C3-GFP. C3-GFP loaded beads (green) show a strong increase in fluorescence intensity compared to non-functionalized beads (black). Upon treatment with 300 mM imidazole solution for 15 min C3-GFP is efficiently eluted from the beads (grey).

Having validated the immobilization and elution protocol the beads could be used for SELEX. The successfully coupled beads were stored at 4 °C until further usage.

3.5.5. SELEX with FT2-N42-EdU library for C3-GFP

The SELEX strategy for the new selection attempt was kept as similar as possible to the previous strategy with the streptavidin beads. Again, 3-(2-azidoethyl)indole **1** was used for the functionalization of the library. However, DNA recovery after the selection step was achieved by affinity elution of the target protein with imidazole. With affinity elution, larger amounts of DNA could be eluted from the beads, allowing a reduction in necessary PCR cycles and reaction volume during the selection.

The selection was carried out as described in section 5.4.3.2 (page 94). Briefly, the beads were washed three times with 200 µl SELEX buffer to remove unspecific binding DNA and the enriched library was eluted from the beads by addition of 100 µl of a 300 mM imidazole solution. The eluted DNA was PCR amplified in a total volume of 800 µl and purified as described in section 5.1.3 (page 87). The PCR product formation was monitored on 4 % agarose gels. 10 % of the purified PCR product from each selection round was stored as a backup. Single strand displacement of the double-stranded PCR product was carried out by λ-exonuclease digestion. The purified single strand DNA was again functionalized with indole azide **1** and applied in the next selection round.

To increase the selection pressure, the amount of magnetic beads was reduced and the washing time was increased during the SELEX. An overview of all the PCR products obtained after each SELEX round is displayed in **figure 3.37**.

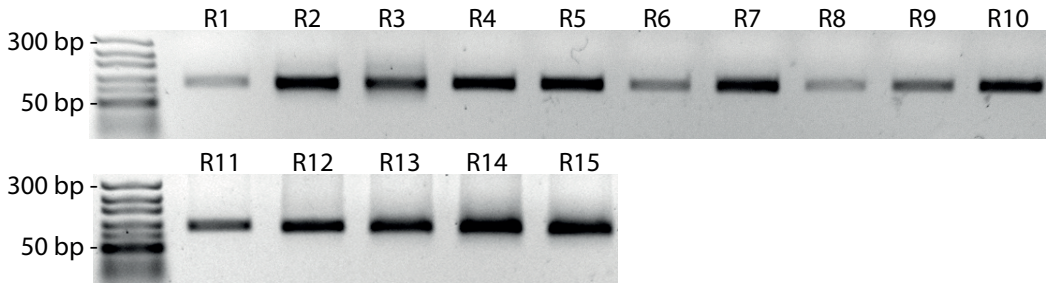

Figure 3.37. – PCR products of the SELEX with EdU against GFP
Ethidium bromide stained 4 %agarose gel showing the PCR products from all 15 SELEX rounds. For every selection round an intense band corresponding to the PCR product of the correct length could be obtained.

After eleven rounds of selection, the enriched library was tested for binding to C3-GFP by flow cytometry. As described in section 5.5.3 (page 96), 45 µl of a 500 nM solution of DNA from the 11th round and the start library (each with and without indole-functionalization) were incubated for 30 min at 37 °C with 5 µl of the C3-GFP loaded beads. The beads were washed two times with 200 µl SELEX buffer and resuspended in D-PBS. 50000 events were measured in the flow cytometer, and the mean Cy5 fluorescence intensity (APC-A channel) was analyzed.

Results are depicted in **figure 3.38**. After eleven rounds of selection, no difference in Cy5 fluorescence could be observed, indicating that no sufficient enrichment against C3-GFP could be obtained.

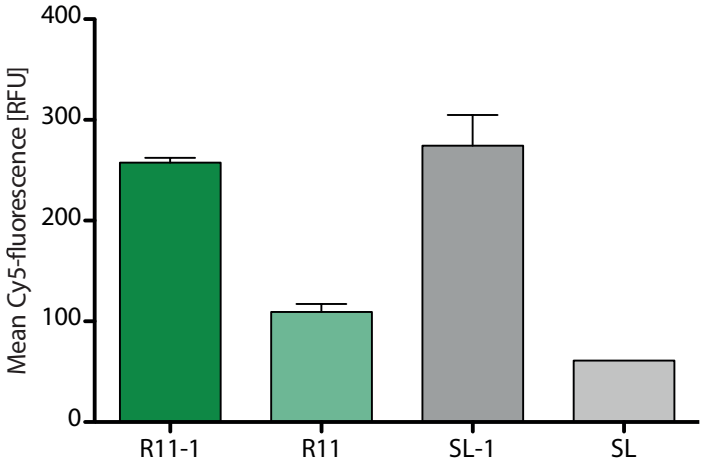

Figure 3.38. – Binding analysis of the 11th round SELEX library (R11-1) by flow cytometry
Compared to the indole-functionalized start library (SL-1), no difference in Cy5 fluorescence could be observed, indicating that not sufficient enrichment against C3-GFP could be obtained. As expected, all samples without indole-functionalization (R11 and SL) did not show binding to C3-GFP.

As no binding of the 11th round could be detected, a more stringent selection pressure had to be applied. Therefore, the addition of an indole-functionalized competitor oligo was conceived. This competitor oligo should be able to mimic the interaction possibilities of the indole-functionalized library closely and therefore, reduce the unspecific binding of the library. To generate such a competitor, a different DNA library (D3, section 6.2 (page 100), containing a different pair of primer binding sites, was PCR amplified with EdUTP instead of TTP, introducing alkyne modifications into this library. The double-stranded library was indole-functionalized as described in section 5.2.1 (page 90) and added in increasing amounts in the subsequent SELEX rounds.

Four additional rounds of selection were performed as described in section 5.4.3.2 (page 94). The enriched library of the 15th round was analyzed for binding to C3-GFP by flow cytometry. 5 µl of the GFP loaded beads were incubated with 45 µl of a DNA solution from the 15th round (R15-1) and the start library (SL-1) (each 500 nM). The beads were washed two times with 200 µl SELEX buffer and resuspended in D-PBS. 50000 events were measured in the flow cytometer, and the mean Cy5 fluorescence intensity (APC-A channel) was analyzed (**figure 3.39**).

After four additional SELEX rounds with the indole-functionalized competitor oligo, the enriched library showed a significant increase in binding to the C3-GFP functionalized beads, indicating a successful enrichment of sequences with the desired properties. As an enrichment of the library could be observed, a sample of the 15th round was prepared for sequencing.

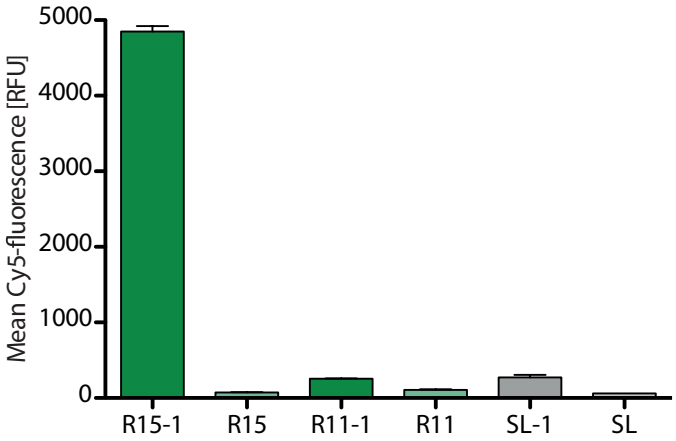

Figure 3.39. – Binding analysis of the 15th round SELEX library (R15-1) by flow cytometry
The enriched library of the 15th round (R15-1) showed a significant increase in binding to the C3-GFP functionalized beads, indicating a successful enrichment of sequences with the desired properties. All samples without indole-functionalization (R15, R11, and SL) did not show binding to C3-GFP.

3.5.6. Identification of aptamer candidates from C3-GFP SELEX

3.5.6.1. Cloning and Sanger sequencing of C3-GFP SELEX

Due to the promising results revealed in the binding analysis of the 15th round library it was decided to sequence the library. Therefore, a sample of the library was cloned into a vector by TOPO-TA cloning as described in section 5.1.6 (page 88). 42 of the obtained clones were amplified, and the plasmid DNA was extracted. 50 µl plasmid-DNA (30 ng/µl) was sent to GATC-Biotech for Sanger sequencing with the M13 reverse primer.

From Sanger sequencing, 36 sequences could be obtained (**figure A.8**, page 108). By comparing the monoclonal sequences according to their sequence homology, one promising family could be identified. This family was comprised of 7 highly homologous members with clone 12 and clone 33 being identical and the other members showing only a few point mutations. As clone 12 (C12) had the highest abundance in the dataset, the sequence family was called C12 family (**figure 3.40**).

```
C12 GXXGGAAGCGACGGGACGGXAAGGCXXGGGCCCCAAGGAGXG
C33 GXXGGAAGCGACGGGACGGXAAGGCXXGGGCCCCAAGGAGXG
C32 GXXGGAAGCGACGGGACGGXAAGGCXCGGGCCCCAAGGAGXG
C38 GXXGGAAGCGACGGAACGGXAAGGCACGGGCCCCAAGGAGXA
C26 GXXGAAAGCGACGGGACGGXAAGGCXXGGGCCCCAAGGAGXG
C27 GCXGGAAGCGACGGGACGGXAAGGCACGGGCCCCAAGGAGXG
C29 GXXGGAAGCGACGGXACGGXAAGGCXXGGACCCCGAXAAGXG
```

Figure 3.40. – The C12 aptamer family
Sequence comparison of the random region within the C12 aptamer family identified from the 15th SELEX round against C3-GFP. X stands for EdU, which was used during the SELEX instead of the canonical thymidine.

3.5.6.2. Next-generation sequencing of C3-GFP SELEX

To learn more about the SELEX and to further investigate the effect of the competitor on the selection pressure a NGS analysis of the SELEX was performed using Illumina sequencing by synthesis technology. Each SELEX round was PCR amplified with a different set of primers allowing for multiplexing 12 different SELEX rounds on one lane of the sequencer. After the indexing PCR, the 12 samples were mixed, and Illumina adapters were ligated. For adapter ligation with some adaptations the TruSeq DNA PCR-Free (LT) sample preparation kit was used.

Sequencing was performed on an Illumina HiSeq 1500 instrument in collaboration with the group of Prof. Joachim Schultze (LIMES, Bonn). 100 bp single end reads were performed. After sequencing the data analysis of the raw sequencing data was done by Dr. Carsten Gröber at AptaIT (Munich) using the COMPAS software.

Unique sequences

In **figure 3.41** the relative amount of unique sequences is plotted over the selection rounds. In the first six rounds, only very little enrichment of sequences could be observed. From round 6 to 10 a slight decrease in library diversity to 75 % unique sequences is observed. The relative amount of unique sequences plateaued until round 14, followed by a strong drop in complexity to 55 % unique sequences in round 15.

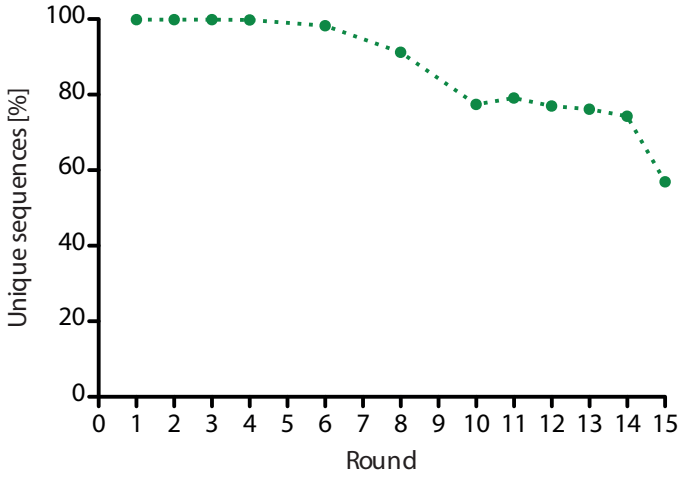

Figure 3.41. – Analysis of the unique sequences for the EdU-SELEX against GFP
In the first six rounds only little enrichment of sequences could be observed. From round 6 to 10 a slight decrease in library diversity is observed. Until round 14 no further decrease in unique sequences is measured, followed by a strong drop in complexity in round 15.

Character statistics

The character statistics for the EdU start library and the round 15 library are depicted in **figure 3.42**. The start library showed a uniform distribution of all four nucleobases over all positions of the random region. However, the relative amount of EdU nucleotides was only approximately 20 % instead of the expected 25 %, resulting in an average of 8.6 EdU nucleotides per strand. After 15 rounds of selection, as the diversity of the library decreases, clear preferences for specific nucleobases at individual positions within the random region emerged. This pattern formation can be seen as the first indication of a specific enrichment of the library. A selection of additional character statistics can be found in **figure A.9** (page 109).

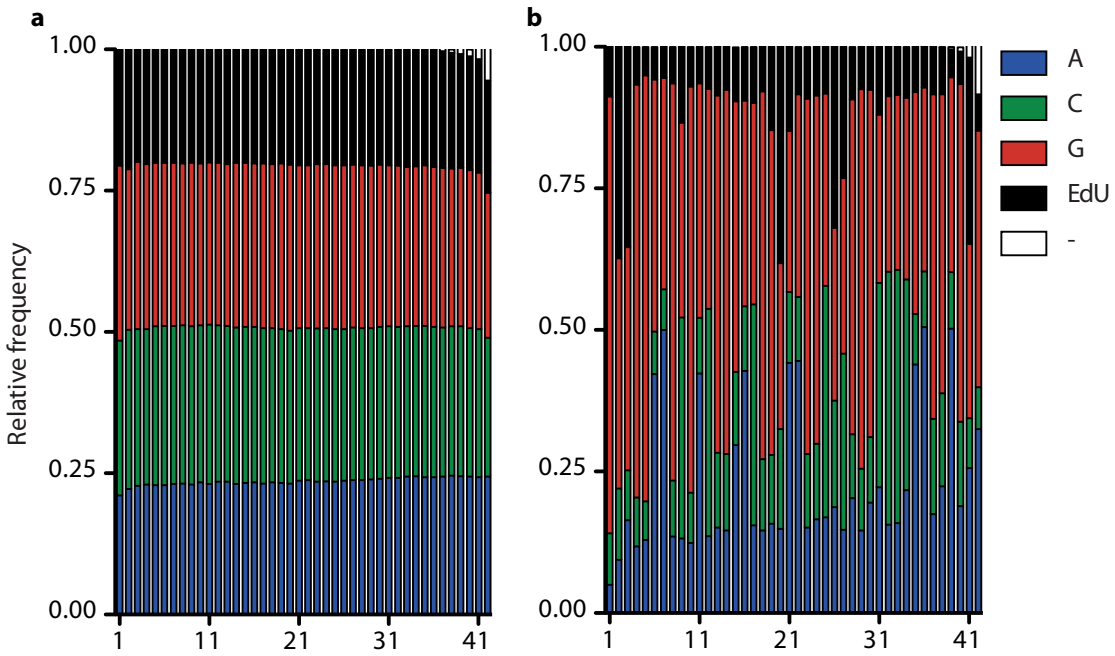

Figure 3.42. – Character statistics for the start library and the 15th round of the GFP-SELEX
a) Statistical base distribution in the EdU start library, showing a uniform distribution of all four nucleobases over the 42 positions of the random region. The relative amount of EdU nucleotides was only approximately 20 % instead of the expected 25 %, resulting in an average of 8.6 EdU nucleotides per strand. b) Statistical base distribution over all 42 positions of the random region after 15 rounds of selection. A clear position specific preference for certain nucleobases had emerged. This pattern formation can be an indication of a specific enrichment of the library.

Pattern and tracing

The main feature of the COMPAS software is to cluster similar sequences into families (”pattern”). Thereby, very similar families which vary in only a few point mutations can be analyzed collectively. Looking at the pattern, unsurprisingly the C12 family (pattern 1) was the most dominant sequence in the 15th round library. The COMPAS software allows for the tracking of the frequency of sequences and patterns over several rounds of selection (tracing).

The tracing for the 15 most abundant patterns is shown in **figure 3.43**, the consensus sequences of these 15 pattern can be found in **figure A.10** (page 110).

By analyzing the tracing data, a frequency and an amplification fold profile for each aptamer candidate was obtained. Interestingly, it could be observed that the C12 family of aptamers is strongly enriched with the addition of the indole-functionalized competitor oligonucleotide in the 12th SELEX round. Before addition of the competitor, the C12 family was represented in only low frequency in the library. In the 11th round the clones C15/C25 (pattern 3) and C7 (pattern 8) were the dominant sequences, but were decreased in the following rounds after addition of the competitor.

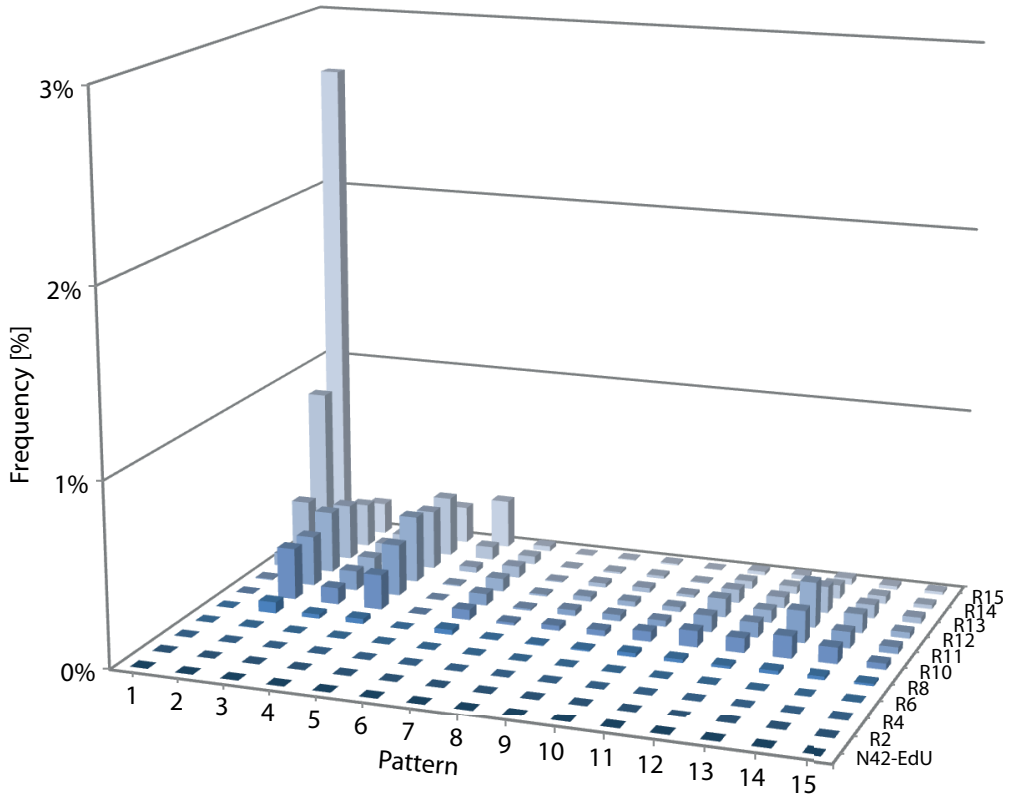

Figure 3.43. – Tracing of the consensus sequence of the 15 most abundant patterns
The C12 family (pattern 1) showed a very strong enrichment from round 12 on, making it the dominant sequence in 15th round library. All other patterns decreased in frequency after addition of the competitor, with pattern 5 (P5) being the only notable exception.

As shown in **figure 3.44**, only one other sequence family (pattern 5 or P5), beside the C12 family, is strongly enriched with the addition of the competitor.

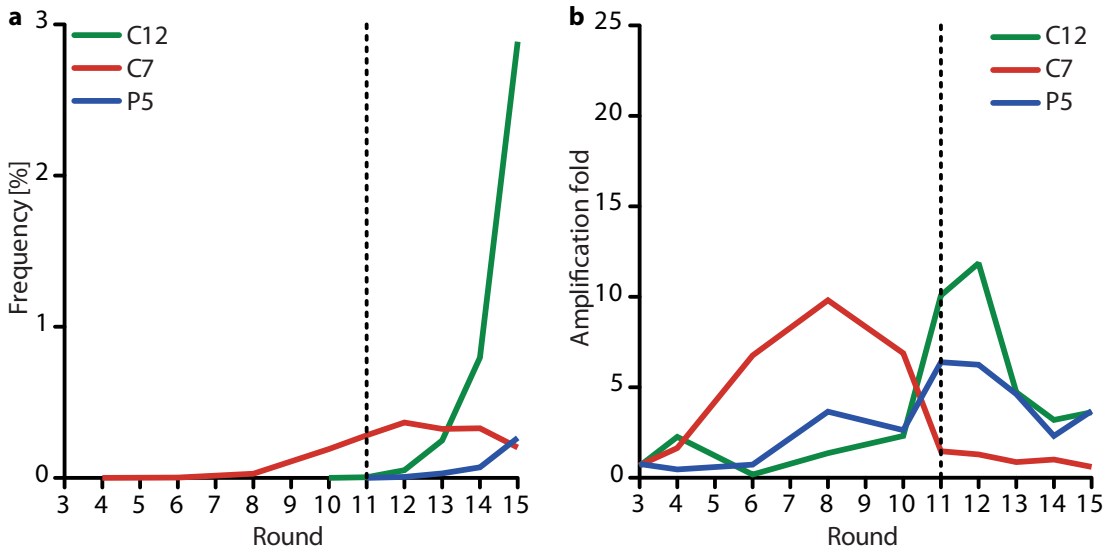

Figure 3.44. – Frequency and amplification fold profile for selected sequences
a) Frequency profile for sequences C12, C7, and P5. Only C12 and P5 showed an increased in frequency after addition of the competitor. C7 (and all other patterns) decreased in frequency after addition of the competitor. b) Amplification fold profile for sequences C12, C7, and P5. C7 (and most other patterns) had their amplification fold maximum before addition of the competitor. C12 and P5 showed the highest amplification after addition of the competitor.

3.5.7. Validation of aptamer candidates from C3-GFP SELEX

3.5.7.1. Binding of C12-1 to immobilized C3-GFP

As described in the previous two chapters, C12 was identified as a the most promising aptamer candidate by Sanger sequencing as well as NGS. To test C12's binding abilities to C3-GFP a flow cytometry binding assay was established. As described in section 5.5.3 (page 96), C12-1 and a scrambled control sequence C12sc-1 (500 nM) were incubated with C3-GFP coated magnetic beads and the retained amount of Cy5-tagged DNA was analyzed by flow cytometry.

As depicted in **figure 3.45**, after incubation with C12, a strong increase in mean fluorescent signal could be detected (green curve), indicating that C12 is capable of interacting with C3-GFP immobilized on the selection matrix. The C12sc-1 treated beads (gray curve) did not show any relevant increase in fluorescence intensity, validating that C12's binding behavior is sequence specific and not solely due to the rise of unspecific interactions, introduced by the addition of the indole-functionalizations.

