

Development of a Screening Platform for the Identification of Macrocyclic Splicing Inhibitors

Dissertation

For the achievement of the academic degree of the

Doctor in Natural Sciences

(*doctor rerum naturalium* – Dr. rer. nat.)

Submitted to

Faculty of Chemistry and Chemical Biology

of the TU Dortmund University

By

Jessica Nowacki-Hansen

born Nowacki

from Dortmund, Germany

Dortmund, 29th of June 2023

This work was performed under supervision of Prof. Dr. Dr. h.c. Herbert Waldmann and Dr. Peter 't Hart at the Faculty of Chemistry and Chemical Biology of the TU Dortmund University and at the Chemical Genomics Centre of the Max Planck Society from November 2018 until March 2023.

Dean: Prof. Dr. Stefan M. Kast

First Referee: Prof. Dr. Dr. h.c. Herbert Waldmann

Second Referee: Prof. Dr. Daniel Summerer

Table of contents

Table of contents.....	4
I. Acknowledgements	10
II. Abstract.....	12
III. Zusammenfassung.....	13
IV. List of Abbreviations.....	14
1. Introduction	16
1.1. RNA splicing, alternative splicing and the contributions of SR and hnRNP splicing factors.....	16
1.2. Aberrant splicing events are a hallmark of cancer.....	19
1.3. Overexpression of splicing regulatory proteins leads to cellular transformation.....	19
1.4. Targeting AS as a novel therapeutic approach to combat diseases	23
1.5. Genetically encoded libraries for the identification of macrocyclic inhibitors	25
2. Aim.....	28
3. Material and methods.....	29
3.1 Material, reagents, devices and cells	29
3.1.1. Material and chemical reagents.....	29
3.1.2. Devices	30
3.1.3. Software.....	31
3.1.4. RNA oligonucleotides	31
3.1.5. Medium and self-made buffers	31
3.1.6. Bacterial cells	33
3.2. Methods.....	33
3.2.1 Bacterial cell culture.....	33

Table of contents

3.2.1.1. Preparation of bacterial cultures	33
3.2.1.2. Preparation of bacterial glycerol stocks for long-term storage.....	33
3.2.1.3. Preparation of electrocompetent cells.....	34
3.2.2.4. Preparation of chemically competent cells	34
3.2.2.5. Transformation by electroporation	34
3.2.2.6. Transformation by heat shock	35
3.2.2. DNA Extraction and Sequencing.....	35
3.2.2.1. Plasmid Extraction by Qiagen QiaPrep Mini- or Midiprep Kit	35
3.2.2.2. Sanger sequencing by Microsynth commercial service.....	35
3.2.2.3. Illumina sequencing by Novogene commercial service	35
3.2.3. Cloning	36
3.2.3.1. Agarose-gel electrophoresis	36
3.2.3.2. Polymerase chain reaction (PCR) with Phusion polymerase	36
3.2.3.3. DpnI digest	36
3.2.3.4. PCR purification	36
3.2.3.5. Gel extraction	37
3.2.3.6. Gel purification	37
3.2.3.7. Restriction and ligation	37
3.2.3.8. Sequence and ligation independent cloning (SLIC)	37
3.2.3.9. Library generation with circular polymerase extension cloning (CPEC)	38
3.2.4. Assays, plate assays, screenings and analysis.....	39

Table of contents

3.2.4.1. Translational repression assay procedure	39
3.2.4.2. Flow cytometry	39
3.2.4.3. Fluorescence polarisation assay.....	40
3.2.4.4. Reverse transcription quantitative polymerase chain reaction (RT-qPCR).....	41
3.2.4.5. TRAP evaluation assay	41
3.2.4.6. RBP consensus sequence screening	42
3.2.4.7. Autorepressor presorted RBP consensus sequence screening	42
3.2.4.8. Data analysis: RBP consensus sequence screening	44
3.2.4.9. Data analysis: Autorepressor presorted RBP consensus sequence screening and 10mer RNA consensus sequence library sequencing.....	44
3.2.4.10. SICLOPPS screening analysed by Sanger sequencing	46
3.2.4.11. SICLOPPS screening analysed by Illumina sequencing	47
3.2.4.12. Enriched SICLOPPS screening analysed by Illumina sequencing	48
3.2.4.13. Data analysis: SICLOPPS screening analysed by Sanger sequencing.....	49
3.2.4.14. Data analysis: SICLOPPS screening and enriched SICLOPPS screening analysed by Illumina Sequencing	49
3.2.5. Protein purification	51
3.2.5.1 Protein expression and purification hnRNP A2B1	51
3.2.5.2 Protein expression and purification SRSF1	51
3.2.5.3. SDS-polyacrylamide gel electrophoresis (SDS-PAGE).....	52
3.2.6. Chemical synthesis.....	52
3.2.6.1. Chlorotriyl chloride resin loading.....	52

Table of contents

2.3.6.2. Fmoc solid phase peptide synthesis	52
3.2.6.3. Allyl ester deprotection	53
3.2.6.4. Cyclization	53
3.2.6.5. Cleavage	54
3.2.6.6. Analysis and purification	54
4. Results	55
4.1. A translational reporter assay for the analysis of RNA-binding protein consensus sites	55
4.1.1. Development of the TRAP assay for SRSF1	57
4.1.2. Development of the TRAP assay for hnRNP A2B1	62
4.1.3. Analysis of the effect of secondary structures in the 5'UTR of the reporter mRNA	65
4.1.4. Correlation of translational repression with binding affinity	67
4.1.5. Analysis of translational repression by flow cytometry	71
4.2. Using the TRAP assay to screen for RNA consensus sequences for RBPs	74
4.2.1. RBP consensus sequence screening for SRSF1 and hnRNPA2B1	76
4.2.2. Autorepressor presorted RBP consensus sequence screening	84
4.3. Identification of hexameric peptide inhibitors for the splicing factors SRSF1 and hnRNP A2B1 using SICLOPPS	89
4.3.1. SICLOPPS screening analysed by Sanger sequencing	90
4.3.2. Synthesis of hexameric, cyclic peptides and evaluation of peptides by FP	97
4.3.3. SICLOPPS screening analysed by Illumina sequencing	101

Table of contents

4.3.4. Evaluation of hit peptides by FP	107
4.3.4.1. Evaluation of hit peptides of the SRSF1 SICLOPPS screenings by FP.....	108
4.3.4.2. Evaluation of hit peptides of the hnRNP A2B1 SICLOPPS screenings by FP	111
5. Discussion.....	116
5.1 A translational reporter assay for the analysis of RNA-binding protein consensus sites	116
5.1.1. Development of the TRAP assay for SRSF1	116
5.1.2. Development of the TRAP assay for hnRNP A2B1.....	118
5.1.3. Analysis of the effect of secondary structures in the 5'UTR of the reporter mRNA	119
5.1.4. Correlation of translational repression with binding affinity	122
5.1.5. Analysis of translational repression by flow cytometry	124
5.2 Using the TRAP assay to screen for RNA consensus sequences for RBPs.....	125
5.2.1. RBP consensus sequence screening for SRSF1 and hnRNPA2B1	125
5.2.2. Autorepressor presorted RBP consensus sequence screening	127
5.3 Identification of hexameric peptide inhibitors for the splicing factors SRSF1 and hnRNP A2B1 using SICLOPPS	129
5.3.1. SICLOPPS screening analysed by Sanger sequencing.....	129
5.3.2. Synthesis of hexameric, cyclic peptides and evaluation of peptides by FP	131
5.3.3. SICLOPPS screening analysed by Illumina sequencing	132
5.3.4. Evaluation of hit peptides by FP	133
6. Summary and Outlook.....	134
7. Literature.....	136

Table of contents

8. Appendix	146
8.1. Tables and Figures	146
8.1.1. Table of figures.....	146
8.1.2. Table of Tables.....	154
8.2. Supplementary figures	155
8.2.1. Supplementary figures: TRAP figures.....	155
8.2.2. Supplementary figures: Secondary structure analysis	160
8.2.3. Supplementary figures: Flow cytometry analysis	169
8.2.4. Supplementary figures: RBP consensus screening	177
8.2.5. Supplementary figures: SICLOPPS screening.....	179
8.3. Supplementary tables	179
8.4. Supplementary material	184
8.4.1. Plasmid sequences	184
8.4.2. SICLOPPS peptides structures, HRMS analyses and yields.	193
8.4.2. Peptide Purity HPLC Chromatograms	194

I. Acknowledgements

First of all, I want to thank Prof. Dr. Dr. h.c. Herbert Waldmann for the opportunity to do my PhD in his department and for being the first examiner of my doctoral thesis.

I thank Prof. Dr. Daniel Summerer for taking over the role as second examiner as well as the capacities at the FACS in the Summerer lab.

My biggest thanks go to my supervisor Dr. Peter 't Hart, who gave me this very cool project and the trust to work on it. I appreciate the empowerment to learn new techniques and level up my skills. I also thank him for all his patience, work, effort and the good guidance. I consider myself to be very lucky to have worked in his group.

I also thank the whole AG Hart for being a nice group offering a good working environment, support, laughter and room for discussion. I thank Stefan Schmeing, Joseph Openy and especially Gulshan Amrahova for being good collaboration partners in my projects. Special thanks also go to Jen-Yao Chang and Dr. Sunit Pal for all the scientific discussions and career-related advice.

I am thanking Damian Schiller, Brinja Kosel and Dr. Tzu-Chen Lin for (late-night) advice on the FACS and their contributions to my projects.

Thanks also go to Dr. Jan-Erik Hoffmann for all the expertise and input concerning the work with my proteins.

Big thanks also go to my Bachelor student Luisa Kurzweg, my master student Mateo Malenica and my intern student Yevfaliia Kopytsia for their contributions to this work.

I am thanking Dr. Peter 't Hart, Lydia Borgelt, Dr. Rachel O'Dea and Jen-Yao Chang for their time and effort on proof-reading this thesis.

I want to thank my former supervisor Dr. Elena Reckzeh who is a role model for me and by that motivated me to become a scientist.

I thank a few people that started as my work colleagues and turned into friends and frequent party gang: Rachel, Kai, Jen-Yao, Jimin, Lydia, Pascal, Christian, Mirko, Joseph, Adrian, Alejandro, Stefan, Katrin, Gulshan, Phillipp, Adrian, Siska, Evie, Zhou, Yang & Cora. Without you work would only have been half as fulfilling.

I. Acknowledgements

I especially appreciate to have met Stefan, Elisabeth, Aylin & Georg in the rooms of department 4 who quickly became my friends, mentors and regular discussion partners for science and non-science related topics.

Last but not least I want to thank my close family, my friends and most important my husband for the time, money and patience they have invested in me. Without you, I would not have undergone the PhD journey.

II. Abstract

Protein-RNA interactions are substance of research of the last decades and are yet to be fully explored. Interactions of RNA-binding proteins (RBPs) with RNA are found in a plethora of fundamental biological processes such as transcription, splicing, capping, polyadenylation, and translation. Alternative splicing (AS) is the major contributor to the high protein diversity in humans, with only ~30,000 genes present, and is regulated by RBPs such as spliceosomal subunits and alternative splicing factors. Aberrant function and expression of these factors disorganise splicing patterns, which can cause neurodegenerative diseases and cancer. Targeting alternative splicing factors to fight cancer has emerged as a novel promising therapeutic approach.

For this purpose, two alternative splicing factors, SRSF1 and hnRNP A2B1, were first explored with the translational repression assay procedure (TRAP) to monitor the proteins' interaction with RNA. The TRAP assay results delivered relative binding affinities, given in repression ratios, and were found to correlate with absolute binding affinities measured with the fluorescence polarisation (FP) assay.

Binding motifs for RBPs are often found by cross-linking or pull-down approaches that are cost-intensive and laborious. The work described here presents a novel approach using the TRAP assay in a high throughput format to screen for 10mer RNA consensus sequence for SRSF1 and hnRNP A2B1.

The TRAP assay, in combination with the plasmid-encoded peptide library SICLOPPS, was used to screen for hexameric cyclic peptides as inhibitors for the RBPs. In a first approach, Sanger sequencing was used to sequence analysis and hit peptides were evaluated using the TRAP assay. Three hit peptides were synthesized chemically and evaluated by FP. In a second, optimised screening approach, Illumina sequencing was used for hit analysis, and the top six peptides were synthesised and evaluated by FP. Two promising candidate peptides for hnRNP A2B1 were found that are currently being analysed and verified as true inhibitors.

III. Zusammenfassung

Protein-RNA-Interaktionen sind Gegenstand der Forschung der letzten Jahrzehnte und müssen nach wie vor untersucht werden. Interaktionen von RNA-bindenden Proteinen (RBPs) mit RNA finden sich in einer Vielzahl von Prozessen wie Transkription, Spleißen, Capping, Polyadenylierung und Translation. Mit nur etwa 30.000 Genen im Menschen trägt Alternatives Spleißen (AS) zur hohen Proteindiversität bei und wird durch RBPs, wie die Untereinheiten des Spleißosoms und alternative Spleißfaktoren, reguliert. Eine anormale Funktion und Expression dieser Faktoren desorganisiert Spleißmuster, die neurodegenerative Erkrankungen und Krebs verursachen können. Alternative Spleißfaktoren als Zielscheibe für Substanzen zur Krebsbekämpfung kann als neuer therapeutischer Ansatz angesehen werden.

Zu diesem Zweck wurden erstmalig zwei alternative Spleißfaktoren, die RBPs SRSF1 und hnRNP A2B1, mit „Translational Repression Assay Procedure“ (TRAP) untersucht, um die Interaktion der Proteine mit RNA zu quantifizieren. Die TRAP-Assay-Ergebnisse lieferten relative Bindungsaffinitäten, die in Repressionsverhältnissen angegeben werden und gut mit der absoluten Bindungsaffinität korrelierten, die mittels Fluoreszenzpolarisationsassays (FP) bestimmt wurden.

Bindungsmotive für RBPs werden oft durch Cross-linking- oder Pulldown-Ansätze gefunden, die kostenintensiv und aufwändig sind. Die hier beschriebene Arbeit präsentiert einen neuartigen Ansatz zur Verwendung des TRAP-Assays in einem Hochdurchsatz-Format zum Screening von 10mer-RNA-Konsensusmotiven für SRSF1 und hnRNP A2B1.

Letzlich wurde der TRAP-Assay in Kombination mit der Plasmid-kodierten Peptidbibliothek SICLOPPS verwendet, um nach hexameren, zyklischen Peptiden als Inhibitoren für die RBPs zu suchen. In einem ersten Ansatz wurde die Sanger-Sequenzierung zur Sequenzanalyse verwendet und Hit-Peptide wurden mithilfe des TRAP-Assays bewertet. Drei Hit-Peptide wurden chemisch synthetisiert und mittels FP evaluiert. In einem zweiten, optimierten Screening-Ansatz wurde die Illumina-Sequenzierung zur Hit-analyse verwendet und die besten sechs Hit-Peptide wurden synthetisiert und mit FP bewertet. Es wurden zwei vielversprechende Kandidatenpeptide für hnRNP A2B1 gefunden, die derzeit weiter analysiert und als echte Inhibitoren verifiziert werden.

IV. List of Abbreviations

Abbreviation	Meaning
2-CTC	2-Chlorotrityl chloride
3'SS	3' splice site
5'SS	5' splice site
AC	Autorepressor corrected
AmpR	Ampicillin resistance gene
AS	Alternative splicing
ASO	Anti-sense oligonucleotide
Boc	tert-butyloxycarbonyl
BP	Branch point
CamR	Chloramphenicol resistance gene
CBD	Chitin-binding domain
Conc.	Concentration
CS	Constitutive splicing
DNA	Desoxyribonucleic acid
<i>E. coli</i>	<i>Escherichia coli</i>
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
Ex/Em	Excitation/Emission
FACS	Fluorescence assisted cell sorting
FDA	Food Drug Administration
Fmoc	Fluoromethoxycarbonyl
FP	Fluorescence polarisation
hnRNP	Heteronuclear ribonuclear particle
hnRNP A2B1	Heteronuclear ribonuclear particle A2B1
HTS	High throughput screening
IPTG	Isopropyl- β -D-thiogalactopyranoside
ISE	Intronic splicing enhancer
ISS	Intronic splicing silencer
ITC	Isothermal titration calorimetry
LB	Luria Broth
mRNA	Messenger ribonucleic acid
nt	Nucleotide
<i>Npu</i>	<i>Nostoc punctiforme</i>
OD	Optical density
ORI	Origin of replication

IV. List of Abbreviations

PCR	Polymerase chain reaction
pre-mRNA	Premature messenger ribonucleic acid
PrLD	Prion-like domain
PPI	Protein-Protein-Interactions
PRI	Protein-RNA-Interactions
RNA	Ribonucleic acid
RBP	RNA-binding protein
RGG box	Arginine-glycine-glycine repeat
RRM	RNA recognition motif
S/D	Shine-Dalgarno
sfGFP	Super folder green fluorescent protein
SICLOPPS	Slit-intein circular ligation of proteins and peptides
SMA	Spinal muscular atrophy
snRNP	small nuclear ribonucleoprotein
SOB	Super Optimal Broth
SOC	Super Optimal Broth with catabolite repression
SR	Serine/arginine-rich
SRSF1	Serine/arginine-rich splicing factor 1
<i>Ssp</i>	<i>Synechocystis species</i>
tagBFP	tag blue fluorescent protein
TB	Terrific Broth
tBu	tert-Butyl
TRAP	Translational repression assays procedure

1. Introduction

1.1. RNA splicing, alternative splicing and the contributions of SR and hnRNP splicing factors

The central dogma of molecular biology, firstly postulated by Francis Crick in 1958, states the flow of genetic information: DNA is transcribed into RNA, which is then translated into protein.^{1,2} The process in which DNA is transcribed into RNA comprises several steps that are essential for the regulation of RNA maturation and protein diversity. These independent events are 5' capping, splicing of introns and 3' cleavage/polyadenylation, which occur cotranscriptionally.³ During splicing, the non-coding introns of the pre-mRNA are excised out, and the coding exons are ligated together, leading to the mature RNA constructs.

Splicing is a conserved and tightly regulated process between transcription and translation that is catalysed by the spliceosome, an RNA-protein multicomponent complex consisting of five small nuclear ribonucleoproteins (snRNPs), the U-complexes, U1, U2, U4, U5 and U6.⁴ It undergoes an ordered process of assembly and rearrangement prior to performing two sequential transesterification reactions.⁵ The spliceosome recognizes consensus sequences within the pre-mRNA transcript and by that initiates the splicing mechanisms. First, the adenosine residue at the pre-mRNA, also called the branch point (BP), attacks the 5' splice site (5'SS), leading to a free 5' exon and a lariat intermediate. Next, the free 5' exon attacks the 3' splice site (3'SS) of the neighbouring exon yielding to a lariat intron and two joined exons, the spliced mRNA (see Figure 1).^{2,4}

1. Introduction

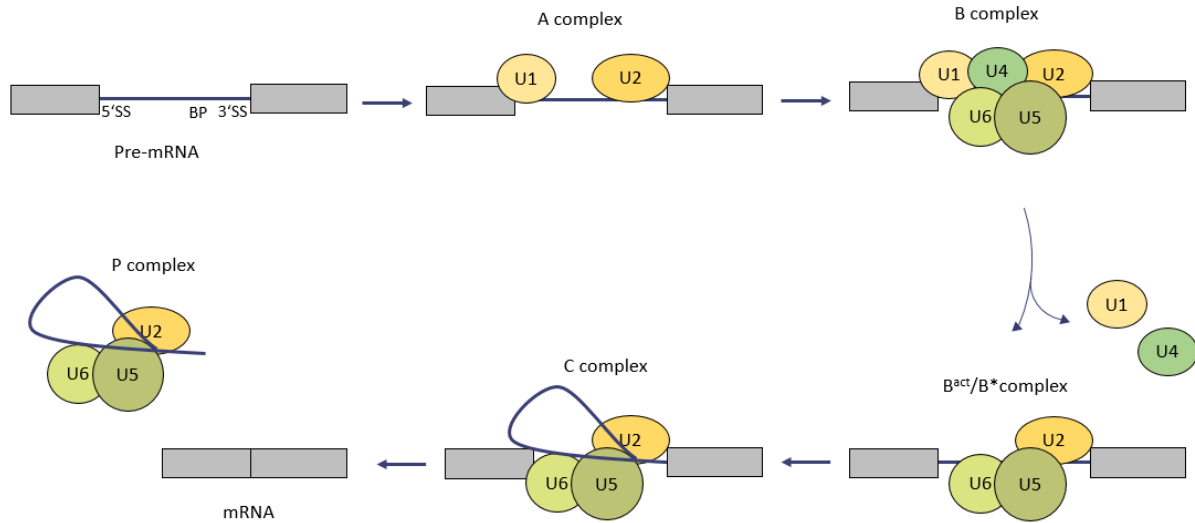


Figure 1: Splicing mechanism performed by the spliceosome. The U1 snRNP binds to the 5'SS, while U2 binds to the BP to form the A complex. U4/U6.U5 (tri-snRNP) joins the A complex, and together they form the B complex. The spliceosome is activated upon the removal of U1 and U4. Once the first splicing reaction is completed, formation of the C complex occurs. After the second splicing reaction, the spliced mRNA is released from the P complex. Adapted from Gehring and Roignant, 2014. ⁴

The general concept of intron removal and exon junction is also known as constitutive splicing (CS), which is distinguished from alternative splicing (AS) mechanisms. Alternative splicing involves a subset of RNA binding proteins (RBPs) to function as further splicing regulatory proteins and can be seen as the leading contributor to the large protein diversity in eukaryotic organisms. ^{6,7} The ability to generate multiple distinct isoforms from a single gene is a key function of the splicing process. ⁶ In fact, genomic studies estimate that ~95 % of genes in mammals undergo alternative splicing mechanisms. ⁸ There are a number of RBPs that control AS by being guided by the sequence composition of a pre-mRNA transcript that determines if a section (intron or exon) is included or excluded and which 5'SS and/or 3'SS is chosen. ⁷ The following figure shows the most common AS mechanisms.

1. Introduction

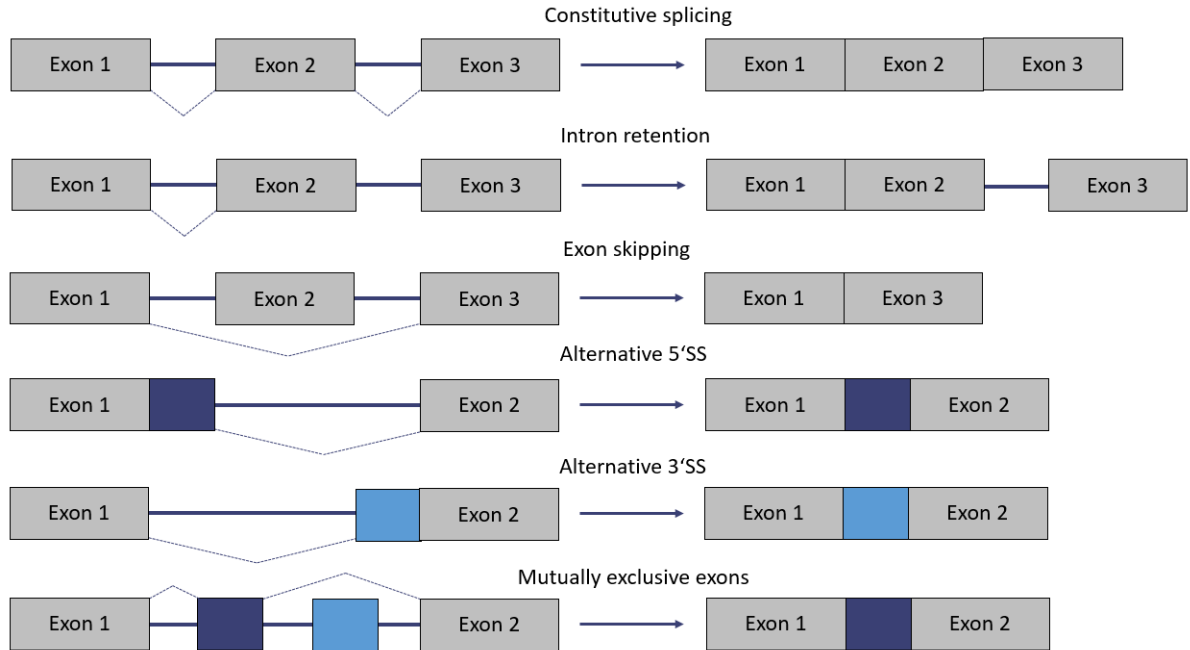


Figure 2: Alternative splicing mechanisms. Blue boxes represent exons, while straight lines indicate introns. Angled lines show the junction path of pre-mRNA to form the spliced mRNA. Adapted from Ren et al., 2019.^{2,9}

Dependent on the transcript to be spliced, a different subset of regulatory proteins is involved in its splicing. These auxiliary splicing factors, called trans-acting factors, have a direct influence on the splice site and enhance or inhibit a splicing event. In contrast to that are the cis-acting sequences whose sequence composition has an influence on the splicing event, known as exonic or intronic splicing enhancer or silencer sequences (ESE/ESS or ISE/ISS).

8

Serine/arginine-rich splicing factors (SR proteins) and the heterogeneous nuclear ribonucleoproteins (hnRNPs) play a major role in alternative splicing regulation. While SR proteins rather promote splicing through binding to ESE/ESS, the hnRNPs can repress splicing by binding ISE/ISS (see figure 3).^{10,11} There are examples where the inhibiting effect on a specific splicing event by representatives of the hnRNP family is found to be counteracted by SR proteins.^{9,12,13}

1. Introduction

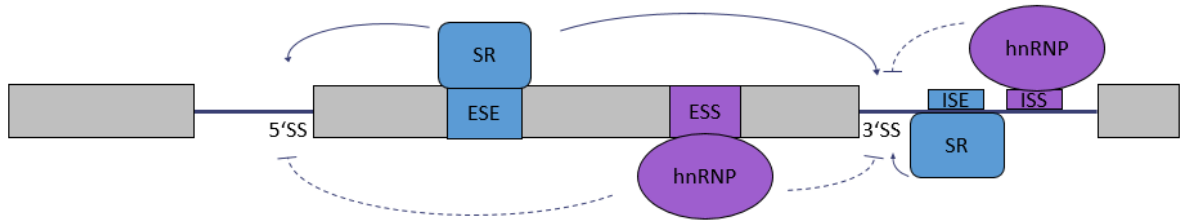


Figure 3: Role of cis and trans-acting factors SR and hnRNP in alternative splicing of pre-mRNAs. Binding of SR proteins to ESE and ISE (blue boxes) promotes exon inclusion (light grey boxes), while hnRNP binding to ESS and ISS elements (purple boxes) promotes exon skipping. ^{8,10}

As AS allows cells to generate protein isoforms with differing or even opposing functions, changes in splicing events such as sequence polymorphisms or deletions or aberrant activity of splicing factors can have severe causes for the whole organism. ^{2,7}

1.2. Aberrant splicing events are a hallmark of cancer

Protein isoforms generated by AS contribute to the regulation of a plethora of processes and metabolic pathways, including cell development, differentiation, cell cycle control and apoptosis. Apoptosis is a key process that is required for the normal development and maintenance of tissue homeostasis and is also tightly regulated by AS. ^{11,14} Dysregulation of apoptosis can lead to diseases like cancer, autoimmune or neurodegenerative diseases. AS produces protein isoforms that have distinct apoptotic regulatory activities. ¹¹ Aberrant expression and activity of regulatory splicing proteins can lead to changes in splicing patterns, causing an imbalance of pro-apoptotic and anti-proliferative splicing isoforms. ^{7,15} Therefore, the dysregulation of AS leading to aberrant mRNA isoforms coding for a mutated protein with altered function can cause neoplastic cellular transformation, cancer development, progression, and metastasis. ^{10,11} In general, several studies found that deregulation of splicing is a major contributor to carcinogenesis and is therefore considered a novel hallmark of cancer. ^{10,15}

1.3. Overexpression of splicing regulatory proteins leads to cellular transformation

The serine/arginine-rich splicing factor 1 (SRSF1) is the first of the twelve-membered SR protein family and is overexpressed in breast and lung cancer and other cancer types. ^{10,16} The oncoprotein is characterized by two RNA recognition motifs (RRMs), an N-terminal canonical RRM (RRM1), a pseudo RRM (RRM2), and a C-terminal domain that consists of

1. Introduction

arginine and serine dipeptide repeats, the RS-domain (see Figure 5).^{17,18} Both RRM1 and RRM2 of the protein are necessary for optimal RNA binding and splicing, while RRM2 dictates substrate specificity *in vivo*.^{18–20} The RRM1 mediates the interaction with the U1-70K component of the U1 complex, whereas the RS domain has a regulatory role.^{18,21} Moreover, the RS domain is necessary for nuclear-cytosolic shuttling and sub-nuclear localization and is subject to post-translational modification, such as serine phosphorylation.^{18,20} Although the RS domain is required for CS activity, deletion studies of SRSF1 lacking the RS domain showed that its AS activity is retained.²²

SRSF1 is a key player in CS and AS and, moreover, plays a role in nonsense-mediated mRNA decay, mRNA transport and translation.^{16,17} It regulates the AS of proto-oncogene *RON*, tumour-suppressor *BIN1* and several other genes that may exhibit oncogenic function.^{16,18} Table 1 shows examples of splicing targets of SRSF1 and the property of the SRSF1-induced isoform.

Table 1: Examples of target genes spliced by SRSF1, their splicing change and induced property. Adapted from Das and Krainer, 2014¹⁸

Target Gene	Function	SRSF1-induced splicing change	Property of the SRSF1-induced isoform	Reference
<i>BCL2L1</i>	Apoptosis regulator	Enhanced inclusion of full-length exon 2	Anti-apoptotic	Leu <i>et al.</i> , 2012
<i>BCL2L11</i>	Apoptosis regulator	Increased inclusion of novel alternative 3' exon	Anti-apoptotic	Anczuków <i>et al.</i> , 2012
<i>BIN1</i>	Apoptosis regulator	Increased 12A inclusion	Anti-apoptotic	Karni <i>et al.</i> , 2007; Anczuków <i>et al.</i> , 2012
<i>CASP2</i>	Apoptosis regulator	Skipping of a novel 61 bp exon	Pro-apoptotic	Jiang <i>et al.</i> , 1998
<i>CASP9</i>	Effector of Apoptosis	Increased inclusion of cassette exons 3-6	Pro-apoptotic	Masiello <i>et al.</i> , 2006
<i>MKNK2</i>	Effector in MAP kinase signaling pathway	Increased inclusion of mutually exclusive exon 13b	Activates p38a-MAPK; Pro-oncogenic	Karni <i>et al.</i> , 2007; Maimon <i>et al.</i> , 2014

1. Introduction

SRSF1 regulates the alternative splicing of many more target genes. Upregulation of SRSF1 generates protein isoforms that stimulate cellular translation and proliferation while also promoting the expression of anti-apoptotic isoforms, both leading to cellular transformation (see Figure 4).

