Click-SELEX enables the selection of Δ9-tetrahydrocannabinol-binding nucleic acids

Dissertation

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

Franziska Pfeiffer

aus

Solingen

Bonn 2018

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

1. Gutachter: Prof. Dr. Günter Mayer
2. Gutachter: Prof. Dr. Irmgard Förster
Tag der Promotion: 22.03.2019
Erscheinungsjahr: 2019

Parts of this thesis have been published in:

Pfeiffer, F. *et al.* Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Scientific reports* **8**, 10950, doi:10.1038/s41598-018-29325-6 (2018).

Pfeiffer, F. et al. Identification and characterization of nucleobase-modified aptamers by click-SELEX. *Nat. Protoc.* **13**, 1153-1180, doi:10.1038/nprot. 2018.023 (2018).

Pfeiffer, F., Rosenthal, M., Siegl, J., Ewers, J. & Mayer, G. Customised nucleic acid libraries for enhanced aptamer selection and performance. *Curr. Opin. Biotechnol.* **48**, 111-118, doi:10.1016/j.copbio.2017.03.026 (2017).

Pfeiffer, F. & Mayer, G. Selection and Biosensor Application of Aptamers for Small Molecules. *Frontiers in chemistry* **4**, 25, doi:10.3389/fchem.2016.00025 (2016).

Tolle, F., Rosenthal, M., Pfeiffer, F. & Mayer, G. Click Reaction on Solid Phase Enables High Fidelity Synthesis of Nucleobase-Modified DNA. *Bioconjug. Chem.* **27**, 500-503, doi:10.1021/acs.bioconjchem.5b00668 (2016).

Contents

List of Figures

List of Tables

List of Abbreviations

11-OH-THC	11-hydroxy-tetrahydrocannabinol
2’F-CTP	2’fluoro-cytidine triphosphate
2’F-UTP	2’fluoro-uridine triphosphate
2-AG	2-arachidonoylglycerol
ACN	acetonitrile
AEA	anandamide
APS	ammonium peroxodisulfate
ATP	adenosine triphosphate
BAC	blood alcohol concentration
BDE	1,4-butanedioldiglycidyl ether
Bn-dU	benzyl-deoxyuridine
bp	base pair
CB1	cannabinoid receptor 1
CB2	cannabinoid receptor 2
CBD	cannabidiol
CF_3	trifluoromethylbenzyl
CIAP	calf intestine alkaline phosphatase
CMC	critical micellar concentration
CNS	central nervous system
CuAAC	copper-catalysed alkyne-azide cycloaddition
dATP	deoxyadenosine triphosphate
DCC	N,N’-dicyclohexylcarbodiimide
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	dexoynucleoside triphosphate
dP	2-amino-8-(1′-β-D-2-deoxyribofuranosyl)-imidazo[1,2-a]-1,3,5-triazin-4(8H)one
DRUID	driving under the influence of drugs, alcohol, and medicines
Ds	7-(2-thienyl)imidazo[4,5- b]pyridine
DTT	1,4-dithiothreitol
dZ	6-amino-5-nitro-3-(1′-β-D-2′-deoxyribofuranosyl)-2(1H)-pyridone
EDTA	ethylenediaminetetraacetic acid
EdU	5-ethynyl-2’-deoxyuridine
EdUTP	5-ethynyl-2’deoxyuridine-triphosphate
EIA	enzyme immunoassay
EtOH	ethanol
EU	European Union
FDA	Food and Drug Administration
Gal3	galectin 3
GC-MS	gas chromatography – mass spectrometry
GC-MS-MS	gas chromatography – tandem mass spectrometry
GPCR	G-protein coupled receptor
GTP	guanosine triphosphate

HCl	hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	high performance liquid chromatography
IL-1	interleukin-1
IL-6	interleukin-6
iPP	inorganic pyrophosphatase
KCl	potassium chloride
K_d	dissociation constant
KH_2PO_4	potassium phosphate monobasic
LB	Lennox broth
LC-MS	liquid chromatography – mass spectrometry
MD	molecular dynamics
MeOH	methanol
$MgCl_2$	magnesium chloride
MS	multiple sclerosis
Na_2HPO_4	sodium phosphate dibasic
NaCl	sodium chloride
NaOAc	sodium acetate
Nap-dU	naphthyl-deoxyuridine
NGF	nerve growth factor
NGS	next-generation sequencing
NH_3	ammonia
NMR	nuclear magnetic resonance
NTP	nucleoside triphosphate
OR	odds ratio
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCB	polychlorinated biphenyl
PCR	polymerase chain reaction
PDB	protein data bank
PDGF(-BB)	platelet-derived growth factor (B)
PEG	polyethylene glycol
pH	negative logarithm of the H^+-concentration
pK_A	acid dissociation constant
PNK	polynucleotide kinase
PNS	peripheral nervous system
Px	2-nitro-4-propynylpyrrole
qPCR	quantitative PCR
R^2	coefficient of determination
RNA	ribonucleic acid
ROSITA	roadside testing assessment
RT	room temperature
RT-PCR	reverse transcription polymerase chain reaction
sal. DNA	salmon sperm DNA
SBS	sequencing by synthesis
SD	standard deviation
SDS	sodium dodecylsulfate
SELEX	systematic evolution of ligands by exponential enrichment

SL	starting library
SOC	super optimal broth
SOMAmer	small off-rate modified aptamer
TEA	triethylamine
TEMED	N,N,N',N'-tetramethylethylenediamine
THC	Δ9-tetrahydrocannabinol
THCA	Δ9-tetrahydrocannabinolic acid
THC-COOH	11-nor-9-carboxy-tetrahydrocannabinol
THC-L	THC-linker
THPTA	tris(3-hydroxypropyltriazolylmethyl)amine
TIPS	triisopropylsilyl ether
TLF	THC-L-FITC
UAE	United Arab Emirates
US	United States

Abstract

Δ9-Tetrahydrocannabinol (THC) is the main psychoactive compound in the plant *Cannabis sativa* and the most widely used illegal drug in the world. Its consumption impairs driving skills such as keeping a vehicle on track and has been shown to increase the risk of accidents. Urine tests are available for the detection of THC in a roadside setting, but necessitate facilities and are prone to sample manipulation. In addition, the correlation between impairment and positive tests is poor as a THC-metabolite instead of the active THC is analysed. Oral fluid sampling has emerged as a simple and quick alternative that permits detection of THC itself and therefore evidence for driving under the acute influence of THC. Existing oral fluid tests that are based on antibodies fail to reach the sensitivity and specificity needed for roadside testing.

We aimed to select a THC-binding aptamer for the development of new oral fluid tests for THC. Aptamers are single-stranded nucleic acids that specifically recognise their target molecule with high affinity. Numerous trials with both DNA and 2'F-modified RNA did not succeed in enriching such sequences for THC. Most likely, this is due to the high hydrophobicity of THC and its thereby constrained ability to interact with nucleic acids. Nucleobase-modified aptamers have been shown to dramatically increase the selection success for protein targets. We thus used a recently developed technique, called click-SELEX, that facilitates modification of DNA with an azide of choice. Emulating the interaction of THC with the human cannabinoid receptor 1, benzyl- and trifluoromethylbenzyl-residues were utilised as nucleobase-modifications. This enabled the immediate selection of sequences that bind THC immobilised on epoxy-sepharose. While none of the sequences could be shown to recognise THC in solution, this thesis established the synthesis of fluorescently-labelled THC-derivatives that can be utilised in fluorescence polarisation assays to ascertain affinity constants of future clickmers. In addition, THC-solubility in aqueous buffers with low amounts of organic solvent was thoroughly investigated. Altogether, the findings of this thesis lay the foundation for the selection of THC-binding clickmers that can be implemented in roadside tests in the future.

Zusammenfassung

Δ9-Tetrahydrocannabinol (THC) ist für den Großteil der psychoaktiven Wirkung der Pflanze *Cannabis sativa* verantwortlich und die meistgenutzte Droge der Welt. Der Konsum beeinträchtigt die Fahrfähigkeit: Fahrzeuge können nicht in der Spur gehalten werden und die Wahrscheinlichkeit von Unfällen steigt. Urintests werden von der Polizei verwendet, können aber verhältnismäßig leicht manipuliert werden und setzen das Vorhandensein von entsprechenden Anlagen voraus. Zusätzlich ist die Korrelation zwischen Einschränkung der Fahrfähigkeit und positiven Ergebnissen schlecht, da ein THC-Metabolit an Stelle des aktiven THCs analysiert wird. Speicheltests sind als eine schnelle und einfache Alternative bekannt geworden, die zudem die Detektion von THC selbst und entsprechend einen viel besseren Nachweis des Fahrens unter akutem Drogeneinfluss gestatten. Die bisher erhältlichen Tests sind Antikörper-basiert, erreichen jedoch nicht die benötigte Sensitivität und Spezifität.

Wir wollten ein THC-bindendes Aptamer für die Entwicklung neuer Speicheltests für THC selektieren. Aptamere sind einzelsträngige Nukleinsäuren, die spezifisch und hochaffin ihr Zielmolekül erkennen. Zahlreiche Versuche mit sowohl DNA als auch 2'F-modifizierter RNA konnten keine THC-Aptamere anreichern. Vermutlich ist dies auf die hohe Hydrophobizität THCs und seine dadurch eingeschränkten Möglichkeiten zur Interaktion mit Nukleinsäuren zurückzuführen. Nukleobasen-modifizierte Aptamere konnten die Erfolgsrate von Selektionen für Protein-bindende Aptamere drastisch erhöhen. Entsprechend verwendeten wir eine vor kurzem entwickelte Technik, die sich Click-SELEX nennt, und die Modifikation der DNA mit einem frei wählbaren Azid ermöglicht. Um die Interaktion von THC mit dem humanen Cannabinoid Rezeptor 1 nachzuahmen, wurden Benzyl- und Trifluoromethylbenzyl-Reste als Nukleobasen-Modifikationen verwendet. Dies ermöglichte die sofortige Selektion von Sequenzen, die auf Epoxy-Sepharose immobilisiertes THC erkennen. Obgleich keine der Sequenzen THC in Lösung bindet, etablierte diese Arbeit die Synthese fluoreszenzmarkierter THC-Derivate, welche für die Bestimmung der Affinität zukünftiger Clickmere mit Hilfe von Fluoreszenz Polarisation genutzt werden können. Zusätzlich wurde die Löslichkeit von THC in wässrigen Puffersystemen mit geringen Mengen organischer Lösungsmittel sorgfältig untersucht. Alles in allem legen die in dieser Dissertation dargelegten Ergebnisse das Fundament für die Selektion THC-bindender Clickmere, die in zukünftige Speicheltests implementiert werden können.

1 Introduction

1.1 Δ9-Tetrahydrocannabinol

Δ9-Tetrahydrocannabinol (THC, **Figure 1A**) is the main psychoactive constituent of the plant *Cannabis sativa*. Due to its psychoactive effect, it has been used for thousands of years for medical and recreational reasons. Nowadays, it is an approved medicine for a number of indications in several countries (section 1.1.3). In addition, THC is still the most widely used illicit drug in the world. A survey conducted in Germany in 2015 revealed that 6.1% of all adults specified that they used Cannabis in the last year. For teenagers (12-17 years) and young adults (18-25 years), the incidence was even higher with 7.3% and 15.3%, respectively.[1] Consumed are either the leaves (marijuana), the resin (hashish), whose THC-concentration is higher, or even more concentrated extracts (hashish oil).[2,3]

Recent years have seen a decrease in the criminalisation of THC. Uruguay and some US states have legalised the drug, while penalties for possession of small quantities and use have been abolished in e.g., Germany (§ 31a BtMG).[4] In contrast, long prison sentences for minimal or death sentences for larger amounts are common in e.g., the Philippines, Singapore, and the United Arab Emirates (UAE).[5-7]

1.1.1 Important cannabinoids of *Cannabis sativa*

THC is only one of a steadily increasing number of known components of *Cannabis sativa*. As of now, 554 compounds have been identified, 113 of which are cannabinoids.[8,9]

While the THC-content is most important for the majority of medicinal and all recreational uses of *Cannabis sativa*, quantification of it usually also encompasses the content of Δ9-tetrahydrocannabinolic acid (THCA, **Figure 1B**). THCA is the biogenetic precursor of THC and can be converted to THC by decarboxylation, which is rapidly achieved by heat exposure (during smoking or baking) or slowly during storage.[10] In its unprocessed form, up to 85% of the plant's THC-content are present as THCA with large variations depending on the origin and type (resin/leaves) of sample.[11]

Apart from THC and THCA, cannabidiol (CBD, **Figure 1C**) is one of the most abundant cannabinoids in *Cannabis sativa*. It amounts to up to 40% of the plant's cannabinoids and is not psychoactive, but known to antagonize the effects of THC. Studies regarding the use of CBD in a number of psychiatric disorders like anxiety, depression, and psychosis have been published.[12] In addition, it is

combined with THC in the oromucosal spray Sativex®, which is approved in a number of countries, including the European Union (EU), but not in the United States (US) (section 1.1.3).[13,14]

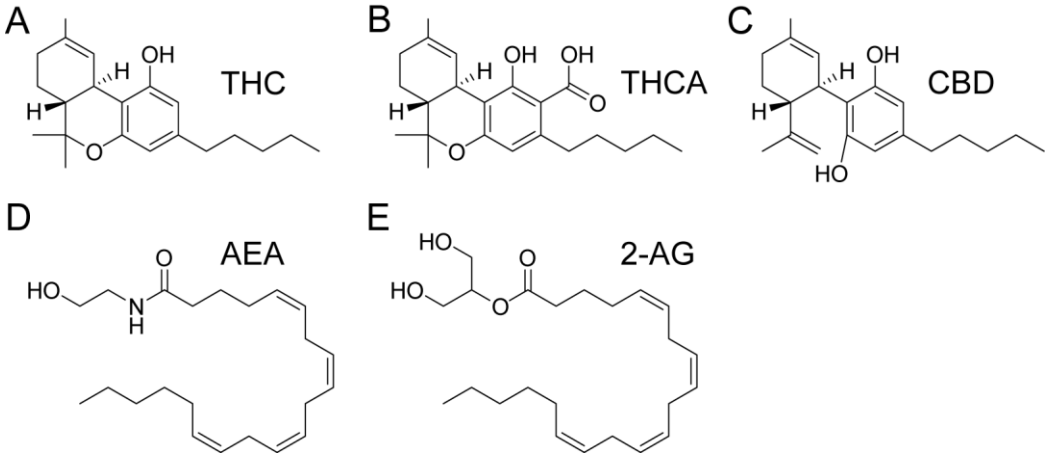


Figure 1: Structures of cannabinoids and endocannabinoids
Chemical structures of the cannabinoids (A) Δ9-tetrahydrocannabinol (THC), (B) Δ9-tetrahydrocannabinolic acid (THCA), and (C) cannabidiol (CBD) and of the endocannabinoids (D) anandamide (AEA) and (E) 2-arachidonoyl glycerol (2-AG).

1.1.2 Human cannabinoid receptors and THC-binding

THC is recognised by two different human cannabinoid receptors: Cannabinoid receptor 1 (CB1)[15] and 2 (CB2)[16]. Both are G-protein coupled receptors (GPCRs). CB1 is mainly expressed in the central and peripheral nervous system (CNS, PNS) and is the primary target of both THC and the endocannabinoids anandamide (AEA, **Figure 1D**) and 2-arachidonoyl glycerol (2-AG, **Figure 1E**).[17] The receptor controls motor function, cognition, memory, and analgesia through inhibition of neurotransmitter release.[18,19] CB2 can be found in peripheral organs with immune functions.[19] It plays a role in immunosuppression, neuroprotection, and neuropathic and inflammatory pain.[20-23]

To increase the understanding of the interaction between CB1 and its ligands, pharmacological evaluations, mutation studies, and molecular dynamics (MD) simulations were performed.[24-26] In 2016 and 2017, two independent groups published crystal structures of CB1 with different ligands (**Table 1**, **Figure 2**).[17,27,28] Nonetheless, none of them were able to co-crystallise CB1 with THC.

Table 1: Crystal structures of CB1

Publication	Year	PDB-ID	Co-crystallised ligand	Resolution [Å]
Hua *et al.*[27]	2016	5TGZ	AM6538 (stabilising antagonist)	2.8
Shao *et al.*[28]	2016	5U09	taranabant (inverse agonist)	2.6
Hua *et al.*[17]	2017	5XR8	AM841 (agonist)	2.95
		5XRA	AM11542 (agonist)	2.8

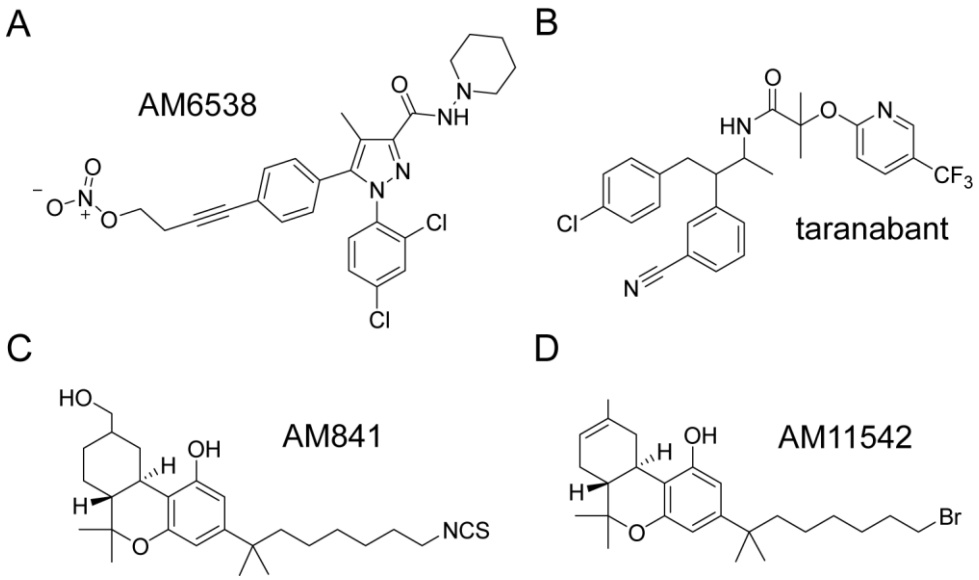


Figure 2: Structures of CB1-ligands used for co-crystallisation
Chemical structures of the CB1-ligands that were used for the co-crystallisation of CB1 with (A) AM6538[27], (B) taranaband[28], (C) AM841, and (D) AM11542.[17]

Therefore, the actual interaction between CB1 and THC remains unknown. All predictions based on the crystal structures and MD simulations agree that phenylalanine-residues of CB1 are undergoing π-stacking with the aromatic ring of THC. They disagree on which specific residues and how many of them are responsible (**Table 2**).[17,25,27,28] As THC is a partial agonist of CB1 and major changes of the ligand binding pocket were observed between the antagonist- and agonist-bound crystal structures, the docking studies based on the agonist-bound crystal structures[17] can probably be presumed closest to reality as long as no CB1-THC co-crystal is available (**Figure 3**).

Table 2: CB1-residues presumed to interact with THC

Publication	Method	CB1-residues interacting with THC
Shim *et al.*[25]	MD simulation	F174, F177
Hua *et al.*[27]	crystal structure	F268
Shao *et al.*[28]	crystal structure	F174 L193, C355, W356, S383
Hua *et al.*[17]	crystal structure	F177, F189, F200, F268, F379 L193, U196, Y275, L359, M363, S383

Identical residues are coloured alike.

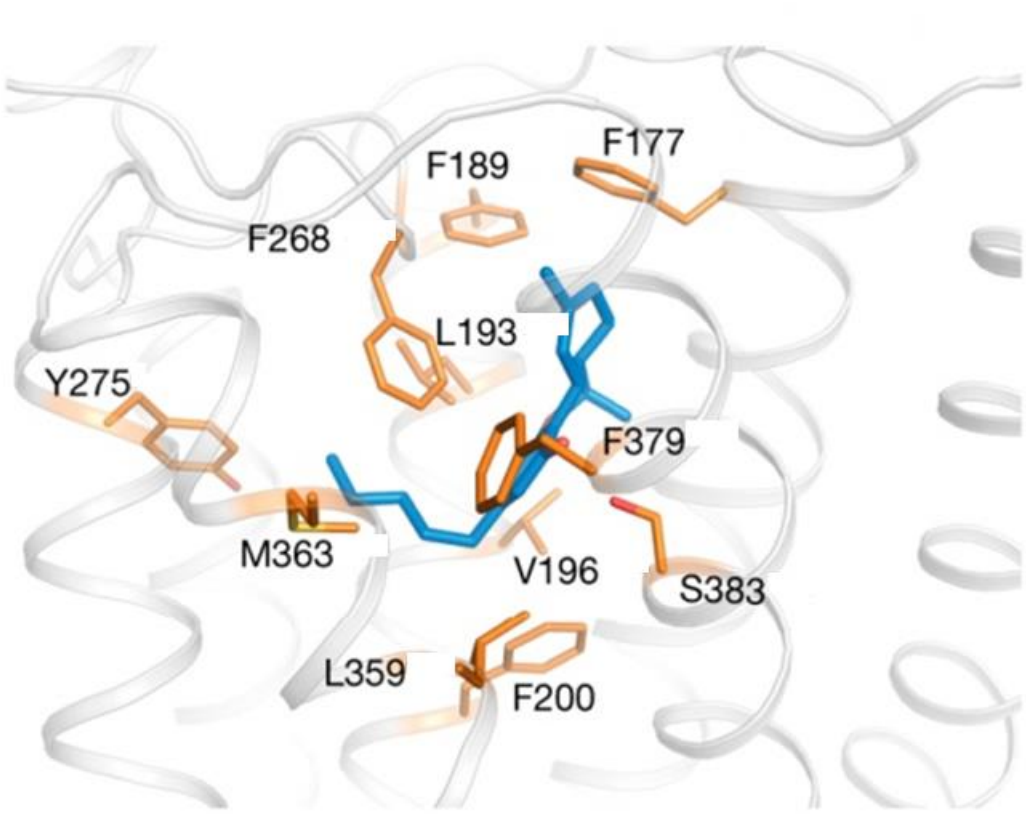


Figure 3: Docking study of the interaction of CB1 with THC modified after Hua *et al*.
The cannabinoid receptor is depicted as a cartoon in grey, with helices representing the transmembrane regions. Residues interacting with THC are displayed in orange. THC is depicted in blue. Most parts of CB1 that do not interact with THC were omitted. Modified after Hua *et al.* 2017.[17]

1.1.3 THC as a medicinal drug

Long before THC or any other cannabinoid had been identified as psychoactive constituent of cannabis, the whole plant was used for medical purposes.

Reports of cannabis use in China date back to 2700 BC.[29] In India, old sacred texts count cannabis as one of the five sacred plants due to its psychoactive effect.[29] In addition, it was used for a multitude of medical reasons that include analgesic, anaesthetic, anti-inflammatory, antibiotic, and appetite stimulant effects since about 1000 BC.[29-31]

In the timeframe until the 18th century, cannabis use spread from India to the Middle East and Africa. While the plant is mentioned in a number of plant books on the European continent during the Middle Ages, including Hildegard von Bingen's *Physica* (1150), far fewer applications are known in comparison with other parts of the world.[30,32]

Cannabis entered western medicine in the 19th century, when an Irish physician brought the knowledge from his service in India and successfully experimented

with the plant's use for rheumatism, convulsions, and muscular spasms that occurred due to tetanus and rabies in 1839.[33] At the start of the 20th century, it was used mainly as a sedative, hypnotic, and analgesic and to improve appetite and digestion.[34]

The first decades of the 20th century saw a decrease in cannabis use. As no knowledge of the constituents responsible for single effects were identified, results were hard to reproduce with the plant material.[34] This problem was eventually solved and standard extracts and tinctures became available in the 1930's.[35] In 1937, the (later on disabused) claim of the American police that cannabis was the cause of crime, violence, insanity, and death led to an extreme tax increase on cannabis products, effectively stopping its use in pharmaceutical formulae.[34]

The identification of the chemical structure of THC in 1964 and of the endocannabinoid system in the early 1990's enabled the investigation of the effects of single cannabinoids.[36,37] Nowadays, three different cannabis preparations are available: Dronabinol (THC), Nabilone (THC-analogue[38,39]), and Nabiximols (THC:CBD 1:1). The approved indications vary between countries. In Germany, Dronabinol (Marinol®) is approved for any pain indication, Nabilone (Canemes®) for chemotherapy-induced emesis, and Nabiximols (Sativex®) for spasticity in multiple sclerosis (MS). In addition, 14 different kinds of cannabis flowers are available for the pain management of patients with severe diseases. In all cases, it is a drug of last resort when other, more traditional medications have failed.[40]

Ongoing studies and an as of now low number of clinical trials further implicate THC as possible treatments for different types of cancer[41], agitation in Alzheimer's disease[42], and CBD for epilepsy.[43] A lot more knowledge has to be gained to confirm the consistency of the observed effects and treatment safety to gain an approval for the respective indication.

1.1.4 Psychoactive effects of THC

The use of THC as a drug is based on the fact that some plant parts contain a multitude of cannabinoids in varying concentrations.[10,11,44] Moreover, the plant has been bred to contain higher levels of THC in comparison to the amounts found in the 1990's.[3] In addition to this, most reports about psychoactive effects origin from people whose drug use history, emotional state, as well as the setting of the use have an impact on the experienced effects and might therefore be less accurate and more subjective than desired.[45]

Altogether, this renders the objective description of cannabis effects difficult. To give future investigators a list of possible effects, Charles T. Tart distributed a

thorough questionnaire to experienced cannabis users. Common effects were reported in nearly every imaginable category (visual, auditory, touch, taste, and smell effects, space-time perception, perception of the body, physical movement, interpersonal relations, sexual effects, thought processes, memory functioning, emotions, self-control, sense of identity, and effects on sleep).[45]

The most abundantly reported influence is an increase in intensity and scope of all senses.[45] Space-time perception[46,47], understanding, memory[48], emotion, and sense of identity are changed.[45,49] Motor coordination and judgment are impaired.[48,50]

1.1.4.1 Pharmacokinetics and THC-metabolites

The administration of THC influences its absorption. Bioavailability after smoking varies between 12-37%, depending on the subject and smoking dynamics (number, duration, spacing, and volume of inhalation).[51,52] Oral bioavailability is reported to be slightly lower with 3-20%.[51,53] Plasma concentrations of THC peak quickly during or shortly after smoking. This is followed by a fast decrease until 30 min after smoking onset and a steady decline over the next hours.[54,55] In contrast, oral THC is absorbed slowly and, after an hour, concentrations decline even slower (**Figure 4**).[51] Subjective THC effects (feeling “high”) have been reported both parallel to the concentration peak after smoking as well as delayed by about 15 min alongside already decreasing plasma levels. Oral administration leads to an even later effect onset (peaking 2 h after ingestion), but longer effect duration, which lasts through lower plasma THC concentrations compared to those eliciting effects after smoking.[51,54]

In plasma, about 97% THC is lipoprotein-conjugated.[56,57] From there, it is taken up by highly perfused tissues like lung, heart, brain, and liver.[58] The amount of THC detectable in the brain, where it can bind to CB1 and result in psychoactive effects, is reported to be less than 1%. This is most likely due to its retention by the blood-brain barrier. Instead, it accumulates in fat tissue.[59-61] A post-mortem study of twelve cannabis users showed consistently higher concentrations of THC in the brain than in the blood. In three cases, THC could not even be detected in blood any longer, while it was still present in the brain samples. It therefore seems that even though bioavailability in the brain is low, THC that arrives there is retained for a longer time. These stores might be the reason for long-lasting THC-effects that have been observed, but more research is needed for a thorough analysis.[62]

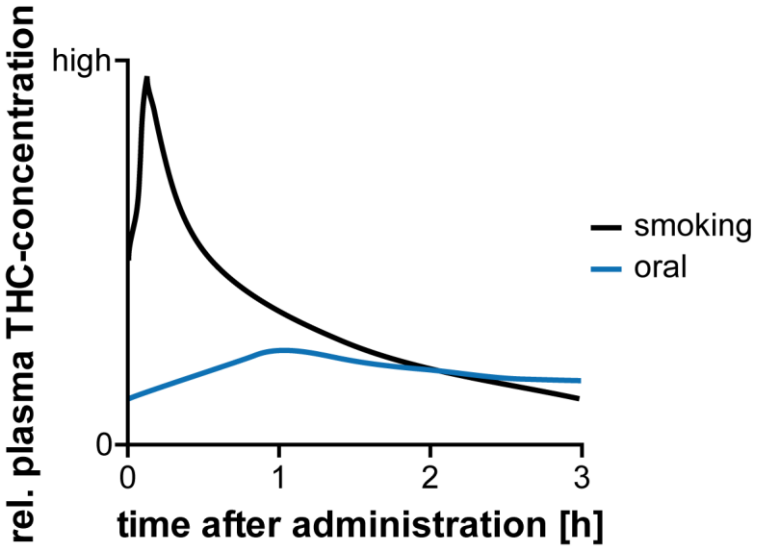


Figure 4: Schematic illustration of the plasma THC-concentration after different routes of administration
THC-concentrations peak during or shortly after smoking (black curve). They rapidly decline until 30 min after smoking and then slowly decline over the subsequent hours. After oral ingestion, plasma THC-concentrations increase to a low peak after about 1 h and then decline even more slowly (blue curve).[51]

THC is metabolised in the liver to the two main metabolites 11-hydroxy-tetrahydrocannabinol (11-OH-THC) (**Figure 5A**) and 11-nor-9-carboxy-tetrahydrocannabinol (THC-COOH) (**Figure 5B**). Both are formed by oxidation, with inactive THC-COOH resulting from oxidation of the psychoactive 11-OH-THC.[63-66] THC-COOH and its phase-II-metabolite, THC-COOH glucuronide are the major end products of the human metabolism.[67,68] After the initial distribution, the rate with which THC is metabolised is determined by the rate with which the drug re-enters the blood from its storage tissue.[69] In addition, THC can be metabolised to hydroxylated forms in brain microsomes, but little is known about these metabolites and their psychoactivity.[70]

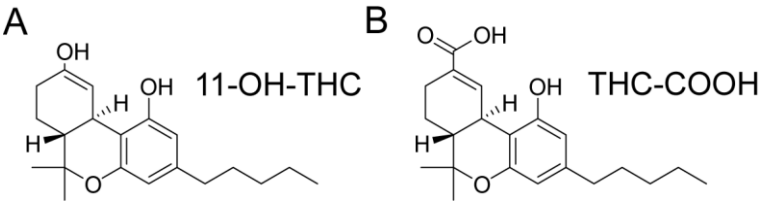


Figure 5: Chemical structures of THC-metabolites
Structures of the main THC-metabolites (A) 11-hydroxy-tetrahydrocannabinol (11-OH-THC) and (B) 11-nor-9-carboxy-tetrahydrocannabinol (THC-COOH).

THC is excreted within days to weeks, depending on the route of administration and abuse history of the subject. THC-COO-glucuronide is the major urinary metabolite, while 11-OH-THC is the main metabolite found in the faeces.[53,67,71] Minor amounts of THC-COOH are also excreted in urine (section 1.1.5.2.1).[58]

1.1.5 THC and driving

The high prevalence of THC-use indicates that it might be commonly consumed by drivers. A large scale study about driving under the influence of drugs, alcohol, and medicines (DRUID) was performed between 2006 and 2011 in 18 European states. About 50,000 drivers were randomly selected and tested for drug and alcohol levels.[72] According to the results, the average prevalence of driving under the influence of THC is 1.32%, which amounts to one third of the prevalence of alcohol.[73]

Considering the psychoactive effects of THC, which include changes in motor coordination, judgement, and sensual as well as space-time-perception (section 1.1.4), it stands to reason that THC affects driving capability. Two different types of studies have tried to quantify the impairment: Experimental ones that test the subject's responses in tests simulating driving situations and epidemiological studies. For the latter, samples from real drivers are taken and analysed to assess the risk of traffic accidents based on the alcohol or drug concentration in the samples.

The most consistent finding in experimental studies is that the ability to keep a vehicle on track is reduced under the influence of THC.[74-78] This is also reported after the use of Dronabinol® (medical THC) by heavy and occasional cannabis users. The impairment after THC-consumption is less pronounced in heavy users, but stronger than the one resulting from a blood alcohol concentration (BAC) of 0.5 g/l, the legal limit for driving in Germany (§ 24a StVG).[79] Two studies report no significant impairment to keep a vehicle on track, even though other signs of impairment were detected in one of them.[80,81] These inconsistencies might result from differences in THC-amounts, subject drug use history, and the delay between drug use and start of the tests.[82] In addition, a number of studies report that subjects that were aware of their impairment compensated by driving slowly, overtaking less, and keeping longer distances.[75,81,83,84] Nevertheless, not every impaired psychomotor skill could be compensated[76,85] and not all subjects tried to: Studies testing judgement were able to show an increase in impulsive behaviour[78,86] and choice of the risky option if subjects were asked to choose between a safe and risky option.[87]

Most epidemiological studies report a 2- to 3-fold increase in accident risk under the influence of cannabis[72,88] and the same for the culpability of the drugged driver.[89,90] Higher blood levels (> 5 µg/ml) have also been associated with a 5- to 7-fold increase in risk.[89,90] This is relatively low compared to the risk reported for alcohol, which varies between a 4- and 15-fold increase, depending on the study.[89,91,92]

Depending on the study, between 30 and 45% of all drivers under the influence of THC had also consumed alcohol.[93,94] For this combination, most studies report a synergistic effect.[77,80,85,95] The drivers are less attentive, even though neither drug

alone has such an impact[96], and the ability to keep a vehicle on track is further impaired.[97] Compensation as mentioned above is no longer possible.[84] The risk factor for the responsibility in car crashes rises accordingly (odds ratio (OR) 11-16).[72,98,99]

1.1.5.1 Legislation for driving under the influence of THC

To prevent people from driving under the influence of THC, limits for the drug concentration have been set. In Europe, three different policies are prevalent: *per se* limits, impairment (obvious from personal behaviour or driving style), and a combination of those two, the two-tier approach, which is also used in Germany.[73,82]

Legal limits for THC range from 0.5 (Italy, *per se*) to 3 ng/ml (Portugal, *per se*) in whole blood and 0.3 (Slovenia, *per se*) to 2 ng/ml (Luxembourg, impairment) in plasma. The limit in Germany is set to 1 ng/ml in plasma.

Only a few countries have also set limits for THC-COOH. Those range from 0.5 (Italy, *per se*) to 50 ng/ml (Poland, *per se*) in whole blood. Slovenia is the only country in Europe to have set plasma limits (5 ng/ml, *per se*).[82]

1.1.5.2 Roadside testing for THC

While the plasma concentration is necessary for any legal sentence[82], sampling of blood is restricted to healthcare practitioners and, in theory, requires a judicial order in Germany. As alcohol and drugs are metabolised and the concentration at a later time point might not be sufficient for sentencing, the judicial order is not mandatory for traffic offenses (§ 81a StPO). Nevertheless, it requires both a healthcare practitioner and a probable cause (§ 81a StPO), which might not always be the case for routine roadside tests. Therefore, alternative testing matrices have been established. The two most common methods are tests using urine or oral fluid. After a positive response in one of those initial tests, blood sampling and quantification of THC using more sophisticated analytical methods like gas chromatography – mass spectrometry (GC-MS) is mandatory.[100,101]

1.1.5.2.1 Urine testing for THC-COOH

Urine tests are based on the detection of THC-COOH.[101] Although the majority of THC-COOH is present in urine as glucuronide, most tests do not differentiate between the two and thus also detect THC-COOH glucuronide. Before laboratory confirmation e.g., by GC-MS, the glucuronide has to be hydrolysed.[58]

Although sensitivity and specificity of available urine based testing devices for THC-COOH are excellent[102], the matrix encompasses some inherent problems for roadside testing. First of all, sample taking is invasive and has to be observed closely to hinder attempts at sample manipulation.[103,104] As the process is quite intimate, a police officer of the same sex is needed and might not always be available in a roadside setting. The same goes for facilities, which have to be sought out, costing time of both subject and officers.[105]

Most importantly, the detected substance is not the pharmacologically active one, but a metabolite. The first detection of THC-COOH in urine occurs 4 (gas chromatography – tandem mass spectrometry, GC-MS-MS) or 6 h (enzyme immunoassay, EIA) after smoking.[101] This is later than the strongest impairment: A subjective intoxication as well as significant negative effects on driving performance are present up to 6-10 h after use, with the majority of the subjective effect in the first two hours.[74,78,106,107] Twenty-four hours after smoking, no effects could be measured.[83] The detection of THC-COOH in urine then, depending on the method, lasts until an average of 33 or 26 h with maximum detection times of 58 or 42 h (**Figure 6**).[101] Chronic daily cannabis users under monitored abstinence were positive for THC-COOH for 13 days (median).[108]

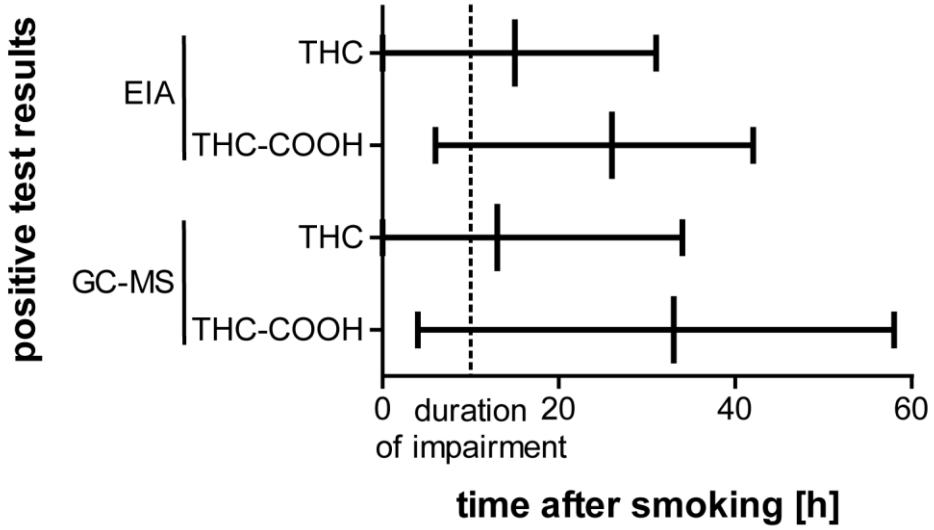


Figure 6: Timeline of positive test results in urine and oral fluid samples
Timeline of positive test results in urine (THC-COOH) and oral fluid (THC) samples. Given is the time point of the first positive test result after smoking of cannabis, as well as the average duration of positive results (line) and the time of the last positive test result for results obtained by GC-MS-MS and EIA. Indicated by a dotted line is the approximate duration for which driving skills are impaired (10 h).[101]

Taken together, those studies show that the timewise correlation between positive urine testing and impaired driving skills is far from optimal. Impaired users will be tested negatively for the majority of the time that impairment lasts and users that are no longer impaired will be tested positively. Confirmation of the

results determining plasma THC will exonerate the latter, but impaired drivers cannot be detected adequately.

1.1.5.2.2 Oral fluid testing for THC

THC enters oral fluid by deposition into the oral mucosa during smoking or oral administration. The concentration decreases slowly as THC is released from the mucosa and moved to the gastro-intestinal tract with the oral fluid. If at all, only minor amounts enter oral fluid by passive diffusion from the blood.[109,110] THC can be detected in oral fluid immediately after smoking and, depending on the method, for a mean of 13 or 15 h. The maximum detection time is 34 or 31 h (**Figure 6**). Samples of chronic daily users under monitored abstinence were positive for up to 48 h.[101,108] Therefore, the timewise correlation between positive results and impairment is much improved over urine sampling, if still longer than the actual impairment of a maximum of 10 h (section 1.1.5.2.1).

Problems associated with oral fluid testing consist of the actual sampling, as THC consumption is known to lead to a dry mouth and collection of the necessitated volume of oral fluid might be difficult.[111,112] The high hydrophobicity of THC complicates the recovery of the drug from oral fluid collection devices. Some devices like the Cozart® and Quantisal® collectors have found a way to circumvent this and ensure high recovery levels.[113,114] Food, beverages, and mouthwashes that are consumed or used can reduce the measured concentration in oral fluid, but are not able to lead to a negative result if the assay's sensitivity is sufficient.[115,116] Passive cannabis smoking was shown to generate measurable THC-concentrations in oral fluid for up to 1 h.[101,117,118] In one study, samples were positive for up to 3 h after the subjects had spent 3 h in a coffee shop.[119] Despite these rare exceptions, oral fluid based assays should be stable enough for initial roadside testing.

Immediately after consumption, THC-concentrations in oral fluid are far higher than in blood. After this initial contamination has dissipated, studies have been able to show correlations between impairment or plasma and oral fluid concentrations, even though inter-subject variations are relatively high.[117,120,121] Nonetheless, the calculation of plasma THC based on oral fluid drug-concentrations remains impossible.[122]

Sensitivity and specificity have long been a problem for oral fluid roadside testing devices for THC[100,123-125]. The last years have seen the development of more promising tests even though they still do not reach the sensitivity and specificity of 90% that have been set as limit for testing devices by the European roadsite testing assessment project (ROSITA).[126] The Dräger DrugTest5000® showed sensitivities of 92 and 93% and specificities of 97 and 71% in two different

studies.[127,128] With those values, it remains the best oral fluid roadside testing device on the market right now.

1.2 Aptamers

Aptamers are oligonucleotide-based affinity tools. They are selected *in vitro* from libraries comprised of single-stranded RNA- or DNA-oligonucleotides to bind a target molecule with high affinity and selectivity.[129,130] Aptamers have been identified for proteins[131], small organic molecules[132], ions[133], cells[134], whole viruses[135], and even *in vivo* for tumours.[136] They fold into a three-dimensional structure, which enables recognition of their target molecule. Therefore, interaction conditions like e.g., buffer system, pH, or temperature cannot be varied without risking partial or complete loss of affinity.[137]

As affinity tools, aptamers combine several advantages: They are thermally stable and can be renatured after denaturation. Their chemical synthesis enables an easy way to incorporate modifications at specific positions combined with relatively low production costs and no batch-to-batch variations.[138] In addition, they have low immunogenicity.[139,140] This explains the increasing attention aptamers have been gaining for the development of diagnostics[141,142] as well as therapeutics.[143-146] Nonetheless, commercialised aptamer-based therapeutics and diagnostics remain rare.[147]

Aptamer-based biosensors, also called aptasensors, are one of the consistently growing uses of aptamers in the last couple of years. A wide variety of mechanisms[148] as well as targets has been explored and many sensors even work in real samples and reach limits of detection and quantification sufficient for real-life use. Problematic remains the stability and reproducibility necessary for commercialisation as the method descriptions in the proof-of-concept papers often lack vital details like exact buffer conditions.[147] Additionally, companies dealing with aptamers are still rare in comparison to those experienced with antibodies.[147] The investment needed to commercialise aptamer-based sensors (or any other kind of therapeutic or diagnostic) is therefore much higher than for antibodies, where well-established protocols and knowledge shorten and cheapen the process.

1.2.1 SELEX

Aptamers are (mostly) selected *in vitro* by a technique called systematic evolution of ligands by exponential enrichment (SELEX). The starting library (SL) consists of single-stranded DNA or RNA with a randomised region flanked on the 5'- and 3'-end with constant primer binding sites for amplification during polymerase chain reaction (PCR). In the first selection cycle, the SL is incubated with the

target molecule of interest. Often, this target molecule is immobilised on some kind of solid phase like magnetic beads or sepharose to enable easy separation of binding and non-binding sequences.[149] Other options include the use of the change in migration between bound and non-bound nucleic acids in electronic fields in capillary electrophoresis [150] or the immobilisation of the SL, which is eluted from the beads upon binding to the target of interest.[151]

After incubation with the target molecule, non-binders are removed and binding sequences recovered and amplified. Depending on the nature of the nucleic acids used for SELEX, the resulting PCR-product is either transcribed into RNA (for RNA-SELEX) or the 3'-strand is digested, resulting in single-stranded DNA (for DNA-SELEX). The enriched library can then be used for the subsequent selection cycle.[149]

After in average 8-12 selection cycles, the enriched DNA/RNA is tested for interaction with the target molecule. If significantly increased interaction in contrast to the DNA of the SL is observed, the DNA from the enriched selection cycle is sequenced. Single sequences are further characterized for affinity and specificity and those that show both for the target molecule are defined as a new aptamer.[152]

1.2.2 Aptamer identification through next-generation sequencing

Sequencing of the DNA from enriched selection cycles is necessary for aptamer identification and can be performed with two different techniques: Sanger and next-generation sequencing. Using traditional Sanger sequencing[153], the DNA from about 50 clones can be sequenced. As the gained number of sequences is low, the final selection cycle has to be highly enriched to enable aptamer identification. In contrast, the millions of reads obtained from next-generation sequencing (NGS) allow the identification of only slightly enriched aptamer candidates. They also create the unique opportunity to gain an insight into the selection process as DNA of several selection cycles can be sequenced. The enrichment of single aptamer candidates can be followed over the course of the SELEX and the impact of changes in selection conditions can be directly evaluated.[154,155] It is therefore not surprising that NGS is used for aptamer identification by an increasing number of research groups.[156-158] In addition, several analysis tools have been described to enable fast and easy analysis of the resulting NGS data.[154,159-162]

NGS can be performed using different techniques.[163] All data in this thesis was obtained using sequencing by synthesis (SBS), which is implemented in Illumina sequencers.[164] For this technique, average error rates between 0.1 and 2.6% have been published with single nucleotide substitutions as the most abundant error type.[165-168] Independent of the sequencing technique, mutations have also

been assigned to result from PCR errors either during sample preparation or sequencing.[165,169]

In addition, some errors are intrinsic to the SBS approach[170]: Overlays of excitation and emission spectra of the different fluorophores lead to colour or laser cross-talk, which can result in base miscalling.[171] The same effect can also be observed between adjacent clusters.[172] Phasing describes two phenomena that lead to the resulting strand being out of phase with the actual sequencing rate: Pre-phasing leads to the strand preceding all others as two or more nucleotides are incorporated in one cycle. This can result from inadequate flushing of the flow cell as leftover dNTPs are incorporated after removal of the terminator. (**Figure 7, right strand**) Post-phasing leads to the strand lagging behind. Here, the terminator is not completely removed, so no nucleotide can be added.[173] (**Figure 7, left strand**) If the terminator cannot be removed, the fluorescence of the cluster is dimmed. This also occurs over the course of the sequencing due to laser damage to the DNA strands, which reduces the number of strands in a cluster.[171]

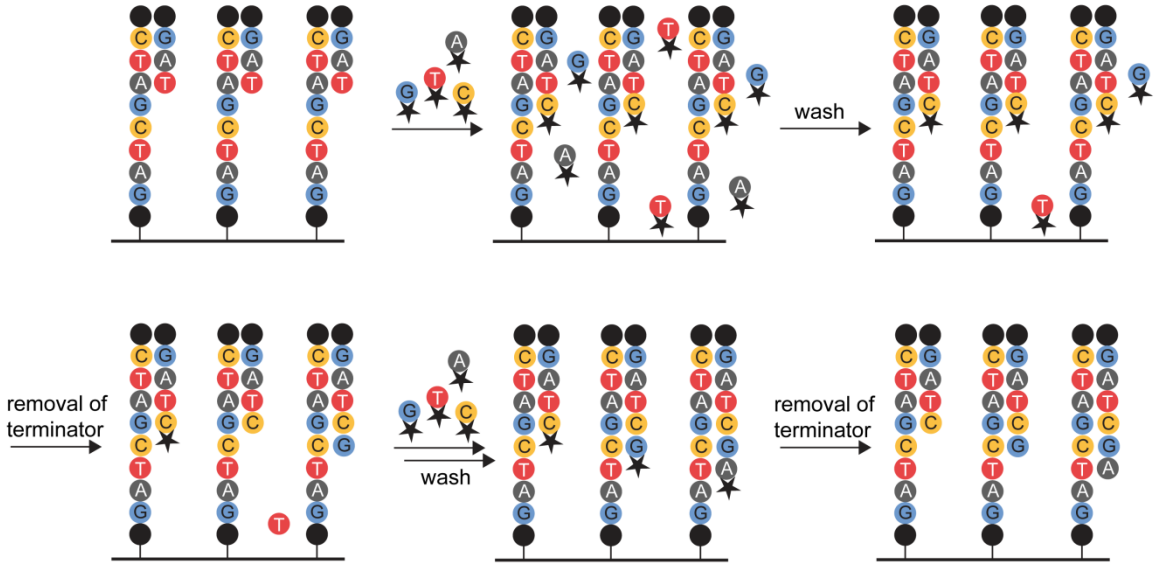


Figure 7: Phasing effects
The left strand depicts a post-phased sequence, the right a pre-phased one. The middle strand continues without phasing effects of any kind. The black dots represent the sequencing primers and the black star the terminator, which prevents addition of the subsequent nucleotide. After incorporation of the correct dNTP (upper middle), the non-incorporated dNTPs are washed away (upper right). If some remain due to inadequate washing, these are incorporated into the growing strand after removal of the terminator (lower left, right strand), resulting in a pre-phased sequence. If the terminator cannot be removed (lower left, left strand), no base will be attached in the next sequencing cycle (lower middle, left strand). This results in a post-phased sequence.

As these errors are well-known, efforts have been undertaken to find methods to correct for them. Bustard, the base-calling software implemented in Illumina sequencers, corrects for phasing based on a constant phasing rate.[174] BayesCall and Ibis also consider the surrounding bases.[174,175] The more complex All Your

Base adjusts for e.g., cycle-wise differences in cross-talk on a run-by-run basis.[171]

1.2.3 Nucleobase-modified aptamers

Specific characteristics of an identified aptamer can be improved by using a wide variety of modifications. While sugar-modifications like 2'-fluoro or 2'-methoxy are mostly used to increase RNA-stability[176], locked nucleic acids (LNA, containing a 2'-O,4'-C-methylene bridge) also have an effect on duplex stability.[177,178] Additions to the 3'- and 5'-end can increase nuclease resistance (by capping the end using an inverted nucleotide) or decrease renal filtration and thereby improve the serum half-life (by addition of a high molecular weight PEG).[176]

Although these modifications can already be included during selection[140], many are added post-selection to improve the stability or pharmacokinetics of the aptamer. A prime example is the FDA-approved (Food and Drug Administration) Pegaptanib, an aptamer against age-related macular degeneration. Of its 27 bases, only two remain unmodified.[144]

Nucleobase-modifications, in contrast, are chosen to improve the interaction between the target molecule and the aptamer. They all build on the premise that the interaction possibilities of nucleic acids are much more restricted than those of antibodies due to the high chemical diversity of the amino acids in comparison with the four canonical nucleobases.[179] This is regarded as the reason for the failure of about 70% of all aptamer selections for protein targets. Modifications of the nucleobase increase the available chemistries and thereby the interaction possibilities. They were shown to elevate the success rate of selections to 84% in the case of slow off-rate modified aptamers (SOMAmers).[180]

The first aptamer containing a nucleobase-modification (5'-(1-pentynyl)-2'-deoxyuridine) was selected for thrombin by Latham *et al.* in 1994. The ability to interact with thrombin depended on the modification, but the affinity of the modified aptamer was no better than those reached with non-modified aptamers.[181,182]

Several options for the selection of nucleobase-modified aptamers have since been established (sections 1.2.3.1 to 1.2.3.3).

1.2.3.1 Aptamers containing an expanded genetic alphabet

While not strictly nucleobase-modifications, the use of an extended genetic alphabet definitely increases the available interaction possibilities. Two different groups published approaches for the selection of aptamers using different additional nucleobases.

Kimoto *et al.* base their work on the additional base pair dDs-dPx (Ds: 7-(2-thienyl)imidazo[4,5- b]pyridine, Px: 2-nitro-4-propynylpyrrole). The base pair interacts through hydrophobic and packing forces in contrast to the hydrogen bonding responsible for the pairing of the natural nucleobases (**Figure 8A**).[183] dDs is incorporated into the DNA used for selection, whereas dPx is only used for PCR-amplification. The resulting library therefore consists of five different nucleobases – the four canonical ones plus dDs.

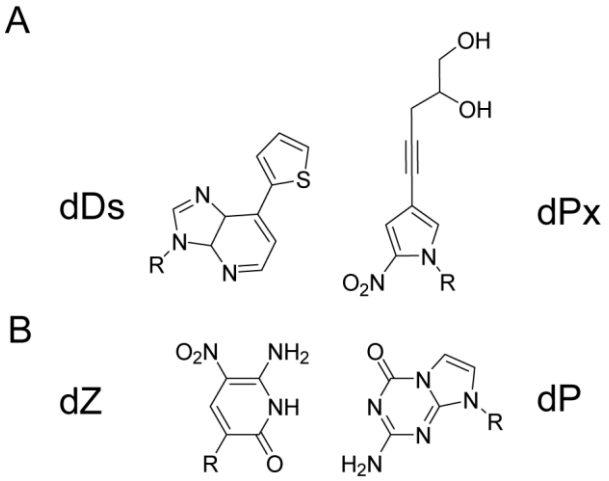


Figure 8: Additional base pairs in aptamers with an extended genetic alphabet
(A) dDs-dPx base pair, whose interaction is based on hydrophobic and packing forces. dDs has been implemented in aptamer selections. (B) dZ-dP base pair, whose interaction is based on hydrogen bonding. The base pair has been used for aptamer selections.

While the original selections constricted dDs to predetermined positions to enable their identification via sequencing of a corresponding index code[184,185], this restriction was recently overcome and aptamers for the von Willebrand factor A1-domain were selected with a completely randomised library.[186] The fact that the affinity of all aptamers selected with the hydrophobic dDs was significantly higher in contrast to those selected with the natural nucleobases and increased further with the complete randomisation of the additional base dDs supports the hypothesis that an increase in interaction possibilities of the nucleobases also improves aptamer selection and characteristics.[184,186,187]

The second unnatural base pair used for aptamer selection is dZ-dP (dZ: 6-amino-5-nitro-3-(1′-β-D-2′-deoxyribofuranosyl)-2(1H)-pyridone, dP: 2-amino-8-(1′-β-D-2-deoxyribofuranosyl)-imidazo[1,2-a]-1,3,5-triazin-4(8H)one). Like the natural base pairs, its interaction is based on hydrogen bonding (**Figure 8B**).[183] Both additional bases were inserted into the random region, resulting in a library composed of six different nucleobases.

Aptamers containing these six nucleobases were published for breast cell cancer cells, for a protein overexpressed on a cell line, and for a purified protein.[188-190] All of them contained fewer unnatural bases than the SL[179], which might result from

an imperfect retention of those bases during PCR.[188] Even though the aptamers had similar affinities as those consisting of only the natural base pairs[188-190], one of them showed an uncharacteristic peak in CD-spectroscopy experiments. This might reflect an as of now unknown folding behaviour induced by the additional base pair, which would also change the interaction possibilities of aptamers containing such base pairs.[190]

1.2.3.2 SOMAmers

SOMAmers have been invented by SomaLogic® as an alternative to conventional aptamers. They consist of nucleobase-modified aptamers that contain uracil carrying an amide-linked modification at position five instead of thymidine. A wide variety of modifications, mostly based on amino acid side chains, has been used for selection successfully (**Figure 9A**).[180,191]

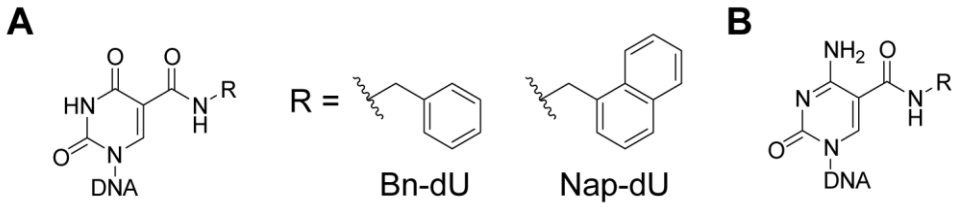


Figure 9: Modified nucleotides in SOMAmers
(A) The modified uridine is used as fourth base for SOMAmer selection. The modification is attached via an amide-linker. Shown are two exemplary hydrophobic, aromatic modifications that are present in the so-far crystallised SOMAmers. (B) The modified deoxycytidine was used for the selection with two different modifications. The same modifications as with the modified uridine were used.

While the technique has some drawbacks – a modification has to be synthesised as both phosphoramidite and triphosphate to be useable for selection and a two-step PCR procedure is needed to gain a fully modified, enriched library[192] – it was the first to prove that nucleobase-modifications are able to dramatically increase selection success for protein targets as mentioned above (section 1.2.3). At the same time, affinities are strongly improved in comparison with aptamers consisting of the canonical nucleobases.[180,191] Of all tested modifications so far, aromatic, hydrophobic side chains like those shown in **Figure 9A** have proven to be the most successful.[191]

A recent publication showed that this technique enables the selection with two different modifications: In addition to uridine, deoxycytidine is also modified with an amide linker at position five (**Figure 9B**). Two different modifications lead to even further improved SOMAmer affinities in comparison with those carrying only one modification.[193]

SOMAmers are the first of the published nucleobase-modified aptamers that were crystallised and therefore give an insight into the impact these types of modifications have on aptamer – or SOMAmer – folding and the molecular interaction with the target molecule. Four different SOMAmer structures have been solved: benzyl-deoxyuridine (Bn-dU, **Figure 9A**) ones for platelet-derived growth factor B (PDGF-BB, PDB-ID: 4HQX and 4HQU)[194], interleukin-6 (IL-6, PDB-ID: 4NI7 and 4NI9)[195], and nerve growth factor (NGF, PDB-ID: 4ZBN)[196], as well as a naphthyl-modified (naphthyl-deoxyuridine, Nap-dU, **Figure 9A**) SOMAmer for interleukin-1 (IL-1, PDB-ID: 5UC6).[197] In all of them, the hydrophobic modifications cluster together and stabilise compact conformations. Often, these contain non-helical zipper structures, in which the modification π-stacks with a second uridine and acts as a pseudo base pair (**Figure 10**).[197] The interaction with the target molecules is based on less polar contacts than in non-modified aptamers as the hydrophobic modifications take over the majority of this interaction.[194,195,197] An exception to this is the NGF-SOMAmer, of which only few modified nucleotides directly interact with the target protein. Most build hydrophobic clusters that stabilise an S-shaped structure made up of three aligned strands.[179,196]

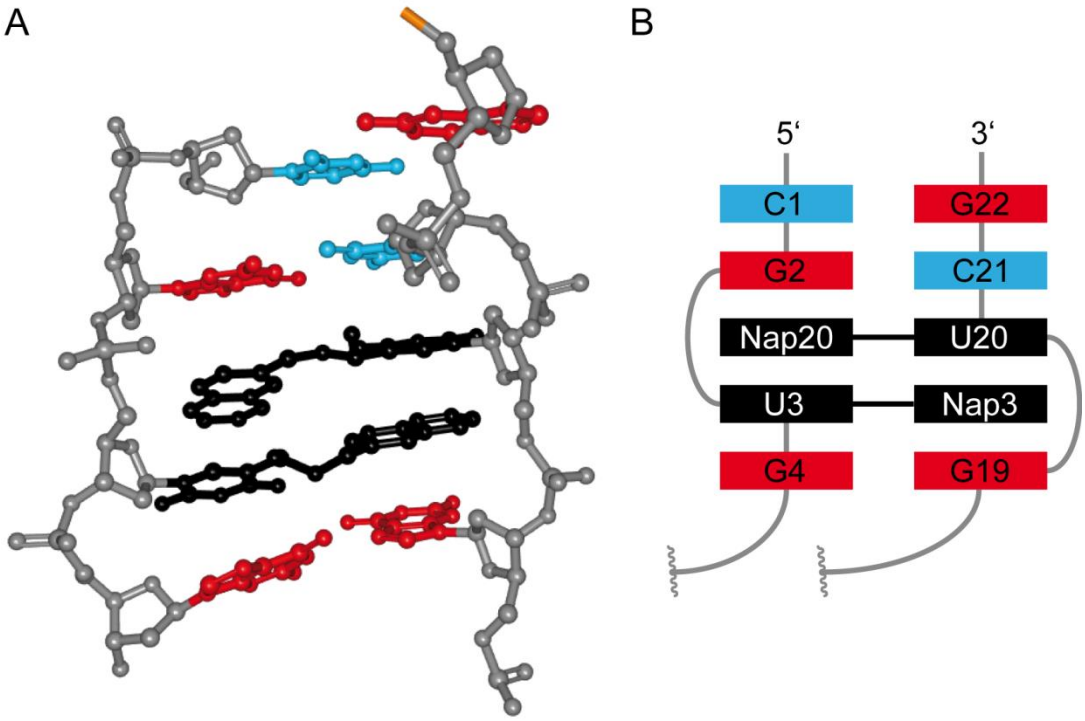


Figure 10: Zipper structure in the IL-1-SOMAmer
(A) Excerpt from the IL1-SOMAmer structure showing the non-helical zipper structure that can be also be found in other SOMAmers. The naphthyl-modification π-stacks with the adjacent uridine. This way, two modified uridines act as base pairs and unwind the helical structure. The structure (PDB-ID: 5UC6) was modified with the freeware program RCSB PDB Protein Workshop 4.2.0.[198] (B) Schematic representation of the structure shown in (A) to illustrate the base pairing. Cytosines are depicted in light blue, guanines in red, and naphthyl-modified uridines in black. The backbone is shown in grey.

1.2.3.3 Clickmers

Copper-catalysed alkyne-azide cycloaddition (CuAAC, click chemistry) has also been used to modify nucleobases for aptamer selection. CuAAC is one of the most widely utilised techniques to functionalise biomolecules nowadays.[199,200] It has become popular because the tags are small and do not react with functional groups intrinsic to biomolecules and the reaction is quantitative.[199] Chelating agents can be used to solve the problem that Cu^{+} in solution damages DNA, facilitating the use of click chemistry to functionalise DNA.[201]

The group of Thomas Carell showed that alkyne-modified DNA-building blocks can be incorporated into DNA, amplified by PCR, and modified by CuAAC with a multitude of functionalisations.[202-204] This enabled the development of a CuAAC-based technique for the selection of nucleobase-modified aptamers, called clickmers.[205]

For this, the SL is synthesised with 5-ethynyl-2'-deoxyuridine (EdU) instead of thymidine and modified with the functionalisation of choice via CuAAC (**Figure 11**). The click-modified DNA is incubated with the target molecule as in a normal SELEX, followed by wash steps to remove unbound DNA and the recovery of the bound DNA. During PCR, unmodified 5-ethynyl-2'-deoxyuridine triphosphate (EdUTP) is incorporated, which enables one-step amplification of the functionalised DNA. Single-strand displacement can be performed as with unmodified DNA because the alkyne does not affect enzymatic compatibilities. Afterwards, the enriched ssDNA is refunctionalised by CuAAC and the thereby gained DNA can be used for the subsequent selection cycle (**Figure 12**).[152,205]

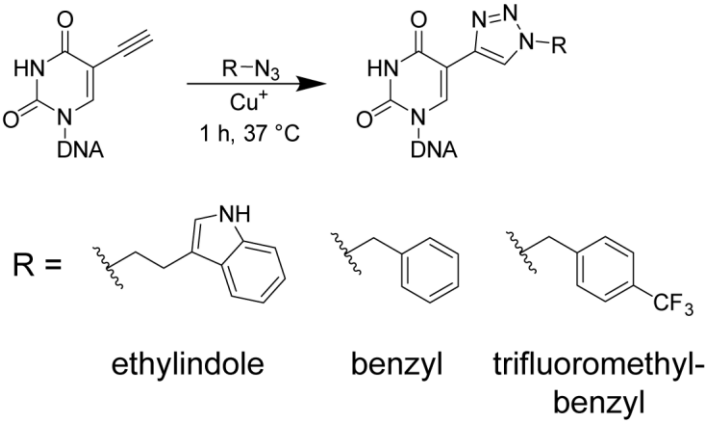


Figure 11: Copper-catalysed alkyne-azide cycloaddition
Indicated are the modified nucleobases resulting from the use of ethylindole-azide as applied in the published click-SELEX[205] as well as those resulting from benzyl- and trifluoromethylbenzyl-azide, both of which are utilised in this thesis.