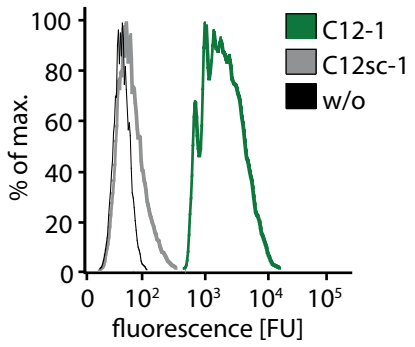

Figure 3.45. – Flow cytometry binding assay for C12 binding to C3-GFP
Shown is the amount of Cy5-labelled DNA (500 nM) bound to C3-GFP-bearing beads and beads in absence of DNA (black curve). C12-1 (green curve) reveals binding to C3-GFP, whereas the scrambled sequence C12sc-1 (gray curve) does not.

3.5.7.2. Binding of other clones to immobilized C3-GFP

Beside C12, several other clones were tested for binding against C3-GFP by flow cytometry. A summary of these experiments is depicted in **figure 3.46**. Four members of the C12 family were tested. C32-1 and C38-1 showed very similar binding profiles as C12-1, suggesting the few point mutations found in these two sequences (C32: EdU48C; C38: G36A, EdU47A, EdU48C, G63A) not to affect the interaction capabilities. However, C27-1 (the third member of the C12 family investigated) did not show any binding towards C3-GFP. This was surprising, as this sequence only contains the EdU23C mutation differentiating it from the rest of the C12 family. This data implies that the EdU at position 23 is critical to maintaining the binding ability of the C12 family aptamers. The nature of the conserved EdU positions was further characterized and described in section 3.5.7.6 (page 67).

None of the other tested sequences obtained by cloning and Sanger sequencing showed any binding to C3-GFP, rendering the C12 family the only aptamer family that could be identified by Sanger sequencing. P5, the consensus sequence of pattern 5, which was identified as a promising candidate by NGS analysis, did also exhibit binding to C3-GFP, making this family a second potential aptamer candidate. Additional binding data for the P5 aptamer is presented in section 3.5.7.8 (page 69) and section 3.5.7.9 (page 70).

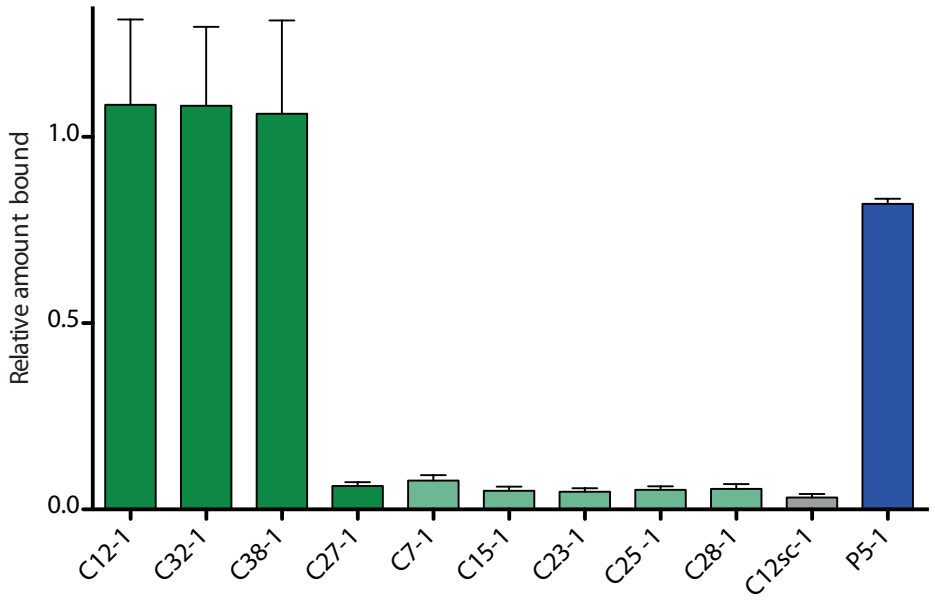

Figure 3.46. – Flow cytometry binding assay to C3-GFP for several monoclones
Given is the amount of Cy5-labelled DNA (200 nM) bound to C3-GFP-bearing beads analyzed by flow cytometry. Two other members of the C12 family (C32-1 and C38-1) showed a similar binding profile to C12-1. P5-1, identified by NGS analysis, also exhibited binding to C3-GFP, making it a second potential aptamer candidate. Values were normalized to C12-1 bound to C3-GFP. Shown is the mean value of at least three independent measurements.

3.5.7.3. Binding of C12-1 to C3-GFP in solution and determination of the dissociation constant

For the future application of the aptamer, it is important to verify its binding abilities in the absence of the selection matrix. To investigate the binding abilities of C12-1 in solution, a filter-retention assay was employed as described in section 5.5.1 (page 95). It is important to notice that the DNA sample has to be carefully purified from any residual primer before kinasation. In addition to the normal purification procedure by silica spin-columns, an additional gel extraction of the desired DNA is recommended. T4 polynucleotide kinase exhibits reduced activity on alkyne modified DNA, favoring the radioactive labeling of unmodified residual DNA, which in turn can lead to high levels of unspecific background in the filter retention assay. All kinasation attempts with already click-functionalized DNA failed completely to yield any labeling of the DNA.

The results of a filter retention experiment with properly purified samples of C12-1 and C12sc-1 towards C3-GFP are depicted in **figure 3.47**. In contrast to C12sc-1 (gray), the C12-1 specimen (green) exhibited a concentration-dependent binding behavior towards C3-GFP, from which a dissociation-constant of 18 ± 3 nM could be determined.

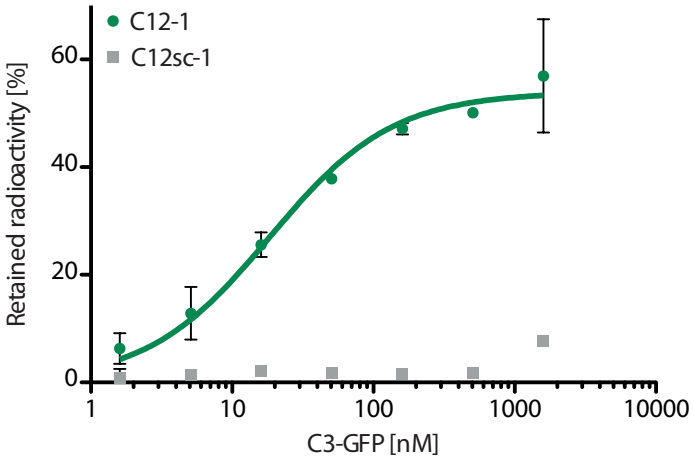

Figure 3.47. – Filter-retention analysis of C12-1 (green) and C12sc-1 (grey) towards C3-GFP
For C12-1 (green curve) a dissociation-constant (K_D) of 18 ± 3 nM could be determined. C12sc-1 (grey squares) did not show binding towards C3-GFP. Shown is the mean value of two independent measurements.

3.5.7.4. Specificity of C12-1 towards other proteins

Next, the specificity of C12-1 towards other proteins was investigated. Therefore, a variety of magnetic beads with different proteins was prepared. This included cobalt functionalized beads coated with ERK2 and mE-GFP, but also a control sample without protein immobilization to rule out unspecific binding to the bead matrix ("Empty"). Also, streptavidin coated magnetic beads were tested.

The results for the specificity assays are summarized in **figure 3.48**. Remarkably, even at a concentration of 500 nM, C12-1 did not bind to any of the non-target proteins tested, exhibiting a high degree of specificity to the original target it was derived from. This is not only true for ERK2 and streptavidin, which are completely unrelated to C3-GFP, but also for mE-GFP which is highly homologous to C3-GFP differing in only nine point mutations, seven of which are located on the surface of the protein (**figure A.7**, page 107).

In addition to the flow cytometric assay, the specificity of C12-1 was also analyzed in a filter-retention assay. In agreement with the flow cytometric data, C12-1 and C12sc-1 displayed no binding to mE-GFP in the filter-retention assay (**figure A.11**, page 110), validating C12-1's capability to discriminate even between very closely related proteins.

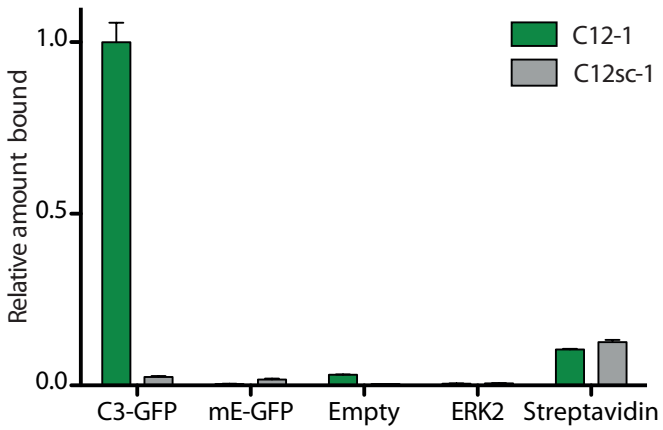

Figure 3.48. – Specificity determination for the aptamer C12-1
Specificity measurements for C12-1 (green bars). C12-1 does not interact with mE-GFP, ERK2, streptavidin nor with non-functionalized cobalt beads ("empty"). C12sc-1 (gray bars) does not specifically interact with any protein investigated. Shown is the amount of Cy5-labelled DNA (500 nM) bound to different protein-bearing beads analyzed by flow cytometry. Values were normalized to C12-1 bound to C3-GFP. Shown is the mean value of two independent measurements.

3.5.7.5. Influence of the chemical modification towards C12's binding behavior

After having investigated the specificity towards the target protein, the influence of the chemical modification towards C12's binding behavior was analyzed. Therefore a variety of additional azides (kindly provided by Dr. Gerhard M. Brändle) namely ethyl-naphthalene azide (**2**), ethyl-phenyl azide (**3**), benzyl azide (**4**), iso-butyl azide (**5**), and ethylamine azide (**6**) were introduced to C12 by CuAAC. The reaction yield for each azide was verified by HPLC-MS. As for indole azide, a full conversion of all alkyne-bearing nucleobases could be observed (**figure A.12**, page 111).

Binding data from the flow cytometry assay for C12 functionalized with a variety of different chemical modifications are summarized in **figure 3.49**. C12 showed very specific requirements towards the functionalization, as none of the tested modifications were able to preserve C12's binding ability towards C3-GFP. This is also true for very similar aromatic groups, such as **2**, **3** and **4**, as they could not substitute indole azide **1** and recover C12's binding properties.

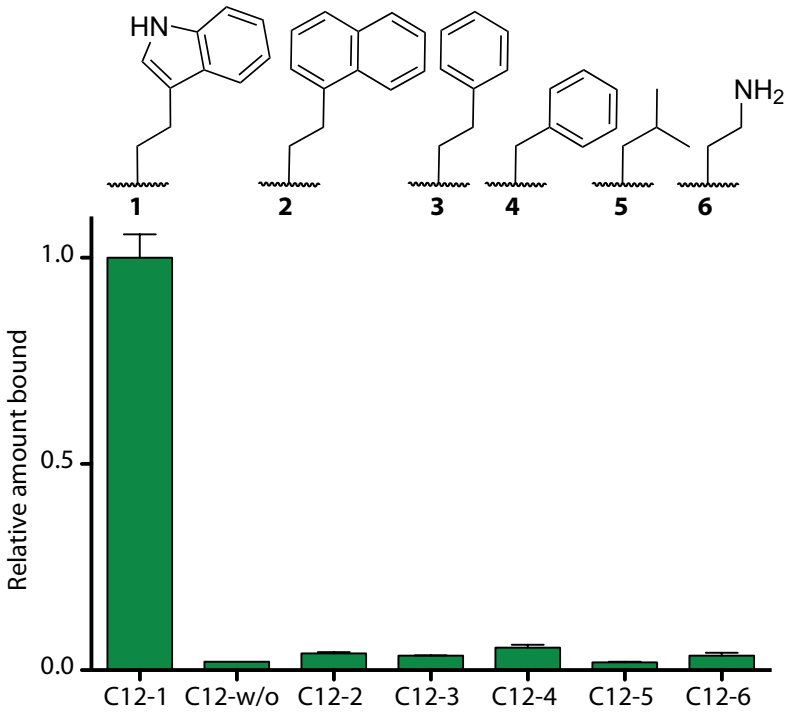

Figure 3.49. – Impact of different functionalizations on C12's binding towards C3-GFP
Shown is the amount of Cy5-labelled C12 (500 nM) bound to C3-GFP-bearing beads analyzed by flow cytometry. C12 was used either unmodified (w/o) or after CuAAC based functionalization with ethyl-indole- (**1**), ethyl-naphthalene- (**2**), ethyl-phenyl- (**3**), benzyl- (**4**), iso-butyl- (**5**), or ethylamine-residues (**6**). Values were normalized to indole functionalized C12-1 bound to C3-GFP. Shown is the mean value of two independent measurements.

3.5.7.6. Determination of essential EdU positions

Sequence analysis of the C12 aptamer family revealed, that EdU at positions 23, 24, 41 and 62 is highly conserved, whereas the positions 47 and 48 are mutated to canonical nucleotides within other C3-GFP binding members of the C12-family (**figure 3.40**, page 58) and **figure 3.46**, page 64). Thus, point-mutated variants of C12-1, with thymidine-residues at position 23 (T23), 24 (T24), 41 (T41), or 62 (T62) were synthesized and analyzed for binding to C3-GFP. Interestingly, positions 23, 24, and 41 were found to be crucial for C12's binding properties, since their substitution with thymidine led to a complete loss of binding to C3-GFP (**figure 3.50**). In contrast, replacing 62 with thymidine had no effect on C12's C3-GFP recognition properties, indicating that an indole-residue at this position is negligible.

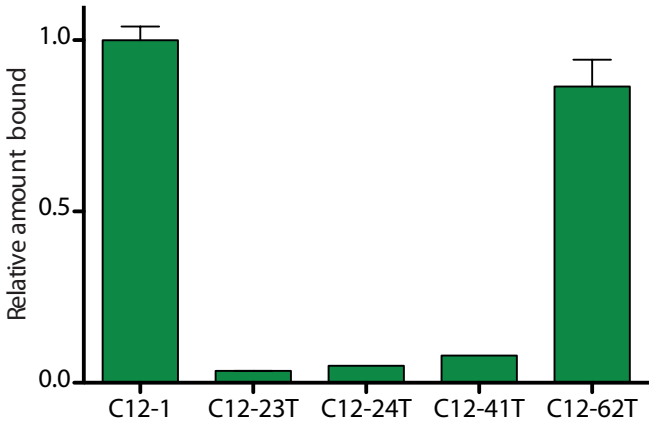

Figure 3.50. – Determination of C12's essential indole residues
Replacing the conserved positions 23, 24 and 41 with thymidine leads to a loss of affinity. Replacing position 62 with thymidine had no impact on C12-1 binding to C3-GFP. Given is the amount of Cy5-labelled and indole-functionalized C12-1 and single point mutated variants thereof (500 nM) bound to C3-GFP-bearing beads analyzed by flow cytometry. Values were normalized to fully indole-functionalized C12-1 bound to C3-GFP. Shown is the mean value of two independent measurements.

3.5.7.7. Proposed secondary structure of the C12 family aptamers

Based on binding data and *in silico* predictions, a secondary structure for the C12-family aptamers was proposed (**figure 3.51**). As neither the influence of the chemical modification nor the formation of possible G-quadruplex structures has been taken into consideration, this structure might not properly reflect the folding of the C12 aptamer in solution. Nevertheless, these *in silico* predictions could be beneficial to a better understanding of critical structural elements, thereby aiding the development of shorter aptamer sequences comprising a minimal binding motif.

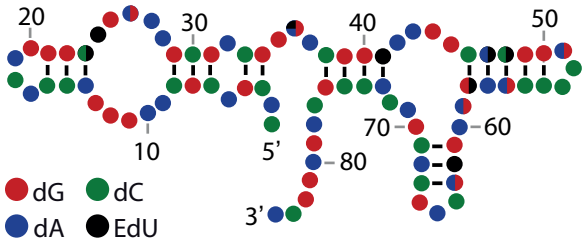

Figure 3.51. – Proposed secondary structure of the C12 family aptamers
Based on binding data and *in silico* predictions, a secondary structure for the C12-family aptamers is proposed. However, as neither the influence of the chemical modification nor the formation of possible G-quadruplex structures has been taken into consideration, this structure might not properly reflect the folding of the C12 aptamer in solution.

3.5.7.8. Characterization of the P5 aptamer

As described in section 3.5.7.2 (page 63), a second promising aptamer family was identified by NGS. This family showed an interesting amplification profile during the SELEX, as it was the only other family, besides the C12 family, that was enriched by the addition of the indole-functionalized competitor. However, even in the last round, the relative frequency was still comparably low. Therefore, unsurprisingly, it had not been observed in the Sanger sequencing data.

To further characterize this family, in analogy to the C12 aptamer, a flow cytometry-based affinity measurement against the target protein C3-GFP immobilized on magnetic beads was performed. The results are summarized in **figure 3.52**. The indole-functionalized P5-1 aptamer bound to C3-GFP in a concentration dependent manner.

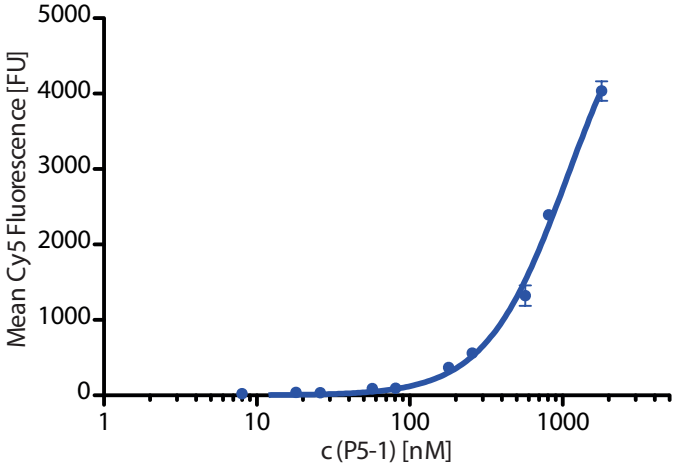

Figure 3.52. – Flow cytometry binding assay for P5-1 binding to C3-GFP
Shown is the mean fluorescence signal of Cy5-labelled P5-1 DNA bound to C3-GFP-bearing beads. Shown is the mean value of two independent measurements.

Next, the specificity of P5-1 was evaluated. A summary of the results is shown in **figure 3.53**. When the binding assay was conducted with the same conditions as the previous specificity assays with C12-1, a high binding of P5-1 to mE-GFP and streptavidin was observed. However, addition of an indole-functionalized competitor DNA in the form of the start library FT2-N42-EdU-1 led to a reduction of unspecific binding to streptavidin, whereas the binding to mE-GFP was unaffected. These results suggest that P5-1 is specific for GFP, but does not as strongly discriminate between the two tested members of the GFP family as C12-1.

In summary, these preliminary findings suggest P5-1 to be an interesting second aptamer candidate. Albeit its relative low binding affinity to the immobilized target protein, its "*pan*-specificity" towards other members of the GFP family might be beneficial for certain application. However, additional research is needed to elucidate P5's full binding characteristics.

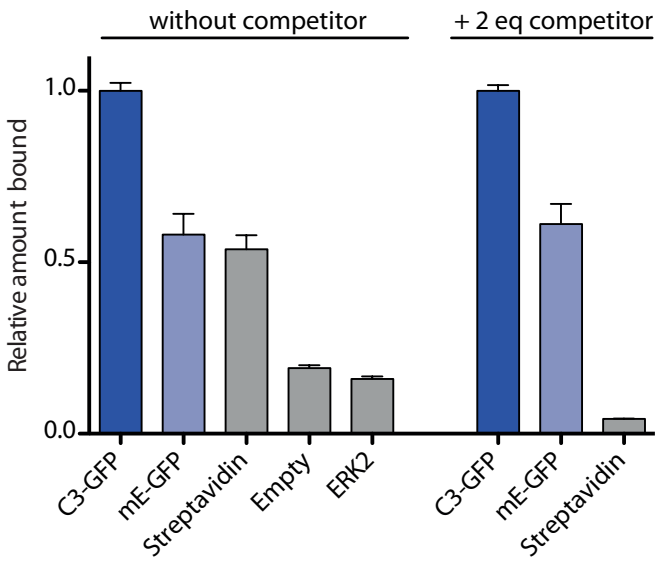

Figure 3.53. – Specificity determination for the aptamer P5-1
Specificity measurements for P5-1 with and without the addition of (1 µM) indole-functionalized competitor oligonucleotide (FT2-N42-EdU-1). Without the addition of the competitor strong unspecific binding to mE-GFP and streptavidin is observed. With the addition of the competitor the unspecific binding to streptavidin is reduced, whereas the binding to mE-GFP is unaffected. Shown is the amount of Cy5-labelled DNA (500 nM) bound to different protein-bearing beads analyzed by flow cytometry. Values were normalized to P5-1 bound to C3-GFP. Shown is the mean value of two independent measurements.

3.5.7.9. pH dependency of the C12 / C3-GFP interaction

During the whole click-SELEX study, for all selections and binding assays, without exception a phosphate buffer system based on Dulbecco's phosphate buffered saline (D-PBS)[135] was used. This D-PBS, being the basis for the SELEX buffer, was prepared from a commercially obtained ten times concentrated (10X) D-PBS solution (D1283, Sigma-Aldrich). To stabilize the magnesium and calcium ions, which would precipitate at this concentration at pH 7, the pH of the solution is adjusted to pH 4.4 by the manufacturer. Upon dilution of the buffer, this results in a pH of 5.3 for the final SELEX buffer. All experiments in this study where performed with this buffer, so that in contrast to most PBS formulations, the actual pH of the here used system was pH 5.3 instead of the commonly found pH 7.

As for some applications, especially in a biological context, a pH around seven might be preferred, the binding behavior of the C12-1 and P5-1 aptamers was analyzed at pH 7.0. A summary of these experiments is presented in **figure 3.54**. As previously described at pH 5.3 C12-1 (green) and P5-1 (blue) interacted with C3-GFP immobilized on magnetic beads, whereas the non-indole functionalized C12 control DNA (gray) showed no interaction. At pH 7 however, there was no detectable binding of any tested oligonucleotide in the flow cytometry binding assay.