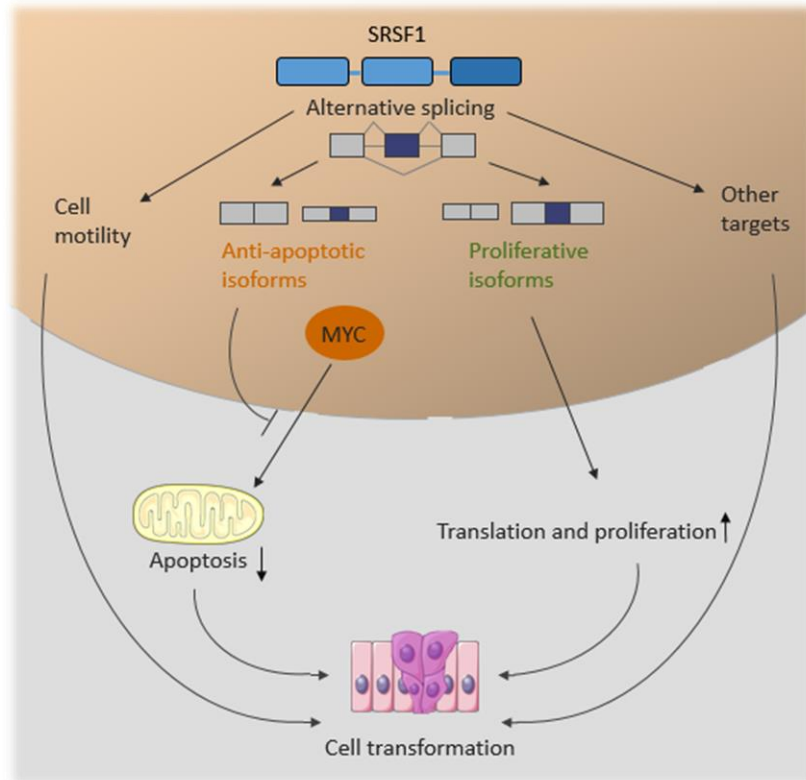


Figure 4: Role of SRSF1 in cellular transformation. Overexpression of SRSF1 generates proliferative and anti-apoptotic isoforms that are unable to interact with pro-apoptotic factors such as MYC. Adapted from Anczuków et al., 2012. ¹⁶

A similar role can be observed for the heterogeneous ribonucleoprotein A2B1 (hnRNP A2B1), which is part of the hnRNP family with 20 major RBPs and was also found to be overexpressed in breast, lung and other cancers. ^{10,23,24} The protein has two N-terminal RRMs, and a C-terminal glycine-rich low-complexity domain, also called the prion-like domain (PrLD). It also has an RGG box with the ability to mediate RNA-binding activity and a proline-tyrosine nuclear localization signal (PY-NLS) which is important for nuclear import (see Figure 5). ^{25,26} Isothermal titration calorimetry (ITC) experiments showed that the two RRMs without the C-terminal domain are able to bind consensus-motif-containing target RNAs with high affinity and that RRM1 and RRM2 can specifically bind AGG and UAG motifs, respectively. Crystallography studies demonstrated that both RRMs are able to bind two strands of RNA in

1. Introduction

an antiparallel orientation.^{26,27} The *hnRNPA2B1* gene codes for two isoforms, both generated by alternative splicing, A2 and B1, that differ in 12 additional N-terminal amino acids in the B1 version.²⁷ Besides its function in AS, similar to SRSF1, it has further use in mRNA stability, export and translation. Furthermore, it plays a role in chromatin remodelling and genome stability.^{10,28}

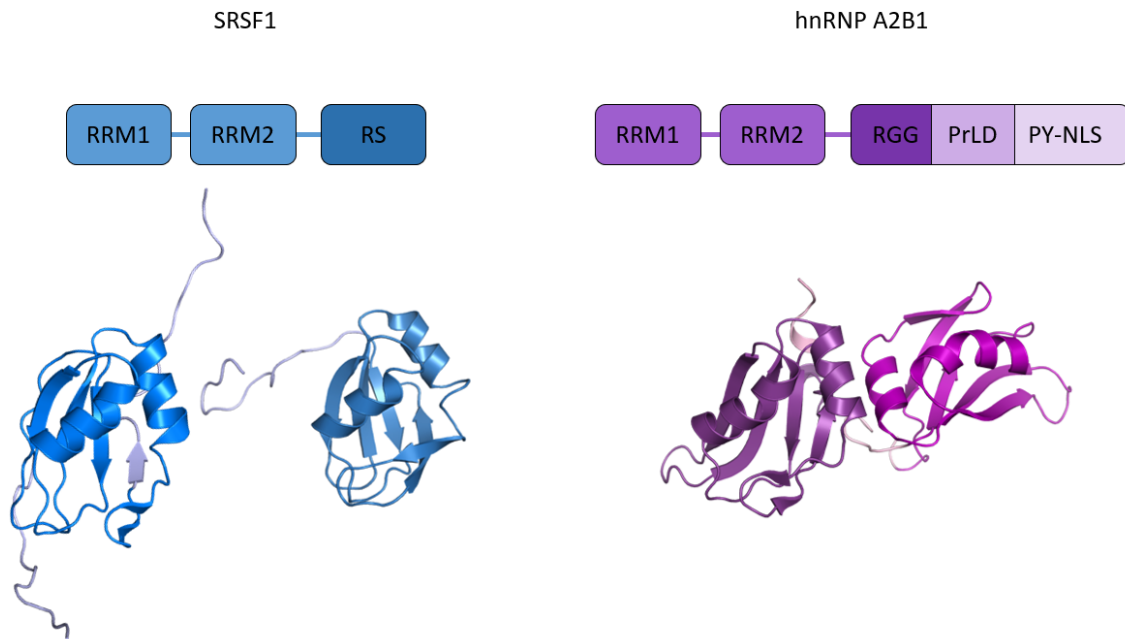


Figure 5: Splicing factors SRSF1 (blue tones) and hnRNP A2B1 (purple tones). Schematic representation of the domains (top, adapted from Anczuków *et al.*, 2012 and Wu *et al.*, 2018^{16,27}) and crystal or NMR structures of RRM1 and RRM2 (bottom). PDB: SRSF1 RRM1: 1X4A, RRM2: 2M8D; hnRNP A2B1 RRM1 and 2: 5HO4.

HnRNP A2B1 modulates the alternative splicing of the tumour suppressors *BIN1*, the anti-apoptotic proteins c-FLIP and caspase-9B and the *RON* proto-oncogene, among others. Just like SRSF1, hnRNP A2B1 overexpression enhances exon 12a inclusion of the *BIN1* gene generating a BIN1 isoform with anti-apoptotic function. In contrast to SRSF1, A2B1 overexpression enhances skipping of exon 3-6, generating the anti-apoptotic isoform caspase-9B.²³

SRSF1 and hnRNP A2B1 are overexpressed in different cancer types, and their knockdown can change or reverse alternative splicing events.^{23,29} HnRNP A2B1 knockdown has been shown to suppress cell proliferation, migration and invasion in HeLa and CaSki cells. Another effect was a higher sensitivity to the anticancer-drugs irinotecan and lobaplatin.³⁰

1. Introduction

In the case of SRSF1, its knockdown decreases the proliferation rate of small cell lung cancer cells. Knockdown also seems to activate caspase-3, a mediator of apoptosis³¹, with a similar effect observed for cisplatin. The combination of SRSF1 knockdown and cisplatin treatment leads to an increase in the effect.³² In myeloma cells, an SRSF1 knockdown promoted apoptosis.³³ Therefore, both splicing regulators could be attractive targets for novel anti-cancer therapeutics.

1.4. Targeting AS as a novel therapeutic approach to combat diseases

The importance of aberrant splicing in cancer formation indicates the need for novel splicing-targeted therapeutics. There are a number of strategies to target deregulated splicing factors and aberrant splicing isoforms, such as small-molecule inhibitors for the core spliceosome (spliceostatsins) derived from bacteria, splicing regulator inhibitors that target kinases and anti-sense oligonucleotides (ASOs) against oncogenic mRNA.¹⁰

One successful story of drug development to target splicing is the ASO drug nusinersen (commercial name: Spinraza), discovered by Krainer and co-workers.^{34,35} Nusinersen is used to treat spinal muscular atrophy (SMA) by targeting the alternative splicing of the *SMN* gene of the survival motor neuron (SMN) protein.^{34,36} SMA is a neurodegenerative disease in children and adults that is caused by the progressive loss of the α -motor neurons in the spinal cord.^{34,35} The reason for the neuronal loss is the reduced expression of the SMN protein. The SMN protein is transcribed from two genes, the *SMN1* and *SMN2* genes. In the disease phenotype, the *SMN1* gene is homozygously deleted or mutated. The presence of *SMN2* allows viability of the patient but is not sufficient to fully compensate for the loss of *SMN1*.³⁴ In the *SMN2* gene, exon 7 is predominantly skipped leading to the expression of a truncated protein version.³⁷ By binding to the mutated RNA region, nusinersen inhibits splicing factors, which leads to increased exon 7 inclusion and restoration of the full-length SMN protein.^{36,38} Nusinersen was the first drug to treat SMA and was approved in the USA by the Food and Drug Administration (FDA) only in 2016.³⁵ The use of ASOs to target diseases, however, suffers from being administrated by injection or via inhalation, oligonucleotide degradation and low cellular delivery.^{34,39}

In 2020 the first small molecule drug risdiplam (Evrysdi) for SMA that targets an RNA structure was approved.⁴⁰ Its mechanism of action is, like nusinersen, to restore the inclusion of exon 7 in the *SMN2* gene. Acting as a molecular glue, the drug stabilizes the transient double-strand RNA formed by the 5'SS of SMN2 exon 7 and the U1 complex.^{40,41} By that, it

1. Introduction

strengthens the weak 5'SS of SMN2 exon 7. The increased binding affinity of the U1 complex to the 5'SS compensates for the sequence mismatch and leads to efficient splicing. Unlike nusinersen, risdiplam is orally bioavailable and does not have to be administered by injection.⁴⁰ Targeting mRNA structure by small molecules is a strategy to modulate biological processes such as splicing, translation and miRNA biogenesis.⁴²

The natural products FR901464, Herboxidiene and pladienolide B, among others, show anticancer activity by targeting cellular splicing mechanisms. All three named compounds target the SF3b1 subunit of the spliceosome.⁴³ Since splicing is essential to the cell, it is not surprising that splicing inhibitors impair cellular function significantly. A few examples of drugs that target the SF3b1 subunit were dropped in clinical phase trials due to side effects related to toxicity. Off-target effects and low specificity of small molecules limit their use in the clinic.⁴⁴ Targeting the spliceosome directly seems to be a high-risk approach since it inhibits an essential cellular process. Instead, using inhibitors to target AS regulators that are overexpressed in cancer to lower their levels to endogenous levels and therefore revert a disease-phenotype could be a more selective strategy.

In 2023, Hu *et al.* proposed a small molecule pseudourea-derivative (XI-011) that downregulates mRNA levels of the oncogene MDMX via disruption of the recruitment of hnRNPA2B1 in gastric cancer cells by that reducing *MDMX* transcription and restoring p53 activity. The molecule showed broad antitumor activity against various gastric cancer cells, which harbour overexpression of hnRNP A2B1. However, XI-011 did not affect hnRNP A2B1 expression, and when tested for binding using biolayer interferometry, it interacted with all individual domains of hnRNP A2B1, including its unstructured PrLD making selective interaction questionable. Furthermore, cellular activity is observed at lower concentrations (<1 μM) than its affinity for hnRNP A2B1 (4.6 μM), further highlighting selectivity issues.⁴⁵

Using small molecules as AS inhibitors seem to be a promising strategy. However, small molecules have their limitations when considering their size (<1000 Da) and small overall surface area that they can occupy (300-1000 Å^2), which makes them ineffective in targeting larger surface areas as found in, e.g. protein-protein (PPI) or protein-RNA interactions (PRI). Larger interaction areas could be addressed by macrocyclic peptides which are significantly bigger in size, usually of a molecular weight of 500-5000 Da. Peptides are an attractive class of therapeutics covering the otherwise “undruggable” space between small molecules and therapeutic proteins to target PPIs. Peptide drugs have shown high specificity, good efficacy and good safety in drug development. Linear peptides, however, are typically not

1. Introduction

metabolically stable and have low cell permeability.^{44,46-49} Cyclization of peptides is an approach to overcome these issues and make linear peptides less susceptible to proteases while increasing cell permeability.⁴⁴ Head-to-tail, backbone-to-side chain or side chain-to-side chain cyclization, as well as stapling approaches, offer ways to generate metabolically more stable and more cell-permeable peptides. Moreover, peptide cyclization allows to mimic and stabilize peptide structures such as α -helices or β -sheets.⁴⁹ Peptides potentially can also interact with RNA motifs which can make peptides valuable candidates for inhibiting PRIs. The absence of structural data to explore the chemical space between protein and RNA can make the rational design difficult to impossible. Instead, screening of genetically encoded cyclic peptide libraries offers an attractive option.⁵⁰

1.5. Genetically encoded libraries for the identification of macrocyclic inhibitors

Genetically encoded libraries have several advantages over synthesized molecules used in high-throughput screenings (HTS). They are readily accessible, deliver a high number of molecules to be used in screening, allow a straightforward hit deconvolution and are lower in cost compared to the HTS counterparts. The use of genetically encoded libraries in a model organism, like e.g. *Escherichia coli* (*E. coli*), allows easy handling and control of a screening campaign.

Variants of such genetically encoded libraries include phage and mRNA display. Both methods generate cyclic peptides that can be used in affinity-based screening.⁵¹ In phage display, a library of phage-vectors is used to genetically manipulate the phage. Once the host, e.g. *E. coli*, is infected, its cell machinery is reprogrammed to express a peptide or protein fragment fused to the phage coat protein. The fusion-coat protein is incorporated into phage particles, the virions, released from the cell and displayed at the phage surface.^{52,53} In mRNA display, a DNA library is used to produce peptides in a cell-free *in vitro* translation system. Once the ribosome reaches the RNA/DNA junction, a puromycin molecule, attached to the DNA linker, enters the A site of the ribosome and stalls the translation. This leads to an expressed peptide chain that is still linked to its mRNA. Reverse-transcription of the mRNA gives a cDNA strand that can be sequenced. Cyclization techniques can be applied post-translationally.⁵⁴ Both display methods directly link the phenotype to the genotype. With up to $\sim 10^9$ members for phage display and $\sim 10^{14}$ members for mRNA display, the libraries produce a huge count of cyclic peptides.^{51,55} However, both methods require large laborious effort,

1. Introduction

including several selection cycles.^{51,55} Peptides are directly fused to either the phage coat protein or in mRNA display to the puromycin-RNA attachment, which could limit binding sites to the target and could lead to non-selective target binding.⁵⁵ mRNA display is also not useful to address proteins that strictly rely on being part of protein complexes and else do not show biological functions. Moreover, mRNA display cannot be used for membrane-bound proteins as the expression of such remains difficult for *in vitro* translation systems.^{56,57} Ultimately, there could be difficulties using mRNA display for the identification of peptide inhibitors for RBPs, possibly due to their potential to bind RNAs rather than peptides.

Another approach to find cyclic peptides as inhibitors is the biotechnology method split-intein circular ligation of peptides and proteins (SICLOPPS) invented by Benkovic and colleagues.⁵⁸ The technique uses the properties of genetically engineered intein, a protein with the ability to self-excision that splices its N- and C-terminus and leaves a cyclic peptide (extein) behind.

51

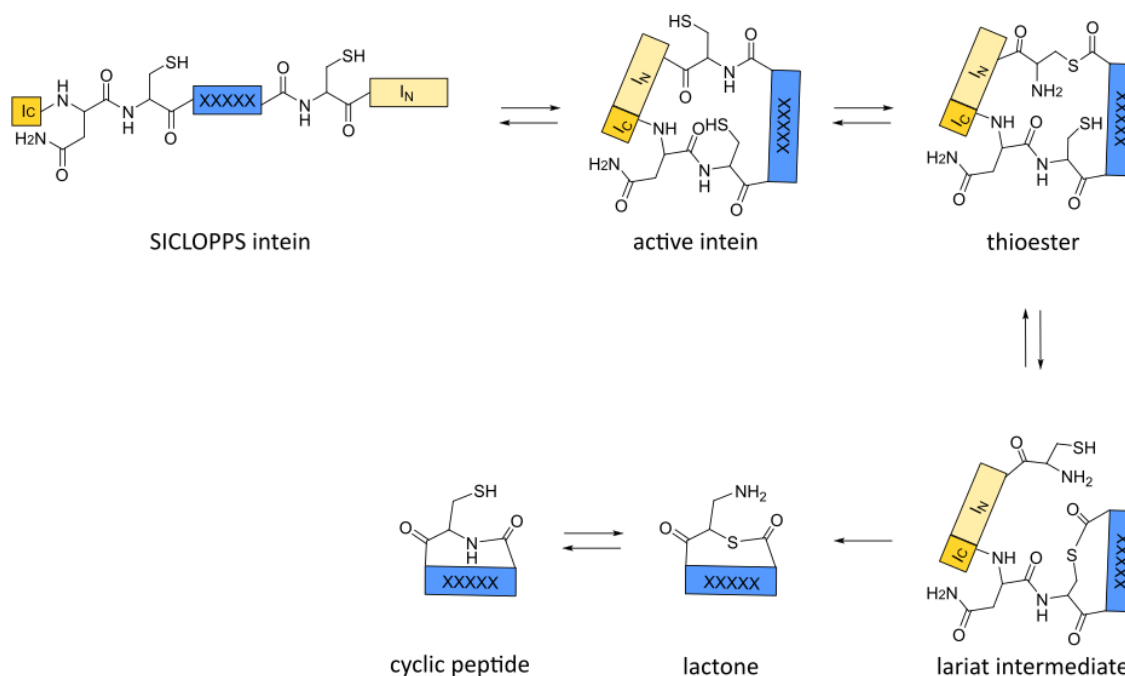


Figure 6: SICLOPPS mechanism (schematic). The two intein domains (I_N and I_C) of the SICLOPPS intein fusion protein fold and generate an active intein. An N-to-S acyl shift at the I_N -junction forms a thioester. In a transesterification reaction with a side chain nucleophile (here: cysteine) at the I_C -junction, a lariat intermediate is produced. An asparagine side chain (I_C -junction) promotes lactone formation, which then generates the thermodynamically favoured lactam product via an S-to-N acyl shift (*in vivo*). The hexameric cyclic peptide contains five randomized amino acids (each indicated by X). Adapted from Tavassoli and Benkovic 2007^{58,59}, created with Biorender.com.

1. Introduction

SICLOPPS libraries are encoded by plasmids that can be readily transformed into *E. coli* cells. Theoretically, any peptide size can be chosen, however, the number of cyclic peptides will be limited by the transformation efficiency of the cells and will be lower than the number of transformants ($\sim 10^9$). A commonly used library of 6 amino acids will give a theoretical diversity of 3.2×10^6 peptides. While the first amino acid side chain is fixed to contain a nucleophile (serine or cysteine) for the splicing to occur, the other 5 amino acids are random. The SICLOPPS technique can be combined with any cell-based assay to screen for inhibitory members in the library that show a desired phenotype. SICLOPPS has no need for laborious *in vitro* affinity-based approaches to screen for hits as peptide translation and cyclization is performed by the cell. In literature, the system has been used to identify a variety of PPI inhibitors. The Tavassoli group used SICLOPPS together with a bacterial reverse-two hybrid system and identified *cyclo*-CLLFVY to be an inhibitor of the hypoxia inducible factor-1 (HIF-1). The peptide successfully inhibits the PPI between HIF-1 α and HIF-1 β and hence transcription activity.^{51,60} Furthermore, the Keiler group used fluorescent activated cell sorting (FACS) together with a fluorescent reporter gene system in *E. coli* to identify an antibacterial inhibitor, *cyclo*-SGWYGRRH, for ClpXP protease.⁶¹ All these findings suggest that SICLOPPS is a promising strategy to screen for macrocyclic peptide inhibitors.

2. Aim

2. Aim

In this work, a PRI assay was developed that demonstrates the interactions of RBPs with RNA and can be used in host organism *E. coli*. The assay was applied to two splicing factors, SRSF1 and hnRNP A2B1, that modulate the alternative splicing of several target genes and are found to be overexpressed in different cancer types. The assay was used to display the interactions with RNA binding motifs that are biologically relevant in splicing mechanisms.

In the second part of the thesis, the assay was used in a high throughput manner to act as a screening platform that allows the identification of short RNA consensus sequences for RNA-binding proteins. For that, a library of RNA plasmids was cloned and screened against the two splicing factors that served as model proteins.

Ultimately, the assay system was used in combination with the biotechnology method SICLOPPS to perform an intracellular screening for the identification of macrocyclic peptide inhibitors for the two splicing factors.

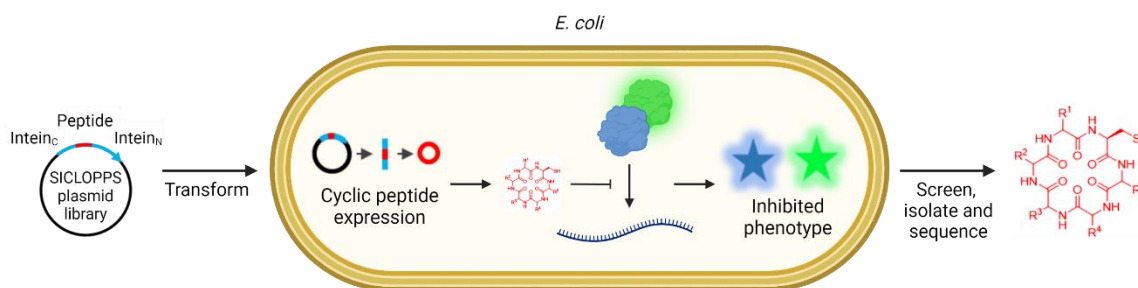


Figure 7: Concept of the SICLOPPS screening using TRAP. The SICLOPPS plasmids are transformed into *E. coli* cells carrying an assays system that allows intracellular screening with FACS. Isolated cells can be sequenced to identify the macrocyclic peptide that caused a phenotypic change. Adapted from Tavassoli, 2017. ⁵¹

3. Material and methods

3.1 Material, reagents, devices and cells

3.1.1. Material and chemical reagents

All chemical reagents, if not else stated, were purchased from Sigma-Aldrich Chemie GmbH, Carl Roth GmbH & Co KG, VWR International GmbH, Fisher Scientific GmbH, Fluorochem Ltd, Carbolution Chemicals GmbH, Th. Geyer GmbH & Co KG.

Material, enzymes and buffers	Supplier
BglII	Thermo Fisher Scientific
96-deep well plate, 2 mL	Thermo Fisher Scientific
Electroporation cuvette, 0.1 cm gap	Bio-Rad
GeneJET Gel-Extraction kit	Thermo Fisher Scientific
GeneRuler 1 kb DNA Ladder	Thermo Fisher Scientific
GelPilot Loading Dye, 5x	Thermo Fisher Scientific
High Capacity cDNA Reverse Transcription Kit	Applied Biosystems
HPLC sample vials	VWR International GmbH
Microplate, 384 well, 4514	Corning
Microplate, 96-well, 655097	Greiner Bio-One
Multichannel pipettes	Thermo Fisher Scientific and Integra Biosciences
NdeI	Thermo Fisher Scientific
NEBuffer 3.1.	New England Biolabs
NEBuffer 2.1.	New England Biolabs
Needles	Braun
Coomassie Protein-Assay-Kit	Thermo Fisher Scientific
Pipetting aid	Integra Biosciences
PowerUp™ SYBR™ Green Master Mix	Applied Biosystems
Phusion polymerase	New England Biolabs or Thermo Fisher Scientific
QIAprep Mini- and Midiprep Kit	Qiagen
QIAquick Gel Extraction Kit	Qiagen
QIAquick PCR Purification Kit	Qiagen
Reaction tubes (0.5, 1 and 2 mL)	Sarstedt
Rec A	Thermo Fisher Scientific

3. Material and methods

Rec A buffer	Thermo Fisher Scientific
RNeasy Mini Kit	Qiagen
RNAprotect Bacterial Reagent	Qiagen
Sample vials	Fisher Scientific GmbH
Sereological pipettes (1, 5, 10, 25 and 50 mL)	Sarstedt
Syringe reactors (5 and 10 mL)	Multisyntech GmbH
Syringes	Braun
Syringe filters (0.22 and 0.45 µm)	Merck Millipore or Fisher Scientific
T4 DNA ligase	Thermo Fisher Scientific
T4 DNA ligase buffer	Thermo Fisher Scientific

3.1.2. Devices

Devices	Company/Supplier and model name
70 µm microfluidic chip (cell sorter)	Sony Biotechnology, Weybridge, U. K.
Cell Sorter	Sony Biotechnology, Weybridge, U. K., SH800SFP
Centrifuges	Beckman Coulter Avanti J-25, Eppendorf Centrifuge 5417 R, Eppendorf Centrifuge 5424 & Eppendorf Centrifuge 5805 R
Gel Imager	Bio-Rad ChemiDoc Imaging System
DNA electrophoresis cell	Bio-Rad Mini-Sub Cell GT Cell
Electrophoresis power supply	Bio-Rad PowerPac™ Basic Power Supply
Electroporation Systems	Bio-Rad Gene Pulser Xcell
Incubator	Binder Drying and heating chamber ED 23
Plate reader	Tecan Sparks
Peptide synthesizer	Multisyntech GmbH Syro I peptide synthesizer
Protein electrophoresis cell	Bio-Rad Mini-PROTEAN Tetra Vertical Electrophoresis Cell
qPCR cyclers	Applied Biosystems 7500 Fast Real-Time PCR System
Shaking incubator	New Brunswick Innova & 42 New Brunswick Innova 443
Thermomixer	Eppendorf Thermomixer comfort 5355
Thermocycler	Eppendorf Thermocycler, Mastercycler
UV/Vis spectrophotometer	Thermo Fisher Scientific NanoDrop 2000c

3. Material and methods

3.1.3. Software

Software	Company/Supplier
Cell Sorter Software Version 2.1.5	Sony Corporation
ChemDraw	PerkinElmer
ESEfinder 3.0	Krainer Lab and Zhang Lab, Cold Spring Harbor Laboratory
GraphPad Prism 9	GraphPad Software, Inc.
R and RStudio	Posit PBC
RNAfold Web Server	ViennaRNA Web Services
RNAstructure and StructureEditor	Mathews group
SnapGene	GSL Biotech LLC
Xcalibur	Thermo Fisher Scientific

3.1.4. RNA oligonucleotides

All RNA oligonucleotides were purchased from Integrated DNA Technologies or Sigma-Aldrich.

3.1.5. Medium and self-made buffers

All mediums and agar plates were provided by the Protein Chemistry Facility (PCF) based at the MPI in Dortmund.

Media, buffers and solutions	Composition
Luria Broth (LB)	1 % Tryptone 0.5 % Yeast extract 171 mM NaCl pH 7.4
M9 Minimal medium	100 mM Na ₂ HPO ₄ 22 mM KH ₂ PO ₄ 8.55 mM NaCl 18.7 mM NH ₄ Cl 1 mM MgSO ₄ 22 mM Glucose 0.1 mM CaCl ₂

3. Material and methods

Super Optimal Broth (SOB)	2 % Tryptone 0.5 % Yeast extract 10 mM NaCl 2.5 mM KCl 10 mM MgCl ₂ 10 mM MgSO ₄
Super Optimal Broth with catabolite repression (SOC)	2 % Tryptone 0.5 % Yeast extract 10 mM NaCl 2.5 mM KCl 10 mM MgCl ₂ 10 mM MgSO ₄ 20 mM Glucose
Terrific broth (TB)	1.2 % Tryptone 2.4 % Yeast Extract 0.4 % Glycerol 72 mM K ₂ HPO ₄ 17 mM KH ₂ PO ₄ pH 7
2xYT	1.6 % Tryptone 1 % Yeast extract 85.5 mM NaCl pH 7.4
1x TAE Buffer	40 mM Tris-acetate 1 mM EDTA
PBS	137 mM NaCl 2.7 mM KCl 8 mM Na ₂ HPO ₄ 2 mM KH ₂ PO ₄
5x SDS sample buffer	200 mM Tris (pH 6.8) 40 % Glycerol

3. Material and methods

	8 % SDS
	10 mM DTE
	0.007 mM Bromophenol blue
SDS running buffer	25 mM Tris
	200 mM Glycin
	1 % SDS
SDS-gel staining solution	0.1 % Coomassie R250
	10 % Acetic acid
	40 % MeOH
SDS-gel destaining solution	10 % Acetic acid in water

3.1.6. Bacterial cells

The bacterial cells used in the assays were *E. coli* Top10F' (Invitrogen). For protein expression, *E. coli* BL21 DE3 or *E. coli* BL21 CodonPlus (DE3) RIPL cells were used. For cloning, *E. coli* Top10F' cells were used.

3.2. Methods

3.2.1 Bacterial cell culture

3.2.1.1. Preparation of bacterial cultures

An adequate volume of medium (LB, TB, SOB or TY) was transferred into an Erlenmeyer-flask or tube of appropriate volume (~4-times higher than the volume of medium). Antibiotics were added in standard working concentrations from a 1000-fold stock solution. With a 10 μ L pipette tip, cells from an agar plate or a glycerol stock were used to inoculate the medium. The culture was grown overnight at 37 °C or 30 °C. After overnight growth, the culture was used for plasmid extraction or as a pre-culture for protein expression.

3.2.1.2. Preparation of bacterial glycerol stocks for long-term storage

In a 2 mL cryotube, 500 μ L of a saturated bacterial culture was mixed with 500 μ L of 50 % sterile glycerol solution. The glycerol stock was stored at -80 °C.