In contrast to the above mentioned technique using an expanded genetic alphabet, the modification can be freely chosen as long as the respective azide is commercially available or can be synthesised. This is also possible for SOMAmers, but the latter requires synthesis of the modified triphosphate for PCR-amplification and the modified phosphoramidite for synthesis of the library. Click-SELEX relies on commercially available building blocks. Modifications can be chosen to suit the target without laborious syntheses and change of the modification can modulate clickmer affinity. In addition, the removal of the modification during amplification facilitates the use of larger modifications that might otherwise lead to enzymatic compatibility issues.[152]

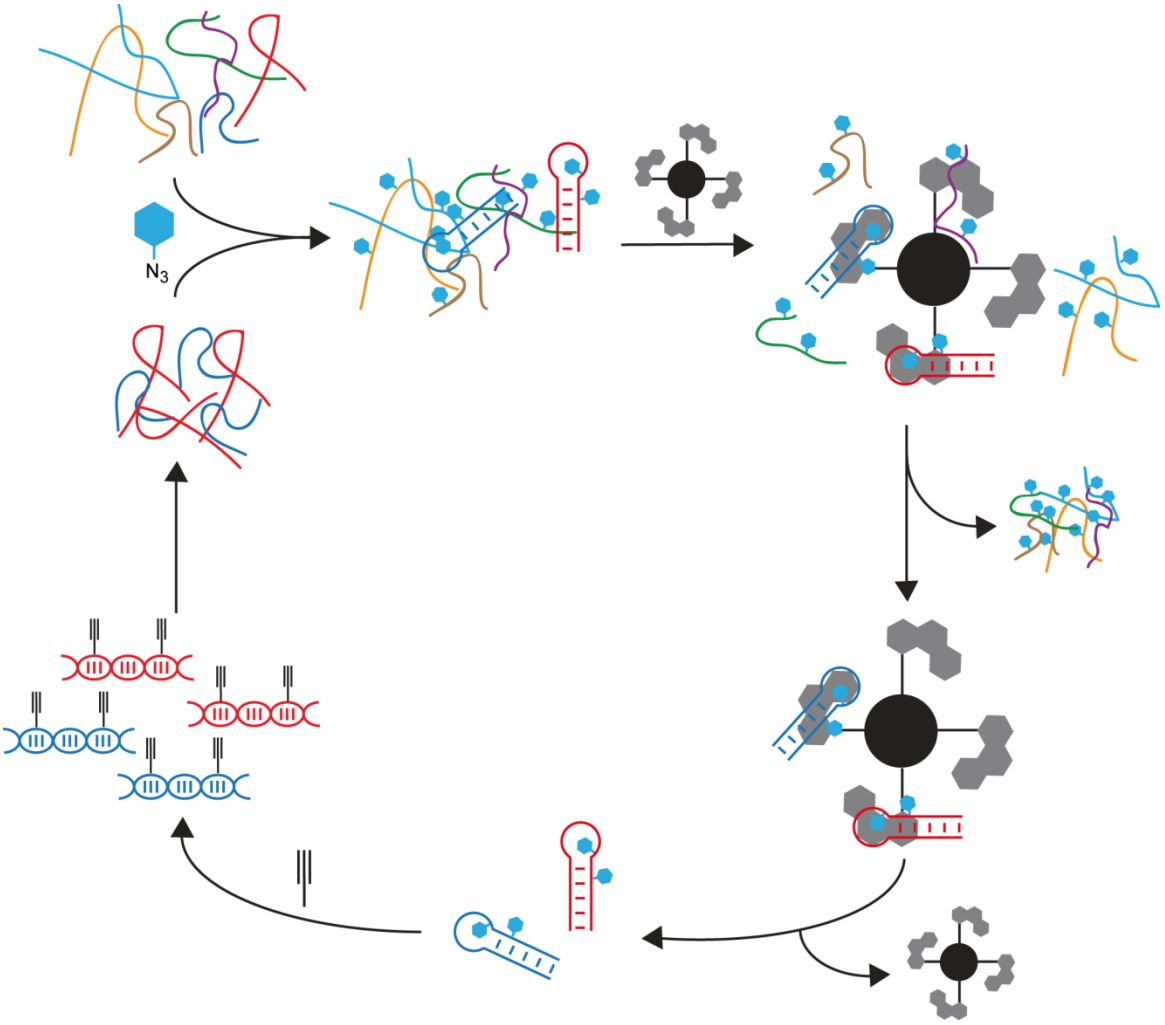


Figure 12: Schematic illustration of a click-SELEX cycle
The SL containing EdU instead of thymidine is modified with the azide of choice and incubated with the immobilised target molecule. After removal of non-binding sequences by washing, the binding sequences are recovered and amplified by PCR using EdUTP. The functionalisation is thereby removed. The alkyne-modified DNA can be enzymatically digested to the single strand and the enriched library refunctionalised by CuAAC before starting the next selection cycle.

As of now, the technique has been shown to work for a model target, GFP, with an ethylindole-azide as clicked-in modification.[205]

2 Aim of the study

THC is the most widely used illegal drug worldwide. Amongst a multitude of effects, its use negatively affects driving skills. This necessitates the ability to detect THC on the roadside. As urine tests fail to detect THC in impaired drivers and oral fluid tests do not reach the needed sensitivity and specificity, a different approach was called for.

The objective of this study was the identification and characterisation of THC-binding aptamers that could then be used for the development of oral fluid-based roadside tests. In contrast to antibodies, aptamers are thermostable and can be stored at room temperature. In addition, their chemical synthesis enables relatively cheap and consistent production. Therefore, such a test would probably be cheaper and more stable than an antibody-based one and could be stored in a police car until needed without necessitating any kind of cooling equipment.

For the selection of THC-binding aptamers, strategies for the immobilisation of THC on a solid support had to be developed. As THC itself only contains a single functional group that can be used for immobilisation, the carboxylated, biological precursor, THCA, was also taken into consideration.

To evaluate binding of potential aptamers in solution, knowledge about THC-solubility was imperative. Due to its high hydrophobicity and the resulting limited solubility in aqueous solutions in combination with the low solubility of DNA or 2’F-RNA in organic solvents, conditions that keep DNA, 2’F-RNA, and THC in solution had to be identified. The most likely candidates were combinations of buffers and low percentages of organic solvents.

3 Results

Selections for THC-binding aptamers were performed both with non-nucleobase-modified nucleic acids (section 3.1) as well as with the technique called click-SELEX (section 1.2.3.3), where click-chemistry is used to introduce modifications into the nucleic acid library used for selection. The respective results of the selection as well as experiments undertaken to characterise the resulting THC-binding sequences can be found in section 3.2. The final section, 3.3, illustrates the results of a toggle-SELEX, for which THCA-modified magnetic beads and THC-sepharose were used alternatingly as solid support during the selection process.

3.1 THC-selections with non-nucleobase-modified DNA and 2'F-RNA

The initial selections for THC-binding aptamers were performed with non-nucleobase modified DNA and 2'F-RNA, both of which are nuclease-resistant and might be usable in saliva. Prior to the actual selection, THC was immobilised. Two different solid supports were used: amine-functionalised magnetic beads and epoxy-activated sepharose. The immobilisation technique as well as the selections performed using the respective solid support and their results are described in section 3.1.1 for the amine-functionalised magnetic beads and in 3.1.2 for epoxy-activated sepharose.

3.1.1 Selections on THCA-functionalised amine-dynabeads

In order to retain the hydroxyl group of THC and its possibility of forming hydrogen bonds with a potential aptamer, selections for a THC-binding aptamer were performed using the biological precursor THCA. The additional carboxyl group could be used for the immobilisation on amine-functionalised magnetic beads (**Figure 1A** and **B**).

Due to the low solubility of THCA in aqueous solutions, the immobilisation was performed in DMSO, which is well tolerated by the magnetic beads.[206] THCA was activated with N,N'-dicyclohexylcarbodiimide (DCC) and the activated carboxylic acid reacted with the amine-groups on the magnetic beads, building an amide bond between magnetic bead and THCA (**Figure 13**).

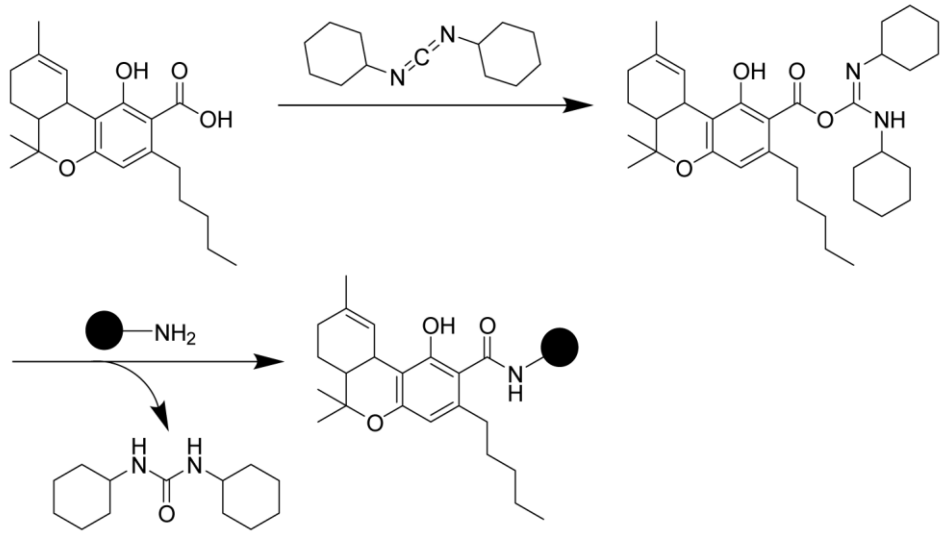


Figure 13: Coupling of THCA to magnetic amine-dynabeads
THCA (upper left) is activated using DCC (upper arrow). Upon addition of amine-functionalised magnetic beads (lower arrow), the activated THCA is coupled to the beads via an amide-bond and dicyclohexylurea is cleaved off.

In order to verify the successful coupling of THCA to the magnetic amine-beads, a THC-binding antibody was used. **Figure 14** shows the signal from both the THCA-coupled beads as well as the non-modified amine-functionalised beads. A clear increase in antibody-signal was detected for the THCA-beads. THCA had therefore been successfully coupled to the magnetic beads and these were used for the selection of THC-binding aptamers.

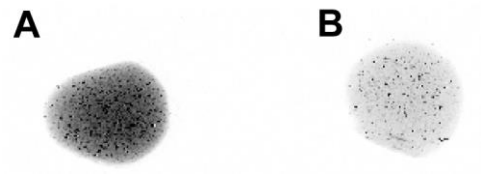


Figure 14: Detection of THCA on magnetic amine-beads
Detection of THCA using a THC-antibody. (A) Signal for THCA on magnetic amine-beads that have been coupled to THCA. (B) Background signal on non-modified magnetic amine-beads.

Overall five different selections were performed using THCA-beads as solid support: Two selections used 2'F-RNA-, two DNA-based libraries, and one selection was performed using 2'F-RNA in a library that is prestructured to form a G-quadruplex. **Table 3** gives an overview over the respective length of the random regions, selection buffers, elution methods, and the number of selection cycles.

The DNA or 2'F-RNA of several selection cycles of each selection (**Table 3**) was tested for its interaction with THCA-beads as well as non-modified magnetic amine-beads. The DNA or 2'F-RNA of the last respective selection cycle did not

show improved interaction with the THCA-beads in comparison with the non-modified beads or the DNA/2'F-RNA from earlier selection cycles (**Figure 15**). For SELEX 1 (**Figure 15A**), the interaction with non-modified beads was not tested as it was already apparent from the tests on THCA-beads that no enrichment of THC-binding sequences had occurred.

Table 3: Selections performed on THCA-modified amine beads with non-nucleobase-modified DNA/RNA

SELEX	Library	Buffer	Elution	Selection cycles
1	N40 2'F-RNA	PBS, pH 7.4 500 mM NaCl, 3 mM $MgCl_2$, 0.1 mg/ml salmon sperm DNA	heat: 3 min 80°C	15
2			affinity: 5 mM THC, 7% Cremophor	
3*	G-quadruplex 2'F-RNA	50 mM Hepes, pH 8.0 50 mM KCl, 3 mM $MgCl_2$ heparin	(Beads into PCR)	10
4	N43 DNA			
5	N75 DNA			

* performed by Patrick Günther

In general, all selections on THCA-coupled magnetic beads were unsuccessful even though THCA was detectable on the beads (**Figure 14**). Therefore, a different immobilisation was chosen for the subsequent selections.

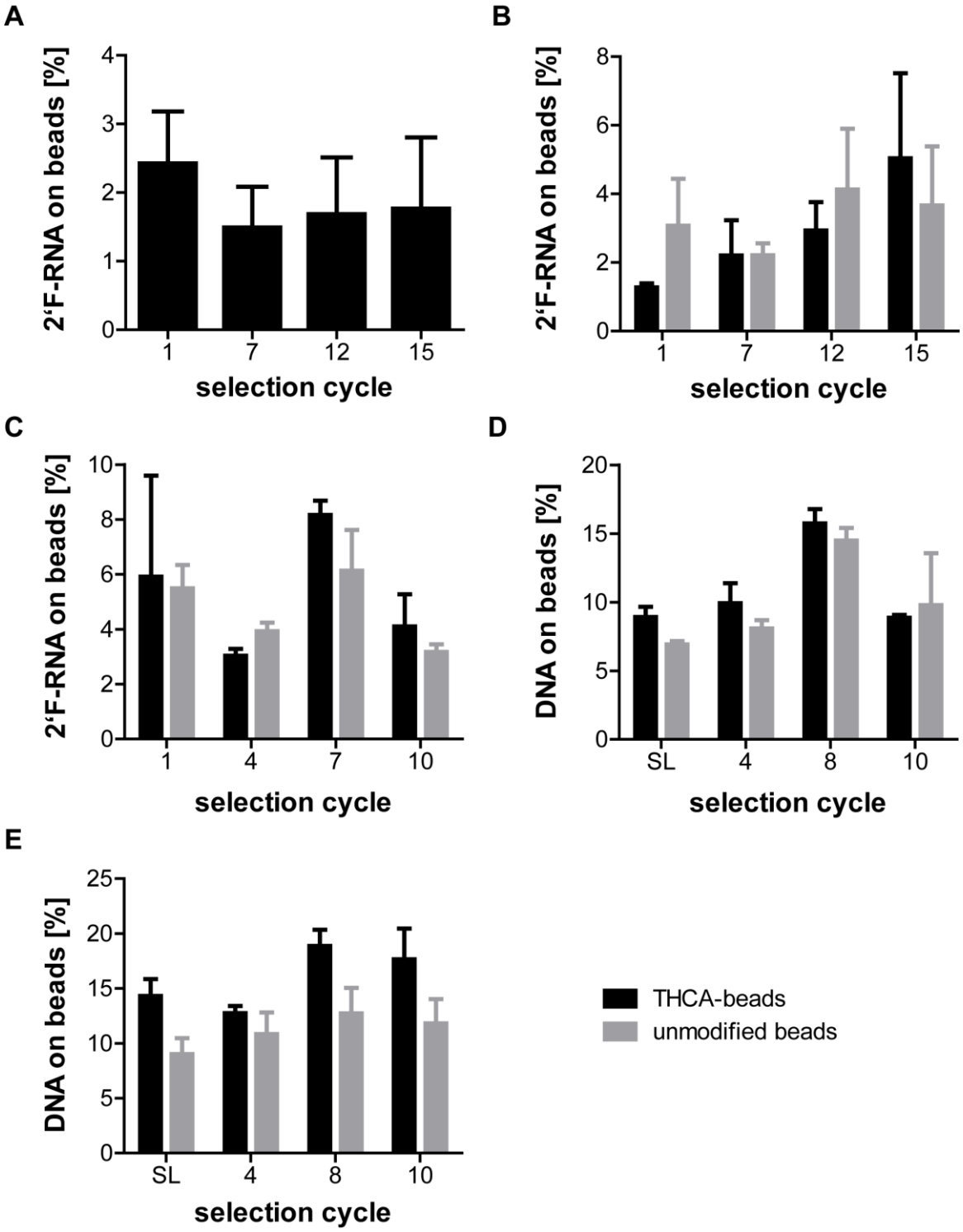


Figure 15: Interaction analysis of THC-selections performed on THCA-modified amine beads with non-nucleobase-modified DNA/RNA

^{32}P-labelled DNA/2'F-RNA was incubated with THCA-beads (black bars) or unmodified beads (grey bars). After two washing steps, the amount of radioactivity on the beads was quantified. Depicted is the interaction analysis of (A) SELEX 1, (B) SELEX 2, (C) SELEX 3, (D) SELEX 4, and (E) SELEX 5. No interaction analysis of SELEX 1 (A) on unmodified beads was performed. The experiments for the data presented in (C) were performed by Patrick Günther.

3.1.2 Selections on THC-sepharose

3.1.2.1 Immobilisation of THC on epoxy-activated sepharose

To optimise the immobilisation of THC on epoxy-activated sepharose, the reaction was first established in solution. The linker that is found on epoxy-activated sepharose is commercially available: 1,4-butanedioldiglycidyl ether (BDE). As with THCA and the magnetic beads, the reaction was carried out in DMSO to ensure THC-solubilisation.

In preliminary tests, 10% triethylamine (TEA) was found to be effective in transferring phenol into the aqueous phase, an indication of its deprotonation. As the acid dissociation constant (pK_A) of phenol (9.99)[207,208] is similar to the one of THC (10.6)[56] identical conditions should be able to deprotonate THC, which should in turn enable its reaction with BDE.

The reaction scheme is depicted in **Figure 16** and includes the exact masses of both educts and the product.

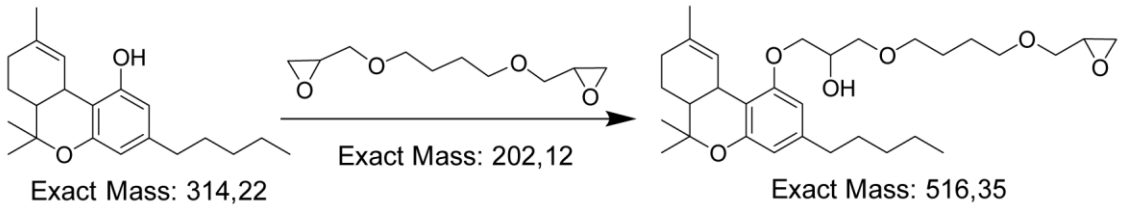


Figure 16: Coupling of THC to BDE in solution
THC (on the left) reacts with BDE (on the arrow) in solution to THC-linker (THC-L, on the right). The reaction takes place in 90% DMSO with 10% TEA as a base at 50°C.

Figure 17 shows the liquid chromatography – mass spectrometry (LC-MS) analysis of the coupling after 1, 3, 6, and 7 days. Peak 1 as indicated in the UV chromatograms in **Figure 17A** represented unreacted THC as apparent by the corresponding mass that was analysed in **Figure 17B**. Peak 2 represented the coupling product, THC-L. The corresponding mass and its sodium adduct (+22 = 539.4) was analysed in **Figure 17C**.

The two peaks between peak 1 and 2 contained masses 513.42 and 515.51, respectively, in addition to the corresponding sodium adducts (data not shown). These masses were only slightly reduced (-4 or -2, respectively) in comparison with the mass of the coupling product, THC-L, and might represent elimination products.

An additional sample with ten times more of the coupling product from day 7 (day7_conc, **Figure 17A**) was also analysed to see if the sample concentration had an influence on the distribution of the peaks and to identify low abundance peaks that might not be visible in the lower concentrated samples. No distinctive differences between the samples from day 7 and day7_conc can be observed.

The tail of the product peak (peak 2) is less pronounced in the concentrated sample, indicating that the tail might result from the low sample concentration and not represent a real contamination.

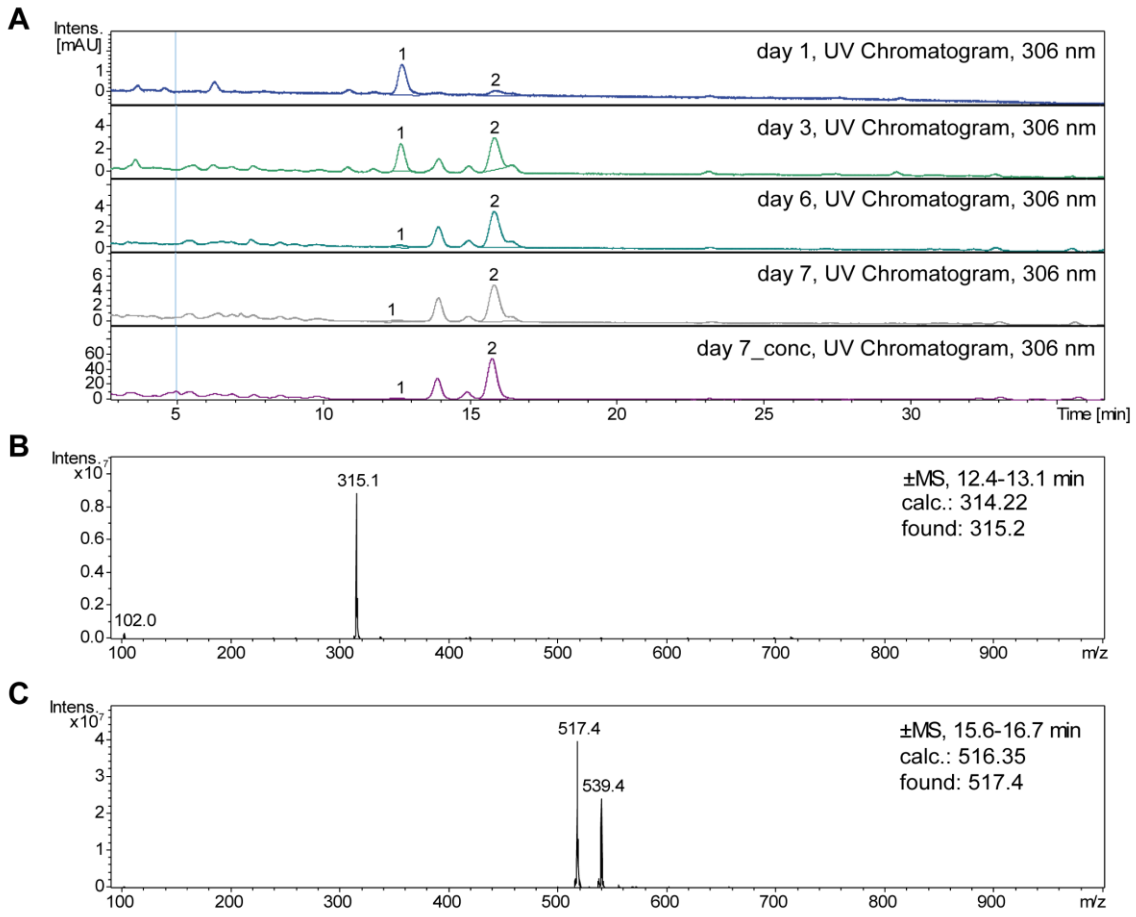


Figure 17: LC-MS analysis of the coupling of THC to BDE after 1, 3, 6, and 7 days
(A) UV-chromatograms of the LC-MS analysis of the coupling of THC to BDE after 1, 3, 6, and 7 days. The last line shows the analysis of a ten times more concentrated, but otherwise identical sample from day 7 (day 7_conc). (B) Positive and negative mass spectrum of peak 1 on day 1. The found mass matches THC. (C) Positive and negative mass spectrum of peak 2 on day 7. The found masses correspond to THC-L and THC-L with Na^+ (+22 = 539.4).

Peak 1 and peak 2 were analysed regarding the area under the curve for quantification of the coupling yields. It was apparent that the percentage of coupling product (peak 2) increased with increasing reaction time (**Figure 17A** and **Figure 18**). The increase in coupling product was only minor between day 6 and 7, indicating that the reaction was completed and longer reaction times would not lead to higher coupling yields. The yield reached 94.3% THC-L after 7 days. In the more concentrated sample, day7_conc, the calculated yield was even higher with 99.4%. As both of these samples were taken after 7 days of coupling, the real yield was probably in between those two values and high enough to enable successful coupling of THC to epoxy-activated sepharose.

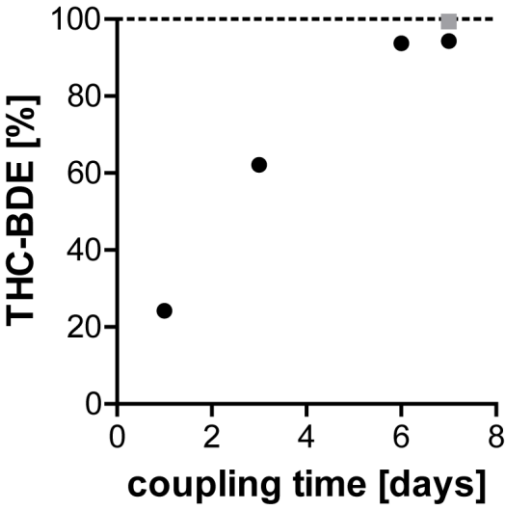


Figure 18: Time course of the coupling of THC to BDE
Yield of coupling product, THC-L, according to the LC-MS-analysis. 100% equals the sum of peak 1 and 2 and is indicated with a dashed line. Black dots represent the normal analysis setup, while the grey rectangle stands for the more concentrated sample (day 7_conc).

The coupling reaction of THC to epoxy-activated sepharose is depicted in **Figure 19**. The linker on the sepharose is identical to BDE, the linker used for the in solution experiments. One of the epoxy rings has opened and reacted with the sepharose.

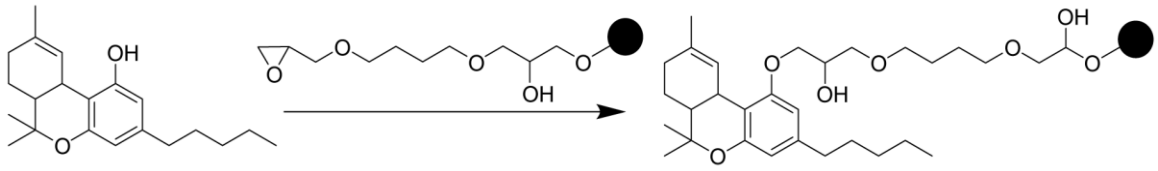


Figure 19: Coupling of THC to epoxy-activated sepharose
THC (on the left) reacts with epoxy-activated sepharose (on the arrow) to THC-sepharose (on the right). The reaction takes place in 90% DMSO with 10% TEA as a base at 50°C for 7 days.

As for the THCA-coupled magnetic beads, the successful coupling was verified using a THC-antibody. THC was clearly detected on the surface of the THC-sepharose, while only a minor background signal was detected on the unmodified epoxy-activated sepharose (**Figure 20**). In comparison to the signal obtained on the THCA-beads (**Figure 14**), the signal on the THC-sepharose was much more intense, while the background on the non-modified sepharose was only slightly more pronounced. This indicates a denser coupling of THC on the sepharose than THCA on the amine-beads.

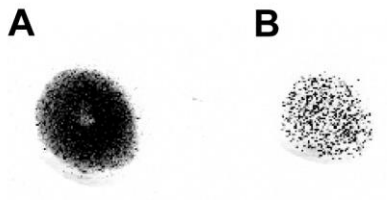


Figure 20: Detection of THC on epoxy-activated sepharose
Detection of THC using a THC-antibody. (A) Signal for THC on epoxy-activated sepharose that has been coupled to THC. (B) Background signal on non-modified epoxy-activated sepharose.

3.1.2.2 Selections performed on THC-sepharose

After successfully coupling THC to epoxy-activated sepharose, two different selections were performed: One with a traditional, unstructured DNA-library and one with a DNA-library that is prestructured to form a G-quadruplex. This library had been used before as a 2'F-RNA-library in SELEX 3 (**Table 3**). **Table 4** gives an overview of the selection libraries, buffer, elution technique, and number of selection cycles.

Table 4: Selections performed on THC-sepharose with non-nucleobase-modified DNA

SELEX	Library	Buffer	Elution	Selection cycles
6	N43 DNA	PBS, pH 7.0 20 mM K^+, 3 mM $MgCl_2$	heat: 10 min 65°C, 4.25 M urea, 12.5 mM EDTA	10
7	G-quadruplex DNA			

The DNA of several selection cycles of each selection was tested for its interaction with THC-sepharose as well as unmodified sepharose. For both selections, a clear increase in interaction with the unmodified sepharose, but not with the THC-sepharose was observed for DNA from higher selection cycles (**Figure 21**). Therefore, the selections most likely only enriched sepharose-binding sequences that are incapable of recognising THC-sepharose.

At this point, an overall number of seven selections with DNA, 2'F-RNA, libraries of varying random region lengths, different buffers, elution techniques, and immobilisation techniques had failed. Therefore, a completely new and innovative approach was needed for the selection of THC-binding aptamers.

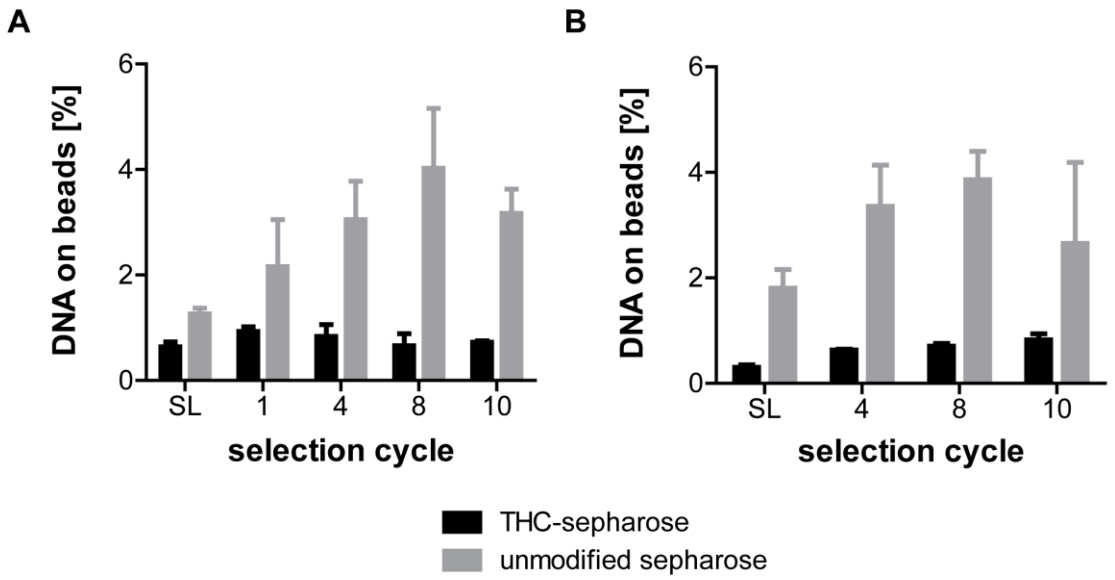


Figure 21: Interaction analysis of THC-selections performed on THC-sepharose with non-nucleobase-modified DNA
^{32}P-labelled DNA from different selection cycles and the SL was incubated with THC-sepharose (black bars) or unmodified sepharose (grey bars). After two washing steps, the amount of radioactivity on the sepharose was quantified. Depicted is the interaction analysis of (A) SELEX 6 and (B) SELEX 7.

3.2 Click-SELEX on THC-sepharose

None of the THC-selections with non-nucleobase-modified nucleic acids had resulted in an enrichment of THC-binding sequences (section 3.1). This might be due to the high hydrophobicity of THC and its low amount of functional groups that allow little interaction with the canonical nucleic acids.

Click-SELEX enables the modular introduction of nucleobase-modifications into the DNA-library that is used for selection (section 1.2.3.3). It therefore presents a possibility to introduce side chains into the nucleic acid library that might enable its interaction with THC. Based on the knowledge that the interaction of THC with the cannabinoid receptor is mainly conferred through π-π-stacking with phenylalanine residues (section 1.1.2, **Figure 3**, **Table 2**), we chose benzyl- and trifluoromethylbenzyl- (CF_3-) azides for the modification of the library and performed two selections in parallel, one with each modification.

The following sections detail the selections with those two modifications (section 3.2.1) as well as the characterisation and sequencing of the DNA from the enriched libraries. The initial characterisation of the THC-sepharose-binding sequences that resulted from those selections follows (section 3.2.2 and 3.2.3). The NGS approach used for sequencing is verified (section 3.2.4) before section 3.2.5 deals with the determination of THC-solubility. This was necessary for the

analysis of the interaction of the THC-sepharose binding sequences with THC in solution.

As none of the sequences showed interaction with THC in solution, a re-selection using affinity elution was performed. Details on the selection itself, the subsequent sequencing, and on the identified sequences can be found in section 3.2.6.

Finally, fluorescence polarisation assays were performed for the interaction analysis between THC in solution and the THC-sepharose-binding sequences. The synthesis and purification of two different THC-FITC-compounds and the corresponding fluorescence polarisation assays are detailed in sections 3.2.7 and 3.2.8.

3.2.1 Click-SELEX

Two click-selections utilising either benzyl- or CF_3-azide were performed to identify THC-binding clickmers. An overview of the changing selection conditions can be found in **Table 5**. The amount of THC-sepharose was kept constant and binding sequences were eluted using heat elution.

Table 5: Selection conditions for the click-SELEX

SELEX cycle	wash volume	PCR-cycles
1	2 mL	B: 14 C: 18
2	3 mL	B: 18 C: 18
3	4 mL	B: 15 C: 17
4	5 mL	B: 16 C: 14
5	6 mL	B: 21 C: 16
6	7 mL	B: 22 C: 19
7	8 mL	B: 18 C: 20
8	9 mL	B: 19 C: 19
9	10 mL	B: 18 C: 18
10	11 mL	B: 15 C: 15

B: Selection using benzyl-modified DNA. C: Selection using CF_3-modified DNA.

For both selections, the selection buffer was composed as follows: 1x D-PBS (pH 5.4, with 4.93 mM $MgCl_2$, and 9 mM $CaCl_2$) with 1 mg/ml BSA, 0.1 mg/ml salmon sperm DNA, and 0.1% Tween-20. The incubation was performed at 20°C for 15 min to imitate a temperature and time frame that are realistic for a roadside testing device.

As a slight decrease in the number of PCR-cycles was detected only in cycle 10 and none of the prior selections with non-nucleobase-modified nucleic acids had succeeded, the wash volume was only slightly and slowly increased over the selection cycles. For the same reason, the amount of THC-sepharose and therefore the amount of target was kept constant throughout the selection process.

3.2.2 Characterisation and sequencing of the benzyl-modified click-SELEX

After ten selection cycles, the interaction between THC-sepharose and the benzyl-modified DNA from different selection cycles was evaluated. DNA from selection cycle 8 and 10 showed a clearly increased binding to THC-sepharose, both in comparison to the interaction with unmodified sepharose as well as to the interaction of the DNA of selection cycle 1 and 4 (**Figure 22A**). This indicates that the selection was successful in enriching THC-sepharose-binding sequences.

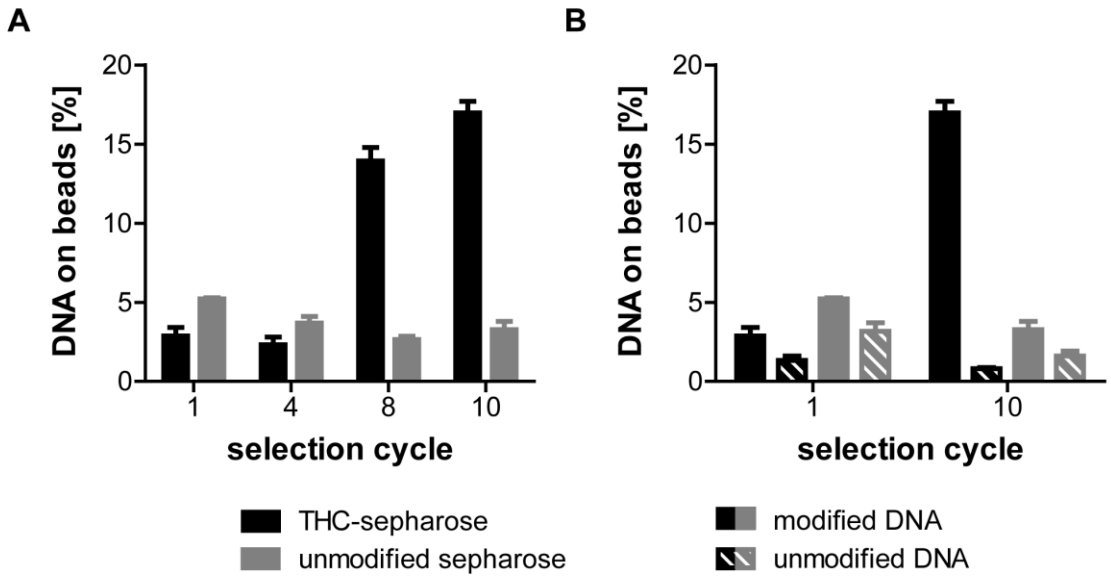


Figure 22: Interaction analysis of THC-selection performed on THC-sepharose with benzyl-modified DNA
^{32}P-labelled DNA with (full bars) or without benzyl-modification (alkyne-modified, striped bars) was incubated with THC-sepharose (black bars) or unmodified sepharose (grey bars). After two washing steps, the amount of radioactivity on the sepharose was quantified. (A) Interaction analysis of DNA from selection cycles 1, 4, 8, and 10. (B) Interaction analysis of benzyl-modified and non-modified (EdU-containing) DNA from selection cycles 1 and 10.

To test if the benzyl-modification is necessary for the interaction with THC-sepharose, DNA from the first and last selection cycles (1 and 10) was tested both as modified (clicked with benzyl-azide) and unmodified (unclicked, containing EdU) DNA on THC- and unmodified sepharose. Contrary to modified DNA, the unmodified DNA from selection cycle 1 and 10 interacted with neither THC- nor unmodified sepharose (**Figure 22B**). Therefore, the benzyl-modification is required for the interaction of the DNA of selection cycle 10 with THC-sepharose.

The DNA from selection cycle 10 was sequenced by Sanger sequencing to identify potential THC-binding sequences. **Table 6** shows the identified sequences that were tested, the number of modified nucleotides (EdUs), and their frequency both in the initial Sanger sequencing as well as in a later-on performed NGS analysis (**Table A 1**). **Figure A 1A** depicts the frequency of the identified sequences in all selection cycles analysed by NGS. The amount of modifications ranged from 5 EdUs in B13 to 9 in B12. The most abundant sequence in both Sanger and NGS was B27, the least abundant B10. The order of the sequences in between varied depending on the sequencing method used.

Table 6: Sequences from the THC-selection with benzyl-modified DNA

Seq.	consensus sequence of the random region	EdUs	Sanger [%]	NGS [%]
B10	AXACXGGACCGCXACACCCCGCACCACCXCGCCAAXXGGXGA	7	5.4	0.5
B11	GAXGCCGGACGXXGGAACGGGCGXGXCAXCAXAGGXACAGG	8	10.8	5.1
B12	CAGCGGXCCXAGCGCGGXACXACGCAXCCCCXAXCXCCCCCX	9	13.5	4.1
B13	GXGGGGCACAXCAGCGACCGXAXCCCCCAGACAGACXCAGGC	5	8.1	1.7
B14	AACCCXGACXGGGAGXGAXGCCXACCGXAACACCXCCCGCAC	7	24.3	1.6
B27	CXACAAXCGXGCGAXCCCCXAXCXCCAGACCXACGGGAACAC	8	35.1	10.6

Seq. = sequence. X = EdU. Sanger = frequency in Sanger sequencing of selection cycle 10. NGS = frequency in NGS of selection cycle 10.

All sequences described in **Table 6** were tested for their interaction with THC- and unmodified sepharose. All sequences but the least abundant sequence, B10, interacted with THC-sepharose to varying degrees (**Figure 23**). The strongest interaction was apparent for B14 and B27, which were also the most abundant sequences in the final selection cycle according to NGS. B11 and B12 had the weakest interaction with THC-sepharose, but the interaction was still far more pronounced than that of B10, the DNA from selection cycle 1, or the interaction of any sequence with unmodified sepharose.

The selection with benzyl-modified DNA thus led to the identification of five THC-sepharose-binding sequences. Sequence B10 did not show binding to THC-modified sepharose and was used as a non-binding control sequence for all subsequent assays.

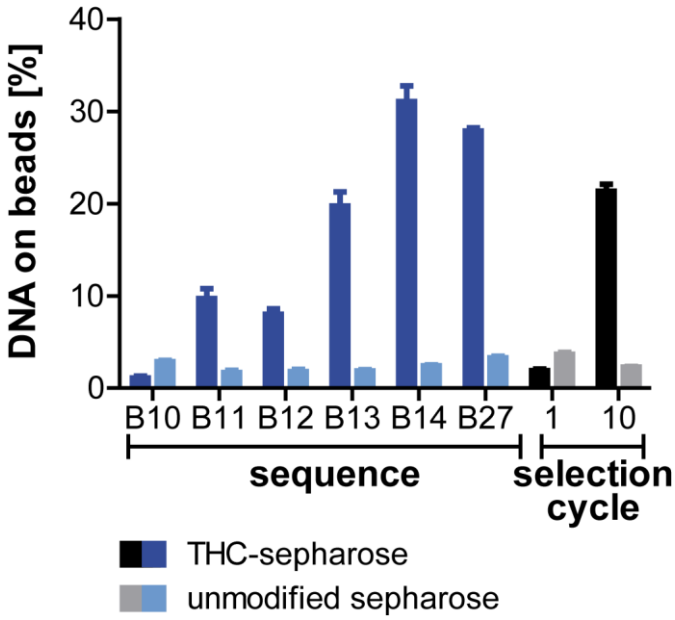


Figure 23: Interaction analysis of the sequences from the selection with benzyl-modified DNA
^{32}P-labelled DNA of the single sequences or from the indicated selection cycle was incubated with THC-sepharose (black or dark blue bars) or unmodified sepharose (grey or light blue bars). After two washing steps, the amount of radioactivity on the sepharose was quantified. All DNA was benzyl-modified.

To analyse the specificity of the benzyl-modified THC-sepharose-binding sequences, their interaction with THCA-functionalised tosylated magnetic beads was tested. In comparison to the THC-sepharose, the hydroxyl-group of THC is unconjugated on the magnetic beads. Like with the THCA on magnetic amine beads used for the initial selections described in this thesis (section 3.1.1), THCA is immobilised via an amide bond with its carboxyl-group (**Figure 24**). The success of the coupling was verified by Malte Rosenthal, who also performed the coupling reaction.[a]

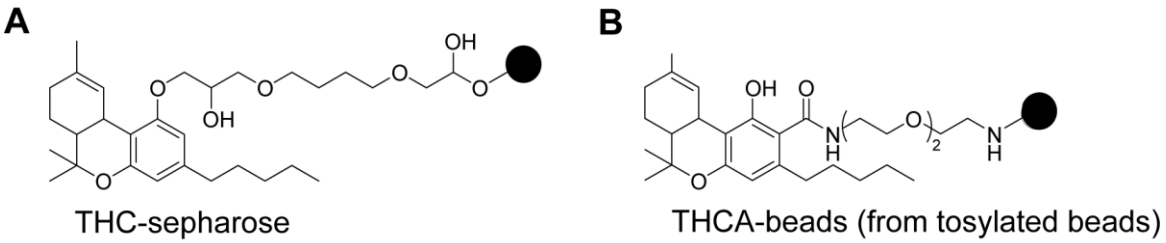


Figure 24: Chemical structure of THC-sepharose and THCA-beads.
(A) Structure of THC-sepharose that has been coupled according to **Figure 19**. (B) Structure of THCA-beads, for which THCA (**Figure 1B**) was coupled to tosylated magnetic beads. The coupling and detection of THCA on the beads according to section 6.2.2.4 was performed by Malte Rosenthal.

[a] Malte Rosenthal, PhD thesis in preparation, Bonn, 2018/2019

The interaction analysis depicted in **Figure 25** shows that only sequence B11 interacted with the THCA-beads. While the control sequence B10 did not bind to the THCA-beads, the DNA from selection cycle 1 recognised THCA with a similar ratio modified:unmodified beads as the DNA from cycle 10. B12 and B27 showed better binding to the THCA- than to the methylamine-blocked tosylated beads, but barely better than the control B10. The percentage of binding of the remaining sequences was reduced in comparison to those on THC-sepharose (**Figure 23**). None of the sequences interacted with the methylamine-blocked tosylated beads.

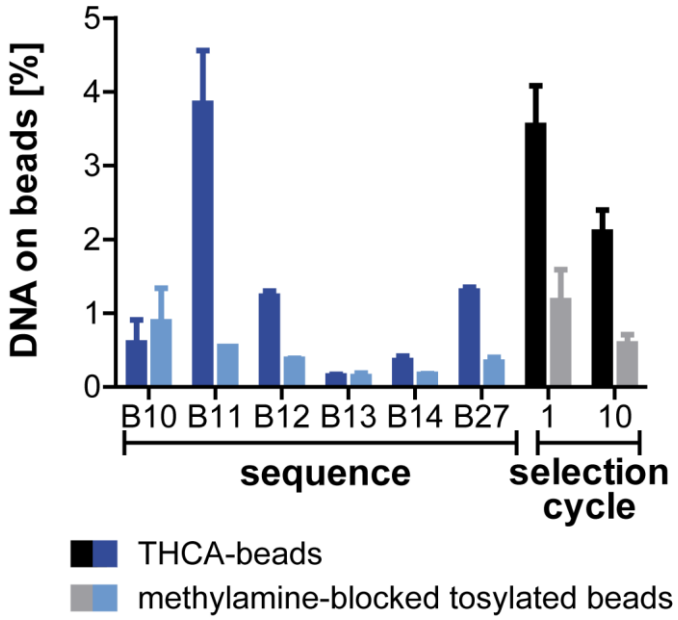


Figure 25: Interaction analysis of the benzyl-modified sequences with THCA-beads
^{32}P-labelled DNA was incubated with magnetic THCA-beads (black or dark blue bars) or methylamine-blocked tosylated beads (grey or light blue bars). After two washing steps, the amount of radioactivity on the beads was quantified. Tested were B10, B11, B12, B13, B14, and B27, as well as DNA from selection cycle 1 and 10. All DNA was benzyl-modified.

3.2.3 Characterisation and sequencing of the CF_3-modified click-SELEX

The DNA from selection cycle 1, 4, 8, and 10 of the selection with CF_3-modified DNA was incubated with THC- and unmodified sepharose. DNA from both selection cycle 8 and 10 showed an increase in binding to THC-sepharose and a decrease in binding to unmodified sepharose, indicating an enrichment of THC-binding sequences in the later selection cycles (**Figure 26A**).

The DNA from selection cycles 1 and 10 was also tested without modification (unclicked, containing EdU). This should elucidate the necessity of the modification for the increase in interaction with THC-sepharose, which had been observed for CF_3-modified DNA from selection cycle 10. Without modification, the

interaction of the DNA from selection cycle 10 with both types of sepharose was even lower than the interaction of the modified DNA with the unmodified sepharose. No difference between unmodified DNA from selection cycle 1 and 10 was detectable (**Figure 26B**). Therefore, the CF_3-modification is necessary for the interaction with THC-sepharose.

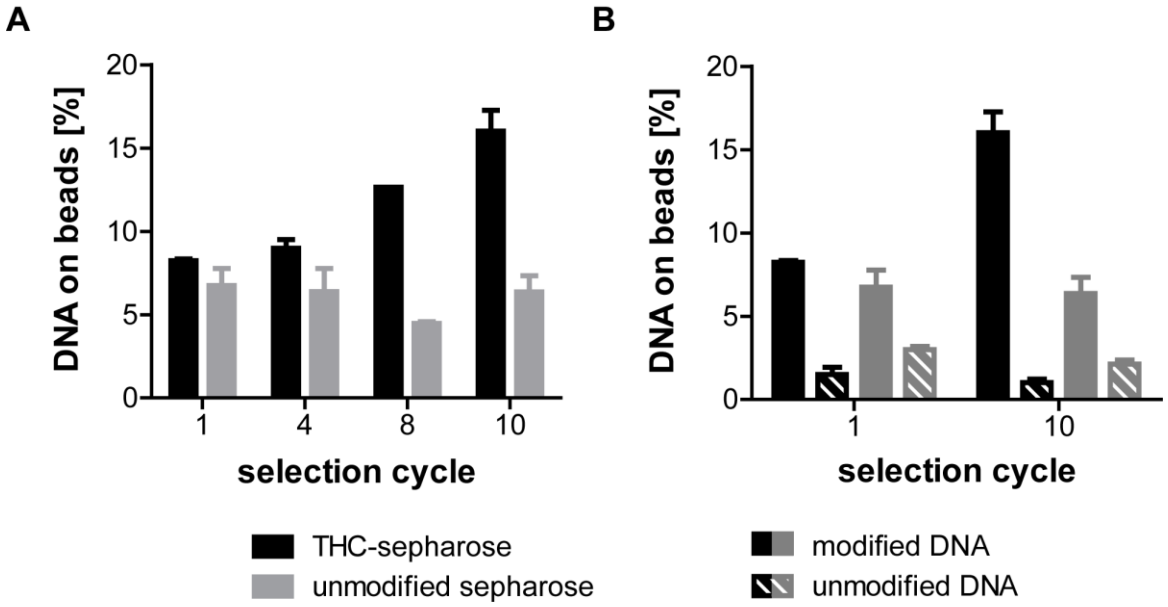


Figure 26: Interaction analysis of THC-selection performed on THC-sepharose with CF_3-modified DNA
^{32}P-labelled DNA with (full bars) or without CF_3-modification (alkyne-modified, striped bars) was incubated with THC-sepharose (black bars) or unmodified sepharose (grey bars). After two washing steps, the amount of radioactivity on the sepharose was quantified. (A) Interaction analysis of DNA from selection cycles 1, 4, 8, and 10. (B) Interaction analysis of CF_3-modified and non-modified (EdU-containing) DNA from selection cycles 1 and 10.

In order to identify the THC-binding sequences, the DNA from selection cycle 10 was sequenced by both Sanger and NGS. Three different sequences were identified as enriched: C15, C30, and C47 (**Table 7** and **Table A 2**). **Figure A 1B** depicts the frequency of the identified sequences in all selection cycles analysed by NGS. They contained between 4 and 9 modifications. Even though the frequency of the identified sequences was far lower in NGS than detected by Sanger sequencing, the order was identical with both methods: C15 was the least abundant sequence, followed by C30 and C47.

All three sequences (**Table 7**) were radioactively labelled and tested for their interaction with THC- and unmodified sepharose. All of them interacted with THC-sepharose to a similar degree as DNA from selection cycle 10 (**Figure 27**).

Table 7: Sequences from the THC-selection with CF_3-modified DNA

Seq.	consensus sequence of the random region	EdUs	Sanger [%]	NGS [%]
C15	XGGCGCGXCCAGAGGCXXAXAGGGGGGGAGCCXACGXXGAAX	9	6.3	0.1
C30	CCAGCCGCGCGGAAGACAAXACACGXXGGGCCACXAAGGAAG	4	9.4	0.6
C47	CGAXAAXACACGXXCGGCCCCXAAAGCCGGXCGGCCCXXGCA	8	43.8	11.4

Seq. = sequence. X = EdU. Sanger = frequency in Sanger sequencing of selection cycle 10. NGS = frequency in NGS of selection cycle 10.

The selection with CF_3-modified DNA enabled the identification of three THC-sepharose-binding sequences. No non-binding sequence was identified. Therefore, DNA from selection cycle 1 was used as a negative control in the subsequent assays.

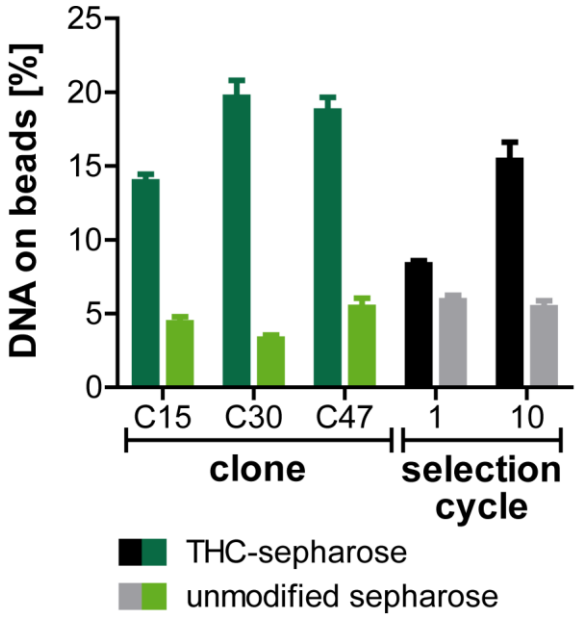


Figure 27: Interaction analysis of the sequences from the selection with CF_3-modified DNA
^{32}P-labelled DNA was incubated with THC-sepharose (black or dark green bars) or unmodified sepharose (grey or light green bars). After two washing steps, the amount of radioactivity on the sepharose was quantified. Tested were C15, C30, and C47 as well as DNA from selection cycles 1 and 10. All DNA was CF_3-modified.

3.2.4 Verification of the NGS approach

The use of NGS for aptamer identification is a relatively new technique (section 1.2.2) and its results can differ vastly from those obtained by Sanger sequencing as is obvious from the results presented in **Table 6**. To verify our analysis, we evaluated different sample preparation techniques and determined average mutation rates by analysing a single sequence with millions of reads. The results are presented in the subsequent sections. The influence of the sample preparation on mutation rates is described in section 3.2.4.1. Repetitive

sequences are analysed in section 3.2.4.2 and the influence of the exclusion of shortened sequences is portrayed in section 3.2.4.3. Error rates after the omission of shortened sequences are calculated (section 3.2.4.4) and the effect of the exclusion of shortened sequences on SELEX samples is displayed in section 3.2.4.5.

3.2.4.1 Influence of the sample preparation on mutation rates

PCRs during sample preparation ('index-PCR') and sequencing have been indicated as a reason for increased mutation rates that result in sequencing errors.[165,169,209] The NGS samples for the analysis of the click-selections presented above (section 3.2.2 and 3.2.3) were prepared using Pwo polymerase. To assess the impact of the index-PCR (section 6.2.6.6), the sequence of C12, the clickmer identified for GFP[205], was prepared for NGS-analysis using Pwo (T_Pwo) and Taq polymerase (T_Taq) as well as without index-PCR (T_w/o). All samples were amplified from a template containing thymidine instead of EdU. A forth sample was prepared from an EdU-containing template with Pwo polymerase (EdU_Pwo) (**Table 8**). The NGS-analysis showed that EdU_Pwo led to the highest frequency of mutated sequences and the highest error rate (average mutation per base). From the samples with thymidine-containing templates, T_Pwo obtained the lowest frequency of mutated sequences, but the highest error rate, while T_Taq had the highest frequency of mutated sequences and T_w/o the lowest error rate (**Table 8**).

Table 8: Frequency of mutations in differently prepared C12-samples

Sample name	EdU/T in template	Polymerase for index-PCR	Mutated seq. [%]	Error rate [%] (mean ± SD)	Analysed seq.
T_Pwo	T	Pwo	12.23	3.04 ± 1.87	1,119,179
T_Taq	T	Taq	12.47	2.85 ± 1.75	3,416,163
T_w/o	T	none[1]	12.43	2.55 ± 1.83	1,872,807
EdU_Pwo	EdU[2]	Pwo	32.02	6.15 ± 4.01	4,593,685

Seq. = sequences. SD = standard deviation.

[1] DNA was solid-phase synthesised including the indices

[2] due to solid-phase synthesis of template, 20% of EdUs are oxidized to KdU[210,211]

We next analysed each position of the sequence. All samples showed an increase in mutation frequency over the length of the sequence, resulting in a roughly ten times higher mutation rate at the end than at the start of the analysed sequence (**Figure 28**). Overall, EdU_Pwo had increased mutation frequencies in comparison with the samples prepared from thymidine-containing templates.

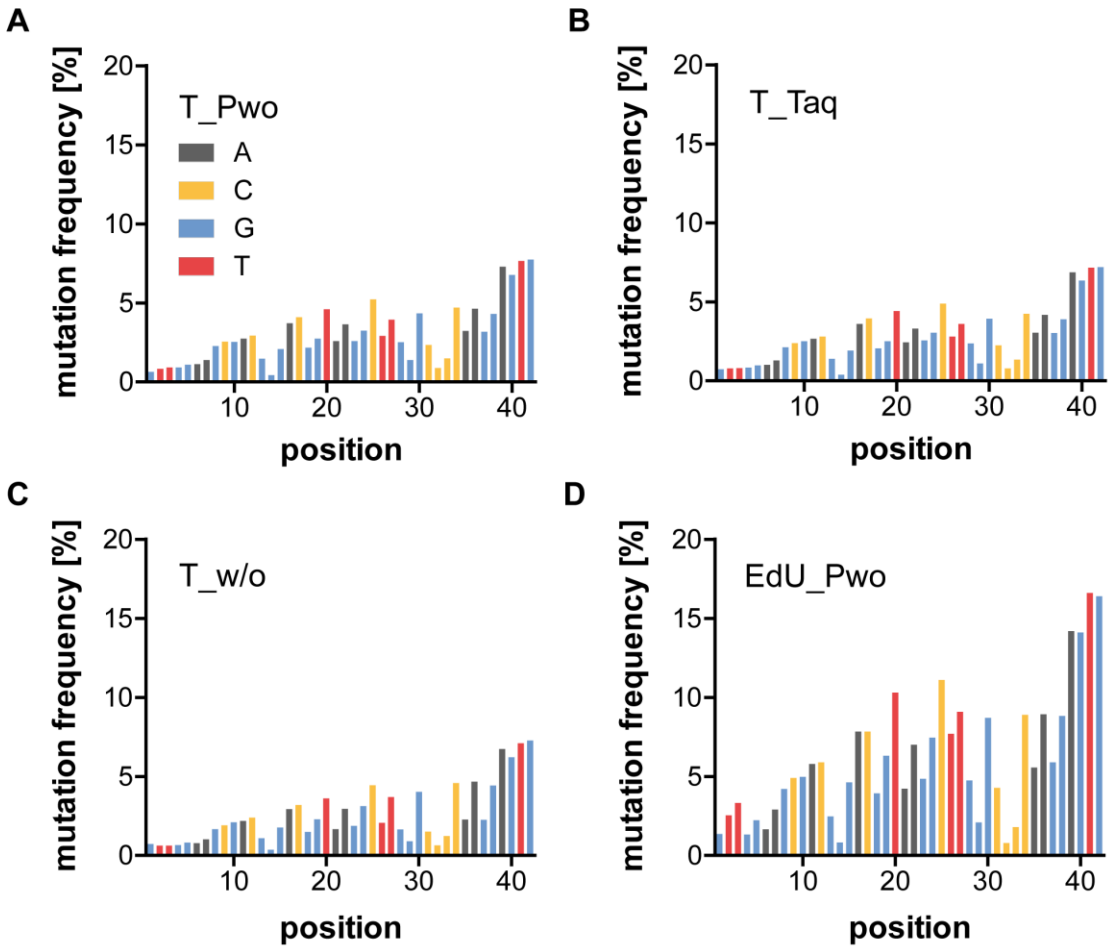


Figure 28: Mutation frequency in differently prepared C12-samples
Mutation frequency at each position of the analysed sequence. Colours indicate the original nucleotide at the respective position. The colour legend in (A) is valid for all four graphs. The analysed samples were prepared using (A) a thymidine-containing template and Pwo polymerase, (B) a thymidine-containing template and Taq polymerase, (C) a thymidine-containing template without index-PCR, and (D) an EdU-containing template and Pwo polymerase.

To further compare the differently prepared samples, the average mutation frequency of the single nucleotides and of the nucleotides, into which the original nucleotide mutated, were analysed. EdU_Pwo had a significantly increased mutation frequency of the mutated nucleotides (**Figure 29A**). As expected, the preparation without index-PCR (T_w/o) showed the lowest mutation rate. However, the slight differences between the three samples with the thymidine-containing template were not significant. The same trends could be seen for the mutation into a specific nucleotide, but the increase for EdU_Pwo was not significant for this analysis (**Figure 29B**). It was discovered that the mutation frequencies into a specific nucleotide reflected the nucleotide distribution of the original sample (**Figure 29C**). To identify the reason for this correlation, we studied samples with repetitive sequences.

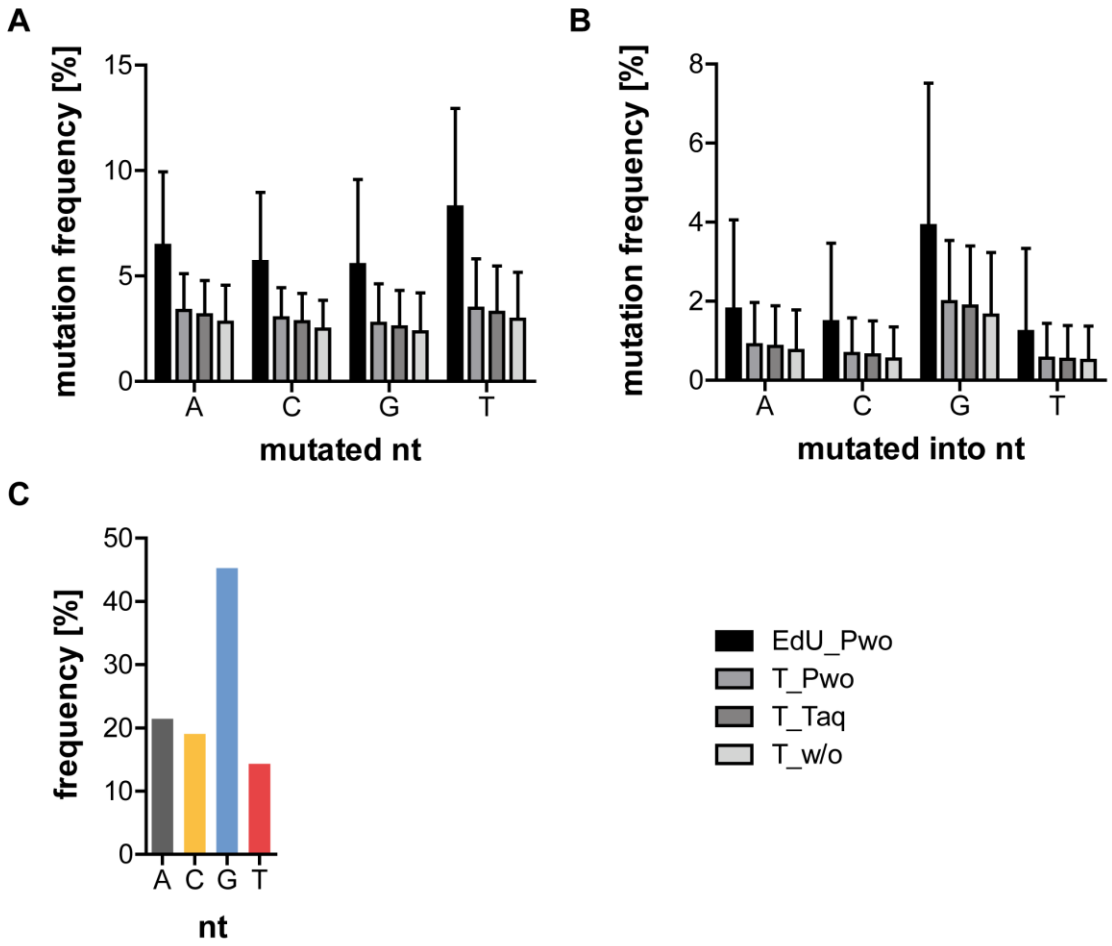


Figure 29: Average mutation frequencies of differently prepared C12-samples and nucleotide distribution of C12
(A) Average mutation frequencies of the original nucleotides. EdU_Pwo has a statistically increased mutation frequency in comparison to the three other data sets: EdU_Pwo vs. T_Pwo, T_Taq, and T_w/o $p = 0.0286$. Mann-Whitney tests, two-tailed. Preliminary Kruskal-Wallis-test $p = 0.0132$. $n = 9$, 8, 19, and 6 for A, C, G, and T, respectively. (B) Average frequencies, with which the original nucleotide mutated into the denoted one. The differences are non-significant: Kruskal-Wallis-test. $n = 33$, 34, 23, and 36 for A, C, G, and T, respectively. (C) Nucleotide distribution in non-mutated C12.

3.2.4.2 Analysis of repetitive sequences

We evaluated repetitive sequences with one or four identical nucleotides in a row: FT2_GATC and FT2_G4A4T4C4. While the frequency of mutated sequences was slightly higher for FT2_G4A4T4C4 than for FT2_GATC, the latter had a higher error rate (**Table 9**). In comparison to the C12-samples, both repetitive samples showed lower frequencies of mutated sequences and lower error rates (**Table 8**).

Table 9: Frequency of mutations in repetitive sequences

Sample name	Mutated sequences [%]	Error rate [%] (mean ± SD)	Number of analysed sequences
FT2_GATC	8.44	1.63 ± 0.82	10,059,713
FT2_GATC_II	6.62	1.48 ± 0.78	2,332,475
FT2_G4A4T4C4	10.87	0.83 ± 0.83	8,235,942
FT2_G4A4T4C4_II	10.15	0.83 ± 0.83	7,288,615
FT2_G2A2T2C2	11.33	1.54 ± 0.96	2,301,791
FT2_G3A3T3C3	11.66	1.46 ± 1.08	6,265,796
FT2_TGCA	10.94	2.18 ± 1.16	7,441,266
D3_TGCA	7.27	1.09 ± 0.56	429,868
FT2_T4G4C4A4	10.79	0.92 ± 0.90	1,956,098
D3_T4G4C4A4	10.90	0.87 ± 0.97	5,930,886

Regarding their mutation frequencies at each position, the trend was identical to what had been observed before: The frequency increased from the start to the end of the sequence by a factor of about ten (**Figure 30A** and **B**). The increase was steady for FT2_GATC. In FT2_G4A4T4C4, only the last nucleotide of each block made up of four identical nucleotides had a high and rising mutation frequency. The first three nucleotides showed a low mutation frequency. This can explain the discrepancy observed in **Table 9**. As only every fourth nucleotide had a high mutation frequency, the average error rate was lower than for FT2_GATC. The high mutation frequency of the fourth nucleotide nonetheless led to a relatively high frequency of mutated sequences.

For both repetitive sequences, clear preferences for the conversion from one nucleotide into another specific nucleotide were observed (**Figure 30C** and **D**). These preferences are depicted in **Figure 31**. The conversions occurred primarily from one nucleotide into the subsequent one.