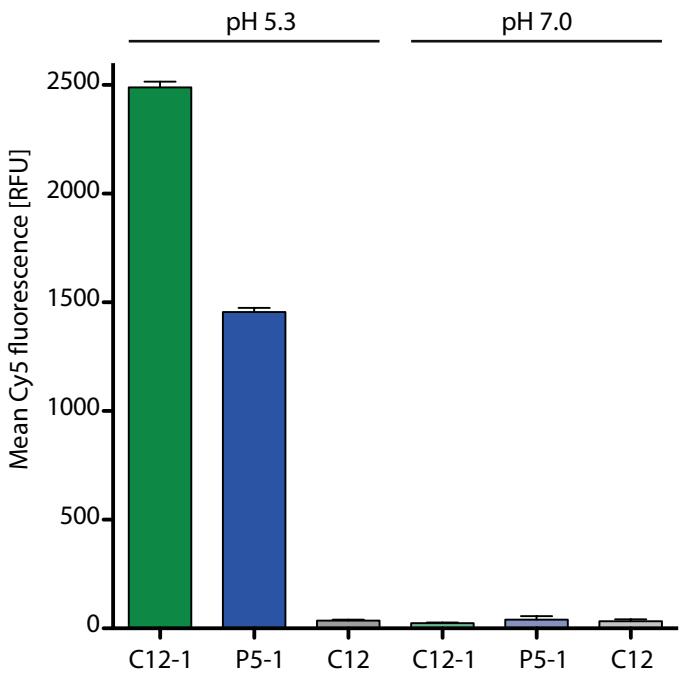

Figure 3.54. – pH dependency of the C12 and P5 aptamers
As previously described, at pH 5.3 C12-1 (green) and P5-1 (blue) strongly interacted with C3-GFP immobilized on magnetic beads, whereas the non-indole functionalized C12 control DNA (gray) showed no interaction. At pH 7.0 none of the tested oligonucleotides showed any binding affinity to C3-GFP. Shown is the amount of Cy5-labelled DNA (500 nM) bound to different protein-bearing beads analyzed by flow cytometry. Shown is the mean value of two independent measurements.

3.5.8. EdU by-product formation

As described in section 3.5.3 (page 50), surprisingly, despite thorough optimization of the click-conditions (including the solvent composition, the concentrations and ratios of the reagents and the reaction parameters), only up to 80 % product yield could be observed for the EdU containing test-oligonucleotide. However, while careful inspecting MS data of this reaction, it was observed that although having the same retention time as the starting material, the remaining DNA showed a slightly different mass, approximately 18 atomic units larger than observed for the starting material (**figure 3.33**, page 51).

Upon closer inspection it was revealed that the start material was already contaminated with this by-product, which in the previous analyzes was mistaken for a sodium adduct of the test-oligonucleotide. The observed by-product formation is consistent with a 2013 publication [145] describing the formation of a ketone by-product (KdU), formed under alkaline conditions, as found in the workup conditions after solid-phase oligonucleotide synthesis (**figure 3.55**).

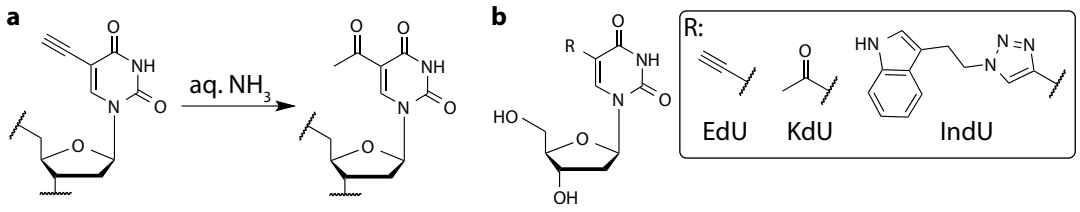

Figure 3.55. – EdU by-product formation under alkaline conditions
a) Under the alkaline condition (aq. 28 % NH_3) used in the workup after solid-phase oligonucleotide synthesis, approximately 20 % of the alkyne reacts to a ketone by-product.[145]. b) Structure of the ethynyl-deoxyuridine nucleoside (EdU), the undesired ketone by-product (KdU) and the desired indole-functionalized product (IndU).

The by-product is only observed in solid-phase synthesized DNA samples, explaining the full conversion of the samples produced by PCR. Although this finding does not effect the click-SELEX process itself (PCR amplified libraries do not contain this by-product), it poses a challenge for the production of large amounts of alkyne-containing aptamers by solid-phase synthesis. Several potential solutions to this problem are discussed in section 4.3.2 (page 80).

One possible alternative synthesis route to overcome the "ketone-problem" in the large scale production of clickmers is depicted in **figure 3.56**. As the by-product is caused by exposing the unprotected alkynes to alkaline conditions during the cleavage from the solid phase (SPS-A), one potential solution to this problem would be to first functionalize the alkyne, while still coupled to the solid support. Upon forming of the triazole product, no undesired by-products will be formed under the alkaline deprotection conditions (SPS-B).

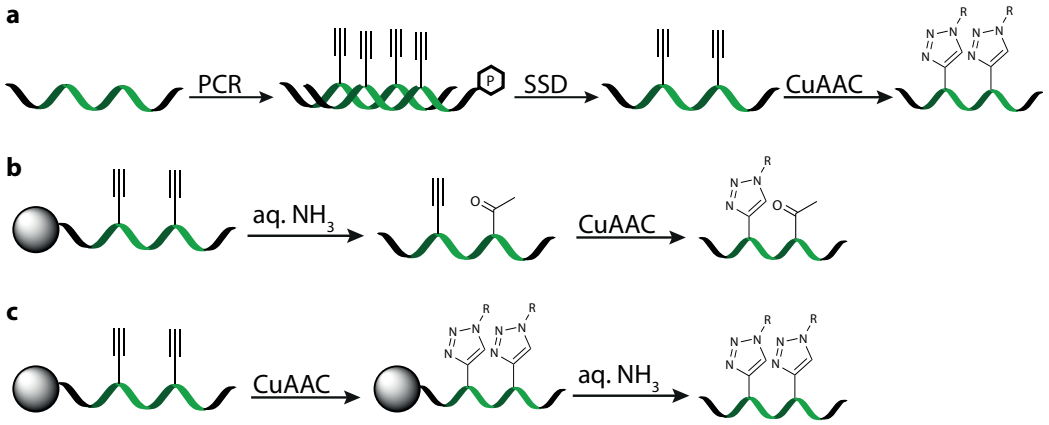

Figure 3.56. – Overview of synthesis routes for the generation of CuAAC functionalized DNA
a) "PCR:" Traditional synthesis route during the SELEX. The DNA is PCR-amplified and enzymatically digested to ssDNA. The DNA is functionalized by CuAAC in solution. b) "SPS-A:" The ssDNA is synthesized on solid phase and chemically cleaved from the solid support under alkaline conditions. The free DNA is functionalized by CuAAC in solution. c) "SPS-B:" The ssDNA is synthesized on solid phase and first functionalized by CuAAC while still connected to the solid support. After full functionalization the ssDNA is released under alkaline conditions.

To test this approach, a novel protocol for the large-scale functionalization of nucleic acids on solid support was developed (section 5.2.2, page 90), enabling the synthesis of C12-1 by solid-phase synthesis, introducing the indole-residue either after (SPS-A) or before (SPS-B) the alkaline workup step. The resulting DNA was analyzed on an agarose gel (**figure 3.57**) and by HPLC-MS (**figure 3.58**).

Both chemically synthesized aptamers revealed a clear and distinct band on the agarose gel, demonstrating the production of pure full-length DNA. However, the HPLC retention profile of the variant modified with indole-azide after DNA workup (C12-I-SPS-A) was broader compared to the one that has been functionalized on the solid phase before the workup (C12-I-SPS-B). Also, the identified mass of the variant synthesized according to strategy SPS-A differed significantly from the calculated mass, indicating an incomplete functionalization of the DNA under these conditions. In turn, C12-I synthesized according to method SPS-B (click reaction on the solid phase) showed a sharp retention profile and matched the calculated mass precisely, indicating full conversion to the desired product.

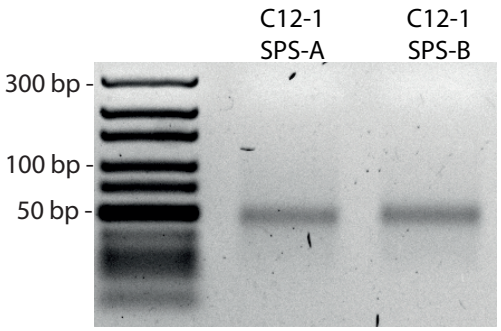

Figure 3.57. – Comparison of the two alternative solid phase based DNA production strategies
Ethidium bromide stained 4 % agarose gel showing the production of pure full-length DNA for both synthesis conditions.

To verify the full functionalization of the DNA synthesized according to the SPS-B procedures, the DNA was further analyzed at the nucleoside level. Therefore, enzymatically digested DNA was analyzed by HPLC-MS (section 5.3.2, page 92). As a control, C12 DNA, synthesized on solid phase and worked up under alkaline conditions, but not modified with the indole-azide (C12-SPS-A) was analyzed, revealing the two characteristic peaks for EdU and the expected KdU by-product. Next, the C12 DNA generated by the same strategy, but further reacted with the indole-azide via CuAAC in solution (C12-I-SPS-A), was analyzed.

As expected, the peak corresponding to EdU vanished, whereas the peak reflecting the indole-modified product (IndU) emerged. Notably, the KdU related peak was still present, underlining the heterogeneity of the resultant DNA molecules. In turn, the modification of the solid-phase

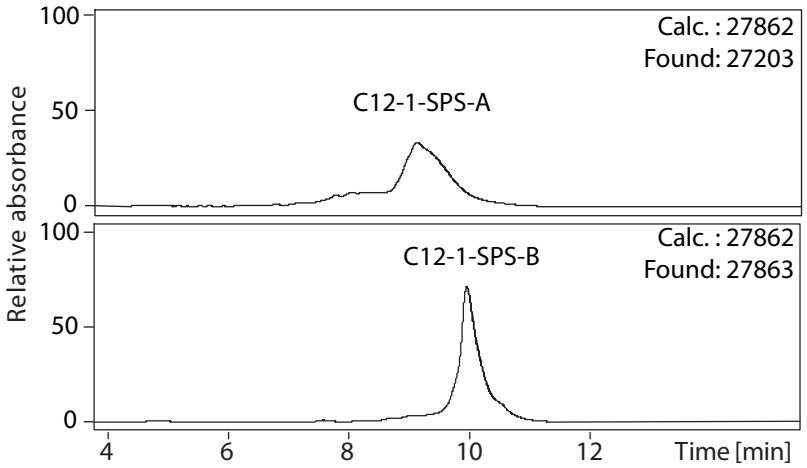

Figure 3.58. – Comparison of the two alternative solid phase based DNA production strategies
HPLC-MS based analysis of the functionalized DNA. C12-I-SPS-A (top) shows a broad retention profile and a deviation from the calculated mass, indicating an incomplete formation of the indole functionalized product. C12-I-SPS-B (bottom) shows a sharp retention profile with the expected mass of the fully functionalized product.

synthesized C12 by indole-azides on the solid matrix before the alkaline workup procedure (C12-I-SPS-B) resulted in a homogeneous DNA population, bearing only the desired IndU residues without any KdU contamination.

Having demonstrated the full functionalization of the solid phase synthesized clickmer, next the impact of the synthesis strategies on the binding behavior of C12-1 was analyzed. Therefore, a flow cytometry analysis was performed. As a reference, the PCR-synthesized C12-1, and the scrambled variant thereof (C12sc-1) were included.

As depicted in **figure 3.60**, the PCR synthesized C12-I (C12-I-PCR, grey squares) behaved as predicted and revealed a concentration-dependent binding to C3-GFP modified beads (K_D) of 216 ± 5 nM. In contrast, the PCR-synthesized scrambled variant (C12sc-I-PCR, gray diamonds) did not show interaction with C3-GFP. The C12-I clickmer, synthesized according to method SPS-B (C12-I-SPS-B, black dots) exhibited a concentration-dependent binding to C3-GFP, very similar to the PCR-synthesized reference (K_D) of 494 ± 44 nM). In contrast, the variant synthesized according to method SPS-A (C12-I-SPS-A, black triangles) showed a strongly reduced binding affinity towards C3-GFP functionalized magnetic beads. These findings underline the quality of the fully functionalized clickmers produced by the novel synthesis route.

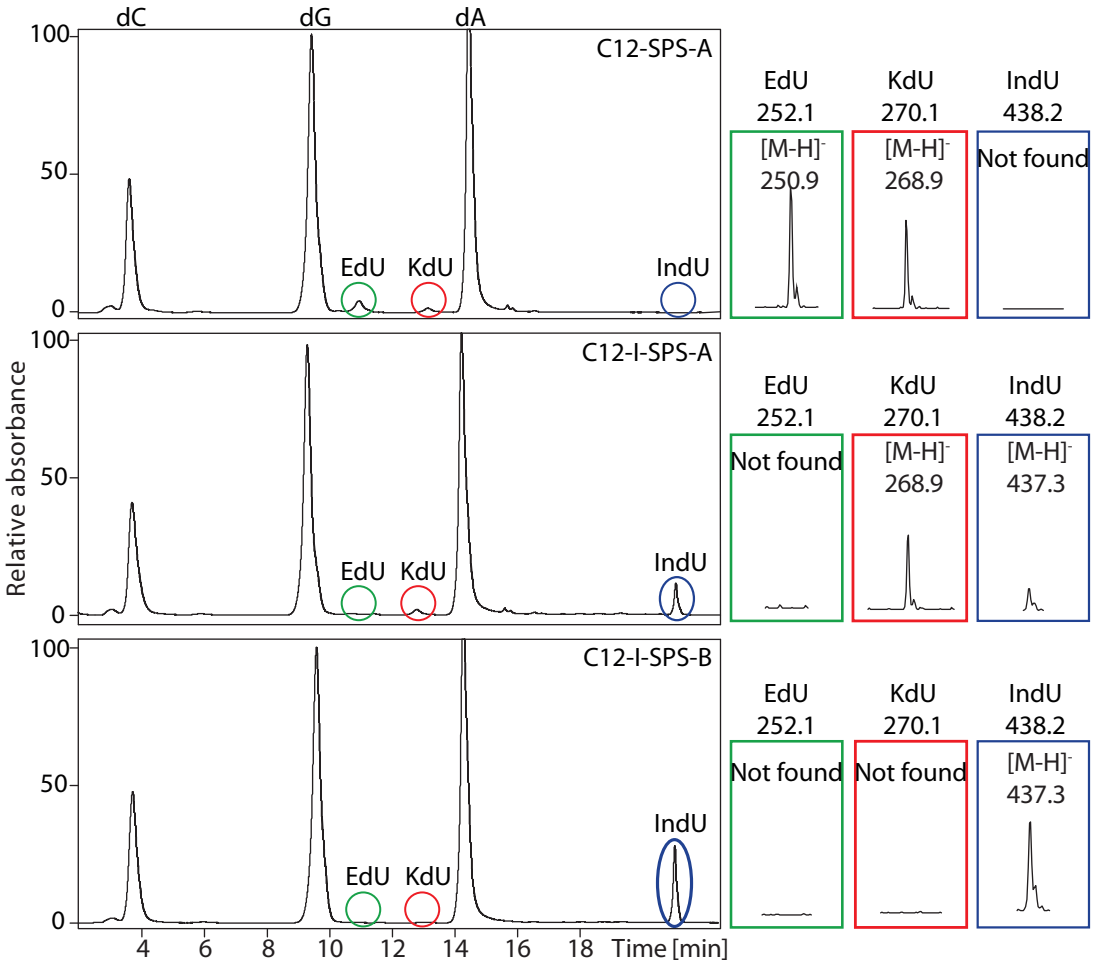

Figure 3.59. – Analysis of by-product formation on nucleoside level
HPLC (left) and ESI-MS (right) based analysis of the nucleoside mixture resulting from the enzymatic digestion of the DNA. For both click reaction strategies, full conversion of EdU to IndU is observed. However, the SPS-A produced DNA is contaminated with KdU (middle), whereas the SPS-B produced DNA is free of the undesired KdU by-product. Thus, SPS-B produced DNA contains a higher yield of the desired IndU product.

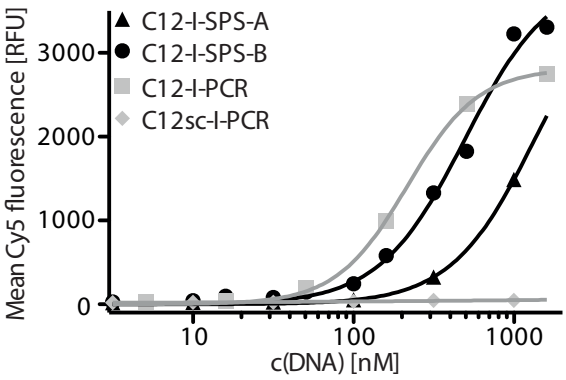

Figure 3.60. – Binding analysis of different C12-I clickmer derivatives towards C3-GFP
Shown is the mean fluorescence of Cy5-labelled DNA bound to C3-GFP-modified beads. The previously described reference C12-I-PCR (gray squares) reveals concentration-dependent binding to C3-GFP, whereas the scrambled sequence C12sc-I-PCR (gray diamonds) does not. The C12-I clickmer, synthesized according to method SPS-B (black dots) exhibits binding to C3-GFP that is very similar to the PCR-synthesized reference. In contrast, the variant synthesized according to method SPS-A (black triangles) shows a strongly reduced binding affinity. The data shown are the mean of two independent experiments.

4. Discussion and Outlook

4.1. Library design for click-SELEX

The alkyne libraries in this study were designed based on the ”traditional” library design utilized in most SELEX experiments.[29] Therefore, as with most selections with nucleobase-modified libraries, a random region with equal distribution of all four nucleotides was desired.

NGS data revealed a reduced incorporation rate for the alkyne modified nucleotides in both libraries. Instead of the expected relative average amount of 10.5 modified nucleotides per strand, only 8.5 (C8-dU, page 39) and 8.6 (EdU, page 58) were observed. This is probably partly due to a reduced coupling efficiency of the modified phosphoramidite. However, a position-depending variation in the base composition can also be observed for ”normal” libraries composed solely of canonical nucleotides (section 3.3.5.2, page 28). As the successful GFP selection demonstrated, this does not appear to be problematic for the selection efficiency of the FT2-N42-EdU library. If, however, a fully equal distribution of all four nucleotides would be desired, the difference in coupling efficiency could be counterbalanced by an increase in the phosphoramidite concentration during the solid phase synthesis of the library.

Nevertheless, due to inevitable minor performance variations during the solid phase synthesis, an entirely randomized library is difficult to obtain. In future implementations of click-SELEX other library designs could be explored. More research will be needed to evaluate if, for some targets, a decrease or increase of the total relative amount of functionalization sites might be beneficial. Also constrained and pre-structured libraries could be used to fine-tune the properties of the start library to the respective selection goals.[146,147] Such libraries could be beneficial to navigate the structural landscape adoptable by functional nucleic acids more efficiently and thereby prove useful in finding new functional motifs and/or more stable high-affinity aptamers.[148,149]

The concept of an *in situ* introduction of modifications can also be extended to other nucleobases. Besides the already discussed substitution of EdU by different alkyne bearing nucleobases such as EdA, the addition of more than one type of modified nucleobase per library should be explored. However, to technically implement an additional functionalization, other orthogonal click reactions would be desirable. Several additional types of click reactions have been developed in the recent years. One very promising reaction is the inverse-electron-demand Diels-Alder reaction of norbornenes with tetrazines.[150–152] But also other reactions like Staudinger ligation[153,154] or even reversible imine chemistry[155] could be used.

Potentially, these additional forms of click chemistry could be implemented with an extended genetic alphabet,[156,157] allowing for the introduction of functionalizations without disrupting the natural base pair system.[158] The use of extended alphabets would also enable a site-specific introduction of modifications.[159]

In addition to alternative coupling chemistry, adaptation to other forms of nucleic acid backbones could be attempted. The first step in this direction would probably be an adjustment of the click-SELEX method to RNA. However, the backbone chemistry is not limited to natural nucleic acids. The click-SELEX concept could also be applied to synthetic genetic polymers such as XNAs and PNAs, enabling the generation of substances with yet unrepresented properties.[93,95]

The click-SELEX methods allow for the generation of entirely new classes of functional tools. As an example, light-sensitive moieties like o-nitro-phenylethyl (NPE) groups[160] could be placed between the azide and the desired modification, enabling the direct selection of light-controllable aptamers.[161] Having the potential to be valuable new tools for biomedical research, light-controllable clickmers could also be used in highly efficient, yet very mild protein purification methods.[162]

4.2. Amplification of alkyne-DNA libraries

4.2.1. PCR with alkyne-nucleotides

The full compatibility of modified triphosphates with the enzymatic amplification during PCR is essential for the iterative nature of every SELEX protocol. One main advantage of the *in situ* introduction of the desired functionalization is that during PCR only a very lightly modified triphosphate has to be incorporated by the polymerase. Especially EdU and C8-dU were reported to be incorporated with high fidelity by various natural polymerases.[122]

In agreement with the published reports, EdU and C8-dU triphosphate were observed to be fully accepted by *PWO* polymerase when natural or alkyne DNA was used as a template. However, it was unclear if the polymerase would be able to accept the modified DNA as a template during the first PCR cycle. Regardless of the functionalization, no problems occurred during PCR amplification, and none of the tested functionalizations inhibited the PCR amplification.

Notably, the relative EdU amount in the library was reduced from ca. 20 % in the start library to ca. 12 % in the library of the last round (**figure 3.42**, page 60 and **figure A.9**, page 109). Since similar observations have also been made for selections with canonical nucleotides (unpublished results), at this point, it can not be fully resolved if this effect is due to an incorporation bias in the PCR or due to a selection pressure towards target binding.

If in the future very large modifications were desired in the library, the template might get sterically shielded, preventing a successful interaction with the polymerase. In this case, a light cleavable linker[161] (e.g. 1-(2-nitrophenyl)ethyl-ester (NPE) derivatives)[163] between the azide and

the desired modification could be used. This would allow the removal of the modification via UV irradiation after the selection step and guarantee compatibility with the click-SELEX protocol even for very bulky modifications.

4.2.2. Single strand displacement of alkyne-DNA

Both types of alkyne-DNA have shown full compatibility with λ-exonuclease digestion, making this fast and convenient method ideal for single strand displacement during click-SELEX (**figure 3.14**, page 33 and **figure 3.31**, page 50). Besides being fast and yielding high recovery yields, λ-exonuclease digestion has the advantage of being scalable and compatible with automatization, allowing this technique to be used on a robotic platform, dramatically increasing the throughput of possible click-SELEX experiments.[57,59]

If desired, the click-SELEX protocol could be adapted to other SSD techniques, for example, the capture and release (CaR) method.[58] This method uses magnetic particles to purify one strand from an asymmetric PCR mixture and would also be fully compatible with an automated workflow.

4.3. Functionalization by click chemistry

4.3.1. CuAAC for DNA functionalization

CuAAC has proven to be a very efficient method for bioconjugation.[164] For C8-dU as well as for EdUA a full functionalization of all alkynes, even in complex libraries, was observed. Although previously non-quantitative reaction yields were reported for EdU, the apparently lower yields could be attributed to a contaminating by-product found only in EdU containing synthetic oligonucleotides (section 4.3.2, page 80).[145]

After the initial evaluation of the poorly water soluble copper stabilizing ligand TBTA, the focus was rapidly changed to the better water soluble, "second generation" ligand THPTA. In agreement with previous reports[120], the commercially available THPTA ligand performed very well for the functionalization of DNA in solution and on a solid support.