3. Material and methods

3.2.1.3. Preparation of electrocompetent cells

A bacterial pre-culture was used as a starter culture to inoculate 100, 500, 1000 or 2000 mL of medium (SOB) in an Erlenmeyer-flask at a concentration of 0.25-0.5 % (v/v). The culture was shaken at 37 °C, and the optical density at 600 nm (OD_{600}) was monitored until an OD_{600} of 0.5-0.7 was reached. Growth of the culture was stopped by placing the culture at 4 °C-8 °C and incubation for 30 min. The culture is centrifuged at 5000xg at 4 °C for 15 minutes. The cell pellet was washed with half of the volume of the original culture with 10 % glycerol, incubated for 15 minutes at 4 °C, and then centrifuged under the same conditions. The procedure was repeated. Then, the cells were pelleted again and resuspended in 1 % of the volume of the initial culture in 10 % glycerol. Snap-frozen 100 μ L aliquots were stored at -80 °C.

3.2.2.4. Preparation of chemically competent cells

A bacterial pre-culture was used as a starter culture to inoculate 500-2000 mL of medium (SOB) in an Erlenmeyer-flask with appropriate antibiotics or no antibiotic at a concentration of 0.25-0.5 % (v/v). The culture was shaken at 37 °C, and the OD_{600} was monitored until an OD_{600} of 0.5-0.7 was reached. Growth of the culture was stopped by placing the culture at 4 °C-8 °C and incubating for 30 min. The culture was centrifuged at 5000 x g at 4 °C for 15 minutes to pellet cells. The cell pellet was washed with half volume of the original culture volume with 0.1 M $MgCl_2$ and incubated for 15 minutes at 4 °C and then centrifuged under the same conditions. The cell pellet was washed with half volume of the original culture volume with 0.1 M $CaCl_2$, incubated for 15 minutes at 4 °C and then centrifuged. Then, the cells were pelleted and resuspended in 1 % of the volume of the initial culture in 0.1 M $CaCl_2$ with 10 % glycerol. Snap-frozen 100 μ L aliquots are stored at -80 °C.

3.2.2.5. Transformation by electroporation

Electrocompetent cells were thawed on ice for 5-10 minutes. 1-5 μ L of the respective plasmid was added to the cells and incubated for 10-15 minutes. The cell solution was transferred to an electroporation chamber and pulsed with electroshocks (1.8 kV, 200 Ω for 1 mm cuvette; 2.4 kV, 200 Ω for 2 mm cuvette) in an electroporator. SOC medium pre-warmed to 37 °C was added subsequently to the cuvette, and the cell suspension was added to a reaction tube. The reaction tube was shaken in a thermomixer (300 rpm) for 1 h at 37 °C. After incubation, 300 μ L of the cell suspension was plated on an agar plate with the corresponding antibiotic. The transformed cells were grown overnight at 37 °C or for 3 days at room temperature.

3. Material and methods

3.2.2.6. Transformation by heat shock

Chemically competent cells were thawed on ice for 5-10 minutes. 1-5 μL of the respective plasmid was added to the cells and incubated for 10-15 minutes. The reaction with the cell suspension tube was transferred to a heat block and heat shocked for 45 seconds at 42 °C. The reaction tube was transferred to ice, and 300 μL SOC medium pre-warmed to 37 °C was added. The reaction tube was shaken in a thermomixer (300 rpm) for 1 h at 37 °C. After incubation, 300 μL of the cell suspension was plated on an agar plate with the corresponding antibiotic. The transformed cells were grown overnight at 37 °C or for 3 days at room temperature.

3.2.2. DNA Extraction and Sequencing

3.2.2.1. Plasmid Extraction by Qiagen QiaPrep Mini- or Midiprep Kit

A bacterial culture of desired volume was prepared (as described in 3.2.1.1.) and used further for plasmid extraction following the manufacturer's instructions using the Qiagen QIAprep Mini- or Midiprep Kit.

3.2.2.2. Sanger sequencing by Microsynth commercial service

After plasmid extraction, the DNA concentration was determined by spectrophotometric measurement using a NanoDrop device. A plasmid DNA concentration of 40-120 ng/ μL in 12 μL volume is prepared in a 1.5 mL reaction tube. For sequencing with premixed primers, the primer of choice is added in a concentration of 10 μM in a 3 μL volume.

3.2.2.3. Illumina sequencing by Novogene commercial service

After plasmid extraction, the DNA was used to PCR amplify the amplicon to be sequenced, as stated in 3.2.3.2. Amplicons of a length between 269 and 495 bp were generated while the varying inserts (10 bp for the RBP consensus sequence screening and 15 bp for the SICLOPPS screening) were amplified with flanking regions. When pool-sequencing was performed, barcodes on one or both primers were used. The template plasmid was digested overnight at room temperature (total volume 140 μL , 15 μL 10xTANGO Buffer, 2 μL DpnI (10 U/ μL). The reaction was PCR purified, as stated in 3.2.3.4. ~1 μg of DNA in a volume of ~30 μL was sent to the service at room temperature.

3. Material and methods

3.2.3. Cloning

3.2.3.1. Agarose-gel electrophoresis

1 % (w/v) agarose (or higher) was dissolved in 1x TAE buffer (see section 3.1.5.), boiled in a microwave oven and cooled to 50-60 °C. Then, a gel dye was added (Midori green or SYBR safe according to the manufacturer's instructions), and the gel was immediately cast. The gel was polymerized for ~30 min. Samples were mixed with 1/5 parts of gel loading dye (gel pilot 5x) and loaded into the gel wells. Gel-electrophoresis was performed at 110 V and 400 mA until sufficient separation was achieved.

3.2.3.2. Polymerase chain reaction (PCR) with Phusion polymerase

In a 200 µL reaction, 133 µL ddH₂O, 40 µL 5x HF buffer, 1 µL (1-50 ng) template, 10 µL of 10 µM of each primer, 4 µL of 10 mM dNTPs and 2 µL of Phusion polymerase (2 U/µL) were added in a 0.5 mL reaction tube and mixed. 20 µL of the reaction mixture was aliquoted into 8 PCR reaction tubes, centrifuged down and transferred into the thermocycler. The PCR reaction was started using the protocol described in Table 2.

Table 2: Thermocycling conditions for Phusion PCR.

Cycle	Denature	Anneal	Extend
1	98 °C, 30 s		
2-26	98 °C, 10 s	Temperature Gradient (55.8 °C to 63.9 °C), 30s	72°C, (0.5 min per kb)
27			72°C , 10 min
Hold (4 °C)			

3.2.3.3. DpnI digest

Sample volumes were measured, and 10x TANGO buffer was added to 1x, then 0.66 µL (for overnight incubation) or 1 µL (for 2 h incubation) DpnI (10 U/µL) was added per 50 µL total reaction volume. The reaction mixture was incubated either at 37 °C for 2 h or at room temperature overnight or over the weekend, then PCR purified or used for gel extraction.

3.2.3.4. PCR purification

PCR purification was performed with the QIAquick PCR purification kit or with the GeneJet PCR purification kit according to the manufacturer's protocols. Purified samples were eluted with MilliQ H₂O in a total volume of 35 or 50 µL.

3. Material and methods

3.2.3.5. Gel extraction

Samples were supplied with the appropriate volume of gel loading dye (gel pilot 5x), and an agarose gel-electrophoresis was performed (see section 3.2.3.1.). Gel bands were visualized with a blue light transilluminator and excised using a plastic gel cutter. The band-containing gel was transferred into a preweighed reaction tube, and its weight was determined.

3.2.3.6. Gel purification

Gel purification was performed with the QIAquick gel purification kit or with the GeneJet gel extraction kit following the manufacturer's protocol. Purified samples were eluted in 35 μ L MilliQ H₂O.

3.2.3.7. Restriction and ligation

For cloning of the TRAP RNA plasmids, the template plasmid S0 was amplified by growing a 50 mL culture of S0 in *E. coli* Top10F' and harvesting the plasmid by Midiprep (see sections 3.2.1.1. and 3.2.2.1. Up to 5 μ g of plasmid DNA was double digested with 3 μ L BglIII and 3 μ L NdeI in 1x NEB Buffer 3.1 in a total volume of 100 μ L at room temperature overnight or over the weekend. The reaction was separated on an agarose gel, gel-extracted and gel-purified (see 3.2.3.1, 3.2.3.5. and 3.2.3.6.). To generate the double-stranded insert for the ligation reaction, oligonucleotide inserts were designed to be reverse-complemented to each other and contain sticky ends that match the restriction sites of BglIII and NdeI. Oligonucleotides were hybridized using 10 μ L of each oligonucleotide in a concentration of 10 μ M and mixed with 1x HF buffer in a total volume of 50 μ L. The reaction mixture was heated up to 95 °C in a thermomixer and gradually cooled down by setting the device to 22 °C. Ligation of both fragments was performed using T4 ligase. In a 10 μ L reaction, a 10:1 insert-to-vector ratio was supplemented with 1x T4 ligase buffer using 1 μ L of T4 ligase (5 U/ μ L). The reaction was incubated at 37 °C for 1-3 h or at room temperature overnight or over the weekend. 1 or 3 μ L were used for transformation into chemically competent cells (see 3.2.2.6. Colonies were picked, cultivated, and plasmid DNA was extracted (as mentioned in 3.2.1.1. and 3.2.2.1.). Correctness of plasmid DNA was checked by sequencing (see 3.2.2.2.).

3.2.3.8. Sequence and ligation independent cloning (SLIC)

For cloning of the sfGFP-RBP fusion construct, SLIC cloning was used according to the protocols of Li and Elledge, 2012.⁶² The backbone plasmid was PCR amplified using primers 1 and 2, while for the insert, the protein template plasmid was used to PCR amplify the protein

3. Material and methods

ORF with primers 3 and 4, which included 20 bp homology regions of the backbone plasmid (see Supplementary table 1). For the other sfGFP-RBP fusion protein plasmids, equivalent primers for the protein ORF were used. For PCR amplification, the instructions in section 3.2.3.2. were used. Both PCR products were then template digested and PCR purified, as mentioned in sections 3.2.3.3. and 3.2.3.4. The backbone and insert were then treated with T4 DNA polymerase using its exonuclease activity to generate 5' overhangs. 1 µg of DNA was treated with 0.5 U/µL T4 DNA polymerase using 2 µL of 10× NEBuffer 2.1 and filled up with water to a total volume of 20 µL at room temperature for 30 min. The reaction was stopped by addition of 2 µL of 10 mM dNTP. Backbone and insert were annealed using a vector to insert ratio of 1:10, 1 µL of 10× T4 DNA polymerase buffer, 330 ng RecA per 50 ng of vector and ddH₂O filled up to a total volume of 10 µL. The reaction was incubated at 37 °C for 1-3 h or at room temperature overnight. 3 µL of the annealing reaction were transformed as stated in 3.2.2.6. 8 mL liquid cultures of resulting colonies were prepared, and plasmids were extracted as mentioned in sections 3.2.1.1. and 3.2.2.1. Presence of the correct plasmid was checked by sequencing (see 3.2.2.2.).

3.2.3.9. Library generation with circular polymerase extension cloning (CPEC)

First, the backbone and insert were PCR amplified with the primers 5 and 6 for the backbone and primers 7 and 8 for the insert (shown in Supplementary table 1) according to entry 3.2.3.2. The combined fractions and the template plasmids were digested using 15 µL 10× TANGO buffer and 2 µL DpnI (10 U/µL) over the weekend at room temperature. The reaction was PCR purified as mentioned in 3.2.3.4. and a CPEC reaction was set up: 200 ng of the backbone and 430.1 ng of the insert (1:10 ratio) were mixed with 4 µL 5x HF buffer, 1.6 µL 10 mM dNTPs and 0.2 µL Phusion polymerase. A control approach was set up accordingly, replacing the insert with ddH₂O. The CPEC reaction was performed in a thermocycler using the following protocol:

Table 3: Thermocycling conditions for the CPEC reaction.

Cycle	Denature	Anneal	Extend
1	98 °C, 30 s		
2-11	98 °C, 10 s	55.1 °C, 30s	72°C, 2 min
12			72°C, 5 min
Hold (4 °C)			

3. Material and methods

1 μL of the crude reaction mixtures was transformed into electrocompetent Top10F' cells. Dilution streaking led to a total of 2×10^6 colonies for the library and 59 colonies for the backbone control. Five transformation approaches of the library were performed, plated onto five LB agar plates (12.5×12.5 cm) with 50 $\mu\text{g}/\text{mL}$ kanamycin for each transformation and grown overnight. All cells were scraped from the plates, dissolved in 50 mL LB medium with 50 $\mu\text{g}/\text{mL}$ kanamycin for each transformation and grown overnight. Library plasmids were harvested by MidiPrep as stated in 3.2.2.1.

3.2.4. Assays, plate assays, screenings and analysis

3.2.4.1. Translational repression assay procedure

E. coli Top10F' cells were cotransformed with the TRAP assay plasmids. A 1 mL preculture of the cells was prepared in LB medium with 50 $\mu\text{g}/\text{mL}$ kanamycin and 34 $\mu\text{g}/\text{mL}$ chloramphenicol in a 96-deep well plate. The culture was either prepared from a glycerol stock, or from the cotransformation plate. The plate was sealed with a semipermeable sticky lid and the cultures were grown overnight at 37 °C in a shaking incubator (160 rpm). The next day the cultures were diluted 1:19 in M9 minimal medium with respective antibiotics in a total volume of 190 μL in a black, 96-well plate with clear bottom. OD_{600} was monitored until a value of ~ 0.2 was reached. The assay was induced with IPTG to a final concentrations of 1 mM and arabinose to a final concentration of 0, 0.125, 0.25, 0.5 or 1 %. Fluorescence of tagBFP (ex/em: 402 nm/457 nm) and sfGFP (ex/em: 485 nm/510 nm) and OD_{600} were recorded every 20 minutes over a time course of ~ 7 h at 30 °C in a Tecan Spark plate reader. Data evaluation was performed according to the report by Katz *et al.* TagBFP production rate is calculated using TagBFP levels divided by the integral of cell density in a time interval within the linear growth phase. SfGFP expression levels were generated using sfGFP fluorescence normalized to cell density and then averaged by the number of time points within the chosen time interval. TagBFP production rates are plotted against the sfGFP expression levels to generate repression graphs.^{63,64} Repression ratios were calculated by dividing the basal TagBFP production rate (non-induced sfGFP-RBP, 0 % arabinose) by the TagBFP production rate at the highest expression level (1 % arabinose).

3.2.4.2. Flow cytometry

E. coli Top10F' cells were cotransformed with the TRAP assay plasmids. A 1 mL preculture of the cells was prepared in LB medium with kanamycin 50 $\mu\text{g}/\text{mL}$ and chloramphenicol

3. Material and methods

34 µg/mL in a 96-deep well plate. The culture was either prepared from a glycerol stock, or from the cotransformation plate. The plate was sealed with a semipermeable sticky lid and the cultures were grown overnight at 37 °C in a shaking incubator (160 rpm). The next day the cultures were diluted 1:19 in LB medium with respective antibiotics in a total volume of 1 mL in a 96-deep-well plate. OD₆₀₀ was monitored until a value of ~0.2 was reached. The assay was induced with 1 mM IPTG and 0.125 % arabinose for 4 h and 30 minutes at 30 °C. The plate was centrifuged at 4500 x g for 5 minutes in a tabletop centrifuge to remove the medium. Cells were washed using 500 µL PBS, centrifuged, and washing was repeated. Finally, cells were centrifuged and resuspended in 800 µL PBS and placed on ice until analysis.

Samples were analysed on an SH800SFP Cell Sorter using a 70 µm microfluidic chip, and tagBFP and sfGFP fluorescence intensities were recorded and compensated using the respective single-colour controls. Each measurement was replicated twice.

3.2.4.3. Fluorescence polarisation assay

Direct FP

Labelled RNA oligonucleotides were dissolved in nuclease-free water at 100 µM according to the manufacturer's instructions and kept at -20 °C until use. The assay was performed in protein storage buffer (see sections 3.2.5.1. and 3.2.5.1.) with 0.01 % Triton, in black, 384 well-plates, with a total volume of 20 µL per well. Protein concentration was determined by spectrophotometric measurement using a NanoDrop device or with the Coomassie (Bradford) Protein-Assay-Kit following the microplate procedures. The 6-FAM labelled RNAs were tested at a final concentration of 1 nM, and the appropriate unlabelled protein was titrated as two-fold dilution series. Fluorescence polarisation was measured at room temperature using a plate reader with (ex/em) 490 nm/520 nm for 6-FAM and 610 nm/670 nm for Cy5. Direct binding experiments for measured after 20 min, if not else stated.

Competition FP

Unlabelled RNA oligonucleotides were dissolved in nuclease-free water at 100 µM according to the manufacturer's instructions and kept at -20 °C until use. The assay was performed in protein storage buffer (see sections 3.2.5.1. and 3.2.5.1.) with 0.01 % Triton, in black, 384 well-plates, with a total volume of 20 µL per well. Protein concentration was determined by spectrophotometric measurement using a NanoDrop device or with the Coomassie (Bradford) Protein-Assay-Kit following the microplate procedures. The 6-FAM labelled RNAs were tested at a final concentration of 1 nM and were mixed with the tested protein at desired

3. Material and methods

concentration. Unlabelled competitor (RNA or peptide) was titrated as two-fold dilution series. Fluorescence polarisation was measured at room temperature using a plate reader with (ex/em) 490 nm/520 nm and 610 nm/670 nm for Cy5.

3.2.4.4. Reverse transcription quantitative polymerase chain reaction (RT-qPCR)

E. coli Top10F' cells were cotransformed with sfGFP-SRSF1/S6 assay plasmids. A 2 mL preculture of cells in LB medium with kanamycin 50 µg/mL and chloramphenicol 34 µg/mL was prepared and grown overnight at 37 °C in a shaking incubator (160 rpm). The next day the cultures were diluted 1:19 in M9 medium with respective antibiotics in a total volume of 1.9 mL. OD₆₀₀ was monitored until a value of ~0.2 was reached. The assay plasmids were induced with 1 mM IPTG and 0, 0.125, 0.25, 0.5 or 1 % arabinose for 4 h and 30 minutes at 30 °C. 500 µL of cell suspension was treated with 1 mL of RNAprotect Bacterial Reagent, incubated for 5 minutes and then centrifuged for 10 minutes at 5000×g. The supernatant was decanted, and the cells were either stored overnight at 20 °C or total RNA was isolated immediately. RNA isolation was performed according to the RNAprotect Bacterial Reagent Handbook following protocols 1, 7 and appendix B. Total RNA was reverse-transcribed using the High Capacity cDNA Reverse Transcription Kit. Quantitative PCR was performed using the PowerUp™ SYBR™ Green Master Mix in a qPCR cycler. Fold expression of genes was calculated with the $\Delta\Delta C_t$ -method and normalized to the gapA gene.⁶⁵ The assay was performed in technical duplicates on two independent days (quadruplicates). Primer sequences, amplicon sizes and efficiencies are reported in Supplementary table 1.

3.2.4.5. TRAP evaluation assay

Frozen glycerol stocks originating from the SICLOPPS screening (see 3.2.4.10.) were used to inoculate a 1 mL preculture in LB medium with 50 µg/mL kanamycin, 34 µg/mL chloramphenicol and 100 µg/mL ampicillin in a 96-deep well plate. Control samples were prepared in a 1 mL preculture in LB medium with 50 µg/mL kanamycin and 34 µg/mL chloramphenicol. The plate was sealed with a semipermeable sticky lid, and the cultures were grown overnight at 37 °C in a shaking incubator (160 rpm). The next day the cultures were diluted 1:20 in 1 mL LB medium with respective antibiotics and grown until OD₆₀₀~0.2 was reached, then induced with 100 ng/mL anhydrotetracycline and grown overnight at 30 °C. The cultures were diluted 1:19 in M9 minimal medium with respective antibiotics in a total volume of 190 µL in a black, 96-well plate with clear bottom and grown at 37 °C until OD₆₀₀~0.2 was reached. The assay was induced with IPTG to a final concentration of 1 mM and arabinose to a final concentration of 0 and 0.125 %. Fluorescence of tagBFP (ex/em: 402 nm/457 nm) and

3. Material and methods

sfGFP (ex/em: 485 nm/510 nm), and OD₆₀₀ were recorded every 20 minutes over a time course of ~7 h at 30 °C in a Tecan Spark plate reader. Data evaluation was performed according to the report by Katz *et al.* TagBFP production rate is calculated using TagBFP levels divided by the integral of cell density in a time interval within the linear growth phase. SfGFP expression levels were generated using sfGFP fluorescence normalized to cell density and then averaged by the number of time points within the chosen time interval.^{63,64}

3.2.4.6. RBP consensus sequence screening

E. coli Top10F' cells were transformed with the sfGFP-RBP fusion protein plasmid and grown overnight at 37°C. A single colony was used to prepare a 10 mL preculture to generate electrocompetent cells according to section 3.2.1.3. Cells pretransformed with the protein plasmid were then transformed with 1 µL (2.66 µg/µL) of the 10mer plasmid library. Cells were plated on one 24.5 x 24.5 cm dish or on five 12.5 x 12.5 cm dishes filled with LB agar with 50 µg/mL kanamycin and 34 µg/mL chloramphenicol and grown overnight at 37 °C. The next day all cells were scraped and transferred into 50 mL LB medium with corresponding antibiotics and grown overnight at 37 °C. The culture was diluted 1/80 into 5 mL LB and was grown at 37 °C until OD₆₀₀~0.2, then induced with 0.125 % arabinose and 1 mM IPTG for 4 h 30 min at 30 °C. The culture was transferred on ice and centrifuged at 14,000xg at 4 °C. Cells were washed with 500 µL PBS, centrifuged, and washing was repeated. Centrifugation was repeated, cells were resuspended in 800 µL PBS and placed on ice until analysis. Samples were sorted with the SH800SFP Cell Sorter using a 70 µm microfluidic chip and the “normal” sorting mode. TagBFP and sfGFP fluorescence intensities were recorded and compensated using the respective single-colour controls. Events that showed a repressed phenotype (low tagBFP levels) were gated for sorting. Cells were coeluted with PBS and sorted into 1.5 mL reaction tubes. Depending on the volume, collected cells were plated on one or more 12.5 x 12.5 cm LB agar plates with 50 µg/mL kanamycin and grown overnight at 37 °C. The next day cells were scraped and transferred into 50 mL LB with 50 µg/mL kanamycin and grown overnight at 37 °C. Plasmids were extracted as mentioned in section 3.2.2.1. and samples were prepared for sequencing as described in 3.2.2.3.

3.2.4.7. Autorepressor presorted RBP consensus sequence screening

E. coli Top10F' cells were transformed with the sfGFP-RBP fusion protein plasmid and grown overnight at 37°C. A single colony was used to prepare a 10 mL preculture to generate electrocompetent cells according to section 3.2.1.3. Electrocompetent cells pretransformed with the protein plasmid were then transformed with 1 µL (2.66 µg/µL) of the 10mer plasmid

3. Material and methods

library. Cells were plated on one 24.5 x 24.5 cm dish or on five 12.5 x 12.5 cm dishes filled with LB agar with 50 µg/mL kanamycin and 34 µg/mL chloramphenicol and grown overnight at 37 °C. The next day all cells were scraped and transferred into 50 mL LB medium with corresponding antibiotics and grown overnight. The culture was diluted 1/80 into 5 mL LB and grown at 37 °C until $OD_{600} \approx 0.2$, then induced with 1 mM IPTG for 4 h 30 min at 30 °C. The culture was transferred on ice and centrifuged at 14,000xg at 4 °C. Cells were washed with 500 µL PBS, centrifuged and washing was repeated. Cells were then centrifuged and resuspended in 800 µL PBS and placed on ice until analysis. Samples were sorted with the SH800SFP Cell Sorter using a 70 µm microfluidic chip and the “purity” sorting mode. TagBFP and sfGFP fluorescence intensities were recorded and compensated using the respective single-colour controls. Events that showed an unrepressed phenotype (high tagBFP levels) were gated for sorting. Cells were coeluted with PBS and sorted into 1.5 mL reaction tubes. Depending on the volume collected, cells were plated on one or more 12.5 x 12.5 cm LB agar plates with 34 µg/mL chloramphenicol and 50 µg/mL kanamycin and grown overnight at 37 °C. The next day cells were scraped and resuspended in 5 mL LB with 34 µg/mL chloramphenicol and 50 µg/mL kanamycin. The optical density was determined, and the culture was mixed with 25 % glycerol, snap-frozen and kept at -80 °C until the procedure was continued. The culture was diluted 1/568 for SRSF1 and 1/781 for PTBP1 into 5 mL LB with corresponding antibiotics and grown at 37 °C until $OD_{600} \approx 0.2$ then induced with 0.125 % arabinose and 1 mM IPTG for 4 h 30 min at 30 °C. The culture was transferred on ice and centrifuged at 14,000xg at 4 °C. Cells were washed with 500 µL PBS, centrifuged and washing was repeated. Cells were then centrifuged and resuspended in 800 µL PBS and placed on ice until analysis. Samples were sorted with the SH800SFP Cell Sorter using a 70 µm microfluidic chip and the “purity” sorting mode. TagBFP and sfGFP fluorescence intensities were recorded and compensated using the respective single-colour controls. This time events that showed a repressed phenotype (low tagBFP levels) were gated for sorting. Cells were coeluted with PBS and sorted into 1.5 mL reaction tubes. Depending on the volume collected, cells were plated on one or more 12.5 x 12.5 cm LB agar plates with 50 µg/mL kanamycin and grown overnight at 37 °C. The next day cells were scraped and transferred into 50 mL LB with 50 µg/mL kanamycin and grown overnight at 37°C. Plasmids were extracted as mentioned in 3.2.2.1. and samples were prepared for sequencing as described in 3.2.2.3.

3.2.4.8. Data analysis: RBP consensus sequence screening

A paired-end sequencing dataset containing the forward and reverse complemented strands was uploaded to the galaxy web platform (usegalaxy.org). First, a quality filter step was performed using the command *Filter by quality* with a “Quality cut-off value” of 20 and “Percent of bases in sequence that must have quality equal to / higher than cut-off value” of 50. As pool-sequencing was performed, a text file with the used barcodes on the reverse complemented strand was uploaded and used for demultiplexing the pool using *Barcode splitter* and “Number of allowed mismatches” was set to 1. The resulting matching files for the reverse complemented strands were downloaded and reuploaded to the platform. The 3’ ends of the sequences were removed using the command *Clip* using the clipping adapter sequence: AGATCTTTTGAATTCTGAAATTGTTA

The “Minimum sequence length (after clipping, sequences shorter than this length will be discarded)” was set to 10. A trim of the 5’ end before the varying insert was performed using *Trim* and “Trim from the beginning up to this position” set to 215 and the “Is input dataset in FASTQ format?” option was set to “Yes” so only the varying inserts were left behind. The datatype was changed using the command *FASTQ to FASTA*. The inserts were filtered by length using *Filter sequences by length* to remove inserts that were shorter or longer than 10 nucleotides setting the “Minimum length” and “Maximum length” each to 10. All unique inserts were added up using the command *Collapse* and reverse complemented using *Reverse-Complement*.

For removal of the “Autorepressors” from the datasets, the “Autorepressors” sequences were converted into a .txt file and used for barcode splitting using *Barcode splitter*. The resulting file that was “unmatched” was downloaded and resembled the autorepressor subtracted data file (w/o auto).

3.2.4.9. Data analysis: Autorepressor presorted RBP consensus sequence screening and 10mer RNA consensus sequence library sequencing

Both files of the paired-end sequencing dataset, containing the forward and reverse complemented strands, were uploaded to the galaxy web platform (usegalaxy.org). First, a quality filter step was performed using the command *Filter by quality* with a “Quality cut-off value” of 20 and “Percent of bases in sequence that must have quality equal to / higher than cut-off value” of 50. When pool-sequencing was performed, a text file with the used barcodes on the forward and the reverse complemented strands was uploaded and used for

3. Material and methods

demultiplexing the pool using *Barcode splitter*. Otherwise, a text file with extended primers on both ends was used to check the library for mismatches. The “*Number of allowed mismatches*” was set to 1. The resulting matching files for the forward and the reverse complemented strands were downloaded and reuploaded to the platform. The datatype was changed using the command *FASTQ to FASTA*, and both forward strand files and both reverse complemented strand files were merged using *Merge.files*. The reverse complemented strand files were reverse complemented using *Reverse-Complement* to achieve the forward strands. The 3’ ends of the sequences were removed using the command *Clip* using the clipping adapter sequence:

```
ATATTTAAGAAGGAGATATACATATGAGCGAGCTGATTAAGGAGAACAT
```

The “*Minimum sequence length (after clipping, sequences shorter than this length will be discarded)*” was set to 10. A trim of the 5’ end before the varying insert was performed for the strands that originate from the forward strand files using *Trim* and “*Trim from the beginning up to this position*” set to 192 and the “*Is input dataset in FASTQ format?*” option was set to “*Yes*”. Another trim of the 5’ end before the varying insert was performed for the strands that originate from the reverse-complemented strand files using *Trim* and “*Trim from the beginning up to this position*” set to 27 and the “*Is input dataset in FASTQ format?*” option was set to “*Yes*” so only the varying inserts were left behind. The two forward strand files were then merged using *Merge.files*, and the inserts were filtered by length using *Filter sequences by length* to remove inserts that were shorter or longer than 10 nucleotides setting the “*Minimum length*” and “*Maximum length*” each to 10. Lastly, all unique inserts were added up using the command *Collapse*. A schematic representation of the workflow is shown in Figure 8.