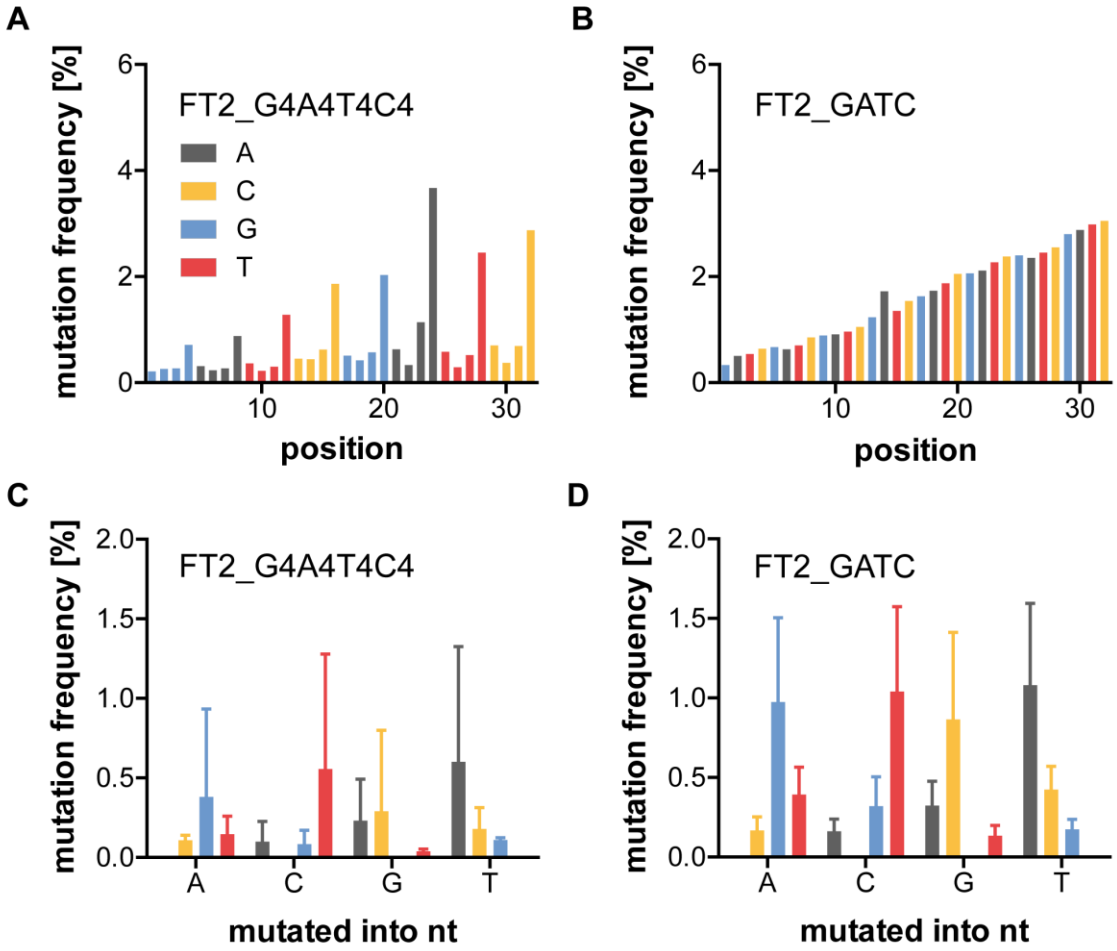


Figure 30: Mutation frequencies of repetitive sequences
The colour legend in (A) denotes the colour of the original nucleotide for all four subpanels. Mutation frequency of FT2_G4A4T4C4 (A) and FT2_GATC (B) at each position of the analysed sequence as well as average mutation frequency, with which the mutation occurred into the denoted nucleotide for FT2_G4A4T4C4 (C) and FT2_GATC (D). The conversions for (C) are non-significant (Kruskal-Wallis test, n=8). Those for (D) are significant with a $p \leq 0.0174$ (t-tests, two-tailed, preliminary one-way ANOVA: p=<0.0001, 0.0002, <0.0001, and 0.0007 for mutated into T, A, C, and G, respectively, n=8).

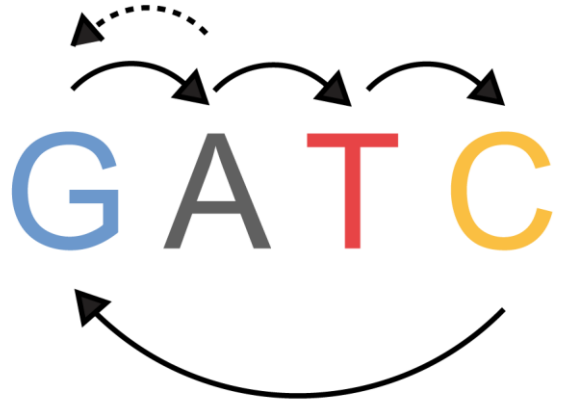


Figure 31: Main conversions of repetitive GATC-samples
The most frequent mutations of one nucleotide into another of the FT2_GATC and FT2_G4A4T4C4-samples. The dotted arrow is only valid for the FT2_G4T4A4C4-sample.

To evaluate this hypothesis, additional repetitive sequences were analysed by NGS. Sequences containing different primer binding sites and different nucleotide orders (D3_TGCA, D3_T4G4C4A4, FT2_TGCA, FT2_T4G4C4A4) as well as some with two or three identical nucleotides in a row (FT2_G2A2T2C2, FT2_G3A3T3C3) were chosen to identify potential biases (**Table 9**). All analysed sequences, including the C12-samples, showed a bias towards mutation to the subsequent nucleotide (**Table 10**): A completely random mutation would result in mutations to the subsequent nucleotide around 33.3% due to the three available options. All samples tested in this study showed mutation frequencies to the subsequent nucleotide ≥ 60%.

Table 10: Frequency of mutation to the subsequent nucleotide

Sample name	Mutation to subsequent nt [%] (mean ± SD)
EdU_Pwo	76.0 ± 14.80
T_Pwo	73.2 ± 17.14
T_Taq	74.1 ± 18.34
T_w/o	76.5 ± 15.14
FT2_GATC	64.3 ± 3.85
FT2_GATC_II	65.0 ± 4.49
FT2_TGCA	60.2 ± 7.98
D3_TGCA	71.7 ± 11.45
FT2_G2A2T2C2	77.2 ± 8.65
FT2_G3A3T3C3	80.1 ± 8.07
FT2_G4A4T4C4	83.8 ± 9.63
FT2_G4A4T4C4_II	83.8 ± 9.63
FT2_T4G4C4A4	86.0 ± 6.57
D3_T4G4C4A4	86.6 ± 6.81

Independent of the nucleotide order and the primer binding site, the mutation frequency to the subsequent nucleotide increased significantly with the number of identical nucleotides in a row from about 65% for a single nucleotide to about 85% for four identical nucleotides in a row (**Figure 32**).

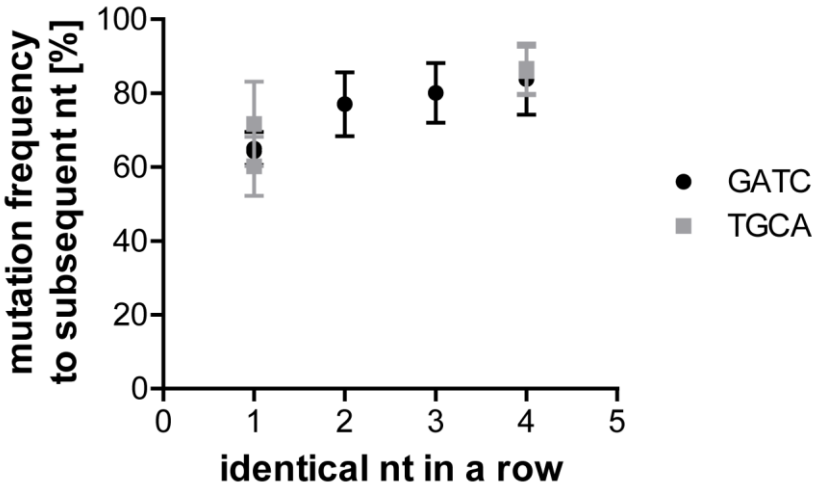


Figure 32: Mutation frequency to the subsequent nucleotide
Depicted is the average frequency, with which a nucleotide mutated to the subsequent nucleotide for all analysed, repetitive sequences. The difference between one consecutive identical nucleotide and four is significant with p=0.0294 (Mann-Whitney test, two-tailed, n=31 and 7 for 1 and 4 nucleotides in a row, respectively).

3.2.4.3 Omission of shortened sequences

Phasing effects (section 1.2.2) are a possible explanation for the increase in mutation frequency over the length of the analysed sequence as well as for the high mutation rates to the subsequent nucleotide. Analysis of the most abundant sequences of each sample showed that pre-phased sequences are shortened (**Table A 3** to **Table A 16**). Although shortening might also occur through deletions, we decided to omit all shortened sequences and reanalyse the samples under these conditions.

Omission of shortened sequences led to a decrease of the number of analysed sequences by about 5% and an increase of the frequency of non-mutated sequences by the same margin. At the same time, the error rate dropped by about 79%. Due to the presence of EdU in EdU_Pwo and the resulting high error rates and mutation frequencies in comparison with all other analysed sequences, EdU_Pwo was excluded for the calculation of all averages (**Table 11**).

Table 11: Change upon omission of shortened sequences

Sample name	Δ analysed seq. [%]	Δ non-mutated seq. [%]	Δ error rate [%]	Mutation to subsequent nt: deviation from 33.3%	
				w/	w/o shortened sequences
EdU_Pwo	-15.76	18.70	-86.83	42.7	3.7
T_Pwo	-7.75	8.41	-95.07	39.9	11.4
T_Taq	-7.17	7.71	-93.33	40.8	12.9
T_w/o	-7.40	7.99	-92.94	43.2	2.1
FT2_GATC	-2.94	3.04	-87.73	31.0	0.5
FT2_GATC_II	-2.80	2.88	-72.97	31.7	0.1
FT2_G4A4T4C4	-3.16	3.26	-66.27	26.9	1.3
FT2_G4A4T4C4_II	-3.25	3.37	-69.88	38.4	6.3
FT2_G2A2T2C2	-3.90	4.06	-81.17	43.9	-0.1
FT2_G3A3T3C3	-4.56	4.78	-43.48	46.8	6.5
FT2_TGCA	-4.00	4.17	-88.99	50.5	2.0
D3_TGCA	-1.97	2.01	-83.49	50.5	-2.1
FT2_T4G4C4A4	-3.86	4.01	-72.83	52.7	13.5
D3_T4G4C4A4	-3.52	3.65	-70.11	53.3	6.5
average	**-5.15**	**5.57**	**-78.93**	**42.31**	**4.61**

Seq. = sequences.
EdU_Pwo was excluded for calculation of the averages.

Both the average mutation frequencies of each nucleotide of the C12-samples as well as the average frequency, with which a nucleotide mutates into the denoted one, were severely reduced for all samples. The mutation frequencies of EdU_Pwo were still significantly higher than those of the other samples (**Figure 33A** and **B**). Most importantly, the mutation frequencies into the denoted nucleotide no longer reflected the nucleotide composition of C12 (**Figure 33B** and **Figure 29C**). One representative of the mutation frequency at each position is depicted in **Figure 33C** and displays that the mutation frequency no longer increased over the length of the analysed sequences. This was the case for all analysed samples after omission of the shortened sequences (data not shown). For FT2_GATC and FT2_G4A4T4C4, specific mutation hotspots became apparent (**Figure 33C**). The mutation frequency to the subsequent nucleotide dropped to values around the expected 33.3%. In addition, it no longer depended on the number of identical nucleotides in a row (**Figure 33D** and **Table 11**).

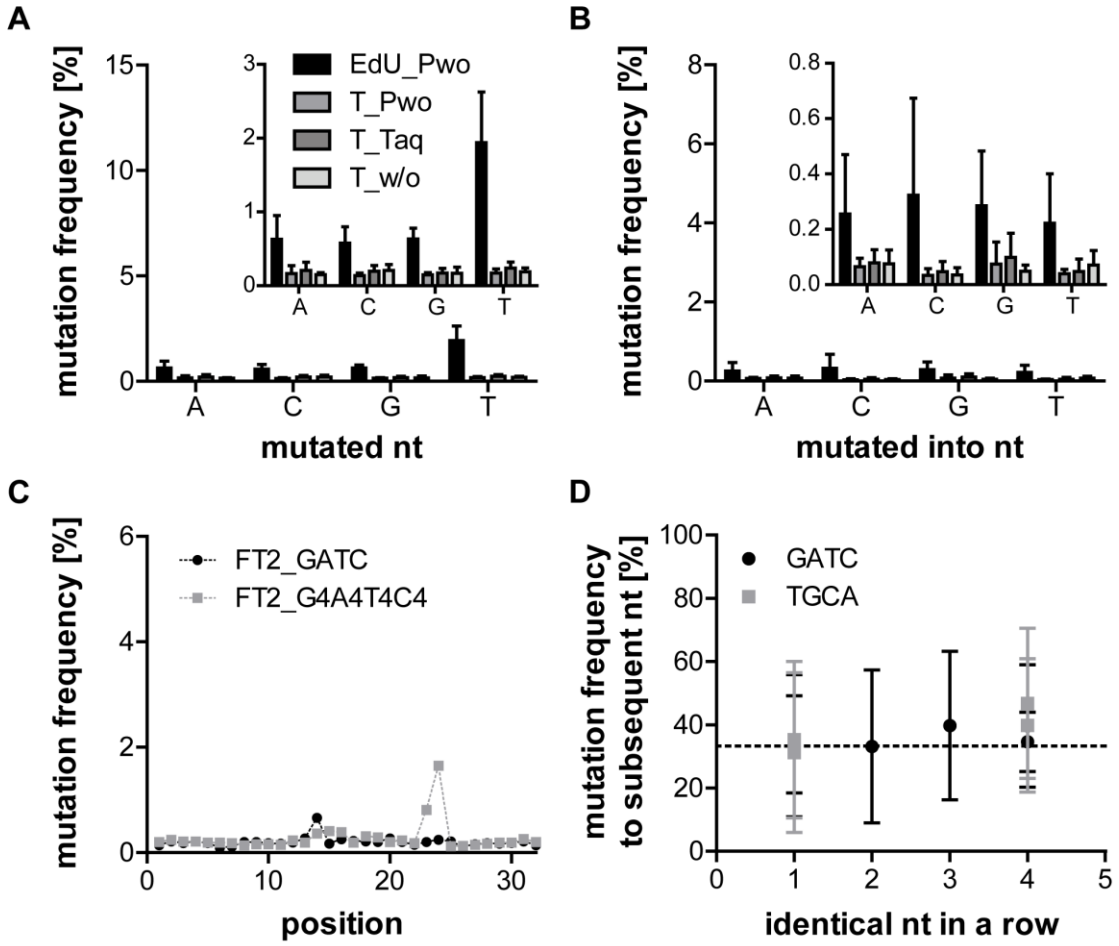


Figure 33: Mutation frequencies after omission of shortened sequences
Average mutation frequencies of the original nucleotides (A) and frequencies, with which the original nucleotide mutated into the denoted one (B) for the four different C12-samples. The legend in (A) is valid for both (A) and (B). The larger graphs are scaled according to **Figure 29A** and **B** to simplify comparisons before and after omission of shortened sequences. The smaller zoom-ins are scaled to allow a detailed view of the results. For (A), EdU_Pwo vs. T_Taq and T_w/o and T_Pwo vs. T_Taq are significant with a p = 0.0067 each. (Mann-Whitney tests, two-tailed, preliminary Kruskal-Wallis test: p = 0.0067, n = 9, 8, 19, and 6 for A, C, G, and T, respectively.) For (B), EdU_Pwo vs. T_Pwo, T_Taq, and T_w/o is significant with a p = 0.0286. (Mann-Whitney tests, two-tailed, preliminary Kruskal-Wallis test: p = 0.026, n = 33, 34, 23, and 36 for A, C, G, and T, respectively.) (C) Mutation frequency for each position of the analysed sequences of FT2_GATC and FT2_G4A4T4C4. The graph is scaled according to **Figure 30A** and **B** to simplify comparisons before and after omission of shortened sequences. (D) Mutation frequency of one nucleotide into the subsequent nucleotide for all repetitive sequences after omission of shortened sequences. The expected average mutation frequency of 33.3% is indicated with a dotted line. The difference between the samples containing one and four identical nucleotides in a row is no longer significant. (Mann-Whitney test, two-tailed, n = 31 and 7 for 1 and 4 nucleotides in a row, respectively.)

All of this indicated that the omission of shortened sequences managed to exclude the majority of the pre-phased sequences without loss of the majority of sequences as only about 5% of all analysed sequences were omitted (**Table 11**).

3.2.4.4 Error analysis after omission of shortened sequences

Having thus excluded pre-phased sequences, we analysed the 'real' error rates in samples sequenced by NGS. EdU_Pwo was excluded due to its high mutation rates. After exclusion, the average percentage of mutated sequences of the remaining sequences was 6.4 ± 1.24% and the average error rate 0.24 ± 0.06% per base. The C12-samples (apart from EdU_Pwo) had slightly lower percentages of mutated sequences and error rates compared to the majority of repetitive sequences.

Table 12: Frequency of mutations if shortened sequences are omitted

Sample name	Mutated seq. [%]	Error rate [%] (mean ± SD)	Mutation to subsequent nt [%] (mean ± SD)	Analysed seq.
EdU_Pwo	19.31	0.81 ± 0.57	37.0 ± 16.87	3,869,868
T_Pwo	4.85	0.15 ± 0.06	44.7 ± 18.29	1,032,398
T_Taq	5.72	0.19 ± 0.08	46.2 ± 21.52	3,171,344
T_w/o	5.43	0.18 ± 0.07	35.4 ± 15.04	1,734,189
FT2_GATC	5.66	0.20 ± 0.09	33.8 ± 15.38	9,763,653
FT2_GATC_II	3.93	0.14 ± 0.05	33.4 ± 22.45	2,267,079
FT2_G4A4T4C4	7.96	0.28 ± 0.28	34.6 ± 9.37	7,975,576
FT2_G4A4T4C4_II	7.12	0.25 ± 0.23	39.6 ± 19.39	7,051,464
FT2_G2A2T2C2	7.73	0.29 ± 0.16	33.2 ± 24.18	2,211,912
FT2_G3A3T3C3	7.44	0.26 ± 0.14	39.8 ± 23.50	5,979,814
FT2_TGCA	7.23	0.24 ± 0.16	35.3 ± 24.73	7,143,566
D3_TGCA	5.41	0.18 ± 0.11	31.2 ± 25.28	421,388
FT2_T4G4C4A4	7.21	0.25 ± 0.17	46.8 ± 23.74	1,880,590
D3_T4G4C4A4	7.65	0.26 ± 0.22	39.8 ± 21.08	5,722,279

Seq. = sequences.

The conversion from each nucleotide to a specific, different nucleotide was analysed for each of the samples (**Table 13**). Again, EdU_Pwo was excluded from the calculation of the averages due to its high mutation rates. As expected, EdU_Pwo showed very high mutation rates of thymidine (in average 0.65%), which was replaced by EdU in the sequencing template. In the other samples, C followed by G mutated most often (in average 0.093 and 0.083%, respectively) and A and T showed the lowest mutation rates (in average 0.067 and 0.06%). This was turned around when analysing the nucleotides into which mutations occurred most abundantly: T and A were most frequently mutated into (both in average 0.107%) and G and C least (in average 0.05 and 0.04%, respectively). Regarding specific conversions, the conversion from C to A was most likely to occur with an average of 0.13% and the conversions from A to C, C to G, G to C, and T to C least likely with an average of 0.04%. FT2_G2A2T2C2 and FT2_G3A3T3C3 had especially high mutation rates from C to A and G to T

(0.27% for both for FT2_G2A2T2C2 and 0.26 and 0.24%, respectively, for FT2_G3A3T3C3).

Table 13: Conversion between nucleotides after omission of shortened sequences

Original	A			C			G			T		
Mutated to [%]	**C**	**G**	**T**	**A**	**G**	**T**	**A**	**C**	**T**	**A**	**C**	**G**
EdU_Pwo	0.36	0.20	0.07	0.29	0.19	0.10	0.13	0.16	0.35	0.61	0.79	0.54
T_Pwo	0.03	0.11	0.03	0.05	0.04	0.05	0.08	0.02	0.04	0.03	0.06	0.08
T_Taq	0.05	0.14	0.02	0.11	0.04	0.05	0.08	0.03	0.06	0.03	0.10	0.10
T_w/o	0.02	0.05	0.07	0.11	0.04	0.06	0.06	0.03	0.07	0.07	0.06	0.06
FT2_GATC	0.06	0.04	0.13	0.08	0.04	0.10	0.08	0.04	0.08	0.09	0.04	0.05
FT2_GATC_II	0.02	0.02	0.10	0.07	0.02	0.07	0.05	0.02	0.09	0.06	0.02	0.02
FT2_G4A4T4C4	0.09	0.18	0.18	0.10	0.07	0.10	0.07	0.05	0.10	0.07	0.05	0.04
FT2_G4A4T4C4_II	0.16	0.05	0.09	0.06	0.04	0.03	0.13	0.05	0.01	0.16	0.06	0.16
FT2_G2A2T2C2	0.01	0.03	0.09	0.27	0.03	0.14	0.11	0.05	0.27	0.13	0.03	0.01
FT2_G3A3T3C3	0.01	0.03	0.09	0.26	0.05	0.13	0.10	0.04	0.24	0.10	0.03	0.01
FT2_TGCA	0.01	0.03	0.20	0.16	0.04	0.14	0.09	0.05	0.10	0.11	0.03	0.01
D3_TGCA	0.01	0.03	0.15	0.09	0.04	0.14	0.10	0.02	0.08	0.04	0.02	0.01
FT2_T4G4C4A4	0.02	0.02	0.08	0.18	0.03	0.20	0.17	0.04	0.14	0.07	0.03	0.03
D3_$T_4G_4C_4A_4$	0.02	0.03	0.07	0.16	0.04	0.19	0.11	0.02	0.16	0.20	0.02	0.02
average	0.04	0.06	0.10	0.13	0.04	0.11	0.10	0.04	0.11	0.09	0.04	0.05

The colouring indicating high (red) or low (green) values was performed separately for A) EdU_Pwo, B) the other samples, and C) the average. EdU_Pwo was excluded for calculation of the averages.

3.2.4.5 Effect of exclusion of shortened sequences on analysis of SELEX samples

To investigate the effect of the exclusion of the shortened sequences on the analysis of complex SELEX samples, we analysed several selection cycles of the selection that yielded C12.[152,205] Displayed in **Figure 34** are patterns, which are clustered using relative information entropy and regarded as sequence families. The NGS-frequency values given in **Table 6** and **Table 7** as well as those displayed in section 3.2.6 were obtained from such patterns that contain the specified sequence family. While slight differences in frequencies resulted from the omission of shortened sequences, the overall trends remained unchanged (**Figure 34**).

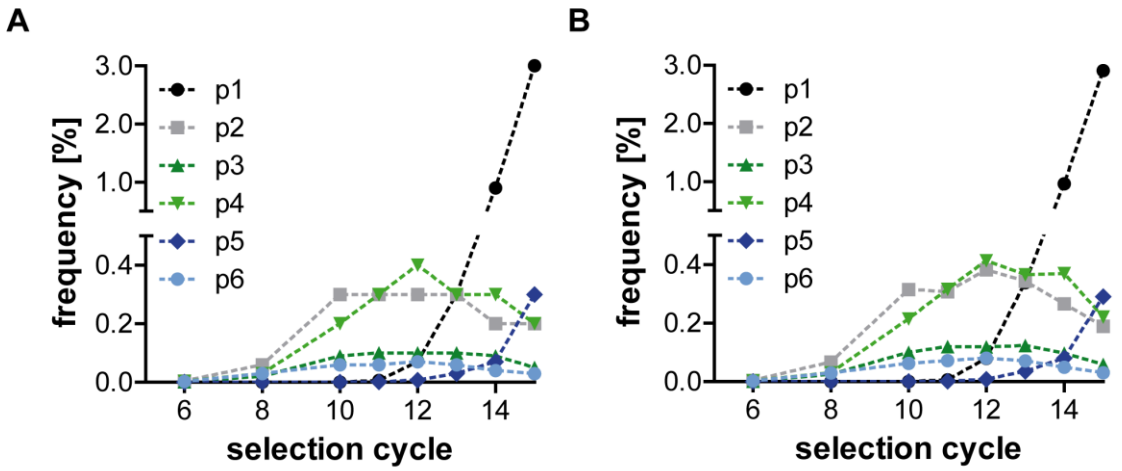


Figure 34: Frequency of different SELEX patterns before and after exclusion of shortened sequences
Frequency of six different SELEX patterns over the course of the GFP-SELEX. Selection cycles before cycle six were omitted as the frequencies were too low to be visible. (A) Frequencies before omission of the shortened sequences. (B) Frequencies after omission of the shortened sequences.[152]

3.2.5 Solubility determination of THC

Having thus validated that NGS can be used for the identification of potential clickmer candidates, we went on to determine the affinity of the identified sequences to THC in solution. As a multitude of assays at low THC-concentrations was not successful (data not shown), higher THC-concentrations had to be evaluated. More knowledge on THC-solubility in the used buffer was required. For this analysis, the area under the curve of the THC-peak as determined by high performance liquid chromatography (HPLC) was measured. Solubility was analysed by linear regression and determination of the coefficient of determination (R^2). An R^2-value of 1 represents perfect linearity and therefore solubility. A similar method has been published by Hazekamp *et al*.[212]

THC has a high hydrophobicity with an octanol/water partition coefficient (logP) of 5.65.[208,213] Accordingly, solubility in aqueous solutions is low with 2.45 µM THC in 0.15 M NaCl.[56] The selection buffer contains slightly less NaCl (137 mM), in which THC-solubility is expected to be about 2.5 µM.

THC-solubility in organic solvents is much better: THC is known to be soluble in methanol (MeOH) and up to 100 mg/ml in ethanol (EtOH).[214,b] As non-PEGylated (polyethylene glycol) nucleic acids are not soluble in pure organic solvents[215], low percentages of organic solvents were added to D-PBS and examined for their ability to enhance THC-solubility.[c,216] First of all, different concentrations of

[b] Personal communication with PD Dr. Andras Bilkei-Gorzo, Institute of Molecular Psychiatry, University of Bonn

DMSO, 100% of which can solubilise 60 mg/ml (190.7 mM) THC, were studied.[c,216]

The detergent Tween-20 was added to some samples to evaluate its effect on THC-solubility as it is part of the selection buffer used for the click-selections. Tween-20 is a non-ionic detergent with a critical micelle concentration (CMC) of 0.06 mM (0.0067% (v/v)).[217]

For the samples containing 100% DMSO, the area under the curve of the THC-peak increased linearly (**Figure 35A**). This indicated, as expected, that all THC was in solution and that this sample can serve as a control. In contrast, the signal for THC in 10% DMSO was not linear and was reduced in comparison to the control. The addition of 0.1% Tween led to similar values as measured for 100% DMSO, indicating that THC-solubility was recovered (**Figure 35A**).

As 0.1% Tween apparently enabled the use of low concentrations of organic solvents to solubilise THC, the best solvent had to be discovered. Acetonitrile (ACN), ethanol, and DMSO were tested. Both 5% EtOH and ACN led to good linearities (> 0.98), which were far higher than those reached with the same amount of DMSO (0.69) (**Figure 35B**). ACN was chosen for further tests. 100% DMSO and ACN both showed good linearities, but the signal intensities varied between the two solvents (**Figure 35B**).

[c] Information regarding solubility in organic solvents could be found on http://www.caymanchem.com (product insert for product 12068, extracted 24.04.2014):

> *[…] Solvents such as ethanol, methanol, DMSO, and dimethyl formamide (DMF) […] can be used. The solubility of Δ^9-THC in these solvents is approximately 35, 30, 60, and 50 mg/ml, respectively.*
>
> *Δ^9-THC is sparingly soluble in aqueous buffers. […] Δ^9-THC has a solubility of approximately 250 µg/ml [0.8 mM, annotation of the author] in a 1:3 solution of DMF:PBS (pH 7.2) […]. […]*

No references were given and the product insert has since been exchanged for one that does not contain any information on THC-solubility.

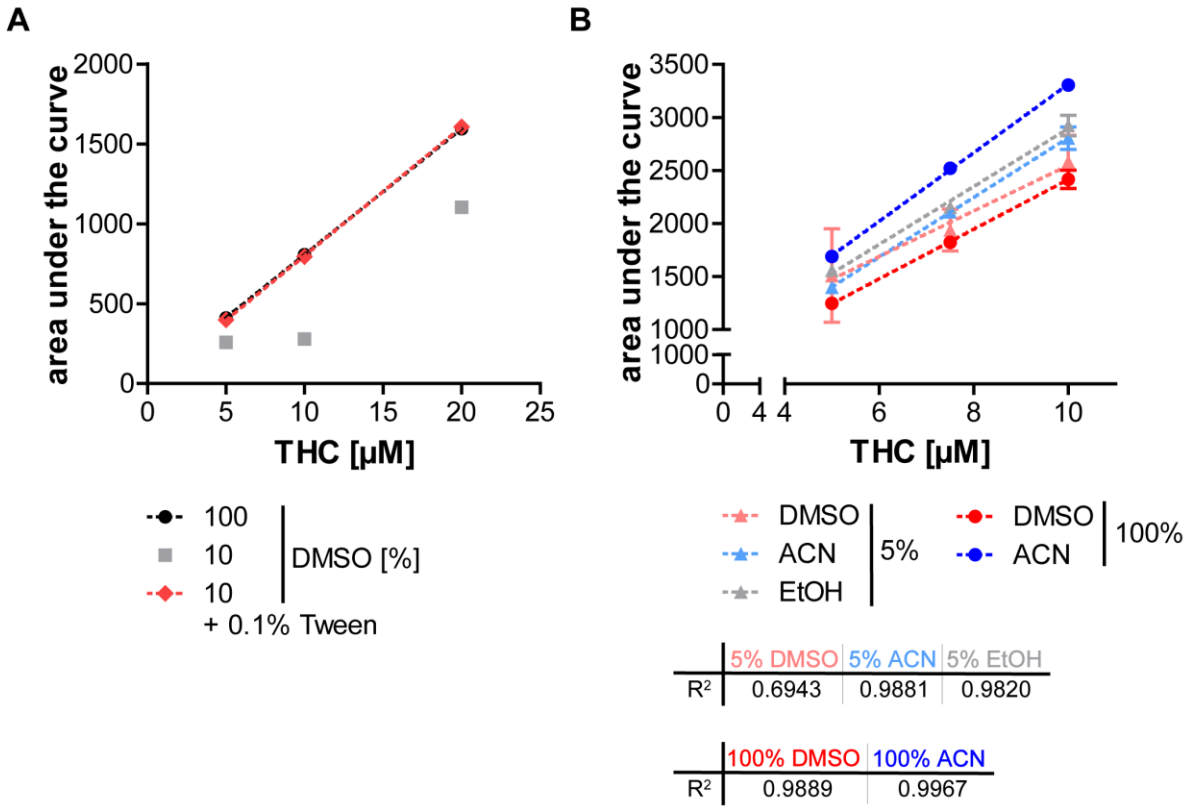


Figure 35: HPLC-measurement of THC-dilutions in DMSO, ACN, and EtOH
(A) Analysis of a single sample each of 5, 10, and 20 µM THC in 100% DMSO, 10% DMSO, or 10% DMSO with 0.1% Tween-20. As only a single sample was measured, no R^2-values were determined. (B) Analysis of THC-concentrations between 5 and 10 µM in 5% DMSO, ACN, and EtOH, all of which also contained 0.1% Tween. Measurements in 100% DMSO and ACN were performed as controls.

Having established that 5% ACN with 0.1% Tween can solubilise THC in concentrations up to 10 µM, higher concentrations were tested to discover the upper limit of THC-solubility under these conditions. As shown in **Figure 36**, the linearity of the regression curve for 5% ACN was satisfactory up to a concentration of 150 µM THC. A measurement of 250 µM THC suffered from very large error bars, indicating relatively unreproducible data points and strongly reduced the obtained linearity. Therefore, 150 µM THC in 5% ACN with 0.1% Tween were set as the maximum THC-concentration used in the subsequent assays.

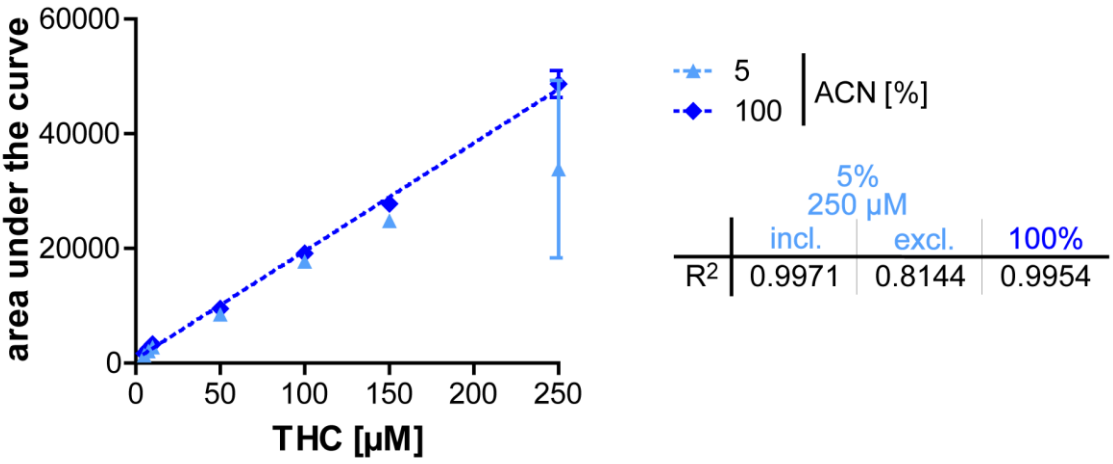


Figure 36: HPLC-measurement of THC-dilutions in 5 and 100% ACN
HPLC-analysis of THC-concentrations between 5 and 250 µM THC in 5% ACN with 0.1% Tween and in 100% ACN as control. The linearity (R^2) for 5% ACN has been calculated both including (incl.) and excluding (excl.) the value measured for 250 µM THC.

Before these conditions could be used for interaction analysis of the identified THC-sepharose-binding sequences, it had to be determined if the addition of 5% ACN (0.1% Tween is already part of the selection buffer) has an influence on the interaction of the sequences with the THC-sepharose. **Figure 37A** shows the results for most of the benzyl-modified sequences: The interaction of B14 and B27 with THC-sepharose was reduced in the presence of 5% ACN, but sufficient to categorise both as binding sequences. B13 did not interact with THC-sepharose in the presence of 5% ACN. Both B11 and B12 bound to THC-sepharose in the presence of 5% ACN (**Figure 37B**). The analysis of the CF_3-modified sequences can be found in **Figure 37C**. All sequences interacted with THC-sepharose in the presence of 5% ACN.

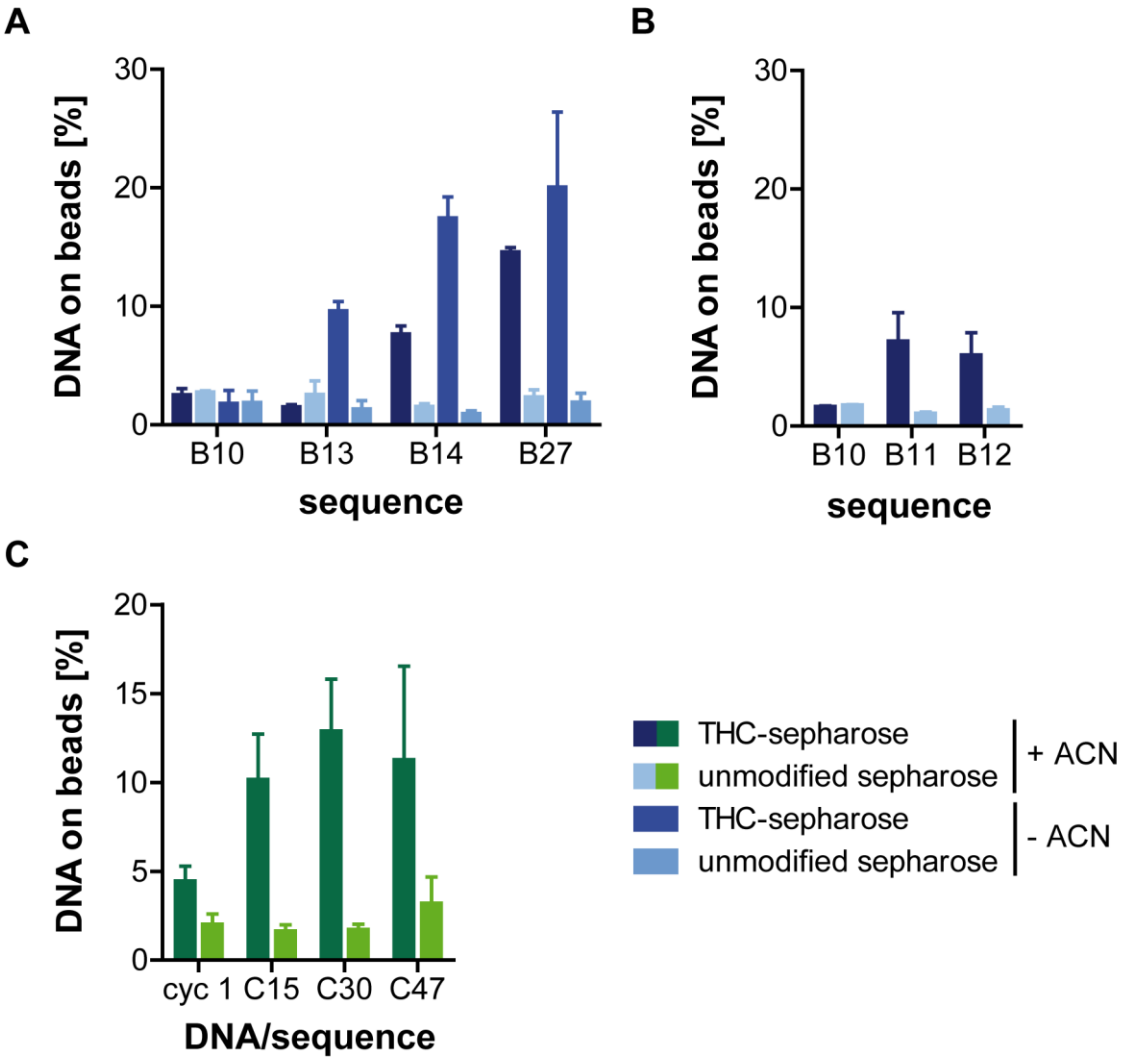


Figure 37: Interaction analysis of all sequences in the presence of 5% ACN
Benzyl-modified DNA is depicted in shades of blue, CF_3-modified DNA in shades of green. ^{32}P-labelled DNA was incubated with THC-sepharose (dark blues and dark green) or unmodified sepharose (light blues or light green). After two washing steps, the amount of radioactivity on the sepharose was quantified. The incubation took place in the presence and absence (A only) of 5% ACN. (A) Test of benzyl-modified sequences B10, B13, B14, and B27. (B) Test of benzyl-modified sequences B10, B11, and B12. (C) Test of CF_3-modified sequences. cyc 1 = CF_3-modified DNA from selection cycle 1.

These successfully established conditions (≤150 µM THC, 5% ACN, 0.1% Tween) were used in a multitude of assays in order to obtain information on the affinity of the THC-sepharose-binding sequences and on their ability to recognise THC in solution. Tested were, amongst others, microscale thermophoresis, filter retention assays using THC immobilised on BSA, lysozyme, or THC-sepharose, competition assays using a combination of immobilised THC (THC-sepharose and THCA-beads (structure see **Figure 24**)) and varying concentrations of THC in solution, and pulldown assays, for which the sequences were biotinylated and immobilised on streptavidin-coated plates. None of the assays was able to

reproducibly and clearly show binding of any of the sequences to THC in solution (data not shown).

3.2.6 Re-selection of THC-binding sequences utilising affinity elution

As none of the assays showed binding of any sequence to THC in solution, the sequences that had been identified so far were suspected to recognise THC only if it was immobilised on sepharose. The original selections had used heat elution to unspecifically recover bound DNA (**Table 5**). In a re-selection approach, bound DNA was eluted by addition of decreasing concentrations of THC in 5% ACN to enrich sequences that interact with non-immobilised THC. The re-selection was performed based on DNA from the old selection cycles 4. The buffer used to wash away unbound or unspecifically binding sequences also contained 5% ACN. **Table 14** details the selection conditions during each selection cycle.

Table 14: Conditions for the re-selection with affinity elution

SELEX cycle	incubation time	wash volume	wash time	elution	PCR-cycles
5	15 min	6 ml	n.d.	150 µM THC	B: 14 C: 15
6	15 min	7 ml	n.d.	15 µM THC	B: 18 C: 12
7	15 min	B: 8 ml C: 14 ml	B: 15 min C: 30 min	1.5 µM THC	B: 15 C: 18
8	10 min	B: 14 ml C: 14 ml	B: 22 min C: 23 min	150 nM THC	B: 14 C: 15
9	B: 10 min C: 5 min	B: 22 ml C: 20 ml	B: 43 min C: 43 min	15 nM THC	B: 10 C: 11
10	5 min	B: 40 ml C: 40 ml	B: 84 min C: 81 min	1.5 nM THC	B: 11 C: 12

B: Selection using benzyl-modified DNA. C: Selection using CF_3-modified DNA. n.d. = not determined.

After selection cycle 10, DNA from selection cycles 1, 5, 8, and 10 was tested for its interaction with THC-sepharose. DNA from selection cycle 10 of the first selections (10*) was used as a positive control. DNA from selection cycles 8 and 10 from both selections showed a clear increase in binding to THC-sepharose. The DNA from cycle 10 interacted with THC-sepharose to an equal (benzyl-modified selection) or even stronger degree (CF_3-modified selection) than the DNA from the original selection (10*) (**Figure 38**).

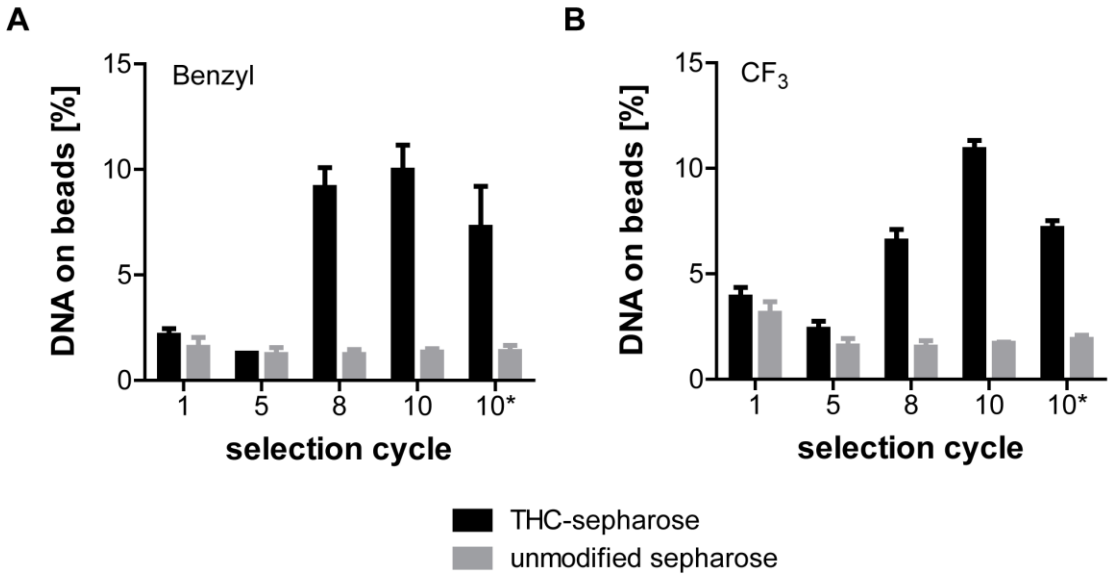


Figure 38: Interaction analysis of THC-selections performed on THC-sepharose using affinity elution
^{32}P-labelled DNA with benzyl- (A) and CF_3- (B) modification was incubated with THC-sepharose (black bars) or unmodified sepharose (grey bars). After two washing steps, the amount of radioactivity on the sepharose was quantified. Indicated with an asterisk is the final selection cycle of the first successful THC-selections with the same modification (compare **Figure 22A** for (A) and **Figure 26A** for (B)).

To identify potential THC-binding sequences, the DNA from both selection cycles 10 was analysed by Sanger sequencing. The identified sequences from the benzyl-modified selection are detailed in **Table 15**. B11, B14, and B27 were known from the first selection. Sequence 3 had also been identified in the first selection, but as a unique sequence and therefore not tested alongside the original sequences (**Table A 1**). Sequence 27 was newly discovered, but only present as a unique sequence (**Table A 17**).

Table 15: Sequences from the affinity THC-selection with benzyl-modified DNA.

Seq.	consensus sequence of the random region	EdUs	Sanger [%]
B11*	GAXGCCGGACGXXGGAACGGGCGXGXCAXCAXAGGXACAGG	8	29.3
B14*	AACCCXGACXGGGAGXGAXGCCXACCGXAACACCXCCCGCAC	7	12.2
B27*	CXACAAXCGXGCGAXCCCCXAXCXCCAGACCXACGGGAACAC	8	48.8
3*	ACACXGGAXAGAXGCGGGGGGCAACGCCXXCAGGCAXCAGX	7	7.3
27	ACGXCCXCAXCACCXCCCAXGXCXCCGGAGXCCACGGACAGG	8	unique

Seq. = sequence. X = EdU. Sanger = frequency in Sanger sequencing of selection cycle 10. * = beforehand identified in the first THC-selection with benzyl-modified DNA

Table 16 compares the frequencies of all known benzyl-modified sequences in the two Sanger sequencings and the NGS (**Table A 1** and **Table A 17**). The non-binding B10 as well as the binding B12 and B13 could not be detected in the Sanger sequencing of the re-selection. B27 was still the most abundant sequence. B11 was much more abundant than in the Sanger sequencing of the first selection, but was already the second most abundant sequence according to

the NGS of the first selection. Both new sequences, 3 and 27, could be found in the NGS data of the first selection cycle 10, but with frequencies close to the 0.5% of the non-binding B10.

Table 16: Frequencies of all benzyl-modified sequences identified from the first (1st) and second (affinity, 2nd) THC-selection

Seq.	1st Sanger [%]	1st NGS [%]	2nd Sanger [%]
B10	5.4	0.5	n.f.
B11	10.8	5.1	29.3
B12	13.5	4.1	n.f.
B13	8.1	1.7	n.f.
B14	24.3	1.6	12.2
B27	35.1	10.6	48.8
3	unique	0.8	7.3
27	n.f.	0.2	unique

Seq. = sequence. Sanger = frequency in Sanger sequencing of the respective selection cycle 10. NGS = frequency in NGS of the respective selection cycle 10. n.f. = sequence not found.

The tested sequences that were identified from the final selection cycle of the new CF_3-SELEX are depicted in **Table 17**. C15 was not found, but both C30 and C47 were present in the new selection (**Table A 18**). A multitude of new sequences was identified: 49, 53, 63, 67, 74, and 77. C47 was the most abundant sequence, but closely followed by the new sequence 53. All other new sequences were at least as abundant as C30.

Table 17: Sequences from the affinity THC-selection with CF_3-modified DNA

Seq.	consensus sequence of the random region	EdUs	Sanger [%]
C30*	CCAGCCGCGCGGAAGACAAXACACGXXGGGCCACXAAGGAAG	4	4.9
C47*	CGAXAAXACACGXXCGGCCCCXAAAGCCGGXCGGCCCXXGCA	8	19.5
49	XGCGCGAGGGAXGAXXAAGAGAAGCXCAACCGCGCGGAGGCC	5	9.8
53	XXCGGACGGAGAGAGCAGGGGGXGAGXAAAGAGXAGGXCCAG	6	14.6
63	GCXCACGCCACACAAGGCAGAXCCXGAGCXCCGGGGGXCCAC	5	7.3
67	AACXGCAGAAAGGGGAAGGCGAGGAAXAGXCCCCGGXXGGCA	5	4.9
74	GCACGGAACAACCGCGAACGXACXAGCAXCCCGCGGCGAACA	3	9.8
77	GGXAAXGGXGGAGGGXAAGGAGGACCCCGAAXXGXCCCCGXG	8	7.3

Seq. = sequence. X = EdU. Sanger = frequency in Sanger sequencing of selection cycle 10. * = beforehand identified in the first THC-selection with CF_3-modified DNA

When comparing all three sequencing approaches for the selections with CF_3-modified DNA, C47 was always the most abundant sequence. Some of the new sequences were identified in the final selection cycle of the first selection by NGS, but their frequencies were extremely low. Sequences 63 and 74 could not be identified at all in the original selection cycle (**Table 18**).

Table 18: Frequencies of all CF_3-modified sequences identified from the first (1st) and second (affinity, 2nd) THC-selection

Seq.	1st Sanger [%]	1st NGS [%]	2nd Sanger [%]
C15	6.3	0.1	n.f.
C30	9.4	0.6	4.9
C47	43.8	11.4	19.5
49	n.f.	0.005	9.8
53	n.f.	0.01	14.6
63	n.f.	n.f.	7.3
67	n.f.	0.03	4.9
74	n.f.	n.f.	9.8
77	n.f.	0.002	7.3

Seq. = sequence. Sanger = frequency in Sanger sequencing of the respective selection cycle 10. NGS = frequency in NGS of the respective selection cycle 10. n.f. = sequence not found.

All newly identified sequences were tested for binding to THC-sepharose in a radioactive assay alongside B10 as a negative control. All sequences interacted with THC-sepharose to varying degrees. Sequence 67 was the weakest binder (**Figure 39**).

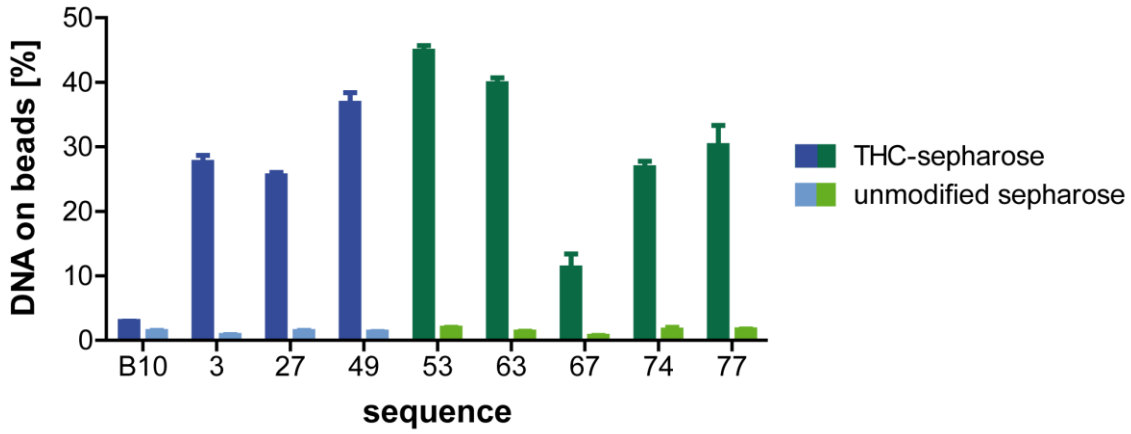


Figure 39: Interaction analysis of the new sequences from the affinity THC-selection
Benzyl-modified DNA is depicted in shades of blue, CF_3-modified DNA in shades of green. ^{32}P-labelled DNA was incubated with THC-sepharose (dark blue or green) or unmodified sepharose (light blue or green). After two washing steps, the amount of radioactivity on the sepharose was quantified. B10 from the first selection was used as a negative control.

Competition as well as pulldown assays were performed with the newly identified sequences, but, as for the old sequences, no binding to THC in solution was detected (data not shown).

3.2.7 Fluorescence polarisation assays with THC-FITC

Fluorescence polarisation assays represent a method to evaluate in solution interaction between two molecules. They are often used with fluorescently-labelled aptamers for the detection of their interaction with much larger

proteins.[218] The smaller interaction partner needs to be fluorescent or fluorescently labelled.

Fluorescence polarisation measures the rotation of the fluorescently labelled molecule. **Figure 40** depicts the reason the linear polarised light, which is used to excite the fluorophore, is either emitted as polarised light or not: Small molecules rotate before the light is emitted. Thus, the resulting light is not polarised any longer. Larger molecules or complexes do not rotate before the light is emitted, leading to a high polarisation of the emitted light.[219]

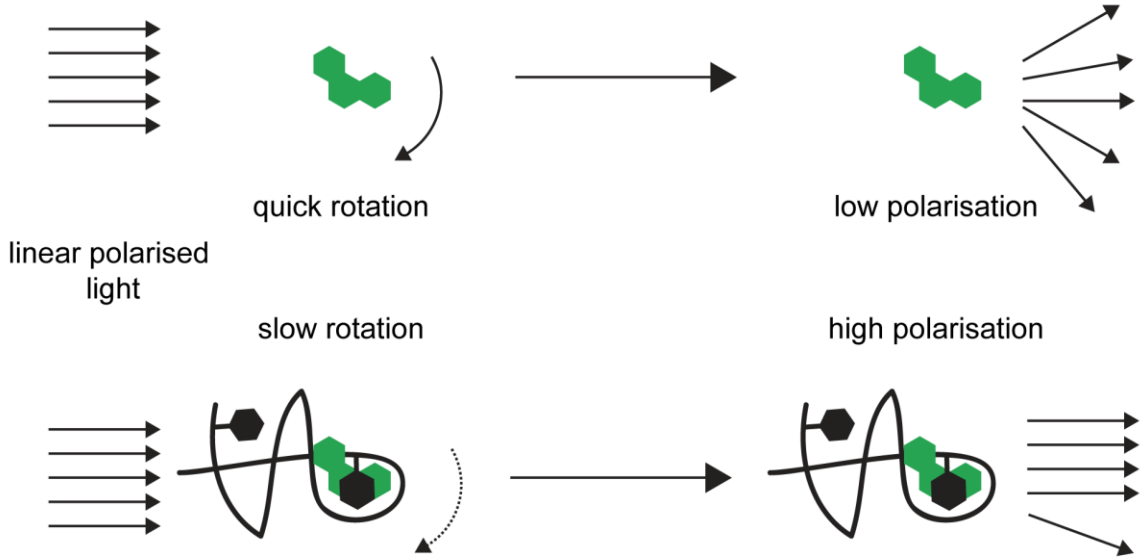


Figure 40: Fluorescence polarisation assay
Linear polarised light is used to excite a fluorophore (green). The small fluorophore rotates quickly and the light is emitted into different directions, resulting in a low polarisation (upper panel). Larger molecules or complexes rotate slowly. The light is emitted before any rotation has taken place and is therefore still polarised, leading to a high polarisation value (lower panel).

3.2.7.1 Synthesis and characterisation of THC-FITC

As THC is the smaller interaction partner and no fluorescently labelled THC-derivative was commercially available, THC-FITC was synthesised (**Figure 41**).

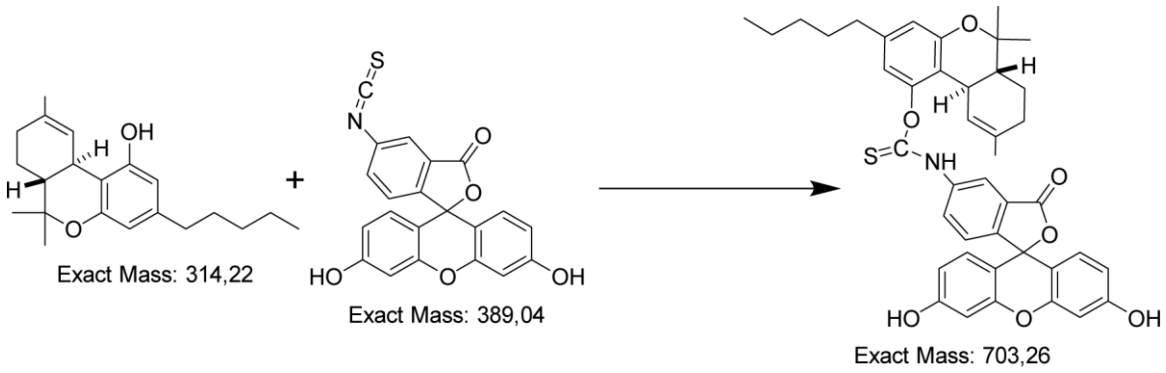


Figure 41: Coupling of THC to FITC
THC is coupled to FITC (molar ratio THC:FITC 5:1) in dimethylformamide (DMF) with 20% TEA for 2 days at 65°C, yielding THC-FITC.

The successful coupling was confirmed using LC-MS analysis (**Figure 42A**). Peak 1 was caused by THC, as identified in the mass spectrum (**Figure 42B**). Peak 2 represented the coupling product, THC-FITC (**Figure 42C**). The two peaks were integrated to analyse the coupling yield: About 75% of THC was successfully coupled to FITC (**Figure 42A**). Due to its high water solubility, the uncoupled FITC was most likely represented by the peak at min 1-2, which was not injected into the mass spectrometer.

To remove unreacted THC and FITC, the crude product was purified using HPLC. LC-MS-analysis of the purified product showed that the THC-peak is no longer detectable and the FITC-peak has been severely reduced, indicating that the purification was successful (**Figure 42A**). The remaining peak was THC-FITC, as ensured by the corresponding mass analysis (**Figure 42D**).

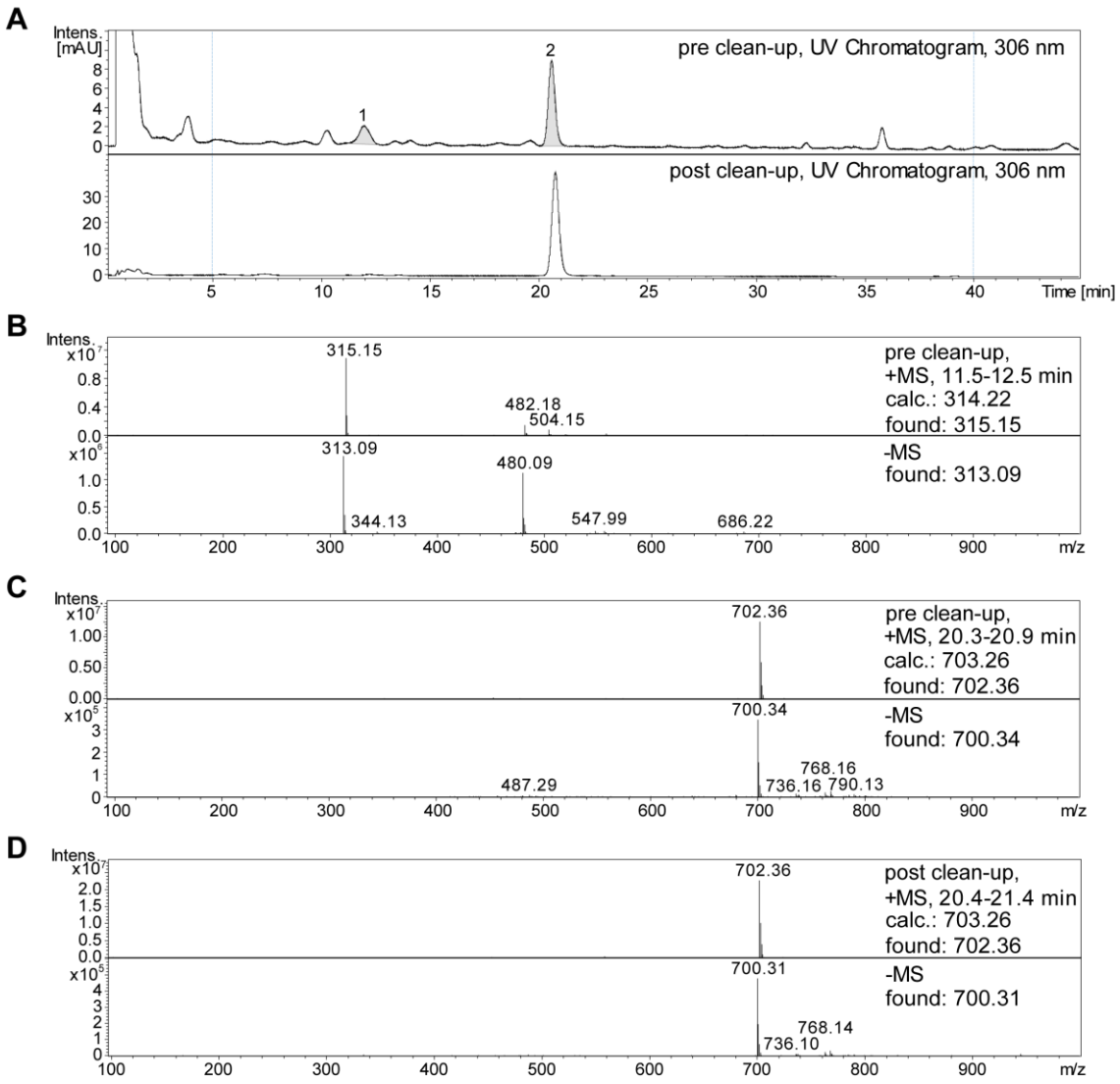


Figure 42: LC-MS analysis of coupling of THC to FITC before and after clean-up
(A) UV-chromatograms of the LC-MS analysis of the coupling of THC to FITC before (upper panel) and after (lower panel) clean-up. Peaks 1 and 2 were analysed regarding the area under the curve for quantification of THC (peak 1) and THC-FITC (peak 2). (B) Positive (upper panel) and negative (lower panel) mass spectrum of peak 1, pre clean-up. The found mass matches THC. (C) Positive (upper panel) and negative (lower panel) mass spectrum of peak 2, pre clean-up. The found mass matches THC-FITC. (D) Positive (upper panel) and negative (lower panel) mass spectrum of peak 2, post clean-up. The found mass matches THC-FITC.

The newly synthesised THC-FITC was then characterised regarding its spectral properties. It showed two absorption maxima at 456 and 492 nm in water with 5% ACN (**Figure 43**). The latter was the known absorption maximum of FITC and was used for the concentration determination of THC-FITC. After excitation at 492 nm, THC-FITC emits at about 520 nm. The emission was strongly reduced at pH 5.4 (D-PBS) in comparison to water (**Figure 43B**). Nonetheless, the fluorescence should be sufficient to measure fluorescence polarisation.

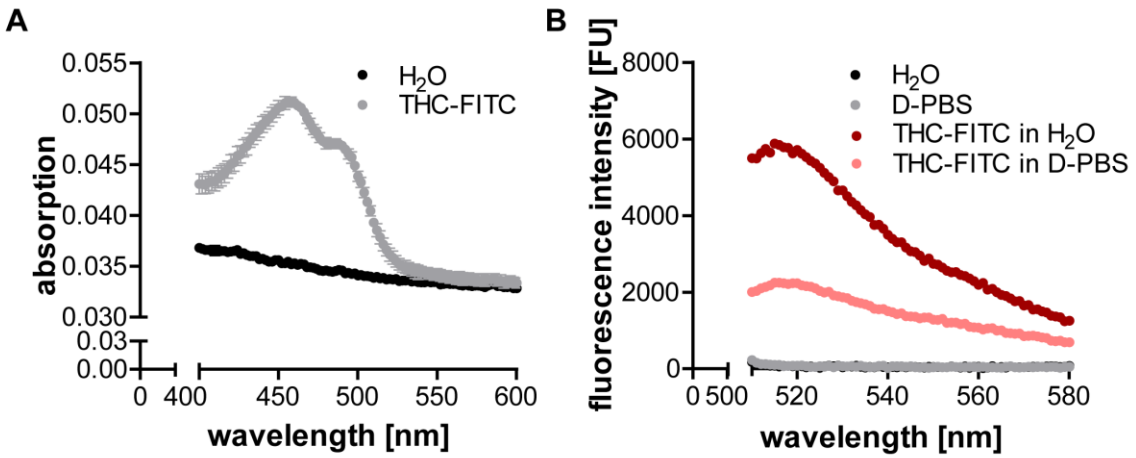


Figure 43: Absorption and emission spectra of THC-FITC
(A) Absorption spectrum of THC-FITC (grey) (d = 0.05 cm) in water with 5% ACN (blank, black). (B) Emission spectrum of THC-FITC (excitation at 492 nm, depicted in shades of red). Fluorescence was measured in D-PBS, pH 5.4 (light red) and ddH_2O (dark red). The blanks are depicted in grey and black, respectively. All samples contained 5% ACN to solubilise THC-FITC.

3.2.7.2 Establishing the fluorescence polarisation assay

To determine the optimal concentration for the interaction assays and to uncover possible differences between the two fluorophores, different concentrations of THC-FITC and FITC were tested for their fluorescence polarisation. The samples without Tween in **Figure 44A** showed that the polarisation of THC-FITC was strongly increased in comparison to FITC alone. Such an increase in polarisation was expected upon ligand-recognition. Therefore, THC-FITC could not be used for fluorescence polarisation assays under these conditions as no interaction with the THC-sepharose-binding sequences would be detectable. Thorough investigations of the reason for the increase in polarisation were conducted to establish conditions, with which THC-FITC could be used in fluorescence polarisation assays. Afterwards, binding of the THC-sepharose-binding sequences under the new conditions had to be verified.

First of all, the influence of Tween was investigated. All assays up to this point contained 0.1% Tween, which was both part of the selection buffer (of all selections with click-modified DNA) as well as known to be necessary to keep THC in solution (**Figure 35A**). Its critical micelle concentration (CMC) of 0.0067% is far below the 0.1% used.[217]

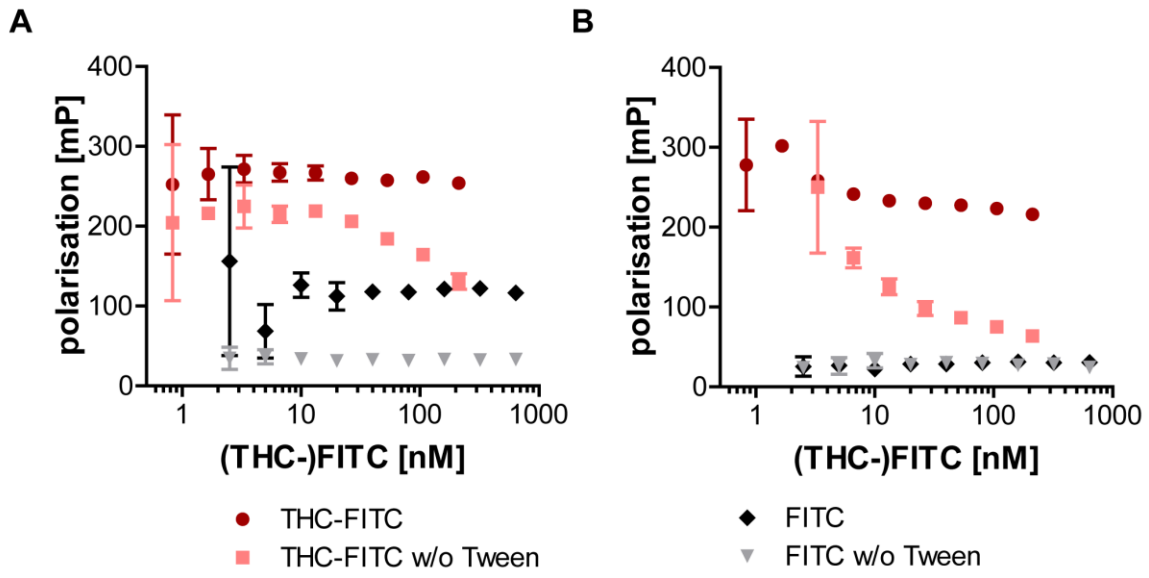


Figure 44: Florescence polarisation of FITC and THC-FITC in the presence and absence of Tween-20

The fluorescence polarisation of varying concentrations of FITC (black and grey) and THC-FITC (shades of red) was measured in the presence (dark red, black) and absence (light red, grey) of 0.1% Tween-20. (A) Measurement in D-PBS, pH 5.4. (B) Measurement in PBS, pH 7.4. Values for 0.8 and 1.7 nM THC-FITC at pH 7.4 in the absence of Tween were omitted as the intensities were not above background intensity.