As described in section 1.4.3.2 (page 15), fragile biomolecules can be damaged during CuAAC due to the formation of reactive oxygen species (ROS). With four equivalents of the protective ligand no damaged DNA was observed via gel electrophoresis or HPLC-MS. This makes CuAAC a suitable and very reliable method for the post-synthetic functionalization of DNA libraries during click-SELEX.

In a very recent publication it was shown that the addition of 10 % DMSO during CuAAC has a protective effect on DNA.[165] It is therefore recommended, as done throughout this study, to always include at least 10 % DMSO during the click reaction. For the click reaction of EdU libraries in solution with THPTA a rapid full conversion without measurable DNA damage was always

observed. However, especially for the large-scale reaction on the solid support, further improvements regarding the reaction conditions and solvent composition could be made. Therefore, the evaluation of the newest "third generation" ligands, which are reported to be beneficial in certain CuAAC scenarios, is encouraged. [128,154]

4.3.2. EdU by-product formation

Although the by-product formation observed with EdU does not directly affect the click-SELEX process itself, it might be detrimental to the synthetic large scale production of clickmers. Here, several strategies to overcome this limitation are proposed:

One strategy, implemented by Sachin Ingale and coworkers, [145] relies on a protection group on the alkyne during the workup under alkaline conditions. This method was successfully tested by introducing triisopropylsilyl ether (TIPS) protection groups at the alkyne of EdU. Unfortunately, the TIPS protected EdU building block is not yet commercially available and would have to be synthesized. Although being an elegant solution to the ketone problem, this solution is currently only available to the few groups with the necessary organic synthesis facilities and preparative know-how.

An alternative strategy, also involving the synthesis of a not yet commercially available triphosphate, would be to switch from EdU to the alkyne modified derivative of adenine (EdA). [166] This building block was described to be unsusceptible to the oxidative by-product formation. [145] Besides potentially eliminating the ketone problem, shifting to a functionalized deazapurine will probably have more profound additional effects on the click-SELEX process. [167,168] Although investigating the impact of this alteration on the library could lead to interesting new insights, it would probably involve the re-evaluation and potentially the adaptation of many of the click-SELEX associated protocols.

An alternative to these two solutions, both involving the synthesis of not commercially available building blocks, would be to functionalize the alkyne DNA directly on the solid support after synthesis, before the alkaline deprotection treatment. As described in section 3.5.8 (page 71) a successful implementation of this strategy could be achieved. With this new protocol large amounts (in the nmol range) of clickmers can be produced. [1] This is paramount for the further implementation of clickmers in e.g. biological systems, as for many interesting applications large amounts, beyond what is practically producible via PCR, are needed.

4.3.3. C8-dU for aptamer selection

In the course of this thesis, no successful enrichment of DNA libraries containing the C8-dU nucleotide could be observed. There is good reason to believe that the very flexible alkyl chain is detrimental for the exact positioning of the functionalization relative to the rest of the DNA. The increasing degree of rotational freedom in a longer linker will come with an entropic cost working

against the probability of the library to contain high-affinity binders. Indeed, empirically most successfully employed modifications have shorter and more rigid side chains.[96] However, the generalization that shorter and more rigid functionalizations are always better suited for selection cannot be made. In one instance it has been reported that shortening of the side-chain length had a dramatic detrimental effect on the activity of a DNAzyme.[169]

Nevertheless, for all selection attempts with C8-dU during this study no selection pressure could be obtained. Instead, in all instances a shortening of the library could be observed (**figure 3.22**, page 40 and **figure 3.28**, page 47). This effect is commonly observed during SELEX when no proper selection pressure towards target binding is generated.[53] In this case, the strongest selection pressure acting on the library population is an efficient amplification during PCR, giving shorter sequences a replicative advantage.[4,52]

Although the unsuitability of C8-dU for click-SELEX would have to be further evaluated, there are no apparent reasons to adopt the C8-dU building block for future click-SELEX experiments. C8-dU was initially chosen solely on the assumption that EdU would not give sufficient reaction yields during CuAAC.[123,124,140] As EdU was now shown to give quantitative product yields, this alkyne building block is recommended over C8-dU for its application in click-SELEX.

4.4. Selection methods for click-SELEX

Today, many SELEX protocols rely on magnetic bead-based selection strategies.[31,170] These have the advantages that they are cheap, easy to set up, have only moderate sample requirements, and can be automated.[32] Therefore, also for click-SELEX, magnetic bead-based selection strategies were evaluated, leading to a successful selection with cobalt functionalized beads and imidazole affinity elution. However, one of the drawbacks of magnetic particle-based separation is the relatively high degree of background binding. Strong background binding of the library to the beads leads to the elution of unspecific strands into the next SELEX round, thereby reducing the selection stringency and increasing the number of selection rounds needed.

The functionalized start libraries used in click-SELEX contain additional chemical groups and are therefore even more prone to unspecific binding due to the increased elemental affinity, not only towards the desired target molecule but also towards the bead matrix. NGS data revealed that the addition of indole-functionalized competitors seemed to be essential for successful enrichment (section 3.5.6.2, page 58). Traditional polyanionic competitors, such as salmon sperm DNA and hydrophobic competitors, such as TWEEN-20, successfully used in selections with unmodified nucleic acids, are apparently not sufficient to efficiently reduce the unspecific interactions of the indole-modified library.

This study showed that chemically functionalized competitors and efficient washing protocols are especially important for chemically modified libraries. To improve the selection process more stringent selection conditions should be explored. Microfluidics-based washing of the magnetic

particles could be beneficial in this context, as it allows for an efficient purification by which weakly attached background binders are removed in a high flow of buffer.[44] A similar "kinetic challenge" effect could be achieved by incubating the beads in large volumes of wash buffer, favoring the dissociation of unspecifically attached molecules.[78,171,172]

Other parameters that could be optimized include the bead surface itself. The use of more negatively charged bead surfaces or blocking methods could help to reduce the general affinity of nucleic acids. Furthermore, it should be tested if, with the right selection conditions, streptavidin beads can be used for click-SELEX or if the presence of an additional protein is detrimental to the selection pressure.

Although the implementation of capillary electrophoresis into the click-SELEX protocol was not yet successful, this method has some interesting properties that would make it an ideal fit for the generation of high-affinity clickmers. The very low degree of background binding would allow the reduction of the number of SELEX rounds, saving time and reagent costs.[173,174] It would also decrease the amount of amplification required, thereby reducing the influence of potential PCR amplification biases.[53]

One disadvantage of CE is that, as only a few nanoliters are injected per run, only a tiny amount of starting material can be used.[175] This results in a substantially reduced sampling of sequence space compared to other selection strategies.[51] Moreover, CE is not suitable for every target or selection buffer. A high degree of initial preparation and optimization of the buffer system is necessary to obtain acceptable migration behaviour. Nevertheless, regarding the benefits of CE it should be encouraged to investigate the general applicability of a CE based click-SELEX in a future SELEX attempt with a functionalized EdU based library.

The ability of click-SELEX to increase the general background affinity of libraries to their target structures demands methods for the reduction of unspecific background binding. Therefore, methods for the generation of high selection pressure, which in turn will lead to a reduction of selection rounds, are needed. The combination of click-SELEX with novel selection techniques will unlock the full potential of this technology, enabling the identification of clickmers with extremely high affinities (and specificities) for their target molecules.[31]

4.5. Sequencing and identification of aptamer candidates

Sanger sequencing has long been the only option for the identification of potential aptamers from enriched SELEX libraries. Although having several distinct advantages, such as being easy to implement, giving fast results and being comparably cheap, Sanger sequencing has the significant disadvantage of very low sample throughput. That results in the limitation that hardly more than several hundred sequences can be sequenced per selection. This implies that a very high degree of enrichment is necessary to be able to identify aptamer families.

Next-generation sequencing, on the other hand, allows for a several orders of magnitude increased sequencing throughput.[137] Thereby, not only the last selection round but also earlier rounds can be sequenced productively. This allows an early detection of promising sequences and offers valuable new insights into the previously inaccessible processes and mechanisms behind SELEX.[176,177] Hence, NGS offers new ways of data analysis. In particular, the three ways of data analysis described below have been used throughout this thesis.

Unique sequences: Plotting the percentage of unique sequences against the SELEX rounds enables a very fast and easy to comprehend assessment of the selection experiment. However, one has to keep in mind that also non-specific background binding sequences which happen to be efficient PCR amplifiers may also be enriched during SELEX, as they can never be entirely eliminated. Therefore, the percentage of unique sequences is a good first indicator of the successful enrichment during SELEX, but does not substitute more vigorous analytical methods.

Character statistics: Another very helpful way of representing NGS data is to plot the relative distribution of all nucleotides over all positions of the random region. This type of analysis is not only helpful to visualize conserved and variable positions in enriched libraries, but can also give valuable insights into the properties of start libraries. Analysing the character statistics can provide valuable insight into the structure-activity relationships of aptamer families, thereby improving the *post*-SELEX aptamer maturation process.[178]

Pattern and tracing: The main advantage of NGS data sets is the ability to trace individual sequences (and sequence families) over several rounds of selection. These amplification profiles are a very useful tool to improve our understanding of how binding sequences behave in the course of *in vitro* selection experiments. Currently, the relative round to round enrichment, especially in the last phase of the selection, is the best predictor for the binding affinity of a sequence.[62,176,179] A better understanding of the amplification profile of desired sequences can help to devise better selection methods and reduce the number of required amplification rounds.

Albeit the enormous rate of improvements currently made to NGS technology, there are still some areas for further improvement. Comparing the sequencing depth to the diversity of the library one has to keep in mind that even the highest throughput sequencing methods available today still can only sample fractions of the sequence diversity present in early SELEX rounds. Therefore, processes happening in the first rounds of selection are still elusive. Also, NGS data has to be analysed very carefully, as there are several sources of biases and contaminations.[180] These include biases in the sample preparation, the formation of clusters in the flow cell, as well as preferences associated with the sequencing method.[181]

As e.g. Illumina sequencing is widely based on PCR, these PCR related biases can be reflected in the dataset and have to be accounted for in the interpretation of complex NGS data sets.[182] Nevertheless, with ever falling prices and increasing accessibility of the instruments, NGS is clearly replacing Sanger sequencing for the analysis of *in vitro* selection experiments and will probably continue to do so in the future.

The technical possibilities associated with NGS will lead to significant improvements in the identification of aptamer candidates. With the ever improving ability to analyze the library composition, new forms of intelligent selection schemes can be realized. With the next generation of sequencing technologies on the horizon (e.g. microdroplet sequencing) and massively falling instrumental costs the online implementation of NGS directly during the selection scheme could be realized. Thereby, NGS data could be used to make informed decisions about the selection conditions for the next round, optimizing the selection pressure and leading to highly efficient selection procedures.

4.6. SELEX for APC

After several unsuccessful selection attempts with C8-dU modified libraries for APC, an EdU based library could be enriched for C3-GFP. As several parameters have been changed in between these selection attempts it is currently impossible to pinpoint the relevant factors contributing to the improvements in selection pressure. However, the transition from C8-dU to EdU and the implementation of a functionalized competitor oligonucleotide are very likely to have had a strong influence on the selection. Nevertheless, it can currently not be excluded that APC simply represents a poorly addressable target for functionalized libraries.

APC is very well targetable by natural DNA, yielding aptamers with very high affinity.[4,58,133] The SELEX libraries tend to enrich a strongly conserved consensus motif containing three essential thymidine residues. It could be speculated that the modification of the canonical thymidine residues disfavors the enrichment of clickmers against APC. In any case, it would be interesting to try to select against APC with an EdU based library (including a suitable competitor) to better understand the factors contributing to an efficient enrichment of functionalized nucleic acid libraries.

4.7. SELEX for GFP

An indole-functionalized EdU library could be enriched for C3-GFP immobilised on magnetic particles. The newly selected C12 aptamer family represents a proof of principle example for the feasibility of the click-SELEX concept. This makes C12-1 the first ever selected clickmer, an aptamer where a nucleobase functionalization is introduced via a rigid triazole linkage during SELEX. C12-1 is also the first DNA based aptamer to be selected against GFP. GFP represents a challenging target with several DNA-based attempts failing to identify aptamers.[142] That strengthens the initial assumption that the addition of chemical modifications can indeed be beneficial for the selection against previously inaccessible targets.

C12-1 was shown to bind very specifically to C3-GFP immobilized on magnetic particles, as well as in solution. It could be demonstrated that the indole functionalizations at positions 23, 24 and 41 are essential to maintaining binding affinity to the target, as the omission or substitution of the chemical modifications was not tolerated. C12-1 (and P5-1) binding was shown to be pH dependent, as almost complete loss of binding affinity was observed when altering the pH from the selection conditions (pH 5.3) to pH 7.0. If this is due to pH-induced changes in the aptamer or the protein cannot be addressed yet. However, previous selection attempts with unmodified ssDNA failed to yield aptamers against GFP at pH 5.2 as well as pH 6.5.[142]

Besides the C12 family, P5-1, a second interesting aptamer candidate, could be identified by NGS. Preliminary results indicate that the P5 aptamer family is less specific towards members of the GFP protein family. This *pan*-specificity might be useful for some future applications of this aptamer, making it a complementary addition to the clickmer toolbox.

Several interesting scientific questions about these two aptamers remain to be explored. One such research topic is the potential binding site and mode of interaction of the clickmers. Additional binding experiments against point mutated GFP variants and competition experiments of C12-1 with P5-1 could give additional insights. As there are no detectable similarities between the P5 and C12 sequences, P5 might potentially address a different epitope on C3-GFP. This could explain the difference in specificity, making P5 less selective towards other GFP variants.

Another interesting area for future research is the elucidation of the structures adopted by the functionalized aptamers. The three currently available crystal structures of nucleobase-modified aptamers have shown that the addition of functional groups can have a huge impact on the structure, massively expanding the already wide range of structural motifs natural RNA and DNA aptamers have been shown to adopt.[105–107] In analogy to the nucleobase-modified IL-6 SOMAmer, two independent domains could be predicted for the C12 aptamer family.[183] In addition to the initially predicted stem-loop motif one could postulate a G-quadruplex structure as found in many other aptamers.[184,185] Crystallography[186], NMR[187,188], or novel electron paramagnetic resonance (EPR) methods, such as "pulsed electron-electron double resonance" (PELDOR)[189], could help to elucidate the actual structure of C12 in complex with C3-GFP and give valuable insight in the fundamental folding and interaction profiles of clickmers. This in turn could inform the future library design and choice of nucleobase modifications.

4.8. Potential fields of application

The successful establishment of a proof of principle example in this study and thereby the validation of the viability of the click-SELEX approach opens up the field for numerous future applications. Especially innovations regarding the library design, new forms of coupling chemistry, as well as future improvements to the selection methods, will be crucial to push the field forward.

Being a versatile and adaptable method for the introduction of chemical modification, click-SELEX can democratize the ability to use the best possible selection conditions to everyone. Greater access to improved libraries will lead to more targets being addressed by aptamers. Especially in the context of cell-SELEX these new libraries might be very beneficial. Modulating the chemical modifications presented could result in the identification of novel biomarkers on cell surfaces.[190] This could improve our understanding of cell biology as well as the associated diseases and help in the development of novel diagnostic platforms.

Clickmers can in principle be applied in most areas where highly specific affinity reagents are needed. Due to their comparatively small size and robust, reproducible synthesis, clickmers could have an enormous impact in e.g. the field of microscopy where good tools are needed to accompany the development of modern super-resolution techniques.[142,191]

However, click-SELEX could have applications far beyond the selection of aptamers. As nucleic acids based catalysts will also benefit from increased chemical diversity,[94,192,193] "clickzymes" could have a tremendous impact in the field of biocatalysis.[194]

5. Methods

5.1. Working with nucleic acids

5.1.1. Agarose gel electrophoresis

4 % agarose gels were used to monitor length and quality of nucleic acids such as PCR products. For a 4 % agarose gel, 8 g agarose were dissolved in 200 ml TBE-buffer and boiled for several minutes in the microwave. After cooling down, ethidium bromide solution was added at a 1:10000 dilution and 45 ml gels were poured. The gels can be stored in TBE-buffer at 4 °C for several weeks. 5 µl of the sample were mixed with 1 µl DNA loading buffer. For electrophoresis, gels were run in TBE-buffer at 150 V for 10 min. Bands were visualized on a UV transilluminator and compared to the DNA ultra low range ladder (Life Technologies).

5.1.2. Polymerase Chain Reaction (PCR)

PCR was done in a Mastercycler Personal (Eppendorf). 0.5 µM of both primers and 250 µM of dNTPs were used. *Pwo* DNA polymerase (Genaxxon) and the supplied buffer containing 2 mM Mg^{2+} were use with the following cycling program (2 min 95 °C; 30 s 95 °C, 30 s 62 °C, 1 min 72 °C; hold 4 °C). The samples were always prepared on ice and the thermocycler was pre-heated to 95 °C to reduce the risk of by-product formation.

5.1.3. Purification

5.1.3.1. Silica spin columns

If not indicated otherwise, nucleic acids (e.g. PCR products) were purified with the commercially available NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) according to the manufacturer's recommendations.

5.1.3.2. Gel filtration

After radioactive kinasations, the DNA was purified with commercially available G25 spin columns (GE Healthcare) according to the manufacturer's recommendation.

5.1.3.3. Size exclusion

After functionalization by click-chemistry, DNA was purified with commercially available Amicon Ultra centrifugal filters (3 kDa cutoff for test-oligos and 10 kDa cutoff for the FT2 library) according to the manufacturer's recommendation.

5.1.4. Concentration measurement

DNA quantification in solution was done by UV absorbance. For concentration determination of DNA (or any other solute) the Beer-Lambert equation can be employed:

$$A = lg(\frac{I_0}{I_1}) = \epsilon_\lambda \times c \times d$$

A : Absorbance at a given wavelength (also called optical density (OD))
I_1 : Intensity of the transmitted light
I_0 : Intensity of the incident light
ϵ_λ : Molar absorption coefficient at a given wavelength [l×mol^{-1}×cm^{-1}]
c : Concentration [mol×l^{-1}]
d : Path length [cm]
DNA exhibits its absorbance maximum at around 260 nm. Therefore this wavelength is commonly used as diagnostic wavelength for DNA quantification. The molar absorption coefficient was calculated from the length and base composition with the online software tool OligoAnalyzer 3.1.

5.1.5. λ-Exonuclease digestion

In order to achieve a selective single strand displacement by degradation of only one strand of the dsDNA, λ-exonuclease digestion was applied. λ-Exonuclease [195] is a 5' to 3' exodeoxyribonuclease. It selectively digests the 5'-phosphorylated strand of dsDNA. The enzyme exhibits low activity on single-stranded DNA and non-phosphorylated DNA. 175 µl DNA solution was mixed with 20 µl λ-exonuclease buffer (10 x) and 5 µl λ-exonuclease (5000 U/ml) was added. This mixture was incubated for 60 min at 37 °C. Afterwards the resultant ssDNA was purified as described in section 5.1.3 (page 87).

5.1.6. TOPO-TA cloning

After SELEX, the enriched libraries were cloned into vectors and transformed into competent cells with the TOPO-TA Cloning kit (Life Technologies) according to the manufacturer's recommendation. Before cloning the samples have to be PCR amplified with *Taq* polymerase (instead of *Pwo* polymerase) as the addition of an 3' overhang deoxyadenosine is crucial for the cloning strategy. It is also critical that unmodified primers, without 5' Cy5- or phosphate groups are used. After plasmid preparation of successfully transformed colonies, the DNA was sent to Sanger sequencing.

5.1.7. Sanger sequencing

50 µl plasmid-DNA (30 ng/µl) was sent to GATC-Biotech for Sanger sequencing with the M13 reverse primer.

5.1.8. Next-generation sequencing (NGS)

NGS was performed using the Illumina sequencing by synthesis technology. A detailed description of the sample preparation has been published as a book chapter in the "Methods in Molecular Biology" series. [2]

5.1.8.1. Indexing PCR

To multiplex several samples in one sequencing run, index sequence primers have to be added to the end of each SELEX sample. Therefore, a set of 24 primers with 12 different index sequences at the 5' end of each primer was ordered. The index sequences are listed in table 5.1. Each SELEX round was PCR amplified with a different set of primers allowing for multiplexing 12 different SELEX round on one lane of the sequencer. After indexing PCR, the 12 samples were mixed, and Illumina adapters were ligated.

Table 5.1. – Index sequences used for NGS.

Index	Sequence	Index	Sequence
Index 1	ATCACG	Index 7	CAGATC
Index 2	CGATGT	Index 8	ACTTGA
Index 3	TTAGGC	Index 9	GATCAG
Index 4	TGACCA	Index 10	TAGCTT
Index 5	ACAGTG	Index 11	GGCTAC
Index 6	GCCAAT	Index 12	CTTGTA

5.1.8.2. Adapter ligation

For adapter ligation with some adaptations the TruSeq DNA PCR-Free (LT) sample preparation kit (Ref.15037063, Illumina) was used. For further details, please refer to the sample preparation chapter in "Methods in Molecular Biology". [2]

5.1.8.3. Sequencing

Sequencing was performed on an Illumina HiSeq 1500 instrument in collaboration with the group of Prof. Joachim Schultze (LIMES, Bonn). 100 bp single end reads were performed.

5.1.8.4. Data analysis

Data analysis of the raw sequencing data was done by Dr. Carsten Gröber at AptaIT (Munich) using the COMPAS software. Visual representations of the data were generated with Prism 5.0f (GraphPad Software) and Excel 2011 (Microsoft).

5.2. Click chemistry

5.2.1. Reaction conditions in solution

A 1 mM Cu^{I} catalyst solution was freshly prepared by addition of 25 µl freshly prepared 100 mM sodium ascorbate solution to 75 µl 1:4 $CuSO_4$/THPTA solution (1 mM and 4 mM final concentration respectively). The click reactions were performed in 100 µl final volume. Therefore, 10 µl of a 10 mM azide solution in DMSO was added to 80 µl of DNA solution in 100 mM phosphate buffer (pH 7). The click reaction was started by addition of 10 µl catalyst solution and the reaction mixture was incubated for 60 min at 37 °C, 800 rpm. After the click reaction the samples were purified and buffer exchanged to D-PBS with Amicon Ultra centrifugal filters (10 kDa cutoff) according to the manufacturer's recommendation.

5.2.2. Reaction conditions on solid phase

A 1 mM Cu^{I} catalyst solution was freshly prepared by addition of 25 µl freshly prepared 100 mM sodium ascorbate solution to 75 µl 1:4 $CuSO_4$/THPTA solution (1 mM and 4 mM final concentration respectively). The click reactions were performed in 300 µl final volume on a thermomixer. Therefore, 75 µl of a 1 M azide solution in DMSO was added to the solid support. 165 µl water, 30 µl 1 M phosphate buffer (pH 7) and 30 µl catalyst solution were mixed and added to the solid support. The suspension was incubated for 60 min at 37 °C, 1400 rpm. The solid support was washed with 500 µl water and 500 µl acetonitrile. To ensure full conversion, the click reaction and washing was repeated twice.