3. Material and methods

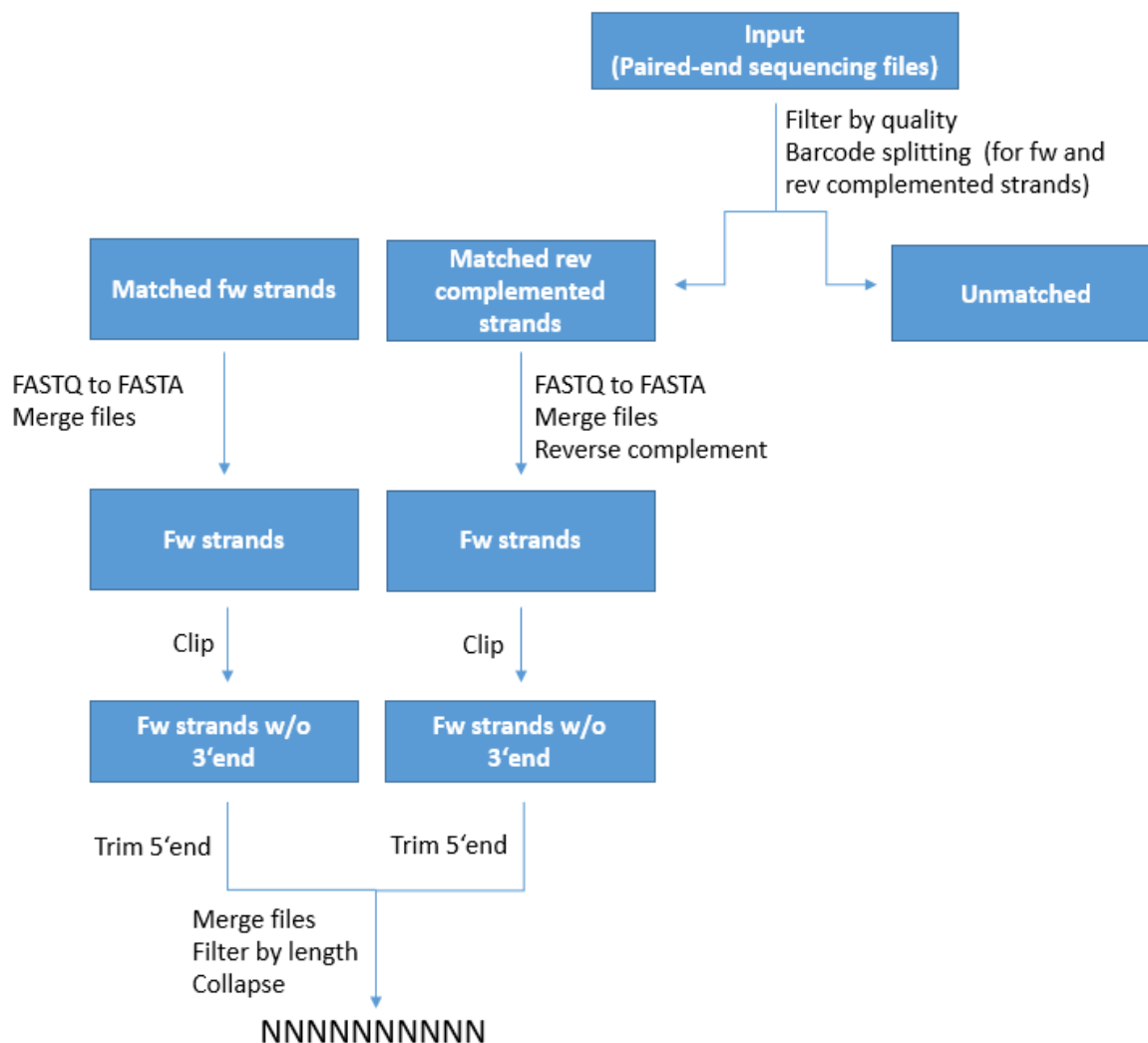


Figure 8: Schematic representation of the analysis steps of the RBP consensus sequence screening dataset using the Galaxy web platform. ⁶⁶

3.2.4.10. SICLOPPS screening analysed by Sanger sequencing

E. coli Top10F' cells were transformed with the sfGFP-RBP fusion protein plasmid and plasmid S6 (for SRSF1) or plasmid H8 (for hnRNP A2B1) and grown overnight at 37 °C. A single colony was used to prepare a 10 mL preculture to generate electrocompetent cells according to section 3.2.1.3. Electrocompetent cells pretransformed with the protein plasmid were then transformed with 1 µL (1.19 µg/µL) of the SICLOPPS plasmid library. Cells were plated on a 10 cm round dish containing LB agar with 50 µg/mL kanamycin, 34 µg/mL chloramphenicol and 100 µg/mL ampicillin. The next day all cells were scraped and transferred into 50 mL LB medium with corresponding antibiotics and grown for 2 h at 37 °C. Cells were then induced with 100 ng/mL anhydrotetracycline overnight at 30 °C. The culture

3. Material and methods

was diluted 1/20 into 1 mL LB with the corresponding antibiotic and grown at 37 °C until $OD_{600} \approx 0.2$. The culture was induced with 0.125 % arabinose and 1 mM IPTG for 7 h 30 min at 30 °C. Cells were transferred on ice and centrifuged at 14,000×g at 4 °C. Cells were washed 500 µL PBS, centrifuged and washing was repeated. Cells were then centrifuged and resuspended in 800 µL PBS and placed on ice until analysis. Samples were sorted with the SH800SFP Cell Sorter using a 70 µm microfluidic chip, and tagBFP and sfGFP fluorescence intensities were recorded and compensated using the respective single-colour controls. Cells were coeluted with PBS and sorted into 1.5 mL reaction tubes. Cells were plated on a 10 cm round dish with LB agar with corresponding antibiotics and grown overnight at 37 °C. 90 colonies in total were counted, while 70 individual colonies were picked and used to inoculate a 2 mL liquid culture of LB with corresponding antibiotics and one well of a 96-well Microsynth *E. coli* plate sequencing plate. Cultures were grown overnight at 37 °C, and the 96-well plate was sealed and sent for sequencing. The next day the cultures were mixed with 25 % glycerol and stored at -80 °C.

3.2.4.11. SICLOPPS screening analysed by Illumina sequencing

E. coli Top10F' cells were transformed with the sfGFP-RBP fusion protein plasmid and plasmid S6 (for SRSF1) or plasmid H8 (for hnRNP A2B1) and grown overnight at 37°C. A single colony was used to prepare a 10 mL preculture to generate electrocompetent cells according to section 3.2.1.3. Electrocompetent cells pretransformed with the protein plasmid were then transformed with 1 µL (1.19 µg/µL) of the SICLOPPS plasmid library. Cells were plated on LB agar with 50 µg/mL kanamycin, 34 µg/mL chloramphenicol and 100 µg/mL ampicillin on a 24.5 x 24.5 cm dish or four 15 cm round dishes. The next day all cells were scraped and transferred into 50 mL LB medium with corresponding antibiotics and grown for 2 h at 37 °C. Cells were then induced with 100 ng/mL anhydrotetracycline overnight at 37 °C. The culture was diluted 1/20 into 10 mL LB with corresponding antibiotic and 100 ng/mL anhydrotetracycline and grown at 37 °C until $OD_{600} \approx 0.2$. The culture was induced with 0.125 % arabinose and 1 mM IPTG for 4 h 30 min at 30 °C. Cells were transferred on ice and centrifuged at 14,000×g at 4 °C. Cells were washed 500 µL PBS, centrifuged and washing was repeated. Cells were then centrifuged and resuspended in 800 µL PBS and placed on ice until analysis. Samples were sorted with the SH800SFP Cell Sorter using a 70 µm microfluidic chip, and tagBFP and sfGFP fluorescence intensities were recorded and compensated using the respective single-colour controls. Cells were coeluted with PBS and sorted into 1.5 mL reaction tubes prefilled with 300 µL 2YT medium. Cells were plated on one

3. Material and methods

15 cm round dish with LB agar with 100 µg/mL ampicillin and grown overnight at 37 °C. The next day cells were scraped and transferred into 50 mL LB with 100 µg/mL ampicillin and grown overnight at 37°C. Plasmids were extracted as mentioned in 3.2.2.1. and samples were prepared for sequencing as described in 3.2.2.3.

3.2.4.12. Enriched SICLOPPS screening analysed by Illumina sequencing

E. coli Top10F' cells were transformed with the sfGFP-RBP fusion protein plasmid and plasmid S6 (for SRSF1) or plasmid H8 (for hnRNP A2B1) and grown overnight at 37°C. A single colony was used to prepare a 10 mL preculture to generate electrocompetent cells according to section 3.2.1.3. Electrocompetent cells pretransformed with the protein plasmid were then transformed with 1 µL (1.19 µg/µL) of the SICLOPPS plasmid library. Cells were plated on LB agar with 50 µg/mL kanamycin, 34 µg/mL chloramphenicol and 100 µg/mL ampicillin on a 24.5 x 24.5 cm dish or four 15 cm round dishes. The next day all cells were scraped and transferred into 50 mL LB medium with corresponding antibiotics and grown for 2 h at 37 °C. Cells were then induced with 100 ng/mL anhydrotetracycline overnight at 37 °C. The culture was diluted 1/20 into 10 mL LB with corresponding antibiotic and 100 ng/mL anhydrotetracycline and grown at 37 °C until $OD_{600} \approx 0.2$. The culture was induced with 0.125 % arabinose and 1 mM IPTG for 4 h 30 min at 30 °C. Cells were transferred on ice and centrifuged at 14,000×g at 4 °C. Cells were washed 500 µL PBS, centrifuged and washing was repeated. Cells were then centrifuged and resuspended in 800 µL PBS and placed on ice until analysis. Samples were sorted with the SH800SFP Cell Sorter using a 70 µm microfluidic chip, and tagBFP and sfGFP fluorescence intensities were recorded and compensated using the respective single-colour controls. Cells were coeluted with PBS and sorted into 1.5 mL reaction tubes prefilled with 300 µL 2YT medium. Cells were plated on one 15 cm round dish with LB agar with corresponding antibiotics and grown overnight at 37 °C. The next day cells were frozen with 25 % glycerol until the procedure was continued. The glycerol stock was thawed, and 500 µL were brought into liquid culture in 50 mL LB with 50 µg/mL kanamycin, 34 µg/mL chloramphenicol and 100 µg/mL ampicillin and grown overnight. Cells were diluted 1/20 in LB, grown for 2 h and induced with 100 ng/mL anhydrotetracycline overnight at 37 °C. The culture was diluted 1/20 into 10 mL LB with corresponding antibiotics and anhydrotetracycline and grown at 37 °C until $OD_{600} \approx 0.2$. The culture was then induced with 0.125 % arabinose and 1 mM IPTG and grown for 4 h and 30 min at 30 °C. Cells were transferred on ice and centrifuged at 14,000×g at 4 °C. Cells were washed 500 µL PBS, centrifuged and washing was repeated. Cells were then centrifuged and

3. Material and methods

resuspended in 800 μ L PBS and placed on ice until analysis. Samples were sorted with the SH800SFP Cell Sorter using a 70 μ m microfluidic chip, and tagBFP and sfGFP fluorescence intensities were recorded and compensated using the respective single-colour controls. Cells were coeluted with PBS and sorted into 1.5 mL reaction tubes prefilled with 300 μ L 2YT medium. Cells were plated on one 15 cm round dish with LB agar with 100 μ g/mL ampicillin and grown overnight at 37 °C. Cells were scraped and transferred into 50 mL LB with 100 μ g/mL ampicillin and grown overnight at 37 °C. Plasmids were extracted as mentioned in 3.2.2.1. and samples were prepared for sequencing as described in 3.2.2.3.

3.2.4.13. Data analysis: SICLOPPS screening analysed by Sanger sequencing

Sequencing results of the Microsynth *E. coli* plate sequencing approach were received as individual ab1-files and used for alignment against the SICLOPPS library plasmid using SnapGene to check for mutations, deletions and insertions. Varying inserts received as nucleotide sequences were translated using the web platform expasy.org.

3.2.4.14. Data analysis: SICLOPPS screening and enriched SICLOPPS screening analysed by Illumina Sequencing

Both files of the paired-end sequencing dataset, containing the forward and reverse complemented strands, were uploaded to the galaxy web platform (usegalaxy.org). First, a quality filter step was performed using the command *Filter by quality* with a “*Quality cut-off value*” of 20 and “*Percent of bases in sequence that must have quality equal to / higher than cut-off value*” of 50. Sequencing data was checked for mismatches using the *Barcode splitter* and an uploaded text file containing the 5' sequence before the variable insert. The “*Number of allowed mismatches*” was set to 0. The resulting matching files for the forward strands were downloaded and reuploaded to the platform. The datatype was changed using the command *FASTQ to FASTA*, and both forward strand files were merged using *Merge.files*. The 3' ends of the sequences were removed using the command *Clip* using the clipping adapter sequence:

```
TGCTTAAGTTTTGGCACCGAAATTTAACCGTTGAGTACGGCCCATTGCCCATGGCAA  
AATTGTGAGTGAAGAAATTAATTGTTCTGTGTACAGTGTTGATCCAGAAGGGAGAGTTT  
ACACCC
```

The “*Minimum sequence length (after clipping, sequences shorter than this length will be discarded)*” was set to 15. Then a trim of the 5' end before the varying insert was performed using *Trim* and “*Trim from the beginning up to this position*” set to 112, and the “*Is input dataset in FASTQ format?*” option was set to “*Yes*” so only the varying inserts were left behind. The

3. Material and methods

inserts were filtered by length using *Filter sequences by length* to remove inserts that were shorter or longer than 15 nucleotides setting the “*Minimum length*” and “*Maximum length*” each to 15. All unique inserts were added up using the command *Collapse* and the nucleotide sequence was translated into amino acid one-letter code using *Transeq* while “*Code to use*” was set to “*Standard*” and “*Change all STOP codon positions from the '*' character to 'X'*” was set to “*Yes*”. A schematic representation of the workflow is shown in Figure 9.

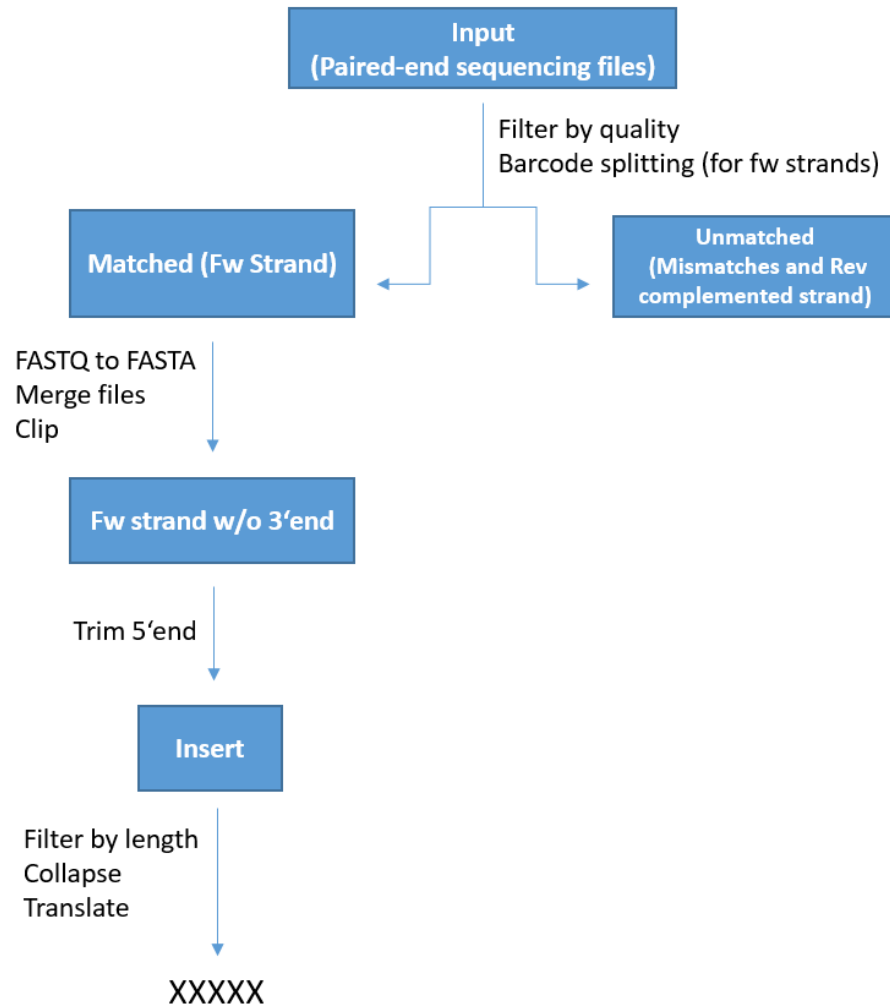


Figure 9: Schematic representation of the analysis steps of the SICLOPPS screening dataset using the Galaxy web platform. ⁶⁶

3.2.5. Protein purification

3.2.5.1 Protein expression and purification hnRNP A2B1

MBP-tagged hnRNP A2B1 (aa1-251) was expressed and purified at the Protein Chemistry Facility (PCF) based in the MPI Dortmund. HnRNP A2B1 (1-251) was sub-cloned into pOPIN-His-MBP multihost expression vectors by SLIC. The MBP fusion was chosen to increase expression yield and solubility. MBP-hnRNP A2B1 was expressed in *E. coli* BL21 CodonPlus (DE3) RIPL. Bacteria with the respective plasmid were cultured in Terrific Broth with 0.01 % lactose, 2 mM MgSO₄, 100 µg/ml ampicillin and 50 µg/ml chloramphenicol. Protein expression was auto-induced, with incubation of the starter-culture (starting OD of ~0.05) at 37 °C for 4 h, followed by an overnight incubation (20-24h) at 25 °C. Bacteria were harvested by centrifugation and lysed (50 mM HEPES, 300 mM NaCl, 20 mM Imidazole, 1 mM TCEP, pH 8) using a Celldisrupter TS 0.75 (Constant Systems) at 1350 bar. Protein purification was performed on HisTrap FF crude 5 ml Ni-based column using ÄKTA Xpress System (Cytiva, former GE Healthcare). For that, a wash buffer (50 mM HEPES, 300 mM NaCl, 30 mM Imidazole, 1 mM TCEP at pH 8) and elution buffer (50 mM HEPES, 300 mM NaCl, 500 mM Imidazole, 1 mM TCEP at pH 8) were used. The protein was then further purified using size exclusion chromatography (HiLoad 26/60 Superdex 75 prep grade column) in protein storage buffer (50 mM HEPES, 100 mM NaCl, 1 mM TCEP at pH 8.0) at 4 °C. Fractions were collected and concentrated with a 50 kDa molecular weight cutoff Amicon spin filter.⁶⁷

3.2.5.2 Protein expression and purification SRSF1

SRSF1 RRM1+2 (aa1-195) was sub-cloned into pOPIN-His multihost expression vectors by SLIC and expressed in *E. coli* BL21 (DE3) and purified using a protocol adapted from Cléry et al.^{68,69} Bacteria with the respective plasmid were cultured in LB medium with 100 µg/ml ampicillin. Protein expression was induced at OD₆₀₀=0.6 with 1 mM isopropyl β-D-thiogalactoside (IPTG) overnight at 18 °C. Cells were harvested by centrifugation and lysed (50 mM Na₂HPO₄, 300 mM KCl, 50 mM L-Arg, 50 mM L-Glu, 1.5 mM MgCl₂, 1 mM PMSF at pH 8) using a microfluidizer. Protein purification was performed on a HisTrap HP 5 ml Ni-based column (Cytiva, former GE Healthcare) using an ÄKTA Explorer System (Cytiva, former GE Healthcare). The protein was dialyzed (MWCO 3.5 kDa) overnight at 4 °C into wash buffer (50 mM Na₂HPO₄, 300 mM KCl, 50 mM L-Arg, 50 mM L-Glu, 1.5 mM MgCl₂, 40 mM Imidazole at pH 8), and a second Ni-based column purification was performed. The protein was again dialyzed (MWCO 3.5 kDa) at 4 °C into wash buffer and treated with His-tagged 3C protease.

3. Material and methods

After overnight cleavage, the protein was loaded onto a Ni-based column for reversed purification. The protein was finally dialyzed (MWCO 3.5 kDa) at 4 °C in the storage buffer (20 mM NaH₂PO₄, 150 mM KCl, 50 mM L-Arg, 50 mM L-Glu, 1.5 mM MgCl₂, 0.2 mM EDTA, 1 mM TCEP at pH 7) and concentrated with a 3 kDa molecular weight cutoff Amicon.⁶⁷

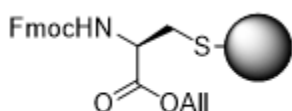
3.2.5.3. SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

15 % (w/v) SDS-gels were prepared, polymerized for 1 h and immediately used or stored at 4 °C for up to one week. Samples were mixed with 1/5 parts of 5x SDS sample buffer and loaded into the gel wells. Gel-electrophoresis was performed at 180 V and 400 mA in SDS running buffer until sufficient separation was achieved. Gel-staining was performed o/n or for 1 h at room temperature using SDS-gel staining solution. The staining solution was drained and destaining was performed with multiple washes of the gel using SDS-gel destaining solution or via o/n incubation on a shaking plate. The solution was drained and the gel was washed with H₂O and imaged.

3.2.6. Chemical synthesis

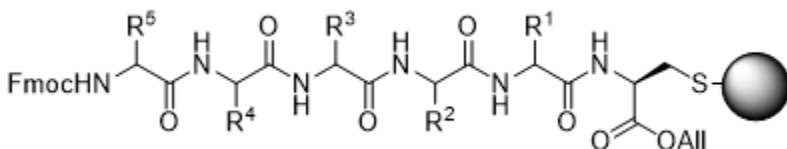
3.2.6.1. Chlorotrityl chloride resin loading

Synthesis of starting material Fmoc-L-Cys-OAll was performed by Joseph Openy.



2-Chlorotrityl chloride resin (polystyrene based; 100-200 mesh) was swollen in DCM for 2 min. The solvent was drained, and a solution of Fmoc-L-Cys-OAll (1 equiv) and DIPEA (2 equiv) in DCM was added. The resin was shaken for 2 h and the solvents were drained. The resin was suspended in DCM:MeOH:DIPEA (17:2:1 v/v) and shaken for 30 min. The solvents were removed, and the resin was washed twice with DMF and DCM and Et₂O, followed by another two washes with DCM and Et₂O to reach a maximum substitution of approx. 0.3 mmol/g.

2.3.6.2. Fmoc solid phase peptide synthesis



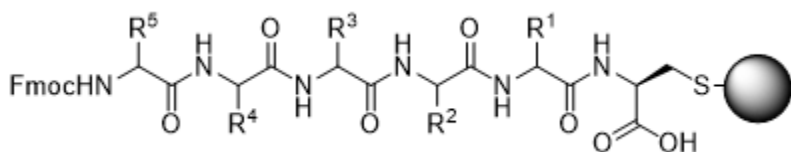
Fmoc removal was performed by addition of 20 % piperidine in DMF and agitation with argon for

5 min. After removal of the liquid, fresh 20 % piperidine in DMF was added and the resin was

3. Material and methods

agitated for 10 min. The resin was washed with DMF (4 x 30 s). Amino acid coupling was performed by using Fmoc-protected amino acid (4 equiv), PyBOP (4 equiv) and DIPEA (8 equiv) in DMF. The reaction mixture was agitated for 1 h and the liquid was drained. The resin was washed with DMF (4 x 30 seconds) and the procedure was repeated until completion of the linear peptide. Alternatively, peptide synthesis was performed automatically using the Syro I peptide synthesiser (MultisynTech GmbH, Germany).

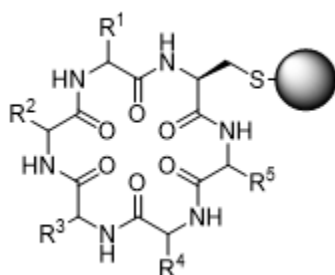
3.2.6.3. Allyl ester deprotection



The resin was washed with DCM and dry DCM. Pd(PPh₃)₄ (4 equiv) and PhSiH₃ (25 equiv) in dry

DCM were mixed. The solution was drawn into the syringe reactor and sealed with a stopcock. The syringe reactor was covered with aluminium foil and shaken for 1 h. The solution was drained and the procedure was repeated using fresh reagents. After the reaction, the solvents were removed and the resin was washed twice each with DCM, DMF, five times with 0.5 % diethyldithiocarbamic acid sodium salt in DCM for 5 minutes and twice with DMF. Further sequential washing was performed with DCM (x2), 1 M pyridinium HCl in DCM/MeOH (19:1 v/v) for 2 minutes (x3), DCM (x2) and DMF (x2).

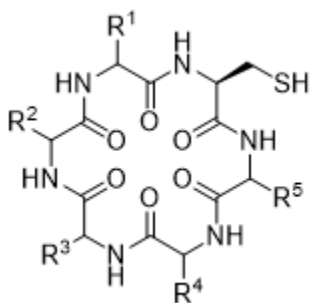
3.2.6.4. Cyclization



The terminal Fmoc protecting group was removed as described above. PyBOP (2 equiv) and Oxyma (2 equiv) in DMF and DIPEA (4 equiv) were combined and added to the syringe reactor. The reaction mixture was shaken overnight and solvents were drained. The reaction progression was checked and, if necessary, the cyclization procedure was repeated.

3. Material and methods

3.2.6.5. Cleavage



A solution of TFA/TIPS/DODT/H₂O (90:2.5:2.5:5 v/v) was prepared and drawn into the syringe reactor fitted with a stopcock. The reaction mixture was shaken for 1 h and the solution was drained through the filter into a 50 mL reaction tube containing cold Et₂O (20 ml). Fresh TFA was drawn into the syringe and shaken for 30 seconds and also added to the ether.

The procedure was repeated once more. The reaction tubes were centrifuged at 3100xg, the supernatant was decanted and fresh Et₂O was added. The mixture was shaken until the pellet was completely resuspended and centrifuged again. This step was repeated once more. All Et₂O was removed, and the pellet was dissolved in H₂O or ACN/H₂O and lyophilized.

3.2.6.6. Analysis and purification

Preparative scale HPLC purification was carried out either on an Agilent Infinity II LC-MS system equipped with a 125 mm x 21 mm, 5 μm or 125 mm x 10 mm, 5 μm Nucleodur C18 Gravity column (Macherey-Nagel GmbH & Co. KG, Germany) with a flow rate of 20 mL/min and detection at 210 nm, or on BÜCHI Pure C-850 FlashPrep equipped with a 125 mm x 10 mm, 5 μm Nucleodur C18 Gravity column (Macherey-Nagel GmbH & Co. KG, Germany) with a flow rate of 20 mL/min and detection at 210 nm.

Purity of the final peptides 44 and 50 was determined at 210 nm with an Agilent Infinity HPLC system applying the elution system: 5 % to 65 % ACN (0.1 % TFA) in H₂O (0.1 % TFA) over 14 min, using 50 mm x 3 mm, 1.8 μm Nucleodur C18 Gravity column and a flow rate of 0.56 ml/min.

For peptide 6, the Agilent Infinity II HPLC system was used equipped with 150 mm x 2.1 mm, 2.7 μm Agilent InfinityLab Poroshell 120 EC-C18, with a flow rate of 0.4 ml/min and the elution system 5 % to 95 % ACN (0.1 % TFA) in H₂O (0.1 % TFA) over 20 min.

HRMS analyses were performed using an LTQ-XL Orbitrap mass spectrometer (Thermo Fisher Scientific, USA) with electrospray ionization coupled to an Accela HPLC System (column: Hypersyl GOLD, 50 mm x 1 mm, 1.9 μm).

4. Results

4.1. A translational reporter assay for the analysis of RNA-binding protein consensus sites

Results presented in this chapter contributed to the following publication:

Jessica Nowacki, Mateo Malenica, Stefan Schmeing, Damian Schiller, Benjamin Buchmuller, Gulshan Amrahova, Peter 't Hart, A Translational Repression Reporter Assay for the Analysis of RNA-Binding Protein Consensus Sites, *RNA Biology*, **2023**, 20 (1), 85-94. ⁶⁷

Experimental data from Mateo Malenica and Gulshan Amrahova will be presented in this chapter.

There are assays for the identification and quantification of PRIs, such as the fluorescence polarisation assay, the systematic evolution of ligands by exponential enrichment (SELEX) experiment, the electrophoresis mobility shift assays (EMSA) and pulldown approaches, among others, which have been proven to be useful for several purposes. ^{70,71} *In vitro* approaches as those mentioned have the need for isolated protein and/or RNA, which can be cost-intensive, laborious and time-consuming considering the good purity and quantities needed. In addition, *in vitro* methods do not resemble the intracellular environment, so *in vitro* data should be considered with caution when predicting *in vivo* behaviour ^{72,73} To overcome these limitations, bacteria or yeast cell-based assays to quantify PRIs such as the antitermination assay, yeast-three-hybrid or bacterial-three-hybrid models as well as the translational repression assay procedure have been developed. ^{63,74–79}

In 1996 Jain and Belasco published a procedure that uses the concept of translational repression to report on protein-RNA binding. RNA constructs were cloned as DNA into β -galactosidase reporter plasmids and, together with an RBP of interest, expressed in *E. coli* cells. Once a protein bound to an RNA binding site, translational repression of the reporter gene was observed. PRI could be measured by the reduction of β -galactosidase activity reporting repression ratios (no protein/protein present). ^{78,80} Paraskeva *et al.* later termed the assay as the Translational Repression Assay Procedure (TRAP) while also replacing the *lacZ* reporter with a green fluorescent protein reporter gene making cell lysis redundant. ⁸¹

4. Results

Katz *et al.* optimized the assay further by introducing the RBP as a fluorescent protein fusion construct which allows monitoring of its expression level. Figure 10 schematically shows the principle of that variant of the assay.

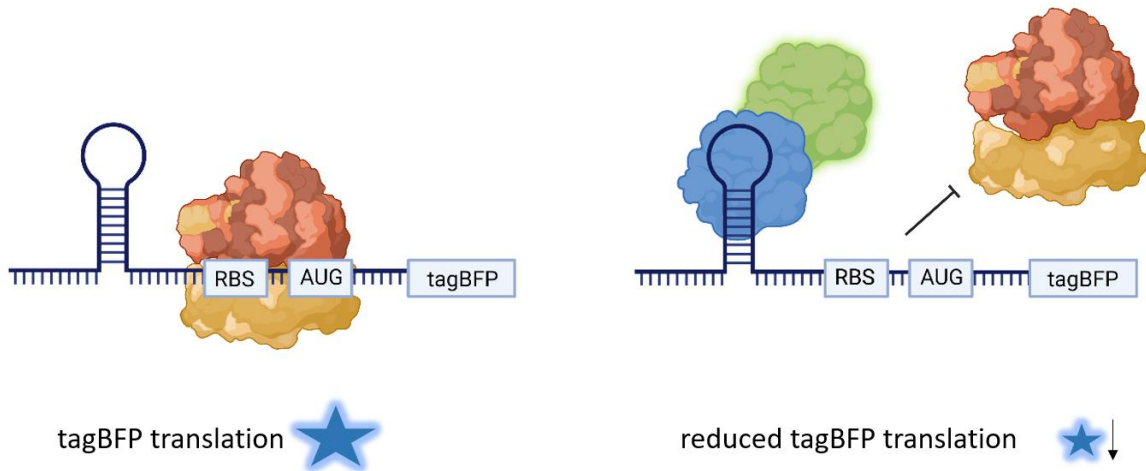


Figure 10: Principle of TRAP. Left: Reporter RNA in the absence of an RBP. Right: Binding of an RBP to its RNA target results in translational repression. Translation is initiated by the recruitment of ribosomes to the ribosomal binding site (RBS). As an improvement of the procedure, Katz *et al.* added a fluorescent tag to the RBP to monitor its expression during the measurement.^{63,82}

The assay is DNA-encoded on two plasmids on the one hand, the “RNA plasmid”, which contains the RNA insert, a Shine-Dalgarno-sequence (S/D sequence) as ribosomal binding site (RBS) and the reporter gene. On the other hand, there is the “protein plasmid”, which codes for the fluorescently tagged RBP-fusion protein. Both plasmids can be seen schematically in the following Figure 11.