The CMC determines the concentration at which a detergents starts to form micelles. At concentrations lower than the CMC, the detergent can be found either in solution or at the surface (**Figure 45A**). With increasing detergent concentration, increasing amounts are adsorbed at the surface and thereby decrease surface tension (**Figure 45B**). Once the surface is saturated, the detergent starts to form micelles (**Figure 45C**).[220]

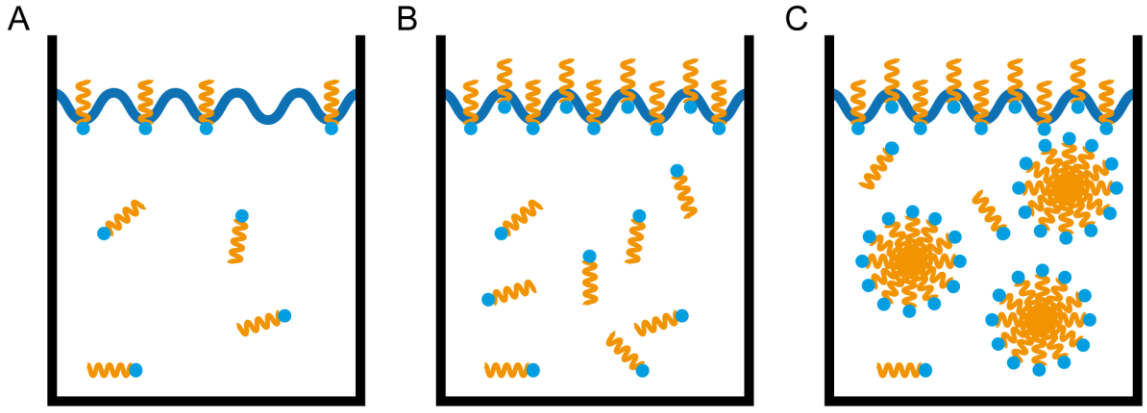


Figure 45: Visualisation of the detergent distribution at concentrations higher and lower than its CMC

The detergent is depicted with an orange hydrophobic tail and a light blue hydrophilic head group. (A) At concentrations far below the CMC, the detergent can be found both in solution and at the surface. (B) With concentrations approaching the CMC, more and more detergent is adsorbed at the surface. The surface tension decreases because of this. (C) Once the surface is saturated with detergent, micelles start to form. The CMC has been reached.

As the Tween-concentration used is above its CMC, it is possible that THC and THC-FITC are contained within Tween-micelles. Repeating the determination of the concentration-dependent polarisation in the absence of Tween led to a decrease of the polarisation-values for both FITC and THC-FITC. The effect became more pronounced at higher concentrations of THC-FITC (**Figure 44A**). The effect was even more noticeable for THC-FITC at pH 7.4 (**Figure 44B**). FITC showed no dependence on the presence or absence of Tween at neutral pH.

To evaluate the hypothesis that THC and THC-FITC are enclosed by Tween-micelles, we tested their solubility in the presence of different concentrations of Tween: 0.1%, which had been used extensively and is far above the CMC, 0.01%, which is above the CMC, and 0.05%, which is just below the CMC. The results of the HPLC-measurement are depicted in **Figure 46**.

For THC-FITC (**Figure 46A**) and THC (**Figure 46B**) both, 100% as well as 5% ACN with 0.1% Tween led to good linearities. 5% ACN without any Tween also resulted in a good linearity for THC-FITC, but the gradient was completely different than the one of the regression curves of the two other conditions. Both lower Tween-concentrations led to a complete loss of linearity for THC-FITC, which mainly resulted from the decrease in signal obtained for the highest concentration of 20 µM.

For THC (**Figure 46B**), 5% ACN alone led to a decrease in linearity and a changed gradient. Both Tween-concentrations above the CMC, 0.1 and 0.01%, led to a good linearity and a gradient similar to the sample containing 100% ACN. The concentration below the CMC, 0.005% Tween, led to a loss of linearity and a gradient similar to the sample containing 5% ACN, but no Tween.

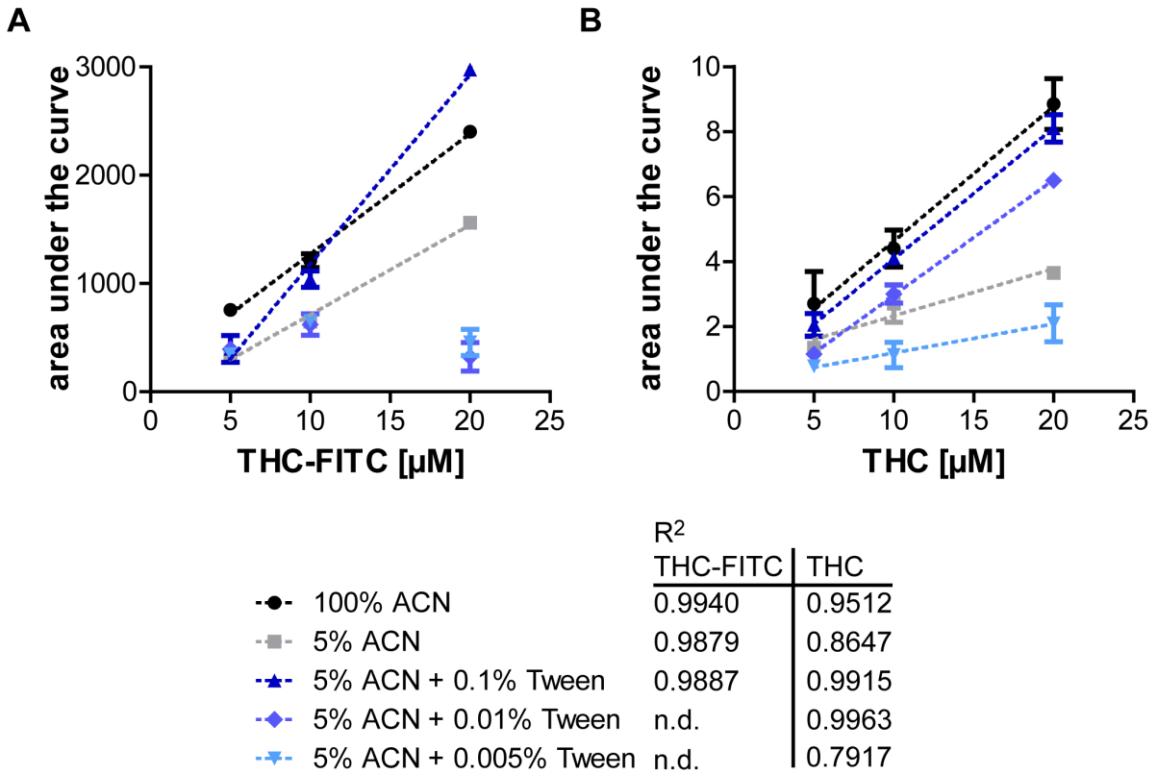


Figure 46: HPLC-measurement of THC-FITC- and THC-dilutions in ACN with different concentrations of Tween
HPLC-analysis of THC-FITC- (A) and THC- (B) concentrations between 5 and 20 µM in 100 (black) and 5% ACN (grey) as controls as well as in 5% ACN with 0.1 (dark blue), 0.01 (medium blue), and 0.005% Tween-20 (light blue).

The addition of Tween was not necessary to keep THC-FITC well solubilised. This would prevent any potential problems with Tween-micelles. Therefore, the maximal solubility of THC-FITC in 5% ACN was determined. The linearity of the resulting regression curve was sufficient for concentrations up to 25 µM (**Figure 47**). As the concentrations of THC-FITC used for fluorescence polarisation assays were in the nanomolar range, no solubility problems should arise.

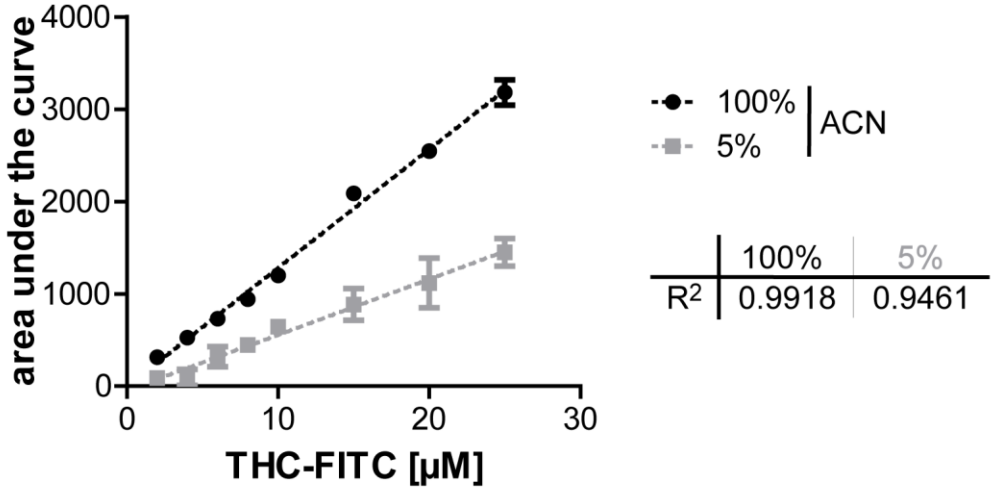


Figure 47: HPLC-measurement of THC-FITC-concentrations in ACN
HPLC-measurement of THC-FITC-concentrations between 2 and 25 µM in 100 (black) and 5% ACN (grey).

To determine the optimal buffer for fluorescence polarisation assays to analyse the interaction of the THC-sepharose-binding sequences with THC-FITC, different components of the selection buffer were tested for their influence on the polarisation. At the concentration used, 0.1% Tween had little effect. In contrast, the presence of BSA increased the polarisation by a factor of 2 to 2.5. It therefore had to be removed from the fluorescence polarisation buffer (**Figure 48**).

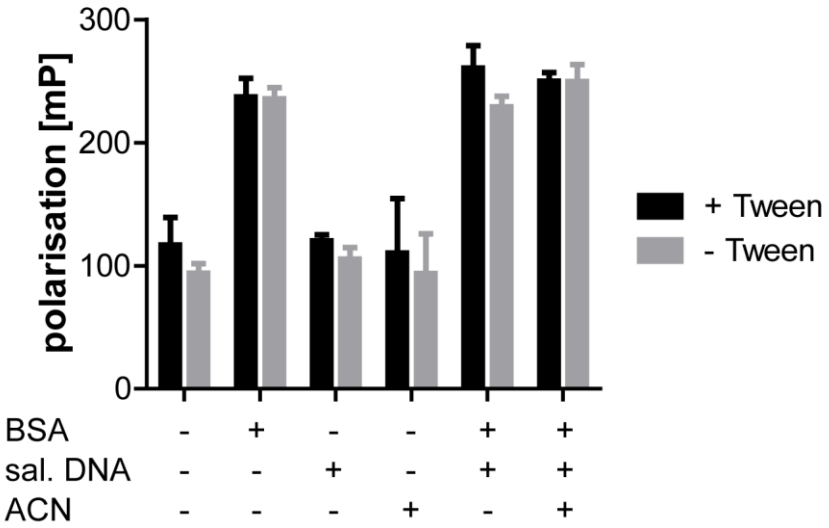


Figure 48: Fluorescence polarisation of THC-FITC in different buffers
Fluorescence polarisation of THC-FITC (26.5 nM) in the presence of different components of the selection buffer: BSA, salmon sperm DNA (sal. DNA), ACN, and Tween.

3.2.7.3 Evaluation of the THC-sepharose-binding sequences under conditions differing from the original selection buffer

Before any fluorescence polarisation assays could be performed, the interaction of the THC-sepharose-binding sequences with THC-sepharose under the changed conditions (no Tween, no BSA) had to be confirmed.

For this, all THC-sepharose binding sequences were incubated with both THC- and unmodified sepharose in the presence and absence of 0.1% Tween. The absence of Tween increased the interaction of all sequences with the THC-sepharose to varying degrees (strongest for C15 and C47). B10 remained non-binding under both conditions (**Figure 49A**).

Figure 49B shows the results of the same assay in the presence and absence of BSA. No Tween was added to any of the samples. As with Tween, C15 and C47 showed the highest differences between samples: BSA increased their interaction with THC-sepharose. Nonetheless, the interaction was sufficient also in the absence of BSA. The interaction of other sequences was only reduced slightly in the absence of BSA. B10 was not influenced at all.

In conclusion, all sequences still interact with THC-sepharose in the absence of both Tween and BSA and are therefore fit to be tested for binding to THC in solution in fluorescence polarisation assays.

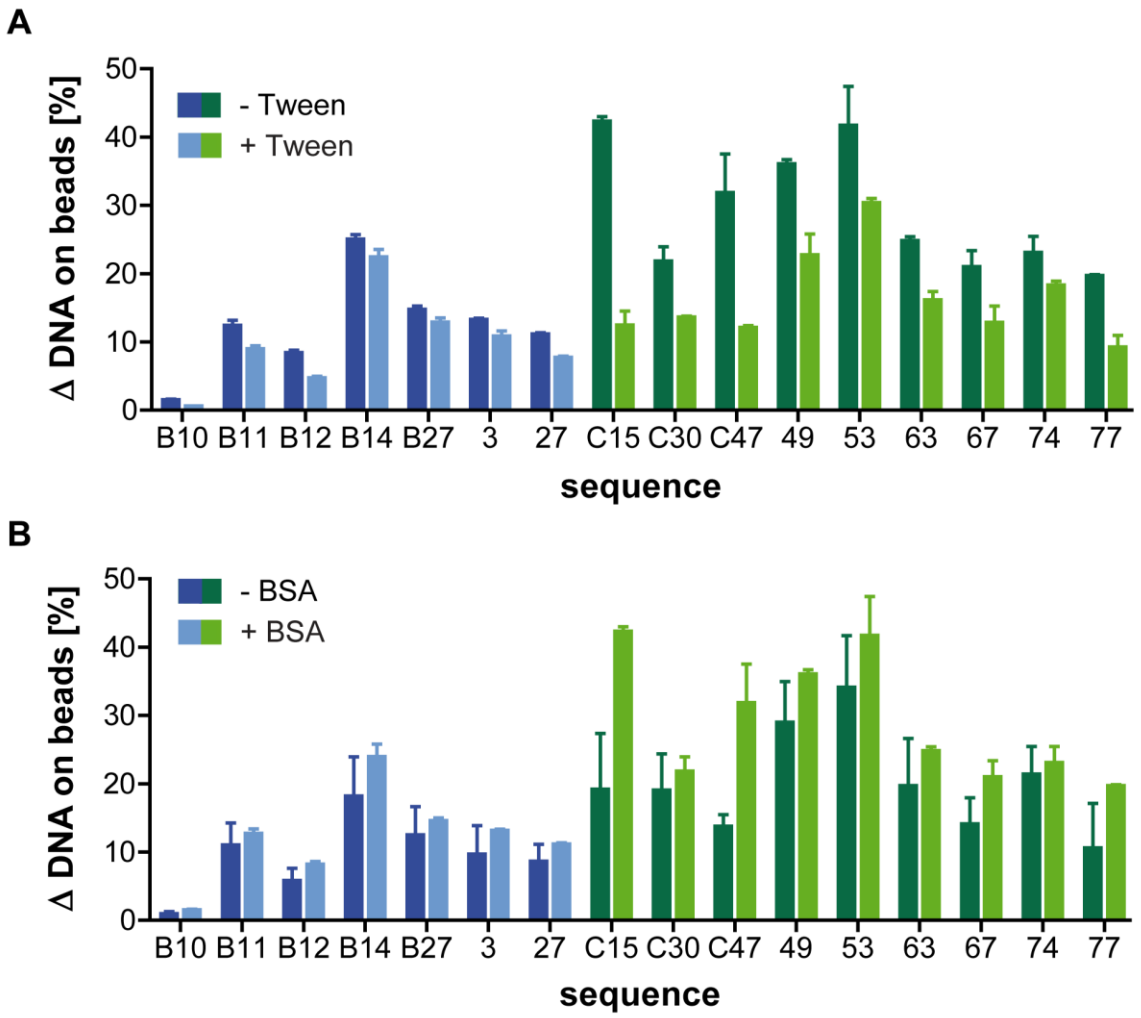


Figure 49: Interaction analysis of all sequences in the presence and absence of Tween and BSA
Benzyl-modified DNA is depicted in shades of blue, CF_3-modified DNA in shades of green. ^{32}P-labelled DNA was incubated with THC-sepharose or unmodified sepharose. The values on unmodified sepharose were subtracted from those on THC-sepharose and the difference depicted. (A) The samples were incubated in D-PBS, pH 5.4, with 0.1 mg/ml salmon sperm DNA, 0.1 mg/ml BSA, and 5% ACN in the presence (light blue and green) or absence (dark blue and green) of 0.1% Tween-20. (B) The samples were incubated in D-PBS, pH 5.4, with 0.1 mg/ml salmon sperm DNA, and 5% ACN in the presence (light blue and green) or absence (dark blue and green) of 0.1 mg/ml BSA.

As a last test, binding of all identified sequences at neutral pH was investigated. Since the fluorescence polarisation of THC-FITC is lower at pH 7.4 (**Figure 44**), the assay sensitivity is increased at this pH and so is the possibility of detecting in solution binding of THC.

Most sequences lost their ability to interact with THC-sepharose at pH 7.4. (**Figure 50**) The non-binding control, B10, showed slightly higher binding values than at pH 5.4. The benzyl-modified sequences B12 and 27 showed equal (B12) or slightly decreased (27) interaction. Of the CF_3-modified sequences, C15, C47, 49, 53, and 77 retained their ability to interact with THC-sepharose, although the

interaction of 49, 53, and 77 is reduced. Overall, two of six benzyl- and five of nine CF_3-modified sequences bind to THC-sepharose at neutral pH. The identified sequences will be tested for interaction with THC-FITC at pH 7.4 in the fluorescence polarisation assay.

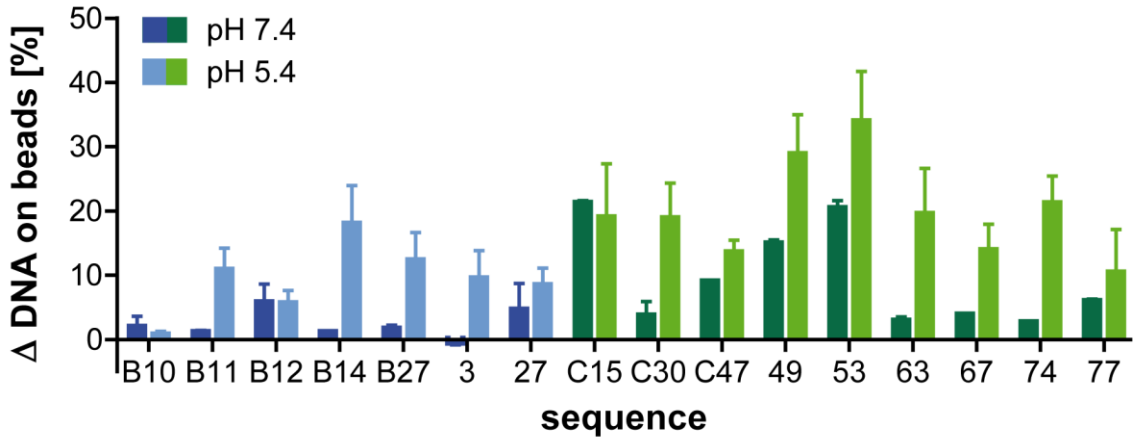


Figure 50: Interaction analysis of all sequences at pH 7.4
Benzyl-modified DNA is depicted in shades of blue, CF_3-modified DNA in shades of green. ^{32}P-labelled DNA was incubated with THC-sepharose or unmodified sepharose. The values on unmodified sepharose were subtracted from those on THC-sepharose and the difference depicted. The samples were incubated in D-PBS with 0.1 mg/ml salmon sperm DNA and 5% ACN at pH 7.4 (dark blue and green) and, as control, 5.4 (light blue and green).

3.2.7.4 Fluorescence polarisation measurements using THC-FITC

Prior to testing the sequences for their recognition of THC in solution, the ability of the fluorescence polarisation assay to measure concentration-dependent interactions had to be verified. Since no DNA – modified or not – is known to interact with THC, we chose the THC-antibody that was also used to detect THC on the THCA-beads and THC-sepharose. A galectin 3 (Gal3)-antibody, which should not interact with THC-FITC, as well as FITC were used as negative controls. The experiment was performed in PBS, pH 7.4**.** As before (**Figure 44**), the polarisation values were higher for THC-FITC than for FITC alone (**Figure 51**). Any combination with either FITC or the Gal3-antibody resulted in antibody-independent, low fluorescence polarisation values. Only the combination of THC-FITC and the THC-antibody led to an increase in fluorescence polarisation that depended on the antibody-concentration. The determined affinity constants were 27.8 ± 28.9 and 62.7 ± 60.5 nM for THC-FITC-concentrations of 26.5 and 53 nM, respectively.

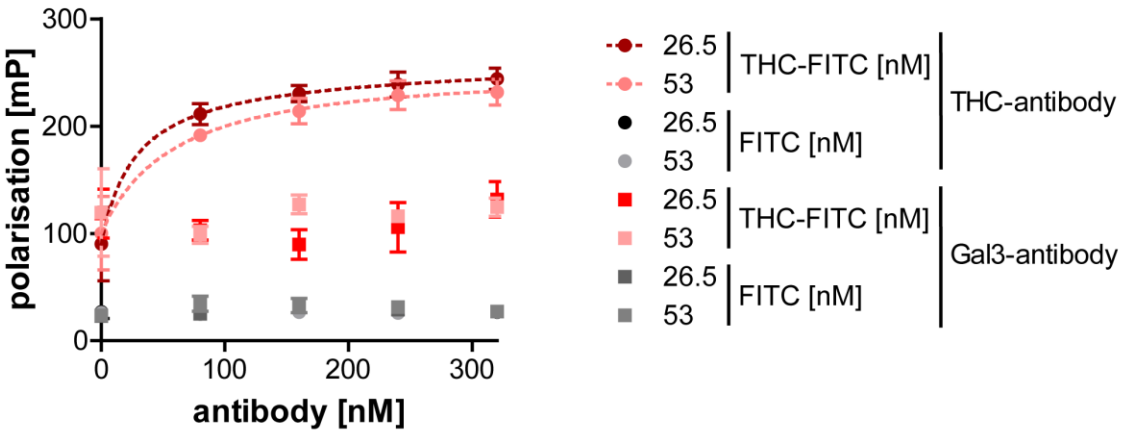


Figure 51: Fluorescence polarisation of FITC and THC-FITC with varying concentrations of antibodies
The fluorescence polarisation of FITC (black, shades of grey) and THC-FITC (shades of red) in the presence of varying concentrations of THC- (dots) and Gal3-antibody (squares) was measured. A K_d was determined for THC-FITC with the THC-antibody: 27.8 ± 28.9 and 62.7 ± 60.5 nM for 26.5 and 53 nM THC-FITC, respectively.

For the assays at pH 5.4, BSA was used as a positive control, as it is not known if the THC-antibody stably interacts with THC at this pH. BSA had been shown to interact unspecifically with THC-FITC and to increase its polarisation in the preliminary buffer test (**Figure 48**).

The results of the fluorescence polarisation assay with the different benzyl-modified sequences are shown in **Figure 52A**. While BSA led to the expected increase in polarisation, none of the sequences had any effect.

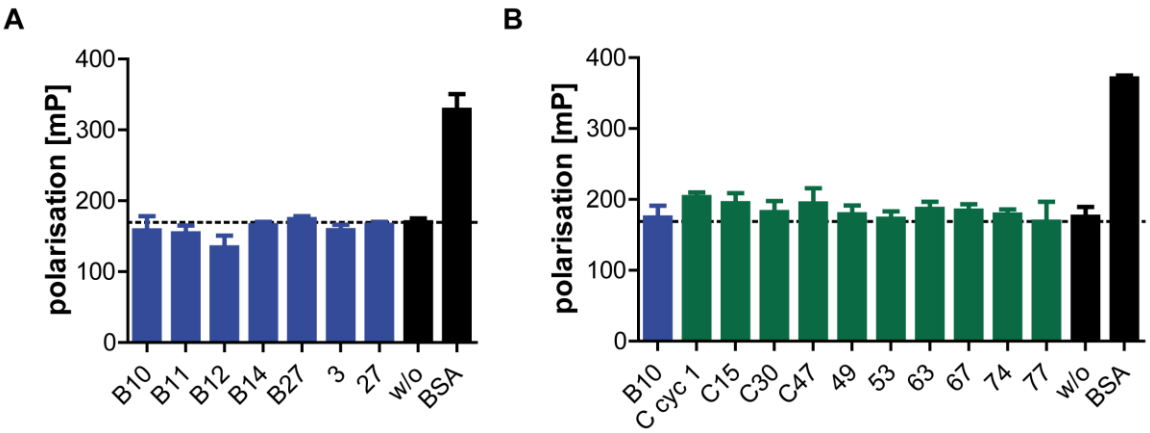


Figure 52: Fluorescence polarisation of THC-FITC in the presence of different sequences at pH 5.4
The fluorescence polarisation of THC-FITC (53 nM) was measured in the presence of 500 nM of the different sequences, 1 mg/ml BSA, or without any sequence (w/o). BSA was used as a positive control. No sequence (w/o) was used as negative control and the dotted line represents the corresponding polarisation. Benzyl-modified DNA is depicted in blue, CF_3-modified DNA in green. (A) Sequences from the selection with benzyl-modified DNA. (B) Sequences from the selection with CF_3-modified DNA. B10 was measured alongside as an additional, negative control. C cyc 1 indicates DNA from the first selection cycle of the selection with CF_3-modified DNA.

The same is true for the assay concerning the CF_3-modified sequences, which is depicted in **Figure 52B**. Some of the sequences led to a slight increase in polarisation in comparison to the non-binding B10 and the sample without DNA (w/o). Nevertheless, no increase was apparent in comparison to CF_3-modified DNA from the first selection cycle (C cyc 1).

The positive control BSA weighs about 66 kDa[221], the THC-antibody has a size of about 150 kDa.[222,223] The sequences consist of around 84 nucleobases and weigh 26 to 28 kDa, depending on the individual length, base composition, and number and type of modification. While unlikely, it is possible that the weight of the sequence alone might not be sufficient to lead to a detectable increase in polarisation. To rule this out, the sequences were 5'-biotinylated and streptavidin (52 kDa[224]) was added to the fluorescence polarisation assay. The combined weight of at least 78 kDa should lead to an increase in assay sensitivity.

Figure 53 illustrates that streptavidin had no influence on the polarisation in the presence of BSA, but led to a slight increase in polarisation in the absence of DNA (w/o), resulting in an increase in background. None of the complexes of streptavidin with the biotinylated benzyl- or CF_3-modified sequences increased the measured polarisation.

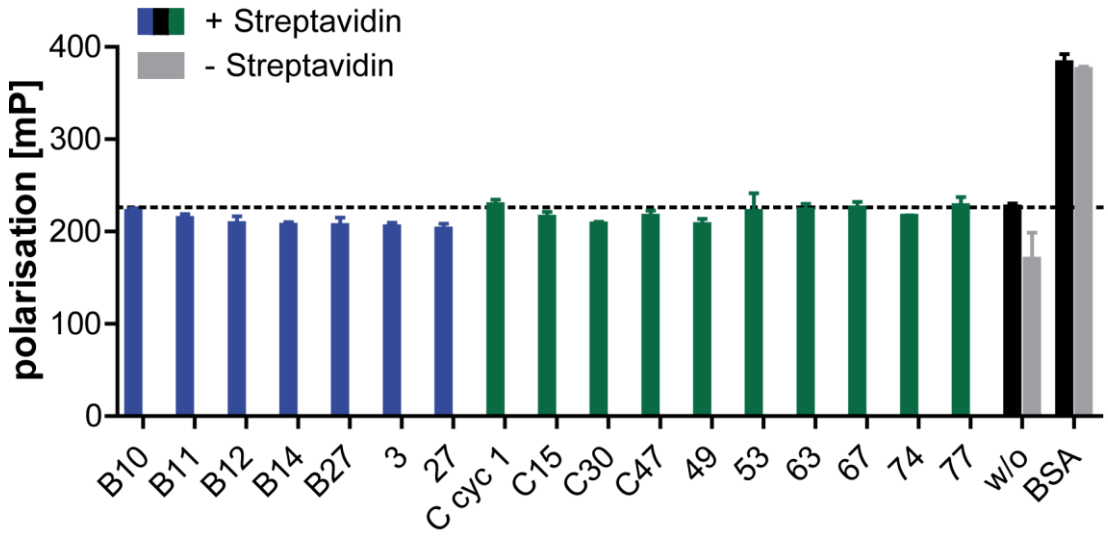


Figure 53: Fluorescence polarisation of THC-FITC in the presence of different biotinylated sequences and streptavidin
The fluorescence polarisation of THC-FITC (53 nM) was measured in the presence of 500 nM of the different, biotinylated sequences, 1 mg/ml BSA, or without any sequence (w/o). All assays but those represented by grey bars contained 750 nM streptavidin. BSA was used as a positive control. No sequence (w/o) was used as negative control and the dotted line represents the corresponding polarisation. Benzyl-modified DNA is depicted in blue, CF_3-modified DNA in green.

The fluorescence polarisation of THC-FITC is lower at pH 7.4 than at 5.4 (**Figure 44**). The resultant increase in assay sensitivity might enable the detection of sequences that bind THC in solution. Therefore, the sequences that interacted

with THC-sepharose at neutral pH in the radioactive assay (**Figure 50**) were also tested at neutral pH in the fluorescence polarisation assay. As in the assays at pH 5.4, no increase in polarisation and therefore no interaction with THC-FITC was detectable (**Figure A 3**).

So far, none of the setups (increase of molecular weight by addition of streptavidin and biotinylation, pH) led to the detection of interaction of THC-FITC with any of the sequences. In comparison with THC-sepharose (chemical structure see **Figure 19**), which all the sequences are known to interact with (**Figure 23**, **Figure 27**, and **Figure 39**), THC-FITC has the relatively bulky FITC directly connected to the oxygen of the former hydroxyl group of THC. This can lead to steric hindrances if any of the sequences normally bind from the coupling side (**Figure 54A**). To rule out that the interaction with THC in solution is blocked by FITC, a different coupling product was synthesised: THC-L-FITC (TLF, **Figure 54B**). Here, THC and FITC are coupled through the same linker that is also on THC-sepharose.

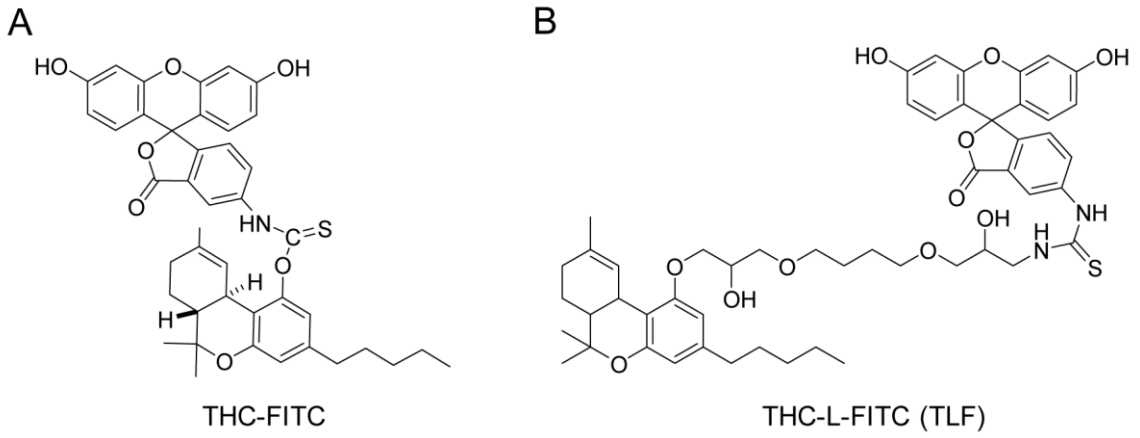


Figure 54: Structures of THC-FITC and THC-L-FITC
Chemical structures of both coupling products of THC and FITC that were used for fluorescence polarisation assays in this thesis. (A) THC-FITC. (B) TLF.

3.2.8 Fluorescence polarisation assays with THC-L-FITC

3.2.8.1 Synthesis and characterisation of THC-L-FITC

TLF is synthesised from THC-L, whose synthesis is depicted in **Figure 16**. THC-L was reacted with NH_3 and FITC to TLF in a two-step synthesis as shown in **Figure 55**.

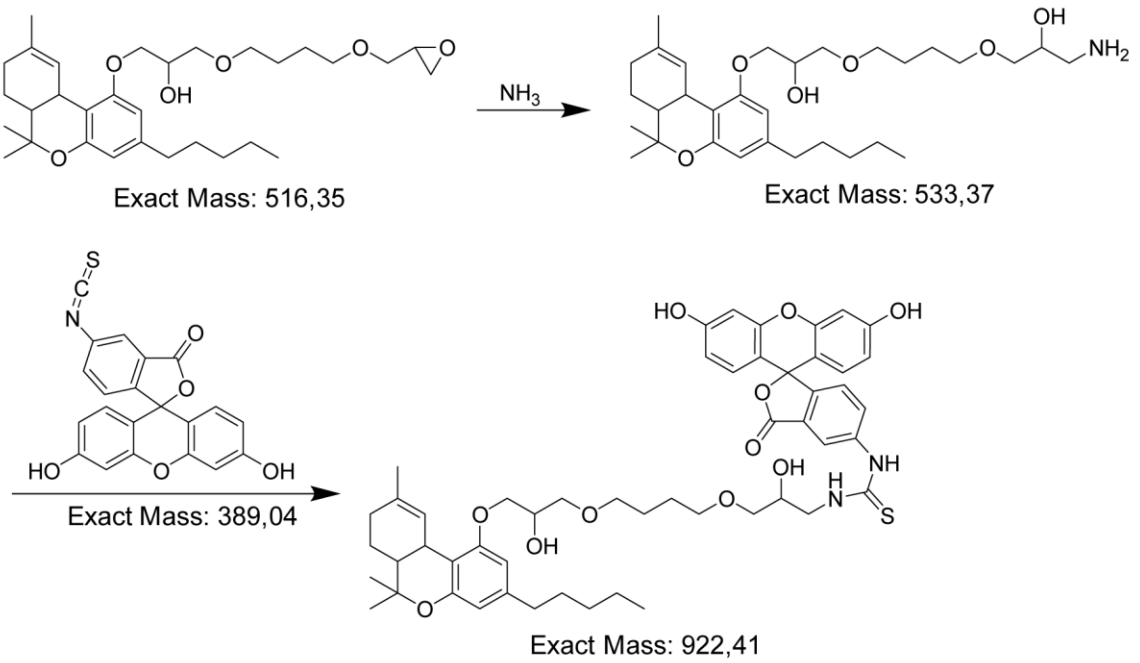


Figure 55: Coupling of THC-L to FITC
THC-L (on the upper left) reacts with ammoniac (on the upper arrow), opening the epoxide ring. The reaction takes place in MeOH for 4 h @ 70°C. After evaporation of the remaining NH_3 and MeOH, FITC (on the lower arrow) is added to the crude product and reacted in MeOH for 3 h at room temperature (RT) to the final product, TLF (lower right).

The crude product was analysed by LC-MS as depicted in **Figure 56A** to **C**. While both the positive (**Figure 56A**) and negative (**Figure 56B**) mass showed a lot of different peaks, the UV chromatogram (**Figure 56C**) contained only two peaks: One that was most likely unreacted FITC (min 0-1, compare **Figure 42A**) and the product peak (min 9-10).

To remove the leftover FITC as well as whatever kind of impurities were represented by the additional mass peaks, the crude product was purified by HPLC. Afterwards, most additional mass peaks had vanished (**Figure 56D** and **E**) and the FITC-peak in the UV chromatogram was dramatically reduced in comparison to the product peak (**Figure 56F**).

The mass analysis of the product peak (min 9-10) both before and after clean-up is depicted in **Figure 57** and verified that the peak represents TLF. The increased purity of the sample was also indicated by the decrease of contaminating mass peaks in this analysis.

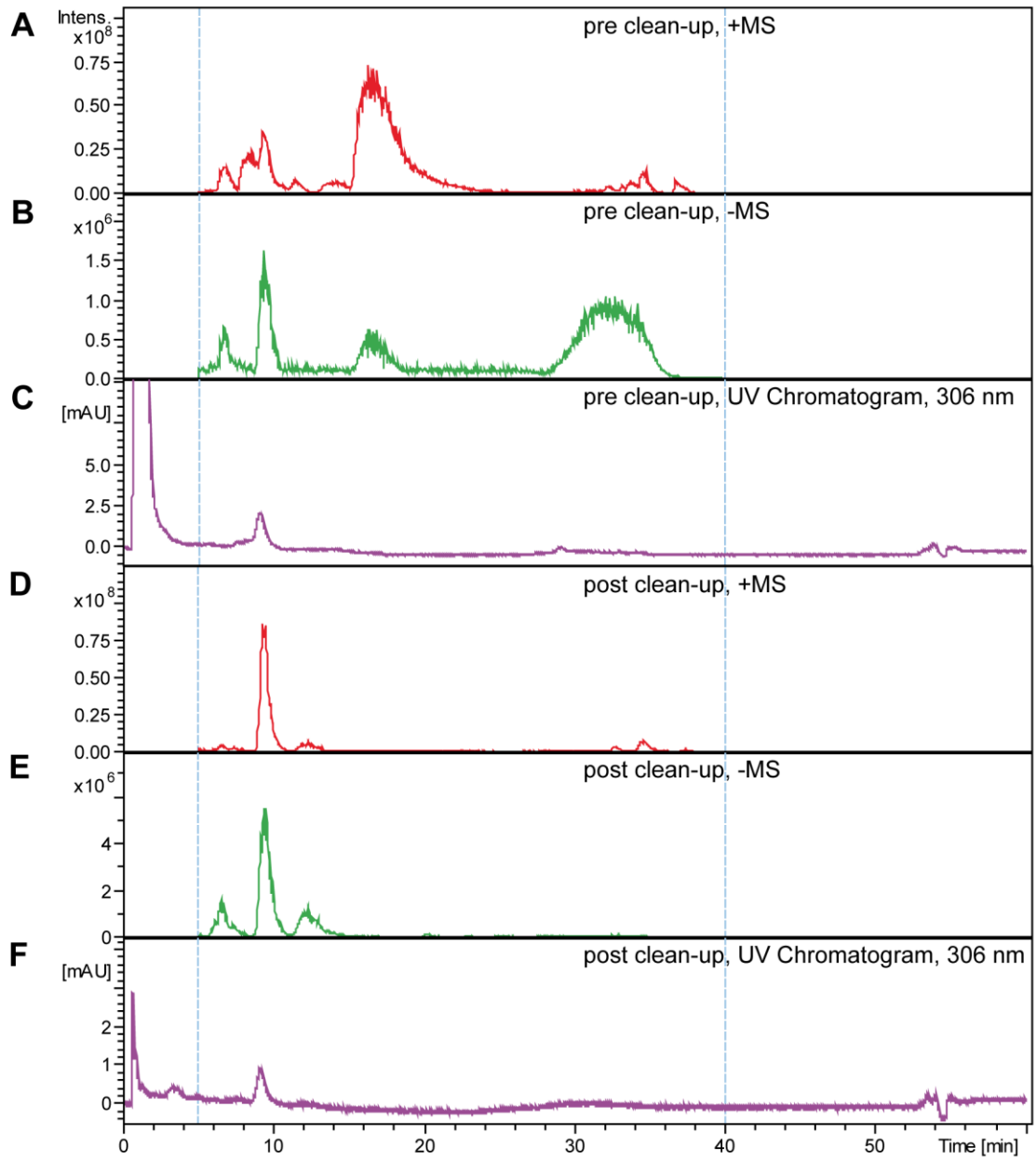


Figure 56: LC-MS analysis of coupling of THC-L to FITC before and after clean-up
Positive (A) and negative (B) mass and UV chromatogram (C) of the coupling before clean-up. Positive (D) and negative (E) mass and UV chromatogram (F) of the coupling after clean-up. The single peak (min 9-10) in the UV chromatograms represents the coupling product.

Some FITC was left in the sample even after purification (**Figure 56F**). As no unspecific interactions were visible between any of the sequences and THC-FITC (**Figure 52** to **Figure 53**) before, this FITC should not lead to false positive measurements. The TLF was therefore deemed pure enough to be used for fluorescence polarisation assays.

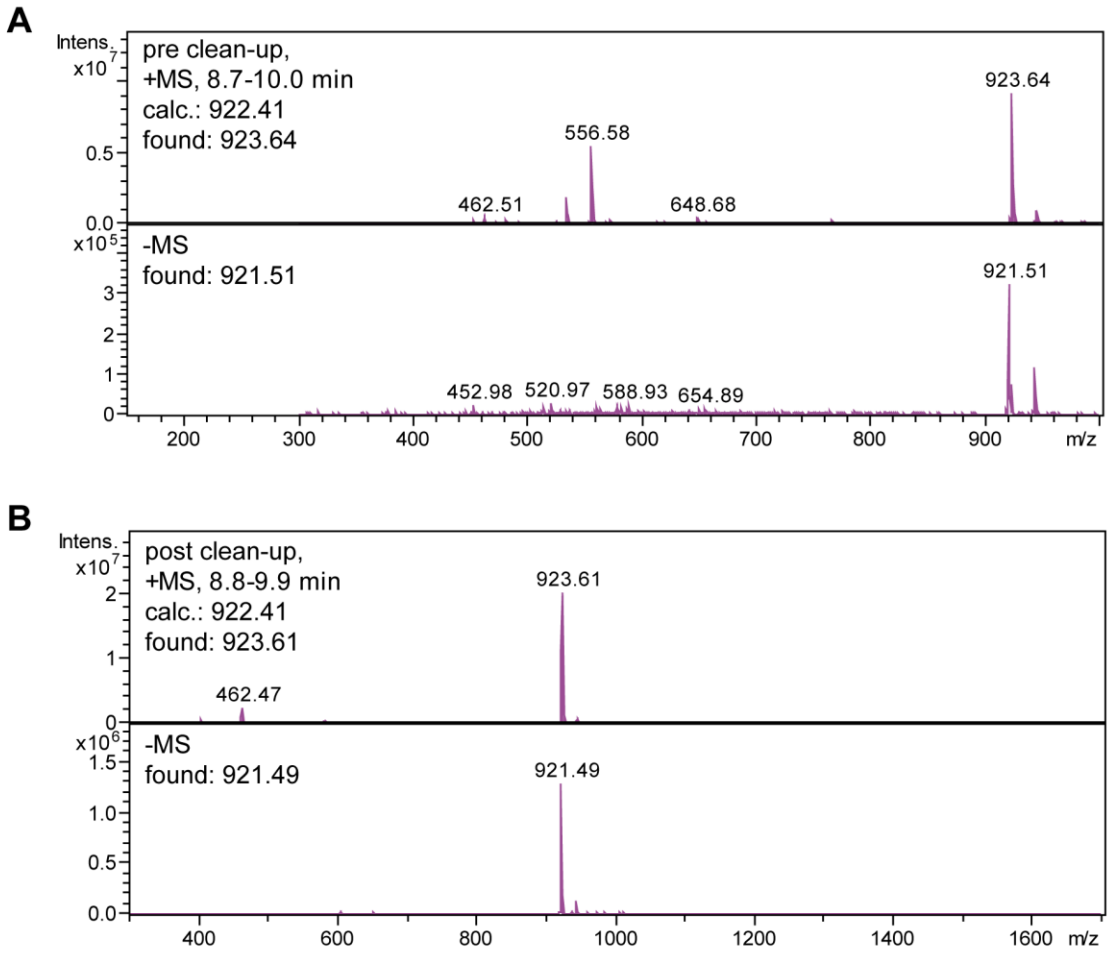


Figure 57: MS analysis of coupling of THC-L to FITC before and after clean-up
(A) Positive (upper panel) and negative (lower panel) mass spectrum of the UV-peak, pre clean-up (**Figure 56C**). The found mass matches TLF. Even though masses below 300 m/z are indicated on the x-axis, they were neither measured nor analysed during the run. (B) Positive (upper panel) and negative (lower panel) mass spectrum of the UV-peak, post clean-up (**Figure 56F**). The found mass matches TLF.

The newly synthesised TLF was characterised regarding its absorption and emission spectra, both of which closely resembled those determined for THC-FITC, including the pH-dependency of the fluorescence intensity (**Figure 43**, **Figure A 2**).

3.2.8.2 Fluorescence polarisation assays using THC-L-FITC

To evaluate the fluorescence polarisation properties of TLF, its polarisation was measured in comparison with FITC at both pH 5.4 and 7.4 (**Figure 58**). The standard deviation was very high at low concentrations of both fluorophores, but improved with increasing concentrations. As before, no pH-dependency of FITC could be determined (compare **Figure 44**). The polarisation of TLF was as low as that of FITC at pH 7.4, and only slightly higher at pH 5.4. Overall, the polarisation

was reduced in comparison with the one of THC-FITC at both pH-values (**Figure 44**).

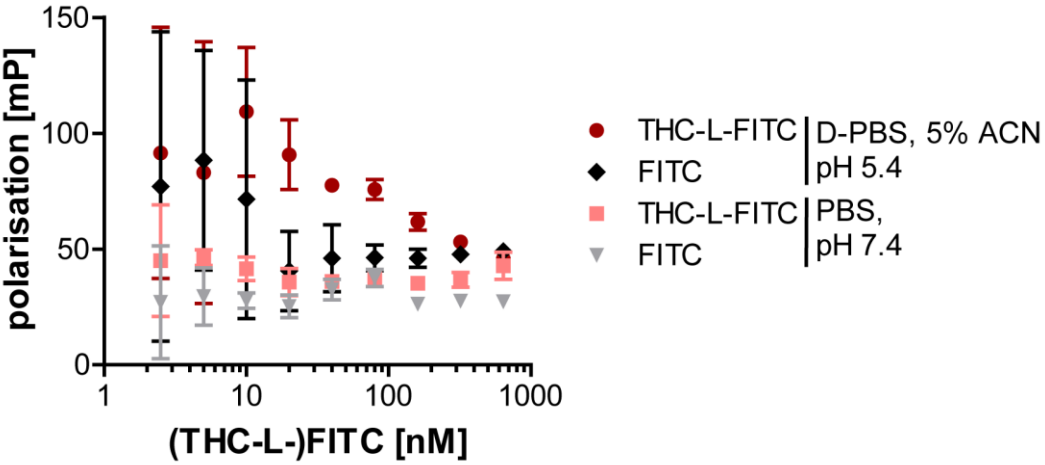


Figure 58: Fluorescence polarization of FITC and THC-L-FITC at pH 5.4 and 7.4
The fluorescence polarisation of varying concentrations of FITC (black and grey) and THC-L-FITC (shades of red) was measured in D-PBS pH 5.4 with 5% ACN (dark red and black) and in PBS, pH 7.4 (light red and grey).

TLF was then tested in a fluorescence polarisation assay using the THC-antibody as a positive and the Gal3-antibody and FITC as negative controls. As expected, FITC did not interact with any of the antibodies; Gal3 interacted with neither of the fluorophores (**Figure 59**). The combination of TLF and THC-antibody led to a concentration-dependent increase in polarisation with a determined affinity of 226.5 ± 175.8 nM.

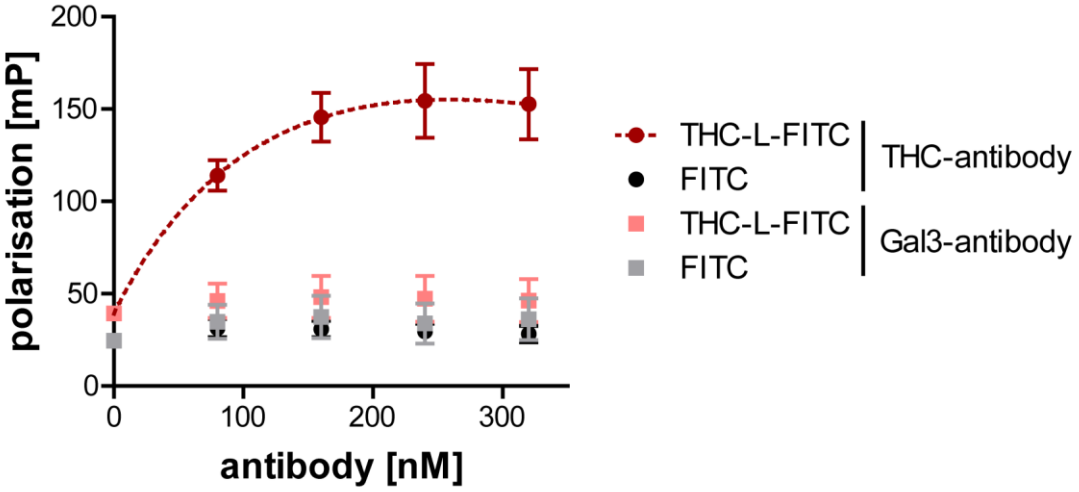


Figure 59: Fluorescence polarisation of FITC and THC-L-FITC with varying concentrations of antibodies
The fluorescence polarisation of 80 nM FITC (black, grey) and THC-L-FITC (shades of red) in the presence of varying concentrations of THC- (dots) and Gal3-antibody (squares) was measured. The K_d determined for the interaction between THC-L-FITC and the THC-antibody is 226.5 ± 175.8 nM.

Having shown that the assay can be used for affinity determination, all sequences were tested at pH 5.4. BSA was again used as a positive control and led to the expected increase in polarisation. The slight increases in polarisation obtained for B27 and C15 were within the range of the respective error bars. No sequence led to a significantly increased polarisation (**Figure 60**).

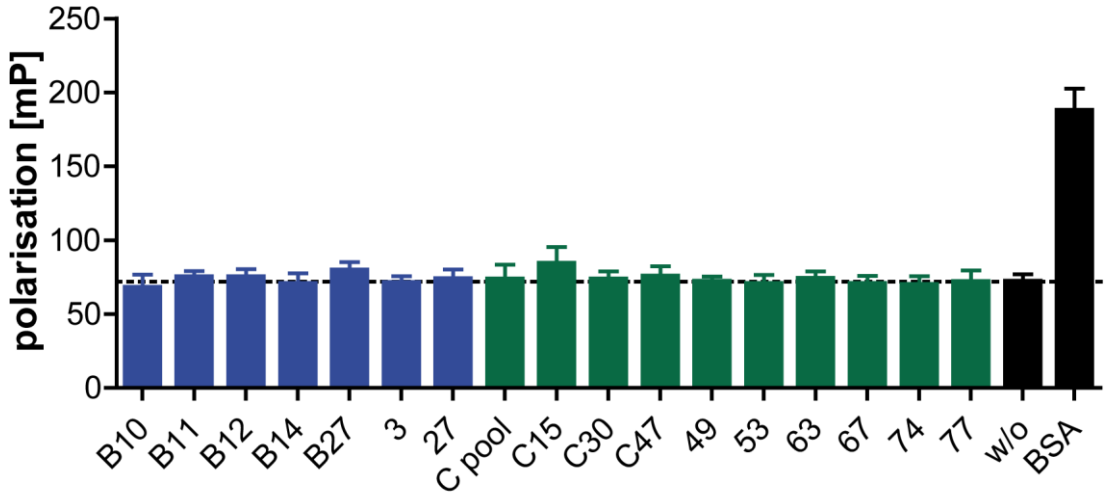


Figure 60: Fluorescence polarisation of THC-L-FITC at pH 5.4 in the presence of different sequences
The fluorescence polarisation of THC-L-FITC (80 nM) was measured in the presence of 500 nM of the different sequences, 1 mg/ml BSA, or without any sequence (w/o). BSA was used as a positive control. No sequence (w/o) was used as negative control and the dotted line represents the corresponding polarisation. Benzyl-modified DNA is depicted in blue, CF_3-modified DNA in green. C pool indicates FT2-SL that has been CF_3-modified.

To increase the sensitivity of the fluorescence polarisation assay, the sequences were biotinylated and incubated with streptavidin, thus raising their molecular weight (section 3.2.7.4). **Figure 61** shows that the addition of streptavidin again led to a slight increase in background. None of the sequences elicited a significant increase in polarisation.

Due to the lower background fluorescence polarisation and the resultant increase in assay sensitivity (**Figure 58**), the sequences that interact with THC-sepharose at pH 7.4 were also tested at this pH. Neither the unlabelled nor the biotinylated sequences in combination with streptavidin induced a detectable increase in polarisation (**Figure A 4** and **Figure A 5**).

Even though the fluorescence polarisation assay worked for the THC-antibody, none of the THC-sepharose-binding sequences could be ascertained as binding THC in solution. Therefore, we decided to perform a toggle-SELEX.

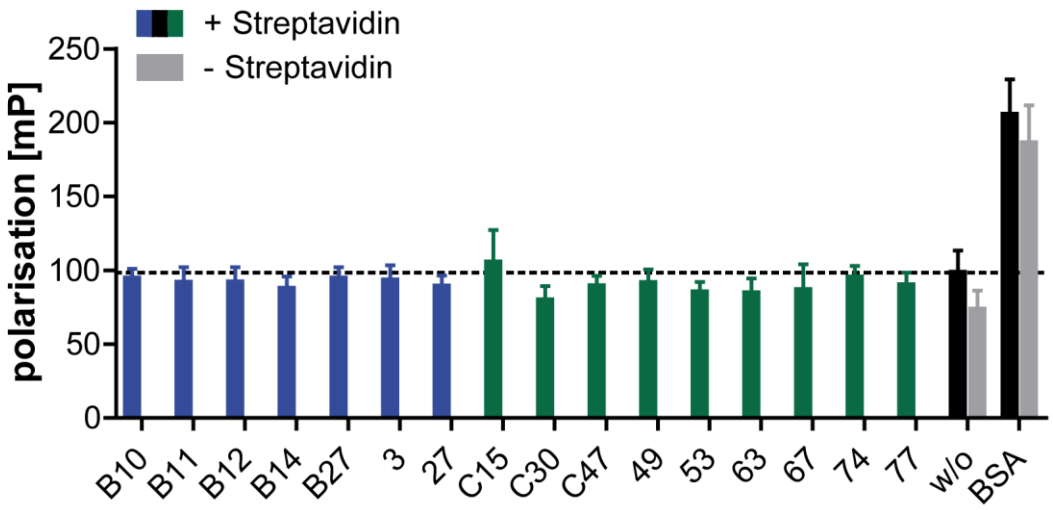
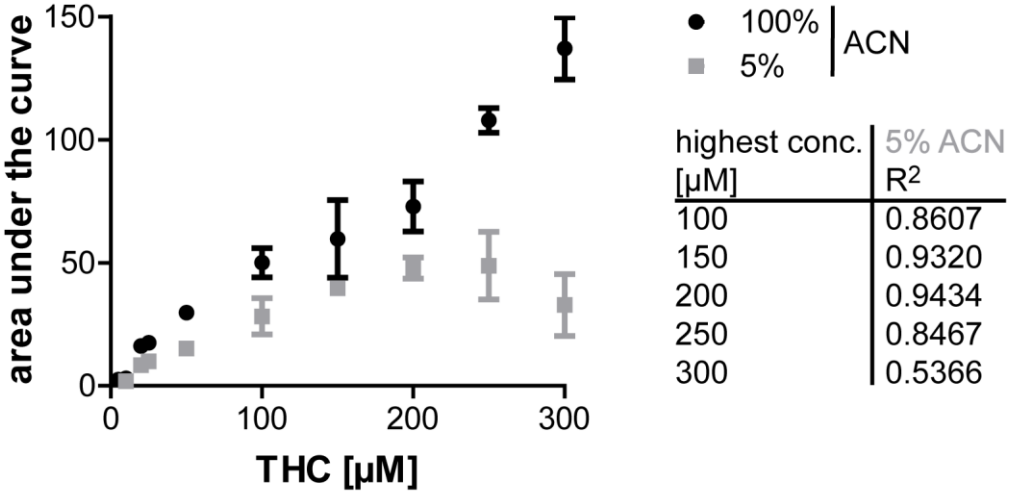


Figure 61: Fluorescence polarisation of THC-L-FITC in the presence of different biotinylated sequences and streptavidin
The fluorescence polarisation of THC-L-FITC (80 nM) was measured in the presence of 500 nM of the different, biotinylated sequences, 1 mg/ml BSA, or without any sequence (w/o). All assays but those represented by grey bars contained 750 nM streptavidin. BSA was used as a positive control. No sequence (w/o) was used as a negative control and the dotted line represents the corresponding polarisation. Benzyl-modified DNA is depicted in blue, CF_3-modified DNA in green.

3.3 Toggle-SELEX

The toggle-SELEX alternated between THC-sepharose and THCA-beads (**Figure 24**) as solid supports ("toggled").[225] In addition, the selection was performed at pH 7.0 in a PBS-buffer with a K^+-concentration of 20 mM. Both are closer to the 'average' human oral fluid[226] and might therefore facilitate the use of any selected aptamer in a roadside testing device. The same buffer was used before for the DNA-selections on THC-sepharose, SELEX 6 and 7 (**Table 4**). No Tween was added to circumvent any potential problems (section 3.2.7.2 and 4.7.2).

The THC-solubility in the selected buffer was determined using HPLC-analysis. The addition of 5% ACN resulted in a relatively good linearity up to 200 µM THC without necessitating further solvents or detergents (**Figure 62**). Therefore, this concentration was set as the upper limit for affinity elution during the toggle-selection. The control, 100% ACN, showed a slightly worse than normal linearity (0.9168).



highest conc. [µM]	5% ACN R^2
100	0.8607
150	0.9320
200	0.9434
250	0.8467
300	0.5366

Figure 62: HPLC-measurement of THC-concentrations in PBS
HPLC-analysis of THC-concentrations between 10 and 300 µM in 100 (black) and 5% (grey) ACN. Depicted in the table are the R^2-values of the 5% ACN-sample, depending on the highest concentration that has been included in the linearity-calculation.

It is possible that the disability to identify sequences that recognise THC in solution is due to the high local concentration of THC on the sepharose, which cannot be reached in solution. To prevent the appearance of this problem in the new toggle-SELEX, three different kinds of THC-sepharose were prepared: One coupled in the presence of the normal excess of THC, which is from here on called 1:1-sepharose. One coupled using only 1/10 the amount of THC that can be coupled according to the available chemical functionalities on the surface[227]: 1:10-sepharose, and one using only 1/100 the amount: 1:100-sepharose. **Figure 63** details the respective intensities as quantified by the THC-antibody. All three THC-sepharoses showed stronger signals than the unmodified sepharose (**Figure 63A**), indicating that the coupling reaction was successful for all three preparations. The 1:1-sepharose gave the strongest signal by a factor of about 2 in comparison with the other two THC-sepharoses (**Figure 63B**). However, no clear difference in signal intensity was detectable between the 1:10- (**Figure 63C**) and 1:100-sepharose (**Figure 63D**). Since they both contain less THC than the 1:1-sepharose, all three were used for the selection.

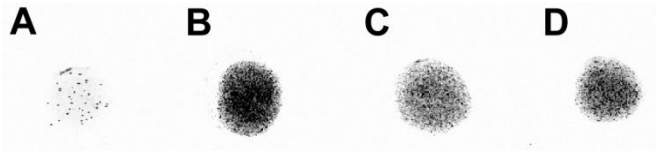


Figure 63: Detection of different amounts of THC coupled to epoxy-activated sepharose
Detection of THC using a THC-antibody. (A) Background signal on non-modified epoxy-activated sepharose. (B) 1:1-coupled sepharose: Signal for THC on epoxy-activated sepharose that has been coupled to THC in the presence of an excess of THC. (C) 1:10-coupled sepharose: Signal for THC that has been coupled using 1/10 THC in comparison with the maximal available chemical functionalities on the sepharose. (C) 1:100-coupled sepharose: Signal for THC that has been coupled using 1/100 THC in comparison with the maximal available chemical functionalities on the sepharose. The quantified signal intensities were 344% of the background-signal for the 1:1-sepharose, 165% for the 1:10- and 184% for the 1:100-coupled sepharose.

Table 19 details the selection conditions used during the toggle-SELEX. First, ten selection cycles were performed. During these, THC-sepharose and THCA-beads were used alternatingly as solid support. The amount and coupling grade of THC-sepharose and the amount of THCA-beads was reduced over the course of the selection, resulting in a reduction in THC(A). The incubation time was reduced from 1 h to 5 min, and the washing steps were increased from about 6.5 min to 3 h. During the first three selection cycles as well as during all cycles performed on THC-sepharose, heat elution was performed for 3 min at 80°C. From cycle four on, affinity elution with decreasing concentrations of THC (starting at 200 µM and stopping at 25 µM THC) were performed on the THCA-beads. Less PCR-cycles were needed to amplify the DNA if the samples were incubated on the THCA-beads than on the THC-sepharose, indicating a higher DNA-concentration before amplification. Apart from this observation, no clear decrease in PCR-cycles was detectable from selection cycle one to ten when evaluating the two different matrices individually.

Table 19: Selection conditions for the toggle-SELEX

SELEX cycle	DNA [pmol]	blocked tosyl-beads	THCA-beads	unmodified sepharose	THC-sepharose	incubation time	max. THC(A) [nmol]	wash steps	elution	PCR-cycles
1	n.d.	-	-	-	25 µl 1:1	1 h	475 to 1000	6.5 min 3 mL	3 min @ 80°C	B: 17 C: 17
2	B: 28.4 C: 19	-	50 µl	-	-	1 h	50 to 100	30 sec 2x 5 min (10 min)	3 min @ 80°C	B: 13 C: 13
3	B: 25.6 C: 26.3	-	-	-	25 µl 1:10	30 min	47,5 to 100	15 min 10 mL	3 min @ 80°C	B: 19 C: 19
4	B: 5 C: 10.6	-	25 µl	-	-	15 min	25 to 50	30 sec 4x 5 min (20 min)	200 µM THC 40 nmol	B: 15 C: 15
5	B: 5.7 C: 27.4	-	-	-	10 µl 1:10	10 min	19 to 40	30 min 8 mL	3 min @ 80°C	B: 20 C: 17
6	B: 23.2 C: 51.2	-	10 µl	-	-	10 min	10 to 20	30 sec 9x 5 min (45 min)	100 µM THC 20 nmol	B: 16 C: 17
7	B: 33 C: 27.3	-	-	-	25 µl 1:100	5 min	4,75 to 10	1 h 13 mL	3 min @ 80°C	B: 19 C: 18
8	B: 24.8 C: 24	-	5 µl	-	-	5 min	5 to 10	30 sec 9x 10 min (1.5 h)	50 µM THC 10 nmol	B: 14 C: 14
9	B: 15.8 C: 28.8	-	-	-	10 µl 1:100	5 min	1,90 to 4	2 h 28 mL	3 min @ 80°C	B: 20 C: 17
10	B: 22.3 C: 27.4	-	2,5 µl	-	-	5 min	2,5 to 5	30 sec 12x 15 min (3 h)	25 µM THC 5 nmol	B: 14 C: 12
11	B: 10.55 C: 14.86	-	-	-	10 µl 1:100	5 min	1,9 to 4	3 h 21 mL	3 min @ 80°C	B: 19 C: 16
12	B: 15.96 C: 10.6	2x 50 µl 15 min each	2.5 µl	-	-	5 min	2,5 to 5	30 sec 12x 15 min (3 h)	25 µM THC 5 nmol	B: 10 C: 12
13	B: 5.69 C: 5.05	-	-	2x 50 µl 15 min each	10 µl 1:100	5 min	1,9 to 4	3 h 30 mL	3 min @ 80°C	B: 22 C: 22
14	B: 10.87 C: 3.07	-	-	2x 50 µl 15 min each	10 µl 1:100	5 min	1,9 to 4	3 h 30 mL	3 min @ 80°C	B: 19 C: 19

B/C: Selection with benzyl-/CF_3-modified DNA. n.d. = not determined.

After ten selection cycles, the DNA from the SL, as well as from selection cycles 4, 8, and 10 was radioactively labelled and tested for interaction with both the THC-sepharose (1:1) and the THCA-beads as well as the respective unmodified solid supports. The results on the sepharose for the selection with benzyl- and CF_3-modified DNA are depicted in **Figure 64A** and **B**, respectively. B12 and C15 were radioactively labelled in parallel and used as positive controls. Their absolute values cannot be compared to the others as the assays were performed in different buffers, but both showed that the two assays worked as expected. For the samples from the new toggle-SELEX, neither the benzyl- nor the CF_3-modified samples showed an increase in interaction with THC-sepharose in comparison to either their interaction with the unmodified sepharose or with the modified DNA from the SL.