5.2.3. Determination of reaction yield (test-oligos)

After the click reaction, short test-oligonucleotides with only one alkyne modification (Test-EdU and Test-C8-dU) were directly analyzed by HPLC-MS as described in section 5.3.1.

5.2.4. Determination of reaction yield (libraries)

Libraries composed of a mixture of different sequences, containing varying amounts of alkyne modifications can not directly be analyzed by HPLC-MS. Instead, the strands were enzymatically digested to the individual nucleosides, which in turn could be analyzed by HPLC-MS as described in section 5.3.2.

5.2.4.1. Enzymatic digestion to nucleosides

For the enzymatic analytical digestion to nucleosides, 3 µl of the supplied 10 x S1 nuclease reaction buffer (Life Technologies) was added to 200 pmol DNA in 27 µl water. After addition of 0.5 µl S1 nuclease (1000 U/µl), the samples were incubated for 60 min at 37 °C, 800 rpm. 3.5 µl 10 x alkaline phosphatase buffer (Promega) and 0.5 µl of alkaline phosphatase (CIAP) (1 U/µl), 0.5 µl snake venom phosphodiesterase I (5 mU/µl) and 0.5 µl Benzonase nuclease (250 U/µl) were added. The samples were incubated for 120 min at 37 °C, 800 rpm. After digestion, the samples were heated to 95 °C for 3 min and centrifuged for 3 min at 12000 rcf. 20 µl of the supernatant was further analyzed by HPLC-MS.

5.3. High-performance liquid chromatography and mass spectrometry (HPLC-MS)

5.3.1. DNA

5.3.1.1. Ion-pairing reversed-phase high performance liquid chromatography (IP-RP-HPLC)

Reversed-phase ion-pairing chromatography was used for separation and purification of DNA strands. The mechanism of reversed-phase chromatography is based on the varying retention of a compound on the non-polar stationary phase as a function of its polarity. Polar compounds interact less with the non-polar stationary phase and are eluted first by the flow of the polar mobile phase. To enhance the retention of the (at pH 7) negatively charged DNA, triethylammonium hexafluoroisopropanol (TEA/HFIP) is added as an ion-pairing reagent to the mobile phase, providing a less polar counter-ion for the DNA. TEA/HFIP has the advantage over the more frequently used triethylammonium acetate (TEA/OAc) that it is better compatible with ESI-MS. For separation of DNA an analytical 1100 series HPLC system with an Hypersil ODS (C18) (Agilent) 2×100 mm, 5 µm reverse phase column was used. As mobile phase a gradient (from 0 % to 30 % ACN in 20 min) of 10 mM triethylammonium and 100 mM hexafluoroisopropanol (TEA/HFIP) solution and acetonitrile (ACN) was used (flow: 0.5 l/min).

5.3.1.2. Electrospray ionization mass spectrometry (ESI-MS)

Mass spectrometry was done with a HTC Esquire (Brucker). Samples were measured using ultra scan in the negative mode with the following settings: nebulizer: 50 psi, dry gas: 10 l/min, dry temperature: 365 °C, SPS: 1000 m/z, ICC: 70000, scan: 500-1500 m/z.

5.3.2. Nucleosides

5.3.2.1. Reversed-phase high performance liquid chromatography (RP-HPLC)

For separation of nucleosides after an enzymatic digestion an analytical 1100 series HPLC system with an Phenomenex Synergi Fusion-RP 2×50 mm, 4 µm reverse phase column was used. As mobile phase a gradient of 0.1 % ammonium acetate in water (pH 7) and acetonitrile (ACN) was used (flow: 0.5 l/min). The following program was used: 10 min 0 % ACN followed by a gradient from 0 % to 30 % ACN in 20 min.

To improve the resolution later the HPLC conditions were optimized as follows: A Hypersil ODS (C18) (Agilent) 2.1×100 mm, 5 µm reverse phase column was used. As mobile phase a gradient of 20 mM ammonium acetate in water (pH 4.5) and ACN was used (flow: 0.3 l/min). The following program was used: from 0 % to 3 % ACN in 9 min followed by a gradient from 3 % to 40 % ACN in 11 min.

5.3.2.2. Electrospray ionization mass spectrometry (ESI-MS)

Mass spectrometry was done with a HTC Esquire (Brucker). Samples were measured using standard enhanced mode (alternating) with the following settings: nebulizer: 40 psi, dry gas: 9 l/min, dry temperature: 365 °C, SPS: 400 m/z, $ICC_{neg.}$: 50000, $ICC_{pos.}$: 100000, scan: 120-800 m/z.

5.4. SELEX

5.4.1. Streptavidin beads

5.4.1.1. Biotinylation

The target protein (APC from Haematologic Technologies) was biotinylated with NHS-PEG4-Biotin (Life Technologies) according to the manufacturer's recommendation. Briefly, 200 pmol APC was incubated with 20 equivalents of NHS-PEG4-Biotin, freshly prepared in DMSO to a final volume of 50 µl. The solution was incubated for 1 h on ice and 15 min at 25 °C and subsequently purified with Amicon Ultra centrifugal filters (10 kDa cutoff) according to the manufacturer's recommendation.

5.4.1.2. Bead preparation

200 pmol biotinylated protein was immobilized on 5 mg streptavidin coated magnetic beads (Dynabeads M-280 Steptavidin) in a total volume of 500 µl SELEX buffer, according to the manufacturer's recommendation.The loaded beads were stored at 4 °C until further usage.

5.4.1.3. Selection conditions

The selections were carried out in SELEX buffer based on Dulbecco's phosphate buffered saline (D-PBS).[135] To reduce unspecific interactions, the buffer was enriched with 1 mg/ml bovine serum albumin (BSA), 0.1 % TWEEN-20 and 0.1 mg/ml salmon sperm DNA. 500 pmol of the start library was incubated with 500 µg of non-conjugated streptavidin beads to remove the unwanted matrix binding sequences. The supernatant was incubated for 30 min at 37 °C and 800 rpm with 500 µg APC-conjugated beads in a total volume of 200 µl. The beads were thoroughly washed two times (30 s and 5 min) with 200 µl SELEX buffer to remove non-specific binding DNA and the enriched library was eluted with 85 µl water from the beads by heat denaturation (80 °C for 3 min). The eluted DNA was PCR amplified in a total volume of 800 µl and purified via spin-columns (see section 5.1.3, page 87). PCR product formation was carefully monitored by gel electrophoresis and stopped as soon as the desired product band could be detected on ethidium bromide stained gels. Single strand displacement of the double-stranded PCR product was carried out by λ-exonuclease digestion, and the resulting single strand DNA was purified via spin-columns. The purified single strand DNA (approx. 10 pmol) was applied in the next SELEX round. For each selection cycle, the PCR product and the purified single strand DNA were analyzed by gel electrophoresis on a 4 % agarose gel. To increase the selection pressure, the stringency of the washing steps was gradually increased up to a maximum of 8 washing steps during the selection.

5.4.2. Capillary electrophoresis (CE)

A PA 800 Capillary Electrophoresis system (Beckman Coulter) with 32 Karat software (Beckman Coulter) was used to acquire CE data. A 50 cm long, 50 µm inner diameter capillary having a 40 cm length to detector was employed for the CE-SELEX selections. Before each CE run the capillary was rinsed with water and B15 buffer (see section 6.1.3, page 99) for 5 min. The capillary cartridge was maintained at 25 °C and the sample chamber was held at 20 °C. Electric fields of 25 kV were utilized to drive the electrophoresis separations. The samples (47.5 nl) were injected into the capillary, pre-filled with the run buffer, by a pressure pulse of 5 s × 4 psi. For the selection and binding assays samples were monitored using UV detection at 254 nm.

5.4.2.1. Selection conditions

First SELEX round: The indole-azide clicked starting library (3 µM final concentration) and APC (3 µM final concentration) were incubated in B15 buffer. 3 consecutive 20 min separation runs were performed as described above and collected in 80 µl B15 buffer. The eluted DNA (enriched library) was PCR amplified as described in section 5.1.2 (page 87). The first PCR product was divided into 16 aliquots and further PCR amplified. As a backup of this SELEX round 10 % of

the purified dsDNA was stored at -20 °C. The purified, amplified double stranded library was processed to single strand DNA as described in section 5.1.5 (page 88). Before the next SELEX round, the ssDNA was functionalized with indole-azide as described in section 5.2.1 (page 90).

SELEX rounds 2-5: The enriched library was incubated with APC in B15 buffer. To increase the selection pressure, the DNA and APC concentration was reduced as summarized in the following table. In total five rounds of SELEX, following the same protocol as in the first round, were performed.

Table 5.2. – Summary of protein and DNA concentrations used for CE-SELEX.

Round	APC [µM]	DNA [µM]
1	3	3
2	0.23	3
3	0.3	0.3
4	0.1	0.06
5	0.3	0.12

5.4.3. Cobalt-beads

5.4.3.1. Bead preparation

25 µl of resuspended cobalt Dynabeads from a His-Tag Isolation and Pulldown kit (40 mg/ml) were washed 3 times with 500 µl SELEX buffer and resuspended in 500 µl SELEX buffer. 500 µl of a 3.5 µM protein solution in SELEX buffer was added and incubated for 30 min at 25 °C, 800 rpm. The protein-loaded beads were washed 3 times with 500 µl SELEX buffer and resuspended in 500 µl SELEX buffer. To confirm C3-GFP conjugation to the cobalt beads, the GFP was verified by flow cytometry.

5.4.3.2. Selection conditions

The selection was carried out in a selection buffer based on Dulbecco's phosphate buffered saline (D-PBS) enriched with 1 mg/ml bovine serum albumin (BSA), 0.1 % TWEEN-20 and 0.1 mg/ml salmon sperm DNA. As a target, His6 tagged C3-GFP immobilized on magnetic beads was used. The library was incubated with C3-GFP coupled magnetic beads for 30 min at 37 °C and 800 rpm. The beads were washed three times with 200 µl SELEX buffer to remove non-specific binding DNA, and the enriched library was eluted from the beads by addition of 100 µl of a 300 mM imidazole solution and incubation for 15 min at 37 °C. The eluted DNA was PCR amplified in a total volume of 800 µl and purified. The PCR product formation was monitored on 4 % agarose gels. 10 % of the purified PCR product from each selection round was stored as a backup. Single strand displacement of the double-stranded PCR product was carried out by λ-exonuclease digestion.

The purified single strand DNA was again functionalized with indole-azide and applied in the next selection round. To increase the selection pressure, the amount of magnetic beads was reduced and the washing time was increased during the SELEX. To further increase the selection pressure, from round 12 on an indole-azide functionalized double-stranded DNA library with different primer binding sites was used as a competitor for non-specific interactions. In order to deplete the library of sequences binding to the bead matrix, from selection round three onwards, the library was pre-incubated with 50 µl non-conjugated magnetic beads for 30 min at 37 °C and 800 rpm. The supernatant was recovered and incubated with the C3-GFP loaded beads. The detailed selection conditions for each selection round are summarized in table 5.3.

Table 5.3. – Summary of GFP-SELEX conditions.

Round	non-conjugated Beads	C3-GFP Beads	Wash	Competitor
1	-	50 µl	3 × 30 s	-
2	-	50 µl	3 × 30 s	-
3	50 µl	50 µl	2 × 30 s + 1 × 5 min	-
4	50 µl	25 µl	2 × 30 s + 1 × 5 min	-
5	50 µl	10 µl	1 × 30 s + 2 × 5 min	-
6 - 11	50 µl	5 µl	3 × 5 min	-
12	50 µl	5 µl	3 × 5 min	5 pmol
13	50 µl	5 µl	3 × 5 min	15 pmol
14	50 µl	5 µl	3 × 5 min	40 pmol
15	50 µl	5 µl	3 × 5 min	40 pmol

5.5. Interaction analysis

5.5.1. Filter-retention assay

Equilibrium dissociation constants were determined by nitrocellulose filter-retention with radioactive labelled DNA.

5.5.1.1. Kinasation

Purified single stranded DNA was 5' labelled with ^{32}P from γ-ATP using T4 polynucleotide kinase (NEB) by incubating 41 µl DNA solution (ca. 10 pmol) with 5 µl T4 PNK buffer, 2 µl T4 PNK and 2 µl radioactive ATP (10 µCi/µl) (Perkin Elmer) for 30 min at 37 °C. The labelled DNA was purified from unreacted ATP with G25 spin columns (GE Healthcare). The DNA was functionalized with indole-azide and further purified with NucleoSpin columns.

5.5.1.2. Measurement

After incubation in selection buffer with increasing protein concentrations for 30 min at 37 °C, the solution was filtered through nitrocellulose and the amount of bound aptamer was quantified by autoradiography in a phosphorimager (FLA 3000 from Fuji).

5.5.1.3. Data analysis

Raw binding data were corrected for non-specific background binding and equilibrium dissociation constants were calculated via nonlinear regression assuming a 1:1 binding model with Prism 5.0f (GraphPad Software).

5.5.2. Microscale thermophoresis (MST)

5.5.2.1. Measurement

All measurements were performed on the Monolith NT.115 (NanoTemper) instrument. For every measurement, 16 glass capillaries were prepared by dipping the capillaries in 10 µl of a protein dilution series containing a constant concentration of Cy5 labeled DNA (10 nM).

5.5.2.2. Data analysis

Raw binding data were baseline subtracted, and equilibrium dissociation constants were calculated via nonlinear regression assuming a 1:1 binding model with Prism 5.0f (GraphPad Software).

5.5.3. Flow cytometry

Binding interactions with proteins immobilized on magnetic beads were measured using a FACSCanto II (BD Biosciences). 5 µl of the respective magnetic beads were incubated for 30 min at 37 °C with 500 nM DNA in SELEX buffer in a total volume of 50 µl. The beads were washed two times with 200 µl SELEX buffer and resuspended in D-PBS. 50000 events were measured in the flow cytometer, and the mean Cy5 fluorescence intensity (APC-A channel) was analyzed.

5.5.3.1. Data analysis

All values were normalized to the value obtained by C12-1. Mean and standard deviation of at least two independent experiments were calculated and visualized with Prism 5.0f (GraphPad Software).

6. Materials

6.1. Reagents

If not stated otherwise all chemicals were bought from Sigma-Aldrich in analytical grade or the highest available purity. All solutions were prepared using (Type1) ultrapure water (18 MΩ × cm).

6.1.1. Chemicals

Name	Details / CAS	Supplier
Agarose	Agarose LE	Genaxxon
Alkaline phosphatase (CIAP)	M2825	Promega
γ-^{32}P-ATP	NEG502A	Perkin Elmer
Benzonase	9025-65-4	Sigma
C8-dUTP	BCT-05	BaseClick
DNA ladder	GeneRuler ultra low range	Life Technologies
D-PBS (10X)	D 1283	Sigma
Dynabeads	M-280 Streptavidin	Life Technologies
Dynabeads	His-Tag Isolation & Pulldown	Life Technologies
EdUTP	BCT-08	BaseClick
Hexafluoro-2-propanol (HFIP)	LC-MS grade	Sigma
λ-Exonuclease	EN0561	Thermo Scientific
NHS-PEG4-Biotin	21329	Life Technologies
Nuclease S1	18001-016	Life Technologies
Neutravidin-800	22853	Life Technologies
PWO DNA polymerase	M3002	Genaxxon
Phosphodiesterase I (PDE I)	9025-82-5	Sigma
T4 Polynucleotide kinase (T4-PNK)	M0201	NEB
Triethyamine (TEA)	LC-MS grade	Sigma
THPTA	BCMI-006	BaseClick

6.1.2. Commercial kits

Application	Kit	Supplier
Adapter ligation	TruSeq DNA PCR-Free (LT) sample prep. kit	Illumina
Gel filtration	G25 spin columns	GE Healthcare
PCR purification	NucleoSpin Gel and PCR Clean-up	Macherey-Nagel
Plasmid purification	NucleoSpin Plasmid	Macherey-Nagel
Size exclusion	Amicon-Ultra	Merck Millipore
TOPO TA cloning	TOPO TA cloning kit for sequencing	Life Technologies

6.1.3. Buffers and solutions

Name	Composition	Concentration
B15 CE-SELEX buffer (pH 8.3)	Tris	25 mM
	NaCl	30 mM
	KCl	1 mM
	$CaCl_2$	1 mM
	$MgCl_2$	1 mM
CuAAC catalyst solution	THPTA	4 mM
	$CuSO_4$	1 mM
	Sodiumascorbate	25 mM
D-PBS (pH 5.3)	NaCl	137.9 mM
	Na_2HPO_4	8.1 mM
	KCl	2.7 mM
	KH_2PO4	1.5 mM
	$CaCl_2$	0.9 mM
	$MgCl_2$	0.5 mM
DNA loading buffer (6x)	Glycerol	60 %
	Tris	10 mM
	Na_2EDTA (pH 8.0)	60 mM
	Xylencyanol	0.03 %
HPLC-MS buffer A (TEA/HFIP)	TEA	10 mM
	HFIP	100 mM
Resuspension buffer	Tris-HCl (pH 8.5)	10 mM
	TWEEN-20	0.1 %
SELEX buffer	D-PBS	1 X
	BSA	0.1 %
	TWEEN-20	0.1 %
	Salmon-sperm	0.1 mg/ml
TBE buffer	Tris	89 mM
	Boric acid	89 mM
	Na_2EDTA (pH 8.0)	2 mM

6.2. Nucleic acids

All nucleic acids were ordered HPLC purified by Ella Biotech.

FT2-F

5'-CACGACGCAAGGGACCACAGG-3'

FT2-F-Cy5

5'-Cy5-CACGACGCAAGGGACCACAGG-3'

FT2-R

5'-TGCCTCTGCGGTGTCGTGCTG-3'

FT2-R-(P)

5'-Phosphate-TGCCTCTGCGGTGTCGTGCTG-3'

FT2-N42 (T) (N: dA:dG:dC:T = 1:1:1:1)

5'-CACGACGCAAGGGACCACAGG-N42-CAGCACGACACCGCAGAGGCA-3'

FT2-N42 (C8-dU) (N: dA:dG:dC:C8-dU = 1:1:1:1)

5'-CACGACGCAAGGGACCACAGG-N42-CAGCACGACACCGCAGAGGCA-3'

FT2-N42 (EdU) (N: dA:dG:dC:EdU = 1:1:1:1)

5'-Cy5-CACGACGCAAGGGACCACAGG-N42-CAGCACGACACCGCAGAGGCA-3'

D3-F

5'-GCTGTGTGACTCCTGCAA-3'

D3-R

5'-GGAGACAAGATACAGCTGC-3'

D3

5'-GCTGTGTGACTCCTGCAA-N43-GCAGCTGTATCTTGTCTCC-3'

C12

5'-Cy5-CACGACGCAAGGGACCACAGGGXXGGAAGCGACGGGACGGXAAGGCXXGGG-CCCCAAGGAGXGCAGCACGACACCGCAGAGGCA-3' (X = EdU)

C12sc

5'-Cy5-CACGACGCAAGGGACCACAGGGAGXGGAGGCGXGAXGGACGAGAGXCGXGG-CGCCAXCAACGGCAGCACGACACCGCAGAGGCA-3' (X = EdU)

P5

5'-Cy5-CACGACGCAAGGGACCACAGGXAGGAGGGAGCAAGGXACGXXGCAGCAXGA-CGGGACCCGACACAGCACGACACCGCAGAGGCA-3' (X = EdU)

Test-C8-dU

5'-GCACTGTXCATTCGCG-3' (X = C8-dU)

Test-EdU

5'-GCACTGTXCATTCGCG-3' (X = EdU)

6.3. Proteins

Activated Protein C (APC), HCAPC-080, Lot: CC0405-1.0MG, Haematologic Technologies
Light chain: ANSFLEELRHSSLERECIEEICDFEEAKEIFQNVDDTLAFWSKHVDGDQCLVLP-
LEHPCASLCCGHGTCIDGIGSFSCDCRSGWEGRFCQREVSFLNCSLDNGGCTHYCLEEVG-
WRRCSCAPGYKLGDDLLQCHPAVKFPCGRPWKRMEKKRSHL
Heavy chain: DTEDQEDQVDPRLIDGKMTRRGDSPWQVVLLDSKKKLACGAVLIHPSWVL-
TAAHCMDESKKLLVRLGEYDLRRWEKWELDLDIKEVFVHPNYSKSTTDNDIALLHLAQP-
ATLSQTIVPICLPDSGLAERELNQAGQETLVTGWGYHSSREKEAKRNRTFVLNFIKIPVVPH-
NECSEVMSNMVSENMLCAGILGDRQDACEGDSGGPMVASFHGTWFLVGLVSWGEGCG-
LLHNYGVYTKVSRYLDWIHGHIRDKEAPQKSWAP

Cycle 3 GFP (C3-GFP) [144], 13105-S07E, Life Technologies
MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTL-
VTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTL-
VNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIKANFKIRHNIEDGSVQLA-
DHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK

Monomeric E-GFP (mE-GFP) [144], kindly provided by Volkmar Fieberg (Caesar, Bonn)
MGSHHHHHHENLYFQGSMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYG-
KLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRPDHMKQHDFFKSAMPEGYVQERTIFFK-
DDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNG-
IKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVL-
LEFVTAAGITLGMDELYK

6.4. Equipment

Equipment	Details	Supplier
Agarose chamber equipment		In house construction
Analytical balance		Sartorius
CE	PA 800	Beckman Coulter
Flowcytometer	FACSCanto II	BD Biosciences
HPLC	1100	Agilent
Hypersil ODS (C18)		Agilent
Imaging system	Odyssey	Li-Cor
Magnetic rack		Life Technologies
Microscale thermophoresis	Monolith NT.115	NanoTemper
MS	Esquire HTC	Bruker
Nitrocellulose membrane	Protran 0.45 µM	Schleicher and Schuell
PCR Thermocycler	Mastercycler Personal	Eppendorf
Phosphorimager (PI)	FLA-3000	Fuji
PI screens and cassettes		Fuji
Spectrophotometer	NanoDrop 2000	Thermo Scientific
Pipets		Eppendorf
Plastic consumables		Sarstedt
Synergi Fusion-RP		Phenomenex
Thermomixer		Eppendorf
UV Transilluminator		VWR
Water purification system		Merck Millipor

A. Appendix

```
C1:  TCATATCCCGTATGGGCTTGAGGAACGTTTCGCGAGATACTA
C2:  TCATATCCCGTATGGGCTTGAGGAACGTTTCGCGAGATACTA
C3:  TCATATCCCGTATGGGCTTGAGGAACGTTTCGCGAGATACTA
C4:  TCATATCCCGTATGGGCTTGAGGAACGTTTCGCGAGATACTA
C5:  TCATATCCCGTATGGGCTTGAGGAACGTTTCGCGAGATACTA
C6:  TCATATCCCGTATGGGCTTGAGGAACGTTTCGCGAGATACTA
C7:  TCATATCCCGTATGGGCTTGAGGAACGTTTCGCGAGATACTA
C8:  TCATATCCCGTATGGGCTTGAGGAACGTTTCGCGAGATACTA
C9:  GCACTCTGCCTTATCCCGTATGGGTGGGGTCTATGCGTCCTA
C10: GCACTCTGCCTTATCCCGTATGGGTGGGGTCTATGCGTCCTA
C11: GCACTCTGCCTTATCCCGTATGGGTGGGGTCTATGCGTCCTA
C12: GCACTCTGCCTTATCCCGTATGGGTGGGGTCTATGCGTCCTA
C13: GCACTCTGCCTTATCCCGTATGGGTGGGGTCTATGCGTCCTA
C14: GCACTCTGCCTTATCCCGTATGGGTGGGGTCTATGCGTCCTA
C15: GCACTCTGCCTTATCCCATATGGGTGGGGTCTATGCGTCCTA
C16: TCATATCCCGTATGGGCTTGCCGGCTAACCTCTTGTTACCT
C17: CCCCTATCCCGTATGGGCTTGCAGAGTCTTCGGGTCACGCTT
C18: TCGTTATCCCGTATGGGCTTGCCGGCTAACCTCTTGTTACCT
C19: TCGTATCCCGTATGGGGTTGCGTTTTCAAACGACAACGTATT
```

Figure A.1. – Sequencing results for the test-SELEX against APC
The sequence of 18 full-length mono-clones found in the 11th SELEX round could be identified.