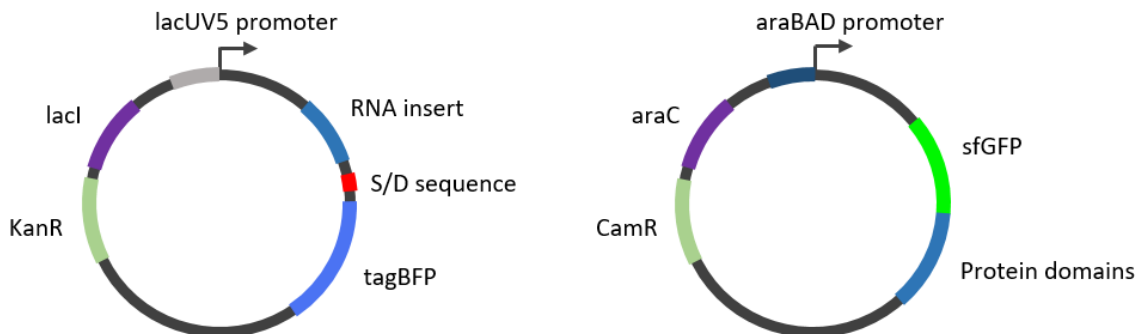


Figure 11: Schematic representation of the two assay plasmids used in TRAP. Left: The “RNA plasmid” contains the genetic information for the RNA insert, the S/D sequence, tagBFP which are under control of the lacUV5 promoter and the kanamycin resistance (KanR). Right: The “protein plasmid” contains

4. Results

the genetic information for the sfGFP-RBP fusion protein which is controlled by the araBAD promoter and the chloramphenicol resistance gene (CamR).

Both plasmids can be combined with an orthogonal induction system to control RNA and protein production independently.^{63,82} Simple readout in a plate reader assaying reporter production while also being able to monitor RBP production lifted the utility of the assay to the next level.

The two alternative splicing factors, SRSF1 and hnRNP A2B1, are attractive targets for therapeutic inhibition. For this purpose, an assay system has to be used that allows to monitor the interaction of the protein with RNA. In the following, the TRAP assay has been chosen to study the interaction of SRSF1 and hnRNP A2B1 with RNA consensus motifs.

4.1.1. Development of the TRAP assay for SRSF1

To design the protein plasmid constructs to study SRSF1 interacting with RNA, the two RRMs of SRSF1 were C-terminally fused to the green fluorescent protein sfGFP (shown in Figure 12). The RS domain was removed as it was shown not to contribute to RNA binding.²¹

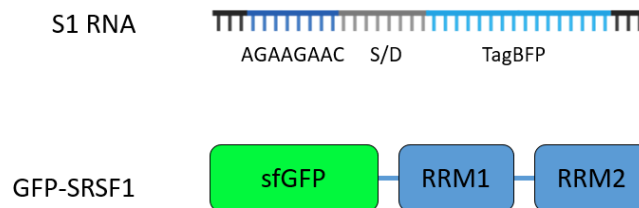


Figure 12: TRAP assay plasmid constructs. The RNA plasmid contains the RNA insert (here: AGAAGAAC), the S/D sequence and the reporter gene *tagBFP*. The protein plasmid carries the gene for the sfGFP-SRSF1 fusion protein.

In literature, SRSF1 and hnRNP A2B1 are both proposed to interact with purine-rich RNA sequences. Consensus motifs described in literature for SRSF1 are the GGAGA-motif, derived from cross-linking immunoprecipitation-sequencing (CLIP-seq) experiments, while RNA-seq experiments found the motif UCAGAGGA.^{83,84} SELEX experiments with SRSF1 RRM1 and RRM2 identified the consensus octamer RGAAGAAC (R=A or G). As a starting point, the latter was picked and inserted in front of the S/D sequence, and translational repression was observed. Placement of the binding motif upstream of the S/D sequence rather than downstream avoids the need to stay in frame with the three-letter code, thus making construct cloning more practicable. This is in contrast to the system reported by Katz et al., who inserted the RBP-binding site within the ribosomal initiation region, 11-13 nt upstream of

4. Results

the AUG start codon.⁸² Figure 12 schematically shows the construct cloning of the AGAAGAAC octamer, here named S1, into the RNA plasmid.

The reporter was induced constantly at 1 mM IPTG while the sfGFP-RBP fusion was induced with increasing concentrations of arabinose (0, 0.125, 0.25, 0.5 and 1 % arabinose). According to the methods by Katz *et al.*, translational repression can be measured by reporting production rates that are plotted against averaged sfGFP-RBP fusion protein expression levels. Reporting production rate instead of steady-state levels of the fluorescent reporter avoid saturation of the signal.^{63,82} Detailed calculation is described in the methods section 3.2.4.1. Dividing basal tagBFP production rate at non-induced sfGFP-RBP (0 % arabinose) conditions by the tagBFP production rate at the highest induction conditions (1 % arabinose) allows the calculation of a “repression ratio”. Such repression ratios have been reported before and provide a straightforward comparison of the influence of RNA-RBP pairs on translational repression.^{81,85,86} As a negative control RNA-RBP pair, an sfGFP-fusion with the RBP polypyrimidine binding tract protein 1 (PTBP1) RRM 3 and 4 together with the same RNA reporter construct S1 has been used to determine the contribution of the RBP alone on translational repression. All PTBP1 RRM (1-4), in contrast to SRSF1 or hnRNP A2B1, bind pyrimidine-rich sequences, presumably making it a suitable negative control.⁸⁷ To control the effect of the RNA insert, the no-insert negative control S0 was prepared. Figure 13 depicts TRAP assay results and the resulting repression ratio of the S0 and S1 reporter with SRSF1 and PTBP1 domains as control.

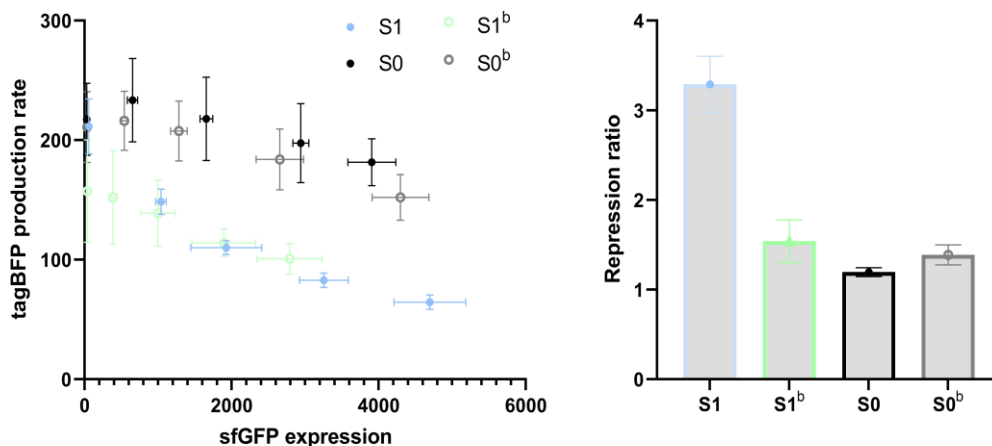


Figure 13: TRAP assay with the reporter constructs S0 and S1. Left: Repression curves for sfGFP-SRSF1 (blue and black), and sfGFP-PTBP1 (green and grey, indicated by b). Right: Repression ratios for sfGFP-SRSF1 (blue and black) and sfGFP-PTBP1 (green and grey, indicated by b). Data are mean values (n=2, N=2).

4. Results

The plot of tagBFP production rate as a function of the sfGFP expression in Figure 13 (left) shows a typical repression response for the sfGFP-SRSF1/S1 pair while the control graphs are more flattened. Construct S1 gives a repression ratio of 3.3 ± 0.3 while the no insert control, as well as the RBP control, are significantly lower (1.2 ± 0.0 and 1.5 ± 0.2 , Figure 13 and Table 4). Experiments with an sfGFP only construct give a similar result (Repression ratio 1.5 ± 0.1 in, graphs in Supplementary figure 2).

Table 4: Reporter constructs used in TRAP and the resulting repression ratios with SRSF1.

Construct	RBP binding sequence	Linker	Repression ratio	Basal tagBFP production rate (1/min)
S0	-	-	1.2 ± 0.0	217.5 ± 30.2
S1	AGAAGAAC	-	3.3 ± 0.3	211.5 ± 23.1
S2	AGAAGAACAGAAGAAC	-	9.0 ± 2.2	260.4 ± 57.3
S3	AGAAGAACAGAAGAACAGAAGAAC	-	15.1 ± 2.2	280.1 ± 15.2
S4	AGAAGAAC	AUA	4.8 ± 1.0	194.5 ± 33.1
S5	AGAAGAACAGAAGAAC	AUA	12.5 ± 1.0	314.4 ± 14.8
S6	AGAAGAACAGAAGAACAGAAGAAC	AUA	17.0 ± 0.4	301.5 ± 12.3
S6-4	AGAAGAACAGAAGAACAGAAGAAC	(AU) ₂	14.5 ± 1.6	283.3 ± 22.5
S6-5	AGAAGAACAGAAGAACAGAAGAAC	(AU) ₂ A	13.5 ± 1.1	276.2 ± 20.8
S6-6	AGAAGAACAGAAGAACAGAAGAAC	(AU) ₃	11.9 ± 2.3	294.1 ± 17.7
S6-7	AGAAGAACAGAAGAACAGAAGAAC	(AU) ₃ A	11.1 ± 2.7	265.0 ± 18.5
S6-8	AGAAGAACAGAAGAACAGAAGAAC	(AU) ₄	9.2 ± 1.6	256.5 ± 16.1
S6-9	AGAAGAACAGAAGAACAGAAGAAC	(AU) ₄ A	8.4 ± 3.1	201.5 ± 13.1
S6-10	AGAAGAACAGAAGAACAGAAGAAC	(AU) ₅	6.5 ± 0.8	155.4 ± 9.9
S7	AGAAG <u>U</u> ACAGAAGAACAGAAGAAC	AUA	14.2 ± 2.1	321.0 ± 20.2
S8	AGAAGAACAGAAG <u>U</u> ACAGAAGAAC	AUA	16.1 ± 4.0	305.3 ± 20.9
S9	AGAAGAACAGAAGAACAGAAG <u>U</u> AC	AUA	11.8 ± 1.3	316.7 ± 18.8
S10	AGAAG <u>U</u> ACAGAAG <u>U</u> ACAGAAG <u>U</u> AC	AUA	3.2 ± 0.4	219.4 ± 6.0

In the presence of the SRSF1-fusion protein, repression can be observed, however, higher differences in repression and a bigger dynamic range would be desirable to allow better discrimination. Tandem insertion of a hairpin binding motif to improve translational repression has been performed by the Saito group.⁸⁶ Denichenko *et al.* also showed that increasing the number of repeats to two or more times improves the binding affinity of SRSF1 to oligonucleotide consensus sequences.⁸⁸ Furthermore, Tacke and Manley describe that the

4. Results

insertion of three copies of AGAAGAAC in a minigene functions as a splicing enhancer sequence in an *in vitro* splicing assay, indicating high-affinity binding to SRSF1.⁸⁹

To gradually explore this effect, two and three repeats of the AGAAGAAC consensus motif were cloned into the RNA plasmid (S2 and S3) that resulted in repression ratios of 9.0 ± 2.2 and 15.1 ± 2.2 , confirming the assumptions of the above-mentioned reports (see Supplementary figure 1).

In the work of Katz *et al.*, the authors describe that distancing between the RBP binding site and the S/D sequence can influence translational repression.⁸² Therefore, a three-nucleotide linker (AUA) was inserted in between those sites resulting in the reporters S4-S6 (see Table 4 and Supplementary figure 1). Slight improvements in the repression ratios were observed, while reporter S6 gave the highest repression ratio (17 ± 0.4) so far. To further explore the influence of different linker lengths, the linker of reporter S6 was increased to up to 10 nucleotides (S6-4 - S6-10). The repression ratios were determined for all constructs and the results are shown in Figure 14 and Table 4.

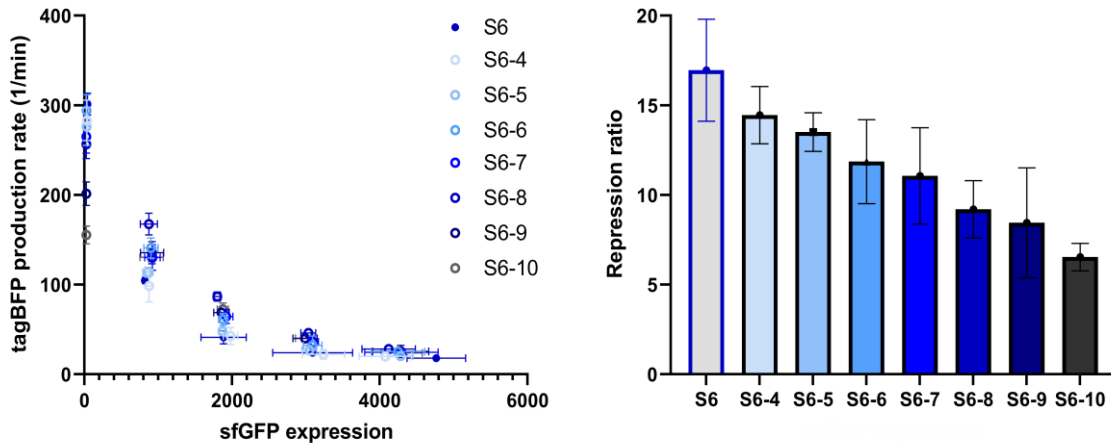


Figure 14: TRAP assay data of reporter constructs S6-4 – S6-10. Left: Repression curves with sfGFP-SRSF1. Right: Repression ratios 10 with sfGFP-SRSF1 (blue tones). Data are mean values ($n=2$, $N=2$).

A repression response is observable for all variants of S6, however, the basal tagBFP production rate is significantly lower for the construct S6-10. The maximum in repression ratio is reached using the three nucleotide linker (construct S6) while adding more nucleotides gradually reduces the repression ratio. Increasing the distance between the RBP-binding site and the S/D sequence reduces the repressive effect.

To analyse if the binding of the SRSF1-fusion protein to the RNA consensus motifs results from the two RRM, four mutations reported in literature, F54D and F58D in RRM1 and Q135A

4. Results

and K138A in RRM2, were inserted into the RRM2s with the aim to reduce binding and therefore translational repression.^{22,68} A repression ratio of 1.6 ± 0.2 was detected (see Supplementary figure 3), which is similar to the RBP-fusion controls (PTBP1 and sfGFP only) and suggests that the RBP binding originates from the RRM2s.

To get an idea of how each consensus motif repeat contributes to the RBP binding point mutations in repeat one (S7), repeat two (S8), repeat three (S9) and in all three repeats (S10) were introduced (see Table 4, mutations highlighted in red). If the mutations were in the first or second repeat (S7 or S8), the drop in translational repression was only small (14.2 ± 2.1 and 16.1 ± 4.0), but if the mutation occurred in the repeat right next to the S/D sequence the drop in repression ratio (11.8 ± 1.3) was more severe and comparable to the two repeat S5 construct (12.5 ± 1.0). Inserting point mutations in all three repeats led to a significant drop in repression (3.2 ± 0.4), similar to the repression ratio of the S1 construct. All control experiments with the sfGFP-PTBP1 fusion showed no influence on translational repression (see Supplementary figure 4).

The discovery of the mRNA as an intermediate to transfer genetic information made clear that it is rather unstable in the cellular environment.^{90,91} mRNA has a short lifetime, which adds another layer of gene expression regulation to cells and allows better adaptation to changing environments. Translation of mRNA into proteins is highly dependent on how long mRNA transcripts stay intact in the cell. Research of decades tried to unravel how mRNA degradation takes place and how it regulates gene expression.⁹¹ In this context, it had to be verified if the reduction of tagBFP fluorescence upon RBP binding is a result of translational repression rather than mRNA degradation. For this purpose, RT-qPCR experiments were performed as described in 3.2.4.4. with the sfGFP-SRSF1/S6 pair, as can be seen in Figure 15.

4. Results

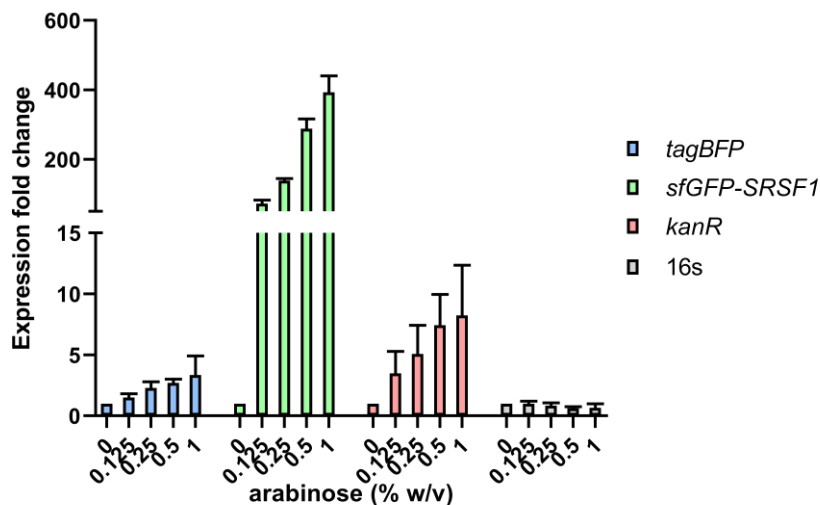


Figure 15: RT-qPCR analysis of *E. coli* cells cotransformed with the sfGFP-SRSF1/S6 plasmid pair. Gene expression levels were normalized to *gapA*. Data are mean values (n=2, N=2).

The RT-qPCR analysis shows a high and concentration-dependent increase of the *sfGFP-SRSF1* gene with increasing arabinose concentrations, as expected. Surprisingly, the increase is also observed for *tagBFP*, although this gene is not regulated by arabinose induction. The housekeeping gene, *kanR*, originating from the same plasmid as the *tagBFP* gene, also shows slight increases in fold expression. The expression of another housekeeping gene, the 16s rRNA, present on the *E. coli* genome instead is stable. As the transcript of the *tagBFP* does not show a decrease in fold expression, the reduction of tagBFP fluorescence can therefore be attributed to translational repression upon binding of the sfGFP-RBP fusion to the RBP-binding site.

4.1.2. Development of the TRAP assay for hnRNP A2B1

As the assay could be successfully developed for SRSF1, the general use of the assay should be explored for another protein. Similar to the design for SRSF1, the RRM1 and RRM2 domain with a C-terminal extension (aa 1-251) of hnRNP A2B1 was cloned into the protein plasmid as an sfGFP-fusion (see Figure 16). As hnRNP A2B1 also binds to purine-rich sequences, a 10mer sequence rich in A and G nucleotides was picked as the RNA insert that showed high binding affinity ($K_D=26.5$ nM) in an ITC experiment carried out by Wu *et al.*, 2018.²⁷

4. Results

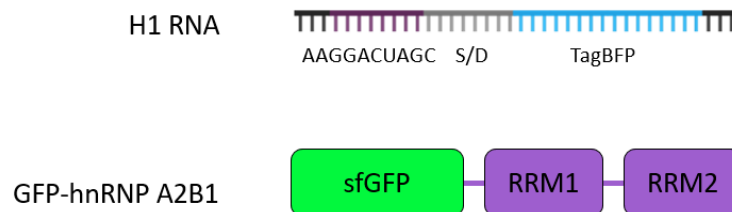


Figure 16: TRAP assay plasmid constructs. The RNA plasmid contains the RNA insert (here: AAGGACUAGC), the S/D sequence and the reporter gene *tagBFP*. The protein plasmid carries the gene for the sfGFP-hnRNP A2B1 fusion protein.

Equivalent to the SRSF1 constructs (S-series), a similar set of constructs was cloned for hnRNP A2B1 (H-series, see Table 5). Single, double and triple repeats (constructs H1-H3) were tested in the TRAP assay and gave repression ratios of 5.3 ± 0.4 , 2.7 ± 0.4 and 5.0 ± 1.1 (see Figure 17 and Table 5). Control experiments with sfGFP alone or the sfGFP-PTBP1 construct did not lead to translational repression (shown in Supplementary figure 5 and Figure 17). Different from the SRSF1 example, where multiplication of the recognition sequence increased the repression ratio, this effect is not visible for the hnRNP A2B1/RNA pairs.

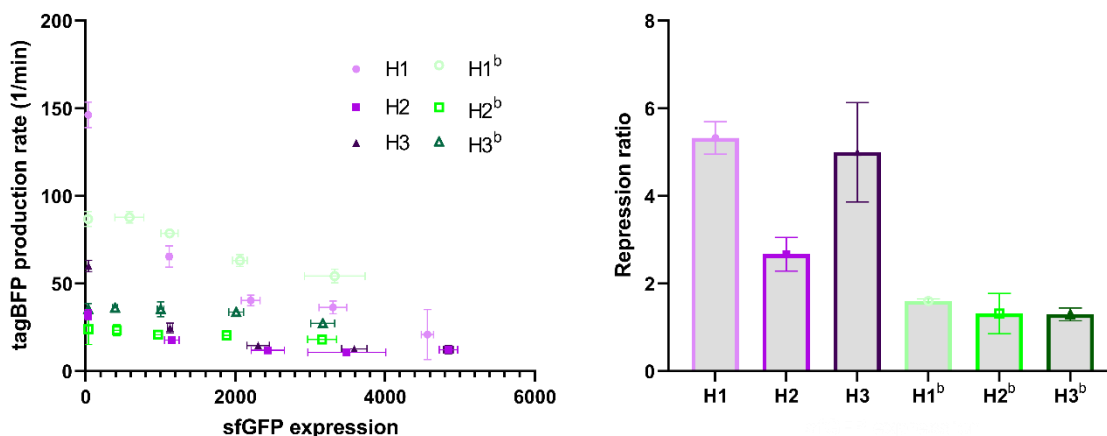


Figure 17: TRAP assay data of reporter constructs H1-H3. Left: Repression curves for sfGFP-A2B1 (pink tones) and sfGFP-PTBP1 (green tones, indicated by b). Right: Repression ratios for sfGFP-A2B1 (pink tones) and sfGFP-PTBP1 (green tones, indicated by b). Data are mean values ($n=2$, $N=2$).

Table 5: Reporter constructs used in TRAP and the resulting repression ratios with hnRNP A2B1.

Construct	RBP binding sequence	Linker	Repression ratio	Basal tagBFP production rate (1/min)
H0	-	-	1.3 ± 0.2	242.5 ± 20.5
H1	AAGGACUAGC	-	5.3 ± 0.4	146.4 ± 7.3

4. Results

H2	AAGGACUAGCAAGGACUAGC	-	2.7 ± 0.4	32.3 ± 2.9
H3	AAGGACUAGCAAGGACUAGCAAGG ACUAGC	-	5.0 ± 1.1	60.0 ± 3.3
H4	AAGGACUAGC	AUA	3.7 ± 0.2	117.4 ± 11.3
H5	AAGGACUAGCAAGGACUAGC	AUA	2.6 ± 0.6	47.7 ± 5.0
H6	AAGGACUAGCAAGGACUAGCAAGG ACUAGC	AUA	4.6 ± 0.7	56.7 ± 5.4
H7	AAGGACUAGCGGGAAAGGACUAGC	-	4.6 ± 0.2	43.3 ± 2.2
H8	AAGGACUAGCGGGAAAGGACUAGC	AUA	6.4 ± 0.1	60.2 ± 3.6
H9	AAGGACUAGCGGGAAAGGACUAGC GGGAAGGACUAGC	AUA	5.0 ± 1.7	49.5 ± 9.7
H10	AAG <u>C</u> ACUAGCGGGAAAGGACUAGC	AUA	4.6 ± 0.3	58.1 ± 6.4
H11	AAGGACUAGCGGGAAAG <u>C</u> ACUAGC	AUA	6.0 ± 0.5	61.4 ± 4.6
H12	AAGGACUAGCGGGCGAUCAGGAA	AUA	7.2 ± 0.6	163.1 ± 6.18
H13	AAG <u>C</u> ACUAGCAAGGACUAGCAAGG ACUAGC	AUA	3.1 ± 0.7	42.7 ± 2.0
H14	AAGGACUAGCAAG <u>C</u> ACUAGCAAGG ACUAGC	AUA	5.9 ± 1.0	114.9 ± 9.6
H15	AAGGACUAGCAAGGACUAGCAAG <u>C</u> ACUAGC	AUA	3.2 ± 0.4	81.0 ± 4.7
H16	AAG <u>C</u> ACUAGCAAG <u>C</u> ACUAGCAAG <u>C</u> ACUAGC	AUA	3.4 ± 0.4	270.9 ± 10.0

The influence of an AUA-linker (constructs H4-H6) was also checked in the TRAP assay (see Supplementary figure 6), but no improvement in repression ratios nor a trend was detectable.

The spacing between two recognition motifs could influence the binding of the protein to the RNA binding site, therefore, a GGG spacer was inserted between two repeats of the 10mer sequence (construct H7). The modification led to an improved repression ratio of 4.6 ± 0.7 in comparison to H2, while addition of an AUA linker (construct H8 and Table 5) even further improved the repression ratio to 6.4 ± 0.1 (see Supplementary figure 7 and Table 5). Insertion of another GGG-spacer and the third repeat (construct H9) of the recognition sequence did not follow the success.

Similar to the SRSF1 construct analysis, mutations were inserted into each repeat to study the influence on translational repression. The studies of Wu *et al.*, 2018 show that the

4. Results

replacement of G₄ to C₄ in the 10mer sequence drastically reduced the affinity (8-fold) of the protein to the RNA.²⁷ The mutation was inserted into the first and second repeat of H8, resulting in the new constructs H10 and H11. Different from the SRSF1 example, the mutations have only minimal effects on the repression ratio (see Supplementary figure 8 and Table 5).

In the TRAP assay, the RBP-fusion is expressed and released into the cellular environment freely, so it is able to bind RNA technically from any orientation. The N-terminal sfGFP fusion could apply steric constraints when placing two recognition motifs side by side, which is why the first recognition motif close to the S/D sequence was inverted (H12, see Table 5) while else keeping the design as in H8. This modification led to an increase in the repression ratio reaching a maximum value of 7.2 ± 0.6 .

Furthermore, the influence of mutations in each repeat and in all three repeats of construct H6 was also investigated (H13-H16, see Supplementary figure 9). The mutations have a slightly decreasing effect on the repression ratio that initially was only moderate while, surprisingly, construct H14 leads to an increase in the repression ratio.

Equally as for SRSF1, the no-insert control H0 (equal to S0) was measured and resulted in a repression ratio of 1.3 ± 0.2 . The control experiments with the sfGFP-PTBP1 fusion also showed no influence on translational repression (see Table 5), indicating that the reporter constructs respond to hnRNP A2B1 binding.

4.1.3. Analysis of the effect of secondary structures in the 5'UTR of the reporter mRNA

One aspect of the assay that needs to be taken into consideration is the basal tagBFP production rate which was observed to be different among all reporters of the S- and H-series, while the H reporters were significantly lower on average (see Table 4 and Table 5). The no insert controls, S0 and H0, give tagBFP production rates above $200 \times 1/\text{min}$, while almost all reporters of the S-series are higher and almost all reporters of the H-series are lower than this value. Previous reports of groups that studied the effect of translational repression in *E. coli* claim that the region where the ribosome binds must be kept free of secondary structures and associated proteins for the translation to occur efficiently.^{80,92,93} Since this could be directly reflected by the basal tagBFP production rate potential secondary structures in the 5'UTR of the reporters starting at the 5' end up until the S/D sequence were predicted with the online algorithm RNAfold.⁹⁴ Figure 18 shows the secondary structure prediction of the no-insert constructs S0/H0, where hairpin formation directly at the 5'UTR is predicted, while the 3'end

4. Results

of the construct is linear. This could be an explanation for the relatively high basal tagBFP production rate of the constructs S0 and H0.

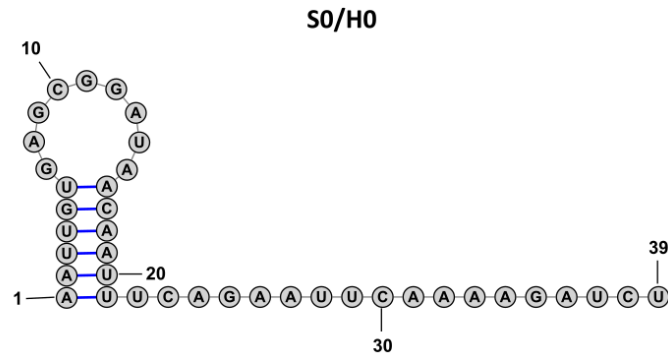


Figure 18: Secondary structure analysis for reporters S0/H0 using RNAfold.

Analogous to the S0/H0 example, the prediction was carried out for all other reporters (see Supplementary figure 10-Supplementary figure 17). All structure predictions show at least one hairpin, while several of the reporters have additional hairpins located in different positions. Comparing the constructs H7 and H8 with H12, the relatively high basal tagBFP production rate of H12 could be a result of the distancing nucleotides (3 and 6 vs 8, see Supplementary table 4) between the hairpin and the S/D sequence (compare Figure 19 with Supplementary figure 7).

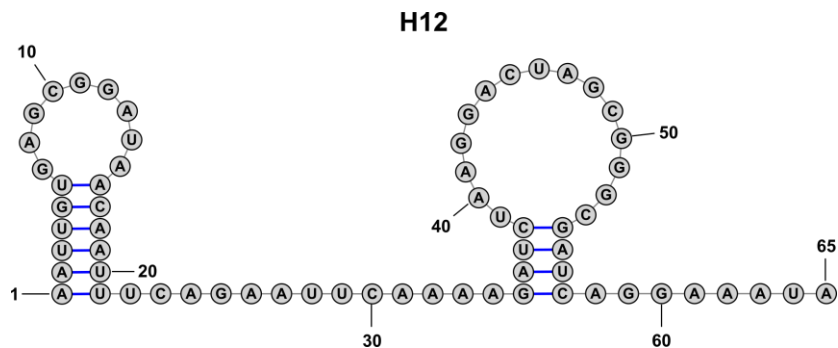


Figure 19: Secondary structure analysis for reporter H12 using RNAfold.