The results on the THCA-beads for the selection with benzyl- and CF_3-modified DNA are shown in **Figure 64C** and **D**, respectively. For both modifications, roughly 60% of the DNA from the SL stuck to the THCA-beads. This high background binding decreased to about 20% for the DNA of selection cycle four of the selection with benzyl-modified DNA and stayed like this for the higher selection cycles (**Figure 64C**). The samples from the selection with CF_3-modified DNA showed a slight reduction of background binding over the selection process, but about 40% of the DNA from the final selection cycle remained on the beads.

In case ten selection cycles were not sufficient to enrich THC-binding clickmers, an additional four selection cycles (11 to 14) were added and are detailed in **Table 19**. For these, the incubation and washes were kept constant. Cycles 11, 13, and 14 were performed on THC-sepharose; cycle 12 on THCA-beads. A negative selection step was added to deplete matrix-binding sequences during selection cycles 12 to 14. For this, the DNA from the previous selection cycle was incubated with the unmodified solid support before non-bound sequences were added to the THC(A)-containing matrix. As for the first ten selection cycles, no decrease in the number of PCR-cycles needed for amplification was detectable (**Table 19**).

To test if the additional selection cycles led to an enrichment of THC-binding sequences, DNA from the SL, selection cycle 10, and 14 was radioactively labelled and incubated with THC- and unmodified sepharose. Due to the high background binding on the THCA- and methylamine-blocked tosylated beads apparent in the previous assay (**Figure 64C**,**D**), the interaction with this matrix was not tested.

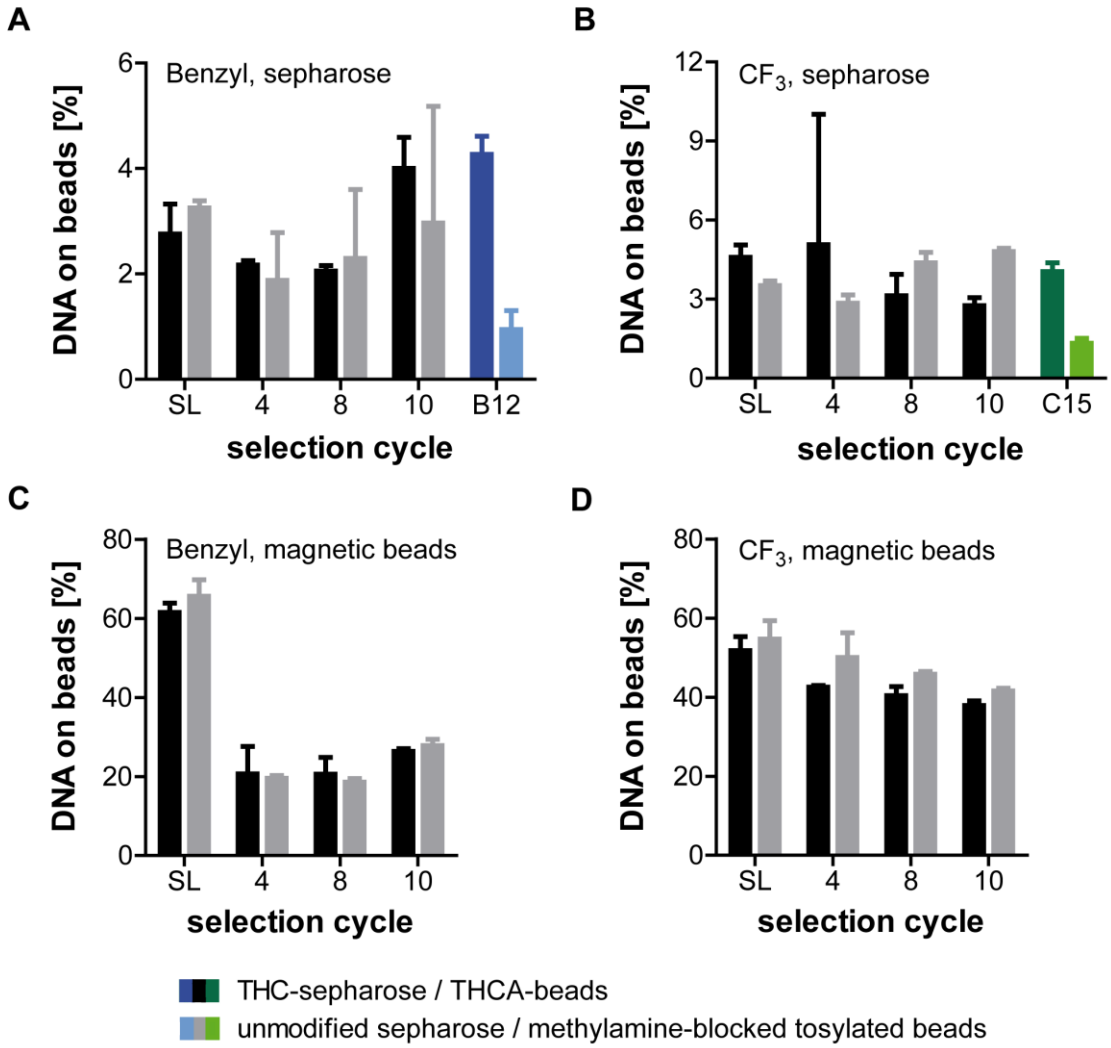


Figure 64: Interaction analysis after ten selection cycles of toggle-selections for THC-binding clickmers
^{32}P-labelled DNA with benzyl- (A, C) and CF_3- (B, D) modification was incubated with sepharose (A, B) or magnetic beads (C, D). THC(A)-modified solid supports are depicted in black, dark blue, and dark green. Non-target solid supports are depicted in grey, light blue, and light green. After two washing steps, the amount of radioactivity on the solid support was quantified. In blue and green are sequences from the first successful click-SELEX that are known to be interacting with THC-sepharose as positive controls. (A) DNA from the SL as well as selection cycles 4, 8, and 10 from the selection with benzyl-modification on sepharose. B12 was used as a positive control. (B) DNA from the SL as well as selection cycles 4, 8, and 10 from the selection with CF_3-modification on sepharose. C15 was used as a positive control. (C) DNA from the SL as well as selection cycles 4, 8, and 10 from the selection with benzyl-modification on magnetic beads. (D) DNA from the SL as well as selection cycles 4, 8, and 10 from the selection with CF_3-modification on magnetic beads.

Figure 65 depicts the results for the samples from both selections. B11 and C30 were tested in parallel as positive controls. Both clearly distinguished between THC- and unmodified sepharose. As their selection buffer differed from the one used for the new selections, the absolute values of the samples cannot be compared. Neither of the samples from selection cycles 14 differentiated between THC- and unmodified sepharose.

Therefore, both toggle-selections were unsuccessful in enriching THC(A)-binding sequences.

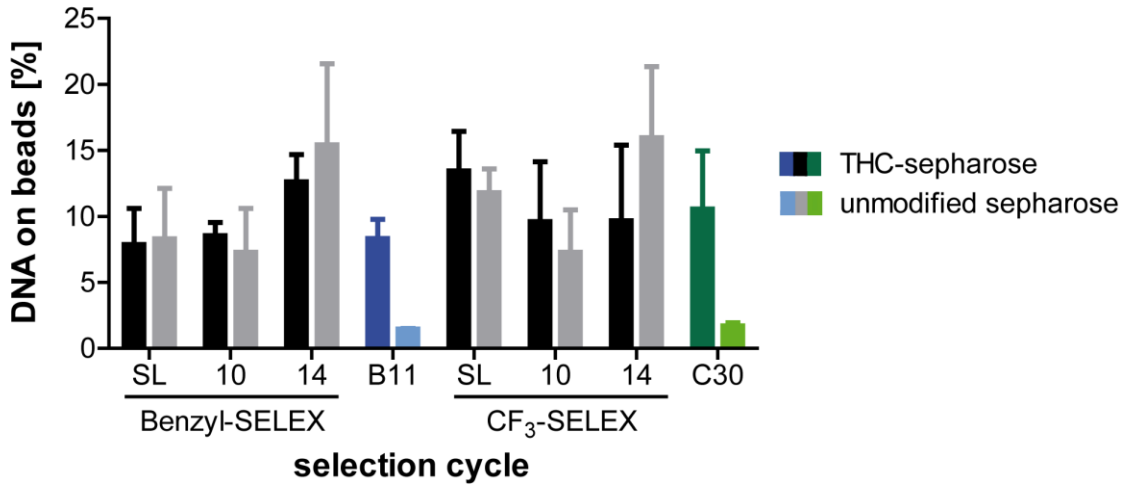


Figure 65: Interaction analysis after 14 selection cycles of toggle-selections for THC-binding clickmers
^{32}P-labelled DNA with benzyl- and CF_3-modification was incubated with THC- (black, dark blue, and dark green) or unmodified sepharose (grey, light blue, and light green). After two washing steps, the amount of radioactivity on the sepharose was quantified. In blue (benzyl-modified) and green (CF_3-modified) are sequences from the first successful click-SELEX that are known to be interacting with THC-sepharose as positive controls. DNA from the SL as well as selection cycles 4, 8, and 10 from the toggle-SELEX with benzyl-modification and CF_3-modification was tested.

4 Discussion

Section 4.1 deals with the different immobilisation techniques used for THC. It is followed by the discussion of all THC-selections performed with non-nucleobase-modified nucleic acids, which stresses possible reasons for the failure of those selections (section 4.2). The two initial click-selections and the identified sequences are discussed in section 4.3. The validation of the use of NGS for aptamer or clickmer identification is evaluated in section 4.4. To evaluate the binding of the THC-sepharose-binding sequences in solution, efforts were undertaken to determine the solubility of THC. These are elucidated on in section 4.5. Since no binding in solution could be determined, an affinity re-selection was performed. The results and the identified sequences are discussed in section 4.6. The setup of the fluorescence polarisation assays that were used to determine interaction with THC in solution is detailed in section 4.7. Section 4.8 contains the discussion of the toggle-SELEX.

4.1 Immobilisation of THC(A)

Three different immobilisation techniques were used for THC(A) during the course of this thesis: THCA on amine-functionalised magnetic beads, THC on epoxy-activated sepharose, and THCA on tosylated magnetic beads, which were prepared by Malte Rosenthal.[d]

For the first selections, THCA was immobilised by coupling to amine-functionalised magnetic beads. THCA was chosen instead of THC due to the additional carboxyl group, which enabled immobilisation without loss of the hydroxyl group. Another option that would have offered the same advantage was the THC-metabolite, THC-COOH (section 1.1.4.1, **Figure 5A**). The choice in favour of THCA was based on the fact that the developed test would also be useable in urine. If the selected aptamer differentiated between THC and THC-COOH, the test would still be able to indicate recent consumption of THC and therefore current impairment of the driving skills.

THC was coupled to epoxy-activated sepharose for the subsequent selections, because none of the selections on THCA-functionalised amine beads had led to the identification of THC-binding aptamers. Sepharose has been the predominant choice as immobilisation matrix for the selection of small molecule-binding aptamers.[147] For small molecule targets, its density of chemical functionalities on the surface results in a higher target molecule density than possible on magnetic beads. For example, the amine-functionalised dynabeads contain 5-10 nmol functional groups on the surface, the epoxy-activated sepharose 0.95-

[d] Malte Rosenthal, PhD thesis in preparation, Bonn, 2018/2019

2 µmol.[147,228] This becomes apparent when comparing the results in **Figure 14** and **Figure 20**: Much more THC can be detected on the sepharose than THCA on the magnetic beads. However, it cannot be excluded that this results from differences in the antibody-affinity for the differentially coupled THC(A). To exclude this possibility, trials to couple THC to three different types of epoxy-modified magnetic beads were performed using the protocol established for the coupling of THC-sepharose. None of them was successful (data not shown, see also section 4.8).

Malte Rosenthal coupled THCA to tosylated magnetic beads.[e] These beads were used to ascertain the specificity of the benzyl-modified THC-sepharose-binding sequences (**Figure 25**) and for the toggle-SELEX (section 3.3). In contrast to the amine beads, they are hydrophobic.[228] This facilitates the selection process as the unspecific binding of DNA/RNA to the beads is reduced.

Even though the toggle-SELEX was as unsuccessful as all selections performed on the THCA-modified amine-beads, Malte Rosenthal selected a benzyl-modified THC(A)-binding clickmer on the THCA-modified tosylated beads.[e] The failure of all selections on magnetic beads presented in this thesis can therefore not be due to the matrix alone and the corresponding low target molecule density on the surface. Possible reasons are discussed in the subsequent section.

4.2 THC-selections with non-nucleobase-modified nucleic acids

Overall, seven different selections for THC-binding aptamers were performed using non-nucleobase-modified nucleic acids. None of them enriched THC-binding sequences.

As the selected aptamer was supposed to be implemented in an oral fluid-based roadside test and oral fluid is known to contain nucleases, RNA was deemed too instable.[229] 2'F-RNA has a higher resistance towards degradation by nucleases and was therefore used for all RNA-selections (**Table 3**).[176]

To increase the selection stringency, competitors such as salmon sperm DNA or polyanions (heparin) were added during incubation of the library with the target molecule in all selections. These reagents unspecifically interact with both the target molecule and the matrix. They thereby compete with unspecifically binding sequences that are part of the selection library. The unspecific binders thus do not bind to the target molecule and are removed from the library during the wash steps. Only specifically binding sequences are amplified.

For SELEX 1 and 2, the same linear, unstructured 2'F-RNA-library was used. The selections were performed in parallel with a PBS-based selection buffer. The only

[e] Malte Rosenthal, PhD thesis in preparation, Bonn, 2018/2019

difference between the two setups was the elution technique: During SELEX 1, 2'F-RNA was recovered by heat, which should unspecifically denature all folded nucleic acids and allow their recovery. For SELEX 2, 5 mM THC was used in combination with 7% Cremophor EL, a polyethoxylated castor oil, as recommended by PD Dr. Andras Bilkei-Gorzo to keep the THC solubilised (**Table 3**).[f] None of them enriched THC-binding sequences in 15 selection cycles (**Figure 15**). While no clear reason for SELEX 1 is apparent, SELEX 2 most probably did not succeed due to the high amount of Cremophor EL during the recovery of the DNA. The CMC of Cremophor EL is reported as 0.02%.[230] Accordingly, all THC molecules are probably inside micelles and not able to be bound by potential aptamers. Instead, 2'F-RNA that dissociated from the THCA-beads has been recovered and amplified. Any initially present THC-binding sequences have thus most likely been depleted.

SELEX 3 to 5 used a 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)-based buffer. Instead of recovering the DNA or 2'F-RNA by heat or affinity elution before PCR, the beads were added to the PCR. During the denaturation phase of the PCR, the DNA or 2'F-RNA is denatured and can serve as template afterwards. The three selections differed regarding the library used for selection: SELEX 3 was performed with a prestructured 2'F-RNA-library. The random region contains G-stretches that predispose the library to form G-quadruplex structures. The DNA-libraries used for SELEX 4 and 5 have random regions of different lengths to enable formation of different three-dimensional structures (**Table 3**).

However, all three selections did not enrich THC-binding sequences in ten selection cycles without obvious reason (**Figure 15**). THC is thermostable[231] and remains unchanged throughout the PCR. Potential THC(A)-recognising sequences might have refolded during annealing and elongation and then re-associated with THC. They would then be unavailable for amplification. G-quadruplex structures are known to be exceptionally stable with melting temperatures reaching 50°C for sequences as short as 15 nucleotides.[232] Accordingly, this explanation for the unsuccessfulness of the selection is most likely for SELEX 3.

Due to the higher density of functional groups on the surface that might allow a higher local target molecule concentration (section 4.1), the matrix was exchanged for THC-sepharose for SELEX 6 and 7.

Both selections (SELEX 6 and 7) were performed in a PBS-buffer at pH 7.0 and with an increased concentration of KCl to emulate oral fluid.[226] A combination of heat (10 min at 65°C) and denaturing agents (urea and ethylenediaminetetraacetic acid (EDTA)) was used to recover binding sequences.

[f] Personal communication with PD Dr. Andras Bilkei-Gorzo, Institute of Molecular Psychiatry, University of Bonn

The selections differed concerning the utilised DNA-libraries: The library for SELEX 6 was unstructured, whereas SELEX 7 made use of the same prestructured library as SELEX 3 that is supposed to preferentially form G-quadruplex structures. Here, the library was used as DNA, not as 2'F-RNA (**Table 4**). None of the two selections enriched THC-binding sequences in ten selection cycles without obvious reasons (**Figure 21**).

Overall, seven out of seven selections failed. A reason for the failure is only obvious for SELEX 2. All others were unsuccessful despite the differences concerning libraries, buffer conditions, and elution and immobilisation techniques. For proteins, it is known that about 30% of all selections do not lead to aptamer identification[180], but even with equally low success rates, one of the selections should have worked. After all, DNA-aptamers have been successfully selected for three hydrophobic hormones, progesterone[233], 17ß-estradiol[234-236], and cortisol.[237] All three are less hydrophobic than THC according to their logP-values (THC: 5.65[213], progesterone: 3.87[238], estradiol: 4.01[239], cortisol: 1.61[240])[208]. Accordingly, they are probably easier targets for the hydrophilic non-nucleobase-modified nucleic acids. Nevertheless, DNA-aptamers have also been selected for polychlorinated biphenyls (PCBs), which are more hydrophobic than THC (logP-values: PCB 77: 6.72[241], PCB 72: 6.4[242], PCB 106: 6.9[243]). They do not differentiate well between the different PCBs, but achieve affinities in the nanomolar range and were able to bind the target molecules in solution.[244,245] The ability of PCBs to interact with aptamers has been hypothesised to result from the aromatic groups.[245,246] While THC also contains an aromatic ring, two of the three rings are non-aromatic (**Figure 1A**) and therefore unable to interact with the nucleobases via π-π-stacking. The two selections (SELEX 3 and 7) performed with the G-quadruplex library were chosen in the hope that these interactions might suffice to enable the selection of THC-binding aptamers. As they were not successful, π-π-stacking with non-modified nucleobases is – at least under the conditions tested – not sufficient for the successful selection of aptamers for THC.

While it is entirely possible that different selection conditions and libraries might lead to the successful selection of aptamers for THC, it has to be deemed unlikely. Our approach of introducing hydrophobic modifications into the DNA to imitate the interaction of THC with the CB1 (section 1.1.2, **Figure 3**) led to the immediate successful selection of THC-sepharose-binding sequences in two different selections. We therefore conclude that non-nucleobase-modified DNA and 2'F-RNA is highly unlikely to form structures that can stably interact with THC.

4.3 Click-SELEX on THC-sepharose

The click-SELEX for GFP published by Tolle *et al.* was the only successful use of the technique at the onset of this part of the thesis.[205] To maximise the chances of success, the selections were performed as similar to the GFP-SELEX as possible, including the use of the same library and selection buffer, regardless of its pH of 5.4. A major change was the inclusion of a clicked competitor from the first selection cycle on. Due to the additional chemical functionalities of clicked DNA, the usual competitors (section 4.2) are not sufficient to remove the majority of unspecifically binding DNA. This was observed in the first click-selection, where only the late addition of a double-stranded click-modified DNA-competitor lead to the enrichment of the two clickmers.[152] For the selections presented in this thesis, a solid phase-synthesised single-stranded DNA of equal length as the random region of the selection library was used (sequence: 5'-N_{42}-A-3'). This click-modified DNA competitor contains EdU and is modified with the same azide as the selection library. Therefore, it is able to imitate any potential unspecific interaction the selection library is capable of and can compete with it. This reduces the amount of unspecifically binding DNA that is amplified and thus increases selection stringency.

The number of PCR-cycles necessary for amplification decreased from 18 to 15 between selection cycles nine and ten for both the selection with benzyl- as well as CF_3-modified DNA (**Table 5**). This is often interpreted as an indication that the selection has been successful. The interaction analysis of both click-selections as depicted in **Figure 22A** and **Figure 26A** shows how unreliable this decrease in PCR-cycles is as readout of selection success: Both selections were already visibly enriched for THC-binding sequences in selection cycle eight, before any apparent decrease of PCR-cycles.

The fact that the benzyl- and CF_3-modified-selections were successful on the first try after the failure of seven conventional selections proofs the merit of the technique for difficult targets. To the best of our knowledge, this is the first example of the successful selection with nucleobase-modified libraries for small molecule targets. More selections have to be performed to determine the increase in selection success as SomaLogic® has done for protein targets (30% without, 84% with nucleobase-modifications).[180] Nonetheless, nucleobase-modifications seem to be a worthwhile possibility if conventional selections for small molecule aptamers fail.

In addition to the nucleobase-modification *per se*, click-SELEX enables the modular and strategic choice of the modification. For THC, we chose benzyl and CF_3 to imitate the interaction between THC and its receptor (section 1.1.2, **Figure 3**, **Table 2**). Gold *et al.* have shown that, at least for protein targets, not all modifications work equally well.[180] It stands to reason that this is also true for click-SELEX and that the thoughtful choice of azide is necessary for selection

success. A first indication for this is given by the fact that the affinity of two different GFP-binding clickmers is influenced by the clicked-in modification, some of which completely preclude binding to the original target molecule.[152,205] The interaction of the SL with THC-modified beads has also been shown to vary depending on the chosen modification.[152]

Both selections performed with benzyl- as well as CF_3-modified DNA were visibly enriched from cycle 8 on (**Figure 22A** and **Figure 26A**). The selections were performed in parallel with identical conditions (apart from the number of PCR-cycles). The ratio of binding between THC- and unmodified sepharose is higher for the benzyl-modified selection (**Figure 22A** and **Figure 26A**). This is an indication that of the two analysed libraries, benzyl seems to be the better choice of modification.

The DNA from both selections does not interact with sepharose if it is not modified with THC (**Figure 22B** and **Figure 26B**). This is further proof of the hypothesis that the nucleobase-modifications are necessary for the selection success and not only the different buffer or elution conditions.

The THC-sepharose-binding sequences that were identified from both selections interact with THC-sepharose to different degrees. The strength of the interaction is independent of the number of modifications (**Table 6** and **Table 7**, **Figure 23** and **Figure 27**). As B10 has as many or more modifications than the THC-binding sequences B14, B13, and C30, it is obvious that the modifications alone are not responsible for the interaction with THC. It has to be a combination of the modifications and the three-dimensional structure the DNA folds into. As of now, no clickmer structures have been published. Due to the unsuccessfulness of all selections with non-nucleobase-modified nucleic acids, it seems likely that – as for most known SOMAmer structures (section 1.2.3.2) – the modification directly interacts with the target molecule. Confirmation of this hypothesis has to await structural studies of clickmers.

In the selection with CF_3-modified DNA, the least abundant sequence, C15, also interacts with THC-sepharose to a slightly lower degree in comparison with the two more abundant sequences (**Figure 27A** and **Table 6**). No such correlation is apparent for the benzyl-modified sequences (**Figure 23** and **Table 7**). Knowledge of the affinity constants might explain the frequency of the sequences and give an insight into the differences between Sanger and NGS. As it was not possible to elucidate the k_D during the course of this thesis, the question will have to remain unanswered for now.

In the specificity test, B11 is the only benzyl-modified sequence that clearly shows interaction with the THCA-beads (**Figure 25**). Therefore, B11 is less specific than the other sequences that only recognise THC-sepharose.

No preliminary specificity test was performed for the CF_3-modified sequences. Because the binding of CF_3-modified DNA from selection cycle 1 to the THCA-beads was as high as for the benzyl-modified DNA from selection cycle 1 (data not shown, **Figure 25**), no negative control for such an assay was available.

4.4 Verification of the NGS approach

Although NGS has been used for aptamer identification before[156-158], discrepancies between the data obtained by Sanger and next-generation sequencing (**Table 6** and **Table 7**) caused us to investigate the actual error rates in our NGS approach.

Analysis of the impact of sample preparation disclosed that the index-PCR had no significant effect on the mutation rates (**Table 8**, **Figure 28**, and **Figure 29**). In contrast, the presence of EdU in the template led to significantly increased error rates (**Table 8**, **Figure 28**, and **Figure 29**). As EdU is replaced by thymidine during index-PCR and therefore before the actual sequencing, the most likely reason for the increased mutation rates is a reduced PCR-fidelity due to the modification. This goes along with a publication by Oyola *et al.* that showed that the PCR-free preparation of AT-rich sequences leads to a significant reduction of mutations.[247] Both AT-rich sequences and those containing EdU apparently increase mutation rates during PCR.

The reduced PCR-fidelity has to be taken into account for click-SELEX in general, despite the incorporation of EdUTP instead of TTP in the PCRs during the selection process. On one hand, the resulting mutations during PCR might be helpful to generate better variants of target-binding sequences during the selection.[248-250] On the other hand, the reduced fidelity might decrease the binding affinity of PCR-amplified clickmers due to the generation of non-binding mutants. Nevertheless, no such problems are described in literature. The only study comparing a solid-phase synthesised with a PCR-amplified clickmer describes no significant differences between the two.[211] Regardless, the identified fidelity-reduction should be kept in mind when working with EdU-modified PCR-templates.

In addition to the strong impact of EdU, an increase in mutation frequency over the length of the analysed sequence was observed (**Figure 28**) and the nucleotides, into which mutations occurred, reflected the composition of the original C12-sequence (**Figure 29B** and **C**). The increase in mutation frequency over the length of the analysed sequence has been published before for samples analysed by SBS, even though the effects were less pronounced than observed in this thesis.[166,168,173,251]

The analysis of repetitive sequences revealed a significantly increased mutation frequency to the subsequent nucleotide (**Figure 30C** and **D**, **Figure 31**, and **Table 10**) that depended on the number of identical nucleotides in a row (**Figure 32**). All these results could be explained by the occurrence of pre-phasing effects. These lead to an apparent shortening of the sequence, as one or more nucleotides are incorporated without readout and therefore skipped in the output sequence (**Figure 7** and section 1.2.2). Analysis of the 25 most abundant sequences in all samples showed that singe-nucleotide deletions were the most frequent mutation type (**Table A 3** to **Table A 16**). This contradicts literature that claims single-nucleotide substitutions as the most prevalent kind of mutation during SBS.[165-167] We concluded that the majority of these deletions resulted from pre-phasing and decided to omit all deletion-containing, shortened sequences from further analyses.

Exclusion of the shortened sequences reduced the error rate by an average of 80%, although only about 5% of all sequences were omitted (**Table 11**). Single-nucleotide deletions as generated by pre-phasing lead to the definition of all following nucleotides as mutated, which results in the observed, high error rates (**Table 8** and **Table 9**). Software solutions for the reduction of phasing effects are available, but do not reach the same percentage of non-mutated sequences as omission of all shortened sequences (77% and 94% non-mutated sequences for All Your Base and omission of shortened sequences, respectively).[171] Nonetheless, this approach is only feasible for sequences of a known length like SELEX samples. For more complex samples, software solutions remain the best option available.

After omission of the shortened sequences, the average error rate (without EdU_Pwo) of 0.24 ± 0.06% is in the lower range of error rates published for samples analysed with different Illumina sequencers (**Table 20**). All values are equal or well below those reported for Sanger sequencing (between 2 and 10%).[165,252]

Table 20: Error rates on Illumina sequencers

publication	instrument	error rate [%]	comments
Fox *et al.* [165]	HiSeq2000	0.1	
Fox *et al.* [165]	MiSeq	0.1	
Dohm *et al.* [166]	1G	0.3	at the start of sequence, increases due to phasing effects
May *et al.* [167]	MiSeq	0.21-2.6	depending on the reference sequence; substitutions only
Kelley *et al.* [168]	not disclosed	0.5-2	

The mutation rates from one nucleotide to another as displayed in **Table 13** confirm the hypothesis that the high mutation rates of EdU_Pwo result from the presence of EdU in the template as thymidine shows much higher mutation rates than any of the other nucleotides. In addition, thymidine is the nucleotide that mutates least when analysing the average of the other samples.

The similarity of their fluorophores' emission spectra leads to a relatively high probability of A and C as well as G and T being miscalled for each other on Illumina sequencers.[173] This is only partly represented by our data: C to A has the highest mutation rate (0.13% in average) and G to T is amongst the highest rates (0.11% in average). In contrast, A to C and T to G are amongst the lowest mutation rates with 0.04 and 0.05%, respectively. Since the mutation rates between identical, re-sequenced samples (FT2_GATC and FT2_GATC_II as well as FT2_G4A4T4C4 and FT2_G4A4T4C4_II) vary by a factor of up to ten, the datasets obtained in this thesis are not sufficient for the comparison of single-nucleotide conversions.

We also evaluated the effect the omission of shortened sequences had on SELEX samples. As patterns instead of single sequences were analysed, the frequencies remained largely unchanged (**Figure 34**). This is because shortened sequences are excluded from every pattern. While the number of sequences a pattern consists of decreases, so does the overall number of analysed sequences. The resulting frequency is barely influenced. Together with the low error rate in comparison to Sanger sequencing, this means that NGS is perfectly useable for the analysis of (click-)SELEX samples. Omission of shortened sequences, which was not performed for the click-selections described in this thesis, is not required when analysing patterns instead of single sequences.

4.5 Solubility determination of THC

Knowledge about THC-solubility was necessary to perform affinity determination of the identified THC-sepharose binding sequences. The initial HPLC-measurements showed that THC-solubility is below 10 µM in D-PBS, pH 5.4 with 10% DMSO (**Figure 35A**). The addition of 0.1% Tween-20 led to values similar to those measured in 100% DMSO, which should keep THC well solubilised (section 3.2.5). While other detergents would most likely also ensure THC-solubility in aqueous buffers, Tween was the obvious choice as it was part of the selection buffer. Because of its apparent ability to keep THC in solution (**Figure 35A**), no additional detergents were tested.

The initial analysis also showed that the area under the curve alone is not a good readout of THC-solubility as it varies depending on the solvent used. This is obvious from the differences between the values obtained for 100% DMSO and 100% ACN (**Figure 35B**), both of which should be able to solubilise THC in

concentrations far above those used in the assay (section 3.2.5). Instead, linearity of the resulting curves was chosen as readout of THC-solubility for all assays.

Although 10% DMSO with 0.1% Tween was promising according to **Figure 35A**, the lowest amount of organic solvent possible should be used to minimise the risk of influences on folding of the modified DNA-sequences.[215] Tests with 5% organic solvent and 0.1% Tween demonstrated that DMSO was the worst option, while both ACN and EtOH led to usable linearities (**Figure 35B**). Due to its slightly lower boiling point and the thereby reduced chance of evaporation of the organic solvent, ACN was chosen.[39,253,254] The evaluation of higher THC-concentrations identified 150 µM as the maximal solubility of THC in 5% ACN with 0.1% Tween. The measurement at 250 µM suffered from high error bars and its inclusion into the linearity determination severely reduced the obtained R^2. This indicates that not all THC is stably solubilised at 250 µM (**Figure 36**).

150 µM THC in 5% ACN with 0.1% Tween was used for all subsequent assays and the re-selection using affinity elution until the discovery of the solubility problems Tween evoked when combined with THC (section 4.7.2). The solubility assays concerning these problems (section 3.2.7) are discussed in section 4.7.2, the single assay performed for the toggle-SELEX in section 4.8.

4.6 Re-selection with affinity elution

All sequences identified from the initial click-selections either do not recognise THC in solution at all or with affinity constants far above the tested concentrations of 150 µM THC maximum (data not shown, **Figure 36**). It is not possible to differentiate between the two problems. Therefore, the re-selection was supposed to select clickmers with improved affinity constants and enforce the recognition of THC in solution. Towards the first end, more stringent selection conditions were chosen than in the initial selections: Incubation time was decreased over the course of the SELEX and the volume and duration of the wash steps increased. The incubation time limits the maximal k_{on}, but is still quite long with 5 min and probably has no effect on the affinity. If a clickmer is successfully selected, the reduced incubation time might reduce the duration of a potential roadside test using this clickmer. In contrast, the duration of the wash steps enforces a higher k_{off} as any sequence with a low value will dissociate and be removed. This should lead to the selection of high affinity binders.[152] The second aim, the selection of clickmers that bind to THC in solution, was supposed to be reached using affinity elution to recover binding sequences instead of the unspecific heat elution used during the initial click-selections. For this, the DNA is denatured by heat and slowly cooled down in the presence of the target molecule both in solution and immobilised. Sequences with a higher affinity for the target in solution will remain in solution and be amplified. Those with a

higher affinity for the immobilised target and those that do not recognise the target molecule in solution at all will bind to the matrix and be removed from the selection pool. The elution buffer contained 5% ACN to solubilise THC. Therefore, the buffer used for washing was also supplemented with 5% ACN.

Later on, it was discovered that the conditions used to solubilise THC most likely led to the formation of micelles that encapsulated THC. This effectively blocks THC from interacting with the DNA sequences (section 4.7.2). Therefore, sequences were not recovered for their ability to interact with THC in solution. Instead, sequences that did not bind to the THC-sepharose after denaturation by heat were recovered and amplified. Nonetheless, the selections were successful in identifying THC-sepharose-binding sequences (**Figure 39**). This is most likely because a small number of THC-binding sequences dissociated and was amplified. Sequences might also have misfolded while cooling down in the presence of Tween and therefore not bound to the sepharose even though they are capable of binding under normal circumstances.

The DNA from selection cycle 8 and 10 of the selections with benzyl- and CF_3-modified DNA showed increased binding to THC-sepharose (**Figure 38**). While the enrichment of the selection with benzyl-modified DNA is similar to the initial selection with benzyl-modified DNA (**Figure 22A**), the CF_3-modified DNA from selection cycles 8 and 10 discriminates better between THC- and unmodified sepharose than the corresponding DNA from the first selection (**Figure 26A**). This indicates that the change in selection conditions, most probably mainly the increase in the duration of the wash steps, improved the enrichment of THC-binding sequences. A second indication is the fact that the non-binding B10 could not be detected by Sanger sequencing of the DNA of the final selection cycle (**Table 15**). B13 could also not be identified, which is explained by the fact that it does not bind THC-sepharose in the presence of the 5% ACN that were used during the new selection (**Figure 37A**). No explanation is obvious for the disappearance of B12. B11 has a higher frequency than in the first Sanger sequencing, but was the second most abundant sequence according to the NGS of the first selection and was probably just underrepresented in the sequences picked for the first Sanger sequencing (**Table 16**).

All sequences identified in the final selection cycle from the benzyl-modified affinity-SELEX, even the unique sequence 27 (1 of 41 sequences = 2.4%), recognise THC-sepharose (**Figure 39**, **Table 15**, and **Table A 17**). All tested sequences from the selection with CF_3-modified DNA bind THC-sepharose, but nine out of the 41 sequenced sequences (22.0%) were unique and not tested due to time constraints (**Table A 18**). As during the first selection (benzyl: 1 of 37 sequences unique = 2.7%, 3.12% unique according to NGS, CF_3: 13 of 32 sequences unique = 40.6%, 12.83% unique according to NGS) (**Table A 1**, **Table A 2,** and **Figure A 6**), the selection with CF_3-modified DNA was less enriched than the one using benzyl-modified DNA. This again indicates that benzyl might

be a better modification for the selection of THC-binding clickmers than CF_3 as the selection conditions were as similar as possible during both setups. The sidechain of benzyl-azide is identical to the one of phenylalanine, which is supposed to interact with THC in the CB1 (**Table 2** and **Figure 3**). CF_3-azide was chosen due to the high hydrophobicity of the trifluoromethyl-group[255] that we expected to increase the interaction with the hydrophobic THC. According to the data presented in this thesis, modifications as close as possible to the one known to interact with the target molecule in nature seem to be the best choice for click-SELEX. To validate this, further click-selections with different targets and modifications chosen to imitate an interaction that occurs in nature are needed.

4.7 Fluorescence polarisation assays

4.7.1 Synthesis and characterisation of THC-FITC-conjugates

Fluorescence polarisation assays allow the interaction analysis without immobilisation of either of the interaction partners. As the smaller interaction partner needs to be fluorescently labelled, THC was coupled to FITC. Two different strategies were employed: Direct coupling and the use of BDE as linker. The latter was performed to imitate the immobilisation of THC on the epoxy-activated sepharose as closely as possible and thereby maximise the probability that the fluorescently labelled THC would be recognised by the THC-sepharose-binding sequences.

Both coupling reactions were successful and the products clean after purification by HPLC (**Figure 42**, **Figure 56**, and **Figure 57**). As expected, the absorption and fluorescence properties remained mostly unchanged from uncoupled FITC (absorption maximum: 485 nm, excitation maximum: 512 nm) (**Figure 43** and **Figure A 2**).[256] The only difference is an additional absorption peak at about 450 nm that both coupling products exhibit if the measurement takes place in water (THC-FITC and TLF, **Figure 43A** and **Figure A 2A**) or D-PBS, pH 5.4 (TLF, **Figure A 2A**) with 5% ACN each. Both are slightly acidic.[257] The absorption spectrum of fluorescein is known to change depending on its protonation status. The protonation at low pH-values leads to a reduction of the absorption at 490 nm and the appearance of a second peak around 450 nm.[258] This phenomenon is what can be observed in the aforementioned figures.

4.7.2 The effect of Tween on THC

Initial fluorescence polarisation assays showed a dependency on the presence of Tween: Especially at pH 7.4, the polarisation of THC-FITC was much higher in the presence than in the absence of Tween. For FITC alone, no difference was apparent at neutral pH (**Figure 44B**).

According to the theory behind fluorescence polarisation (**Figure 40**), THC-FITC must rotate more quickly in the absence of Tween than in its presence. This can only be the case if THC-FITC is larger in the presence of Tween. As the percentage of Tween used in the assay is above its CMC, the only valid explanation is the hypothesis that THC-FITC is inside of Tween-micelles, thus rotating more slowly and leading to the measured, higher polarisation.

To verify this hypothesis, HPLC-measurements of THC- and THC-FITC-solubility in the presence of different concentrations of Tween were performed. As presented in **Figure 46**, the solubility of 20 µM THC-FITC is reduced in the presence of Tween below (0.005%) and slightly above its CMC (0.01%). No Tween at all is better to keep THC-FITC in solution than those low concentrations of detergent. The same is true for THC: Concentrations of Tween below its CMC are worse for the THC-solubility than no Tween. However, any concentration of Tween above its CMC is able to keep THC in solution (**Figure 46**).

Due to the distribution of detergents at different concentrations as presented in **Figure 45**, increasing amounts of the detergent are adsorbed at the surface of the liquid if concentrations approach the CMC. During the HPLC-measurements, 1/3 of the prepared solution is injected from the bottom of the vial. Any THC or THC-FITC on the surface is therefore not injected.

Taken together, the following hypothesis is the most likely explanation for the results depicted in **Figure 44** and **Figure 46**:

THC-FITC is soluble in 5% ACN without Tween as both **Figure 46A** and **Figure 47** show. At Tween concentrations below or slightly above its CMC, THC-FITC interacts with Tween at the surface and is therefore not injected. The effect is apparent at 20 µM, as lower concentrations are easier to solubilise. Above its CMC, Tween forms micelles, which enclose THC-FITC. It is therefore ‘in solution’ and can be injected and detected by HPLC (**Figure 46A**).

Like THC-FITC, THC is also less soluble in the presence of Tween below its CMC than in its complete absence. Slight amounts of THC are therefore soluble even in 5% ACN, but as soon as the detergent is present, THC interacts with it and is adsorbed at the liquid surface. Since it is not injected, the measured solubility decreases. For THC, this effect is obvious at all three measured THC-concentrations (**Figure 46B**). This is probably due to its higher hydrophobicity in comparison with THC-FITC. This means that THC interacts with the hydrophobic detergent as soon as any is present regardless of the THC-concentration. Both Tween-concentrations above the CMC lead to an immediate increase in THC-solubility as THC is enclosed by the formed micelles and stays ‘in solution’. The fact that THC-solubility is good as soon as the CMC is reached, while THC-FITC-solubility is still insufficient in the presence of 0.01% Tween, has to be attributed to the higher hydrophobicity of THC. It enters the micelles as soon as any are

present, since they present the most hydrophobic environment available. The slightly more hydrophilic THC-FITC is relatively well solubilised by the Tween at the surface and only enclosed by micelles when their concentration increases (**Figure 46**).

All this indicates that while THC and THC-FITC can be found in solution and measured by HPLC in the presence of 0.1% Tween, both are not in solution as single, free molecules. This is necessary to enable the interaction between THC-sepharose-binding sequences and THC or THC-FITC. All assays performed in the presence of Tween can therefore not be evaluated properly and were not mentioned in detail in this thesis. As already apparent from **Figure 46A** and confirmed in **Figure 47**, THC-FITC is soluble in 5% ACN without addition of any detergent up to a concentration of 25 µM. Since much lower concentrations are necessary for the fluorescence polarisation assays, these conditions were chosen to continue.

4.7.3 Evaluation of the THC-sepharose-binding sequences under conditions differing from the original selection buffer

Since the problems associated with the addition of Tween to solutions containing THC (section 4.7.2) necessitated the absence of Tween in any future binding assay, the THC-sepharose-binding sequences had to be tested for their interaction with THC-sepharose in the absence of Tween. For most sequences, the absence leads to a slight increase in binding to THC-sepharose. Tween most likely interacts unspecifically with the hydrophobic modifications of the DNA and thereby keeps them from the THC-sepharose. The effect is stronger for the CF_3- than for the benzyl-modified sequences and most apparent for C15 and C47 (**Figure 49A**). As CF_3 is more hydrophobic than benzyl, it stands to reason that Tween has a stronger effect on the CF_3-modified sequences. Individual differences between sequences are probably due to their folding and depend on their interaction with THC. Sequences, whose (hydrophobic) modifications are on the surface, have a higher probability to interact with Tween and be kept apart from the THC-sepharose. In addition, sequences with a high affinity for THC-sepharose are less likely to dissociate and therefore have a lower probability of ending up in Tween-micelles. Therefore, the effect of Tween on C15 and C47 might indicate that their affinity for THC-sepharose is worse than that of the other sequences.

In addition to the effect of Tween, different buffer components were tested for their effect on the fluorescence polarisation. **Figure 48** shows that BSA increases the fluorescence polarisation of THC-FITC. This is due to the fact that BSA interacts with THC.[259] None of the other components (salmon sperm DNA, ACN) had an effect (**Figure 48**). To enable fluorescence polarisation assays, BSA had to be removed from the buffer. Therefore, the THC-sepharose-binding sequences

had to be investigated in the absence of BSA. In contrast to the effect of Tween, BSA slightly increases the interaction of all sequences with THC-sepharose. Again, the effect is most prominent for C15 and C47 (**Figure 49B**). As BSA interacts with THC, it should bind to THC-sepharose. BSA is mainly hydrophilic[260] and most probably unspecifically interacts with the DNA, leading to the observed increase in binding to THC-sepharose. The higher influence on the interaction of C15 and C47 can – again – be an indication of a weaker affinity or specificity of these sequences as they are more prone to unspecifically interact with the BSA on THC-sepharose than the other sequences.

As the initial selections were the second click-selections after the GFP-selection published by *Tolle et al.*[205], the selection conditions were kept as close to the original ones as possible to increase the chance of selection success. That also meant selecting at a pH of 5.4. The fluorescence of THC-FITC and TLF is higher (**Figure 43B** and **Figure A 2B**) and their polarisation lower at neutral than at acidic pH (**Figure 44** and **Figure 58**). Therefore, the sequences were also tested for their interaction with THC-sepharose at neutral pH to enable fluorescence polarisation assays under these – improved – conditions. In comparison to the absence of either Tween or BSA, the change in pH has a drastic effect: Binding of a large number of sequences is completely abolished. Only two of the six benzyl- (B12 and 27) and five of the nine CF_3-modified sequences (C15, C47, 49, 53, and 77) interact with THC-sepharose at pH 7.4 and most (all but B12 and C15) to a reduced degree. The change in pH should have no effect on either THC, especially as it is coupled to the sepharose via the phenolic hydroxyl, nor on the nucleosides or click-modifications (pK_A-values: THC 10.6[56], cytosine 4.2, guanosine 1.6 and 9.2, adenosine 3.5, uridine 9.2[261], triazole 9.4[262], benzyl (toluene) 40.9[263]. No value has been published for CF_3, but it is probably similar to benzyl). The phosphate group of the nucleic acid backbone can be protonated at pH 5.4 and deprotonated at pH 7.4 (pK_A 1.0 and 6-7[261]). This indicates that the sequences that do not bind at pH 7.4 might depend on salt bridges with their phosphate backbone for folding and/or binding. An influence on folding seems more likely as THC does not present opportunities for salt bridges. It is also possible that the change in pH leads to a change somewhere on the surface of the sepharose, which somehow changes the way THC is presented. According to the manufacturer, the sepharose should be stable at pH-values as low as two.[264] No information on the surface chemistry apart from the functional group and linker is available to evaluate the effect a change in pH might have on it.

4.7.4 Fluorescence polarisation measurements using THC-FITC and TLF

After exclusion of effects that resulted from the micelle-formation of Tween (section 4.7.2) and the addition of BSA (section 4.7.3), both THC-FITC and TLF

lead to polarisation values that allow interaction assays. Nonetheless, the polarisation of the light emitted by THC-FITC is about twice as high as that of TLF at both tested pH-values (**Figure 44** and **Figure 58**). According to the theory behind fluorescence polarisation assays as depicted in **Figure 40**, this indicates a slower rotation and larger size of THC-FITC. Most likely, this is due to the rigid connection between THC to FITC, which means that the molecule can only rotate as a whole. In comparison, the linker between THC and FITC in TLF allows more rotational freedom for FITC, resulting in lower polarisation values that are equal to those obtained with FITC alone (**Figure 54**, **Figure 58**, and **Figure 59**).

To validate the fluorescence polarisation assay for K_d measurements, the affinity constant of the THC-antibody was determined. The divergence between the two measurements with THC-FITC is negligible (**Figure 51**, 27.8 ± 28.9 and 62.7 ± 60.5 nM) and would probably be reduced along with the respective standard deviations if more concentrations in the range below 100 nM were measured. In comparison, the affinity for TLF is severely reduced by a factor of four to eight (**Figure 59**, 226.5 ± 175.8 nM). Since no structural knowledge about the interaction of the antibody and THC is available, further experiments are necessary to discover the reason for the affinity decrease. The evaluation of additional concentrations is needed to solidify the measured affinity discrepancy.

Fluorescence polarisation assays were performed with both THC-FITC and TLF for all identified THC-sepharose-binding sequences (**Figure 52** and **Figure 60**). In addition, the sequences that were determined to recognise THC-sepharose at pH 7.4 (section 4.7.3 and **Figure 50**) were also evaluated at this pH (**Figure A 3** and **Figure A 4**). None of the sequences led to any increase in polarisation at a DNA-concentration of 500 nM. Higher concentrations could not be measured due to the high cost of preparing the benzyl- and CF_3-modified DNA by PCR. If single sequences had been deemed promising after this initial screening (**Figure 52**, **Figure 60**, **Figure A 3**, and **Figure A 4**), they could have been synthesised by solid-phase synthesis. For a screening of all sequences, the costs are too high.

In an attempt to increase the sensitivity of the assay, the THC-sepharose-binding sequences were biotinylated and incubated with streptavidin during the fluorescence polarisation assay. As mentioned before (section 3.2.7.4), streptavidin was chosen as a simple way to increase the molecular weight of the sequences. This increases the polarisation if any interaction between sequence and THC-FITC or TLF occurred (section 3.2.7). Nonetheless, no effect on the polarisation could be observed with any tested sequence (**Figure 53**, **Figure 61**, and **Figure A 5**).

Altogether, the fluorescence polarisation assay was shown to work well with the THC-antibody (**Figure 51** and **Figure 59**). None of the controls ever led to unexpected results. Therefore, no doubts remain that the assay worked as it should have and that it would have shown an increase in polarisation if any of the

THC-sepharose-binding sequences interacted with THC-FITC or TLF. Two possible explanations for this are available:

The affinity of the sequences is too low to enable the observation of interactions at the tested concentration of 500 nM. A number of aptamers for small molecules is known to have affinities in the low to high micromolar range.[147,265] This explanation seems especially likely for the sequences identified during the initial selections (**Table 6** and **Table 7**) as the selection conditions were not particularly stringent. During the re-selections with affinity elution (section 4.6), care was taken to increase the duration of the wash steps to deplete sequences with a low affinity for THC-sepharose.[152] Nonetheless, no prior knowledge about selections for small molecule clickmers is available. SOMAmers, which contain a nucleobase-modification similar to clickmers, have only been published for protein targets. While they have affinities in the low nano- to picomolar range[180], the increase in affinity in comparison to non-nucleobase-modified aptamers might not be transferrable to clickmers for small molecules. Further click-selections for small molecule targets are necessary to get an idea about the affinity distribution. If possible, the parallel and – apart from the click-modification – identical selection for one target with both unmodified and click-modified DNA would help to elucidate the effects of the modification on affinity. For now, it is entirely possible that the THC-sepharose-binding sequences are capable of recognising THC in solution with affinities in the high micro- or even millimolar range. They might then even be implementable into roadside testing devices for THC according to the original aim of this thesis, as some test designs allow significant increases in sensitivity in comparison to the original affinity of used aptamers.[235,266]

The second explanation is that the sequences do not recognise THC-FITC and TLF. This might again have two different reasons: The sequences either do not bind their target molecule in solution at all or they would recognise THC itself, but do not bind THC-FITC or TLF. The latter seems rather unlikely as TLF was engineered to imitate THC-sepharose, which is recognised by all of the sequences. The most likely option might again be a lacking affinity, as the local concentration of THC on the sepharose is very high (section 4.8).

If the sequences lack affinity for THC or if they are unable to recognise THC in solution cannot be answered on the basis of this thesis. While further experiments might be able to gain the required knowledge, even the best case scenario – the sequences recognise THC with a low affinity – necessitates the increase of sensitivity for any possible roadside testing device. Although the legal limit for THC in Germany of 1 ng/ml (3.18 nM) is only set in plasma[82], any roadside testing device for oral fluid should operate with a similar sensitivity. The required effort would probably be tremendous. Therefore, any further evaluation of the THC-sepharose-binding sequences was foregone at this point in preference of new click-selections that utilise all so-far gained information.

4.8 Toggle-SELEX

The toggle-SELEX was performed for two reasons: To circumvent the dependence on sepharose as a sole matrix and to utilise the gained knowledge on THC-solubility by performing affinity elution.

THCA immobilised on tosylated magnetic beads was chosen as second matrix next to the well-established THC-sepharose. The different matrices in combination with the different immobilisation chemistries were supposed to maximise the chances of selecting clickmers that bind in solution. White *et al.*, the inventors of toggle-SELEX, used both human and porcine thrombin as targets to circumvent the selection of species-specific aptamers.[225] The aim of a decrease of specificity – in our case for THC immobilised on sepharose – is the same. To the best of our knowledge, this is the first time different matrices and immobilisation strategies instead of different target molecules were used in a toggle-SELEX.

The selection buffer of the initial click-selections was chosen to be identical to the one used in the single successful, published click-SELEX (section 4.3).[267] As all four click-selections for THC had successfully enriched THC-sepharose-binding sequences, we chose to change the buffer to one more appropriate to the final application: PBS, pH 7.0, with an increased potassium ion concentration to imitate oral fluid (section 3.3).[226] In this buffer, THC-solubility was determined to be 200 µM (**Figure 62**). This is similar to the 150 µM determined in the selection buffer of the initial click-selections (**Figure 36**), as 200 µM were not tested for the initial buffer. The change in pH does not seem to affect THC-solubility at the tested concentrations. Alkaline conditions should deprotonate THC and therefore increase its solubility in aqueous buffers (section 3.1.2.1), but deviate from the pH in human oral fluid and were therefore not considered.

A decrease of the concentration of the target molecule during the incubation period should enforce the selection of clickmers with high affinity.[152] To decrease both the local as well as the global concentration of THC, sepharose with different amounts of THC on the surface was prepared and the amount of THC-sepharose and THCA-beads reduced over the course of the SELEX (**Table 19**). No differences in the amount of THC on the surface could be determined between the 1:10- and 1:100-sepharose and the signal of the THC-antibody was only twice as strong for 1:1-sepharose as for the other two (**Figure 63**). 1:10-sepharose can contain THC coupled to maximal 10% of the functional groups on the surface of the sepharose, 1:100-sepharose to maximal 1%. The affinity constant determined for the THC-antibody and TLF, which resembles THC-sepharose more closely than THC-FITC, during the course of this thesis amounted to about 230 nM (**Figure 59**) and is therefore slightly lower than the 250 nM used for the detection. Even though slight differences in coupling density

might not be detectable, the sensitivity should be high enough this close to the k_D to discriminate between differences of a factor of 10.

Assuming that the antibody signal obtained for 1:100-sepharose represents THC coupled to about 1% of the functional groups (the maximum possible), 1:10-sepharose would also contain THC coupled to about 1% of the functional group as no difference in the signal intensity was observed (**Figure 63**). 1:1-sepharose would then carry THC on about 2% of the functional groups on the surface of the sepharose. This is far from the yield obtained for the coupling in solution, where 99.4% THC was coupled to the linker (**Figure 18**). This difference must result from surface-specific effects that decreased the coupling efficiency. After all, the same coupling chemistry did not work at all on magnetic epoxy-beads, indicating a major effect of the solid support (data not shown).

Due to the low coupling efficiency, the values given for THC in **Table 19** are too high. The amount was reduced from maximal 9.5-20 nmol in selection cycle 1 to 190 to 400 pmol THC from selection cycle 9 onwards. These values are still approximated from the signals of the antibody on the sepharose and no corresponding values are available for the THCA-beads. Therefore, the values in the table were not adjusted and left as the maximal possible amounts of THC(A).

As for the affinity-selections, the incubation time was reduced and the duration of the wash steps increased over the course of the toggle-SELEX to enforce the selection of high affinity clickmers (section 4.6). Affinity elution was not performed on the THC-sepharose, as the matrix cannot be heated above 80°C without starting to melt and it could not be guaranteed that all DNA would be denatured at 80°C.

After ten selection cycles, interaction assays with both THC-sepharose and THCA-beads showed no enrichment for either of the selections (**Figure 64**). The values obtained for THC-sepharose are similar to those of the SL for the first click-selections (**Figure 22A** and **Figure 26A**). DNA-binding to the THCA- and methylamine-blocked tosylated beads is surprisingly high: While about 3.5% of the benzyl-modified DNA from the first selection cycle of the initial SELEX interacted with the THCA-beads and exhibited a clear preference for the modified beads (**Figure 25**), 50-60% of the DNA of the SL remain on both types of beads (**Figure 64C**, **D**). This observation is the most likely explanation for the fact that less PCR-cycles were necessary for amplification during the selection cycles performed on THCA-beads (**Table 19**).

DNA-libraries are not amplified before use, resulting in a variation of the sequence composition between single aliquots. Nonetheless, it seems unlikely that the new aliquot should contain so many sequences that bind to tosylated beads. It is also improbable that the single selection cycle performed with the DNA presented in **Figure 25** reduced sequences that bind to tosylated beads so

thoroughly. The remaining difference between the two assays is the selection buffer: The increase from pH 5.4 to 7.0 as well as the removal of Tween and addition of ACN are the most likely explanation for the increase in unspecific binding.

The high unspecific interaction with both forms of tosylated beads also explains the failure of the toggle-SELEX. Every selection cycle performed on THCA-beads enriched sequences non-specifically binding to the magnetic beads and not to THCA. The parallel selection on THC-sepharose managed to slowly (CF_3-modified DNA) or rapidly (between the start and cycle four, benzyl-modified DNA) reduce the unspecific interaction with the magnetic beads. The difference between the two selections has to be due to the modification. Further knowledge about what kind of interaction leads to the high binding to the tosylated beads and how the sequence compositions change between the different selection cycles is necessary to elucidate how exactly the modifications have such a strong impact.

In a last attempt to enrich THC-binding sequences, four selection cycles were added. Most of these were performed on THC-sepharose and included a pre-selection step to deplete matrix-binding sequences (**Table 19**). Regardless of these efforts, the DNA from the final selection cycle did not show increased binding to THC-sepharose (**Figure 65**).

Altogether, both toggle-selections were unsuccessful. While other changes (e.g., the buffer with its neutral instead of acidic pH) from the successful click-selections before cannot be excluded as causes, the most probable reason is the high unspecific binding of the modified DNA to the tosylated magnetic beads.

5 Outlook

While this thesis was successful in enabling the selection of THC-sepharose-binding sequences through the use of click-SELEX and the structure-guided choice of DNA-modifications, none of the selected sequences could be shown to recognise THC in solution. For the desired application, the incorporation of a clickmer into a roadside oral fluid test, in-solution binding is required. Immobilisation of THC via its hydroxyl group during the roadside test is not possible, as the reaction takes days (**Figure 18**).

Therefore, the development of a roadside test first has to focus on the selection of a clickmer that is able to recognise THC in solution.

Two different approaches appear to be the most promising towards this aim: another toggle-SELEX or a capture-SELEX.

5.1 Toggle-SELEX

To reach selection-success with another toggle-SELEX, a second matrix without any problems due to unspecific interactions of the DNA with the matrix itself is necessary. The THCA-functionalised amine beads used in the initial selections with non-nucleobase-modified nucleic acids (section 3.1.1) are already well established. Nonetheless, they are accompanied by their own problem regarding unspecific interactions with DNA and 2'F-RNA due to their hydrophilic coating. High salt concentrations and competitors as used in SELEX 1 to 5 can reduce the unspecific interactions sufficiently to enable selection (**Table 3**). In addition, the hydrophobic modifications of the benzyl- and CF_3-modified DNA might lessen the problem. In any case, thorough investigations of the unspecific interactions are recommended before starting any selection.

The immobilisation of THC on epoxy-functionalised magnetic beads presents another alternative. The chemistry is well established and should be transferable from THC-sepharose, but has already failed with three different kinds of magnetic beads (data not shown). Since information on the composition and surface of the magnetic beads apart from the functional group itself is rare, it cannot be excluded that a fourth or fifth epoxy-functionalised magnetic bead would be modifiable as desired.

Establishing a new immobilisation technique is the final option. A possibility that bypasses magnetic beads completely is adipic acid dihydrazide–agarose.[268] The linker, adipic acid dihydrazide, is commercially available and could be used to determine the optimal conditions for the coupling reaction, analogue to what has been presented for THC-sepharose in this thesis (section 3.1.2.1). Adipic acid

dihydrazide reacts with carboxyl groups under formation of an amide bond and could be used in combination with THCA or THC-COOH.

If two similar matrices with different surface chemistries are used to 'toggle' (e.g., epoxy-activated sepharose and adipic acid dihydrazide-agarose), a pre-selection is recommended to deplete matrix-binding sequences. While the problem did not arise during the click-selections in this thesis (**Figure 22A** and **Figure 26A**), two different presentations of the target molecule might lead to a stronger enrichment to the non-changing matrix.

A number of selections in this thesis were performed using affinity elution to specifically recover sequences that recognise THC in solution. However, none of them was successful. In all cases, some DNA or 2'F-RNA was unspecifically recovered and amplified. Accordingly, the effectiveness of a specific elution is not apparent during the selection itself. Such an unplanned, unspecific elution enriches sequences that do not bind to the immobilised target again after denaturation and is therefore worse than a completely unspecific heat elution. To exclude this risk, two selections can be performed in parallel with either heat or affinity elution as exemplified by SELEX 1 and 2 (**Table 3**).

5.2 Capture-SELEX

For capture-SELEX, the DNA-library contains a docking sequence that is complementary to a capture oligo. The latter is immobilised onto magnetic beads. The library is annealed with the capture oligo and remains on the magnetic beads. Only if a sequence interacts with the target molecule in such a way that the parallel basepairing with the capture oligo is no longer possible is the sequence removed from the solid support and amplified for the subsequent selection cycles.[151,269] Since the target molecule is not immobilised during the selection process, this technique guarantees the selection of aptamers that recognise the target molecule in solution.

Most publications utilise a target concentration of 1 mM or higher.[151,270-273] This will not be achievable with THC in an aqueous solution. However, there are some examples, in which lower target concentrations were used to successfully select specific aptamers. For instance, Nutiu *et al.* used 100 µM of each of the nucleoside triphosphates (NTPs), Spiga *et al.* 200 µM tobramycin for selection cycle one to four and 20 µM for the remaining ones.[269,274] These concentration ranges are realistic according to the data in this thesis (**Figure 36** and **Figure 62**), indicating that capture-SELEX might be a valid approach to select a clickmer for THC.

Capture-SELEX has been used for the selection of both DNA-[151,269-272,274] and RNA-aptamers[275], but not in combination with nucleobase-modified nucleic acids.

To combine capture- and click-SELEX, special care has to be taken concerning library design. The base pairing of the docking sequence and the capture oligo has to be stable enough to prevent dissociation of the library during incubation with the target molecule, while still enabling the dissociation if it results from the interaction between sequence and target molecule.[151] Since the effect of the click-modifications on base pair stability is unknown, melting point studies should precede library design.

6 Material & Methods

6.1 Material

6.1.1 Equipment

Equipment	Manufacturer
Agarose gel camera, UV-transilluminator	Bio-Rad Laboratories
Agarose running chamber	In house construction, peqlab
Amicon ultra centrifugal filter units, MWCO 10k	Merck
Analytical balances	Sartorius
Analytical HPLC system, 1100	Agilent
Analytical HPLC system, 1260 Infinity	Agilent
Autoclave	Systec
Bacterial incubator shaker, 4430	Innova
BAS cassettes	Fuji
Centrifuges	Eppendorf, Sigma
Cling film	Roth
Counter	Berthold
Electrophoresis power supply, E865	Consort
Empty micro Bio-Spin columns	Bio-Rad
EnSpire multimode reader	PerkinElmer
Eppi racks	Roth
Glass frit	RobuGlas
Glass pipette tips	Hirschmann
Glass plates for PAGE	Baack
Glass wool	Sigma-Aldrich
Heating cabinet	WTB binder
HiSeq 1500	Illumina
HPLC vials N11 with snap ring caps	Macherey-Nagel
HPLC-column EC/50/4.6 Nucleodur 100-5 18 ec	Macherey-Nagel
HPLC-column Zorbax_50x2.1_5µm	Agilent
HTC Esquire	Bruker
Image eraser	Raytest
Liquid Scintillation counter, WinSpectral 1414	PerkinElmer
Lyophiliser (Alpha 2-4 LD plus)	Christ
Magnetic rack, Dynamag 2	Life Technologies
Microspin G25 columns	GE healthcare
Microwave	Bosch
NanoDrop	Peqlab
NanoQuant infinite M200	Tecan
NanoQuant Plate	Tecan
Odyssey blot imager	LI-COR
PAGE running chamber	In house construction
Parafilm	Faust
Pasteur pipette with rubber bulb	Sigma-Aldrich

PCR 8er Deckelkette flach	Sarstedt
PCR multiply strip 0.2 ml Kette	Sarstedt
PCR Thermocycler	Biometra, Eppendorf
Petri dishes (94 x 16 mm)	Labomedic
pH meter	Inolab
Phosphoimager screens	Fujifilm
Phosphoimager, FLA-3000	Fujifilm
Pierce streptavidin coated high binding capacity black 96-well plate	Thermo Fisher Scientific
Pipette tips	Sarstedt
Pipetteboy Accu-jet Pro	Brand
Pipettes	Eppendorf
Proxi Plate 384 F Plus	PerkinElmer
Radioactive protection shield	Nalgene
Reaction tubes 0.2 ml, 0.5 ml, 1.5 ml, 2 ml	Eppendorf
Refrigerators and Freezer 4°C, -20°C, -80°C	AEG, New Brunswick, Liebherr
Scalpel blades	Labormedic
Speedvac (Concentrator 5301)	eppendorf
Syringe, 5 ml	Braun
Tecan ultra	Tecan
Thermomixer 1.5 ml, 2 ml	Eppendorf, HLC
Thermomixer HLC 2 ml	Ditabis
Tube rotator, REAX 2	Heidolph
Vacuum distillation rotator	Heidolph
Vials for liquid scintillation counter, Maxi-Vial,18 ml	PerkinElmer
Vortex	Neolab
Water bath	GFL, Julabo

6.1.2 Chemicals

Reagents	**Manufacturer**
1,4-dithiothreitol (DTT)	Roth
2'fluoro cytidine triphosphate (2'F-CTP)	Metkinen
2'fluoro uridine triphosphate (2'F-UTP)	Metkinen
7 N ammonia (NH_3) in MeOH	Sigma-Aldrich
ACN, HPLC grade	Sigma-Aldrich
ACN, LC-MS grade	Sigma-Aldrich
Adenosine triphosphate (ATP)	Jena Bioscience
Agar, bacteriology grade	AppliChem
Agarose	Bio-Budget technologies
Ammonium peroxodisulfate (APS)	Roth
Ampicillin	Roth
BDE	Sigma-Aldrich
Benzyl-azide	Sigma-Aldrich
Boric acid	AppliChem
Bovine serum albumin	Calbiochem
Bromphenol blue	Merck

Calf intestine alkaline phosphatase (CIAP) with 10x buffer	Promega
CF_3-azide	In house production, AK Mayer; Malte Rosenthal
Chloroform	Fisher Scientific
Copper(II) sulfate	Sigma-Aldrich
Cremophor EL	Sigma-Aldrich
DCC	Sigma-Aldrich
Deoxyadenosine triphosphate (dATP)	Jena Bioscience
Deoxycytidine triphosphate (dCTP)	Jena Bioscience
Deoxyguanosine triphosphate (dGTP)	Jena Bioscience
Deoxynucleoside triphosphate (dNTP)-mix	Jena Bioscience
DMF	Fisher Scientific
DMSO	Sigma-Aldrich
DNA Polymerase I, Large (Klenow) Fragment	NEB
D-PBS, 10x (cat. no. RNBF8385)	Sigma-Aldrich
Dynabeads M-270 Amine	Thermo Fisher Scientific
EDTA	AppliChem
EdUTP	BaseClick
Epoxy-activated sepharose 6B	GE Healthcare
Ethanol	VWR
Ethidium bromide	Roth
Formamide	Sigma-Aldrich
Formic acid, HPLC and LC-MS grade	Merck
Glycogen	Roche
Goat anti-mouse IgG (L+H) Dylight 800 conjugated	Thermo Fisher Scientific
GoTaq DNA Polymerase with 10x buffer	Promega
GoTaq Flexi DNA Polymerase with 10x buffer	Promega
Guanosine triphosphate (GTP)	Jena Bioscience
Heparin	Sigma-Aldrich
Hydrochlorid acid (HCl)	Roth
Hydroxylamine	Sigma-Aldrich
Inorganic pyrophosphatase (iPP)	Roche
Lennox broth (LB)	Roth
Magnesium chloride ($MgCl_2$)	Roth
MeOH	Fisher Scientific
N,N,N,N-Tetramethylethylendiamide (TEMED)	Roth
Phenol (Roti-Phenol)	Roth
Potassium chloride (KCl)	Roth
Potassium phohsphate monobasic ($KH_2PO_{4)}$	Fluka
Pwo DNA Polymerase with 10x buffer	Genaxxon
RNAsin	Promega
Rotiphorese sequencing gel concentrate	Roth
Salmon sperm DNA	Invitrogen
Sodium acetate (NaOAc)	VWR
Sodium chloride (NaCl)	Roth
Sodium dodecylsulfate (SDS)	Roth
Sodium L-ascorbate	Sigma-Adrich

Sodium phosphate dibasic (Na_2HPO_4)	Roth
Super optimal growth (SOC) medium	NEB
SuperScript II reverse transcriptase with 5x first strand buffer	Thermo Fisher Scientific
T4 Polynucleotide kinase (PNK) with 10x buffer	NEB
T7 Y639F RNA polymerase	In house production, AK Famulok; Nicole Krämer
Taq polymerase with 10x buffer	In house production, AK Famulok; Nicole Krämer
TEA	Acros Organics
THC (dronabinol)	THC Ohara
THCA	Lipomed
THC-antibody, sequence M4020412 (cat. no. 10-T43F)	Fitzgerald
Tris	Roth
Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA)	BaseClick
Tween-20	Roth
Urea	Roth
Xylenecyanole	Merck
λ-exonuclease with 10x buffer	Thermo Fisher Scientific

6.1.3 Buffers

All buffers were solubilised in ddH_2O and the pH (negative logarithm of the H^+-concentration) adjusted using HCl and NaOH.