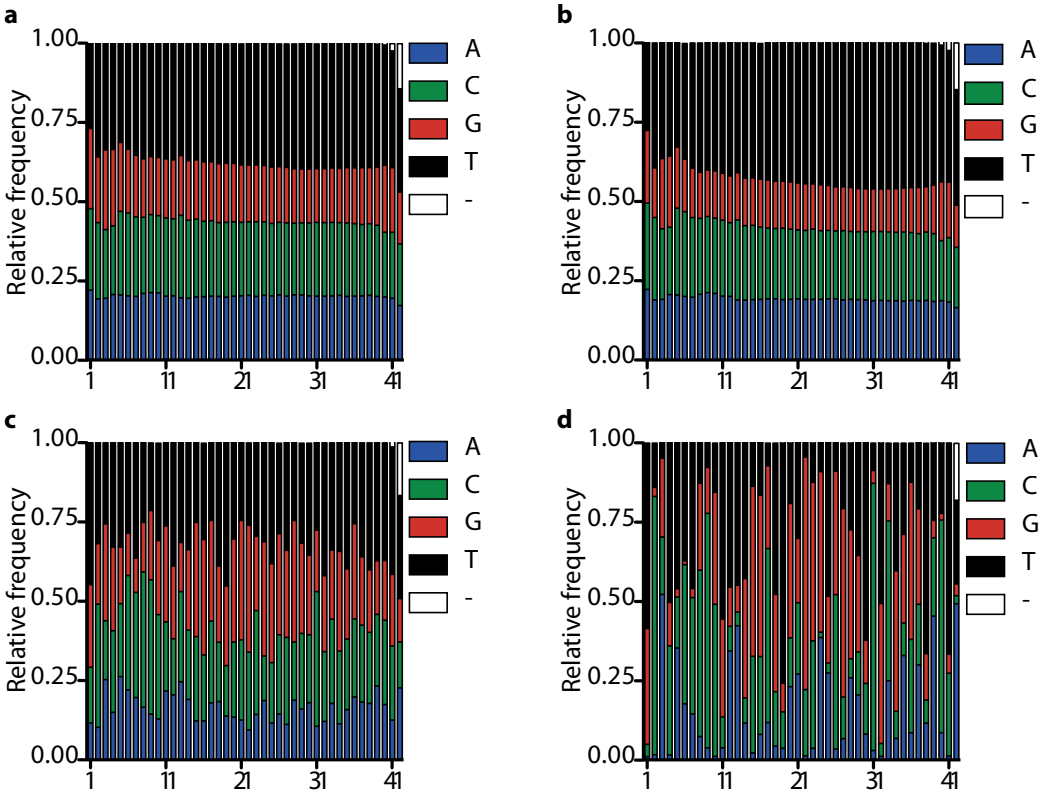

Figure A.2. – Character statistics for the 3rd, 5th, 7th and 9th round of the test-SELEX
a) Statistical base distribution in the 3rd selection round. b) Statistical base distribution in the 5th selection round. c) Statistical base distribution in the 7th selection round. d) Statistical base distribution in the 9th selection round. Until round 5 only little enrichment could be observed, from round 7 on, a strong enrichment was achieved.

```
Patter 1:  TCATATCCCGTATGGGCTTGAGGAACGTTTCGCGAGATACTA
Patter 2:  GCACTCTGCCTTATCCCGTATGGGTGGGGTCTATGCGTCCTA
Patter 3:  TCGTTATCCCGTATGGGCTTGCCGGCTAACCTCTTGTTACCT
Patter 4:  TTCCTCTATCCCGCCATGGGTGCGTCAATGTTTACGGGTTT
Patter 5:  GCCGCCATGTGGTATTATCCCGCATGGGCTCGCGTTCATTC
Patter 6:  TCGTATCCCGTATGGGGTTGCGTTTTCAAACGGCAACGTATT
Patter 7:  CCCCTATCCCGTATGGGCTTGCAGAGTCTTCGGGTCACGCTT
Patter 8:  TCATATCCCGTATGGGCTTGCCGGCTAACCTCTTGTTACCT
Patter 9:  TTGCCTAGATGTATGTGATCCCTATCCCGGATGGGCTTACGG
Patter 10: GCACTCTGCCTATCCCGTATGGGTGGGGTCTATGCGTCCTA
Patter 11: TCATATCCCGTATGGGCTTGTGAGGGCCATGTATGCCGGAGT
Patter 12: GTTTCCGGGCCTGGTACTATCCCGGATGGGCTTGCGTCATTT
Patter 13: CAGGTCCACGATAGGTGGACATATCCCGGATGGGTGCCATCC
Patter 14: TTGGATTTTCTCCGATCTGCGGCCCATATCCCGGATGGGTTG
Patter 15: CCCGTATGGGCTTGCAGAGTCTTCGGGTCACGCTT
```

Figure A.3. – Consensus sequence of the 15 most abundant pattern identified by NGS from the test-SELEX against APC

```
C1:  GCATACCCCGAGCCTAACTGCGAAGCAATGCCAGCGACTGGC
C2:  AAAAGAGGGGGGAGGCGGGAGCGCACGACTTTGGGAACCCGA
C3:  GAGGCGGACACCGCGCGGAGCTGAGGACGCCGAATCGCTTCA
C4:  GCTCGATCGAGGTAAAAAGAGGGAGCCGACGGGAGCCTGGAA
C5:  GGAGACCAAGGGCGGGGAGAGGCCGCACGGCTTTCGGAAGCG
C6:  TAGGCCGCGGGAACTGGGGCCAACTGTCGGCAGCCCAAGTTG
C7:  AGCAGAGCCGGCGAGGCTGAACCCTAGAGACGCGATCATCGG
C8:  GGCGAGTATGGCCATCCGGCGCGCCAAGAGAAGAATACAAAG
C9:  ACGCTGAGGGGCCTAAAGGCGCAGGGGATCTAGGA
C10: TGCGTAACGAGGCCTGTGAGACCAGGTAACGGTAGGCCAA
C11: GTGTATAGCGGC
C12: ACAATGGTAGC
C13: TTTCGGG
C14: CTTAGGAGGCGGGGATA
C15: TTTCGGAA
C16: GTGTATAGCGGC
C17: GGCATAAAGTCTGGG
```

Figure A.4. – Sequencing results for the SELEX with C8-dU against APC
The sequence of 8 full-length mono-clones found in the 10th SELEX round could be identified

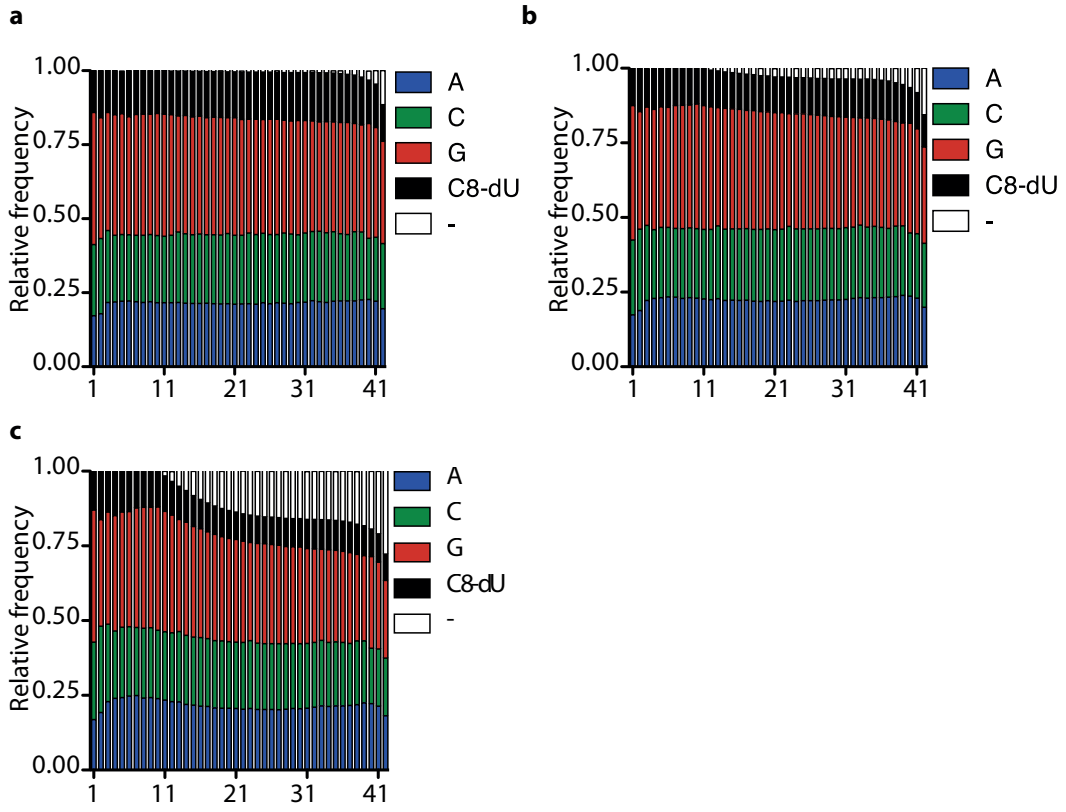

Figure A.5. – Character statistics for the 4th, 6th and 8th round of the SELEX with C8-dU
a) Statistical base distribution in the 4th selection round. b) Statistical base distribution in the 6th selection round. c) Statistical base distribution in the 8th selection round. No specific enrichment could be observed.

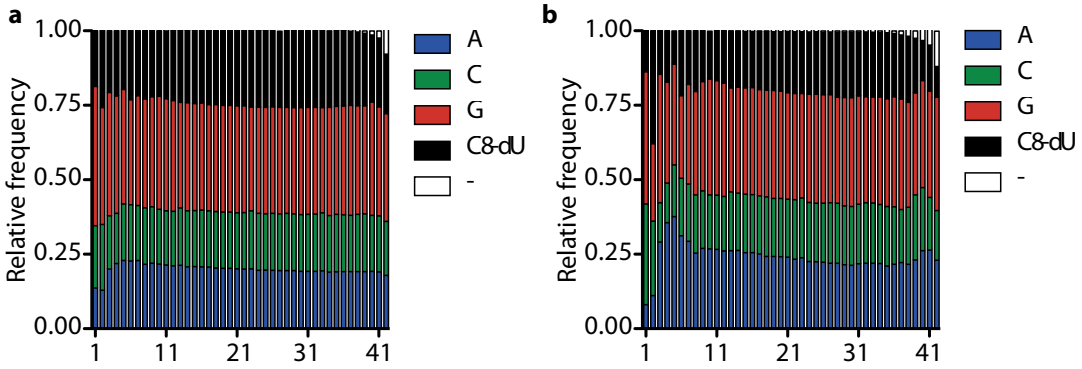

Figure A.6. – Character statistics for the 1st and 3rd round of the CE-SELEX
a) Statistical base distribution in the 1st selection round. b) Statistical base distribution in the 3rd selection round. No specific enrichment could be observed.

a

```
          1           11          21
C3-GFP: MSKGEELFTG VVPILVELDG DVNGHKFSVS
mE-GFP: VSKGEELFTG VVPILVELDG DVNGHKFSVS

31          41          51          61
GEGEGDATYG KLTLKFICTT GKLPVPWPTL VTTFSYGVQC
GEGEGDATYG KLTLKFICTT GKLPVPWPTL VTTLTYGVQC

71          81          91          101
FSRYPDHMKR HDFFKSAMPE GYVQERTISF KDDGNYKTRA
FSRYPDHMKQ HDFFKSAMPE GYVQERTIFF KDDGNYKTRA

111         121         131         141
EVKFEGDTLV NRIELKGIDF KEDGNILGHK LEYNYNSHNV
EVKFEGDTLV NRIELKGIDF KEDGNILGHK LEYNYNSHNV

151         161         171         181
YITADKQKNG IKANFKIRHN IEDGSVQLAD HYQQNTPIGD
YIMADKQKNG IKVNFKIRHN IEDGSVQLAD HYQQNTPIGD

191         201         211         221
GPVLLPDNHY LSTQSALSKD PNEKRDHMVL LEFVTAAGIT
GPVLLPDNHY LSTQSKLSKD PNEKRDHMVL LEFVTAAGIT

231
HGMDELYK
LGMDELYK
```

b

V1a	Optimization of translation initiation
F64L	Improve folding at 37°C
S65T	Promote chromophor ionization
Q80R	Neutral (PCR Error)
F99S	Improve folding at 37°C / Reduce aggregation
M153T	Improve folding at 37°C / Reduce aggregation
V163A	Improve folding at 37°C / Reduce aggregation
A206K	Interference with dimer interface
H231L	Neutral

Figure A.7. – Comparison of C3-GFP and mE-GFP
a) Sequence alignment of ”cycle 3” GFP (C3-GFP) and monomeric enhanced GFP (mE-GFP). b) Phenotypical description associated with these point mutations compared to the wild type protein. In red are mutations on the surface of the protein. In grey are mutations inside the barrel structure.

```
C12 GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG
C33 GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG
C32 GTTGGAAGCGACGGGACGGTAAGGCTCGGGCCCCAAGGAGTG
C38 GTTGGAAGCGACGGAACGGTAAGGCACGGGCCCCAAGAAGTA
C27 GCTGGAAGCGACGGGACGGTAAGGCACGGGCCCCAAGGAGTG
C26 GTTGAAAGCGACGGGACGGTAAGGCTCGGGCCCCAAGGAGTG
C29 GTTGGAAGCGACGGTACGGTAAGGCTTGGACCCCGATAAGTG
C15 CCGGGGAGTGGGGGAGGGTCTGGGGCCAGAACCAAACCACAA
C25 CCGGGGAGGGGGGGAGGGTCTGGGGCTAGGATCGAACCGCAA
C7  TGACAAACATGGACTAGGAAGGTGGGCTCCTAGGTCAGAACT
C23 GCGCGGGGGTGGGGGGGAGGGGGATGGCCAGGGCGTAAAAAT
C28 GGAGGAGGGGGATCCGGCCTCAGTAGGGGGCTTGGGAAACAA
C1  GGGGGACTGTCAGCCGGGGAGTGGCGGGGGGAAACAGAGGCC
C2  GGGGGCATCACACAGATATGCTCCCGAGGGGGGGGGACAGGG
C4  TTGGAGACAAGGAACGTAAAGAGGGCATGTGGGCTTAAATAG
C5  TGGGGATGGGCACAGTGGACTAGTCGGGGCAAGAGGGGGGAT
C8  GGGGGCGGTCATATAAACCCAAGGGGGCCTGGGCATAGGGGG
C10 GCGGGATGTGGGCCGTCACGACTGGGGCGGGAGGGGGAGAGG
C11 GGGGGCACGGTACGTAGGGGCGGGGAGGTACGACAGTGAGGG
C13 GGGGAGAGTGGGACCGTAGCAGAACATGGGGTCTCGGAAAGG
C16 GGAGAATTGGGGCAGGGACTCAGACAGTAGCCAATAGGGAAA
C18 GTATGGGGACGAGAGATTGGGGACATGGCCCAACAGTGGGAA
C19 GGGTACACCTGGGTACTAAGGGGAGGGGGGGGCACCGGGG
C20 GGGGGGGGGGCGCGAGTAAGTACAAGGGAGAGGGAGGGCCCAGC
C22 GGGTACAGGGTACAAGTACAAATACCGGGGGAGGGAGGGGAA
C24 AAATGGGGGGGGTAGGGAGTGATAGACTGGAGCTAGGGGTGAA
C30 GGTGTGAGGGGCAACGGGGAAGGGGAGGTCACGACACTGGGG
C31 GGCAACAGGGTCGTGGGGTACAAGAACGAAGTACATGAGTGA
C34 GGGGGAGGGGGGCGCTCGAGACCAAACGGGGGGTTGGTAAC
C35 GGGGGCAGTGTCATCGTCGGGGAGGGGAGGGGGAAGCCACAT
C36 GACTACAGGGACAAAGTCAATGGGGAAAATAGGGGACACAAC
C37 GGAGGAGGGTGATCTGGCCTCAGCAGGGGGCTTGGGAAACAG
C39 CAATCAATGGTCGGGGAGGACGGGGGGGAATTGGGAAGCTGG
C40 GGGGTTGGGGAAGAAAACGGTGGGATTGGGCAATGGGGCTAGG
C41 GGGGGGAAATCGCCAGTGGGGCAGATCGGGGACAGTGGCCTA
C42 GGTCAGGGGCCTTAACAATGGGCAATTGGGGAACAACAAAAT
```

Figure A.8. – Sequencing results for SELEX with EdU against GFP
The sequence of 37 full-length mono-clones found in the 15th SELEX round could be identified

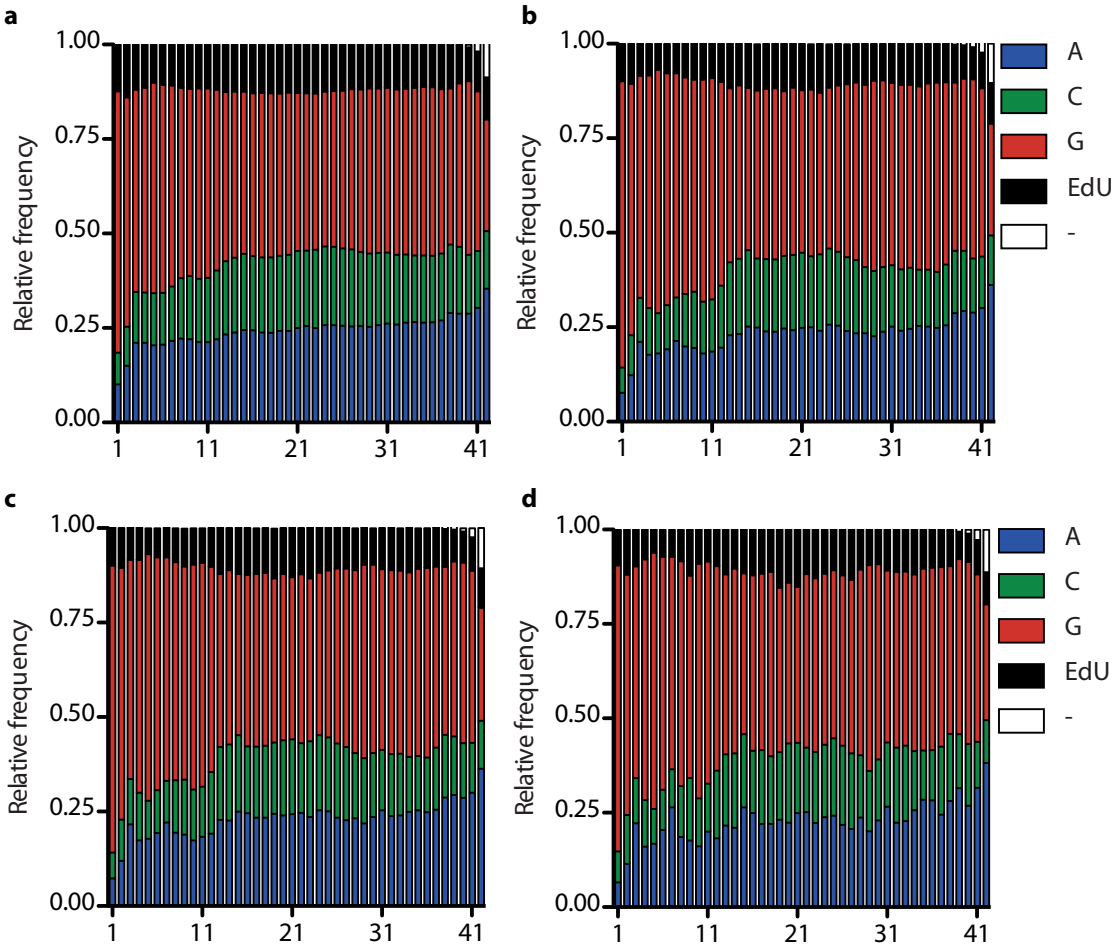

Figure A.9. – Character statistics for the 8th, 10th, 11th and 13th round of the SELEX with EdU againt GFP

a) Statistical base distribution in the 8th selection round. b) Statistical base distribution in the 10th selection round. c) Statistical base distribution in the 11th selection round. d) Statistical base distribution in the 13th selection round.

```
Patter 1:  GTTGGAAGCGACGGGACGGTAAGGCTCGGGCCCCAAGGAGTG
Patter 2:  GGACGGGGCGACCGGGGGAGGGACAGGGGCATCTAACGGCAA
Patter 3:  GGATGGACCGGGGCGGAGGGGGGTAGGCGATGGGGGCACTGGG
Patter 4:  TGACAAACATGGACTAGGAAGGTGGGCTCCTAGGTCAGAACT
Patter 5:  TAGGAGGGAGCAAGGTACGTTGCAGCATGACGGGACCCGACA
Patter 6:  GGAGGAGGGGGATCCGGCCTCAGCAGGGGGCTTGGGAAACAG
Patter 7:  GGCGGCGGGGGGTGCGAGCGCACGGGGGGGACTGGGCTCAGA
Patter 8:  GGGGGGAGGGGGACCGTAGCAGGACAGGGGGGCTTGGGAAGG
Patter 9:  GGGGGGGGCGGACGGGGCGAATGGGCGCGTCAGTGGGAGGGG
Patter 10: GCGGGGCGGGGGCAATGGCCAGTACCAGGGGGGAGGGAGGAC
Patter 11: GGCGGGGAGGGGGGGAGGGATCGGGACAGACCATATGCGTGAG
Patter 12: GGCGGCGGGGGTCGCACAAGGGATCGGGGGGACTGGGCACGT
Patter 13: GCGCGGGGGCGGGGGGGAGGGGGATGGCCAGGGCGTAAGAAT
Patter 14: GGCGGGGAGGGGGGGAGGGATCGGGACAGACCATATGCGTGAG
Patter 15: GTGGGGCATCAACCGTAGGGAGGGCGGGGGACAGGGGCAGGA
```

Figure A.10. – Consensus sequence of the 15 most abundant patterns identified by NGS from the SELEX with EdU against GFP

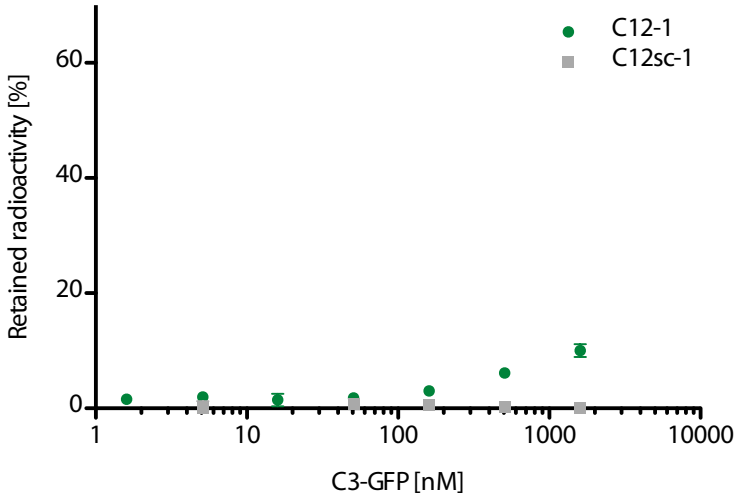

Figure A.11. – Filter-retention analysis of C12-1 and C12sc-1 towards mE-GFP
Neither the aptamer C12-1 (green) nor the control sequence C12sc-1 (grey) bind to mE-GFP.