When plotting the distance between the hairpin closest to the S/D and the S/D sequence itself against the tagBFP production rate (see Figure 20), a trend is visible. For both construct series, a maximum of the tagBFP production rate is reached after longer distancing (21 nucleotides for SRSF1 and 16 nucleotides for hnRNP A2B1), while a decrease in tagBFP production rate starts to happen when the maximum distancing is exceeded.

4. Results

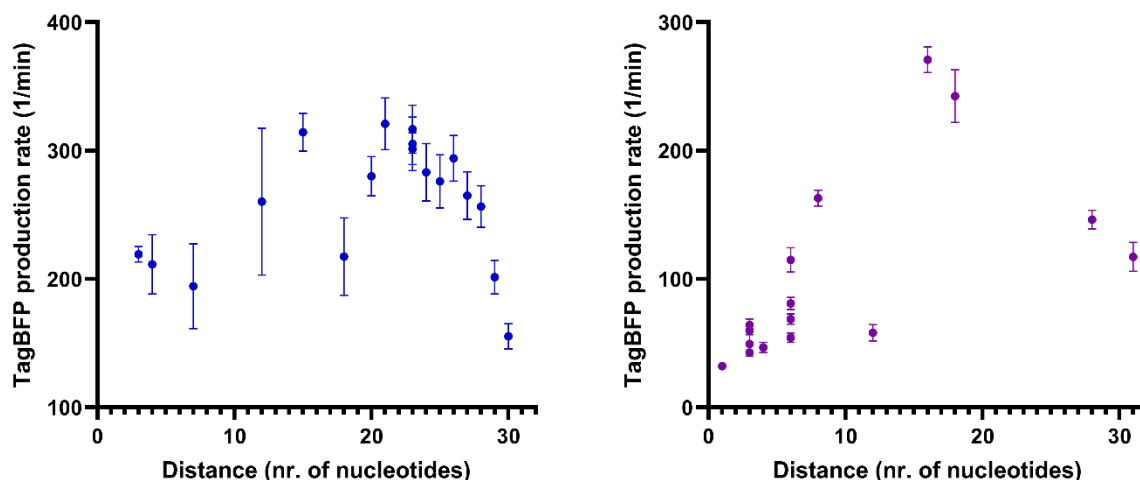


Figure 20: Correlation between basal tagBFP production rate (1/min) of all reporters for SRSF1 (left, blue) and hnRNP A2B1 (right, purple) vs the distance of nucleotides between the first occurring secondary structure and the S/D sequence.

4.1.4. Correlation of translational repression with binding affinity

Reporting repression ratios as a measure for the relative binding affinity of an RBP-fusion protein with an RNA reporter seems to be a good parameter for quick result assessment and has been done by previous groups.^{78,80} However, it remains elusive if translational repression correlates with absolute binding affinities generated by *in vitro* methods. For this reason, fluorescence polarisation (FP) assays were performed with fluorescein amine (6-FAM) labelled RNA and recombinantly expressed SRSF1 and hnRNP A2B1 using the same protein constructs as in the TRAP assay but without the N-terminal sfGFP fusion. While SRSF1 was left untagged, hnRNP A2B1 was N-terminally MBP tagged to improve purification success and yields. As RNA constructs, sequences equivalent to the reporters S1, S2, S3 and H1, H2, H3, H8 and H12 were chosen (see Supplementary figure 18-Supplementary figure 20 and Table 6). For SRSF1, the S1 construct was not bound by the protein, presumably because the RNA with 8 nucleotides is too short to be recognized by both RRM domains. Therefore, the sequence S1 was extended on both ends by an additional 3 nucleotides. On the 5' end, a part of the backbone and on the 3' end, an AUA extension was added (FAM-S1ext). This modification led to a K_D of 92.3 ± 44.1 nM, while FAM-S2 and FAM-S3 give even better binding affinities (see Figure 21 and Table 6). The results follow the logic of the TRAP results, where multiplication of repeats improves the binding affinity and, thereby, the repression ratio. The Hill slopes of the FAM-S1ext curve in the FP measurement is 1.33 ± 0.32 , while the Hill slopes for FAM-S2 and FAM-S3 are approximately halved (Table 6). This may indicate that FAM-

4. Results

S1ext can accommodate a single protein and the FAM-S2 and FAM-S3 constructs accommodate two or more protein copies.⁹⁵

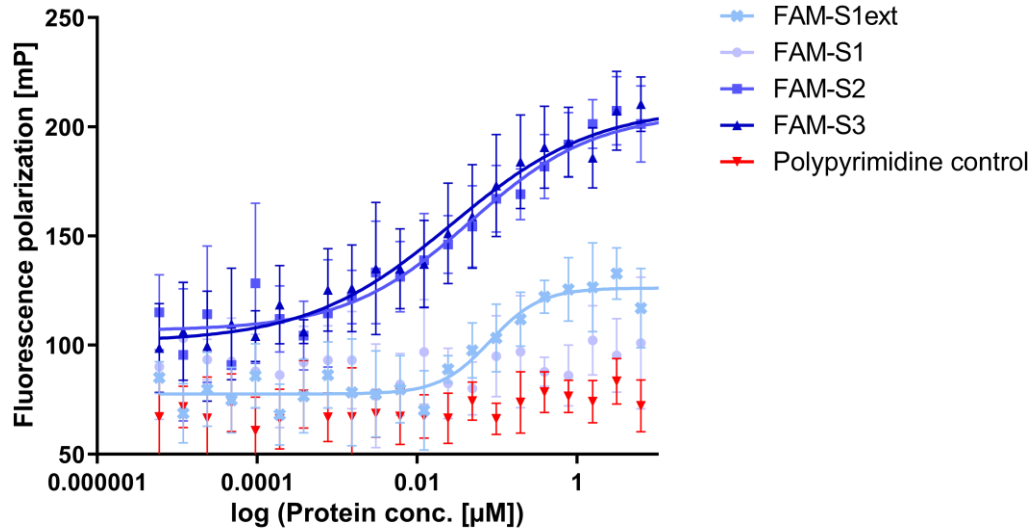


Figure 21: Fluorescence polarisation binding curves of FAM-S1, FAM-S1ext, FAM-S2, FAM-S3 and a Polypyrimidine RNA with SRSF1. Polypyrimidine RNA was used as a negative control. Data are mean values (n=3, N=2).

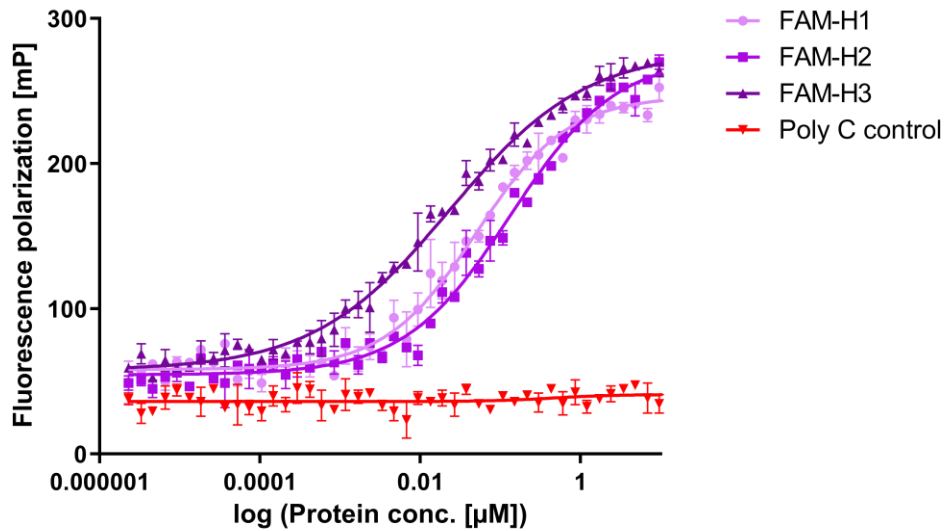


Figure 22: Fluorescence polarisation binding curves of FAM-H1, FAM-H2, FAM-H3, and Poly C RNA with MBP-hnRNP A2B1. Poly C RNA was used as a negative control. Data are mean values (n=2, N=2).

FP experiments with the FAM-H-constructs show high affinity binding for the FAM-H1 and FAM-H3 RNA, while reduced binding affinity is observed for the FAM-H2 constructs (see

4. Results

Figure 22 and Table 6). These findings also correlate well with the TRAP results, where moderate repression for constructs H1 and H3 are visible, while H2 shows only low repression. In general, it is surprising that construct H2 has a reduced binding affinity, in FP and TRAP, in comparison to construct H1. A similar effect as for SRSF1 would be expected where multiplication of repeats improves binding affinity. Secondary structure prediction for FAM-H1-3 (see Supplementary figure 19) shows that FAM-H2 and FAM-H3 contain secondary structures that could impair protein binding. However, construct FAM-H3 does contain the AGGACU sequence in a non-base paired form, which is recognized by the two RRM in the crystal structure by Wu *et al.*, 2018 and seems to be sufficient to show good binding in FP.²⁷ For FAM-H2, a few nucleotides of the AGGACU sequence are embedded into the stem of the secondary structure which could have resulted in lower binding affinity.

FP measurements of the constructs H8 and H12, which gave the highest repression values for the H-series, also follow the trends in absolute binding affinity. The secondary structure predictions are very similar for H8 in both TRAP construct context and FP (see Supplementary figure 7 and Supplementary figure 20). H8 possesses an AGG motif in its terminal loop that, in comparison to H2, is in a more relaxed conformation (7 vs 6 nucleotides in the terminal loop). It remains unclear why binding of H8 is slightly favoured over H3 according to the measured K_D s. H12 gives a very high K_D in the FP measurement, which could be explained through its structure analysis in Supplementary figure 20, which exhibits a fully linear construct.

The Hill slopes of all measured curves for hnRNP A2B1 do not differ as evident as for SRSF1, which makes assumptions about the number of proteins that could potentially bind difficult. Possibly, the number of bound proteins is similar for each tested RNA construct. It can only be hypothesized that the reason for the high-affinity binding of hnRNP A2B1 to H12 could lie in the fact that two or more protein copies could bind to the construct. Alternatively, the RNA secondary structure could lead to less flexibility and, therefore, to a smaller entropic penalty upon binding of the protein, which could improve binding affinity.

4. Results

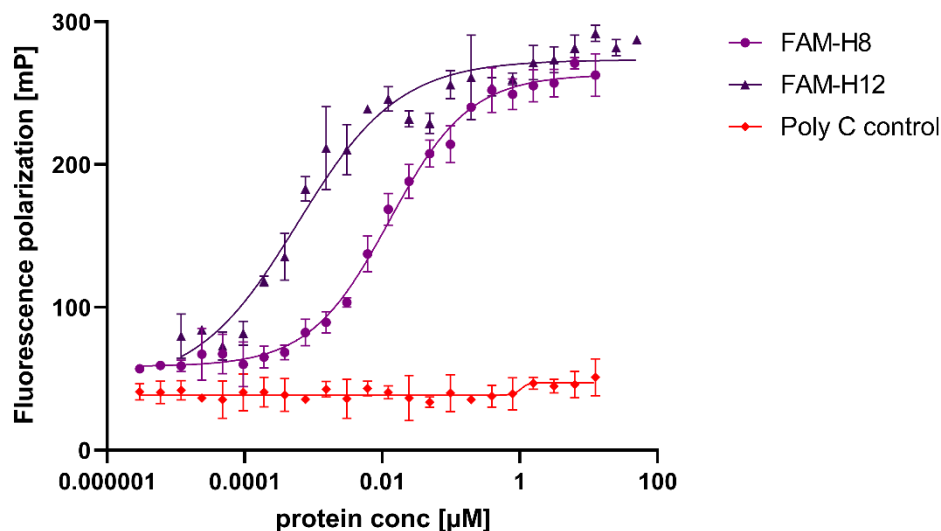


Figure 23: Fluorescence polarisation binding curves of FAM-H8, FAM-H12 and Poly C RNA with MBP-hnRNP A2B1. Poly C RNA was used as a negative control. Data are mean values (for FAM-H8: n=2, N=2; for FAM-H12: n=1, N=2).

Table 6: Affinities of SRSF1 and hnRNP A2B1 for representative RNA sequences measured by fluorescence polarisation.

RNA	Protein	K _D (nM)	Hill slope
FAM-S1	SRSF1	>25000	-
FAM-S1ext	SRSF1	92.3 ± 44.1	1.33 ± 0.32
FAM-S2	SRSF1	56.1 ± 17.5	0.62 ± 0.30
FAM-S3	SRSF1	29.1 ± 15.7	0.53 ± 0.11
FAM-H1	hnRNP A2B1	51.7 ± 9.0	0.75 ± 0.09
FAM-H2	hnRNP A2B1	134.8 ± 9.0	0.66 ± 0.05
FAM-H3	hnRNP A2B1	24.0 ± 8.1	0.52 ± 0.04
FAM-H8	hnRNP A2B1	13.0 ± 2.1	0.75 ± 0.02
FAM-H12	hnRNP A2B1	0.5 ± 0.1	0.65 ± 0.35

Plotting the binding affinity measured by FP against the repression ratios measured by the TRAP assay for the constructs containing the AUA-linker (S4-6 and H4-6, H8 and H12) shows a correlation for both the SRSF1 and the hnRNP A2B1 constructs as visible in the following Figure 24. The correlation for SRSF1 is very precise ($R^2=0.995$), indicating that TRAP assay

4. Results

construct design is straightforward and that repression ratios are directly reflected by absolute binding affinities measured by FP. For hnRNP A2B1, the correlation graph is less precise ($R^2=0.775$), which could demonstrate that construct design is highly sequence dependent. Constructs of the H-series, in comparison to the S-series, harbour more secondary structures whose formation is context-dependent. When comparing construct H12 (depicted in Figure 19) with the FAM-labelled constructs used in FP (see Supplementary figure 20), formation of the secondary structure does not take place at all. For clearer assay-readout secondary structure formation, therefore, should be considered when designing constructs for TRAP and FP.

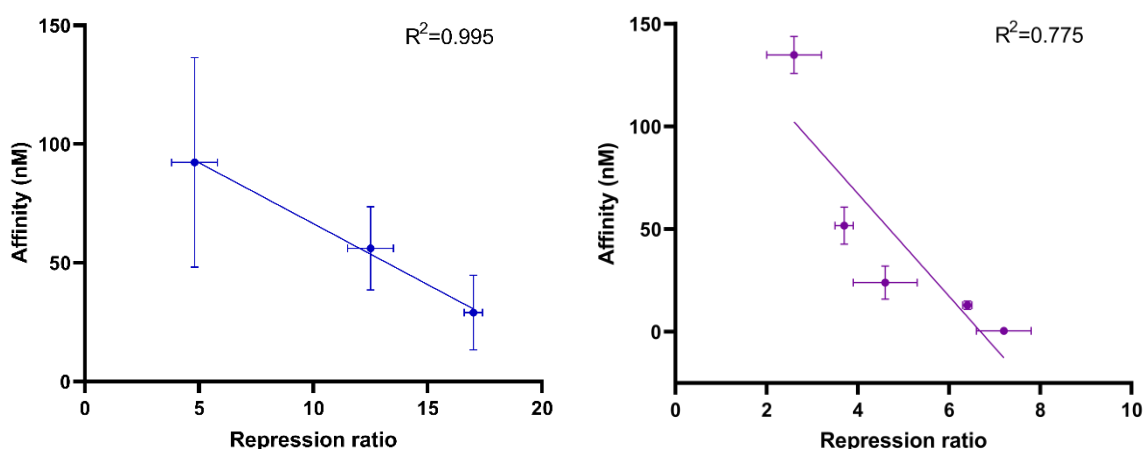


Figure 24: Repression ratios measured with the TRAP assay correlate with binding affinities generated by FP. Left: Correlation graph for SRSF1; Right: Correlation graph for hnRNP A2B1.

Nevertheless, TRAP assay results given in repression ratios are good indicators to assess relative binding affinity and the findings described above suggest that it can be used reliably to study RBP-RNA binding.

4.1.5. Analysis of translational repression by flow cytometry

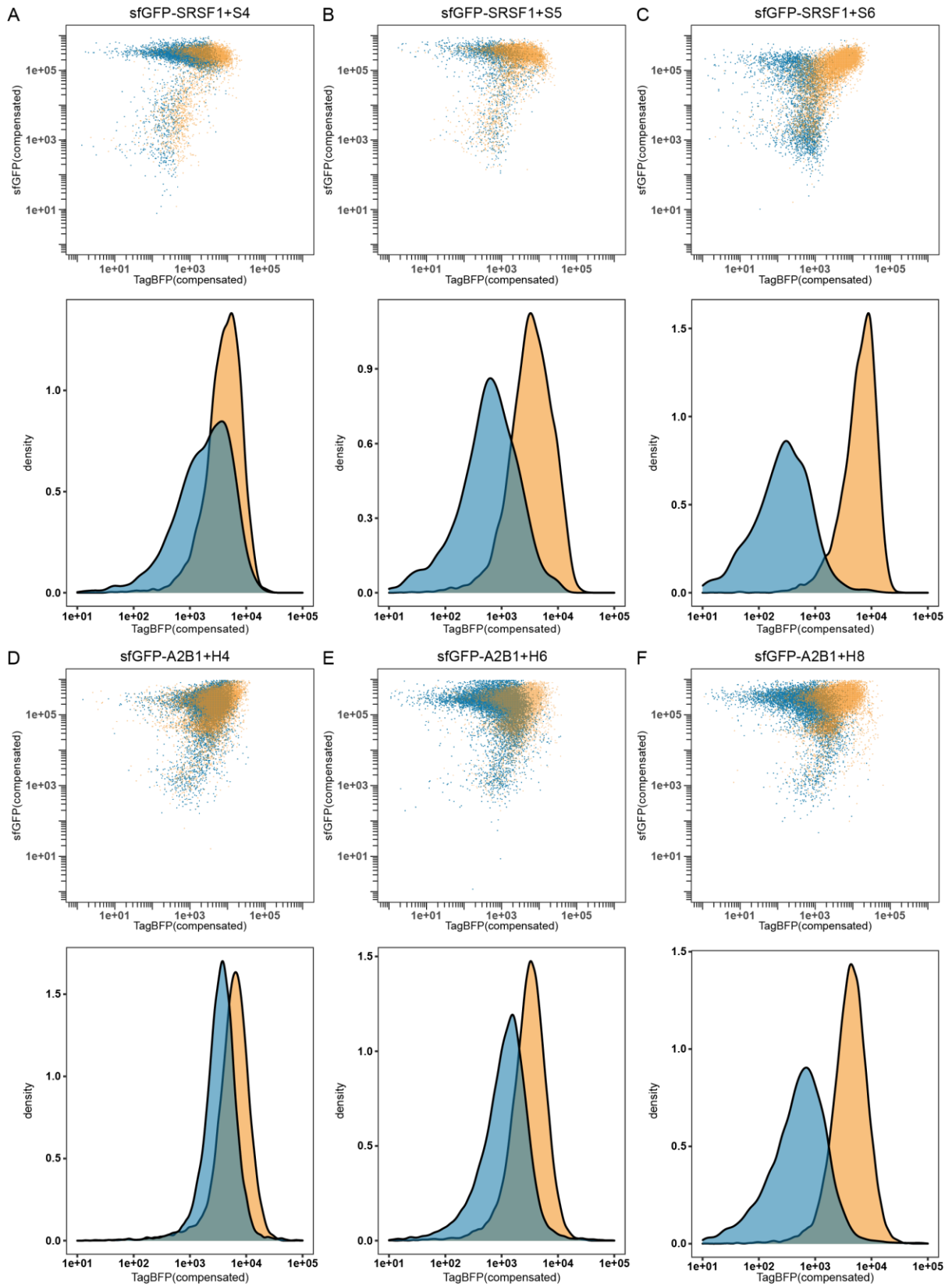
The use of the TRAP assay to study RBP-RNA interactions was successfully performed using repression ratios to evaluate RBP-RNA binding. Repression ratios are calculated using tagBFP production rate within a certain time interval. Paraskeva *et al.* used flow cytometry to observe translational repression as an endpoint assay, however, only the fluorescence reporter was taken into account.⁸¹ In the following, sfGFP-RBP/RNA construct pairs will be evaluated using flow cytometry, including two fluorescent channels to analyse the fluorescent properties of a bacterial cell population carrying both assay plasmids. For SRSF1, reporter

4. Results

constructs S0-S10 were analysed, while for hnRNP A2B1, constructs H0-H11 were measured (see Supplementary figure 23-Supplementary figure 28 and Figure 25). As control populations, the corresponding reporter was combined with the sfGFP-PTBP1 fusion protein construct (termed sfGFP-PTBP1). Population shifts were more pronounced when comparing the histograms at sufficiently high sfGFP expression levels ($>1 \times 10^5$) to reflect good expression levels of the RBPs. Figure 25 shows that the populations carrying the reporters S4-S6 can be distinct from the control population and an increase in repression ratio is also reflected by a larger distance between the peaks in the histograms of the repressed and non-repressed population. Construct S6 gave the highest repression ratio among all constructs analysed in this study, which can also be seen in the flow cytometry analysis as the two populations are almost completely separated. Similar effects can be observed for the constructs S0-S3 and S7-S10, where measured repression ratios seem to correlate with population changes (see Supplementary figure 21 and Supplementary figure 24).

Analysis of the H4, H6 and H8 reporters in combination with sfGFP-hnRNP A2B1 or sfGFP-PTBP1 showed similar shifts. For construct H4, the population changes are minimal, while constructs that gave higher repression ratios also show bigger population changes (see Figure 25). The remaining constructs H0-H3, H7 and H9-H11 in Supplementary figure 25 and Supplementary figure 28 show smaller shifts still correlating with repression ratios as these were lower for the H-series. Relatively high repression ratios of about 6 and higher seem to be required for good discrimination between populations in flow cytometry.

4. Results



4. Results

Figure 25: Flow cytometry results for sfGFP-SRSF1 (blue) or sfGFP-PTBP1 (orange) in combination with reporter S4 (A), reporter S5 (B) or reporter S6 (C). As well as sfGFP-A2B1 (blue) or sfGFP-PTBP1 (orange) in combination with reporter H4 (D), reporter H6 (E), or reporter H8 (F). Histograms were produced by analysing all events $>1 \times 10^5$ in the sfGFP channel.

4.2. Using the TRAP assay to screen for RNA consensus sequences for RBPs

Experimental data from Luisa Kurzweg and Yevfaliia Kopytsia will be presented in this chapter.

Cellular processes such as RNA splicing, localization, stability and translation are regulated by RBPs which specifically recognize short RNA sequence elements and most regulatory steps in gene expression are influenced by the sequence of the RNA.⁹⁶ RBPs participate in the formation of ribonucleoprotein complexes by binding to sequence or structural motifs in RNA via RNA-binding domains (RBD). Moreover, proteins without classical RBDs have been found to take part in PRIs, such as in ribosomal or spliceosomal complexes.^{97,98} Both the role of RBPs and RNA in various cellular processes have stimulated interest in studying RBPs and their target RNAs.⁹⁹ Ribonucleoprotein immunoprecipitation (RIP) and cross-linking and immunoprecipitation (CLIP) variants, as well as SELEX, are techniques that allow the identification of RNA targets of an RBP.^{70,99,100} While RIP and CLIP, in combination with deep-sequencing technologies (-seq), identify PRIs in cellular context, SELEX's *in vitro* properties allow the identification of RNA sequences from a random oligonucleotide pool, expanding the number of interacting RNAs. Computational approaches such as SSMART that use synthetic datasets generated from experimental data complement available approaches.¹⁰¹

The TRAP assay was presented as an excellent tool to monitor the RBP-RNA binding in cellular context providing repression ratios as a reliable parameter that correlates well with absolute binding affinities. As described above, the TRAP assay makes use of two plasmids, the “protein” and the “RNA” plasmid, which both can easily be transformed and manipulated by cloning techniques. The assumption lies close to hand that it may be used in a high throughput format to select preferred RNA sequences by using a library of the “RNA” plasmid.

In literature, consensus sequence motifs of RBPs, such as members of the SR and hnRNP family, are usually of short lengths (3-10mer sequences).^{27,83,89,102–104} To follow TRAP assay design as presented above, 10mer sequences were chosen. The above-presented results of the RNA motifs designed for the TRAP assay also revealed that usage of a three-nucleotide AUA-linker improves translational repression. Therefore a randomized N₁₀-AUA insert was cloned upstream of the S/D sequence of the S0/H0 plasmid and transformed into *E. coli* cells

4. Results

carrying the sfGFP-RBP fusion plasmid, sfGFP-SRSF1 or sfGFP-hnRNP A2B1. When using 10mer inserts, a library diversity of $4^{10}=1,048,576$ is potentially possible. However, checking the colony count via two independent dilution plating experiments after transformation led to a count of $\sim 7 \times 10^5$ colonies, while repetition of the experiment gave $\sim 2 \times 10^5$ colonies. Sequencing the library with a sequencing depth of ~ 2 million reads revealed $\sim 1.13 \times 10^5$ unique sequences. It has to be considered that the sequencing depth is probably not sufficient to cover the entirety of the plasmid pool, as a 10-fold coverage is necessary to allow minimization of the effects of PCR amplification and sequencing errors.¹⁰⁵ Nevertheless, the library was used for further experiments since further optimization of cloning and transformation efficiency was not successful at this stage.

The above-presented results demonstrate that high-affinity binding PRIs lead to translational repression resulting in low tagBFP production rates and relatively high sfGFP expression. Equivalent effects can also be observed with end-point measurements using flow cytometry. Therefore, screening was performed using fluorescence-activated cell sorting (FACS) gating for the repressed phenotype (reduced tagBFP fluorescence levels and high sfGFP fluorescence levels), as portrayed in Figure 26.

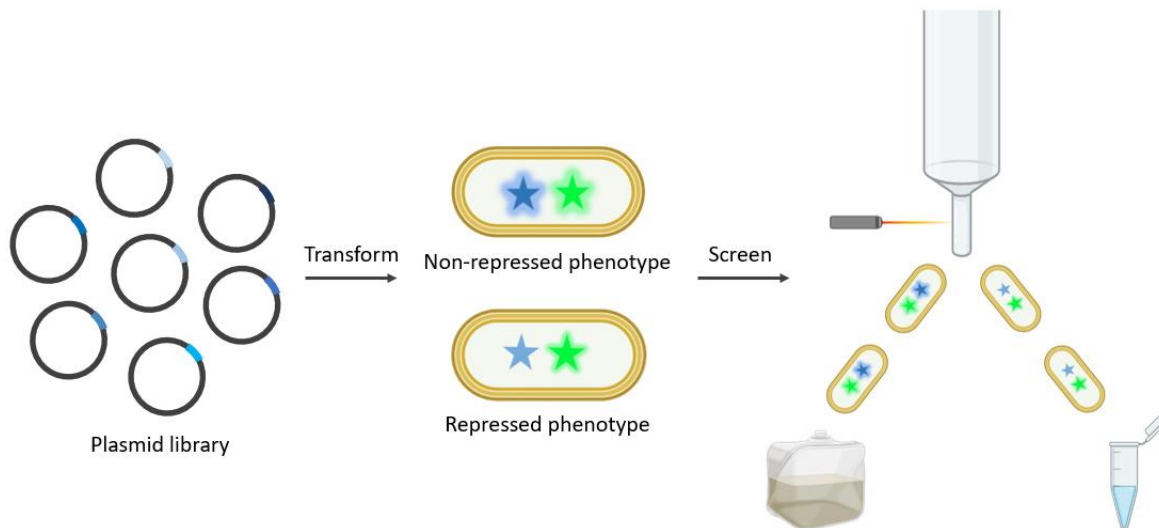


Figure 26: Concept of the RBP consensus sequence screening. A plasmid library with a randomized 10mer insert is transformed into *E. coli* cells carrying the sfGFP-RBP fusion plasmid and screened using flow cytometry. Cells with the repressed phenotype are sorted and sequenced. Illustration was created with Biorender.com.

4. Results

Since SRSF1 and hnRNP A2B1 are now well characterized with the TRAP assay, and a variety of consensus motifs are presented in literature, the two proteins have been picked for proof of concept of the screening.

4.2.1. RBP consensus sequence screening for SRSF1 and hnRNPA2B1

Considering the repression response of cells carrying the protein/RNA plasmid pair upon treatment with the induction agents, it becomes evident that high-affinity interactions already respond to low arabinose concentrations (0.125%), resulting in strong translational repression. Moreover, milder treatment conditions resemble less cellular stress as less protein is expressed and should lead to better recovery of the cells and less cell death compared to higher concentration treatment, so 0.125 % arabinose and 1 mM IPTG were chosen as the screening conditions. As a theoretical diversity of $\sim 1 \times 10^6$ plasmids is possible, a gated event count of 10×10^6 cells was chosen for all screening approaches (further sorting statistics can be seen in Supplementary table 7). Firstly, SRSF1 was used for the consensus sequence screening, demonstrated in Figure 27.

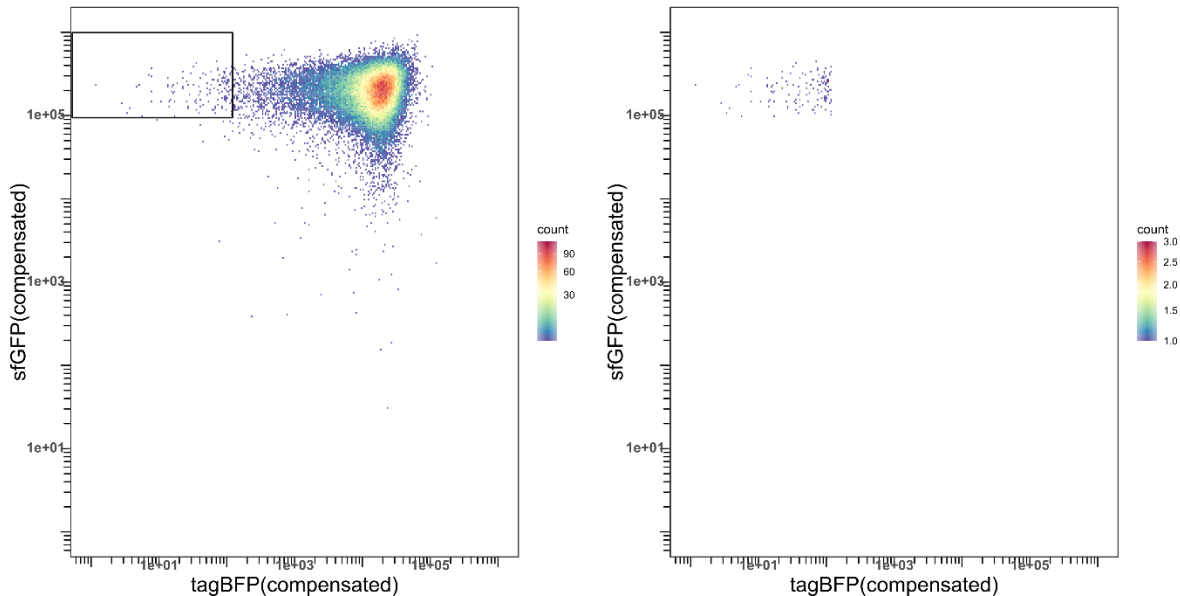


Figure 27: SRSF1 consensus sequence screening. Cells carrying the sfGFP-SRSF1 plasmid were transformed with the 10mer plasmid library and were treated with 1 mM IPTG and 0.125 % arabinose. Left: Whole population of the cells in the analysis. Cells that produced high sfGFP levels and low levels of tagBFP were gated for sorting (black box). Right: Sorted cells with gated fluorescence properties.