Buffer	**composition**
0.1 M phosphate buffer	61.5 mM K_2HPO_4 38.5 mM KH_2PO_4, pH 7.0
10x TBE buffer	890 mM Tris/HCl 890 mM boric acid 20 mM Na_2EDTA, pH 8.0
2x RNA loading dye	95% (v/v) formamide 0.5 mM EDTA 0.025% (w/v) SDS 0.025% (w/v) Xylencyanol
4x PAGE loading dye	9 M urea 50 mM EDTA, pH 8.0
6x DNA loading dye	60% (v/v) glycerol 10 mM Tris/HCl 0.03% (w/v) Xylencyanol 60 mM Na_2EDTA, pH 8.0
PBS, pH 7.4	137 mM NaCl 3.93 mM $Na_2HPO_4 * 2\ H_2O$ 1.47 mM KH_2PO_4, pH 7.4

6.1.4 Kits

Product	Manufacturer
Gel and PCR clean-up kit	Macherey-Nagel
NTC buffer for Gel and PCR clean-up kit	Macherey-Nagel
NucleoSpin Plasmid	Macherey-Nagel
TOPO TA cloning kit with One Shot TOP10 chemically competent *Escherichia coli*	Life Technologies
TruSeq DNA PCR-Free Sample Preparation Kit LT	Illumina

6.1.5 Standards, radioactive nucleotides

Type	Length	Manufacturer
Ultra low range DNA ladder	10 – 300 base pairs (bp)	Thermo Fisher Scientific

γ ^{32}P-ATP purchased from Perkin Elmer.

6.1.6 Synthetic oligos

Name	Sequence (5´ - 3´)
Click-competitor	N_{42} – A
D3-F	GCT GTG TGA CTC CTG CAA
D3-library	GCT GTG TGA CTC CTG CAA - N_{43} – GCA GCT GTA TCT TGT CTC C
D3-R-P	Phosphate – GGA GAC AAG ATA CAG CTG C
FT2-F	CAC GAC GCA AGG GAC CAC AGG
FT2-F-bio	Biotin – CAC GAC GCA AGG GAC CAC AGG
FT2-F-Cy5	Cy5 – CAC GAC GCA AGG GAC CAC AGG
FT2-library	CAC GAC GCA AGG GAC CAC AGG – N_{42} – CAG CAC GAC ACC GCA GAG GCA
FT2-R	TGC CTC TGC GGT GTC GTG CTG
FT2-R-P	Phosphate – TGC CTC TGC GGT GTC GTG CTG
GDNA-F	GGG AGA GGA GGG AAG TCT ACA TCT T
GDNA-library	GGG AGA GGA GGG AAG TCT ACA TCT T NNN GGG NNN GGG NNN GGG NNN GGG NNN AAG ATG TCT GGA GTT GAC GAA GCT T
GDNA-R-P	Phosphate – AAG CTT CGT CAA CTC CAG ACA TCT T
GRNA-F	ATA GCT AAT ACG ACT CAC TAT AGG GAG AGG AGG GAA GTC TAC ATC TT
GRNA-library	ATA GCT AAT ACG ACT CAC TAT AGG GAG AGG AGG GAA GTC TAC ATC TT

	NNN GGG NNN GGG NNN GGG NNN GGG NNN AAG ATG TCT GGA GTT GAC GAA GCT T
GRNA-R	AAG CTT CGT CAA CTC CAG ACA TCT T
M111-F	GCG CCA GTC TAG GGC ACC
M111-library	GCG CCA GTC TAG GGC ACC - N_{75} – CAT TGA CTC GGT GGA TCC
M111-R-P	Phosphate – GGA GGA TCC ACC GAG TCA ATG
SulI-F	GGG GGA ATT CTA ATA CGA CTC ACT ATA GGG AGG ACG ATG CGG
SulI-library	GGG GGA ATT CTA ATA CGA CTC ACT ATA GGG AGG ACG ATG CGG - N_{40} – CAG ACG ACT CGC TGA GGA TCC GAG A
SulI-R	TCT CGG ATC CTC AGC GAG TCG TC
T_w/o	CTT GTA CAC GAC GCA AGG GAC CAC AGG GTT GGA AGC GAC GGG ACG GTA AGG CTT GGG CCC CAA GGA GTG CAG CAC GAC ACC GCA GAG GCA TAC AAG
C12_EdU	CAC GAC GCA AGG GAC CAC AGG GXX GGA AGC GAC GGG ACG GXA AGG CXX GGG CCC CAA GGA GXG CAG CAC GAC ACC GCA GAG GCA
C12_T	CAC GAC GCA AGG GAC CAC AGG GTT GGA AGC GAC GGG ACG GTA AGG CTT GGG CCC CAA GGA GTG CAG CAC GAC ACC GCA GAG GCA
FT2_GATC_sense	CTT GTA CAC GAC GCA AGG GAC CAC AGG GAT CGA TCG ATC GAT CGA TCG ATC GAT CGA TCC AGC ACG ACA CCG CAG AGG CAT ACA AG
FT2_G4A4T4C4_sense	GGC TAC CAC GAC GCA AGG GAC CAC AGG GGG GAA AAT TTT CCC CGG GGA AAA TTT TCC CCC AGC ACG ACA CCG CAG AGG CAG TAG CC
FT2_G2A2T2C2_sense	GCC AAT CAC GAC GCA AGG GAC CAC AGG GGA ATT CCG GAA TTC CGG AAT TCC GGA ATT CCC AGC ACG ACA CCG CAG AGG CAA TTG GC
FT2_G3A3T3C3_sense	ACA GTG CAC GAC GCA AGG GAC CAC AGG GGG AAA TTT CCC GGG AAA TTT CCC GGG AAA TTC AGC ACG ACA CCG CAG AGG CAC ACT GT
FT2-TGCA_sense	ACT TGA CAC GAC GCA AGG GAC CAC AGG TGC ATG CAT GCA TGC

	ATG CAT GCA TGC ATG CAC AGC ACG ACA CCG CAG AGG CAT CAA GT
D3-TGCA_sense	TAG CTT GCT GTG TGA CTC CTG CAA TGC ATG CAT GCA TGC ATG CAT GCA TGC ATG CAG CAG CTG TAT CTT GTC TCC AAG CTA
FT2-T4G4C4A4_sense	CAG ATC CAC GAC GCA AGG GAC CAC AGG TTT TGG GGC CCC AAA ATT TTG GGG CCC CAA AAC AGC ACG ACA CCG CAG AGG CAG ATC TG
D3-T4G4C4A4_sense	GAT CAG GCT GTG TGA CTC CTG CAA TTT TGG GGC CCC AAA ATT TTG GGG CCC CAA AAG CAG CTG TAT CTT GTC TCC CTG ATC
FT2_GATC_antisense	CTT GTA TGC CTC TGC GGT GTC GTG CTG GAT CGA TCG ATC GAT CGA TCG ATC GAT CGA TCC CTG TGG TCC CTT GCG TCG TGT ACA AG
FT2_G4A4T4C4_antisense	GGC TAC TGC CTC TGC GGT GTC GTG CTG GGG GAA AAT TTT CCC CGG GGA AAA TTT TCC CCC CTG TGG TCC CTT GCG TCG TGG TAG CC
FT2_G2A2T2C2_antisense	GCC AAT TGC CTC TGC GGT GTC GTG CTG GGA ATT CCG GAA TTC CGG AAT TCC GGA ATT CCC CTG TGG TCC CTT GCG TCG TGA TTG GC
FT2_G3A3T3C3_antisense	ACA GTG TGC CTC TGC GGT GTC GTG CTG AAT TTC CCG GGA AAT TTC CCG GGA AAT TTC CCC CTG TGG TCC CTT GCG TCG TGC ACT GT
FT2-TGCA_antisense	ACT TGA TGC CTC TGC GGT GTC GTG CTG TGC ATG CAT GCA TGC ATG CAT GCA TGC ATG CAC CTG TGG TCC CTT GCG TCG TGT CAA GT
D3-TGCA_antisense	TAG CTT GGA GAC AAG ATA CAG CTG CTG CAT GCA TGC ATG CAT GCA TGC ATG CAT GCA TTG CAG GAG TCA CAC AGC AAG CTA
FT2-T4G4C4A4_antisense	CAG ATC TGC CTC TGC GGT GTC GTG CTG TTT TGG GGC CCC AAA ATT TTG GGG CCC CAA AAC CTG TGG TCC CTT GCG TCG TGG ATC TG
D3-T4G4C4A4_antisense	GAT CAG GGA GAC AAG ATA CAG CTG CTT TTG GGG CCC CAA AAT TTT GGG GCC CCA AAA TTG CAG GAG TCA CAC AGC CTG ATC

All oligos were purchased from Ella Biotech GmbH. For RNA-libraries, the start of the transcribed RNA-sequence is underlined. X = EdU.

6.1.7 Software

Software	Manufacture
Adobe Illustrator	Adobe Systems
AIDA Biopackage	Raytest
ChemDraw	PerkinElmer
COMPAS	AptaIT
Compass DataAnalyser	Bruker
GraphPad Prism	GraphPad Software
Magellan data analysis software	Tecan
Microsoft office package	Microsoft
Odyssey 2.1	LI-COR
QuantityOne	Bio-Rad

6.2 Methods

6.2.1 General procedures

6.2.1.1 Agarose gel electrophoresis

4% (w/v) agarose gels were prepared by dissolving the respective amount of agarose in water and solubilisation in a microwave. The resulting liquid was supplemented with 0.1% (v/v) ethidium bromide and left to solidify in gel casting chambers. Samples for agarose gel electrophoresis were prepared as follows: 5 µl DNA were mixed with 1 µl 6x DNA loading dye or 5 µl RNA were mixed with 5 µl 2x RNA loading dye. 4 µl ultra-low range DNA ladder were used as standard for both RNA- and DNA-gels. The gels were run for 13 min at 150 V and the DNA/RNA visualised by UV-transillumination.

6.2.1.2 Phenol-chloroform extraction

One volume phenol was added to the sample, vortexed vigorously, and centrifuged at 14000 rpm for 5 min. The upper phase was transferred to a new vial. Two volumes chloroform were added, vortexed vigorously, and centrifuged at 14000 rpm for 3 min. The upper phase was transferred to a new vial and ethanol precipitated (section 6.2.1.3).

6.2.1.3 Ethanol precipitation

1/10 volume 3 M NaOAc, pH 5.4, 0.5 µl glycogen, and three volumes EtOH were added to the sample. The precipitation was incubated for 10 min at -80°C and centrifuged for 20 min at 4°C and 14000 rpm. The supernatant was discarded, the pellet was washed with 1 volume 70% (v/v) EtOH, and centrifuged for 3 min

at RT, 14000 rpm. The supernatant was discarded and the pellet dried on air before resuspension in ddH_2O.

6.2.1.4 5’-Dephosphorylation of 2’F-RNA

The dephosphorylation mix was setup as described in **Table 21** and added up to 50 µl with ddH_2O. The mix was incubated at 37°C for 15 min. 0.425 µl 20 u/µl CIAP were added and incubated at 55°C for 15 min. 0.5 µl 0.5 M EDTA, pH 8.0 were added and incubated at 75°C for 10 min. Finally, 150 µl ddH_2O were added and the RNA purified by phenol-chloroform extraction and ethanol precipitation. The pellet was resuspended in 5 µl, 3 of which were used for radioactive phosphorylation (section 6.2.1.5).

Table 21: Setup of 5'-dephosphorylation

Reagent	Stock concentration	Volume for 1 reaction	Final concentration in 50 µl
CIAP buffer	10x	5 µl	1x
BSA	10 mg/ml	5 µl	1 mg/ml
RNA			75 pmol
RNAsin	40 u/µl	0.5 µl	20 u
CIAP	20 u/µl	0.85 µl	17 u

6.2.1.5 Radioactive phosphorylation of single-stranded nucleic acids

The dephosphorylated 2’F-RNA or DNA (either solid-phase synthesised or amplified by PCR, both of which are already 5’-dephosphorylated) was mixed for the radioactive phosphorylation according to **Table 22** and filled up to 50 µl with ddH_2O. The reaction was incubated at 37°C for 1 h and purified using MicroSpin G-25 columns according to the manufacturer’s instructions.

Table 22: Setup of 5' radioactive phosphorylation

Reagent	Stock concentration	Volume for 1 reaction	Final concentration in 50 µl
2’F-RNA or DNA		3 µl	15 pmol
T4 PNK buffer	10x	5 µl	1x
T4 PNK	10 u/µl	2 µl	0.4 u/µl
γ-^{32}P-ATP	10 µCi/µl	1 µl	10 µCi

6.2.1.6 Radioactive polyacrylamide gel electrophoresis

10% Polyacrylamide gel electrophoresis (PAGE) gels were prepared according to **Table 23** and polymerised. The gel was inserted into a gel running chamber and pre-run at 375 V, 15 W for 30 min. The samples were prepared by mixing 0.5 µl radioactively labelled 2'F-RNA or DNA with 14.5 µl ddH_2O and 5 µl 4x PAGE loading buffer. The samples were heated at 95°C for 1 min and loaded onto the gel. The gel was run at 350 V, 15 W for 45 min to 1.5 h. It was carefully removed from the glass plates, wrapped in cling film and placed into a BAS cassette with a Fujifilm phosphoimager screen on top. The screen was exposed overnight at -80°C and read out using a Phosphoimager FLA-3000.

Table 23: Composition of 10% PAGE-gels

Reagent	Volume for 1 gel
8.3 M Urea in 10x TBE	4 ml
Rotiphorese Sequenziergel Konzentrat	16 ml
8.3 M Urea	20 ml
10% (w/v) APS	320 µl
TEMED	16 µl

6.2.1.7 Concentration determination of nucleic acids

The concentration of the nucleic acids was determined with a NanoQuant Infinite M200 and 2 µl sample on a NanoQuant Plate using the absorption at 260 nm. The molar concentration was calculated using the concentration in ng/µl obtained by the NanoQuant and the molar mass of the respective nucleic acid.

6.2.1.8 Single-strand displacement

6 µl 10 u/µl λ-exonuclease and 1/10 volume 10x λ-exonuclease buffer were added to the PCR-product and incubated for 1 h at 37°C, 650 rpm for single-strand displacement. The displacement was monitored by agarose gel electrophoresis according to section 6.2.1.1. The ssDNA was purified with a Gel and PCR clean-up kit according to the manufacturer's instructions using one column and four volumes NTC buffer. The purified ssDNA was eluted in two times 25 µl ddH_2O.

For the selections with non-nucleobase-modified DNA-libraries, 5 µl instead of 6 µl λ-exonuclease were used and the incubation performed until no more dsDNA was detectable on the agarose gel (45 min and 2 h).

For the amplification of DNA from selection cycles and sequences, 2.7 µl 10 u/µl λ-exonuclease were used for 100 pmol dsDNA.

6.2.2 Immobilisation of THC(A)

6.2.2.1 Immobilisation of THCA on amine-functionalised magnetic dynabeads

THCA was coupled to Dynabeads® M270 Amine as depicted in **Figure 13**. To prepare sufficient beads for a selection, 400 µl 30 mg/ml magnetic beads were washed three times with 1.5 ml DMSO each and resuspended in 1370 µl DMSO. After addition of 190 µl 5 mg/ml THCA and 24 µl 10 mg/ml DCC (both in DMSO), the coupling reaction was rotated overnight at RT. Then, 175 µl 100 mM hydroxylamine (in DMSO) was added according to the manufacturer's instructions and the sample rotated for 15 min at RT. The coupled beads were washed three times with 1.5 ml DMSO and three times with 1.5 ml PBS before resuspension in 1.2 ml PBS, pH 7.4, for a final concentration of 10 mg/ml.

Unmodified beads as negative control were prepared accordingly, but without addition of THCA and DCC and without overnight incubation.

6.2.2.2 Coupling of THC to 1,4-butanedioldiglycidyl ether

THC was coupled to BDE as depicted in **Figure 16**. 12.5 µmol THC (39.3 µl 100 mg/ml in EtOH) were dried in a speedvac and resuspended in 93.1 µl 90% DMSO with 10% (v/v) TEA. 6.9 µl BDE were added and rotated for one to seven days at 50°C. After 1, 3, and 6 days, a sample containing 0.53 µl of the coupling solution was taken and stored at -20°C after addition of 99.5 µl ddH_2O for LC-MS-analysis. Two additional samples were taken on coupling day 7: One contained 0.53 µl of the coupling solution and 99.5 µl 80% ACN with 20% ddH_2O with 0.1% (v/v) formic acid. The more concentrated sample contained 5.3 µl of the coupling solution and 94.7 µl 80% ACN with 20% ddH_2O with 0.1% (v/v) formic acid.

6.2.2.3 Coupling of THC to epoxy-activated sepharose

For the preparation of 1:1-THC-sepharose, 2.3 g epoxy-activated sepharose 6B were swelled in 50 ml ddH_2O. Once the sepharose had taken up the water, it was washed with 550 ml ddH_2O on a glass frit and with three times 20 ml 90% DMSO with 10% (v/v) TEA. The sepharose was transferred to a 50 ml falcon in overall 4 ml 90% DMSO with 10% (v/v) TEA.

800 µmol THC (2515 µl 100 mg/ml in EtOH) were dried in a speedvac and resuspended in the sepharose solution. The sample was rotated at 50°C for seven days.

The THC-sepharose was transferred to a glass frit and washed three times with 30 ml 90% DMSO with 10% (v/v) TEA. Afterwards, it was washed three times

with 50 ml ddH_2O. The THC-sepharose was washed three times alternatingly with 45 ml each 0.1 M NaOAc, pH 4.0 with 500 mM NaCl and 0.1 M Tris-HCl, pH 8.0 with 500 mM NaCl. Finally, the THC-sepharose was washed three more times with 50 ml ddH_2O and resuspended in ca. 6 ml ddH_2O for a final concentration of 50% (v/v) resin (sepharose). The THC-sepharose was stored at 4°C.

For the same amount of 1:10- and 1:100-THC-sepharose, 11.99 and 1.2 µmol THC were used, respectively.

Unmodified sepharose was prepared by letting it swell in ddH_2O and washing it on a glass frit with an amount of water equal to what was used for THC-sepharose at the start. The sepharose was resuspended to a final concentration of 50% (v/v) resin in ddH_2O

6.2.2.4 Detection of THC(A) on the solid support

50 µl sepharose were washed in an empty micro BioSpin column with 2 ml PBS, pH 7.4. The sepharose was resuspended in 150 µl PBS with 0.1% (v/v) Tween-20 and 1 mg/ml BSA and rotated for 30 min at RT. The sepharose was transferred to a column and washed with 3 ml PBS with 0.1% (v/v) Tween. It was resuspended in 148.5 µl PBS and 1.5 µl 3.75 mg/ml THC-antibody (final 1:100-dilution). The sample was rotated for 1 h at RT. The sepharose was transferred to a column and washed with 3 ml PBS with 0.1% (v/v) Tween. It was resuspended in 149 µl PBS and 1 µl 1:100-diluted goat anti-mouse IgG, Dylight 800 conjugated (stock 1 mg/ml) secondary antibody. The sample was protected from light from here on and rotated for 30 min at RT. The sepharose was transferred once more to a column and washed with 3 ml PBS with 0.1% (v/v) Tween, before resuspension in 150 µl PBS. 50 µl were spotted onto the glass plate of an Odyssey reader for evaluation. AIDA was used for quantification.

For the magnetic beads, the assay was performed using 18.75 µl 10 mg/ml magnetic beads and without the transferral to the empty BioSpin columns for the wash steps.

6.2.3 Selections with non-nucleobase-modified nucleic acids

6.2.3.1 General procedure for 2'F-RNA-selections

All selections with 2'F-RNA were performed as described here. 50 µl 10 mg/ml non-modified amine dynabeads (section 6.2.2.1) were washed twice with 100 µl selection buffer. The beads were resuspended in 1 nmol 2'F-RNA-library in overall 200 µl selection buffer and incubated for 30 min at 20°C. To keep the beads in suspension, they were pipetted up and down every 5 min. The supernatant, which contains all sequences not binding to the matrix itself, was

transferred onto 50 µl 10 mg/ml THCA-functionalised dynabeads that had also been washed twice with 100 µl selection buffer beforehand. The remaining library was incubated with the target-modified beads for 30 min at 20°C. As before, the beads were pipetted up and down every 5 min to keep them in suspension. The beads were then washed with 100 µl selection buffer for 30 sec, followed by a second wash step that lasted 3 min. Depending on the selection cycle, this wash procedure was repeated as detailed in **Table 24**.

Table 24: Repeats of the wash procedure during SELEX 1 to 5

Selection cycle	Repeats wash procedure
1 and 2	1
3 and 4	2
5 and 6	3
7 and 8	4
9 to 15	5

Only selection cycle 1 to 10 were performed for SELEX 3 to 5.

The binding sequences were then recovered as detailed in the respective SELEX-section. 50 µl eluate was added to 49 µl of the respective RT-PCR-mastermix (**Table 29** and **Table 30**) and incubated for 5 min at 65°C. The PCR was cooled down to 4°C and 0.5 µl SuperScript II Reverse Transcriptase (200 U/µl) and 0.5 µl GoTaq DNA Polymerase (5 U/µl) were added. The PCR was continued according to **Table 25**.

Table 25: PCR-method for 2'F-RNA-selections

Step	Temperature [°C]	Duration
1	54	10 min
2	95	30 sec
3	SELEX 1 + 2: 60	30 sec
	SELEX 3: 71	30 sec
4	72	30
5	72	2 min

Steps 2 to 4 were repeated according to **Table 26**.

The PCR was monitored by agarose gel electrophoresis as detailed in section 6.2.1.1.

10 µl of the PCR-product was mixed with the transcription mastermix (**Table 27**), 2.5 µl 50 u/µl T7 Y639F RNA polymerase, and 1 µl 40 u/µl RNAsin. The remaining PCR-product was stored at -20°C. The transcription was incubated for 20 min at 37°C and its success verified by agarose gel electrophoresis (section 6.2.1.1).

Table 26: Number of RT-PCR-cycles needed for amplification during SELEX 1, 2, and 3

Selection cycle	RT-PCR-cycles SELEX 1	RT-PCR-cycles SELEX 2	RT-PCR-cycles SELEX 3
1	13	5	11
2	19	9	24
3	19	11	21
4	19	11	16
5	19	11	16
6	16	10	18
7	18	10	20
8	18	10	20
9	17	9	22
10	14	7	21
11	16	8	-
12	16	7	-
13	18	7	-
14	16	8	-
15	16	8	-

Table 27: Composition of the transcription-master mix for SELEX 1, 2, and 3

Reagent	Stock concentration	Volume for 1 reaction	Final concentration in 100 µl
Tris, pH 7.9	200 mM	20 µl	40 mM
DTT	100 mM	5 µl	5 mM
ATP, GTP (each)	100 mM	1 µl	1 mM
2'F-CTP, 2'F-UTP (each)	100 mM	3 µl	3 mM
$MgCl_2$	100 mM	15 µl	15 mM
ddH_2O		38.5 µl	

For all subsequent selection cycles, 20 µl transcription product was added to 50 µl washed, non-modified beads in overall 100 µl selection buffer.

6.2.3.2 Preparation of radioactive 2'F-RNA for interaction assays

2'F-RNA of the selection cycles was prepared by transcription according to **Table 28**. The transcription was incubated at 37°C for 4 h. The resulting 2'F-RNA was purified using the Gel and PCR clean-up kit according to the manufacturer's instructions.

The purified 2'F-RNA was dephosphorylated as detailed in section 6.2.1.4 and 3 µl of the resulting dephosphorylated 2'F-RNA radioactively phosphorylated according to section 6.2.1.5. The phosphorylation was then verified by PAGE as described in section 0.

Table 28: Setup of transcription

Reagent	Stock concentration	Volume for 1 reaction	Final concentration in 100 µl
Tris, pH 7.9	200 mM	20 µl	40 mM
DTT	100 mM	5 µl	5 mM
ATP, GTP (each)	100 mM	1 µl	1 mM
2'F-CTP, 2'F-UTP (each)	100 mM	3 µl	3 mM
$MgCl_2$	100 mM	25 µl	25 mM
RNAsin	40 u/µl	1 µl	0.4 u/µl
iPP	200 u/µl	0.2 µl	2 u/µl
T7 Y639F RNA polymerase	50 u/µl	2.5 µl	1.25 u/µl
RT-PCR-product		10 µl	
ddH_2O		28.3 µl	

6.2.3.3 SELEX 1

The selection buffer for SELEX 1 contained 3 mM $MgCl_2$, 500 mM NaCl (overall, including the NaCl in PBS), 0.1 mg/ml salmon sperm DNA in PBS, pH 7.4. SulI was used as selection library.

To recover the bound sequences, the beads were resuspended in 55 µl ddH_2O and incubated at 80°C for 3 min. The supernatant was quickly removed and used as eluate as described above.

The composition of the RT-PCR-mastermix is given in **Table 29**.

Table 29: Composition of the RT-PCR-mastermix for SELEX 1 and 2

Reagent	Stock concentration	Volume for 1 reaction	Final concentration in 100 µl
Colorless GoTaq buffer	5x	20 µl	1x
First strand buffer	5x	4 µl	0.2x
DTT	100 mM	2 µl	2 mM
SulI-F	100 µM	1 µl	1 µM
SulI-R	100 µM	1 µl	1 µM
$MgCl_2$	100 mM	1.5 µl	1.5 mM
dNTPs	25 mM (each)	1.2 µl	0.3 mM (each)
ddH_2O		17.3	

6.2.3.4 SELEX 2

The selection buffer for SELEX 2 contained 3 mM $MgCl_2$, 500 mM NaCl (overall, including the NaCl in PBS), 0.1 mg/ml salmon sperm DNA in PBS, pH 7.4. SulI was used as selection library.

To recover the bound sequences, the beads were resuspended in 100 µl affinity elution buffer (3 mM $MgCl_2$, 500 mM NaCl (overall, including the NaCl in PBS), 0.1 mg/ml salmon sperm DNA, 3% (v/v) ethanol, 7% (v/v) Cremophor EL, 5 mM THC in PBS, pH 7.4) and incubated for 30 min at 20°C. The beads were carefully pipetted up and down every 5 min to keep them in suspension and to prevent the formation of bubbles.

The supernatant was transferred to a new vial and precipitated by ethanol precipitation according to section 6.2.1.3. The pellet was resuspended in 55 µl ddH_2O, which were used as eluate as described above.

The composition of the RT-PCR-mastermix is given in **Table 29**.

6.2.3.5 SELEX 3

The selection was performed by Patrick Günther.

The selection buffer for SELEX 3 contained 50 mM KCl, 3 mM $MgCl_2$, and 1 mg/ml heparin in 50 mM HEPES, pH 8.0. GRNA was used as selection library.

After incubation and the washes, the beads were resuspended in 50 µl ddH_2O and used as eluate (including the beads) as described above.

The composition of the RT-PCR-mastermix is given in **Table 30**.

Table 30: Composition of the RT-PCR-mastermix for SELEX 3

Reagent	Stock concentration	Volume for 1 reaction	Final concentration in 100 µl
Colorless GoTaq buffer	5x	20 µl	1x
First strand buffer	5x	4 µl	0.2x
DTT	100 mM	2 µl	2 mM
GRNA-F	100 µM	1 µl	1 µM
GRNA-R	100 µM	1 µl	1 µM
$MgCl_2$	100 mM	1.5 µl	1.5 mM
dNTPs	25 mM (each)	1.2 µl	0.3 mM (each)
ddH_2O		17.3	

6.2.3.6 DNA-selections on amine-functionalised magnetic beads

The selection buffer for both SELEX 4 and 5 contained 50 mM KCl, 1 mg/ml heparin, and 3 mM $MgCl_2$ in 50 mM Hepes, pH 8.0. SELEX 4 used the D3-library, SELEX 5 was performed with the M111-library.

50 µl 10 mg/ml THCA-functionalised amine dynabeads (section 6.2.2.1) were washed twice with 100 µl selection buffer. The beads were resuspended in

1 nmol DNA-library in overall 200 µl selection buffer and incubated for 30 min at 20°C. To keep the beads in suspension, they were pipetted up and down every 5 min. The beads were then washed with 100 µl selection buffer for 30 sec, followed by a second wash step that lasted 3 min. Depending on the selection cycle, this wash procedure was repeated as detailed in **Table 24**.

The beads were resuspended in 50 µl ddH_2O and added to 945 µl of the PCR-mastermix (**Table 31**) along with 5 µl 5 u/µl GoTaq Flexi DNA polymerase. SELEX 4 used D3-F and D3-R-P, SELEX 5 M111-F and M111-R-P as primers. The sample was aliquoted in 10 samples à 100 µl and the PCR continued according to **Table 32**.

Table 31: Composition of the PCR-mastermix for SELEX 4 and 5

Reagent	Stock concentration	Volume for 10 reactions	Final concentration in 1000 µl
GoTaq Flexi buffer	5x	200 µl	1x
DTT	100 mM	20 µl	2 mM
5' primer	100 µM	10 µl	1 µM
3' primer	100 µM	10 µl	1 µM
$MgCl_2$	100 mM	15 µl	1.5 mM
dNTPs	25 mM (each)	12 µl	0.3 mM (each)
ddH_2O		678 µl	

Table 32: PCR-method for SELEX 4 to 7

Step	Temperature [°C]	Duration
1	95	3 min
2	95	30 sec
3	SELEX 4 and 6: 64	30 sec
	SELEX 5: 50	30 sec
	SELEX 7: 55	30 sec
4	SELEX 4 and 7: 72	30 sec
	SELEX 5: 72	1 min
	SELEX 6: 72	45 sec
5	72	2 min

Steps 2 to 4 were repeated according to **Table 33**.

The PCR was monitored by agarose gel electrophoresis as detailed in section 6.2.1.1. The beads were removed and the PCR-product cleaned up with a Gel and PCR clean-up kit according to the manufacturer's instructions. 200 µl PCR-product were loaded onto one column and eluted with two times 25 µl ddH_2O for overall 250 µl purified PCR-product. 25 µl were stored as backup at -20°C.

The single-strand displacement was performed according section 6.2.1.8 and the concentration of the eluate determined as detailed in section 6.2.1.7. 150 pmol DNA were used for selection cycle 2 to 5, 100 pmol for selection cycle 6 to 10.

Table 33: Number of PCR-cycles needed for amplification during SELEX 4 to 7

Selection cycle	PCR-cycles SELEX 4	PCR-cycles SELEX 5	PCR-cycles SELEX 6	PCR-cycles SELEX 7
1	21	14	28	19
2	21	13	22	18
3	20	14	17	16
4	20	15	18	18
5	16	16	17	15
6	16	17	15	15
7	16	12	14	13
8	16	13	15	15
9	16	15	14	15
10	17	14	14	15

For all subsequent selection cycles, a pre-selection step was performed. For this, 50 µl 10 mg/ml non-modified dynabeads were washed twice with 100 µl selection buffer. The ssDNA was added in final 100 µl selection buffer and incubated at 20°C for 30 min. To keep the beads in suspension, they were pipetted up and down every 5 min. The supernatant, which contained all sequences not binding to the matrix itself, was transferred onto 50 µl 10 mg/ml THCA-functionalised dynabeads that had also been washed twice with selection buffer beforehand. It was proceeded as detailed above.

6.2.3.7 DNA-selections on sepharose

The selection buffer for both SELEX 6 and 7 contained 137 mM NaCl, 3.93 mM Na_2HPO_4 * 2 H_2O, 1.47 mM KH_2PO_4, 18.53 mM KCl, and 3 mM $MgCl_2$ at a pH of 7.0. SELEX 6 was performed with the D3-library, SELEX 7 with the GDNA-library.

100 µl THC-sepharose (1:1-coupled, 50 µl pure resin, prepared according to section 6.2.2.3) was washed with 2 ml selection buffer. The THC-sepharose was resuspended in 1 nmol DNA-library in overall 200 µl selection buffer and incubated at 20°C for 15 min. To keep the THC-sepharose in suspension, it was pipetted up and down every 5 min. It was placed into an empty micro Bio-Spin column and the supernatant forced through by applying pressure with a Pasteur pipette rubber bulb. The THC-sepharose was washed with the volume of selection buffer indicated in **Table 34**. During washing, no external pressure was applied to increase the flow rate.

The THC-sepharose was resuspended in 200 µl 4.25 M Urea with 12.5 mM EDTA, pH 8.0, transferred to a new vial, and incubated at 65°C for 10 min. To keep the THC-sepharose in suspension, it was pipetted up and down after 5 min. The sample was then placed into an empty micro Bio-Spin column and the eluate collected quickly by applying pressure with a Pasteur pipette rubber bulb. The

eluate was precipitated by ethanol precipitation according to section 6.2.1.3 and resuspended in 50 µl ddH_2O.

Table 34: Volume of selection buffer used to wash the sepharose during SELEX 6 and 7

Selection cycle	Volume selection buffer [ml]
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10

The eluate was added to 945 µl of the PCR-mastermix (**Table 31**) along with 5 µl 5 u/µl GoTaq Flexi DNA polymerase. SELEX 6 used D3-F and D3-R-P, SELEX 7 GDNA-F and GDNA-R-P as primers. The sample was aliquoted in 10 samples à 100 µl and the PCR continued according to **Table 32**.

The PCR was monitored by agarose gel electrophoresis as detailed in section 6.2.1.1 and the PCR-product cleaned up with a Gel and PCR clean-up kit according to the manufacturer's instructions. 200 µl PCR-product were loaded onto one column and eluted with two times 25 µl ddH_2O for overall 250 µl purified PCR-product. 25 µl were stored as backup at -20°C.

The remaining sample was single-strand displaced according to section 6.2.1.8 and the concentration of the eluate determined as detailed in section 6.2.1.7. 150 pmol DNA were used for selection cycle 2 to 6, 100 pmol for selection cycles 7 and 8, and 65 pmol for selection cycles 9 and 10.

6.2.3.8 Preparation of radioactive DNA for interaction assays

DNA was prepared by PCR and single-strand displacement as detailed in sections 6.2.3.6, 6.2.3.7, and 6.2.1.8. 15 pmol ssDNA was radioactively phosphorylated (section 6.2.1.5) and the phosphorylation verified with a 10% PAGE-gel (section 0).

For SELEX 4 and 5, all radioactively labelled and purified DNA was loaded onto a preparative 10% PAGE-gel and proceeded as explained in section 0. Instead of overnight, the screen was only exposed for 30 min and a printout of the gel image used to localise and cut out the radioactive DNA bands. The resulting gel pieces were smashed with a pipette tip and soaked in 0.3 M NaOAc, pH 5.4 at 65°C and 1200 rpm for 1.5 h. The liquid was filtered through a syringe filled with glass wool

and the remaining gel pieces washed several times with 0.3 M NaOAc, pH 5.4. The sample was precipitated by ethanol precipitation (section 6.2.1.3) and the purified DNA resuspended in 50 µl ddH_2O and used for interaction analysis.

6.2.4 Click-selections

6.2.4.1 Cupper-catalysed alkyne-azide cycloaddition

A catalyst solution was prepared by addition of 4 µl 100 mM THPTA, 1 µl 100 mM $CuSO_4$, and 25 µl freshly prepared 100 mM sodium ascorbate to 70 µl ddH_2O. To enable reduction of Cu^{II} to Cu^{I}, the catalyst was incubated for 10 to 15 min at RT before use. The EdU-modified DNA (maximal 500 pmol) was mixed with 10 µl 10 mM azide and 10 µl 0.1 M phosphate buffer, pH 7 before addition of 10 µl freshly prepared catalyst solution. The sample was incubated for 1 h at 37°C and 650 rpm. For purification, Amicon Ultra centrifugal filter unit with MWCO 10 kDa were used according to the manufacturer's instructions.

6.2.4.2 PCR

PCRs with click-modified DNA were prepared according to **Table 35**. The eluate served as template during the selections. For amplification of the DNA of selection cycles, 2 µl of a 1:100-dilution of the backup of the DNA of the respective selection cycle were used as template for 100 µl PCR. For amplification of the DNA of single sequences, 1 µl of a 1:100-dilution of PCR-product I (section 6.2.6.5) were used as template for 100 µl PCR.

The PCR was aliquoted in 10 samples à 100 µl and the PCR continued according to **Table 36** using a PCR-block that had been preheated to 95°C. Steps 2 to 4 were repeated according to **Table 5** during the initial click-selections, according to **Table 14** during the affinity re-selections, and according to **Table 19** during the toggle-SELEX. Twenty PCR-cycles were performed for the amplification of DNA of single sequences and of selection cycles.

Table 35: Composition of the PCR-mastermix for the click-selections

Reagent	Stock concentration	Volume for 10 reactions	Final concentration in 1000 µl
Pwo PCR buffer	10x	100 µl	1x
FT2-F-Cy5	50 µM	10 µl	0.5 µM
FT2-R-P	50 µM	10 µl	0.5 µM
Pwo DNA polymerase	2.5 u/µl	10 µl	0.025 u/µl
dN*TPs	25 mM (each)	10 µl	0.25 mM (each)
ddH_2O		add up to 1000 µl	

dN*TPs = 1:1:1:1-mix of dATP, dGTP, dCTP, and EdUTP

Table 36: PCR-method for the click-selections

Step	Temperature [°C]	Duration
1	95	2 min
2	95	30 sec
3	62	30 sec
4	72	30 sec
5	72	2 min

Steps 2 to 4 were repeated according to the text.

The PCR was monitored by agarose gel electrophoresis as detailed in section 6.2.1.1 and the PCR-product cleaned up with a Gel and PCR clean-up kit according to the manufacturer's instructions. 200 µl PCR-product were loaded onto one column and eluted with two times 25 µl ddH_2O for overall 250 µl purified PCR-product.

6.2.4.3 General procedure for click-selections on sepharose

All click-selections were performed using the FT2-library. For each selection approach, one selection was performed using benzyl- and one CF_3-azide.

500 pmol of the library and the click-competitor were click-modified according to section 6.2.4.1 using D-PBS for the purification on the Amicon filters. Library and competitor were heated separately to 95°C for 3 min and cooled down for 5 min at RT. Library and competitor were mixed and 15 µl 10x selection buffer added for a final volume of ~150 µl.

100 µl THC-sepharose (1:1-coupled, 50 µl pure resin, prepared according to section 6.2.2.3) was washed with 2 ml selection buffer. The sepharose was resuspended in the 150 µl library and competitor in selection buffer and incubated at 20°C for 15 min. To keep the sepharose in suspension, it was pipetted up and down every 5 min. The sepharose was placed into an empty micro Bio-Spin column and the supernatant forced through by applying pressure with a Pasteur pipette rubber bulb. The sepharose was washed with the volume of selection buffer indicated in **Table 5**, **Table 14**, or **Table 19** depending on the selection. During washing, no external pressure was applied to increase the flow rate.

The recovery of the binding sequences is explained in sections 6.2.4.6 to 6.2.4.8.

PCR was performed according to section 6.2.4.2 using the eluate as template. 25 µl of the purified PCR-product were stored as backup at -20°C. The remaining PCR-product was single-strand displaced as detailed in section 6.2.1.8.

All purified ssDNA as well as 500 pmol click-competitor were click-modified for the subsequent selection cycle according to the start of this section.

6.2.4.4 Preparation of radioactive click-modified nucleic acids for interaction assays

DNA was prepared by PCR and single-strand displacement according to sections 6.2.4.2 and 6.2.1.8. After single-strand displacement, the DNA was eluted two times with 15 µl ddH_2O and all purified ssDNA was loaded onto an agarose gel (section 6.2.1.1). The gel was run for 30 min at 100 V and the bands were excised and purified using a Gel and PCR clean-up kit according to the manufacturer's instructions using four volumes NTC buffer. Fifteen pmol ssDNA was radioactively phosphorylated (section 6.2.1.5). The radioactively labelled DNA was click-modified according to section 6.2.4.1, but purified using the Gel and PCR clean-up kit with four volumes NTC buffer instead of Amicon filters. The ssDNA was eluted in four times 25 µl ddH_2O. The phosphorylation was verified with a 10% PAGE-gel (section 0).

6.2.4.5 Preparation of click-modified nucleic acids for fluorescence polarisation assays

DNA was prepared by PCR and single-strand displacement according to sections 6.2.4.2 and 6.2.1.8. The concentration was determined according to section 6.2.1.7.

6.2.4.6 Initial click-selections

The selection buffer for the two initial click-selections was composed of 1 mg/ml BSA, 0.1 % (v/v) Tween-20, and 0.1 mg/ml salmon sperm DNA in 1x D-PBS, pH 5.4. 10x selection buffer contained 10 mg/ml BSA, 1% (v/v) Tween-20, and 1 mg/ml salmon sperm DNA in 1x D-PBS, pH 5.4.

For recovery of the binding sequences, the THC-sepharose was resuspended in 200 µl ddH_2O, transferred to a new vial, and incubated at 80°C for 3.5 min. The sample was then placed into an empty micro Bio-Spin column and the eluate collected quickly by applying pressure with a Pasteur pipette rubber bulb. The eluate was used as PCR-template as described in section 6.2.4.3.

6.2.4.7 Re-selections using affinity elution

The re-selections were performed using the same selection buffer as the initial click-selections (section 6.2.4.6).

The backup DNA from selection cycle 4 of the initial click-selections was re-amplified in 1 ml total PCR-volume per selection and single-strand displaced

according to sections 6.2.4.2 and 6.2.1.8. The resultant ssDNA was used for selection cycle 5.

The selection was performed as explained in section 6.2.4.3 with incubation times according to **Table 14**.

For recovery of the binding sequences, the THC-sepharose was resuspended in 150 µl D-PBS with 5% (v/v) ACN, 0.1% (v/v) Tween, and the concentration of THC indicated in **Table 14**. The sample was heated to 80°C for 3 min and cooled down for 5 min at 20°C. It was then incubated at 20°C for a time equal to the incubation time during that selection cycle (**Table 14**). The THC-sepharose was transferred to a new empty micro BioSpin column and the flow-through collected by applying pressure with a Pasteur pipette rubber bulb. The flow-through was used as eluate for the PCR as described in section 6.2.4.3.

6.2.4.8 Toggle-SELEX

The selection buffer for the toggle-SELEX consisted of 137 mM NaCl, 3.93 mM Na_2HPO_4 * 2 H_2O, 1.47 mM KH_2PO_4, 18.53 mM KCl, 3 mM $MgCl_2$, and 0.1 mg/ml salmon sperm DNA at a pH of 7.0.

500 pmol of the library and the click-competitor were click-modified according to section 6.2.4.1 using the selection buffer without salmon sperm DNA for the purification on the Amicon filters. Library and competitor were heated separately to 95°C for 3 min and cooled down for 5 min at RT. Library and competitor were mixed, filled up to 148.5 µl and 1.5 µl 10 mg/ml salmon sperm DNA added for a final volume of ~150 µl.

The volume of solid support (sepharose: 50% (v/v) resin, 50% ddH_2O, prepared according to section 6.2.2.3) indicated in **Table 19** was washed with **A** ml selection buffer without salmon sperm DNA (**Table 37**). The matrix was resuspended in the 150 µl library and competitor in selection buffer and incubated at 20°C for the amount of time indicated in **Table 19**. To keep the solid support in suspension, it was pipetted up and down every 5 min.

The wash steps were performed as described in section 6.2.4.3. Magnetic beads were placed onto a magnet.

The solid support was washed with selection buffer with 5% (v/v) ACN and without salmon sperm DNA as detailed in **B** (**Table 37**).

Table 37: Volume of wash steps in selection cycles performed with sepharose and magnetic beads.

Step	THC-sepharose	THCA-beads
A	1 ml	0.2 ml
B	Time and volume according to **Table 19**	Time according to **Table 19,** 200 µl for each step

For heat elution, the solid support was resuspended in 200 µl ddH_2O and heated for 3 min at 80°C. The eluate was collected quickly.

For affinity elution, the magnetic beads were resuspended in 200 µl selection buffer with 5% (v/v) ACN and the THC-concentration indicated in **Table 19**, but without salmon sperm DNA. They were incubated at 80°C for 3 min and cooled at RT for 5 min. The beads were incubated for an amount of time equal to the incubation time of the respective selection cycle at 20°C (**Table 19**). The supernatant was used as eluate.

PCR was performed according to section 6.2.4.2 using the eluate as template. 25 µl of the purified PCR-product were stored as backup at -20°C. The remaining PCR-product was single-strand displaced as detailed in section 6.2.1.8.

All purified ssDNA as well as 500 pmol click-competitor were click-modified for the subsequent selection cycle.

If selection cycles contained a pre-selection step on the unmodified solid support (selection cycle 12 to 15), the ~150 µl library and click-competitor were added to the unmodified solid support as indicated in **Table 19**, which had been washed with **A** ml selection buffer without salmon sperm DNA beforehand. The sample was incubated at 20°C for the amount of time indicated in **Table 19**. To keep the solid support in suspension, it was pipetted up and down every 5 min. The supernatant was collected and added again to unmodified solid support, which had been washed with **A** ml selection buffer without salmon sperm DNA beforehand. The sample was incubated at 20°C for the amount of time indicated in **Table 19**. To keep the solid support in suspension, it was pipetted up and down every 5 min. The supernatant was collected and added to the modified solid support. The selection proceeded as described above.

6.2.5 Cherenkov for interaction analysis of radioactively labelled nucleic acids

1/10 volume 10x selection buffer (10x D-PBS for all click-selections but the toggle-SELEX) was added to the radioactively labelled DNA and to the click-modified click-competitor with the same modification (prepared according to section 6.2.4.1) for a final concentration of 1x D-PBS or 1x selection buffer. 5 µl

radioactively labelled DNA and the click-competitor were separately heated to 95°C for 3 min and cooled at RT for 5 min.

100 µl THC- or unmodified sepharose (50 µl pure resin) was placed in an empty micro BioSpin column and washed with 1 ml selection buffer. The sepharose was resuspended in 5 µl 2'F-RNA or DNA and 1 µl 10 µM clicked competitor (only for click-selections) in final 150 µl selection buffer. The sample was incubated for 15 min at 20°C. To keep the sepharose in suspension, the sample was pipetted up and down every 5 min. The sepharose was transferred into an empty column and the flow-through collected by applying pressure with a Pasteur pipette rubber bulb. Two washes with 1 ml selection buffer each were left to run through without application of pressure and collected separately. The sepharose was transferred into a new vial using two times 200 µl ddH_2O.

All samples were added up to 1 ml with ddH_2O and the vials place in liquid scintillation vials. The ^{32}P-signal was measured using a liquid scintillation counter. Counts per minute were measured for 3 min per sample.

The Cherenkov for samples containing ACN, no Tween, no BSA, and those measured at pH 7.4 was performed identically with the modifications only applying to the selection buffer.

For assays using non-nucleobase-modified nucleic acids (2'F-RNA and DNA), the initial addition of 10x selection buffer and the folding step (95°C 3 min, RT 5 min) were not performed and the click-competitor not added.

Assays using magnetic beads were performed using 25 µl 10 mg/ml magnetic beads and two washes with 100 µl instead of 1 ml selection buffer each. Instead of the columns, the vials containing the beads were placed on a magnet to remove the supernatant.

All assays were performed at least twice.

6.2.6 Cloning and sequencing

6.2.6.1 Cloning

The DNA of the selection cycle was amplified in a 100 µl PCR as detailed in **Table 38**. 0.3 µl of a 1:100-dilution of the backup DNA of the respective selection cycle were used as template. The PCR was performed in overall 18 cycles according to **Table 36**. The final elongation step at 72°C was set to last 20 instead of 2 min.

The success of the PCR was verified by agarose gel electrophoresis (section 6.2.1.1).

Table 38: Composition of the PCR-mastermix for cloning

Reagent	Stock concentration	Volume for 1 reactions	Final concentration in 100 µl
Taq buffer	10x	10 µl	1x
FT2-F	50 µM	1 µl	0.5 µM
FT2-R	50 µM	1 µl	0.5 µM
Taq DNA polymerase	2.5 u/µl	1 µl	0.025 u/µl
dNTPs	25 mM (each)	1 µl	0.25 mM (each)
$MgCl_2$	25 mM	6 µl	1.5 mM
ddH_2O		79.7 µl	

1 µl PCR-product was mixed with 1 µl salt solution (1.2 M NaCl, 0.06 M $MgCl_2$, part of the kit), 3 µl ddH_2O, and 1 µl TOPO TA vector. The sample was incubated at 23°C and 300 rpm for 30 min and placed on ice afterwards.

One aliquot TOP10 bacteria was thawed slowly on ice. 2 µl of the cloning reaction were added to 1/3 aliquot bacteria without pipetting up and down and incubated on ice for 10 min. The bacteria were heat shocked for 30 sec at 42°C and placed on ice for 5 min immediately afterwards. 250 µl room-tempered SOC medium was added to the sample and incubated for 1 h at 37°C, 300 rpm. The bacteria were spread onto a 100 µg/ml ampicillin plate and grown over night at 37°C.

6.2.6.2 Colony-PCR

Colonies were numbered and placed with a pipette tip into 5 µl ddH_2O. After heat treatment for 10 min at 95°C, the bacteria were spun down for 5 min at 14,000 rpm. 3 µl of the supernatant was used for a 20 µl PCR according to **Table 38** with 17 PCR-cycles as described in **Table 36**. The success of the PCR was verified by agarose gel electrophoresis (section 6.2.1.1) and the colonies with a PCR-product of the expected length, which contain the desired insert, identified.

6.2.6.3 Overnight culture and plasmid preparation

The insert-containing colonies were picked again and transferred into a falcon containing 5 ml LB medium with 100 µg/ml ampicillin. They were grown overnight at 37°C and 115 rpm.

The overnight culture was spun down for 15 min at 3250 rpm and purified with a Plasmid kit according to the manufacturer's instructions. The plasmid was eluted in two times 25 µl ddH_2O.

6.2.6.4 Sanger sequencing

Sanger sequencing was performed by GATC Biotech, Konstanz, Germany. The random region was identified using the known primer binding sites of the library.

The obtained random region sequences were analysed for sequence similarity by multiple sequence alignment (Clustal Omega[276,277]) and motif recognition software (MEME[278]).

6.2.6.5 Preparation of DNA for interaction assays from plasmids

1 µl 1:10-diluted purified plasmid was used as a template for a 50 µl PCR according to **Table 38** with 18 PCR-cycles as described in **Table 36**. The resultant PCR-product I was diluted 1:100 and 1 µl used for 100 µl PCR.

6.2.6.6 Preparation of next-generation sequencing samples

Samples for NGS were prepared and sequenced in different runs: 1) THC-click-SELEX samples, 2) EdU_Pwo, 3) T_Pwo, T_Taq, and T_w/o, 4) FT2_GATC and FT2_G4A4T4C4, 5) FT2_GATC_II, FT2_G4A4T4C4_II, FT2_G3A3T3C3, FT2_G2A2T2C2, FT2_TGCA, D3_TGCA, FT2_T4G4C4A4, and D3_T4G4C4A4, and 5) GFP-click-SELEX samples for **Figure 34**.

All samples were prepared as published.[279] Briefly, the DNA from the respective selection cycles was amplified by PCR with primers containing 5'-indices (index-PCR). **Table 39** describes which index was used for which selection cycle of the THC-click-SELEX. **Table 40** describes the indices used for the mutational analysis study. The purified PCR-products with different indices were pooled and adaptors, which enable hybridisation to the sequencing flow cell, annealed to the 5'-ends. After gel purification, the sample-concentration was measured by qPCR.

Index-PCR was performed using Pwo DNA polymerase for all samples but T_Taq, T_w/o, and all repetitive sequences. C12_EdU was used as template for EdU_Pwo. C12_T was used as template for T_Taq and T_Pwo. For T_Taq, Taq DNA polymerase was used for amplification and large Klenow fragment was utilised according to the manufacturer's instructions for blunt end generation. T_w/o and all repetitive sequences were obtained as sense and antisense strand including the indices and annealed. For annealing, 100 pmol of each strand were mixed in 40 mM Tris, pH 7.9 and heated to 95°C for 5 min. Afterwards, the strands were cooled down to 4°C in 30 min (0.05°C/s). Annealing was confirmed by agarose gel electrophoresis (section 6.2.1.1).

Table 39: Indices used for NGS-analysis of the initial click-selections

Selection cycle	Lane	Index number	Index sequence[279]
4 (benzyl)	2	9	GATCAG
6 (benzyl)	2	10	TAGCTT
8 (benzyl)	2	11	GGCTAC
10 (benzyl)	2	12	CTTGTA
4 (CF_3)	1	6	GCCAAT
6 (CF_3)	1	7	CAGATC
8 (CF_3)	2	6	GCCAAT
10 (CF_3)	2	7	CAGATC

Table 40: Indices used for NGS-analysis of the mutational analysis study

sample name	index[279]	index seq.	primer sites from library	random region
EdU_Pwo	6	GCCAAT	FT2[205]	see C12 in section 6.1.6
T_Pwo	10	TAGCTT	FT2	see C12 in section 6.1.6
T_Taq	11	GGCTAC	FT2	see C12 in section 6.1.6
T_w/o	12	CTTGTA	FT2	see C12 in section 6.1.6
FT2_GATC	12	CTTGTA	FT2	$(GATC)_8$
FT2_GATC_II	12	CTTGTA	FT2	$(GATC)_8$
FT2_G4A4T4C4	11	GGCTAC	FT2	$(GGGGAAAATTTTCCCC)_2$
FT2_G4A4T4C4_II	11	GGCTAC	FT2	$(GGGGAAAATTTTCCCC)_2$
FT2_G2A2T2C2	6	GCCAAT	FT2	$(GGGAAATTTCCC)_2$GGGAAATT
FT2_G3A3T3C3	5	ACAGTG	FT2	$(GGAATTCC)_4$
FT2-TGCA	8	ACTTGA	FT2	$(TGCA)_8$
D3-TGCA	10	TAGCTT	D3[280]	$(TGCA)_8$
FT2-T4G4C4A4	7	CAGATC	FT2	$(TTTTGGGGCCCCAAAA)_2$
D3-T4G4C4A4	9	GATCAG	D3	$(TTTTGGGGCCCCAAAA)_2$

6.2.6.7 Next-generation sequencing

Next-generation sequencing was performed by Joachim L. Schultze, Kristian Händler, and Marc Beyer (University of Bonn, Germany) on a HiSeq1500 as described in Pfeiffer *et al*.[281] 76 base pairs and 7 index bases were sequenced.

6.2.6.8 Analysis of next-generation sequencing

Analysis of NGS-data was performed by Carsten Gröber and Michael Blank (AptaIT GmbH, Planegg, Germany) using the software COMPAS.[154,282] The analysis is detailed in Pfeiffer *et al*.[281] For the analysis of samples after exclusion

of shortened sequences, only sequences of the correct length and longer were analysed.

The frequency of unique sequences was calculated by dividing the number of unique sequences by the total number of sequences analysed in the respective selection cycle and multiplying with 100.

The frequency of mutated sequences was calculated by calculating the percentage of non-mutated sequences and subtracting it from 100%.

The error rate was calculated by subtracting the frequency of the correct nucleotide at the specified position from 1. The mutation frequency of a specific nucleotide was defined as ‘mutated nt’. The average and SD of the mutation frequency of all nucleotides in a sequence is the ‘error rate’.

The ‘mutated into nt’ was calculated by averaging the frequency of that nucleotide at all positions where it was not the original nucleotide. The conversion from one nucleotide into another nucleotide was calculated alike, but by restricting the positions to those of the original nucleotide.

The mutation frequency into the subsequent nucleotide was calculated by calculating 100/(1-frequency correct nucleotide)*frequency subsequent nt for each position and then calculating the average.

All data is given as average ± SD.

Statistical analysis was performed by testing normality using the Shapiro-Wilk normality test. Normally distributed datasets were analysed by one-way ANOVA followed by two-tailed t-tests. Not normally distributed datasets were analysed using the Kruskal-Wallis test followed by two-tailed Mann-Whitney tests. Alpha was set to 0.05 for all tests.

6.2.7 HPLC-(MS)-analysis

6.2.7.1 HPLC-measurements for solubility determination

Apart from the measurements leading to **Figure 35A**, two samples of 30 µl each were prepared with the respective buffer/organic solvent, Tween-20- and THC(-FITC)-concentration. For all samples but those for **Figure 62**, D-PBS, pH 5.4 was used as buffer. For **Figure 62**, the buffer consisted of 137 mM NaCl, 3.93 mM Na_2HPO_4 * 2 H_2O, 1.47 mM KH_2PO_4, and 18.53 mM KCl at a pH of 7.0. The samples were prepared in HPLC-vials. Each sample was injected twice, leading to overall four measurements per concentration. The samples for **Figure 35A** were pipetted once and injected once.

For the measurement, 10 µl of the sample were injected on an analytical HPLC system (1260 Infinity) with an EC/50/4.6 Nucleodur 100-5 C18ec column. Buffer A consisted of ddH_2O with 0.1% (v/v) formic acid, buffer B of ACN. The gradient was described for both the first and second method in **Table 41** and the total runtime of the first method was 56 min and of the second method 30 min with a flowrate of 0.8 ml/min each.

The samples leading to **Figure 35** and **Figure 36** were measured according to the first method, all others according to the second method.

Table 41: Timetable of the HPLC-measurements

Time [min] – 1st	Time [min] – 2nd	Buffer A [%]	Buffer B [%]
0	**0**	60	40
10	**5**	30	70
40	**20**	0	100
45	**23**	0	100
46	**24**	60	40

6.2.7.2 LC-MS-analysis

For THC-L, the samples from the different days (section 6.2.2.2) were centrifuged at 14000 rpm for 5 min and the supernatant discarded. The pellet was resuspended in 80% ACN with 20% ddH_2O with 0.1% (v/v) formic acid. The samples from day 7 were left untreated.

For THC-FITC, 0.3 µl sample were mixed with 100 µl 80% ACN with 20% ddH_2O with 0.1% (v/v) formic acid for analysis before clean-up. After clean-up, 2 µl of the purified sample were mixed with 48 µl 80% ACN with 20% ddH_2O with 0.1% (v/v) formic acid.

For TLF, the sample before clean-up was taken according to section 6.2.8.3, the one after clean-up according to section 6.2.8.4.

10 µl of each sample were injected for analysis on an analytical 1100 series HPLC system with a Zorbax_50x2.1_5 µm column. Buffer A consisted of ddH_2O with 0.1% (v/v) formic acid, buffer B of ACN. The gradient was described in **Table 42** and the total runtime of the method 60 min with a flowrate of 0.2 ml/min. Mass spectrometry was performed from min 5 to min 30 with an HTC Esquire using ultra mode (alternating). The settings were as follows: nebulizer: 40 psi, dry gas: 9 l/min , dry temperature 365°C, target mass 400 m/z, ICC negative: 70000, ICC positive: 200000, analysed mass 90-1000 m/z.

For TLF, target mass was set to 900 m/z with analysed mass 300-1700 m/z. Everything else was as described in the preceding paragraph.

The UV-peaks were integrating using Compass DataAnalyser.

Table 42: Timetable of the LC-MS-analysis

Time [min]	Buffer A [%]	Buffer B [%]
0	40	60
30	0	100
48	0	100
50	40	60

6.2.8 Synthesis of fluorescently labelled THC-compounds

6.2.8.1 THC-FITC

The coupling of THC to FITC was based on the publications by Taldone *et al.* and Llauger-Bufi *et al.* and is depicted in **Figure 41**.[283,284] 9.54 µmol THC were dried in a speedvac at 45°C and solubilized in 30 µl DMF. 47.7 µmol FITC (5 times molar excess) were solubilised in 50 µl DMF. THC and FITC were mixed, 20 µl TEA added, and the reaction rotated at 65°C for 2 days.

0.5 µl sample were mixed with 100 µl 80% ACN with 20% ddH_2O with 0.1% (v/v) formic acid, spun down for 5 min at 18000 rpm and 30 µl of the supernatant used for LC-MS-analysis (section 6.2.7.2).

6.2.8.2 HPLC-purification of THC-FITC

1137.5 µl 50% ddH_2O with 50% (v/v) ACN were added to the remaining coupling reaction.

99 µl of the sample were injected per run on an analytical HPLC system (1260 Infinity) with an EC/50/4.6 Nucleodur 100-5 C18ec column. Buffer A consisted of ddH_2O with 0.1% (v/v) formic acid, buffer B of ACN. The gradient was described in **Table 43** and the total runtime of the method 70 min with a flowrate of 0.8 ml/min. The peak at minute 36.5 to 37.5 was collected.

Table 43: Timetable of the HPLC-purification of THC-FITC

Time [min]	Buffer A [%]	Buffer B [%]
0	60	40
10	40	60
50	0	100
55	0	100
56	60	40

All fractions were dried in a speedvac at 45°C and resuspended in overall 100 µl ACN. 2 µl were mixed with 48 µl 80% ACN with 20% ddH_2O with 0.1% (v/v)

formic acid for LC-MS-analysis (section 6.2.7.2). The remaining sample was stored at 4°C.

6.2.8.3 TLF

The synthesis of TLF is depicted in **Figure 55**. 12.5 µmol THC-L (non-purified, prepared according to section 6.2.2.2) were lyophilised overnight and resuspended in 50 µl 7 N NH_3 in MeOH. The reaction was incubated for 4 h at 70°C in a PCR-cycler with the lid heated to 110°C to prevent evaporation. Afterwards, the sample was cooled down to 4°C and dried in a speedvac overnight at 4°C to remove the unreacted NH_3.

100 µmol FITC (38.94 mg) were solubilised in 1.56 ml MeOH. The reaction product (THC-L-NH_2) was solubilised in 2.34 ml DMF and FITC (in MeOH) and reaction product mixed. The reaction was rotated for 3 h at RT (25.5°C) and dried in a speedvac.

The reaction product was resuspended in 100 µl ACN, centrifuged for 5 min at 18000 rpm and 2.2 µl of the supernatant mixed with 47.8 µl 80% ACN with 20% ddH_2O with 0.1% (v/v) formic acid for LC-MS-analysis (section 6.2.7.2).

6.2.8.4 HPLC-purification of TLF

33 ml 80% ACN with 20% ddH_2O with 0.1% (v/v) formic acid were added to the coupling reaction (supernatant and pellet combined). The sample was incubated for 30 min at 65°C, cooled down and centrifuged for 10 min at 5000 g. The supernatant was used for purification and the small pellet discarded.

99 µl of the sample were injected per run on an analytical HPLC system (1260 Infinity) with an EC/50/4.6 Nucleodur 100-5 C18ec column. Buffer A consisted of ddH_2O with 0.1% (v/v) formic acid, buffer B of ACN. The gradient was described in **Table 44** and the total runtime of the method 30 min with a flowrate of 1 ml/min. The peak at minute 11 to 12 was collected.

Table 44: Timetable of the HPLC-purification of TLF

Time [min]	Buffer A [%]	Buffer B [%]
0	60	40
5	30	70
20	0	100
23	0	100
24	60	40

All collected samples were pooled and dried, first in a vacuum distillation rotator to reduce the volume, then in a speedvac overnight at RT.

TLF was resuspended in 500 µl ACN, centrifuged for 3 min at 14000 rpm and 2.2 µl of the supernatant mixed with 47.8 µl 80% ACN with 20% ddH_2O with 0.1% (v/v) formic acid used for LC-MS-analysis (section 6.2.7.2).

500 µl ddH_2O were added to the TLF (pellet and supernatant), incubated at 65°C for 30 min, and centrifuged for 10 min at 14000 rpm. The supernatant was used for all assays and stored at 4°C. The pellet was discarded.

6.2.9 Absorption and emission scans of fluorescently labelled THC-compounds

6.2.9.1 Absorption scans

THC-FITC (prepared according to sections 6.2.8.1 and 6.2.8.2) was diluted 1:20 in ddH_2O (final 5% (v/v) ACN). The blank consisted of 5% (v/v) ACN in ddH_2O.

TLF (prepared according to sections 6.2.8.3 and 6.2.8.4, 50% ddH_2O, 50% (v/v) ACN) was diluted 1:10 in either ddH_2O, D-PBS, pH 5.4, or PBS, pH 7.4. The blanks consisted of equally prepared dilutions of 50% ddH_2O with 50% (v/v) ACN.