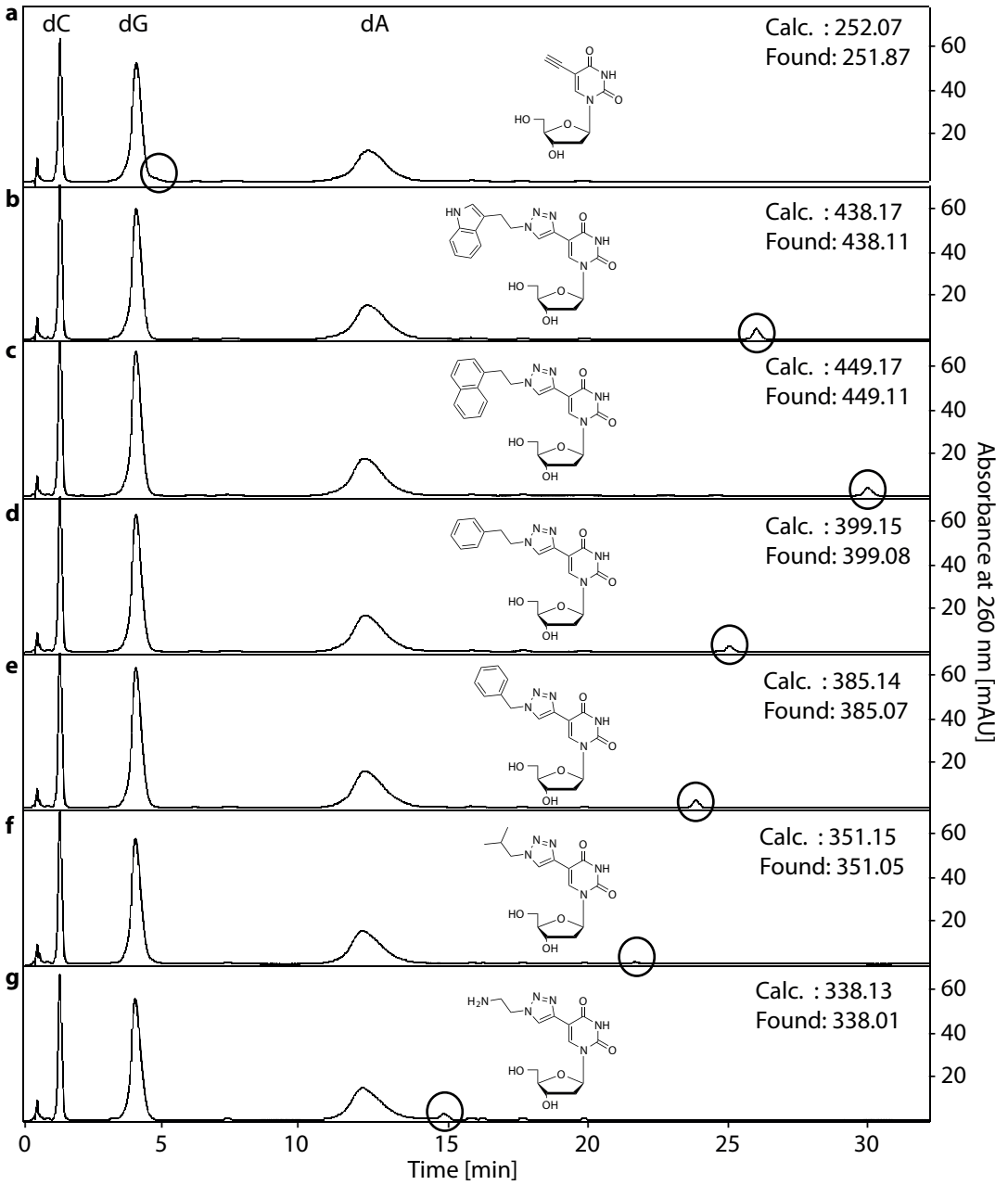

Figure A.12. – HPLC-MS analysis of C12 after enzymatic degradation to single nucleosides
a) Unfunctionalized C12. b) Indole-azide **1** functionalized C12-1. After the reaction no unreacted EdU could be detected by mass spectrometry. c) 1-ethyl-naphthalene-azide **2** functionalized C12-2. After the reaction no unreacted EdU could be detected by mass spectrometry. d) Ethyl-phenyl-azide **3** functionalized C12-3. After the reaction no unreacted EdU could be detected by mass spectrometry. e) Benzyl-azide **4** functionalized C12-4. After the reaction no unreacted EdU could be detected by mass spectrometry. f) Iso-butyl-azide **5** functionalized C12-5. After the reaction no unreacted EdU could be detected by mass spectrometry. g) Ethylamine-azide **6** functionalized C12-6. After the reaction no unreacted EdU could be detected by mass spectrometry.

List of Abbreviations

Abbreviation	Full form
A	2'-Deoxyadenosine
ACN	Acetonitrile
AU	Absorbance units (HPLC)
bp	Base pairs
C	2'-Deoxycytidine
C8-dU	5-(Octa-1,7-diynyl)-2'-deoxyuridine
C8-dUTP	5-(Octa-1,7-diynyl)-5'-O-triphosphate-2'-deoxyuridine
CuAAC	Copper-catalysed azide-alkyne cycloaddition
DCM	Dichloromethane
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide
dsDNA	Double standed DNA
EdU	5-ethynyl-2'-deoxyuridine
EdUTP	5-ethynyl-5'-O-triphosphate-2'-deoxyuridine
EDTA	Ethylendiaminetetraacetate
EPR	Electron paramagnetic resonance
EtOH	Ethanol
g	Gram
G	2'-Deoxyguanosine
HPLC	High performance liquid chromatography
HV	High vacuum
λ_{em}	Emission wavelength
λ_{ex}	Excitation wavelength
l	Litre
K_D	Dissociation constant
NMR	Nuclear magnetic resonance spectroscopy

Abbreviation	Full form
M	Molar mass [g/mol] or Molar concentration [mol/l]
MeOH	Methanol
MS	Mass spectrum
nt	Nucleotides
NTC	No template control
PCR	Polymerase chain reaction
PELDOR	Pulsed electron-electron double resonance
rcf	Relative centrifugal force
RFU	Relative fluorescence units
RNA	Ribonucleic acid
RT	Room temperature (20-25 °C)
SELEX	Systematic evolutions of ligands by exponential enrichment
ssDNA	Single stranded DNA
T	Thymidine
TBTA	Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine
TEA/HFIP	Triethylammonium hexafluroisopropanol
THPTA	Tris(4-(3-hydroxy-propoyl)-[1,2,3]triazol-1-ylmethyl)amine
Tris	Tris(hydroxymethyl)aminomethane
U	Units
UV	Ultraviolet
V	Volt
w/o	Without
W	Watt

List of Figures

List of Tables

Bibliography

[1] Tolle, F., Rosenthal, M., Pfeiffer, F. & Mayer, G. Click Reaction on Solid Phase Enables High Fidelity Synthesis of Nucleobase-Modified DNA. *Bioconjugate chemistry* **27**, 500–503 (2016).

[2] Tolle, F. & Mayer, G. Preparation of SELEX Samples for Next-Generation Sequencing. *Methods Mol. Biol.* **1380**, 77–84 (2015).

[3] Tolle, F., Brändle, G. M., Matzner, D. & Mayer, G. A Versatile Approach Towards Nucleobase-Modified Aptamers. *Angew Chem Int Ed Engl* **54**, 10971–10974 (2015).

[4] Tolle, F., Wilke, J., Wengel, J. & Mayer, G. By-Product Formation in Repetitive PCR Amplification of DNA Libraries during SELEX. *PLoS ONE* **9**, e114693 (2014).

[5] Miller, A. D. & Tanner, J. *Essentials of Chemical Biology*. Structure and Dynamics of Biological Macromolecules (John Wiley & Sons, 2013).

[6] Waldmann, H. & Janning, P. *Concepts and Case Studies in Chemical Biology* (John Wiley & Sons, 2014).

[7] Klußmann, S. *The Aptamer Handbook*. Functional Oligonucleotides and Their Applications (John Wiley & Sons, 2006).

[8] Ellington, A. D. & Szostak, J. W. Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures. *Nature* **355**, 850–852 (1992).

[9] Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H. & Toole, J. J. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **355**, 564–566 (1992).

[10] Raddatz, M.-S. L. *et al.* Enrichment of cell-targeting and population-specific aptamers by fluorescence-activated cell sorting. *Angew Chem Int Ed Engl* **47**, 5190–5193 (2008).

[11] Veedu, R. N. Medicinal chemistry of aptamers. *curr top med chem* **15**, 1065 (2015).

[12] Gopinath, S. C. B. Methods developed for SELEX. *Anal Bioanal Chem* **387**, 171–182 (2007).

[13] Merrifield, B. Solid phase synthesis. *Science* **232**, 341–347 (1986).

[14] Keefe, A. D., Pai, S. & Ellington, A. Aptamers as therapeutics. *Nature Reviews Drug Discovery* **9**, 537–550 (2010).

[15] Tolle, F. & Mayer, G. Dressed for success–applying chemistry to modulate aptamer functionality. *Chem. Sci.* **4**, 60–67 (2013).

[16] Mayer, G. & Jenne, A. Aptamers in Research and Drug Development. *BioDrugs* **18**, 351–359 (2004).

[17] Meyer, C., Hahn, U. & Rentmeister, A. Cell-specific aptamers as emerging therapeutics. *Journal of Nucleic Acids* **2011**, 904750–18 (2011).

[18] Ng, E. W. M. *et al.* Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. *Nature Reviews Drug Discovery* **5**, 123–132 (2006).

[19] Healy, J. M. *et al.* Pharmacokinetics and biodistribution of novel aptamer compositions. *Pharm. Res.* **21**, 2234–2246 (2004).

[20] Famulok, M., Blind, M. & Mayer, G. Intramers as promising new tools in functional proteomics. *Chemistry & biology* **8**, 931–939 (2001).

[21] Blind, M., Kolanus, W. & Famulok, M. Cytoplasmic RNA modulators of an inside-out signal-transduction cascade. *Proc Natl Acad Sci U S A* **96**, 3606–3610 (1999).

[22] Mayer, G. *et al.* Controlling small guanine-nucleotide-exchange factor function through cytoplasmic RNA intramers. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 4961–4965 (2001).

[23] Lennarz, S. *et al.* Selective Aptamer-Based Control of Intraneuronal Signaling. *Angew Chem Int Ed Engl* **54**, 5369–5373 (2015).

[24] Famulok, M. & Mayer, G. Aptamer Modules as Sensors and Detectors. *Acc. Chem. Res.* **44**, 1349–1358 (2011).

[25] Pinto, A. *et al.* Functional detection of proteins by caged aptamers. *ACS chemical biology* **7**, 360–366 (2012).

[26] Plaxco, K. W. & Soh, H. T. Switch-based biosensors: a new approach towards real-time, in vivo molecular detection. *Trends Biotechnol.* **29**, 1–5 (2011).

[27] Ferguson, B. S. *et al.* Real-time, aptamer-based tracking of circulating therapeutic agents in living animals. *Sci Transl Med* **5**, 213ra165–213ra165 (2013).

[28] Famulok, M. & Mayer, G. Aptamers and SELEX in Chemistry & Biology. *Chemistry & biology* **21**, 1055–1058 (2014).

[29] Stoltenburg, R., Reinemann, C. & Strehlitz, B. SELEX–a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol. Eng.* **24**, 381–403 (2007).

[30] Sun, H. & Zu, Y. A Highlight of Recent Advances in Aptamer Technology and Its Application. *Molecules* **20**, 11959–11980 (2015).

[31] Darmostuk, M., Rimpelová, S., Gbelcová, H. & Ruml, T. Current approaches in SELEX: An update to aptamer selection technology. *Biotechnol. Adv.* **33**, 1141–1161 (2015).

[32] Yüce, M., Ullah, N. & Budak, H. Trends in aptamer selection methods and applications. *Analyst* **140**, 5379–5399 (2015).

[33] Robertson, D. L. & Joyce, G. F. Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* **344**, 467–468 (1990).

[34] Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).

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

[36] Famulok, M. & Jenne, A. Oligonucleotide libraries - variatio delectat. *Current opinion in chemical biology* **2**, 320–327 (1998).

[37] McKeague, M. *et al.* Analysis of In Vitro Aptamer Selection Parameters. *J. Mol. Evol.* (2015).

[38] Blind, M. & Blank, M. Aptamer Selection Technology and Recent Advances. *Mol Ther Nucleic Acids* **4**, e223 (2015).

[39] Diafa, S. & Hollenstein, M. Generation of Aptamers with an Expanded Chemical Repertoire. *Molecules* **20**, 16643–16671 (2015).

[40] Turcheniuk, K., Tarasevych, A. V., Kukhar, V. P., Boukherroub, R. & Szunerits, S. Recent advances in surface chemistry strategies for the fabrication of functional iron oxide based magnetic nanoparticles. *Nanoscale* **5**, 10729–10752 (2013).

[41] Seo, Y.-J., Nilsen-Hamilton, M. & Levine, H. A. A computational study of alternate SELEX. *Bull Math Biol* **76**, 1455–1521 (2014).

[42] Hünniger, T., Wessels, H., Fischer, C., Paschke-Kratzin, A. & Fischer, M. just in time-Selection: A rapid semi-automated SELEX of DNA aptamers using magnetic separation and BEAMing. *Anal. Chem.* **86**, 10940–10947 (2014).

[43] Levine, H. A. & Nilsen-Hamilton, M. A mathematical analysis of SELEX. *Comput Biol Chem* **31**, 11–35 (2007).

[44] Lou, X. *et al.* Micromagnetic selection of aptamers in microfluidic channels. *Proc Natl Acad Sci U S A* **106**, 2989–2994 (2009).

[45] Oh, S. S. *et al.* Generation of highly specific aptamers via micromagnetic selection. *Anal. Chem.* **81**, 5490–5495 (2009).

[46] Oh, S. S. *et al.* Improving aptamer selection efficiency through volume dilution, magnetic concentration, and continuous washing in microfluidic channels. *Anal. Chem.* **83**, 6883–6889 (2011).

[47] Mendonsa, S. D. & Bowser, M. T. In vitro evolution of functional DNA using capillary electrophoresis. *J Am Chem Soc* **126**, 20–21 (2004).

[48] Berezovski, M. *et al.* Nonequilibrium capillary electrophoresis of equilibrium mixtures: a universal tool for development of aptamers. *J Am Chem Soc* **127**, 3165–3171 (2005).

[49] Yang, J. & Bowser, M. T. Capillary electrophoresis-SELEX selection of catalytic DNA aptamers for a small-molecule porphyrin target. *Anal. Chem.* **85**, 1525–1530 (2013).

[50] Kasahara, Y., Irisawa, Y., Ozaki, H., Obika, S. & Kuwahara, M. 2',4'-BNA/LNA aptamers: CE-SELEX using a DNA-based library of full-length 2'-O,4'-C-methylene-bridged/linked bicyclic ribonucleotides. *Bioorganic & medicinal chemistry letters* **23**, 1288–1292 (2013).

[51] Musheev, M. U. & Krylov, S. N. Selection of aptamers by systematic evolution of ligands by exponential enrichment: addressing the polymerase chain reaction issue. *Analytica Chimica Acta* **564**, 91–96 (2006).

[52] Marshall, K. A. & Ellington, A. D. Molecular parasites that evolve longer genomes. *J. Mol. Evol.* **49**, 656–663 (1999).

[53] Jing, M. & Bowser, M. T. Tracking the Emergence of High Affinity Aptamers for rhVEGF165 During CE-SELEX Using High Throughput Sequencing. *Anal. Chem.* **85**, 10761–10770 (2013).

[54] Yufa, R. *et al.* Emulsion PCR Significantly Improves NECEEM-based Aptamer Selection, Allowing for Efficient and Rapid Selection of Aptamer to Unmodified ABH2 Protein. *Anal. Chem.* **87**, 1411–1419 (2014).

[55] Ouellet, E., Foley, J. H., Conway, E. M. & Haynes, C. Hi-Fi SELEX: A high-fidelity digital-PCR based therapeutic aptamer discovery platform. *Biotechnol. Bioeng.* **112**, 1506–1522 (2015).

[56] Obeid, S., Bußkamp, H., Welte, W., Diederichs, K. & Marx, A. Interactions of non-polar and "Click-able" nucleotides in the confines of a DNA polymerase active site. *Chem. Commun. (Camb.)* **48**, 8320–8322 (2012).

[57] Svobodová, M., Pinto, A., Nadal, P. & O' Sullivan, C. K. Comparison of different methods for generation of single-stranded DNA for SELEX processes. *Anal Bioanal Chem* **404**, 835–842 (2012).

[58] Hamedani, N. S. *et al.* Capture and Release (CaR): a simplified procedure for one-tube isolation and concentration of single-stranded DNA during SELEX. *Chem. Commun. (Camb.)* **51**, 1135–1138 (2015).

[59] Liang, C. *et al.* Comparison of the methods for generating single-stranded DNA in SELEX. *Analyst* **140**, 3439–3444 (2015).

[60] Avci-Adali, M., Paul, A., Wilhelm, N., Ziemer, G. & Wendel, H. P. Upgrading SELEX technology by using lambda exonuclease digestion for single-stranded DNA generation. *Molecules* **15**, 1–11 (2010).

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

[62] Cho, M. *et al.* Quantitative selection of DNA aptamers through microfluidic selection and high-throughput sequencing. *Proc Natl Acad Sci U S A* **107**, 15373–15378 (2010).

[63] Hoinka, J. & Przytycka, T. AptaPLEX - A dedicated, multithreaded demultiplexer for HT-SELEX data. *Methods* (2016).

[64] Beier, R., Boschke, E. & Labudde, D. New Strategies for Evaluation and Analysis of SELEX Experiments. *Biomed Res Int* **2014**, 849743–12 (2014).

[65] Hoinka, J., Berezhnoy, A., Sauna, Z. E., Gilboa, E. & Przytycka, T. M. AptaCluster - A Method to Cluster HT-SELEX Aptamer Pools and Lessons from its Application. *Res Comput Mol Biol* **8394**, 115–128 (2014).

[66] Alam, K. K., Chang, J. L. & Burke, D. H. FASTAptamer: A Bioinformatic Toolkit for High-throughput Sequence Analysis of Combinatorial Selections. *Mol Ther Nucleic Acids* **4**, e230 (2015).

[67] Caroli, J., Taccioli, C., De La Fuente, A., Serafini, P. & Bicciato, S. APTANI: a computational tool to select aptamers through sequence-structure motif analysis of HT-SELEX data. *Bioinformatics* (2015).

[68] Deng, B. *et al.* Aptamer binding assays for proteins: The thrombin example-A review. *Analytica Chimica Acta* **837C**, 1–15 (2014).

[69] McKeague, M. *et al.* Comprehensive Analytical Comparison of Strategies Used for Small Molecule Aptamer Evaluation. *Anal. Chem.* 150815080118005 (2015).

[70] Helwa, R. & Hoheisel, J. D. Analysis of DNA-protein interactions: from nitrocellulose filter binding assays to microarray studies. *Anal Bioanal Chem* **398**, 2551–2561 (2010).

[71] Strauss, H. S., Boston, R. S., Record Jr., M. T. & Burgess, R. R. Variables affecting the selectivity and efficiency of retention of DNA fragments by E. coli RNA polymerase in the nitrocellulose-filter-binding assay. *Gene* **13**, 75–87 (1981).

[72] Oehler, S., Alex, R. & Barker, A. Is nitrocellulose filter binding really a universal assay for protein-DNA interactions? *Analytical biochemistry* **268**, 330–336 (1999).

[73] Tan, S. Y. *et al.* SELEX modifications and bioanalytical techniques for aptamer-target binding characterisation. *Crit Rev Anal Chem* 0–00 (2016).

[74] Entzian, C. & Schubert, T. Studying small molecule-aptamer interactions using MicroScale Thermophoresis (MST). *Methods* **97**, 27–34 (2016).

[75] Duhr, S. & Braun, D. Why molecules move along a temperature gradient. *Proc Natl Acad Sci U S A* **103**, 19678–19682 (2006).

[76] Jerabek-Willemsen, M. *et al.* MicroScale Thermophoresis: Interaction analysis and beyond. *J Mol Struct* **1077**, – (2011).

[77] Vaught, J. D. *et al.* Expanding the chemistry of DNA for in vitro selection. *J Am Chem Soc* **132**, 4141–4151 (2010).

[78] Gold, L. *et al.* Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery. *PLoS ONE* **5**, e15004 (2010).

[79] Eaton, B. The joys of in vitro selection: chemically dressing oligonucleotides to satiate protein targets. *Current opinion in chemical biology* **1**, 10–16 (1997).

[80] Meek, K. N., Rangel, A. E. & Heemstra, J. M. Enhancing aptamer function and stability via in vitro selection using modified nucleic acids. *Methods* (2016).

[81] Singh, S. K., Koshkin, A. A., Wengel, J. & Nielsen, P. LNA (locked nucleic acids): synthesis and high-affinity nucleic acid recognition. *Chemical Communications* 455–456 (1998).