The density plot shows a widespread population that is centred towards high tagBFP levels between 1×10^4 and 1×10^5 and high sfGFP levels between 1×10^5 and 1×10^6 , indicating that the majority of cells do not show a repressed response upon induction. A fraction of events is

4. Results

spread towards lower tagBFP levels ($<1 \times 10^3$), possibly indicating repressed reporter production. A gate was chosen based on a control population, carrying the sfGFP-SRSF1/S0 plasmid pair as a negative control (visible in Supplementary figure 30). Events that overlapped with the control population were excluded as they were assumed not to show a repressed phenotype. Also, events that were $<1 \times 10^5$ in the sfGFP level were excluded as a sufficiently high sfGFP expression level was desired as an indicator of good SRSF1 expression. These considerations resulted in the gate visible in Figure 27, which comprises 2.66 % of the gated event count.

TRAP experiments with the reporters of the H-series showed that secondary structure formation might lead to low basal tagBFP production rates, which could give false positive hits when gating for the screening phenotype (reduced tagBFP levels, high sfGFP-RBP levels). For this reason, a control population termed “Autorepressor sequences” was sorted, which was not induced with arabinose to identify sequences that intrinsically result in low tagBFP levels and therefore produce false positive hits (see Figure 30).

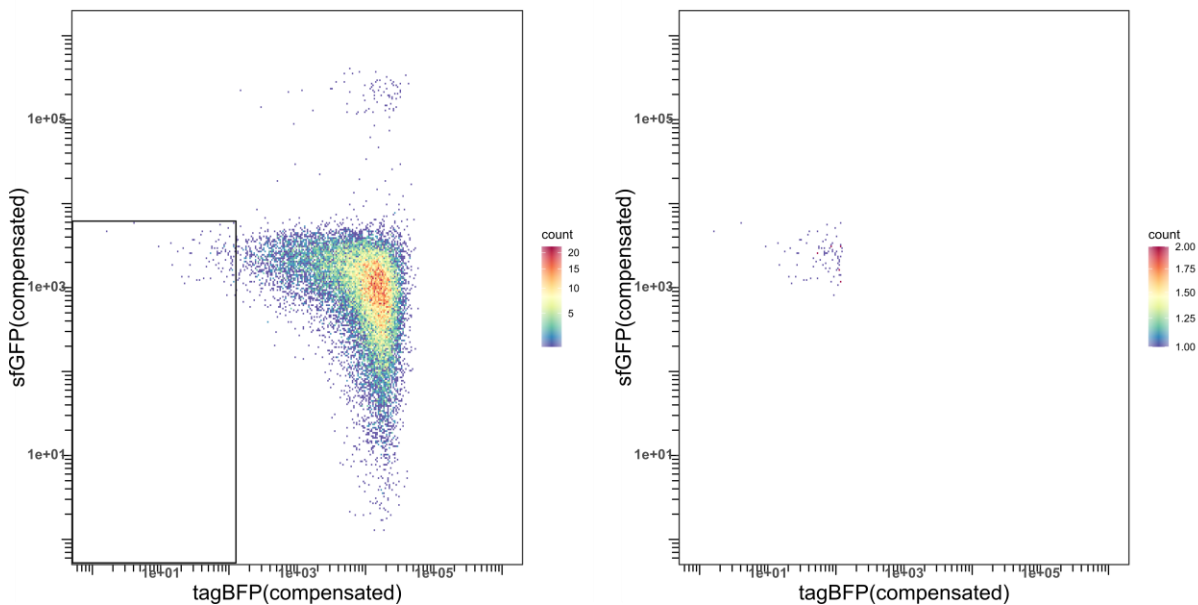


Figure 28: Autorepressor sequence selection using FACS. Cells carrying the sfGFP-SRSF1 plasmid were transformed with the 10mer plasmid library and were treated with 1 mM IPTG but no arabinose. Left: Whole population of the cells in the analysis. Cells that produced low sfGFP and tagBFP levels were gated for sorting (black box). Right: Sorted cells with gated fluorescence properties.

The density plot shows a widespread population that is centred towards high tagBFP levels between 1×10^3 and 1×10^5 and low sfGFP levels ($<1 \times 10^2$) is visible. As sfGFP was not induced, the majority of cells do not show a repressed response, while a small fraction of

4. Results

events is spread towards lower tagBFP levels ($<1 \times 10^3$), possibly indicating autorepressed reporter production. The autorepressor gate was chosen based on the gate of the SRSF1 consensus sequence screening gate (see Figure 27) downshifted towards lower sfGFP levels to sort cells with the same tagBFP fluorescence properties and, therefore, a false positive phenotype, comprising 0.58 % of the gated event count. The Autorepressor sequences were then subtracted from the datasets of the performed screenings to remove false positive hits. The hit list with the first 10 most abundant sequences and the first 10 autorepressor corrected (AC) sequences are listed in Table 7. The obtained hit sequences were each cloned into the S0 plasmid, and the PRI was evaluated using the TRAP assay, while the sfGFP-SRSF1/S6 and the sfGFP-PTBP1/S6 interaction were used as positive and negative controls. Resulting repression ratios, basal tagBFP production rates and sfGFP expression can be seen in Figure 29- Figure 31 and Table 7.

Table 7: Hit sequence constructs of the SRSF1 consensus sequence screening used in TRAP and the resulting repression ratios. The negative control measurement was performed with the sfGFP-PTBP1 fusion protein.

Construct	RBP binding sequence	Linker	Repression ratio	Basal tagBFP production rate (1/min)	Sequence abundance (%)
1	ACCAUGCUCU	AUA	1.2 ± 0.1	136.3 ± 14.6	4.64
2	CUAAUACUCG	AUA	1.4 ± 0.0	151.3 ± 12.5	2.61
3	AAAUGAGUGC	AUA	0.9 ± 0.1	73.4 ± 6.7	1.86
4	GCCUGAGCUC	AUA	0.2 ± 0.4	1.9 ± 3.1	1.85
5	CAUAAUGCAA	AUA	1.3 ± 0.2	144.3 ± 15.3	1.75
6	CACAACAUAG	AUA	1.3 ± 0.1	90.2 ± 4.8	1.73
7	CGACGCGUAA	AUA	1.7 ± 0.1	168.3 ± 12.2	1.72
8	GUUUGAGCAA	AUA	1.4 ± 0.3	154.1 ± 9.0	1.68
9	UCGUCCACCA	AUA	0.8 ± 0.1	121.2 ± 17.0	1.59
10	GGCCUCACUC	AUA	0.6 ± 0.1	4.9 ± 1.8	1.51
1_AC	GAAUACAGAU	AUA	1.2 ± 0.2	127.6 ± 10.4	1.03
2_AC	AGUUUGUCUC	AUA	0.4 ± 0.4	3.0 ± 3.1	0.96
3_AC	UCAUAGCACG	AUA	1.8 ± 0.1	186.9 ± 15.9	0.74
4_AC	UUAGAAUGAG	AUA	1.5 ± 0.2	103.9 ± 12.4	0.69
5_AC	UUAUGCUACC	AUA	1.3 ± 0.1	157.7 ± 4.0	0.66
6_AC	UAUCAUAUAC	AUA	1.4 ± 0.3	189.9 ± 7.7	0.62

4. Results

7_AC	ACCAAAUUG	AUA	1.2 ± 0.2	230.6 ± 13.8	0.61
8_AC	CGGACCAACU	AUA	1.7 ± 0.1	159.4 ± 41.1	0.60
9_AC	UAACGGAGCC	AUA	1.5 ± 0.2	42.8 ± 3.3	0.58
10_AC	AACCUCGCAC	AUA	0.1 ± 0.4	1.6 ± 1.7	0.56
Positive control	AGAAGAACAGAA GAACAGAAGAAC	AUA	16.1 ± 1.9	205.2 ± 5.2	-
Negative control	AGAAGAACAGAA GAACAGAAGAAC	AUA	1.7 ± 0.1	198.0 ± 12.2	-

Surprisingly, neither the initial hit sequences nor the autorepressor corrected sequences do respond to SRSF1 binding in the TRAP assay as no repression ratio significantly above the negative control value can be detected, while both controls result in similar repression ratios as described above. SfGFP expression levels of all tested sequences are above 3000, indicating sufficient sfGFP-RBP expression. Basal tagBFP production rates, however, show strong fluctuation ranging from ~1.9-230.6×1/min. Samples 4 and 10 are particularly low in tagBFP production rate, possibly indicating an autorepressor sequence. When applying the autorepressor subtraction, all sequences 1-10 were removed from the data set resulting in the entries 1_AC-10_AC. The basal tagBFP production rates of the sequences are, on average, slightly higher than those of the initial hit list, however, strongly autorepressed tagBFP production rates are still present (compare samples 2_AC, 9_AC and 10_AC in Figure 30 and Table 7).

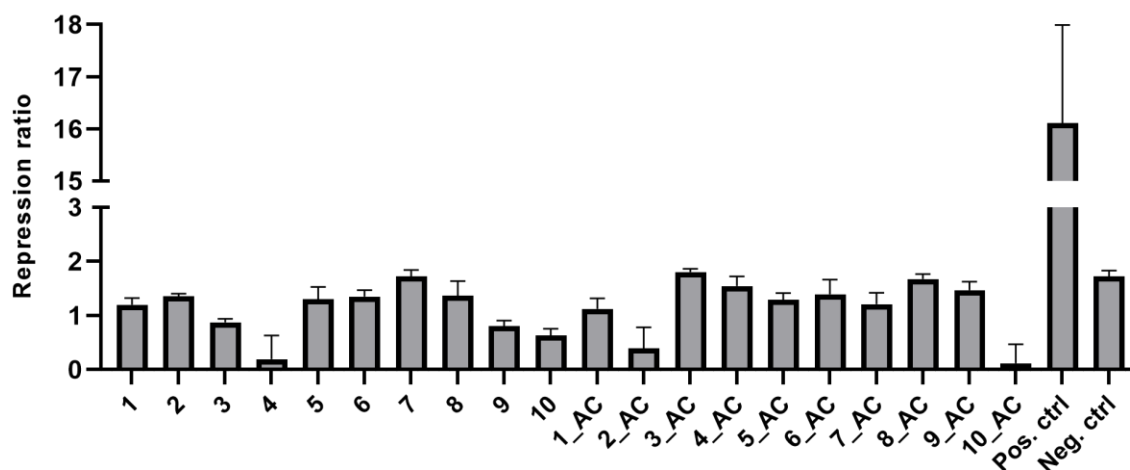


Figure 29: Repression ratios of the hit sequences 1-10 and the autorepressor corrected samples 1_AC-10_AC for the SRSF1 consensus sequence screening. Data are mean values (n=2, N=2).

4. Results

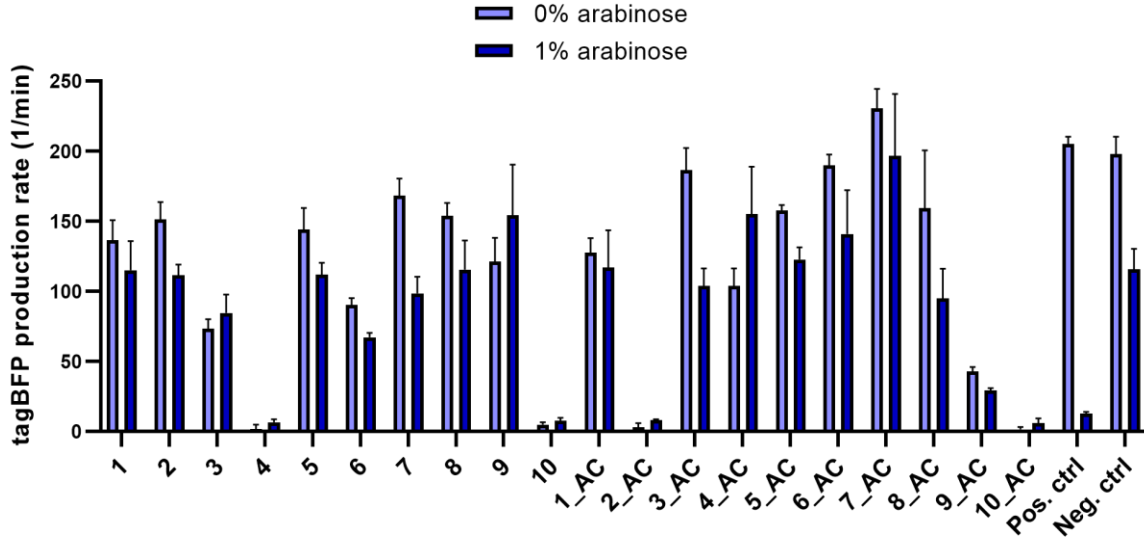


Figure 30: TagBFP production rate [1/min] of the hit sequences 1-10 and the autorepressor corrected samples 1_AC-10_AC upon arabinose induction at 0 or 1 % for the SRSF1 consensus sequence screening. Data are mean values (n=2, N=2).

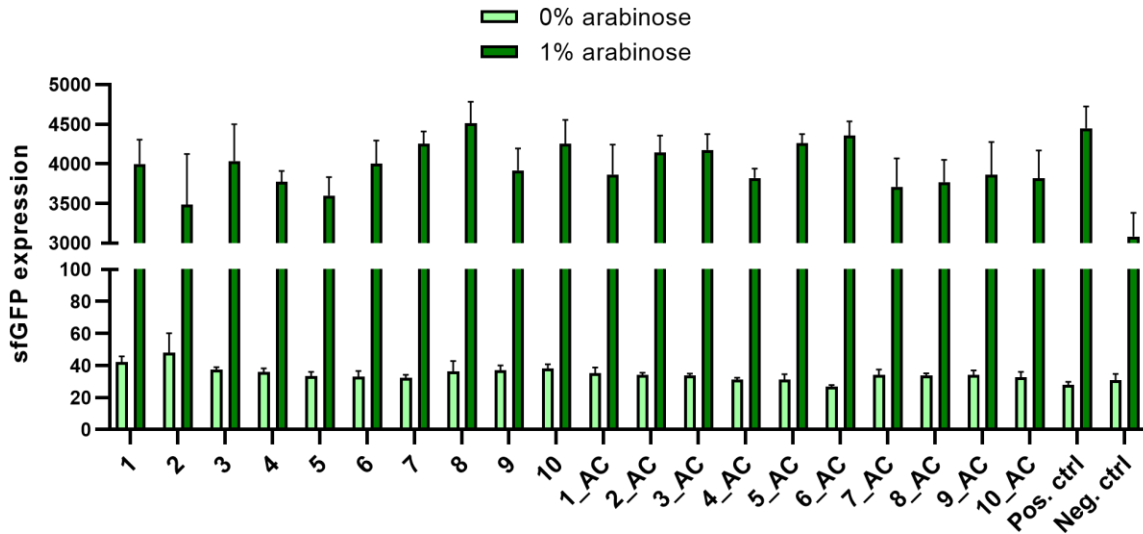


Figure 31: SfGFP expression of the hit sequences 1-10 and the autorepressor corrected samples 1_AC-10_AC upon arabinose induction at 0 or 1 % for the SRSF1 consensus sequence screening. Data are mean values (n=2, N=2).

As no hit sequences could be found amongst the most abundant hits neither in the autorepressor subtracted hit list, the SRSF1 consensus sequence screening is considered unsuccessful.

To exclude that screening failure is in conjunction with the sfGFP-SRSF1 fusion protein construct, another screening was performed with hnRNP A2B1. The gate set for sorting was

4. Results

chosen similarly to the gate for the SRSF1 sorting by excluding events that overlapped with the control population (sfGFP-hnRNP A2B1/S0 plasmid pair) and excluding events that were $<1 \times 10^5$ in the sfGFP expression level (shown in Figure 32) yielding in 2.41 % of the gated event count.

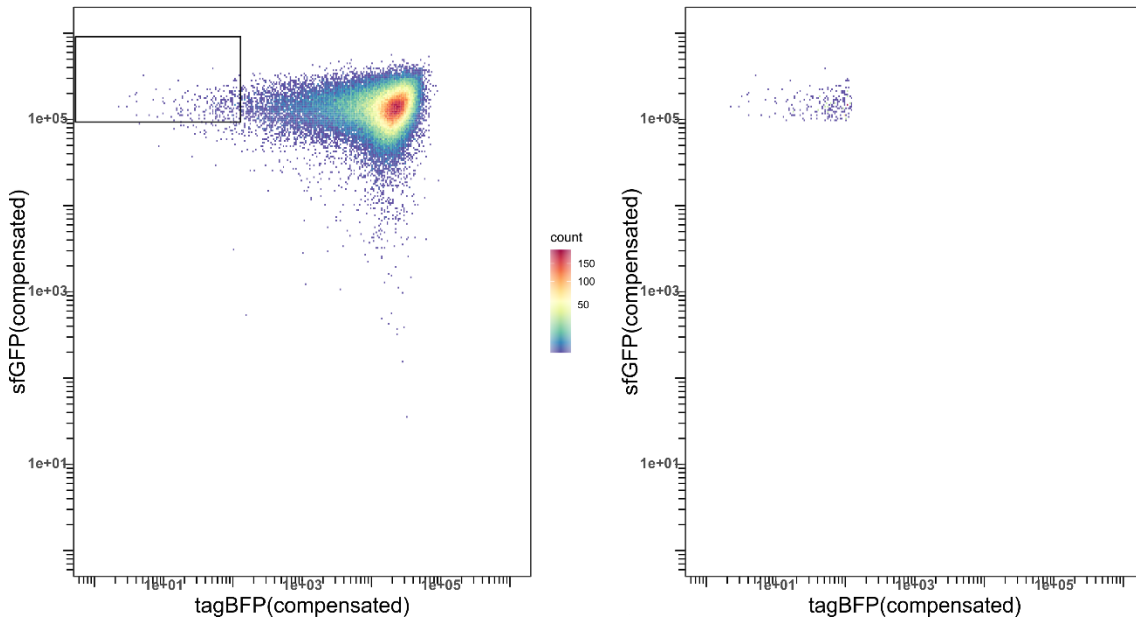


Figure 32: HnRNP A2B1 consensus sequence screening. Cells carrying the sfGFP-hnRNP A2B1 plasmid were transformed with the 10mer plasmid library and were treated with 1 mM IPTG and 0.125 % arabinose. Left: Whole population of the cells in the analysis. Cells that produced high sfGFP levels and low levels of tagBFP were gated for sorting (black box). Right: Sorted cells with gated fluorescence properties.

Similar to the SRSF1 consensus sequence screening, the density plot of hnRNP A2B1 shows a widespread population that is centred towards high tagBFP levels between 1×10^4 and 1×10^5 and high sfGFP levels between 1×10^5 and 1×10^6 . Again a small fraction of events is spread towards lower tagBFP levels ($<1 \times 10^3$), possibly indicating repressed reporter production. The sorting gate was chosen based on the gate of the SRSF1 consensus sequence screening gate (see Figure 27) but slightly shifted towards lower sfGFP levels to adjust to the whole population that was slightly reduced in the sfGFP level.

Experimental data presented in the following were produced by Luisa Kurzweg.

The first 10 most abundant hit sequences, as well as the first 10 autorepressor-corrected sequences, are listed in Table 8.

4. Results

Table 8: Hit sequence constructs of the hnRNP A2B1 consensus sequence screening used in TRAP and the resulting repression ratios. The negative control measurement was performed with the sfGFP-PTBP1 fusion protein.

Construct	RBP binding sequence	Linker	Repression ratio	Basal tagBFP production rate (1/min)	Sequence abundance (%)
1	CAUAAUGCAA	AUA	1.7 ± 0.2	175.4 ± 12.2	4.14
2	CGUCAUUCGU	AUA	0.7 ± 0.5	3.3 ± 2.2	2.62
3	UCAUUUAGUU	AUA	2.4 ± 0.4	189.8 ± 20.3	2.48
4	UGGCGUAUGU	AUA	0.0 ± 0.4	-0.4 ± 2.9	2.32
5	CUAAUACUCG	AUA	1.1 ± 0.2	184.5 ± 37.1	2.30
6	CGUUUGGUGU	AUA	1.1 ± 1.5	3.0 ± 3.2	2.29
7	ACGGGAUGCA	AUA	1.8 ± 0.1	47.1 ± 7.3	2.21
8	UGGCAACAAC	AUA	1.2 ± 0.2	206.6 ± 15.5	2.13
9	CAAACACACU	AUA	1.3 ± 0.2	269.7 ± 28.1	2.09
10	CAGACGACUC	AUA	0.5 ± 0.4	2.8 ± 1.1	2.00
1_AC	GUUCUCGCGA	AUA	0.9 ± 0.1	20.7 ± 5.4	0.57
2_AC	CUCGUUAUGG	AUA	0.6 ± 0.1	32.7 ± 4.6	0.50
3_AC	GACCUCGGCC	AUA	0.7 ± 0.0	43.3 ± 3.6	0.46
4_AC	AAAAUUAUGU	AUA	1.7 ± 0.3	38.7 ± 7.4	0.46
5_AC	AAGAAUCGCC	AUA	0.7 ± 0.0	28.4 ± 4.7	0.44
6_AC	GGCUCUUCCC	AUA	1.0 ± 0.1	76.7 ± 15.8	0.37
7_AC	GUGCCCUUCA	AUA	1.9 ± 0.3	20.2 ± 3.8	0.37
8_AC	CCAGUCAAAC	AUA	0.8 ± 0.1	22.8 ± 5.3	0.35
9_AC	UUUGGUCCUC	AUA	0.5 ± 0.0	40.4 ± 4.5	0.34
10_AC	AGCCAGGCGC	AUA	2.9 ± 0.8	52.3 ± 8.4	0.33
Positive control	AAGGACUAGCGGG AAGGACUAGC	AUA	9.8 ± 3.3	55.1 ± 9.9	-
Negative control	AAGGACUAGCGGG AAGGACUAGC	AUA	1.5 ± 0.2	55.9 ± 7.2	-

Again all listed sequences were cloned in the plasmid S0/H0 and evaluated using the TRAP assay. The sfGFP-hnRNP A2B1/H8 plasmid pair served as a positive control measurement, while the sfGFP-PTBP1/H8 plasmid pair was used as a negative control. Resulting repression ratios, tagBFP production rates and sfGFP expression can be found in Figure 33, Supplementary figure 31, Supplementary figure 32, and Table 8.

4. Results

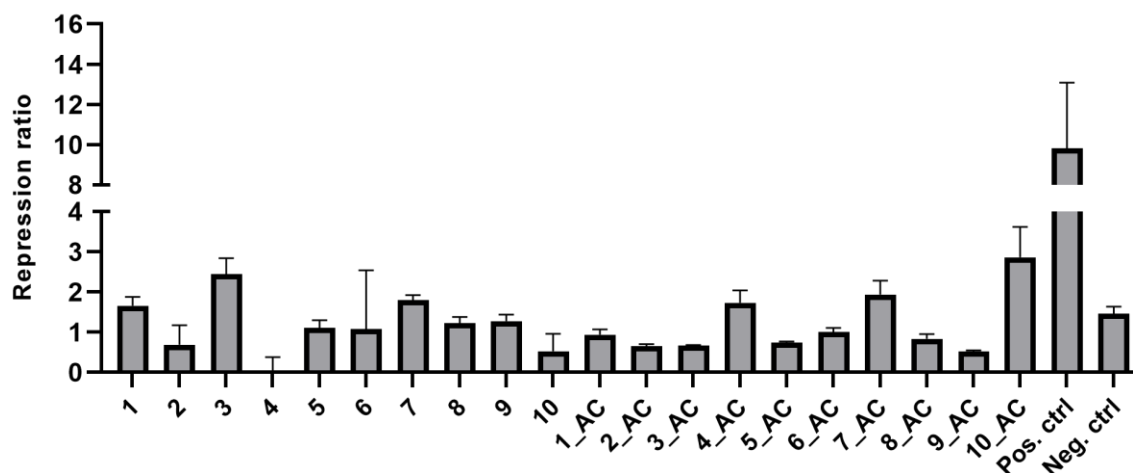


Figure 33: Repression ratios of the hit sequences 1-10 and the autorepressor corrected samples 1_AC-10_AC for the hnRNP A2B1 consensus sequence screening. Data are mean values (n=2, N=2).

The positive and negative control samples result in similar repression ratios, as described above. Most of the samples listed again do not respond to hnRNP A2B1 binding except samples 3 and 10_AC, which give repression ratios of 2.4 ± 0.4 and 2.9 ± 0.8 . SfGFP expression levels of all samples are higher than 3000, indicating sufficiently high levels of the sfGFP-RBP fusion protein (see Supplementary figure 32). Basal tagBFP production rates of samples 2, 4, 6, 7 and 10 are particularly low, probably indicating strong autorepression. Subtraction of the autorepressor sequences again removes all hit sequences 1-10. However, the overall basal tagBFP production rate of the hit sequences 1-10 is much higher than the one of the sequences 1_AC-10_AC. In addition, the autorepressor subtraction removes the hit sample 3, which responds to RBP binding, albeit with a low repression ratio, even though its basal tagBFP production rate is moderately high ($189.8 \pm 20.3 \times 1/\text{min}$). This could indicate that mathematical subtraction is possibly not a correct approach to remove autorepressor sequences. Sample 10_AC generates similar assay results as sample H5 (described in 4.1.2. Development of the TRAP assay for hnRNP A2B1) with regard to repression ratio and basal tagBFP production rate, which is why this sample should be designated as a true binding sequence of hnRNP A2B1 albeit not showing a high-affinity interaction.

Nevertheless, no high-affinity sequences could be found using the screening platform. One major obstacle may be the presence of autorepressor sequences that falsify the list of hit sequences. Removal of those via mathematical subtraction does not appear successful. Another possibility of removing autorepressor sequences could be done by presorting the screening population, while gating for a population that does not show a repressed phenotype

4. Results

using the same conditions as chosen for the autorepressor sequence selection. By that, autorepressor sequences are experimentally removed from the screening population, which then can be screened in a second approach using screening conditions (see Figure 34).

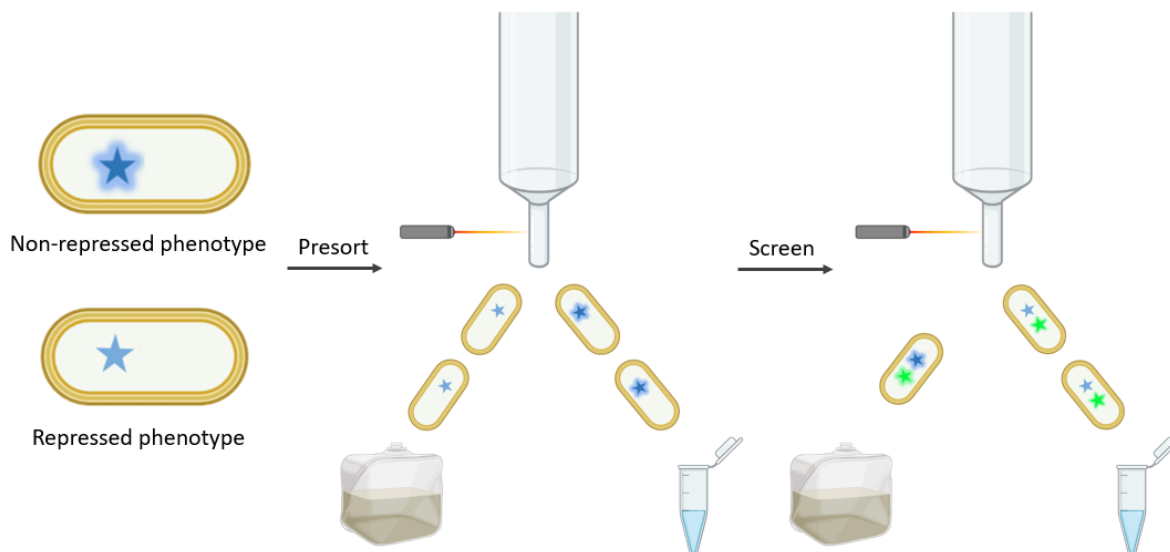


Figure 34: Concept of the autorepressor presorted RBP consensus sequence screening. *E. coli* cells carrying the sfGFP-RBP fusion plasmid are transformed with the 10mer library, induced with IPTG but without arabinose and presorted to remove autorepressor sequences. Cells that do not show a repressed phenotype are collected, while cells that show a repressed phenotype are discarded. In a second sorting round, cells are induced with IPTG and arabinose, and those with the repressed phenotype are sorted and sequenced. Illustration was created with Biorender.com.

A consensus sequence screening of such kind was performed for SRSF1 and will be described in the following section.

4.2.2. Autorepressor presorted RBP consensus sequence screening

Presorting was performed by inducing cells with 1 mM IPTG but no arabinose sorting for the cells that do not show a repressed phenotype. As no protein was expressed, translational repression should not happen and all cells that still showed repressed reporter production were removed. A gated event count of $\sim 11 \times 10^6$ was chosen for the presorting of the SRSF1 consensus sequence screening, as seen in Figure 35 (further sorting statistics can be seen in Supplementary table 7). For practical reasons, the sorting was performed in two steps resulting in two density plots with the same settings so one experiment will be shown representatively.