For both compounds, 2 µl in duplicates were spotted onto a NanoQuant Plate and the absorption measured between 400 and 600 nm with a NanoQuant infinite M200.

The concentration was determined using the absorption at 492 nm (in ddH_2O for TLF) and the molar extinction coefficient of FITC (ε = 75000 L/(mol*cm)).[285] The blank (in ddH_2O for TLF) was subtracted from the absorption before calculation.

6.2.9.2 Emission scans

The fluorescence intensity was measured with an EnSpire using a 384 well Perkin Elmer Proxi Plate 384 F Plus. 15 µl of a 1:20-dilution (THC-FITC) or a 1:10-dilution (TLF) in either ddH_2O, D-PBS, pH 5.4, or PBS, pH 7.4, were loaded per well and all samples measured in duplicates. The samples were excited at 492 nm and the emission measured from 510 to 600 nm.

6.2.10 Fluorescence polarisation assays

For all assays, a final volume of 15 µl in one well of a black 384 well Perkin Elmer Proxi Plate 384 F Plus was measured in a Tecan Ultra. Excitation was set to 485 nm, emission to 535 nm. The number of flashes was 20, lag time 0 µs, integration time 40 µs, z-position 12300 µm, and G-factor 1.06. The gain varied depending on the experiment and is indicated in the respective subsection.

Fluorescence polarisation was calculated as follows: The parallel and perpendicular fluorescence intensity of the blank (without fluorophore, but identical buffer to the rest of the samples) was subtracted from the parallel and perpendicular intensities of the samples. Then, polarisation in mP was calculated according to **Equation 1**. $F_{||}$ is the parallel intensity, $F_{\perp}$ the perpendicular intensity.

Equation 1: Calculation of fluorescence polarisation

$$\frac{F_{\parallel} + F_{\perp}}{F_{\parallel} - F_{\perp}} \times 1000 = FP$$

Measurements were performed at least as n = 2 with duplicates each. Each plate was measured twice.

6.2.10.1 Fluorophore concentration ranges

The desired concentration of fluorophore was measured in D-PBS, pH 5.4 with 5% (v/v) ACN and with or without 0.1% (v/v) Tween-20. For measurements at pH 7.4, PBS, pH 7.4 with or without 0.1% (v/v) Tween-20 was used. The gain was set to 55.

6.2.10.2 Test of different buffer components

All experiments were performed in D-PBS, pH 5.4 with the indicated additives. The gain was set to 45.

6.2.10.3 K_d-determination of the THC-antibody

PBS, pH 7.4 was used as buffer. Antibody-dilutions of 1.5 times the desired concentration and fluorophore-dilutions of 3 times the desired concentration were prepared. 10 µl antibody-dilution was mixed with 5 µl fluorophore-dilution in the 384-well plate and incubated for 30 min at RT. The gain was set to 45 for the measurement using THC-FITC and to 52 for TLF.

6.2.10.4 Interaction analysis of the THC-sepharose-binding sequences

For measurements at pH 5.4, 7.5 pmol DNA (in ddH_2O) was diluted to a final volume of 10 µl with ddH_2O in a well of a 384-well plate. For the control, 1.5 µl 10 mg/ml BSA was mixed with 8.5 µl ddH_2O.

A fluorophore-mastermix was prepared according to **Table 45**. 5 µl fluorophore-mastermix were added per well for final 15 µl sample per well and the sample was incubated for 15 min at RT before measurement.

Table 45: Composition of the fluorophore-mastermix at pH 5.4

Reagent	Stock concentration	Volume for 10 reactions	Final concentration in 1 well
THC-FITC (in ACN)	6.625 µM	1.2 µl	53 nM
TLF (in ACN)	41.13 µM	0.29 µl	80 nM
D-PBS, pH 5.4	10x	14.25 µl	0.95x
Salmon sperm DNA	10 mg/ml	1.5 µl	0.1 mg/ml
ACN	100%	Add volume THC-FITC/TLF up to 7.5 µl	5% (v/v) (incl. ACN from THC-FITC/TLF)
ddH_2O		Add up to 50 µl	

For measurements at pH 7.4, the DNA was dried in a speedvac and 7.5 pmol DNA resuspended in 10 µl D-PBS, pH 7.4. The antibody was diluted 1:10 in D-PBS, pH 7.4. 0.48 µl dilution were mixed with 9.52 µl D-PBS, pH 7.4 per well. 1.5 µl 10 mg/ml BSA were mixed with 8.5 µl D-PBS, pH 7.4 for the control. The fluorophore-mastermix was prepared as described in **Table 46** and 5 µl added per sample. The plate was incubated for 15 min at RT before measurement.

Assays including the THC-antibody as positive control were incubated for 30 instead of 15 min at RT.

Table 46: Composition of the fluorophore-mastermix at pH 7.4

Reagent	Stock concentration	Volume for 10 reactions	Final concentration in 1 well
THC-FITC (in ACN)	6.625 µM	1.2 µl	53 nM
TLF (in ACN)	41.13 µM	0.29 µl	80 nM
Salmon sperm DNA	10 mg/ml	1.5 µl	0.1 mg/ml
ACN	100%	Add volume THC-FITC/TLF up to 7.5 µl	5% (v/v) (incl. ACN from THC-FITC/TLF)
D-PBS, pH 7.4	1x	41 µl	

Measurements using THC-FITC at either pH were performed using a gain of 45. Measurements using TLF were performed using a gain of 69 at pH 5.4 and 56 at pH 7.4.

For assays containing streptavidin, 0.67 µl 16.7 µM (1.5x molar excess) streptavidin were added to the sample before addition of the fluorophore-mastermix. The amount of ddH_2O or D-PBS, pH 7.4 was decreased accordingly for a final volume of 10 µl per well. The plate was incubated for 40 min at RT.

Then, 5 µl of the fluorophore-mastermix was added and the plate incubated for 15 min at RT before measurement.

7 Appendix

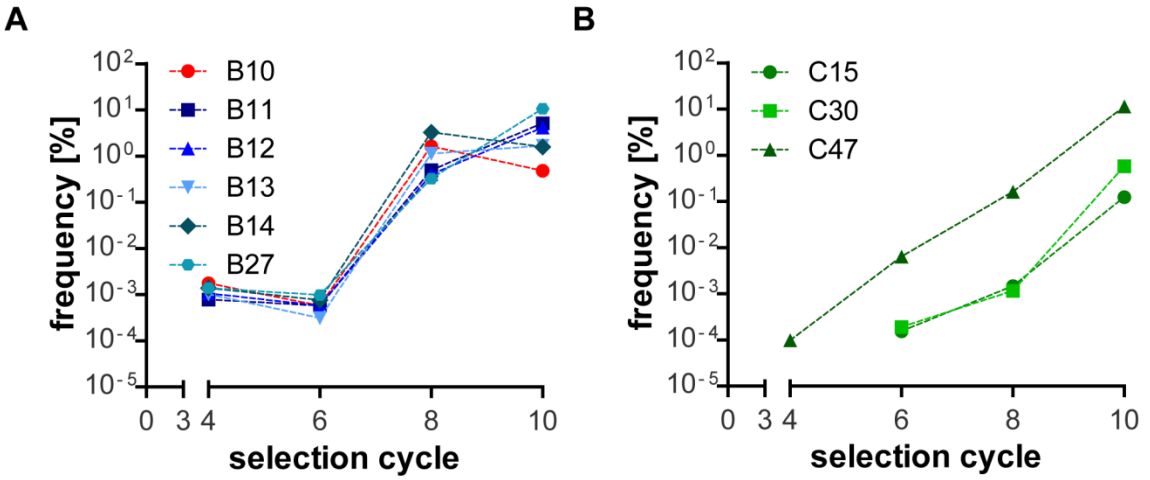


Figure A 1: Frequency of the tested sequences according to the NGS-analysis
Frequency in the analysed selection cycles 4, 6, 8, and 10. (A) Frequency of all tested sequences from the initial selection with benzyl-modified DNA. The non-binding B10 is depicted in red, the binding sequences in shades of blue. (B) Frequency of all tested sequences from the initial selection with CF_3-modified DNA. The binding sequences are depicted in shades of green. C15 and C30 could not be detected in selection cycle 4.

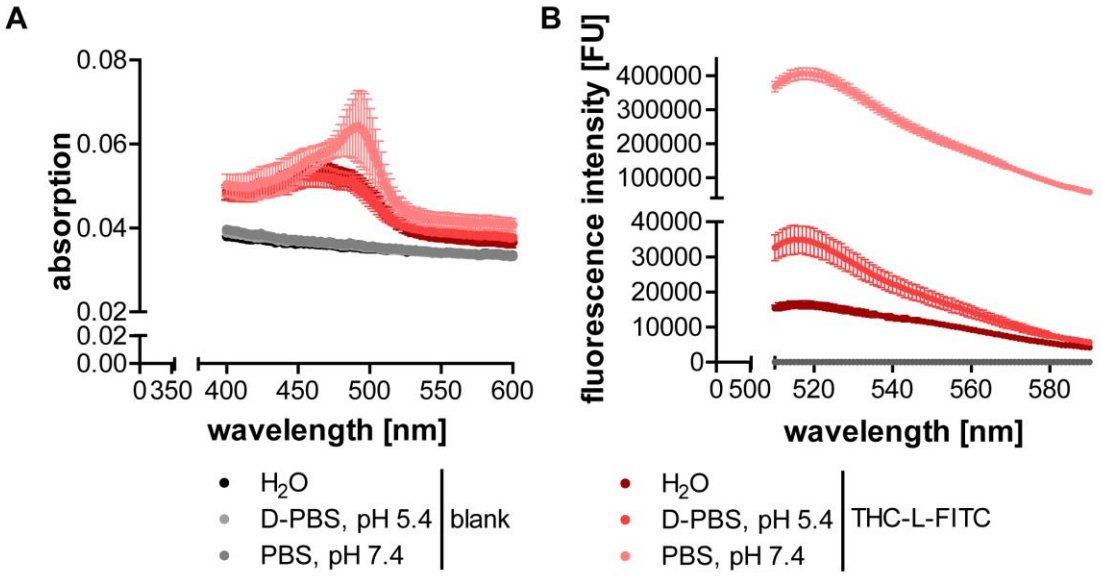


Figure A 2: Absorption and emission spectra of THC-L-FITC
(A) Absorption spectrum of THC-L-FITC (shades of red) (d = 0.05 cm) in different solvents (shades of grey and black). (B) Emission spectrum of THC-L-FITC (excitation at 492 nm, depicted in shades of red) in different solvents (shades of grey and black). All samples contained 5% ACN to solubilize THC-L-FITC. Samples were measured in the following solvents: ddH2O (dark red and black), D-PBS, pH 5.4 (medium red and light grey), and PBS, pH 7.4 (light red and dark grey).

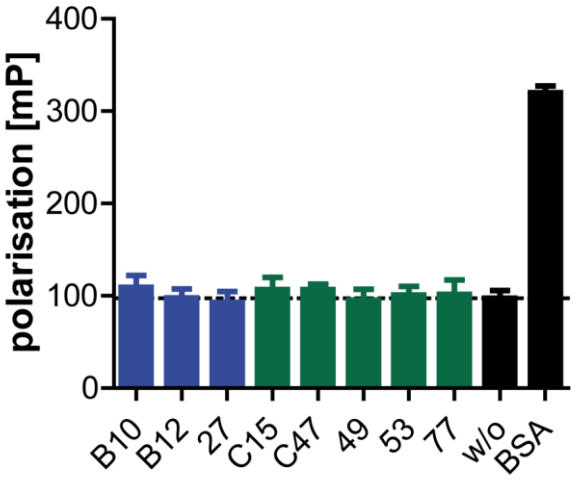


Figure A 3: Fluorescence polarisation of THC-FITC at pH 7.4 in the presence of different sequences

The fluorescence polarisation of THC-FITC (53 nM) was measured in the presence of 500 nM of the different sequences that bind at pH 7.4, 1 mg/ml BSA, or without any sequence (w/o). BSA was used as a positive control. No sequence (w/o) was used as negative control and the dotted line represents the corresponding polarisation. Benzyl-modified DNA is depicted in blue, CF_3-modified DNA in green.

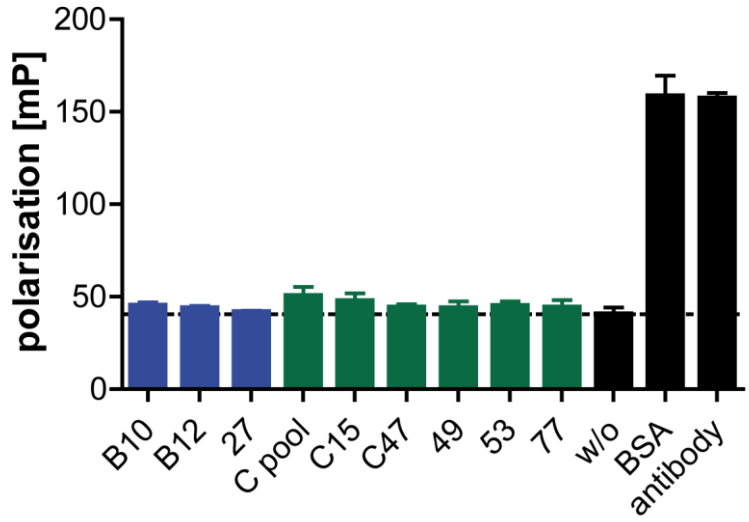


Figure A 4: Fluorescence polarisation of THC-L-FITC at pH 7.4 in the presence of different sequences

The fluorescence polarisation of THC-L-FITC (80 nM) was measured in the presence of 500 nM of the different sequences that bind at pH 7.4, 1 mg/ml BSA, 80 nM THC-antibody, or without any sequence (w/o). BSA and the THC-antibody were used as positive controls. No sequence (w/o) was used as negative control and the dotted line represents the corresponding polarisation. Benzyl-modified DNA is depicted in blue, CF_3-modified DNA in green. C pool indicates FT2-pool that has been CF_3-modified.

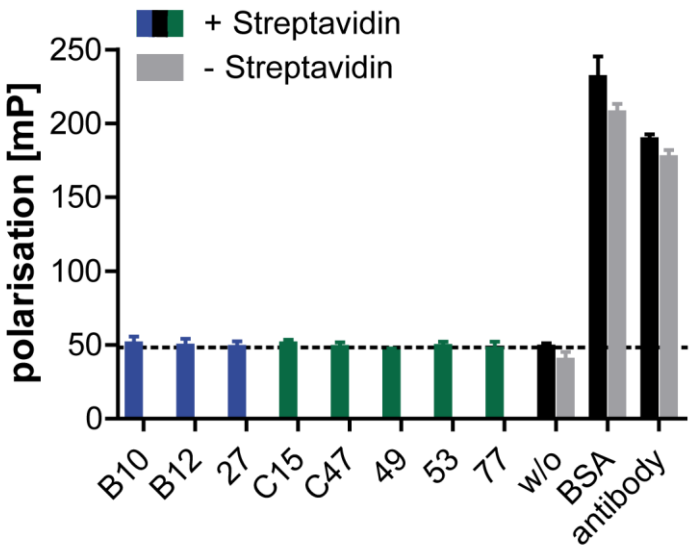


Figure A 5: Fluorescence polarisation of THC-L-FITC at pH 7.4 in the presence of different biotinylated sequences and streptavidin
The fluorescence polarisation of THC-L-FITC (80 nM) was measured in the presence of 500 nM of the different, biotinylated sequences, 1 mg/ml BSA, 80 nM THC-antibody, or without any sequence (w/o). All assays but those represented by grey bars contained 750 nM streptavidin. BSA and the THC-antibody were used as positive controls. No sequence (w/o) was used as a negative control and the dotted line represents the corresponding polarisation. Benzyl-modified DNA is depicted in blue, CF_3-modified DNA in green.

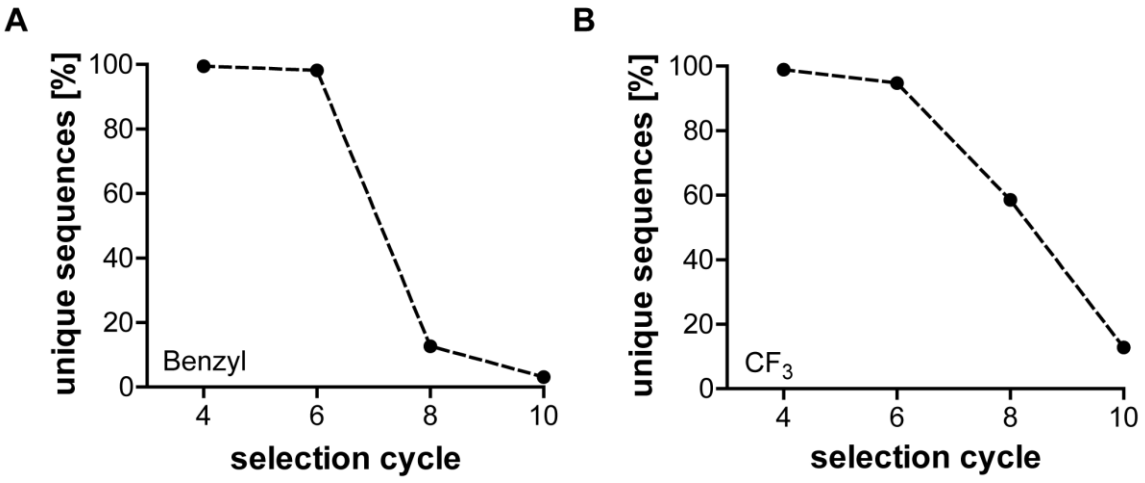


Figure A 6: Frequency of unique sequences in the selection cycles according to NGS
Frequency of unique sequences in the analysed selection cycles of the initial selection with benzyl- (A) and CF_3-modified DNA (B).

Table A 1: Multiple sequence alignment of all sequences identified by Sanger sequencing in the initial click-SELEX with benzyl-modified DNA

Seq.	Sequence of the random region
B27	**-CTACAATC---GTGCG---------ATC--CCCTATCTCC-AGACCTACGGGAACAC**
B37	-CTACAATC---GTGCG---------ATC--CCCTATCTCC-AGACCTACGGGAACAC
B20	-CTACAATC---GTGCG---------ATC--CCCTATCTCC-AGACCTACGGGAACAC
B18	-CTACAATC---GTGCG---------ATC--CCCTATCTCC-AGACCTACGGGAACAC
B26	-CTACAATC---GTGCA---------ATC--CCCTATCTCC-AGACCTACGGGAACAC
B22	-CCACAATT---GTGCG---------ATC--CCCTATCTCC-GGACCTACGGAAACAC
B34	-CCACAATT---GTGCG---------ATC--CCCTATCTCC-GGACCTACGGAAACAC
B29	-CCACAATT---GTGCG---------ATC--CCCTATCTCC-GAACCTACGGAAACAC
B40	-CCACAATC---GTGCG---------ATC--CCCTATCTCC-AGACCTACGGNAACAC
B9	-NTNCAATC---GTGCG---------ATC--CCCTATNTCC-AGACNTACGGAAACAC
B38	-CTACAATC---GTGCG---------ATC--CCCTATCTCC-GGACCTACGGTAACAC
B42	-CTACAATC---GTGCG---------ATC--CCCTATCTCC-GGACCTACGGGAACAC
B17	-CTATGATC---GTGCG---------ATC--CCCTATCTCC-GAACCTACGGGAACAC
B12	**-CAGCGGTCCTAGCGCGGTACTACGCATC--CCCTATCTCC---CCCT----------**
B25	-CAGCGGTCCTAGCGCGGTACTACGCATC--CCCTATCTCC---CCCT----------
B46	-CAGCGGTCCTAGCGCAGTACTACGCATC--CCCTATCTCC---CCCT----------
B41	-CAGCGGTCCTAGCGCGGNNTTACNCANC--CCCTATCTCC---CCCT----------
B36	-CAGCGGTCCTAGTACGGTACTATGTATC--CCCTATCTCC---CCCT----------
B13	**---GTGG----GGCGCA-------TCAGCGACCGTATCCCCCAGACAGACTCAGGC--**
B15	---GTGG----GGCACA-------TCAGCGACCGTATCCCCCAGACAGACTCAGGC--
B3	---GTGG----GGTACA-------TCAGCGACCGTATCCCCCAGACAGACTCAGGC—-
B10	**-ATACTG-----GACCG-------CTACACCCCGCACCACCTC--GCCAATTGGTGA-**
B35	-ATACTG-----GACCG-------TTACACCCCGCACCACCTC--GCCAGTTGGTGA-
B32	-ACACTG-----GATAG-------ATGCGGGGGGCAACGCCTTCAGGCA-TCAGT---
B14	**AACCCTGACTGGGAGTG-------ATGCCTACCGTAACACCTCCCG-CAC--------**
B33	AACCCTGACTGGGAGTG-------ATGCCTACCGTAACACCTCCCG-CAC--------
B6	AACCCTGACTGGGAGTG-------ATGCCTACCGTAACACCTCCCG-CAC--------
B44	AACCCTGACTGGNAGTG-------ATGCCTACCGTAACACCTCCCG-CAC--------
B28	AACCCTGACTGGGAGTG-------ATGCCGACCGTAACACCTCCCG-CAC--------
B7	AACCCTGACTGGGAGTG-------ATGCCAACCGTAACACCTCCCA-CAC--------
B23	AGCCCTGACTGGGAGTA-------CTGCCGACCGTAACACCTCCCA-CAC--------
B24	AGCCCTGACTGGGAGTA-------CTGCCGACCGTAACACCTCCCA-CAC--------
B16	AGCCCTGACTGGGAGTN-------CTGCCAACCGTAACACCTCCCG-CAC--------
B11	**GATGCCG---GACGTTG-------GAACGGGC-GTGTCATCATAGG-TACAGG-----**
B39	GATGCCG---GACGTTG-------GAACGGGC-GTGTCATCATAGG-TACAGG-----
B5	GATGCCG---GACGTTG-------GAACGGGT-GTGTCATCATAGG-TACAGG-----
B19	GATTCCG---GACATTG-------GAACGGGC-GTGTCATCATAGG-TACAGA-----

Seq. = sequence. T = EdU in the tested sequences. N = unidentified nucleotide. Tested sequences are indicated in bold letters. The sequences were analysed using Clusta Omega.[276,277]

Table A 2: Multiple sequence alignment of all sequences identified by Sanger sequencing in the initial click-SELEX with CF_3-modified DNA

Seq.	Sequence of the random region
C47	**----CGATAATACACGTTCGGCCCCTAAAGCCGGTCGG---CCCT----TGCA------**
C9	----CGATAATACACGTTCGGCCCCTAAAGCCGGTCGG---CCCT----TGCA------
C27	----CGATAATACACGTTCGGCCCCTAAAGCCGGTCGG---CCCT----TGCA------
C31	----CGATAATACACGTTCGGTCCCTAAAGCCGGTCGG---CCCT----TGCA------
C12	----CGATAATACACGTTCGGCCCCTAAAGCCGGCCGG---CCCT----TGCA------
C42	----CGATAATACACGTTCGGCCCCTAAAGCCGGCCGG---CCCT----TGCA------
C44	----CGATAATACACGTTCGGCCCCTAAAGCCGGCCGG---CCCT----TGCA------

C40	----CGATAATACACGTTCGGCCCCTAAAACCGGCCGG---CCCT----TGCA------
C45	----CAATAATACACGTTCGGCCCCTAAAACCGGCCGG---CCCT----TGCA------
C4	----CGATAATACACGTTCGGCCCCTAAAACCTGCCGG---CCCT----TGCA------
C41	----CGATAATACACGTTCGGCCCCTAAANCCGGCCGG---CCCT----TGCA------
C29	----CGATAATACACGTTCGGCCCCTAAAGCCGATCGG---CCCT----TACA------
C16	----CGATAATACACGTTCGGCCCCTAAAGCCGACCGG---CCCT----TGCA------
C8	----CGATAATACACGTTCGGTCCCTAAAGCCGACCTA---CCCT----TGCA------
C30	**--CCAGC---CGCGCGGAAGACAA---TACACGTTGGG---CCACTAA--GGAAG----**
C7	--CCGGC---CGCGCGTAAGACAA---TACACGTTGGG---CCACTAA--GGAAG----
C28	--CNNNT---CNNNNGGAAGACAA---TACACGTTGGG---CCACTAA--GGAAG----
C15	---------TGGCGCGTCCAGAGGCT-TATAGGGGGGGAGCCTACG--TTGAAT-----
C25	---------TGGCGCGTCCAGAGGCT-TATAGGGGGGGAGCCTACG--TTGAAT-----
C17	--------TACACACAAGCGGCCGA-GTAGACACCCCTTCCCCCCT---TTAAC-----
C10	----AACATAG--GGAGT-G---GGTGTGACCAT-TAACCGCCACTAGTTAC-------
C46	-----GTCTGG--GATGT-GTAGGGTATTCCCAAATAAC--CCAGCGGCCAA-------
C11	----GACCAGG-AAAGGCCGTCGGGTCTCACCCG--TCCTCCGAG-CTTC---------
C6	-------CAAG-CGAAGC-GTGGGGGTAAACCCA--AATTGAAAGACATCAAC------
C35	GCCACGCCAAGTCGGGGGCGTGAGGCTTACCCCA--ACTTACGC--------------
C21	---GCGCGAACACACCTTTACCCCCTTC-CCCGGCGCA---CGGC----CGGG------
C19	---------ATACACGTTGGGTCACTA--CCCTGGGGA---CCATG---TACAATACCC
C38	-----GT---AGCACGGAATACACGTTTGAACACTAGA---CCAAAGA--GAATA----
C34	---CGGTACACACACATTGGGCAA---TATACGTTGGG---TCGAT-A--GGAT-----
C32	--TGAAC-----TAGGTAGGA------TACACGTTGGGGCCCTAGTCACCGAAC-----
C36	-CTAGATGCATGCTCGAGCGGCCGCC-AGTGTGATGGA---TATNT--GCAGAATTC--
C5	----GTTGCA--CCCGG--GACAACC-TGCCTGGTAAA---TCTC-----GTACCTCAG

Seq. = sequence. T = EdU in the tested sequences. N = unidentified nucleotide. Tested sequences are indicated in bold letters. The sequences were analysed using Clusta Omega.[276,277]

Table A 3: Original sequence and 25 most frequent mutations of T_Pwo

Sequence	Sequence count
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	**982317**
GGTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	4753
TTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	4240
GTTGGAAGCGACGGGACGGTAGGGCTTGGGCCCCAAGGAGTG	3792
GTTGGAAGCGACGGGACGGTAAGGCTTGGCCCCAAGGAGTG	3722
GTTGGAAGCGACGGACGGTAAGGCTTGGGCCCCAAGGAGTG	3515
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCAAGGAGTG	2732
GTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	2226
GTTGGAAGCGACGGGACGTAAGGCTTGGGCCCCAAGGAGTG	1970
GTTGGAAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	1943
GTTGGAAGCGACGGGACGGTAGGCTTGGGCCCCAAGGAGTG	1886
GTTGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	1727
GTTGGAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	1711
GTTGGAAGCGACGGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	1688
GTTGGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	1641
GTTGGAAGCGACGGGACGGTAAGGCTGGGCCCCAAGGAGTG	1590
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCCAAGGAGTG	1412
GTTGGAAGCGACGGGCGGTAAGGCTTGGGCCCCAAGGAGTG	1319
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAAGGAGTG	1301
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGAGTG	1291

GTTGGAAGCGACGGGACGGTAAGCTTGGGCCCCAAGGAGTG	1239
GTTGGAAGCGACGGGACGGTAAGGCTGGGGCCCCAAGGAGTG	1193
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGAAGTG	1137
GTTGGAAGCGACGGGACGGTGAGGCTTGGGCCCCAAGGAGTG	1123
GTTGGAAGCGACGGGAGGTAAGGCTTGGGCCCCAAGGAGTG	1085
GTTGGAAGCGACGGGACGGTAAGGTTGGGCCCCAAGGAGTG	1059

Yellow indicates substitutions, green insertions, and turquoise deletions. The original sequence is indicated in bold.

Table A 4: Original sequence and 25 most frequent mutations of T_Taq

Sequence	Sequence count
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	**2990041**
GGTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	13977
GTTGGAAGCGACGGGACGGTAGGGCTTGGGCCCCAAGGAGTG	12302
TTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	11493
GTTGGAAGCGACGGGACGGTAAGGCTTGGCCCCAAGGAGTG	11343
GTTGGAAGCGACGGACGGTAAGGCTTGGGCCCCAAGGAGTG	10431
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCAAGGAGTG	7847
GTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	6353
GTTGGAAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	6098
GTTGGAAGCGACGGGACGTAAGGCTTGGGCCCCAAGGAGTG	5541
GTTGGAAGCGACGGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	5406
GTTGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	5179
GTTGGAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	5091
GTTGGAAGCGACGGGACGGTAGGCTTGGGCCCCAAGGAGTG	5059
GTTGGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	4909
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCCAAGGAGTG	4721
GTTGGAAGCGACGGGACGGTAAGGCTGGGCCCCAAGGAGTG	4541
GTTGGAAGCGACGGGACGGTGAGGCTTGGGCCCCAAGGAGTG	4053
GCTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	3972
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAAGGAGTG	3948
GTTGGAAGCGACGGGACGGTAAGGCTGGGGCCCCAAGGAGTG	3914
GTTGGAGGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	3772
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGAGTG	3694
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	3635
GTTGGAAGCGACGGGACGGTAAGCTTGGGCCCCAAGGAGTG	3560
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGAAGTG	3536

Yellow indicates substitutions, green insertions, and turquoise deletions. The original sequence is indicated in bold.

Table A 5: Original sequence and 25 most frequent mutations of T_w/o

Sequence	Sequence count
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	**1640067**
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCAAGGAGTG	7565
GTTGGAAGCGACGGGACGGTAAGGCTTGGCCCCAAGGAGTG	5878

GGTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	5444
TTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	4726
GTTGGAAGCGACGGACGGTAAGGCTTGGGCCCCAAGGAGTG	4683
GTTGGAAGCGACGGGACGGTAGGCTTGGGCCCCAAGGAGTG	4033
GTTGGAAGCGACGGGACGGTAAGCTTGGGCCCCAAGGAGTG	3930
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTGC	3777
GTTGGAAGCGACGGGACGGTAAGGCTGGGCCCCAAGGAGTG	3700
GTTGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	3346
GTTGGAAGCGACGGGACGGAAGGCTTGGGCCCCAAGGAGTG	3342
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAGGAGTG	3119
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGT	2986
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGATG	2715
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCACAAGGAGTG	2713
GTTGGAAGCGACGGGACGTAAGGCTTGGGCCCCAAGGAGTG	2659
GTTGGAAGCGACGGGACGGTAAGGTTGGGCCCCAAGGAGTG	2635
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGAGTG	2609
GTTGGAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	2500
TTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	2437
GTTGGAAGCGACGGGCGGTAAGGCTTGGGCCCCAAGGAGTG	1991
GTTGGAAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	1989
GTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	1842
GTTGGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	1839
GCTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	1740

Yellow indicates substitutions, green insertions, and turquoise deletions. The original sequence is indicated in bold.

Table A 6: Original sequence and 25 most frequent mutations of EdU _Pwo

Sequence	Sequence count
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	**3122777**
GTTGGAAGCGACGGGACGGTAAGGCCTGGGCCCCAAGGAGTG	51393
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGATG	42097
GTTGGAAGCGACGGGACGTAAGGCTTGGGCCCCAAGGAGTG	35342
TTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	29569
GTCGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	27323
GCTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	25350
GTTGGAAGCGACGGGACGGTAAGGCATGGGCCCCAAGGAGTG	24557
GTAGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	24291
GTTGGAAGCGACGGACGGTAAGGCTTGGGCCCCAAGGAGTG	23940
GGTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	21405
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGG	20464
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCAAGGAGTG	18605
GATGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	18243
GTTGGAAGCGCCGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	18183
GTTGGAAGCGACGGGCCGGTAAGGCTTGGGCCCCAAGGAGTG	17191
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGAGTG	16764
GTTGGAAGCGACGGGACGGTAAGGCGTGGGCCCCAAGGAGTG	15361

GTTGGAAGCGACGGGACGGTAAGGCTAGGGCCCCAAGGAGTG	15273
GTTGGAAGCGACGGGACGGTAGGCTTGGGCCCCAAGGAGTG	15044
GTTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGAG	14747
GTTGGAAGCGACGGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	14639
GTGGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	14417
GTTGGAAGCGACGGGACGGTAAGGCTTGGCCCCAAGGAGTG	13760
GTTGAAGCGACGGGACGGTAAGGCTTGGGCCCCAAGGAGTG	13186
GTTGGAAGCGACGGGACGGTAAGGCTTTGGCCCCAAGGAGTG	13117

Yellow indicates substitutions, green insertions, and turquoise deletions. The original sequence is indicated in bold.

Table A 7: Original sequence and 25 most frequent mutations of FT2_GATC

Sequence	Sequence count
GATCGATCGATCGATCGATCGATCGATC	**9210580**
GATCGATCGATCGTTCGATCGATCGATCGATC	40756
GATCGATCGATCGATCGATCGATCGATC	16625
ATCGATCGATCGATCGATCGATCGATC	15189
GATCGATCGATCGATCGATCGATCGAT	12245
GATCGATCGATCGATTGATCGATCGATCGATC	11192
GATCGATCGATCGATCGATCGATAGATCGATC	10571
GATCGATCGATCGATCGTTCGATCGATCGATC	10186
GATCGATTGATCGATCGATCGATCGATCGATC	10022
GATCGATCGATCGCTCGATCGATCGATCGATC	9607
GATCGATCGATCTATCGATCGATCGATCGATC	9462
GATCGATCGATCGATCGATCGATCAATCGATC	9387
GATCGATCGATCGATCGTCGATCGATCGATC	8657
GATCGATCGATCGATCGATCGATCGTTC	8628
GATTGATCGATCGATCGATCGATCGATCGATC	8511
GATCGATCGATCGATCGACGATCGATCGATC	8406
GATCGATCGATCGATCGATCGAACGATCGATC	8334
GATCGATCGATCGATCGATTGATCGATCGATC	8294
GATCGATCGATCGATCGATCGACGATCGATC	8159
GTTCGATCGATCGATCGATCGATCGATCGATC	8081
GATCGATCGATCGATCTATCGATCGATCGATC	7884
GATCGATCGATCGATCAATCGATCGATCGATC	7740
GATCGATCTATCGATCGATCGATCGATCGATC	7715
GATCGATCGATCGATCGATAGATCGATCGATC	7604
GATCGATCGATCGATCGAACGATCGATCGATC	7408
GATCGATCGATTGATCGATCGATCGATCGATC	7398

Yellow indicates substitutions, green insertions, and turquoise deletions. The original sequence is indicated in bold.

Table A 8: Original sequence and 25 most frequent mutations of FT2_GATC_II

Sequence	Sequence count
GATCGATCGATCGATCGATCGATCGATC	**2178012**

GATCGATCGATCGTTCGATCGATCGATCGATC	6052
GATCGATCGATCGATCGATCGATCGATC	4924
ATCGATCGATCGATCGATCGATCGATCGATC	3450
GATCGATCGATCGATCGATCGATCGATCGAT	2870
GATCGATCGATCGATCTATCGATCGATCGATC	2324
GATCGATCGATCGATCGTTCGATCGATCGATC	2019
GATCGATCGATCGATCGTCGATCGATCGATC	2003
GATCTATCGATCGATCGATCGATCGATCGATC	1992
GATCGATCGATCGATTGATCGATCGATCGATC	1972
GATCGATCGATCGATCGACGATCGATCGATC	1948
GATCGATCGATCTATCGATCGATCGATCGATC	1942
GATCGATTGATCGATCGATCGATCGATCGATC	1921
GATCGATCGATCGATCGATCGATCGTTC	1905
GATCGATCGATCGATCGATCGATAGATCGATC	1851
GATCGATCGATCGATCGATCGATCAATCGATC	1740
GATCGATCGATCGATCGATCGACGATCGATC	1721
GTTCGATCGATCGATCGATCGATCGATCGATC	1715
TATCGATCGATCGATCGATCGATCGATCGATC	1676
GATCGATCGATCGATCGATTGATCGATCGATC	1640
GATCGATCGATCGATCGATCGTCGATCGATC	1596
GATCGATCGATCGATGATCGATCGATCGATC	1594
GATCGATCGATCATCGATCGATCGATCGATC	1588
GATCGATCGATCGTCGATCGATCGATCGATC	1583
GATCGATCGATCGACGATCGATCGATCGATC	1577
GATTGATCGATCGATCGATCGATCGATCGATC	1565

Yellow indicates substitutions, green insertions, and turquoise deletions. The original sequence is indicated in bold.

Table A 9: Original sequence and 25 most frequent mutations of FT2_G4A4T4C4

Sequence	Sequence count
GGGGAAAATTTTCCCCGGGGAAAATTTTCCCC	**7340975**
GGGGAAAATTTTCCCCGGGGAAAGTTTTCCCC	62115
GGGGAAAATTTTCCCCGGGGAAATTTTTCCCC	29603
GGGGAAAATTTTCCCCGGGGAAACTTTTCCCC	28885
GGGAAAATTTTCCCCGGGGAAAATTTTCCCC	27528
GGGGAAAATTTTCCCCGGGGAAAATTTTCCC	26734
GGGGAAAATTTTCCCGGGGAAAATTTTCCCC	25567
GGGGAAAATTTTCCCCGGGGAATATTTTCCCC	25363
GGGGAAAATTTTCCCCGGGAAAATTTTCCCC	22129
GGGGAAAATTTTCCCCGGGGAAGATTTTCCCC	20528
GGGGAAAATTTCCCCGGGGAAAATTTTCCCC	18488
GGGGAAAATTTTCCCCGGGGAAATTTTCCCC	14789
GGGGAAATTTTCCCCGGGGAAAATTTTCCCC	13789
GGGGAAAATTTTCCCCGGGGAAAATTTCCCC	13618
GGGGAAAATTTTCCCGGGGGAAAATTTTCCCC	12188
GGGGAAAATTTTCCACGGGGAAAATTTTCCCC	11202

GGGGAAAATTTTCCTCGGGGAAAATTTTCCCC	10651
GGGGAAAATTTTCCCCGGGGAACATTTTCCCC	9828
GGGGAAAATTTTCCCCGGGGTAAATTTTCCCC	9817
GGGGAAAATTTTCACCGGGGAAAATTTTCCCC	9097
GGGGAAAATTTTCTCCGGGGAAAATTTTCCCC	8683
GGGGAAAATTTTCCCCGAGGAAAATTTTCCCC	8659
GGGGATAATTTTCCCCGGGGAAAATTTTCCCC	8512
TGGGAAAATTTTCCCCGGGGAAAATTTTCCCC	8250
GGGGAAAATTTTCCCCGTGGAAAATTTTCCCC	8185
GGGGAAAATTTTCCCCGGGGATAATTTTCCCC	8177

Yellow indicates substitutions, green insertions, and turquoise deletions. The original sequence is indicated in bold.

Table A 10: Original sequence and 25 most frequent mutations of FT2_G4A4T4C4_II

Sequence	Sequence count
GGGGAAAATTTTCCCCGGGGAAAATTTTCCCC	**6549166**
GGGGAAAATTTTCCCCGGGGAAAGTTTTCCCC	45606
GGGGAAAATTTTCCCCGGGGAAAATTTTCCC	25874
GGGAAAATTTTCCCCGGGGAAAATTTTCCCC	25002
GGGGAAAATTTTCCCCGGGGAAATTTTTCCCC	24447
GGGGAAAATTTTCCCGGGGAAAATTTTCCCC	23700
GGGGAAAATTTTCCCCGGGAAAATTTTCCCC	20951
GGGGAAAATTTTCCCCGGGGAAACTTTTCCCC	20084
GGGTAAAATTTTCCCCGGGGAAAATTTTCCCC	19289
GGGGAAAATTTTCCCCGGGGAATATTTTCCCC	18918
GGGGAAAATTTCCCCGGGGAAAATTTTCCCC	16348
GGGGAAAATTTTCCCCGGGGAAAATTTTACCC	15108
GGGGAAAATTTTCCCCGGGGAAGATTTTCCCC	14713
GGGGAAAATTTTCCCCGGGGAAATTTTCCCC	13451
GGGGAAAATTTTCCCCGGGGAAAATTTCCCC	13426
GGGGAAATTTTCCCCGGGGAAAATTTTCCCC	12803
GGGGAAAATTTTCCCGGGGGAAAATTTTCCCC	11949
GGGGAAAATTTTACCCGGGGAAAATTTTCCCC	11730
TGGGAAAATTTTCCCCGGGGAAAATTTTCCCC	8438
GGTGAAAATTTTCCCCGGGGAAAATTTTCCCC	8037
GGGGAAAATTTTCCCCGGGTAAAATTTTCCCC	7069
GGGGATAATTTTCCCCGGGGAAAATTTTCCCC	6921
GGGGAAAATTTTCACCGGGGAAAATTTTCCCC	6831
GGGGAAAATTTTCCCCGGGGAAAATTTTCACC	6576
GGGGAATATTTTCCCCGGGGAAAATTTTCCCC	6558
GGGGAAAATTTTCCCCGGGGAAAATTTTCCAC	6555

Yellow indicates substitutions, green insertions, and turquoise deletions. The original sequence is indicated in bold.

Table A 11: Original sequence and 25 most frequent mutations of FT2_G2A2T2C2

Sequence	Sequence count
GGAATTCCGGAATTCCGGAATTCCGGAATTCC	**2040896**
GAATTCCGGAATTCCGGAATTCCGGAATTCC	6073
GGAATTCCGGAATTCCGTAATTCCGGAATTCC	5990
GGAATTCCGGAATTACGGAATTCCGGAATTCC	5539
GGAATTCCGGAATTCCGGAATTCCGGAATTC	5123
GGAATTCCGGAATTCCGGAATTCCGTAATTCC	5085
GGAATTTCGGAATTCCGGAATTCCGGAATTCC	5080
GGAATTCCGGAATTCCGGAATTCCGGAATTAC	4697
GGAATTCCGTAATTCCGGAATTCCGGAATTCC	4674
GGAATTCCGGAATTCCGGAATTACGGAATTCC	4548
GGAATTACGGAATTCCGGAATTCCGGAATTCC	4547
GTAATTCCGGAATTCCGGAATTCCGGAATTCC	4165
GGAATTCCGGAATTCGGAATTCCGGAATTCC	4057
GGAATTCCGGAATTCCGAATTCCGGAATTCC	3949
GGAATTCCGGAATTCAGGAATTCCGGAATTCC	3664
GGAATTCCGGAATTCCGGAATTCAGGAATTCC	3571
GGAATTCCGGAATTCCGGAATTCCGGAATTCA	3541
TGAATTCCGGAATTCCGGAATTCCGGAATTCC	3522
GGAATTCCGGAATTCCTGAATTCCGGAATTCC	3473
GGAATTCCTGAATTCCGGAATTCCGGAATTCC	3450
GGAATTCCGAATTCCGGAATTCCGGAATTCC	3374
GGAATTCCGGAATCCGGAATTCCGGAATTCC	3215
GGAATTCTGGAATTCCGGAATTCCGGAATTCC	3196
GGAATTCCGGAATTCCGGAATCCGGAATTCC	3180
GGAATTCCGGAATTCTGGAATTCCGGAATTCC	3161
GGAATTCCGGAATTCCGGATTCCGGAATTCC	3152

Yellow indicates substitutions, green insertions, and turquoise deletions. The original sequence is indicated in bold.

Table A 12: Original sequence and 25 most frequent mutations of FT2_G3A3T3C3

Sequence	Sequence count
GGGAAATTTCCCGGGAAATTTCCCGGGAAATT	**5535182**
GGAAATTTCCCGGGAAATTTCCCGGGAAATT	21093
GGGAAATTTCCCGGGAAATTCCCGGGAAATT	19864
GGGAAATTTCCGGGAAATTTCCCGGGAAATT	19079
GGGAAATTTCCCGGGAATTTCCCGGGAAATT	18549
GGGAAATTTCCCGGGAAATTTACCGGGAAATT	17688
GGGAAATTTCCCGGAAATTTCCCGGGAAATT	16754
GGGAAATTTCCAGGGAAATTTCCCGGGAAATT	16257
GGGAAATTTCCCGGGAAATTTCCAGGGAAATT	16206
GGGAAATTTCCCTGGAAATTTCCCGGGAAATT	15353
GGGAAATTTCCCGGGAAATTTCCGGGAAATT	14511
GGTAAATTTCCCGGGAAATTTCCCGGGAAATT	13353
GGGAAATTTCCCGGGAAATTTCCCGGAATT	12584
TGGAAATTTCCCGGGAAATTTCCCGGGAAATT	12115

GGGAAATTTCCCGGTAAATTTCCCGGGAAATT	12041
GGGAAATTTCCCGGGAAATTTCCCTGGAAATT	11707
GGGAAATTCCCGGGAAATTTCCCGGGAAATT	11475
GGGAAATTTCCCGTGAAATTTCCCGGGAAATT	11190
GGGAAATTTCCCGGGAAATTTCCCGGGAATT	10863
GGGAAATTTCACGGGAAATTTCCCGGGAAATT	10839
GGGAAATTTCCCGGGAAATTTCACGGGAAATT	10796
GGGAAATTTCCCGGGAAATTTCCCGTGAAATT	10637
GGGAATTTCCCGGGAAATTTCCCGGGAAATT	10332
GTGAAATTTCCCGGGAAATTTCCCGGGAAATT	10015
GGGAAATTTACCGGGAAATTTCCCGGGAAATT	9290
GGGAAATTTCCCGGGAAATTTCCCGGTAAATT	8848

Yellow indicates substitutions, green insertions, and turquoise deletions. The original sequence is indicated in bold.

Table A 13: Original sequence and 25 most frequent mutations of FT2_TGCA

Sequence	Sequence count
TGCATGCATGCATGCATGCATGCATGCATGCA	**6627344**
TGCATGCATGCATGCATGCTTGCATGCATGCA	45693
TGCATGCATGCTTGCATGCATGCATGCATGCA	36139
TGCAAGCATGCATGCATGCATGCATGCATGCA	26613
TGCATGCATGCATGCATGCATGCATGCA	24728
TGCATGCATGCATGCATGCATGCATGCATGAA	19114
TGCATGCATGAATGCATGCATGCATGCATGCA	18958
TGTATGCATGCATGCATGCATGCATGCATGCA	12451
TGCATGCATTCATGCATGCATGCATGCATGCA	11260
TGCATGTATGCATGCATGCATGCATGCATGCA	10812
TGCATGCATGCATGTATGCATGCATGCATGCA	10365
TGCATGCATGTATGCATGCATGCATGCATGCA	10306
TGCATGCATGCATGAATGCATGCATGCATGCA	9163
TGCATGCATGCATGCATGTATGCATGCATGCA	8532
TGCATGCATGCATGCATGCATGCATGTATGCA	7920
TGCATGCATGCATGCATGCATGTATGCATGCA	7487
TGCATGCATGCTGCATGCATGCATGCATGCA	7444
TGCATGCATGCATGCATGCATGCATGCATACA	7208
TGCATGCATCATGCATGCATGCATGCATGCA	7143
TGCATGCATGCATGCATGAATGCATGCATGCA	7044
TGCATGCATGCATGCATGCATGCATGAATGCA	6912
TGCATGAATGCATGCATGCATGCATGCATGCA	6769
TGAATGCATGCATGCATGCATGCATGCATGCA	6711
TGCATGCATGCATGCATGCATGATGCATGCA	6706
TGCATGCATGCATCATGCATGCATGCATGCA	6549
TGCATGCATGCATGCATGCATGCATACATGCA	6531

Yellow indicates substitutions, green insertions, and turquoise deletions. The original sequence is indicated in bold.

Table A 14: Original sequence and 25 most frequent mutations of D3_TGCA

Sequence	Sequence count
TGCATGCATGCATGCATGCATGCATGCA	**398611**
TGCATGCATGCATGCATGCTTGCATGCATGCA	1654
TGCATGCATGCATGCTTGCATGCATGCATGCA	1560
TGCATGTATGCATGCATGCATGCATGCATGCA	1190
TGCATGCATGCATGCATGCATGCATGCA	1026
TGCATGCATGCATGCATGCATGTATGCATGCA	588
TGCATGCTTGCATGCATGCATGCATGCATGCA	555
TGCATGCATGCATGCATGTATGCATGCATGCA	545
TGCATGCATGCATGCATGCATGCATGCATACA	530
TGTATGCATGCATGCATGCATGCATGCATGCA	518
TGCATGCATGCATGCATGCATGCATACATGCA	498
TGCATGCATGTATGCATGCATGCATGCATGCA	459
TGCATGCATTCATGCATGCATGCATGCATGCA	455
TGAATGCATGCATGCATGCATGCATGCATGCA	418
TGCATGCATGCATGCATACATGCATGCATGCA	395
TGCATGCATGCATACATGCATGCATGCATGCA	393
TGCATGCATGCATGTATGCATGCATGCATGCA	392
TGCATGCATGCATGCATGCATACATGCATGCA	390
TGCATGAATGCATGCATGCATGCATGCATGCA	379
TGCATGCTGCATGCATGCATGCATGCATGCA	368
TGCATGCATGCATGAATGCATGCATGCATGCA	364
TGCATGCATGCATGCATTCATGCATGCATGCA	345
TGCATGCATGCATGCATGCATGCATGAA	344
TGCATACATGCATGCATGCATGCATGCATGCA	335
TGCATGCATGCATGCATGAATGCATGCATGCA	332
TGCATGCATGCATGCATGCATGAATGCA	306

Yellow indicates substitutions, green insertions, and turquoise deletions. The original sequence is indicated in bold.

Table A 15: Original sequence and 25 most frequent mutations of FT2_T4G4C4A4

Sequence	Sequence count
TTTTGGGGCCCCAAAATTTTGGGGCCCCAAAA	**1745068**
TTTTGGGGCCCAAAAATTTTGGGGCCCCAAAA	9556
TTTTGGGGCCCCAAAATTTGGGGCCCCAAAA	7973
TTTTGGGGCCCCAAATTTTGGGGCCCCAAAA	7625
TTTTGGGGCCCCAAAATTTTGGGCCCCAAAA	7388
TTTTGGGGCCCAAAATTTTGGGGCCCCAAAA	6884
TTTTGGGCCCCAAAATTTTGGGGCCCCAAAA	4855
TTTTGGGGCCCCAAAATTTTGGGGCCCAAAA	4841
TTTTGGGGCTCCAAAATTTTGGGGCCCCAAAA	4576
TTTTGGGGCCTCAAAATTTTGGGGCCCCAAAA	4488
TTTTGGGGCCCCAAAATTTTGGGGCCCCAAA	4334
TTTGGGGCCCCAAAATTTTGGGGCCCCAAAA	4234
TTTTGGGGCCCCAAAATTTTTGGGCCCCAAAA	4229
TTTTGGGGCCCTAAAATTTTGGGGCCCCAAAA	4204

TTTTGGGGCCCCAAAATTTTGGGGCTCCAAAA	3987
TTTTGGGGTCCCAAAATTTTGGGGCCCCAAAA	3946
TTTTGGGGCCCCAAAATTTTAGGGCCCCAAAA	3878
TTTTGGGGCCCCAAAATTTTGGAGCCCCAAAA	3860
TTTTGGGGCCCCAAAATTTTGAGGCCCCAAAA	3606
TTTTGGGGCCCCAAAATTTTGGGACCCCAAAA	3308
TTTTGGGGCCCCAAAATTTTGTGGCCCCAAAA	2968
TTTTGGGGCCACAAAATTTTGGGGCCCCAAAA	2940
TTTTGTGGCCCCAAAATTTTGGGGCCCCAAAA	2649
TTTTGGGGCCCCAAATTTTTGGGGCCCCAAAA	2524
TTTTGGGGCCCCAAAAATTTTGGGGCCCCAAAA	2469
TTTTGGGGCCCCAAAATTTTGGGGCCACAAAA	2430

Yellow indicates substitutions, green insertions, and turquoise deletions. The original sequence is indicated in bold.

Table A 16: Original sequence and 25 most frequent mutations of D3_T4G4C4A4

Sequence	Sequence count
TTTTGGGGCCCCAAAATTTTGGGGCCCCAAAA	**5284258**
TTTTGGGGCCCCAAAAATTTGGGGCCCCAAAA	56226
TTTTGGGGCCCCAAATTTTGGGGCCCCAAAA	33501
TTTTGGGGCCCCAAAATTTGGGGCCCCAAAA	31991
TTTTGGGGCCCAAAAATTTTGGGGCCCCAAAA	19746
TTTTGGGGCCCCAAAATTTTGGGCCCCAAAA	18308
TTTTGGGGCCCCAAAATTTTGGGGCCCAAAA	18108
TTTTGGGGCCCAAAATTTTGGGGCCCCAAAA	16172
TTTTGGGGCCCCAAAATTTTTGGGCCCCAAAA	16010
TTTTGGGCCCCAAAATTTTGGGGCCCCAAAA	15538
TTTTGGGGCCCTAAAATTTTGGGGCCCCAAAA	14219
TTTTGGGGCTCCAAAATTTTGGGGCCCCAAAA	14016
TTTTGGGGCCTCAAAATTTTGGGGCCCCAAAA	13073
TTTTGGGGCCCCAAAATTTTGGGGCCCCAAAA	12625
TTTTGGGGTCCCAAAATTTTGGGGCCCCAAAA	10980
TTTGGGGCCCCAAAATTTTGGGGCCCCAAAA	10697
TTTTGGGGCCACAAAATTTTGGGGCCCCAAAA	9072
TTTTGGGGCCCCAAAATTTTGTGGCCCCAAAA	8441
TTTTGTGGCCCCAAAATTTTGGGGCCCCAAAA	8378
TTTTGGGGCCCCAAAATTTTAGGGCCCCAAAA	7925
TTTTGGGGCCCCAAAATTTTGGGGCCACAAAA	7891
TTTTTGGGCCCCAAAATTTTGGGGCCCCAAAA	7708
TTTTGGGGCCCCAAAATTTTGGGACCCCAAAA	7669
TTTTGGTGCCCCAAAATTTTGGGGCCCCAAAA	7171
TTTTGGGGCCCCAAAATTTAGGGGCCCCAAAA	7018
TTTTGGGGCCCCAAAATTTTGAGGCCCCAAAA	6677

Yellow indicates substitutions, green insertions, and turquoise deletions. The original sequence is indicated in bold.

Table A 17: Multiple sequence alignment of all sequences identified by Sanger sequencing in the re-selection with benzyl-modified DNA

Seq.	Sequence of the random region
27	**----ACGTCCTCATCACCTCCCATGTCTCCGGAGTCCACGGACAGG-**
B27	**----CTACAATCGTGCGATCCCCTATCTCCAGACCTACGGGAACAC-**
43	----CTACAGTCGTGCGATCCCCTATCTCCAGACCTACGGGAACAC-
35	----CTACAATCGTGCGATCCCCTATCTCCGAACCTACGGGAACAC-
29	----CTACAATCGCGCGATCCCCTATCTCCAGACCTACGGGAACAC-
33	----CTACAATCGCGCGATCCCCTATCTCCAGACCTACGGGAACAC-
22	----CTACAATCGTGCAATCCCCTATCTCCAGACCTACGGGAACAC-
36	----CTACAATCGTGCAATCCCCTATCTCCAGACCTACGGGAACAC-
17	----CTACAATCGTGCGATCCCCTATCTCCAAACCTACGGGAACAC-
32	----CTACAATCGTGCGATCCCCTATCTCCAAACCTACGGGAACAC-
14	----CTACAATCGTGCGATCCCCTATCTCCAGACCTACGGAAACAC-
9	----CTACAATCGTGCGATCCCCTATCTCCAGACCTACGGAAACAC-
18	----CTACAATCGTGCGATCCCCTATCTCCAGACCTACGGGAACAC-
26	----CTACAATCGTGCGATCCCCTATCTCCAGACCTACGGGAACAC-
39	----CTACAATCGTGCGATCCCCTATCTCCAGACCTACGGGAACAC-
4	----CTACAATCGTGCGATCCCCTATCTCCAGACCTACGGGAACAC-
6	----CTACAATCGTGCGATCCCCTATCTCCAGACCTACGGGAACAC-
10	----CCACAATCGTGCGATCCCCTATCTCCAGACTTACGGAAACAC-
23	----CCACAATCGTGCGATCCCCTATCTCCAGACCTACGGAAACAC-
45	----CCACAATCGTGCGATCCCCTATCTCCAGACCTACGGAAACAC-
31	----CCACAATTGTGCGATCCCCTATCTCCAAACCTACGGAAACAC-
7	----CCACAATCGTGCGATCCCCTATCTCCAAACCTACGGAAACAC-
B11	**----GATGCCGGACG--TTGGAACGGGCGTGTCATCATAGGTACAGG**
37	----GATACCAAACG--TTGGAACGGGTGTGTCAACATAGGTACAGG
42	----GATGCCGGACG--TTGGAACGGGTGTGTCATCATAGGTACAGG
12	----GATGCCGGACG--TTGGAACGGGTGTGTCAACATAGGTACAGG
24	----GATGCCGGACG--TTGGAACGGGCGTGTCAACATAGGTACAGG
44	----GATGCCGAACG--TTGGAACGGGTGTGTCATCATAGGTACAGG
21	----GATGCCGGACG--TTGGAACGGGCGTGTCATCATAGGTATAGG
15	----GATGCCGGACG--TTGGAACGGGCGTGTCATCATAGGTACAGG
40	----GATGCCGGACG--TTGGAACGGGCGTGTCATCATAGGTACAGG
41	----GATGCCGGACG--TTGGAACGGGCGTGTCATCATAGGTACAGG
8	----GATGCCGGACG--TTGGAACGGGCGTGTCATCATAGGTACAGG
16	----GATGCCGGACG--TTGGAACGGGCGTGTCATCATAGGTACAGA
2	----GATGCCGAACG—-TTGGAACGGGCGTGTCATCATAGGTACAGG
B32	-ACACTGGATAGA-----TGCGGGGGGCAACGCCTTCAGGCATCAGT
3	**-ACACTGGATAGA-----TGCGGGGGGCAACGCCTTCAGGCATCAGT**
5	-ACACTGGATAGA-----TGCGGGGGGCAACGCCTTCAGGCATCAGT
30	-ACACTGGATAGA-----TGCGGGAGGCAACGCCTTCAGGCATCAGT
B14	**AACCCTGACTGGGAGTGATGCCTACCGTAACACCTCCCGCAC-----**
28	AACCCTGACTGGAAGTGATGCCAACCGTAACACCTCCCACAC-----
47	AACCCTGACTGGGAGTGATGCCAACCGTAACACCTCCCGCAC-----
38	AATCCTGACTGGGAGTGATGCCTACCGTAACACCTCCCGCAC-----
19	AACCCTGACTGGGAGTGATGTCTACCGTAACACCTCCCGCAC-----
46	AACCCTGACTGGGAGTGATGCCTACCGTAACACCTCCCGCAC-----

Seq. = sequence. T = EdU in the tested sequences. N = unidentified nucleotide. Tested sequences are indicated in bold letters. Identical sequences from the initial click-selection are

given for comparison and separated by an empty line. The sequences were analysed using Clusta Omega.[276,277]

Table A 18: Multiple sequence alignment of all sequences identified by Sanger sequencing in the re-selection with CF_3-modified DNA

Seq.	Sequence of the random region
73	---------GCACGGGACAACCGCAAACGTACTAGCATC----------CCGCGGCAAACA-
69	---------GCACGGGACAACCGCGAACGTACTAGCATC----------CCGCGGCGAATA-
74	**---------GCACGGAACAACCGCGAACGTACTAGCATC----------CCGCGGCGAACA-**
75	---------GCACGGGACAACCGTGAACGTACTAGCATC----------CCGCGGCGAACA-
61	--------------------GTTGCACCCGGGACAACCTGCC—TGGTAAATCCCGTACCTCAG
80	------------------AACTGCAGAAAGGGGAAGGCGAG--GAATAGTCCCCGGTTGGCA-
67	**------------------AACTGCAGAAAGGGGAAGGC--GAGGAATAGTCCCCGGTTGGCA-**
77	**-------------GGTAATGGTGGAGGGTAAGGAGGACCCC--GAATTGTCCCCGTG------**
48	-------------GGTAACGGTGGAGGGTAATGAGGGCCCC--GAATTGTCCCCGTG------
60	-------------AGTCACGGTGGAGGGTAAGGAGGGCCCC--GAATTGTCCCCGTG------
87	-AATGCATACGGAGGGCTGGCTACCCAGGCTGAGTCCCC----ATCA----------------
88	CGGTGATTAGGATGGGCT--GGAAATT----------TG----AGTCCCACCAGCGAG-----
83	--------------GAAC--CGACAGTGGACACATAACC----CTAACTTCCAGTGTAGGCA-
53	**------TTCGGACGGAGA--GAGCAGGGGGTGAGTAAAG----AGTAGGTCCAG--------**
54	------TTCGGACGGAGA--GAGCAGGGGGTGAGTAAAG----AGTAGGTCCAG---------
57	------TTCGGACGGAGA--GAGCAGGGGGTGAGTAAAG----AGTAGGTCCAG---------
58	------TTCGGACGGAGA--GAGCAGGGGGTGAGTAAAG----AGTAGGTCCAG---------
64	------TTCGGACGGAGA--GAGCAGGGGGTGAGTAAAG----AGTAGGTCCAG---------
86	------TTCGGACGGAGA--GAGCAGGGGGTGAGTAAAG----AGTAGGTCCAG---------
55	---GACTTTCGGAGGAGG--GGACCGGGGAT-----AAC----AGTAGATCCGCAGA------
63	**---------GCTCACGCCACACAAGGCAGATCCTGAGCTCCGGGGGTC---CAC--------**
76	---------GCTCACGCCGCACAAGGCAGATCCTGAGTTCCGGGGGTC---CGC---------
85	---------GCTCACGCCGCACAAGGCAGATCCTGAGCTCCGGGGGTC---TGC---------
52	----GACCAATACGTGCTGAGTTCCAGACACAAAT----------------GGGCTGATGA—
49	**------TGCGCGAGGGATGATTAAGAGAAGCTCAACCGCGCGGAGGCC--------------**
59	------TGCGCGAGGGATGATTAAGAGAAGCTCAACCGCGCGGAGGCC---------------
68	------TGCGCGAGGGATGATTAAGAGAAGCTCAACCGCGCGGAGGCC---------------
92	------TGCGCGAGGGATGATTAAGAGAAGTTCAACCGCGCGGAGACC---------------
62	----------CGCGGGCCAAGTACCGGGACCGTCTGCTAACTGAGGTCCAAC-----------
89	-------CATAANGGGCTGGTCAGACTACACGNNAGGACACTAGGAAAA--------------
81	--------------GATCACACACGATACACGTTCGGACACTAGGTCACTACACCA-------
C30	**-------CCAGCCGCGCGGAAGACAATACACGTTGGGCCACTAAGGAAG-------------**
66	-------CCGGCCGTGCGGAAGACAATACACGTTGGGTCACTAAGGAAG--------------
79	-------CCGGCCGCGCGGAAGACAATACACGTTGGGCCACTAAGGAAG--------------
C47	**--------------------CGATAATACACGTTCGGCCCCTAAAGCCGGTCGGCCCTTGCA-**
78	--------------------CGATAATACACGTTCGGCCACTAAAGCCGTCCATCCCTTGCA-
71	--------------------CAATAATACACGTTCGGTCACTAAAGCCGGCCGGCCCTTGCA-
94	--------------------CGATAATACACGTTCGGCCACTAAAGCCGTCCGGCCCTTACA-
56	--------------------CGATAATACACGTTCGGCCATTAAAGCCGGCCGGCCCTNGCA-
91	--------------------CGATAATACACGTTCGGCCACTAAAGCCGGCCGACCCTTGCA-
50	--------------------CGATAATACACGTTCGGCCACTAAAGCCGGCCGGCCCTTGCA-
90	--------------------CGATAATACACGTTCGGCCACTAAAGCCGGCCGGCCCTTGCA-
81	--------------------CGATAATACACGTTCGGCCACTAAAGCCGACCGGCCCTTGCA-

Seq. = sequence. T = EdU in the tested sequences. N = unidentified nucleotide. Tested sequences are indicated in bold letters. Identical sequences from the initial click-selection are given for comparison and separated by an empty line. The sequences were analysed using Clusta Omega.[276,277]

8 References

1 Bundesministerium für Gesundheit. *Drogen- und Suchtbericht 2017*, <https://www.drogenbeauftragte.de/presse/pressekontakt-und-mitteilungen/2017/2017-3-quartal/drogen-und-suchtbericht-der-bundesregierung-2017/> (2017).

2 Palamar, J. J., Lee, L. & Weitzman, M. Prevalence and Correlates of Hashish Use in a National Sample of High School Seniors in the United States. *The American journal of drug and alcohol abuse* **41**, 197-205, doi:10.3109/00952990.2015.1011745 (2015).

3 ElSohly, M. A. *et al.* Changes in Cannabis Potency over the Last Two Decades (1995-2014) - Analysis of Current Data in the United States. *Biol. Psychiatry* **79**, 613-619, doi:10.1016/j.biopsych.2016.01.004 (2016).

4 Cressey, D. The cannabis experiment. *Nature* **524**, 280-283 (2015).

5 Congress of the Philippines. *Republic Act 9165*, <http://pdea.gov.ph/images/Laws/RA9165.pdf> (2002).

6 BBC News. *Tourists warned of UAE drug laws*, <http://news.bbc.co.uk/2/hi/uk_news/7234786.stm> (2008).

7 Government of Singapore. *Misuse of Drugs Act*, 2008).

8 Russo, E. B. *Cannabis and cannabinoids: pharmacology, toxicology, and therapeutic potential*. (Routledge, 2013).

9 Calvi, L. *et al.* Comprehensive quality evaluation of medical Cannabis sativa L. inflorescence and macerated oils based on HS-SPME coupled to GC–MS and LC-HRMS (q-exactive orbitrap®) approach. *J. Pharm. Biomed. Anal.* **150**, 208-219, doi:https://doi.org/10.1016/j.jpba.2017.11.073 (2018).