[82] Koshkin, A. A. *et al.* LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition. *Tetrahedron* **54**, 3607–3630 (1998).

[83] Darfeuille, F., Hansen, J. B., Orum, H., Di Primo, C. & Toulme, J.-J. LNA/DNA chimeric oligomers mimic RNA aptamers targeted to the TAR RNA element of HIV-1. *Nucleic Acids Research* **32**, 3101–3107 (2004).

[84] Schmidt, K. S. *et al.* Application of locked nucleic acids to improve aptamer in vivo stability and targeting function. *Nucleic Acids Research* **32**, 5757–5765 (2004).

[85] Elle, I. C. *et al.* Selection of LNA-containing DNA aptamers against recombinant human CD73. *Mol Biosyst* **11**, 1260–1270 (2015).

[86] Klussmann, S., Nolte, A., Bald, R., Erdmann, V. A. & Fürste, J. P. Mirror-image RNA that binds D-adenosine. *Nat Biotechnol* **14**, 1112–1115 (1996).

[87] Nolte, A., Klussmann, S., Bald, R., Erdmann, V. A. & Fürste, J. P. Mirror-design of L-oligonucleotide ligands binding to L-arginine. *Nat Biotechnol* **14**, 1116–1119 (1996).

[88] Vater, A. & Klußmann, S. Turning mirror-image oligonucleotides into drugs: the evolution of Spiegelmer(®) therapeutics. *Drug Discov. Today* **20**, 147–155 (2015).

[89] King, D. J., Ventura, D. A., Brasier, A. R. & Gorenstein, D. G. Novel combinatorial selection of phosphorothioate oligonucleotide aptamers. *Biochemistry* **37**, 16489–16493 (1998).

[90] Yang, X. *et al.* Aptamers containing thymidine 3'-O-phosphorodithioates: synthesis and binding to nuclear factor-kappaB. *Bioorganic & medicinal chemistry letters* **9**, 3357–3362 (1999).

[91] Mann, A. P. *et al.* Identification of Thioaptamer Ligand against E-Selectin: Potential Application for Inflamed Vasculature Targeting. *PLoS ONE* **5**, e13050 (2010).

[92] Eckstein, F. Phosphorothioate oligodeoxynucleotides: what is their origin and what is unique about them? *Antisense and Nucleic Acid Drug Development* (2000).

[93] Pinheiro, V. B. *et al.* Synthetic genetic polymers capable of heredity and evolution. *Science* **336**, 341–344 (2012).

[94] Taylor, A. I. *et al.* Catalysts from synthetic genetic polymers. *Nature* **518**, 427–430 (2015).

[95] Alves Ferreira-Bravo, I., Cozens, C., Holliger, P. & Destefano, J. J. Selection of 2'-deoxy-2'-fluoroarabinonucleotide (FANA) aptamers that bind HIV-1 reverse transcriptase with picomolar affinity. *Nucleic Acids Research* gkv1057 (2015).

[96] Lapa, S. A., Chudinov, A. V. & Timofeev, E. N. The Toolbox for Modified Aptamers. *Molecular Biotechnology* 1–14 (2015).

[97] Kuwahara, M. & Sugimoto, N. Molecular evolution of functional nucleic acids with chemical modifications. *Molecules* **15**, 5423–5444 (2010).

[98] Zhang, L. Unnatural Nucleic Acids for Aptamer Selection. *Aptamers Selected by Cell-SELEX for Theranostics* 35–65 (2015).

[99] Latham, J. A., Johnson, R. & Toole, J. J. The application of a modified nucleotide in aptamer selection: novel thrombin aptamers containing 5-(1-pentynyl)-2'-deoxyuridine. *Nucleic Acids Res* **22**, 2817–2822 (1994).

[100] Battersby, T. R., Ang, D. N. & Burgstaller, P. Quantitative analysis of receptors for adenosine nucleotides obtained via in vitro selection from a library incorporating a cationic nucleotide analog. *Journal of the …* (1999).

[101] Vaish, N. K., Fraley, A. W., Szostak, J. W. & McLaughlin, L. W. Expanding the structural and functional diversity of RNA: analog uridine triphosphates as candidates for in vitro selection of nucleic acids. *Nucleic Acids Res* **28**, 3316–3322 (2000).

[102] Li, M. *et al.* Selecting Aptamers for a Glycoprotein through the Incorporation of the Boronic Acid Moiety. *Journal of the …* **130**, 12636–12638 (2008).

[103] Imaizumi, Y. *et al.* Efficacy of Base-Modification on Target Binding of Small Molecule DNA Aptamers. *J Am Chem Soc* **135**, 9412–9419 (2013).

[104] Temme, J. S., Drzyzga, M. G., MacPherson, I. S. & Krauss, I. J. Directed evolution of 2G12-targeted nonamannose glycoclusters by SELMA. *Chem. Eur. J.* **19**, 17291–17295 (2013).

[105] Davies, D. R. *et al.* Unique motifs and hydrophobic interactions shape the binding of modified DNA ligands to protein targets. *Proc Natl Acad Sci U S A* **109**, 19971–19976 (2012).

[106] Gelinas, A. D. *et al.* Crystal Structure of Interleukin-6 in Complex with a Modified Nucleic Acid Ligand. *J Biol Chem* **289**, 8720–8734 (2014).

[107] Jarvis, T. C. *et al.* Non-helical DNA Triplex Forms a Unique Aptamer Scaffold for High Affinity Recognition of Nerve Growth Factor. *Structure* **23**, 1293–1304 (2015).

[108] Gold, L. SELEX: How It Happened and Where It will Go. *J. Mol. Evol.* 1–4 (2015).

[109] Hermanson, G. T. *Bioconjugate Techniques* (Academic Press, 2013).

[110] Sletten, E. M. & Bertozzi, C. R. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angew Chem Int Ed Engl* **48**, 6974–6998 (2009).

[111] Lim, R. K. V. & Lin, Q. Bioorthogonal chemistry: recent progress and future directions. *Chemical Communications* **46**, 1589–1600 (2010).

[112] Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew Chem Int Ed Engl* **40**, 2004–2021 (2001).

[113] Huisgen, R. 1, 3-Dipolar Cycloadditions. Past and Future. *Angewandte Chemie International Edition in …* **2**, 565–598 (1963).

[114] Huisgen, R. Kinetics and Mechanism of 1,3-Dipolr Cycloadditions. *Angew Chem Int Ed Engl* **2**, 633–645 (1963).

[115] Tornøe, C. W., Christensen, C. & Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J Org Chem* **67**, 3057–3064 (2002).

[116] Rostovtsev, V., Green, L., Fokin, V. & Sharpless, K. A Stepwise Huisgen Cycloaddition Process: Copper (I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes. *Angewandte Chemie International Edition* **41**, 2596–2599 (2002).

[117] Fry, S. C. Oxidative scission of plant cell wall polysaccharides by ascorbate-induced hydroxyl radicals. *Biochemical Journal* **332**, 507–515 (1998).

[118] Prescher, J. A. & Bertozzi, C. R. Chemistry in living systems. *Nat. Chem. Biol.* **1**, 13–21 (2005).

[119] Chan, T. R., Hilgraf, R., Sharpless, K. B. & Fokin, V. V. Polytriazoles as Copper(I)-Stabilizing Ligands in Catalysis. *Org. Lett* **6**, 2853–2855 (2004).

[120] Hong, V., Presolski, S. I., Ma, C. & Finn, M. G. Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angew Chem Int Ed Engl* **48**, 9879–9883 (2009).

[121] El-Sagheer, A. H. & Brown, T. Click chemistry with DNA. *Chem Soc Rev* **39**, 1388 (2010).

[122] Burley, G. A. *et al.* Directed DNA metallization. *J Am Chem Soc* **128**, 1398–1399 (2006).

[123] Gierlich, J., Burley, G. A., Gramlich, P. M. E., Hammond, D. M. & Carell, T. Click chemistry as a reliable method for the high-density postsynthetic functionalization of alkyne-modified DNA. *Org. Lett* **8**, 3639–3642 (2006).

[124] Gierlich, J. *et al.* Synthesis of Highly Modified DNA by a Combination of PCR with Alkyne-Bearing Triphosphates and Click Chemistry. *Chemistry* **13**, 9486–9494 (2007).

[125] Rodionov, V. O., Presolski, S. I., Gardinier, S., Lim, Y.-H. & Finn, M. G. Benzimidazole and Related Ligands for Cu-Catalyzed Azide–Alkyne Cycloaddition. *J Am Chem Soc* **129**, 12696–12704 (2007).

[126] Besanceney-Webler, C. *et al.* Increasing the Efficacy of Bioorthogonal Click Reactions for Bioconjugation: A Comparative Study. *Angew Chem Int Ed Engl* **50**, 8051–8056 (2011).

[127] Uttamapinant, C. *et al.* Fast, Cell-Compatible Click Chemistry with Copper-Chelating Azides for Biomolecular Labeling. *Angewandte Chemie International Edition* **51**, 5852–5856 (2012).

[128] Rudolf, G. C. & Sieber, S. A. Copper-assisted click reactions for activity-based proteomics: fine-tuned ligands and refined conditions extend the scope of application. *ChemBioChem* **14**, 2447–2455 (2013).

[129] Mann, K. G., Butenas, S. & Brummel, K. The dynamics of thrombin formation. *Arterioscler. Thromb. Vasc. Biol.* **23**, 17–25 (2003).

[130] Griffin, J. H., Zlokovic, B. V. & Mosnier, L. O. Protein C anticoagulant and cytoprotective pathways. *Int. J. Hematol.* **95**, 333–345 (2012).

[131] Bernard, G. R. *et al.* Efficacy and Safety of Recombinant Human Activated Protein C for Severe Sepsis. *N Engl J Med* **344**, 699–709 (2001).

[132] Levi, M. Activated protein C in sepsis: a critical review. *Current Opinion in Hematology* **15**, 481–486 (2008).

[133] Müller, J. *et al.* An exosite-specific ssDNA aptamer inhibits the anticoagulant functions of activated protein C and enhances inhibition by protein C inhibitor. *Chemistry & biology* **16**, 442–451 (2009).

[134] Müller, J. *et al.* Monitoring of Plasma Levels of Activated Protein C Using A Clinically Applicable Oligonucleotide-Based Enzyme Capture Assay (OECA). *Journal of Thrombosis and Haemostasis* **10**, no–no (2012).

[135] Dulbecco, R. & Vogt, M. Plaque formation and isolation of pure lines with poliomyelitis viruses. *Journal of Experimental Medicine* **99**, 167–182 (1954).

[136] Balasubramanian, S. Sequencing nucleic acids: from chemistry to medicine. *Chemical Communications* **47**, 7281–7286 (2011).

[137] Blank, M. Next-Generation Analysis of Deep Sequencing Data: Bringing Light into the Black Box of SELEX Experiments. *Methods Mol. Biol.* **1380**, 85–95 (2016).

[138] Mian, I. S., Bradwell, A. R. & Olson, A. J. Structure, function and properties of antibody binding sites. *J. Mol. Biol.* (1991).

[139] Ramaraj, T., Angel, T., Dratz, E. A., Jesaitis, A. J. & Mumey, B. Antigen–antibody interface properties: Composition, residue interactions, and features of 53 non-redundant structures. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1824**, 520–532 (2012).

[140] Wirges, C. T. *et al.* Pronounced Effect of DNA Hybridization on Click Reaction Efficiency. *QSAR Comb. Sci.* **26**, 1159–1164 (2007).

[141] Shui, B. *et al.* RNA aptamers that functionally interact with green fluorescent protein and its derivatives. *Nucleic Acids Research* **40**, e39–e39 (2012).

[142] Stanlis, K. K. H. & McIntosh, J. R. Single-strand DNA aptamers as probes for protein localization in cells. *J Histochem Cytochem* **51**, 797–808 (2003).

[143] Yang, F., Moss, L. G. & Phillips, G. N. The molecular structure of green fluorescent protein. *Nat Biotechnol* **14**, 1246–1251 (1996).

[144] Crameri, A., Whitehorn, E. A., Tate, E. & Stemmer, W. P. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat Biotechnol* **14**, 315–319 (1996).

[145] Ingale, S. A., Mei, H., Leonard, P. & Seela, F. Ethynyl side chain hydration during synthesis and workup of "clickable" oligonucleotides: bypassing acetyl group formation by triisopropylsilyl protection. *J Org Chem* **78**, 11271–11282 (2013).

[146] Davis, J. H. & Szostak, J. W. Isolation of high-affinity GTP aptamers from partially structured RNA libraries. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 11616–11621 (2002).

[147] Ruff, K. M., Snyder, T. M. & Liu, D. R. Enhanced functional potential of nucleic acid aptamer libraries patterned to increase secondary structure. *J Am Chem Soc* **132**, 9453–9464 (2010).

[148] Luo, X. *et al.* Computational approaches toward the design of pools for the in vitro selection of complex aptamers. *RNA* **16**, 2252–2262 (2010).

[149] Jarvis, T. C. *et al.* Non-helical DNA Triplex Forms a Unique Aptamer Scaffold for High Affinity Recognition of Nerve Growth Factor. *Structure* **23**, 1293–1304 (2015).

[150] Vrabel, M. *et al.* Norbornenes in Inverse Electron-Demand Diels–Alder Reactions. *Chem. Eur. J.* **19**, 13309–13312 (2013).

[151] Domnick, C., Eggert, F. & Kath-Schorr, S. Site-specific enzymatic introduction of a norbornene modified unnatural base into RNA and application in post-transcriptional labeling. *Chem. Commun. (Camb.)* **51**, 8253–8256 (2015).

[152] Kath-Schorr, S. Cycloadditions for Studying Nucleic Acids. *Top Curr Chem (Z)* **374**, 1–27 (2016).

[153] Jewett, J. C. & Bertozzi, C. R. Cu-free click cycloaddition reactions in chemical biology. *Chem Soc Rev* **39**, 1272–1279 (2010).

[154] Ramil, C. P. & Lin, Q. Bioorthogonal chemistry: strategies and recent developments. *Chem. Commun. (Camb.)* **49**, 11007–11022 (2013).

[155] Kaul, C., Müller, M., Wagner, M., Schneider, S. & Carell, T. Reversible bond formation enables the replication and amplification of a crosslinking salen complex as an orthogonal base pair. *Nat Chem* **3**, 794–800 (2011).

[156] Kimoto, M., Yamashige, R., Matsunaga, K.-i., Yokoyama, S. & Hirao, I. Generation of high-affinity DNA aptamers using an expanded genetic alphabet. *Nat Biotechnol* **31**, 453–457 (2013).

[157] Sefah, K. *et al.* In vitro selection with artificial expanded genetic information systems. *Proc Natl Acad Sci U S A* **111**, 1449–1454 (2014).

[158] Okamoto, I., Miyatake, Y., Kimoto, M. & Hirao, I. High fidelity, efficiency and functionalization of Ds-Px unnatural base pairs in PCR amplification for a genetic alphabet expansion system. *ACS Synth Biol* acssynbio.5b00253 (2016).

[159] Someya, T., Ando, A., Kimoto, M. & Hirao, I. Site-specific labeling of RNA by combining genetic alphabet expansion transcription and copper-free click chemistry. *Nucleic Acids Research* **43**, 6665–6676 (2015).

[160] Mayer, G. & Heckel, A. Biologically active molecules with a "light switch". *Angew Chem Int Ed Engl* **45**, 4900–4921 (2006).

[161] Brieke, C., Rohrbach, F., Gottschalk, A., Mayer, G. & Heckel, A. Light-Controlled Tools. *Angewandte Chemie International Edition* **51**, 8446–8476 (2012).

[162] Bartnicki, F., Kowalska, E., Pels, K. & Strzalka, W. Imidazole-free purification of His3-tagged recombinant proteins using ssDNA aptamer-based affinity chromatography. *J Chromatogr A* **1418**, 130–139 (2015).

[163] Walker, J. W., Reid, G. P., McCray, J. A. & Trentham, D. R. Photolabile 1-(2-nitrophenyl)ethyl phosphate esters of adenine nucleotide analogs. Synthesis and mechanism of photolysis. *J Am Chem Soc* **110**, 7170–7177 (1988).

[164] Gramlich, P. M. E., Wirges, C. T., Manetto, A. & Carell, T. Postsynthetic DNA Modification through the Copper-Catalyzed Azide-Alkyne Cycloaddition Reaction. *Angew Chem Int Ed Engl* **47**, 8350–8358 (2008).

[165] Gary R Abel, J., Calabrese, Z. A., Ayco, J., Hein, J. E. & Ye, T. Measuring and Suppressing the Oxidative Damage to DNA During Cu(I)-Catalyzed Azide–Alkyne Cycloaddition. *Bioconjugate chemistry* acs.bioconjchem.5b00665 (2016).

[166] Hocek, M. Synthesis of base-modified 2'-deoxyribonucleoside triphosphates and their use in enzymatic synthesis of modified DNA for applications in bioanalysis and chemical biology. *J Org Chem* **79**, 9914–9921 (2014).

[167] Rohloff, J. C., Fowler, C., Ream, B. & Carter, J. D. Practical Synthesis of Cytidine-5-Carboxamide-Modified Nucleotide Reagents. … *and Nucleic Acids* **34**, 180–198 (2015).

[168] Wolk, S. K., Shoemaker, R. K., Mayfield, W. S., Mestdagh, A. L. & Janjic, N. Influence of 5-N-carboxamide modifications on the thermodynamic stability of oligonucleotides. *Nucleic Acids Research* (2015).

[169] Hipolito, C. J., Hollenstein, M., Lam, C. H. & Perrin, D. M. Protein-inspired modified DNAzymes: dramatic effects of shortening side-chain length of 8-imidazolyl modified deoxyadenosines in selecting RNaseA mimicking DNAzymes. *Organic & Biomolecular Chemistry* **9**, 2266–2273 (2011).

[170] Bruno, J. G. & Kiel, J. L. Use of magnetic beads in selection and detection of biotoxin aptamers by electrochemiluminescence and enzymatic methods. *BioTechniques* **32**, 178–80, 182–3 (2002).

[171] Gold, L. *et al.* Aptamers and the RNA world, past and present. *Cold Spring Harb Perspect Biol* **4**, a003582–a003582 (2012).

[172] Gupta, S. *et al.* Chemically-Modified DNA Aptamers Bind Interleukin-6 with High Affinity and Inhibit Signaling by Blocking its Interaction with Interleukin-6 Receptor. *J Biol Chem* **289**, 8706–8719 (2014).

[173] Bowser, M. T. SELEX: just another separation? *Analyst* **130**, 128–130 (2005).

[174] Drabovich, A. P., Berezovski, M., Okhonin, V. & Krylov, S. N. Selection of Smart Aptamers by Methods of Kinetic Capillary Electrophoresis. *Anal. Chem.* **78**, 3171–3178 (2006).

[175] Berezovski, M., Musheev, M., Drabovich, A. & Krylov, S. N. Non-SELEX Selection of Aptamers. *J Am Chem Soc* **128**, 1410–1411 (2006).

[176] Schütze, T. *et al.* Probing the SELEX Process with Next-Generation Sequencing. *PLoS ONE* **6**, e29604 (2011).

[177] Hoon, S., Zhou, B., Janda, K. D., Brenner, S. & Scolnick, J. Aptamer selection by high-throughput sequencing and informatic analysis. *BioTechniques* **51**, 413–416 (2011).

[178] Hoinka, J., Zotenko, E., Friedman, A., Sauna, Z. E. & Przytycka, T. M. Identification of sequence-structure RNA binding motifs for SELEX-derived aptamers. *Bioinformatics* **28**, i215–i223 (2012).

[179] Hoinka, J. *et al.* Large scale analysis of the mutational landscape in HT-SELEX improves aptamer discovery. *Nucleic Acids Research* **43**, 5699–5707 (2015).

[180] Shin, S. & Park, J. Characterization of sequence-specific errors in various next-generation sequencing systems. *Mol Biosyst* (2016).

[181] van Dijk, E. L., Jaszczyszyn, Y. & Thermes, C. Library preparation methods for next-generation sequencing: Tone down the bias. *Experimental Cell Research* **322**, 12–20 (2014).

[182] Erhard, F. & Zimmer, R. Count ratio model reveals bias affecting NGS fold changes. *Nucleic Acids Research* **43**, e136–e136 (2015).

[183] Gelinas, A. D., Davies, D. R. & Janjic, N. Embracing proteins: structural themes in aptamer–protein complexes. *Current Opinion in Structural Biology* **36**, 122–132 (2016).

[184] Kelly, J. A., Feigon, J. & Yeates, T. O. Reconciliation of the X-ray and NMR structures of the thrombin-binding aptamer d(GGTTGGTGTGGTTGG). *J. Mol. Biol.* **256**, 417–422 (1996).

[185] Opazo, F. *et al.* Modular Assembly of Cell-targeting Devices Based on an Uncommon G-quadruplex Aptamer. *Mol Ther Nucleic Acids* **4**, e251–e251 (2014).

[186] Ponce-Salvatierra, A., Wawrzyniak-Turek, K., Steuerwald, U., Höbartner, C. & Pena, V. Crystal structure of a DNA catalyst. *Nature* **529**, 231–234 (2016).

[187] Wolter, A. C. *et al.* NMR resonance assignments for the class II GTP binding RNA aptamer in complex with GTP. *Biomol NMR Assign* 1–5 (2015).

[188] Sochor, F. *et al.* (19)F-labeling of the adenine H2-site to study large RNAs by NMR spectroscopy. *J. Biomol. NMR* **64**, 63–74 (2016).

[189] Grytz, C. M., Marko, A., Cekan, P., Sigurdsson, S. T. & Prisner, T. F. Flexibility and conformation of the cocaine aptamer studied by PELDOR. - PubMed - NCBI. *Phys. Chem. Chem. Phys.* **18**, 2993–3002 (2016).

[190] Berezovski, M. V., Lechmann, M., Musheev, M. U., Mak, T. W. & Krylov, S. N. Aptamer-Facilitated Biomarker Discovery (AptaBiD). *J Am Chem Soc* **130**, 9137–9143 (2008).

[191] Chen, F., Tillberg, P. W. & Boyden, E. S. Optical imaging. Expansion microscopy. *Science* **347**, 543–548 (2015).

[192] Silverman, S. K. Pursuing DNA catalysts for protein modification. *Acc. Chem. Res.* **48**, 1369–1379 (2015).

[193] Zhou, C. *et al.* DNA-Catalyzed Amide Hydrolysis. *J Am Chem Soc* jacs.5b12647–6 (2016).

[194] Hollenstein, M. DNA Catalysis: The Chemical Repertoire of DNAzymes. *Molecules* **20**, 20777–20804 (2015).

[195] Kujau, M. J. & Wölfl, S. Efficient preparation of single-stranded DNA for in vitro selection. *Molecular Biotechnology* **7**, 333–335 (1997).

[196] Mayer, G., Lennarz, S., Rohrbach, F. & Tolle, F. Plug and play with RNA. *Angewandte Chemie. International edition in English* **50**, 12400–12401 (2011).