10 Baker, P. B., Fowler, R., Bagon, K. R. & Gough, T. A. Determination of the distribution of cannabinoids in cannabis resin using high performance liquid chromatography. *J. Anal. Toxicol.* **4**, 145-152 (1980).

11 Baker, P. B., Taylor, B. J. & Gough, T. A. The tetrahydrocannabinol and tetrahydrocannabinolic acid content of cannabis products. *J. Pharm. Pharmacol.* **33**, 369-372 (1981).

12 Campos, A. C., Moreira, F. A., Gomes, F. V., Del Bel, E. A. & Guimarães, F. S. Multiple mechanisms involved in the large-spectrum therapeutic potential of cannabidiol in psychiatric disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 3364-3378, doi:10.1098/rstb.2011.0389 (2012).

13 Perez, J. Combined cannabinoid therapy via an oromucosal spray. *Drugs of today (Barcelona, Spain : 1998)* **42**, 495-503, doi:10.1358/dot.2006.42.8.1021517 (2006).

14 GW pharmaceuticals. *Sativex (delta-9-tetrahydrocannabinol and cannabidiol)*, <https://www.gwpharm.com/products-pipeline/sativex-delta-9-tetrahydrocannabinol-and-cannabidiol> (2016).

15 Devane, W. A., Dysarz, F. A., 3rd, Johnson, M. R., Melvin, L. S. & Howlett, A. C. Determination and characterization of a cannabinoid receptor in rat brain. *Mol. Pharmacol.* **34**, 605-613 (1988).

16 Munro, S., Thomas, K. L. & Abu-Shaar, M. Molecular characterization of a peripheral receptor for cannabinoids. *Nature* **365**, 61-65, doi:10.1038/365061a0 (1993).

17 Hua, T. *et al.* Crystal structures of agonist-bound human cannabinoid receptor CB1. *Nature* **547**, 468-471, doi:10.1038/nature23272 (2017).
18 Szabo, B. & Schlicker, E. Effects of cannabinoids on neurotransmission. *Handb. Exp. Pharmacol.*, 327-365 (2005).
19 Howlett, A. C. & Abood, M. E. CB1 and CB2 Receptor Pharmacology. *Adv. Pharmacol.* **80**, 169-206, doi:10.1016/bs.apha.2017.03.007 (2017).
20 Pacher, P. & Mechoulam, R. Is lipid signaling through cannabinoid 2 receptors part of a protective system? *Prog. Lipid Res.* **50**, 193-211, doi:10.1016/j.plipres.2011.01.001 (2011).
21 Anand, P., Whiteside, G., Fowler, C. J. & Hohmann, A. G. Targeting CB2 receptors and the endocannabinoid system for the treatment of pain. *Brain Res. Rev.* **60**, 255-266, doi:10.1016/j.brainresrev.2008.12.003 (2009).
22 Fernandez-Ruiz, J., Pazos, M. R., Garcia-Arencibia, M., Sagredo, O. & Ramos, J. A. Role of CB2 receptors in neuroprotective effects of cannabinoids. *Mol. Cell. Endocrinol.* **286**, S91-96, doi:10.1016/j.mce.2008.01.001 (2008).
23 Mechoulam, R., Hanus, L. O., Pertwee, R. & Howlett, A. C. Early phytocannabinoid chemistry to endocannabinoids and beyond. *Nat. Rev. Neurosci.* **15**, 757-764, doi:10.1038/nrn3811 (2014).
24 Martin, B. R., Compton, D. R., Prescott, W. R., Barrett, R. L. & Razdan, R. K. Pharmacological evaluation of dimethylheptyl analogs of delta 9-THC: reassessment of the putative three-point cannabinoid-receptor interaction. *Drug Alcohol Depend.* **37**, 231-240, doi:10.1016/0376-8716(94)01081-U (1995).
25 Shim, J.-Y., Bertalovitz, A. C. & Kendall, D. a. Identification of essential cannabinoid-binding domains: structural insights into early dynamic events in receptor activation. *The Journal of biological chemistry* **286**, 33422-33435, doi:10.1074/jbc.M111.261651 (2011).
26 Kapur, A. *et al.* Mutation Studies of Ser7.39 and Ser2.60 in the Human CB1 Cannabinoid Receptor: Evidence for a Serine-Induced Bend in CB1 Transmembrane Helix 7. *Mol. Pharmacol.* **71**, 1512-1524, doi:10.1124/mol.107.034645.al. (2007).
27 Hua, T. *et al.* Crystal Structure of the Human Cannabinoid Receptor CB1. *Cell* **167**, 750-762.e714, doi:10.1016/j.cell.2016.10.004 (2016).
28 Shao, Z. *et al.* High-resolution crystal structure of the human CB1 cannabinoid receptor. *Nature*, doi:10.1038/nature20613 (2016).
29 Touw, M. The religious and medicinal uses of Cannabis in China, India and Tibet. *J. Psychoactive Drugs* **13**, 23-34, doi:10.1080/02791072.1981.10471447 (1981).
30 Zuardi, A. W. History of cannabis as a medicine: A review. *Revista Brasileira de Psiquiatria* **28**, 153-157, doi:10.1590/S1516-44462006000200015 (2006).
31 Li, H.-L. An archaeological and historical account of cannabis in China. *Econ. Bot.* **28**, 437-448, doi:10.1007/BF02862859 (1973).
32 Frankhauser, M. in *Cannabis and Cannabinoids* (ed Franjo; Russo Grotenhermen, Ethan B) Ch. 4, 37-51 (The Haworth Integrative Healing Press, 2002).
33 Mikuriya, T. H. Marijuana in medicine: past, present and future. *Calif. Med.* **110**, 34-40 (1969).

34 Aldrich, M. in *Cannabis in medical practice* (ed ML Mathre) 35-55 (MC Farland, 1997).

35 Mikuriya, T. H. & Aldrich, M. R. Cannabis 1988. Old drug, new dangers. The potency question. *J. Psychoactive Drugs* **20**, 47-55, doi:10.1080/02791072.1988.10524371 (1988).

36 Gaoni, Y. & Mechoulam, R. Isolation, structure, and partial synthesis of an active constituent of hashish. *J. Am. Chem. Soc.* **86**, 1646-1647 (1964).

37 Martin, B. R., Mechoulam, R. & Razdan, R. K. Discovery and characterization of endogenous cannabinoids. *Life Sci.* **65**, 573-595, doi:https://doi.org/10.1016/S0024-3205(99)00281-7 (1999).

38 National Center for Biotechnology Information. *Nabilone*, <https://pubchem.ncbi.nlm.nih.gov/compound/5284592> (

39 Kim, S. *et al.* PubChem Substance and Compound databases. *Nucleic Acids Res.* **44**, D1202-D1213, doi:10.1093/nar/gkv951 (2016).

40 Krcevski-Skvarc, N., Wells, C. & Häuser, W. Availability and approval of cannabis-based medicines for chronic pain management and palliative/supportive care in Europe: A survey of the status in the chapters of the European Pain Federation. *Eur. J. Pain* **22**, 440-454, doi:10.1002/ejp.1147 (2017).

41 Śledziński, P., Zeyland, J., Słomski, R. & Nowak, A. The current state and future perspectives of cannabinoids in cancer biology. *Cancer Medicine* **7**, 765-775, doi:10.1002/cam4.1312 (2018).

42 Porsteinsson, A. P. & Antonsdottir, I. M. An update on the advancements in the treatment of agitation in Alzheimer's disease. *Expert Opin. Pharmacother.* **18**, 611-620, doi:10.1080/14656566.2017.1307340 (2017).

43 Rosenberg, E. C., Patra, P. H. & Whalley, B. J. Therapeutic effects of cannabinoids in animal models of seizures, epilepsy, epileptogenesis, and epilepsy-related neuroprotection. *Epilepsy & behavior : E&B* **70**, 319-327, doi:10.1016/j.yebeh.2016.11.006 (2017).

44 Baker, P. B., Gough, T. A., Johncock, S. I., Taylor, B. J. & Wyles, L. T. Variation in the THC content in illicitly imported Cannabis products--Part II. *Bull. Narc.* **34**, 101-108 (1982).

45 Tart, C. T. Marijuana Intoxication : Common Experiences. *Nature* **226**, 701, doi:10.1038/226701a0 (1970).

46 Sewell, R. A. *et al.* Acute Effects of THC on Time Perception in Frequent and Infrequent Cannabis Users. *Psychopharmacology* **226**, 401-413, doi:10.1007/s00213-012-2915-6 (2013).

47 Karniol, I. G., Shirakawa, I., Takahashi, R. N., Knobel, E. & Musty, R. E. Effects of delta9-tetrahydrocannabinol and cannabinol in man. *Pharmacology* **13**, 502-512, doi:10.1159/000136944 (1975).

48 Volkow, N. D., Baler, R. D., Compton, W. M. & Weiss, S. R. B. Adverse health effects of marijuana use. *The New England journal of medicine* **370**, 2219-2227, doi:10.1056/NEJMra1402309 (2014).

49 Klonoff, H., Low, M. & Marcus, A. Neuropsychological effects of marijuana. *Can. Med. Assoc. J.* **108**, 150-156 passim (1973).

50 Metrik, J. *et al.* Balanced placebo design with marijuana: pharmacological and expectancy effects on impulsivity and risk taking. *Psychopharmacology (Berl.)* **223**, 489-499, doi:10.1007/s00213-012-2740-y (2012).

51 Ohlsson, A. *et al.* Plasma delta-9 tetrahydrocannabinol concentrations and clinical effects after oral and intravenous administration and smoking. *Clin. Pharmacol. Ther.* **28**, 409-416 (1980).

52 Ohlsson, A. *et al.* Single dose kinetics of deuterium labelled delta 1-tetrahydrocannabinol in heavy and light cannabis users. *Biomed. Mass Spectrom.* **9**, 6-10, doi:10.1002/bms.1200090103 (1982).

53 Wall, M. E., Sadler, B. M., Brine, D., Taylor, H. & Perez-Reyes, M. Metabolism, disposition, and kinetics of delta-9-tetrahydrocannabinol in men and women. *Clin. Pharmacol. Ther.* **34**, 352-363 (1983).

54 Huestis, M. A., Sampson, A. H., Holicky, B. J., Henningfield, J. E. & Cone, E. J. Characterization of the absorption phase of marijuana smoking. *Clin. Pharmacol. Ther.* **52**, 31-41 (1992).

55 Huestis, M. A., Henningfield, J. E. & Cone, E. J. Blood cannabinoids. I. Absorption of THC and formation of 11-OH-THC and THCCOOH during and after smoking marijuana. *J. Anal. Toxicol.* **16**, 276-282, doi:10.1093/jat/16.5.276 (1992).

56 Garrett, E. R. & Hunt, C. A. Physicochemical Properties, Solubility , and Protein Binding of Delta-9-Tetrahydrocannabinol. *J. Pharm. Sci.* **63**, 1056-1064, doi:10.1002/jps.2600630705 (1974).

57 Wahlqvist, M., Nilsson, I. M., Sandberg, F. & Agurell, S. Binding of delta-1-tetrahydrocannabinol to human plasma proteins. *Biochem. Pharmacol.* **19**, 2579-2584 (1970).

58 Huestis, M. A. Human Cannabinoid Pharmacokinetics. *Chem. Biodivers.* **4**, 1770-1804, doi:10.1002/cbdv.200790152 (2007).

59 Kreuz, D. S. & Axelrod, J. Delta-9-tetrahydrocannabinol: localization in body fat. *Science* **179**, 391-393 (1973).

60 Johansson, E., Noren, K., Sjovall, J. & Halldin, M. M. Determination of delta 1-tetrahydrocannabinol in human fat biopsies from marihuana users by gas chromatography-mass spectrometry. *Biomed. Chromatogr.* **3**, 35-38, doi:10.1002/bmc.1130030109 (1989).

61 Nahas, G. G., Frick, H. C., Lattimer, J. K., Latour, C. & Harvey, D. Pharmacokinetics of THC in brain and testis, male gametotoxicity and premature apoptosis of spermatozoa. *Hum. Psychopharmacol.* **17**, 103-113, doi:10.1002/hup.369 (2002).

62 Mura, P., Kintz, P., Dumestre, V., Raul, S. & Hauet, T. THC can be detected in brain while absent in blood. *J. Anal. Toxicol.* **29**, 842-843 (2005).

63 Matsunaga, T. *et al.* Metabolism of delta 9-tetrahydrocannabinol by cytochrome P450 isozymes purified from hepatic microsomes of monkeys. *Life Sci.* **56**, 2089-2095 (1995).

64 Halldin, M. M., Widman, M., C, V. B., Lindgren, J. E. & Martin, B. R. Identification of in vitro metabolites of delta 1-tetrahydrocannabinol formed by human livers. *Drug Metab. Dispos.* **10**, 297-301 (1982).

65 Wall, M. E. The in vitro and in vivo metabolism of tetrahydrocannabinol (THC). *Ann. N.Y. Acad. Sci.* **191**, 23-39, doi:doi:10.1111/j.1749-6632.1971.tb13984.x (1971).

66 Wall, M. E. & Perez-Reyes, M. The metabolism of delta 9-tetrahydrocannabinol and related cannabinoids in man. *J. Clin. Pharmacol.* **21**, 178s-189s (1981).

67 Williams, P. L. & Moffat, A. C. Identification in human urine of delta 9-tetrahydrocannabinol-11-oic acid glucuronide: a tetrahydrocannabinol metabolite. *J. Pharm. Pharmacol.* **32**, 445-448 (1980).

68 Mazur, A. *et al.* Characterization of human hepatic and extrahepatic UDP-glucuronosyltransferase enzymes involved in the metabolism of classic cannabinoids. *Drug Metab. Dispos.* **37**, 1496-1504, doi:10.1124/dmd.109.026898 (2009).

69 Garrett, E. R. & Hunt, C. A. Pharmacokinetics of delta9-tetrahydrocannabinol in dogs. *J. Pharm. Sci.* **66**, 395-407 (1977).

70 Watanabe, K., Tanaka, T., Yamamoto, I. & Yoshimura, H. Brain microsomal oxidation of delta 8- and delta 9-tetrahydrocannabinol. *Biochem. Biophys. Res. Commun.* **157**, 75-80 (1988).

71 Ellis, G. M., Jr., Mann, M. A., Judson, B. A., Schramm, N. T. & Tashchian, A. Excretion patterns of cannabinoid metabolites after last use in a group of chronic users. *Clin. Pharmacol. Ther.* **38**, 572-578 (1985).

72 Federal Highway Research Insitute. *Summary of Main DRUID Results*, <https://www.bast.de/Druid/EN/Dissemination/downloads_and_links/2012_Washington_Brochure.html?nn=613804> (2012).

73 Schulze, H., Schumacher, M., Urmeew, R. & Auerbach, K. DRUID Final Report: Work performed, main results and recommendations. (Federal Highway Research Institute, 2012).

74 Menetrey, A. *et al.* Assessment of driving capability through the use of clinical and psychomotor tests in relation to blood cannabinoids levels following oral administration of 20 mg dronabinol or of a cannabis decoction made with 20 or 60 mg Delta9-THC. *J. Anal. Toxicol.* **29**, 327-338 (2005).

75 Sexton, B. *et al.* The influence of cannabis on driving. *TRL report* **477**, 106 (2000).

76 Lenne, M. G. *et al.* The effects of cannabis and alcohol on simulated arterial driving: Influences of driving experience and task demand. *Accid. Anal. Prev.* **42**, 859-866, doi:10.1016/j.aap.2009.04.021 (2010).

77 Ramaekers, J. G., Robbe, H. W. & O'Hanlon, J. F. Marijuana, alcohol and actual driving performance. *Hum. Psychopharmacol.* **15**, 551-558, doi:10.1002/1099-1077(200010)15:7<551::aid-hup236>3.0.co;2-p (2000).

78 Ramaekers, J. G. *et al.* High-potency marijuana impairs executive function and inhibitory motor control. *Neuropsychopharmacology* **31**, 2296-2303, doi:10.1038/sj.npp.1301068 (2006).

79 Bosker, W. M. *et al.* Medicinal Delta(9) -tetrahydrocannabinol (dronabinol) impairs on-the-road driving performance of occasional and heavy cannabis users but is not detected in Standard Field Sobriety Tests. *Addiction* **107**, 1837-1844, doi:10.1111/j.1360-0443.2012.03928.x (2012).

80 Ronen, A. *et al.* The effect of alcohol, THC and their combination on perceived effects, willingness to drive and performance of driving and non-driving tasks. *Accid. Anal. Prev.* **42**, 1855-1865, doi:10.1016/j.aap.2010.05.006 (2010).

81 Anderson, B. M., Rizzo, M., Block, R. I., Pearlson, G. D. & O'Leary, D. S. Sex differences in the effects of marijuana on simulated driving performance. *J. Psychoactive Drugs* **42**, 19-30, doi:10.1080/02791072.2010.10399782 (2010).

82 Bondallaz, P. *et al.* Cannabis and its effects on driving skills. *Forensic Sci. Int.* **268**, 92-102, doi:10.1016/j.forsciint.2016.09.007 (2016).
83 Ronen, A. *et al.* Effects of THC on driving performance, physiological state and subjective feelings relative to alcohol. *Accid. Anal. Prev.* **40**, 926-934, doi:10.1016/j.aap.2007.10.011 (2008).
84 Sewell, R. A., Poling, J. & Sofuoglu, M. The effect of cannabis compared with alcohol on driving. *Am. J. Addict.* **18**, 185-193, doi:10.1080/10550490902786934 (2009).
85 Sexton, B. *et al.* (Crowthorne, Transport Research Laboratory, 2002).
86 McDonald, J., Schleifer, L., Richards, J. B. & de Wit, H. Effects of THC on behavioral measures of impulsivity in humans. *Neuropsychopharmacology* **28**, 1356-1365, doi:10.1038/sj.npp.1300176 (2003).
87 Lane, S. D., Cherek, D. R., Tcheremissine, O. V., Lieving, L. M. & Pietras, C. J. Acute marijuana effects on human risk taking. *Neuropsychopharmacology* **30**, 800-809, doi:10.1038/sj.npp.1300620 (2005).
88 Asbridge, M., Hayden, J. A. & Cartwright, J. L. Acute cannabis consumption and motor vehicle collision risk: systematic review of observational studies and meta-analysis. *BMJ* **344**, e536, doi:10.1136/bmj.e536 (2012).
89 Drummer, O. H. *et al.* The involvement of drugs in drivers of motor vehicles killed in Australian road traffic crashes. *Accid. Anal. Prev.* **36**, 239-248 (2004).
90 Laumon, B., Gadegbeku, B., Martin, J. L. & Biecheler, M. B. Cannabis intoxication and fatal road crashes in France: population based case-control study. *BMJ* **331**, 1371, doi:10.1136/bmj.38648.617986.1F (2005).
91 Hall, W. What has research over the past two decades revealed about the adverse health effects of recreational cannabis use? *Addiction* **110**, 19-35, doi:10.1111/add.12703 (2015).
92 Ramaekers, J. G., Berghaus, G., van Laar, M. & Drummer, O. H. Dose related risk of motor vehicle crashes after cannabis use. *Drug Alcohol Depend.* **73**, 109-119 (2004).
93 Poulsen, H., Moar, R. & Troncoso, C. The incidence of alcohol and other drugs in drivers killed in New Zealand road crashes 2004-2009. *Forensic Sci. Int.* **223**, 364-370, doi:10.1016/j.forsciint.2012.10.026 (2012).
94 Azofeifa, A., Mattson, M. E. & Lyerla, R. Driving under the influence of alcohol, marijuana, and alcohol and marijuana combined among persons aged 16–25 years—United States, 2002–2014. *MMWR* **64**, 1325-1329 (2015).
95 Downey, L. A. *et al.* The effects of cannabis and alcohol on simulated driving: Influences of dose and experience. *Accident; analysis and prevention* **50**, 879-886, doi:10.1016/j.aap.2012.07.016 (2013).
96 Lamers, C. T. & Ramaekers, J. G. Visual search and urban driving under the influence of marijuana and alcohol. *Hum. Psychopharmacol.* **16**, 393-401, doi:10.1002/hup.307 (2001).
97 Hartman, R. L. & Huestis, M. A. Cannabis effects on driving skills. *Clin. Chem.* **59**, 478-492, doi:10.1373/clinchem.2012.194381 (2013).
98 Gadegbeku, B., Amoros, E. & Laumon, B. Responsibility study: main illicit psychoactive substances among car drivers involved in fatal road crashes. *Annals of advances in automotive medicine. Association for the*

Advancement of Automotive Medicine. Annual Scientific Conference **55**, 293-300 (2011).

99 Pulido, J. *et al.* Cannabis use and traffic injuries. *Epidemiology* **22**, 609-610, doi:10.1097/EDE.0b013e31821db0c2 (2011).

100 Musshoff, F., Hokamp, E. G., Bott, U. & Madea, B. Performance evaluation of on-site oral fluid drug screening devices in normal police procedure in Germany. *Forensic Sci. Int.* **238**, 120-124, doi:10.1016/j.forsciint.2014.02.005 (2014).

101 Niedbala, R. S. *et al.* Detection of marijuana use by oral fluid and urine analysis following single-dose administration of smoked and oral marijuana. *J. Anal. Toxicol.* **25**, 289-303, doi:10.1093/jat/25.5.289 (2001).

102 Agius, R., Nadulski, T., Kahl, H. G. & Dufaux, B. Comparison of LUCIO(R)-direct ELISA with CEDIA immunoassay for 'zero tolerance' drug screening in urine as required by the German re-licensing guidelines. *Drug testing and analysis* **5**, 390-399, doi:10.1002/dta.1455 (2013).

103 Gallardo, E. & Queiroz, J. A. The role of alternative specimens in toxicological analysis. *Biomedical chromatography : BMC* **22**, 795-821, doi:10.1002/bmc.1009 (2008).

104 Choo, R. E. & Huestis, M. A. Oral fluid as a diagnostic tool. *Clin. Chem. Lab. Med.* **42**, 1273-1287, doi:10.1515/cclm.2004.248 (2004).

105 Bosker, W. M. & Huestis, M. A. Oral fluid testing for drugs of abuse. *Clin. Chem.* **55**, 1910-1931, doi:10.1373/clinchem.2008.108670 (2009).

106 Hart, C. L., van Gorp, W., Haney, M., Foltin, R. W. & Fischman, M. W. Effects of acute smoked marijuana on complex cognitive performance. *Neuropsychopharmacology* **25**, 757-765, doi:10.1016/s0893-133x(01)00273-1 (2001).

107 Nicholson, A. N., Turner, C., Stone, B. M. & Robson, P. J. Effect of Delta-9-tetrahydrocannabinol and cannabidiol on nocturnal sleep and early-morning behavior in young adults. *J. Clin. Psychopharmacol.* **24**, 305-313 (2004).

108 Lee, D. *et al.* Oral fluid cannabinoids in chronic, daily Cannabis smokers during sustained, monitored abstinence. *Clin. Chem.* **57**, 1127-1136, doi:10.1373/clinchem.2011.164822 (2011).

109 Hawks, R. L. in *The cannabinoids: chemical, pharmacologic, and therapeutic aspects* (eds Stig Agurell, William L Dewey, & Robert E Willette) 123-134 (Academic Press, 2012).

110 Hawks, R. L. The constituents of cannabis and the disposition and metabolism of cannabinoids. *NIDA Res. Monogr.* **42**, 125-137 (1982).

111 Ahmed, A. I. *et al.* Safety and pharmacokinetics of oral delta-9-tetrahydrocannabinol in healthy older subjects: a randomized controlled trial. *Eur. Neuropsychopharmacol.* **24**, 1475-1482, doi:10.1016/j.euroneuro.2014.06.007 (2014).

112 Milman, G., Schwope, D. M., Gorelick, D. a. & Huestis, M. a. Cannabinoids and metabolites in expectorated oral fluid following controlled smoked cannabis. *Clinica chimica acta; international journal of clinical chemistry* **413**, 765-770, doi:10.1016/j.cca.2012.01.011 (2012).

113 Dickson, S. *et al.* The recovery of illicit drugs from oral fluid sampling devices. *Forensic Sci. Int.* **165**, 78-84, doi:10.1016/j.forsciint.2006.03.004 (2007).

114 Quintela, O., Crouch, D. J. & Andrenyak, D. M. Recovery of drugs of abuse from the Immunalysis Quantisal oral fluid collection device. *J. Anal. Toxicol.* **30**, 614-616 (2006).

115 Maseda, C. *et al.* Detection of delta 9-THC in saliva by capillary GC/ECD after marihuana smoking. *Forensic Sci. Int.* **32**, 259-266 (1986).

116 Wong, R. C., Tran, M. & Tung, J. K. Oral fluid drug tests: effects of adulterants and foodstuffs. *Forensic Sci. Int.* **150**, 175-180, doi:10.1016/j.forsciint.2005.02.023 (2005).

117 Huestis, M. A. & Cone, E. J. Relationship of Delta 9-tetrahydrocannabinol concentrations in oral fluid and plasma after controlled administration of smoked cannabis. *J. Anal. Toxicol.* **28**, 394-399, doi:10.1093/jat/28.6.394 (2004).

118 Niedbala, S. *et al.* Passive cannabis smoke exposure and oral fluid testing. *J. Anal. Toxicol.* **28**, 546-552 (2004).

119 Moore, C. *et al.* Cannabinoids in oral fluid following passive exposure to marijuana smoke. *Forensic Sci. Int.* **212**, 227-230, doi:10.1016/j.forsciint.2011.06.019 (2011).

120 Menkes, D. B., Howard, R. C., Spears, G. F. & Cairns, E. R. Salivary THC following cannabis smoking correlates with subjective intoxication and heart rate. *Psychopharmacology (Berl.)* **103**, 277-279, doi:10.1007/BF02244217 (1991).

121 Ramaekers, J. G. *et al.* Cognition and motor control as a function of Delta9-THC concentration in serum and oral fluid: limits of impairment. *Drug Alcohol Depend.* **85**, 114-122, doi:10.1016/j.drugalcdep.2006.03.015 (2006).

122 Kauert, G. F., Ramaekers, J. G., Schneider, E., Moeller, M. R. & Toennes, S. W. Pharmacokinetic properties of delta9-tetrahydrocannabinol in serum and oral fluid. *J. Anal. Toxicol.* **31**, 288-293 (2007).

123 Pehrsson, A. *et al.* Roadside oral fluid testing: comparison of the results of drugwipe 5 and drugwipe benzodiazepines on-site tests with laboratory confirmation results of oral fluid and whole blood. *Forensic Sci. Int.* **175**, 140-148, doi:10.1016/j.forsciint.2007.05.022 (2008).

124 Goessaert, A.-S., Pil, K., Veramme, J. & Verstraete, A. Analytical evaluation of a rapid on-site oral fluid drug test. *Anal. Bioanal. Chem.* **396**, 2461-2468, doi:10.1007/s00216-010-3463-8 (2010).

125 Blencowe, T. *et al.* An analytical evaluation of eight on-site oral fluid drug screening devices using laboratory confirmation results from oral fluid. *Forensic Sci. Int.* **208**, 173-179, doi:10.1016/j.forsciint.2010.11.026 (2011).

126 Verstraete, A. ROSITA Roadsite Testing Assessment. (Ghent University, European Union, 2000).

127 Strano-Rossi, S. *et al.* Evaluation of four oral fluid devices (DDS®, Drugtest 5000®, Drugwipe 5+® and RapidSTAT®) for on-site monitoring drugged driving in comparison with UHPLC-MS/MS analysis. *Forensic Sci. Int.* **221**, 70-76, doi:10.1016/j.forsciint.2012.04.003 (2012).

128 Wille, S. M. R., Samyn, N., Ramírez-Fernández, M. D. M. & De Boeck, G. Evaluation of on-site oral fluid screening using Drugwipe-5(+), RapidSTAT and Drug Test 5000 for the detection of drugs of abuse in drivers. *Forensic Sci. Int.* **198**, 2-6, doi:10.1016/j.forsciint.2009.10.012 (2010).

129 Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505-510, doi:10.1126/science.2200121 (1990).
130 Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822, doi:10.1038/346818a0 (1990).
131 Tuerk, C., MacDougal, S. & Gold, L. RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 6988-6992 (1992).
132 Jenison, R. D., Gill, S. C., Pardi, A. & Polisky, B. High-resolution molecular discrimination by RNA. *Science* **263**, 1425-1429, doi:10.1126/science.7510417 (1994).
133 Hofmann, H. P., Limmer, S., Hornung, V. & Sprinzl, M. Ni2+-binding RNA motifs with an asymmetric purine-rich internal loop and a G-A base pair. *RNA (New York, N.Y.)* **3**, 1289-1300 (1997).
134 Mayer, G. *et al.* Fluorescence-activated cell sorting for aptamer SELEX with cell mixtures. *Nat. Protoc.* **5**, 1993-2004, doi:10.1038/nprot.2010.163 (2010).
135 London, G. M., Mayosi, B. M. & Khati, M. Isolation and characterization of 2'-F-RNA aptamers against whole HIV-1 subtype C envelope pseudovirus. *Biochem. Biophys. Res. Commun.* **456**, 428-433, doi:10.1016/j.bbrc.2014.11.101 (2015).
136 Mi, J. *et al.* In vivo selection of tumor-targeting RNA motifs. *Nat. Chem. Biol.* **6**, 22-24, doi:10.1038/nchembio.277 (2010).
137 Cho, E. J., Lee, J.-W. & Ellington, A. D. Applications of aptamers as sensors. *Annual review of analytical chemistry (Palo Alto, Calif.)* **2**, 241-264, doi:10.1146/annurev.anchem.1.031207.112851 (2009).
138 Ruigrok, V. J. B., Levisson, M., Eppink, M. H. M., Smidt, H. & van der Oost, J. Alternative affinity tools: more attractive than antibodies? *The Biochemical journal* **436**, 1-13, doi:10.1042/BJ20101860 (2011).
139 White, R. R., Sullenger, B. A. & Rusconi, C. P. Developing aptamers into therapeutics. *The Journal of Clinical Investigation* **106**, 929-934, doi:10.1172/JCI11325 (2000).
140 Fulle, L. *et al.* RNA Aptamers Recognizing Murine CCL17 Inhibit T Cell Chemotaxis and Reduce Contact Hypersensitivity In Vivo. *Mol. Ther.* **26**, 95-104, doi:10.1016/j.ymthe.2017.10.005 (2018).
141 Park, K. S. Nucleic acid aptamer-based methods for diagnosis of infections. *Biosens. Bioelectron.* **102**, 179-188, doi:10.1016/j.bios.2017.11.028 (2018).
142 Hori, S. I., Herrera, A., Rossi, J. J. & Zhou, J. Current Advances in Aptamers for Cancer Diagnosis and Therapy. *Cancers* **10**, doi:10.3390/cancers10010009 (2018).
143 Gragoudas, E. S., Adamis, A. P., Cunningham, E. T., Jr., Feinsod, M. & Guyer, D. R. Pegaptanib for neovascular age-related macular degeneration. *N. Engl. J. Med.* **351**, 2805-2816, doi:10.1056/NEJMoa042760 (2004).
144 Ng, E. W. *et al.* Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. *Nat. Rev. Drug Discov.* **5**, 123-132, doi:10.1038/nrd1955 (2006).

145 Poolsup, S. & Kim, C. Y. Therapeutic applications of synthetic nucleic acid aptamers. *Curr. Opin. Biotechnol.* **48**, 180-186, doi:10.1016/j.copbio.2017.05.004 (2017).
146 Nimjee, S. M., White, R. R., Becker, R. C. & Sullenger, B. A. Aptamers as Therapeutics. *Annu. Rev. Pharmacol. Toxicol.* **57**, 61-79, doi:10.1146/annurev-pharmtox-010716-104558 (2017).
147 Pfeiffer, F. & Mayer, G. Selection and Biosensor Application of Aptamers for Small Molecules. *Frontiers in chemistry* **4**, 25, doi:10.3389/fchem.2016.00025 (2016).
148 Han, K., Liang, Z. & Zhou, N. Design strategies for aptamer-based biosensors. *Sensors (Basel, Switzerland)* **10**, 4541-4557, doi:10.3390/s100504541 (2010).
149 Stoltenburg, R., Reinemann, C. & Strehlitz, B. SELEX--a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol. Eng* **24**, 381-403, doi:10.1016/j.bioeng.2007.06.001 (2007).
150 Mendonsa, S. D. & Bowser, M. T. In vitro evolution of functional DNA using capillary electrophoresis. *J. Am. Chem. Soc.* **126**, 20-21, doi:10.1021/ja037832s (2004).
151 Stoltenburg, R., Nikolaus, N. & Strehlitz, B. Capture-SELEX: Selection of DNA Aptamers for Aminoglycoside Antibiotics. *Journal of analytical methods in chemistry* **2012**, 415697, doi:10.1155/2012/415697 (2012).
152 Pfeiffer, F. *et al.* Identification and characterization of nucleobase-modified aptamers by click-SELEX. *Nat. Protoc.* **13**, 1153-1180, doi:10.1038/nprot.2018.023 (2018).
153 Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463-5467 (1977).
154 Blank, M. Next-Generation Analysis of Deep Sequencing Data: Bringing Light into the Black Box of SELEX Experiments. *Methods Mol. Biol.* **1380**, 85-95, doi:10.1007/978-1-4939-3197-2_7 (2016).
155 Dupont, D. M., Larsen, N., Jensen, J. K., Andreasen, P. A. & Kjems, J. Characterisation of aptamer-target interactions by branched selection and high-throughput sequencing of SELEX pools. *Nucleic Acids Res.* **43**, e139, doi:10.1093/nar/gkv700 (2015).
156 Zimmermann, B., Gesell, T., Chen, D., Lorenz, C. & Schroeder, R. Monitoring Genomic Sequences during SELEX Using High-Throughput Sequencing: Neutral SELEX. *PLoS ONE* **5**, e9169, doi:10.1371/journal.pone.0009169 (2010).
157 Beier, R. *et al.* Selection of a DNA aptamer against norovirus capsid protein VP1. *FEMS Microbiol. Lett.* **351**, 162-169, doi:10.1111/1574-6968.12366 (2014).
158 Thiel, W. H. *et al.* Nucleotide bias observed with a short SELEX RNA aptamer library. *Nucleic Acid Ther* **21**, 253-263, doi:10.1089/nat.2011.0288 (2011).
159 Hoinka, J. & Przytycka, T. AptaPLEX - A dedicated, multithreaded demultiplexer for HT-SELEX data. *Methods* **106**, 82-85, doi:10.1016/j.ymeth.2016.04.011 (2016).
160 Thiel, W. H. & Giangrande, P. H. Analyzing HT-SELEX data with the Galaxy Project tools--A web based bioinformatics platform for biomedical research. *Methods* **97**, 3-10, doi:10.1016/j.ymeth.2015.10.008 (2016).

161 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* **32**, 161-164, doi:10.1093/bioinformatics/btv545 (2016).

162 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. *Research in computational molecular biology : ... Annual International Conference, RECOMB ... : proceedings. RECOMB (Conference : 2005-)* **8394**, 115-128, doi:10.1007/978-3-319-05269-4_9 (2014).

163 Goldfeder, R. L., Wall, D. P., Khoury, M. J., Ioannidis, J. P. A. & Ashley, E. A. Human Genome Sequencing at the Population Scale: A Primer on High-Throughput DNA Sequencing and Analysis. *Am. J. Epidemiol.* **186**, 1000-1009, doi:10.1093/aje/kww224 (2017).

164 Illumina. *Next-Generation Sequencing: Sequencing by Synthesis (SBS) Technology*, <https://emea.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html> (2018).

165 Fox, E. J., Reid-Bayliss, K. S., Emond, M. J. & Loeb, L. A. Accuracy of Next Generation Sequencing Platforms. *Next generation, sequencing & applications* **1**, doi:10.4172/jngsa.1000106 (2014).

166 Dohm, J. C., Lottaz, C., Borodina, T. & Himmelbauer, H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.* **36**, e105, doi:10.1093/nar/gkn425 (2008).

167 May, A. *et al.* NGS-eval: NGS Error analysis and novel sequence VAriant detection tooL. *Nucleic Acids Res.* **43**, W301-305, doi:10.1093/nar/gkv346 (2015).

168 Kelley, D. R., Schatz, M. C. & Salzberg, S. L. Quake: quality-aware detection and correction of sequencing errors. *Genome Biol.* **11**, R116, doi:10.1186/gb-2010-11-11-r116 (2010).

169 Gundry, M. & Vijg, J. Direct mutation analysis by high-throughput sequencing: from germline to low-abundant, somatic variants. *Mutat. Res.* **729**, 1-15, doi:10.1016/mrfmmm.2011.10.001 (2012).

170 Fuller, C. W. *et al.* The challenges of sequencing by synthesis. *Nat. Biotechnol.* **27**, 1013-1023, doi:10.1038/nbt.1585 (2009).

171 Massingham, T. & Goldman, N. All Your Base: a fast and accurate probabilistic approach to base calling. *Genome Biol.* **13**, R13, doi:10.1186/gb-2012-13-2-r13 (2012).

172 Wang, B., Wan, L., Wang, A. & Li, L. M. An adaptive decorrelation method removes Illumina DNA base-calling errors caused by crosstalk between adjacent clusters. *Scientific reports* **7**, 41348, doi:10.1038/srep41348 (2017).

173 Schirmer, M. *et al.* Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res.* **43**, e37-e37, doi:10.1093/nar/gku1341 (2015).

174 Kao, W. C., Stevens, K. & Song, Y. S. BayesCall: A model-based base-calling algorithm for high-throughput short-read sequencing. *Genome Res.* **19**, 1884-1895, doi:10.1101/gr.095299.109 (2009).

175 Kircher, M., Stenzel, U. & Kelso, J. Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biol.* **10**, R83, doi:10.1186/gb-2009-10-8-r83 (2009).

176 Keefe, A. D., Pai, S. & Ellington, A. Aptamers as therapeutics. *Nat. Rev. Drug Discov.* **9**, 537-550, doi:10.1038/nrd3141 (2010).
177 Hernandez, F. J., Kalra, N., Wengel, J. & Vester, B. Aptamers as a model for functional evaluation of LNA and 2'-amino LNA. *Bioorg. Med. Chem. Lett.* **19**, 6585-6587, doi:10.1016/j.bmcl.2009.10.039 (2009).
178 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, doi:https://doi.org/10.1016/S0040-4020(98)00094-5 (1998).
179 Pfeiffer, F., Rosenthal, M., Siegl, J., Ewers, J. & Mayer, G. Customised nucleic acid libraries for enhanced aptamer selection and performance. *Curr. Opin. Biotechnol.* **48**, 111-118, doi:10.1016/j.copbio.2017.03.026 (2017).
180 Gold, L. *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS ONE* **5**, e15004, doi:10.1371/journal.pone.0015004 (2010).
181 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).
182 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, doi:10.1038/355564a0 (1992).
183 Chen, T., Hongdilokkul, N., Liu, Z., Thirunavukarasu, D. & Romesberg, F. E. The expanding world of DNA and RNA. *Curr. Opin. Chem. Biol.* **34**, 80-87, doi:10.1016/j.cbpa.2016.08.001 (2016).
184 Kimoto, M., Yamashige, R., Matsunaga, K., Yokoyama, S. & Hirao, I. Generation of high-affinity DNA aptamers using an expanded genetic alphabet. *Nat. Biotechnol.* **31**, 453-457, doi:10.1038/nbt.2556 (2013).
185 Kimoto, M., Matsunaga, K. & Hirao, I. DNA Aptamer Generation by Genetic Alphabet Expansion SELEX (ExSELEX) Using an Unnatural Base Pair System. *Methods Mol. Biol.* **1380**, 47-60, doi:10.1007/978-1-4939-3197-2_4 (2016).
186 Matsunaga, K. I., Kimoto, M. & Hirao, I. High-Affinity DNA Aptamer Generation Targeting von Willebrand Factor A1-Domain by Genetic Alphabet Expansion for Systematic Evolution of Ligands by Exponential Enrichment Using Two Types of Libraries Composed of Five Different Bases. *J. Am. Chem. Soc.*, doi:10.1021/jacs.6b10767 (2016).
187 Kimoto, M., Nakamura, M. & Hirao, I. Post-ExSELEX stabilization of an unnatural-base DNA aptamer targeting VEGF165 toward pharmaceutical applications. *Nucleic Acids Res.* **44**, 7487-7494, doi:10.1093/nar/gkw619 (2016).
188 Sefah, K. *et al.* In vitro selection with artificial expanded genetic information systems. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 1449-1454, doi:10.1073/pnas.1311778111 (2014).
189 Zhang, L. *et al.* Aptamers against Cells Overexpressing Glypican 3 from Expanded Genetic Systems Combined with Cell Engineering and Laboratory Evolution. *Angew. Chem. Int. Ed. Engl.* **55**, 12372-12375, doi:10.1002/anie.201605058 (2016).

190 Biondi, E. *et al.* Laboratory evolution of artificially expanded DNA gives redesignable aptamers that target the toxic form of anthrax protective antigen. *Nucleic Acids Res.*, doi:10.1093/nar/gkw890 (2016).
191 Rohloff, J. C. *et al.* Nucleic Acid Ligands With Protein-like Side Chains: Modified Aptamers and Their Use as Diagnostic and Therapeutic Agents. *Molecular therapy. Nucleic acids* **3**, e201, doi:10.1038/mtna.2014.49 (2014).
192 Vaught, J. D. *et al.* Expanding the chemistry of DNA for in vitro selection. *J. Am. Chem. Soc.* **132**, 4141-4151, doi:10.1021/ja908035g (2010).
193 Gawande, B. N. *et al.* Selection of DNA aptamers with two modified bases. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 2898-2903, doi:10.1073/pnas.1615475114 (2017).
194 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, doi:10.1073/pnas.1213933109 (2012).
195 Gelinas, A. D. *et al.* Crystal structure of interleukin-6 in complex with a modified nucleic acid ligand. *J. Biol. Chem.* **289**, 8720-8734, doi:10.1074/jbc.M113.532697 (2014).
196 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, doi:10.1016/j.str.2015.03.027 (2015).
197 Ren, X., Gelinas, A. D., von Carlowitz, I., Janjic, N. & Pyle, A. M. Structural basis for IL-1alpha recognition by a modified DNA aptamer that specifically inhibits IL-1alpha signaling. *Nat Commun* **8**, 810, doi:10.1038/s41467-017-00864-2 (2017).
198 Moreland, J. L., Gramada, A., Buzko, O. V., Zhang, Q. & Bourne, P. E. The Molecular Biology Toolkit (MBT): a modular platform for developing molecular visualization applications. *BMC Bioinformatics* **6**, 21, doi:10.1186/1471-2105-6-21 (2005).
199 Horisawa, K. Specific and quantitative labeling of biomolecules using click chemistry. *Front Physiol* **5**, doi:ARTN 45710.3389/fphys.2014.00457 (2014).
200 Li, L. & Zhang, Z. Development and Applications of the Copper-Catalyzed Azide-Alkyne Cycloaddition (CuAAC) as a Bioorthogonal Reaction. *Molecules* **21**, doi:10.3390/molecules21101393 (2016).
201 Chan, T. R., Hilgraf, R., Sharpless, K. B. & Fokin, V. V. Polytriazoles as copper(I)-stabilizing ligands in catalysis. *Org. Lett.* **6**, 2853-2855, doi:10.1021/ol0493094 (2004).
202 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, doi:10.1021/ol0610946 (2006).
203 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, doi:10.1002/chem.200700502 (2007).
204 Gramlich, P. M., 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, doi:10.1002/anie.200802077 (2008).

205 Tolle, F., Brändle, G. M., Matzner, D. & Mayer, G. A Versatile Approach Towards Nucleobase-Modified Aptamers. *Angewandte Chemie (International ed. in English)* **54**, 10971-10974, doi:10.1002/anie.201503652 (2015).
206 ThermoFisher Scientific. *Protocols M-270 Amine*, <https://www.thermofisher.com/de/de/home/references/protocols/proteins-expression-isolation-and-analysis/protein-isolation-protocol/m-270-amine.html> (2007).
207 DrugBank. *Phenol*, <https://www.drugbank.ca/drugs/DB03255> (2005).
208 Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074-d1082, doi:10.1093/nar/gkx1037 (2018).
209 Beltman, J. B. *et al.* Reproducibility of Illumina platform deep sequencing errors allows accurate determination of DNA barcodes in cells. *BMC Bioinformatics* **17**, 151, doi:10.1186/s12859-016-0999-4 (2016).
210 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, doi:10.1021/jo401780u (2013).
211 Tolle, F., Rosenthal, M., Pfeiffer, F. & Mayer, G. Click Reaction on Solid Phase Enables High Fidelity Synthesis of Nucleobase-Modified DNA. *Bioconjug. Chem.* **27**, 500-503, doi:10.1021/acs.bioconjchem.5b00668 (2016).
212 Hazekamp, A. & Verpoorte, R. Structure elucidation of the tetrahydrocannabinol complex with randomly methylated beta-cyclodextrin. *Eur. J. Pharm. Sci.* **29**, 340-347, doi:10.1016/j.ejps.2006.07.001 (2006).
213 DrugBank. *Dronabinol*, <https://www.drugbank.ca/drugs/DB00470> (2005).
214 Aphios Corporation. *Δ9-Tetrahydrocannabinol (Δ9-THC)*, <https://www.aphios.com/products/research-chemicals/d9-thc.html> (2012).
215 Abe, H. *et al.* Structure formation and catalytic activity of DNA dissolved in organic solvents. *Angewandte Chemie (International ed. in English)* **51**, 6475-6479, doi:10.1002/anie.201201111 (2012).
216 Cayman Chemical. *Δ9-THC*, <https://www.caymanchem.com/product/12068> (2018).
217 ThermoFisher Scientific. *Tween™ 20 Surfact-Amps™ Detergent Solution*, <http://www.thermofisher.com/order/catalog/product/28320> (2018).
218 Hafner, M. *et al.* Displacement of protein-bound aptamers with small molecules screened by fluorescence polarization. *Nat. Protoc.* **3**, 579-587, doi:10.1038/nprot.2008.15 (2008).
219 Arkin, M. R., Glicksman, M. A., Fu, H., Havel, J. J. & Du, Y. in *Assay Guidance Manual* (eds G. S. Sittampalam *et al.*) (Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2004).
220 Krüss GmbH. *Critical micelle concentration (CMC) and surfactant concentration*, <https://www.kruss-scientific.com/services/education-theory/glossary/critical-micelle-concentration-cmc-and-surfactant-concentration/> (2018).
221 Merck KGaA. *Bovine Serum Albumin*, <https://www.sigmaaldrich.com/catalog/product/sigma/05482> (2018).

222 Fitzgerald Industries International. *THC antibody (10-T43F)*, <https://www.fitzgerald-fii.com/thc-antibody-10-t43f.html> (2018).
223 Janeway, C. A., Travers, P., Walport, M. & Shlomchik, M. J. in *Immunobiology: the immune system in health and disease* (Garland Science, 2001).
224 Merck KGaA. *Streptavidin, recombinant*, <https://www.sigmaaldrich.com/catalog/product/roche/rstrepro> (2018).
225 White, R. *et al.* Generation of species cross-reactive aptamers using "toggle" SELEX. *Mol. Ther.* **4**, 567-573, doi:10.1006/mthe.2001.0495 (2001).
226 Leung, V. W. & Darvell, B. W. Artificial salivas for in vitro studies of dental materials. *J. Dent.* **25**, 475-484, doi:10.1016/S0300-5712(96)00068-1 (1997).
227 Merck KGaA. *Epoxy-activated−Sepharose® 6B* <https://www.sigmaaldrich.com/catalog/product/sigma/e6754> (2018).
228 invitrogen & DYNAL. *Surface-activated dynabeads*, <http://www.thermofisher.com/order/catalog/product/14307D> (2010).
229 Park, N. J., Li, Y., Yu, T., Brinkman, B. M. N. & Wong, D. T. Characterization of RNA in Saliva. *Clin. Chem.* **52**, 988-994, doi:10.1373/clinchem.2005.063206 (2006).
230 Merck KGaA. *238470 | Cremophor EL® - CAS 61791-12-6 - Calbiochem*, <http://www.merckmillipore.com/DE/de/product/Cremophor-EL-CAS-61791-12-6-Calbiochem,EMD_BIO-238470?ReferrerURL=https%3A%2F%2Fwww.researchgate.net%2F&bd=1> (2018).
231 Repka, M. A., Munjal, M., ElSohly, M. A. & Ross, S. A. Temperature Stability and Bioadhesive Properties of Δ(9)-Tetrahydrocannabinol Incorporated Hydroxypropylcellulose Polymer Matrix Systems. *Drug Dev. Ind. Pharm.* **32**, 21-32, doi:10.1080/03639040500387914 (2006).
232 Olsen, C. M., Lee, H. T. & Marky, L. A. Unfolding thermodynamics of intramolecular G-quadruplexes: base sequence contributions of the loops. *J. Phys. Chem. B* **113**, 2587-2595, doi:10.1021/jp806853n (2009).
233 Contreras Jiménez, G. *et al.* Aptamer-Based Label-Free Impedimetric Biosensor for Detection of Progesterone. *Anal. Chem.* **87**, 1075-1082, doi:10.1021/ac503639s (2015).
234 Kim, Y. S. *et al.* Electrochemical detection of 17β-estradiol using DNA aptamer immobilized gold electrode chip. *Biosens. Bioelectron.* **22**, 2525-2531, doi:http://dx.doi.org/10.1016/j.bios.2006.10.004 (2007).
235 Alsager, O. a., Kumar, S., Willmott, G. R., McNatty, K. P. & Hodgkiss, J. M. Small molecule detection in solution via the size contraction response of aptamer functionalized nanoparticles. *Biosensors & bioelectronics* **57**, 262-268, doi:10.1016/j.bios.2014.02.004 (2014).
236 Akki, S. U., Werth, C. J. & Silverman, S. K. Selective Aptamers for Detection of Estradiol and Ethynylestradiol in Natural Waters. *Environ. Sci. Technol.* **49**, 9905-9913, doi:10.1021/acs.est.5b02401 (2015).
237 Martin, J. A. *et al.* Tunable stringency aptamer selection and gold nanoparticle assay for detection of cortisol. *Anal. Bioanal. Chem.* **406**, 4637-4647, doi:10.1007/s00216-014-7883-8 (2014).
238 DrugBank. *Progesterone*, <https://www.drugbank.ca/drugs/DB00396> (2018).

239 DrugBank. *Estradiol*, <https://www.drugbank.ca/drugs/DB00783> (2018).
240 DrugBank. *Hydrocortisone*, <https://www.drugbank.ca/drugs/DB00741> (2018).
241 National Center for Biotechnology Information. *3,3',4,4'-Tetrachlorobiphenyl*, <https://pubchem.ncbi.nlm.nih.gov/compound/3_3__4_4_-Tetrachlorobiphenyl> (
242 National Center for Biotechnology Information. *2,3',5,5'-Tetrachlorobiphenyl*, <https://pubchem.ncbi.nlm.nih.gov/compound/2_3__5_5_-tetrachlorobiphenyl> (
243 National Center for Biotechnology Information. *2,3,3',4,5-Pentachlorobiphenyl*, <https://pubchem.ncbi.nlm.nih.gov/compound/2_3_3__4_5-pentachlorobiphenyl> (
244 Xu, S. *et al.* Selection of DNA aptamers against polychlorinated biphenyls as potential biorecognition elements for environmental analysis. *Anal. Biochem.* **423**, 195-201, doi:http://dx.doi.org/10.1016/j.ab.2012.01.026 (2012).
245 Mehta, J. *et al.* Selection and characterization of PCB-binding DNA aptamers. *Anal. Chem.* **84**, 1669-1676, doi:10.1021/ac202960b (2012).
246 Yang, Q., Goldstein, I. J., Mei, H. Y. & Engelke, D. R. DNA ligands that bind tightly and selectively to cellobiose. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5462-5467 (1998).
247 Oyola, S. O. *et al.* Optimizing Illumina next-generation sequencing library preparation for extremely AT-biased genomes. *BMC Genomics* **13**, 1, doi:10.1186/1471-2164-13-1 (2012).
248 Piganeau, N., Jenne, A., Thuillier, V. & Famulok, M. An Allosteric Ribozyme Regulated by Doxycyline. *Angew. Chem. Int. Ed. Engl.* **39**, 4369-4373, doi:10.1002/1521-3773(20001201)39:23<4369::aid-anie4369>3.0.co;2-n (2000).
249 McGinness, K. E., Wright, M. C. & Joyce, G. F. Continuous In Vitro Evolution of a Ribozyme that Catalyzes Three Successive Nucleotidyl Addition Reactions. *Chem. Biol.* **9**, 585-596, doi:https://doi.org/10.1016/S1074-5521(02)00136-9 (2002).
250 Tsang, J. & Joyce, G. F. Specialization of the DNA-cleaving Activity of a Group I Ribozyme ThroughIn VitroEvolution. *J. Mol. Biol.* **262**, 31-42, doi:https://doi.org/10.1006/jmbi.1996.0496 (1996).
251 Li, B. *et al.* QPLOT: a quality assessment tool for next generation sequencing data. *BioMed research international* **2013**, 865181, doi:10.1155/2013/865181 (2013).
252 Lee, M. *et al.* Mutation Analysis of Synthetic DNA Barcodes in a Fission Yeast Gene Deletion Library by Sanger Sequencing. *Genomics & informatics* **16**, 22-29, doi:10.5808/gi.2018.16.2.22 (2018).
253 National Center for Biotechnology Information. *Ethanol*, <https://pubchem.ncbi.nlm.nih.gov/compound/702#section=Top> (
254 National Center for Biotechnology Information. *Acetonitrile*, <https://pubchem.ncbi.nlm.nih.gov/compound/acetonitrile#section=Top> (
255 Oliver, M. *et al.* Trifluoromethylated proline analogues as efficient tools to enhance the hydrophobicity and to promote passive diffusion transport of

the l-prolyl-l-leucyl glycinamide (PLG) tripeptide. *RSC Advances* **8**, 14597-14602, doi:10.1039/C8RA02511H (2018).

256 Panchompoo, J., Aldous, L., Baker, M., Wallace, M. I. & Compton, R. G. One-step synthesis of fluorescein modified nano-carbon for Pd(II) detection via fluorescence quenching. *Analyst* **137**, 2054-2062, doi:10.1039/c2an16261j (2012).

257 Riche, E., Carrie, A., Andin, N. & Mabic, S. *High-Purity Water and pH*, <https://www.americanlaboratory.com/913-Technical-Articles/35755-High-Purity-Water-and-pH/> (2006).

258 Vodolazkaya, N. A., Gurina, Y. A., Salamanova, N. V. & Mchedlov-Petrossyan, N. O. Spectroscopic study of acid–base ionization and tautomerism of fluorescein dyes in direct microemulsions at high bulk ionic strength. *J. Mol. Liq.* **145**, 188-196 (2009).

259 Haque, S. & Poddar, M. Interaction of delta-9-tetrahydrocannabinol with bovine serum albumin. *IRCS Medical Science* **11**, 290-291 (1983).

260 Jeyachandran, Y. L., Mielczarski, E., Rai, B. & Mielczarski, J. A. Quantitative and qualitative evaluation of adsorption/desorption of bovine serum albumin on hydrophilic and hydrophobic surfaces. *Langmuir* **25**, 11614-11620, doi:10.1021/la901453a (2009).

261 Thaplyal, P. & Bevilacqua, P. C. Experimental Approaches for Measuring pK(a)'s in RNA and DNA. *Methods Enzymol.* **549**, 189-219, doi:10.1016/B978-0-12-801122-5.00009-X (2014).

262 Yet, L. in *Modern Heterocyclic Chemistry* Vol. 2 (eds Julio Alvarez-Builla, Juan Jose Vaquero, & José Barluenga) (John Wiley & Sons, 2011).

263 Streitwieser, A., Granger, M., Mares, F. & Wolf, R. Acidity of hydrocarbons. XLVIII. Kinetic acidities of mono-, di-, and triarylmethanes toward lithium cyclohexylamide. *J. Am. Chem. Soc.* **95**, 4257-4261 (1973).

264 GE Healthcare Bio-Sciences AB. *Epoxy-activated sepharose 6B*, <https://www.gelifesciences.com/en/ca/shop/chromatography/resins/affinity-specific-groups/epoxy-activated-sepharose-6b-p-05691> (2010).

265 McKeague, M. & Derosa, M. C. Challenges and opportunities for small molecule aptamer development. *Journal of nucleic acids* **2012**, 748913, doi:10.1155/2012/748913 (2012).

266 Zhao, T. *et al.* Nanoprobe-Enhanced, Split Aptamer-Based Electrochemical Sandwich Assay for Ultrasensitive Detection of Small Molecules. *Anal. Chem.* **87**, 7712-7719, doi:10.1021/acs.analchem.5b01178 (2015).

267 Tolle, F. *Click-SELEX: A versatile approach towards nucleobase-modified aptamers* PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, (2016).

268 Merck KGaA. *Adipic acid dihydrazide–Agarose*, <https://www.sigmaaldrich.com/catalog/product/sigma/a0802> (2018).

269 Nutiu, R. & Li, Y. In Vitro Selection of Structure-Switching Signaling Aptamers. *Angew. Chem. Int. Ed.* **44**, 1061-1065, doi:10.1002/anie.200461848 (2005).

270 Nikolaus, N. & Strehlitz, B. DNA-Aptamers Binding Aminoglycoside Antibiotics. *Sensors (Basel, Switzerland)* **14**, 3737-3755, doi:10.3390/s140203737 (2014).

271 Reinemann, C., Freiin von Fritsch, U., Rudolph, S. & Strehlitz, B. Generation and characterization of quinolone-specific DNA aptamers

suitable for water monitoring. *Biosens. Bioelectron.* **77**, 1039-1047, doi:10.1016/j.bios.2015.10.069 (2016).

272 Paniel, N. *et al.* Selection of DNA aptamers against penicillin G using Capture-SELEX for the development of an impedimetric sensor. *Talanta* **162**, 232-240, doi:10.1016/j.talanta.2016.09.058 (2017).

273 Kuznetsov, A., Komarova, N., Andrianova, M., Grudtsov, V. & Kuznetsov, E. Aptamer based vanillin sensor using an ion-sensitive field-effect transistor. *Mikrochim. Acta* **185**, 3, doi:10.1007/s00604-017-2586-4 (2017).

274 Spiga, F. M., Maietta, P. & Guiducci, C. More DNA-Aptamers for Small Drugs: A Capture-SELEX Coupled with Surface Plasmon Resonance and High-Throughput Sequencing. *ACS Comb Sci* **17**, 326-333, doi:10.1021/acscombsci.5b00023 (2015).

275 Lauridsen, L. H., Doessing, H. B., Long, K. S. & Nielsen, A. T. A Capture-SELEX Strategy for Multiplexed Selection of RNA Aptamers Against Small Molecules. *Methods Mol. Biol.* **1671**, 291-306, doi:10.1007/978-1-4939-7295-1_18 (2018).

276 Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, doi:10.1038/msb.2011.75 (2011).

277 Goujon, M. *et al.* A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Res.* **38**, W695-W699, doi:10.1093/nar/gkq313 (2010).

278 Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28-36 (1994).

279 Tolle, F. & Mayer, G. Preparation of SELEX Samples for Next-Generation Sequencing. *Methods Mol. Biol.* **1380**, 77-84, doi:10.1007/978-1-4939-3197-2_6 (2016).

280 Civit, L. *et al.* Systematic evaluation of cell-SELEX enriched aptamers binding to breast cancer cells. *Biochimie*, doi:10.1016/j.biochi.2017.10.007 (2017).

281 Pfeiffer, F. *et al.* Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Scientific reports* **8**, 10950, doi:10.1038/s41598-018-29325-6 (2018).

282 Blind, M. & Blank, M. Aptamer Selection Technology and Recent Advances. *Molecular therapy. Nucleic acids* **4**, e223, doi:10.1038/mtna.2014.74 (2015).

283 Taldone, T. *et al.* Synthesis of purine-scaffold fluorescent probes for heat shock protein 90 with use in flow cytometry and fluorescence microscopy. *Bioorg. Med. Chem. Lett.* **21**, 5347-5352, doi:10.1016/j.bmcl.2011.07.026 (2011).

284 Llauger-Bufi, L., Felts, S. J., Huezo, H., Rosen, N. & Chiosis, G. Synthesis of novel fluorescent probes for the molecular chaperone Hsp90. *Bioorg. Med. Chem. Lett.* **13**, 3975-3978 (2003).

285 MP Biomedicals, L. *Technical Information Fluorescein Isothiocyanate*, <https://www.mpbio.com/includes/msds/02100276/MP_DS_02100276.pdf > (2018).

Acknowledgment

A multitude of people were part of the development of this thesis in one way or another.

My utmost thanks go to Prof. Dr. Günter Mayer for this topic, many fruitful discussions, and his never-ending support during as well as before this thesis. You have had the most significant influence on my development as a scientist and I am grateful for all advice given, as hard as it was sometimes.

My gratitude also goes to Prof. Dr. Irmgard Förster for helpful discussions, a wonderful collaboration and the willingness to become my co-corrector.

Thank you to Prof. Dr. Christoph Thiele and Prof. Dr. Ulrich Ettinger for becoming part of my committee.

The work presented in this thesis was financially supported by ZIM – Zentrales Innovationsprogramm Mittelstand. Thank you for enabling this project.

For the supply of THC and many helpful comments, I want to thank PD Dr. Andras Bilkei-Gorzo. For obvious reasons, this project would not have been possible without you.

The development of a protocol for next-generation sequencing had a small impact on this final thesis, but a large one on some of my experiments and my time in the lab. Thank you to PD Dr. Marc Beyer for his help in establishing the protocol for sample preparation and to Dr. Kristian Händler and Kathrin Klee for the actual sequencing. Thank you to Dr. Michael Blank and Dr. Carsten Gröber from AptaIT GmbH for long and fruitful discussions that finally resulted in a working NGS-workflow, advice whenever needed, and the willingness to reanalyse my samples as often as I needed.

I would like to thank my cooperation partners, Bernd Tiemann and Christoph Protzek from Protzek Gesellschaft für biomedizinische Technik mbH and Prof. Dr. Rolf Aderjan for their help whenever needed. I would have liked to give you a THC-binding clickmer to use for assay development and sincerely hope that will happen in the future, even though I am not to be the one selecting it.

An enormous thank you to Malte Rosenthal, who suffered through this topic with me and provided me with THCA-beads as well as CF_3-azide.

For the patient advice and willingness to help at my start of this thesis I would like to thank Björn Niebel. I also want to thank my student, Patrick Günther, who performed SELEX 3.

Thank you to Dr. Daniel Matzner, Dr. Jeffrey Hannam, Dr. Gerhard Brändle, and Jörg Ewers for coaching in chemistry whenever needed and teaching me how to use an HPLC and LC-MS.

I want to thank all members of the Mayer, Famulok, and Kath-Schorr-groups. You are what made my five years in the lab an amazing time. Thanks for advice, discussions, support, and barbecues, and for making conference attendances so much more enjoyable. Without you, it would not have been the same.

Special thanks – for pretty much everything – to Anna Schüller (now Esser), Julia Siegl, Malte Rosenthal, Anna Weber, and Gerhard Brändle. Thank you for coffee breaks, lunches, ICE Breslau, and sparkling wine Thursdays. Thank you for proof-reading of my thesis and helping me through the valley of tears that seems to be unavoidable when one tries to become a PhD.

Thank you to my ‘career advice panel’, otherwise known as the cocktail ladies that now meet for weekend brunches: Franziska Frey, Anna Esser, Christine Wosnitza, and Martina Bettio. I sincerely hope we stay in contact through many jobs and weddings to come.

Last, but not least, thank you to my family and friends. You supported me and kept me going and motivated. Thank you to my parents, Astrid and Wolfgang Pfeiffer, and my sister, Jasmin Pfeiffer. I know that I can always count on you and that you will support me no matter what. Thank you to Tunç F. Ersoy. Words are not enough, but I am looking forward to spending the rest of my life with you.

Publikationsliste

Pfeiffer F, Gröber C, Blank M, Händler K, Beyer M, Schultze JL, Mayer G: Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Scientific reports* 2018, 8:10950.

Pfeiffer F, Tolle F, Rosenthal M, Brändle G, Ewers J, Mayer G: Identification and characterization of nucleobase-modified aptamers by click-SELEX. *Nat Protoc* 2018, 13:1153-1180.

Fülle L, Steiner N, Funke M, Gondorf F, **Pfeiffer F**, Siegl S, Opitz FV, Haßel S, Erazo AB, Schanz O, Stunden HJ, Blank M, Gröber C, Händler K, Beyer M, Weighardt H, Latz E, Schultze JL, Mayer G, Förster I: RNA aptamers recognizing murine CCL17 inhibit T cell chemotaxis and reduce contact hypersensitivity in vivo. *Molecular Therapy* 2017, 26:95-104.

Pfeiffer F, Rosenthal M, Siegl J, Ewers J, Mayer G: Customised nucleic acid libraries for enhanced aptamer selection and performance. *Curr Opin Biotechnol* 2017, 48:111-118.

Pfeiffer F, Mayer G: Selection and Biosensor Application of Aptamers for Small Molecules. *Front Chem* 2016, 4:25.

Tolle F, Rosenthal M, **Pfeiffer F**, Mayer G: Click Reaction on Solid Phase Enables High Fidelity Synthesis of Nucleobase-Modified DNA. *Bioconjug Chem* 2016, 27:500-503.

Long Y, **Pfeiffer F**, Mayer G, Schroder TD, Ozalp VC, Olsen LF: Selection of Aptamers for Metabolite Sensing and Construction of Optical Nanosensors. *Methods Mol Biol* 2016, 1380:3-19.

Rohrbach F, Schäfer F, Fichte MAH, **Pfeiffer F**, Müller J, Pötzsch B, Heckel A, Mayer G: Aptamer-guided caging for selective masking of protein domains. *Angewandte Chemie - International Edition* 2013, 52:11912-11915.