4. Results

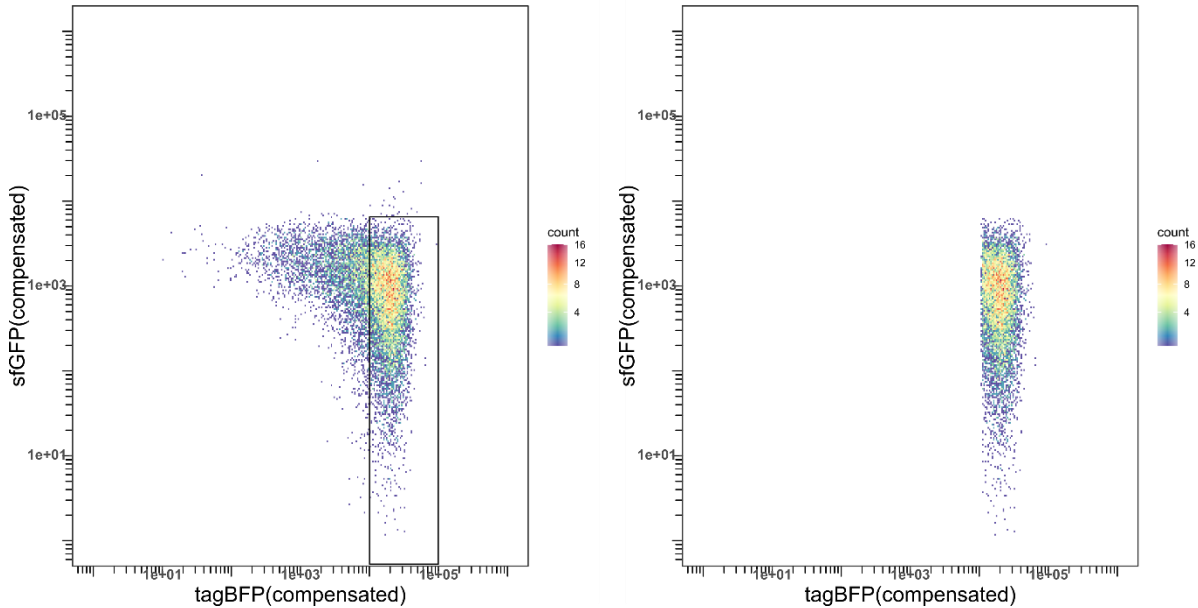


Figure 35: Presorting of the SRSF1 consensus sequencing screening. Cells carrying the sfGFP-SRSF1 plasmid were transformed with the 10mer plasmid library and were treated with 1 mM IPTG but no arabinose. Left: Whole population of the cells in the analysis. Cells that produced low sfGFP levels and high tagBFP levels were gated for sorting (black box). Right: Sorted cells with gated fluorescence properties.

The density plot shows the same population spread as for the autorepressor sequence selection (see Figure 28). This time the chosen gate was shifted towards high tagBFP levels (1×10^4 - 1×10^5) while keeping the same sfGFP levels as before to preselect cells that do not show an autorepressed response. By that, 83.94 % and 84.08 % of the gated event count were selected for sorting.

The presorted population was regrown, treated (1 mM IPTG, 0.125 % arabinose) and screened while a gated event count of $\sim 5 \times 10^6$ was chosen, as seen in Figure 36 (further sorting statistics can be seen in Supplementary table 7).

4. Results

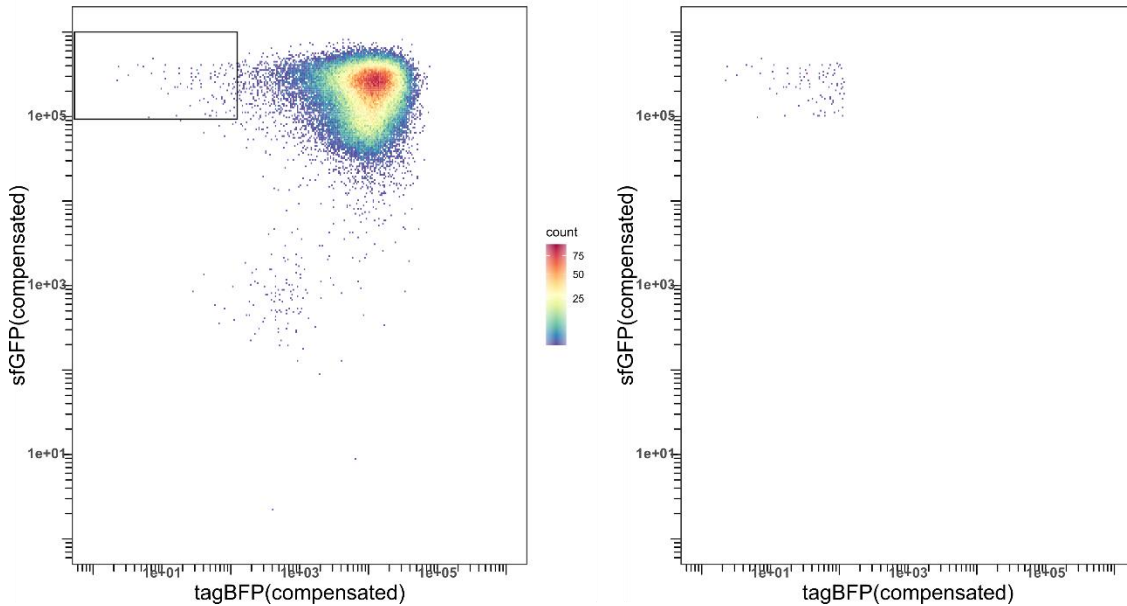
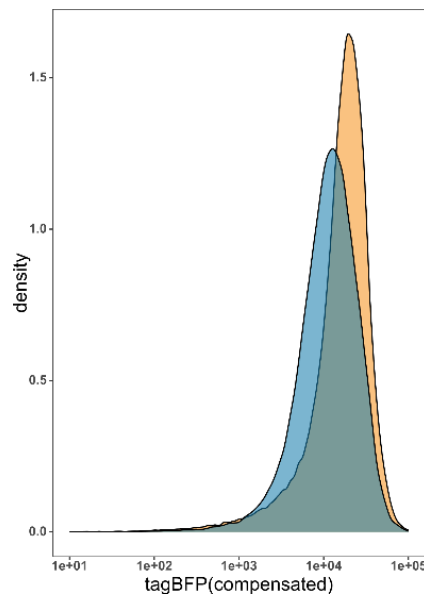


Figure 36: Presorted SRSF1 consensus sequencing screening. Presorted cells were treated with 1 mM IPTG and 0.125 % arabinose. Left: Whole population of the cells in the analysis. Cells that produced high sfGFP levels and low tagBFP levels were gated for sorting (black box). Right: Sorted cells with gated fluorescence properties.

The density plot shows a more evenly spread population as for the SRSF1 consensus sequencing screening without a presorting step (see Figure 27) and the spreading of events towards lower tagBFP levels is less prominent. The effect is more pronounced when comparing the histogram plots of both screenings, where the histogram of the presorted SRSF1 screening is more evenly distributed and shows less of a shoulder formation than the SRSF1 screening without presorting.



4. Results

Figure 37: Histograms for the presorted SRSF1 consensus sequence screening (blue) and the SRSF1 consensus sequence screening without presorting (orange).

The gate was picked as in the screening before and resulted in 1.22 % of the gated event count, while before, 2.66 % of events were gated. This indicates, that indeed a fraction of cells was eliminated performing the presorting step.

The hit list of the most abundant sequences in the sequencing data is presented in Table 9.

Table 9: Hit sequence constructs of the Autorepressor presorted SRSF1 consensus sequence screening used in TRAP and the resulting repression ratios. The negative control measurement was performed with the sfGFP-PTBP1 fusion protein.

Construct	RBP binding sequence	Linker	Repression ratio	Basal tagBFP production rate (1/min)	Sequence abundance (%)
1	AACGUAUCCG	AUA	1.7 ± 0.4	144.0 ± 30.6	4.54
2	UGGCGUAUGU	AUA	0.2 ± 0.4	0.8 ± 1.2	3.13
3	ACCGACGGAC	AUA	4.6 ± 0.5	123.1 ± 14.2	2.81
4	CAAUGCCACA	AUA	1.8 ± 0.3	74.3 ± 10.6	2.74
5	UAGAUAAAGUC	AUA	2.5 ± 0.2	186.7 ± 18.3	2.46
6	GUACAAUAUA	AUA	1.1 ± 0.1	119.1 ± 13.9	2.24
7	CCAGACGACA	AUA	0.8 ± 0.5	7.8 ± 1.9	2.17
8	CGAGCCCUAG	AUA	1.6 ± 0.2	72.4 ± 10.4	2.01
9	UUUCUCAACC	AUA	1.2 ± 0.2	202.3 ± 17.5	1.97
10	UCUUUCCGAC	AUA	1.4 ± 0.2	25.9 ± 4.1	1.97
11	CAACAGAACG	AUA	1.2 ± 0.4	93.9 ± 15.8	1.92
12	UCUUUCUCUA	AUA	1.5 ± 0.2	19.3 ± 3.3	1.92
13	ACGCCGUAAU	AUA	1.1 ± 0.6	118.0 ± 9.6	1.64
14	CCCUUUUGUC	AUA	1.4 ± 0.1	7.4 ± 0.5	1.57
15	CAACCAGUAU	AUA	1.1 ± 0.7	75.5 ± 46.0	1.51
Positive control	AGAAGAACAGAAG AACAGAAGAAC	AUA	13.2 ± 4.5	179.2 ± 7.5	-
Negative control	AGAAGAACAGAAG AACAGAAGAAC	AUA	1.5 ± 0.5	157.2 ± 32.7	-

As before, the hit sequences were cloned each into the S0 plasmid and cotransformed with the sfGFP-SRSF1 plasmid for TRAP assay analysis. Resulting repression ratios, tagBFP

4. Results

production rates and sfGFP expression levels can be seen in Figure 38, Supplementary figure 33, Supplementary figure 34 and Table 9.

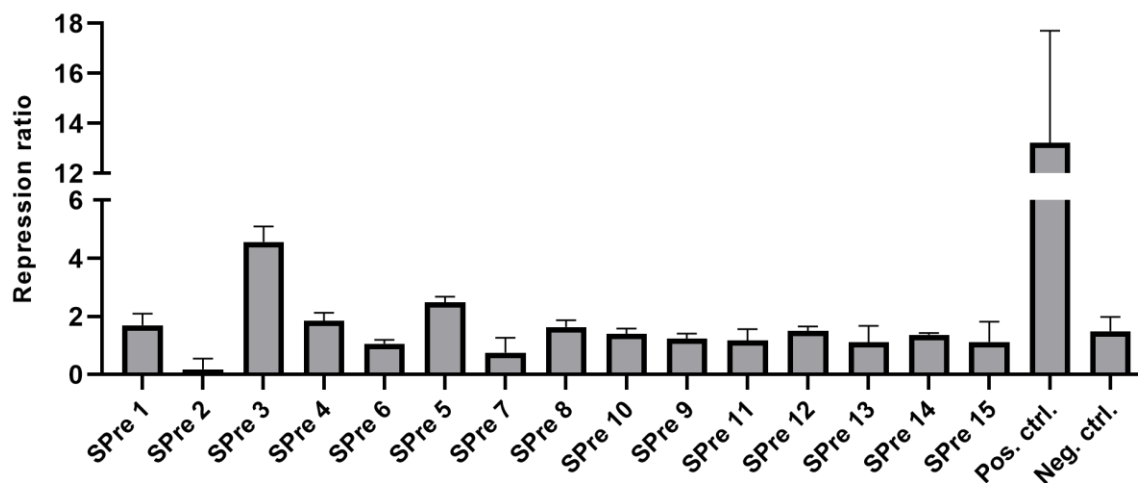


Figure 38: Repression ratios of the hit sequences 1-15 for the autorepressor presorted SRSF1 consensus sequence screening. Data are mean values ($n=2$, $N=2$).

Assay evaluation revealed two hit sequences, sample SPre 3 and 5, with moderate repression ratios of 4.6 ± 0.5 and 2.5 ± 0.2 , which are significantly higher than the repression ratio of the negative control, indicating moderate binding affinity of the RNA sequence to the protein. All other samples do not produce significantly higher repression ratios than the negative control suggesting no binding events between the sequences and the protein, as previous negative control samples also produced such repression ratios (see Supplementary table 3). The repression ratio of the positive control sample is lower than before, whereas the error is larger. TagBFP production rates, on average, are neither significantly lower nor higher than in the SRSF1 screenings before, while samples SPre 2, 7, 10, 12 and 14 are strongly autorepressed, demonstrating that the autorepressor presorting was not fully successful (shown in Supplementary figure 33). SfGFP expression levels are again ~ 3000 and higher, which indicates sufficiently high levels of the sfGFP-RBP fusion protein are present (see Supplementary figure 34).

In comparison to the SRSF1 consensus sequence screening without the presorting, where no binding sequences were identified, the improved screening design of the autorepressor presorted screening identified two sequences as SRSF1 binders and can therefore be considered a success, albeit a minor one. The repression ratio of sample SPre3 is similar to that of construct S4 and higher than construct S1 (4.6 ± 0.5 vs. 4.8 ± 1.0 vs. 3.3 ± 0.3) where

4. Results

the latter generates a K_D of 92.3 ± 44.1 nM. As binding affinities and repression ratios correlate well (see Figure 24), it can be assumed that the screening identified a sequence with an affinity in a low nanomolar range. Nevertheless, it is questionable why the screening still identified many autorepressed sequences and why the overall success is only minor despite an improved assay design. The performed screenings did not identify better binding sequences as those constructed by rational design, therefore, protein/RNA constructs with good repression ratios will be combined with the SICLOPPS method as a screening platform.

4.3. Identification of hexameric peptide inhibitors for the splicing factors SRSF1 and hnRNP A2B1 using SICLOPPS

Inhibiting molecules and chemical probes for RBPs were identified previously using conventional biophysical methods such as Förster Resonance Energy Transfer (FRET), FP or AlphaScreen® assays while screening small molecule libraries in a high throughput manner. While they have been fruitful, these methods are limited as they are incompatible with cell-based screenings. The development of luciferase assays has facilitated HTS of compound libraries in a cellular setting and has successfully identified AS modulators Risdiplam and Branaplam.¹⁰⁶ HTS against compound libraries is, however, limited by the number of molecules that can be screened in a single campaign, which does not exceed a million compounds and requires specialized automation and is therefore costly.¹⁰⁷ The use of SICLOPPS libraries, mostly used in combination with bacterial reverse two-hybrid systems, generate screening platforms that allow screening of hundreds of millions of molecules in cellular context, which is limited only by the transformation efficiency of the host (10^9 plasmids for *E. coli*).¹⁰⁸

The TRAP assay is another *in vivo* screening method that, when combined with SICLOPPS, potentially offers an approach to screen for cyclic peptides inhibitors for PRIs in a high-throughput format, therefore expanding the scope of SICLOPPS from PPIs to PRIs. The design of SICLOPPS plasmids orthogonally to the TRAP assay plasmids required introduction of a third antibiotic resistance gene as well as an orthogonal promoter and origin of replication (ORI), allowing for straightforward transformation of all three plasmids into *E. coli*. Cells that carry a protein/RNA plasmid pair that results in repression of the tagBFP production rate can be used as a basis for a screening platform. Inhibition of the RBP/RNA interaction by a cyclic peptide prevents binding of the RBP to its RNA target that logically causes a restoration of reporter production, as demonstrated in Figure 39.

4. Results

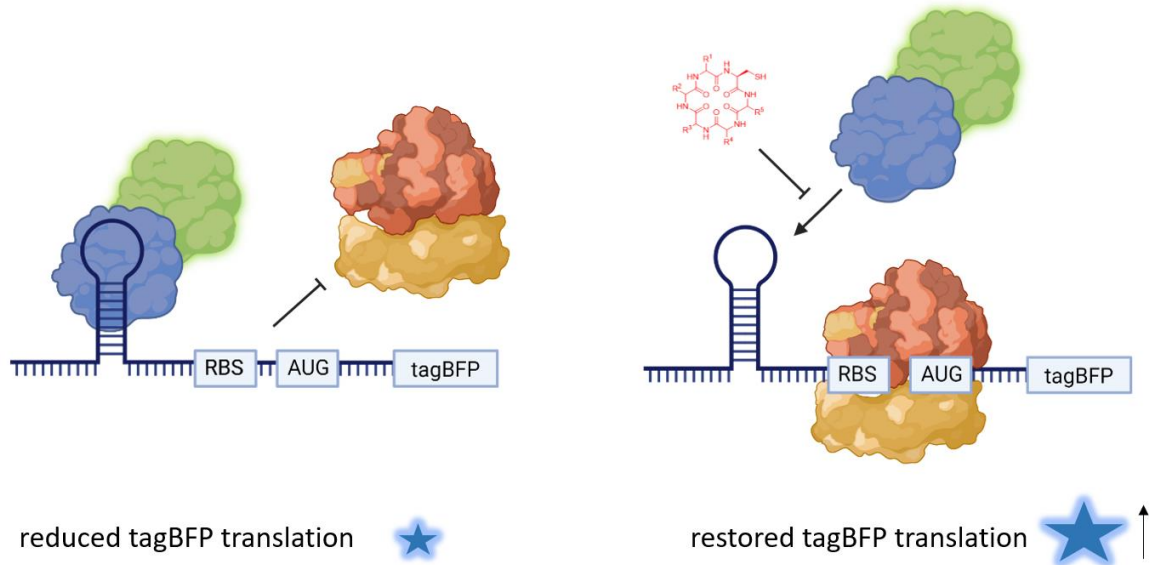


Figure 39: The TRAP assay as a screening platform for SICLOPPS. Left: Reporter RNA/RBP interaction results in translational repression and low reporter production. Right: Inhibition of the RNA/RBP interaction with a peptide inhibitor, highlighted in red, allows ribosome binding and restores reporter translation, leading to high tagBFP production rates.

As shown in section 4.1., the protein/RNA interaction of SRSF1 with S6 and hnRNP A2B1 with H8 are high-affinity interactions that also generate high repression ratios, suggesting a good dynamic range for signal restoration. These two interactions were targeted with cyclic peptides generated by SICLOPPS plasmids (see Supplementary figure 35) while using the TRAP assay as a screening platform.

4.3.1. SICLOPPS screening analysed by Sanger sequencing

*Cloning of the SICLOPPS library used in the following chapters was performed by Stefan Schmeing, according to the report by Tavassoli and Benkovic.*⁵⁹

To identify possible peptide inhibitors of SRSF1, a SICLOPPS screening using the TRAP assay as a screening platform was performed. Here, the cells were transformed with the described assay plasmids and the SICLOPPS plasmid library. The screening was performed in accordance with RBP consensus sequence screening, where proteins were induced with 0.125 % arabinose while reporter production was induced with 1 mM IPTG. To ensure sufficient cyclic peptide production, SICLOPPS peptide formation was induced beforehand with 100 ng/mL anhydrotetracycline and incubated overnight at 30 °C. The SICLOPPS library was designed to produce hexameric cyclic peptides using a CXXXXX insert where X defines an NNS codon (N=A, T, G or C; S=C or G) and has a theoretical diversity of 3.2×10^6 peptides.

4. Results

A 10- to 25-fold count of the library population during FACS selection is considered to fully cover the library (>99.99-100 %).¹⁰⁹ However, a 10-fold library coverage would have increased sorting times remarkably, and time capacities at the device were limited, so a lower fold coverage of total events was chosen. In addition, a 3-fold factor is sufficient to reach 95 % of library completeness.^{109,110} First, the SRSF1 SICLOPPS screening was performed and $\sim 10 \times 10^6$ total events were screened. Further sorting statistics can be seen in Supplementary table 7.

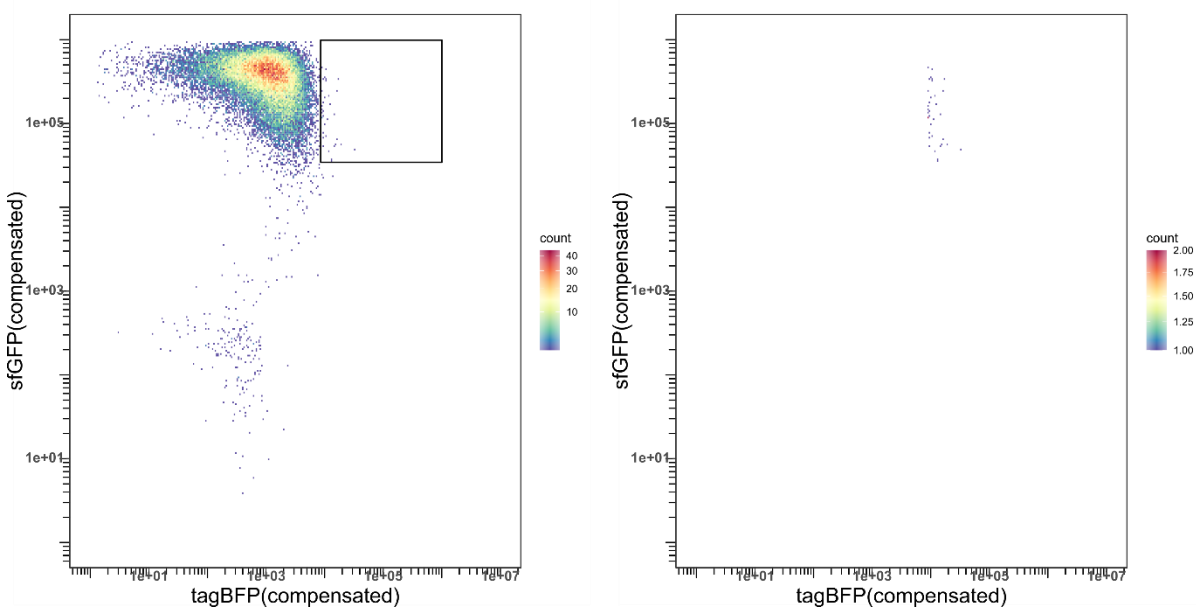


Figure 40: SRSF1 SICLOPPS screening. Cells were treated with 100 ng/mL anhydrotetracycline prior to treatment with 1 mM IPTG and 0.125 % arabinose. Left: Whole population of the cells in the analysis. Cells that produced high sfGFP levels and high tagBFP levels were gated for sorting (black box). Right: Sorted cells with gated fluorescence properties.

The density plot shows a widespread population that is centred towards lower tagBFP levels between 1×10^3 and 1×10^4 and high sfGFP levels between 1×10^5 and 1×10^6 , indicating that the majority of cells show a repressed response. Single events are spread towards higher tagBFP levels ($>1 \times 10^4$), possibly demonstrating restored reporter production. The gate for sorting was set highly focused, comprising 0.04 % of total events yielding 3389 sorted cells to allow better handling during colony counting and picking. After plating and incubation, a total of 90 colonies was counted, indicating low survival of cells (2.66 %), assuming that each sorted cell is represented by a colony-forming unit. 75 colonies were analysed by Sanger sequencing, cultivated and frozen for further analysis. After elimination of stop codon-containing inserts (five inserts, 6.58 %), sequencing revealed 39 unique peptide sequences

4. Results

since a number of peptides appeared more often (up to 5-fold). 16 plasmids showed a deletion in the backbone and were therefore excluded from the analysis. Interestingly, peptide 61 contains seven amino acids instead of six and may be annotated to errors during oligonucleotide synthesis at the commercial service. The full list of peptides that did not contain a stop codon and their occurrence can be seen in Table 10.

Table 10: Hit-peptides, their sequences and occurrence in the screen.

Peptide	Sequence	Occurrence in screen	Peptide	Sequence	Occurrence in screen
5	C G D F L Y	Two	39	C V Y P R C	One
6	C Q I H G W	Five	41	C M S L T V	One
8	C L I T V H	Two	44	C Y Q S Y M	One
9	C R D F Y W	One	45	C D W S G K	One
11	C T V S L G	One	46	C S Q W R R	One
12	C S W F C S	One	50	C E S L C H	One
14	C C R S L R	One	51	C R T L C H	One
16	C R W G F H	Three	53	C V R W M W	One
18	C Q L S N R	One	57	C T V V P Q	One
20	C L T V R Q	One	58	C A S M G R	One
21	C R S N F H	Two	59	C W V L E W	One
22	C G K L V M	One	60	C N A R L T	One
24	C G S L P T	One	61	C L L G A A D	One
26	C L L A Y L	One	63	C S R R T D	One
28	C L I P I P	Three	65	C R Q V E V	One
30	C R F T C C	One	66	C G L G G P	One
31	C S G S W R	Two	67	C I S G Y F	One
35	C R I E R G	One	71	C P Y M F Q	One
36	C G L C R A	One	72	C Q R W T T	One
37	C H W D H R	Two			

To evaluate the performance of SICLOPPS peptides inside the cell, the TRAP assay was used as a modified version (see 3.2.4.5) that allows ranking of peptides according to their capability to restore the tagBFP production rate. For that, cells were induced as performed in the screening protocol, while the reporter production and sfGFP-RBP expression were also

4. Results

observed at an arabinose non-induced state. As positive controls, the sfGFP-SRSF1_{mut}/S6 plasmid pair (PC) and the sfGFP-SRSF1_{mut}/S6 plasmid pair including a SICLOPPS plasmid that codes for an alanine-rich cyclic peptide, cyclo-CAAAAA, (PC+B) were employed. The CAAAAA peptide was considered to represent a “blank” peptide as its -CH₃ side chains do not have the possibility for extensive interactions with the target protein.¹¹¹ Substitution of amino acids with alanine is a concept used in Alanine Scans that allows dissection of functional side chains in peptide compounds or protein target sites and is a widely used method.¹¹² The cysteine side chain was kept as it was required for cyclisation and allowed better comparison to hit peptides. As negative controls, the sfGFP-SRSF1/S6 plasmid pair (NC) and the sfGFP-SRSF1/S6 plasmid pair, including cyclo-CAAAAA (NC+B), were used.

All 39 unique peptides were analysed in this manner. The first 12 peptides that resulted in the highest tagBFP production rate restoration are presented in the following figures. To mimic screening conditions, cells were induced with 1 mM IPTG and 0.125 % arabinose and the ability of the hit peptides to restore tagBFP production rates was observed, as can be seen Figure 41.

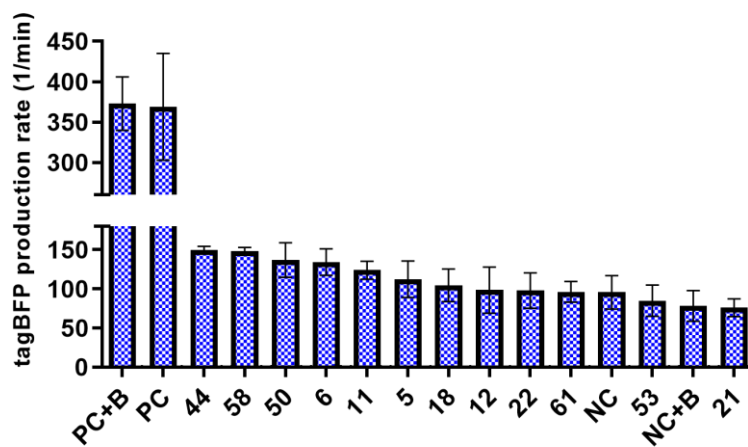


Figure 41: TagBFP production rate of cells carrying the sfGFP-SRSF1/S6 plasmid pair and a SICLOPPS plasmid (indicated by number). Cells were treated with 100 ng/mL anhydrotetracycline prior to treatment with 1 mM IPTG and 0.125 % arabinose. Data are mean values (n=2, N=2).

The negative control samples NC+B and NC led to repression of tagBFP production rates, as described above. Expression of the blank peptide leads to slight reductions in tagBFP production rate, suggesting that peptide expression may influence tagBFP production rates, albeit only to a minor extent. Formation of the SICLOPPS peptides 44, 58, 50, 6, 11, 5, 18, 12, 22, 61 and 53 led to tagBFP production rate restoration higher than the negative control

4. Results

NC+B. Peptides 11, 6, 50, 58 and 44 led to a >50 % increase in tagBFP production rate compared to NC+B control and are therefore considered especially effective.

To control whether the hit peptides lead to changes in tagBFP production rates irrespective of RBP expression levels, cells were induced with IPTG but not with arabinose. Most samples led to a tagBFP production rate of around 250 - 350x1/min, while samples 61 and 21 were significantly lower (see Figure 42).

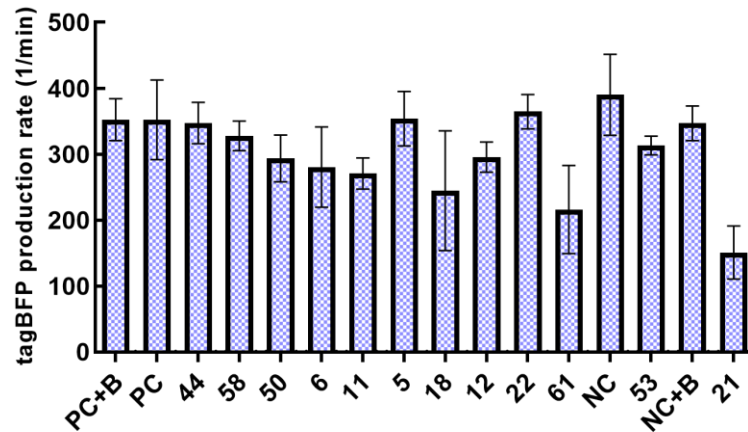


Figure 42: TagBFP production rate of cells carrying the sfGFP-SRSF1/S6 plasmid pair and a SICLOPPS plasmid (indicated by number). Cells were treated with 100 ng/mL anhydrotetracycline prior to treatment with 1 mM IPTG but no arabinose. Data are mean values (n=2, N=2).

Under consideration of the error, sample 21 has a highly reduced tagBFP production that cannot be attributed to RBP binding and can only be hypothesized to result from peptide expression. Overall, although errors are high for some samples, tagBFP production rates are relatively homogenous, indicating functional reporter expression.

SfGFP-RBP fusion protein expression of the samples at 0 % arabinose results in expression levels <100, which is in line with the findings described above (compare Figure 43 and Figure 13).

4. Results

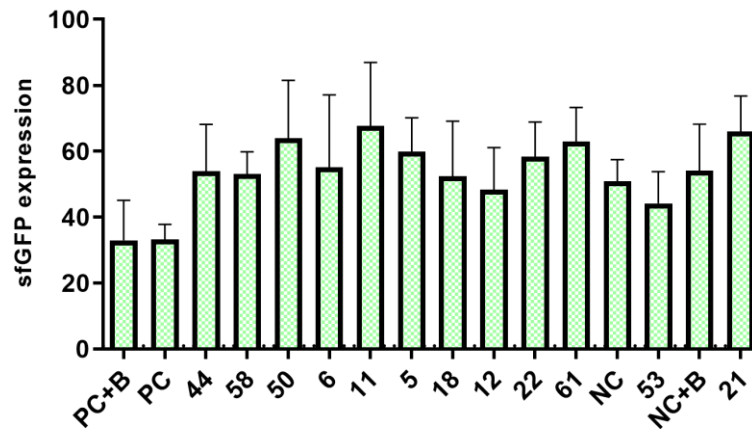


Figure 43: SfGFP expression of cells carrying the sfGFP-SRSF1/S6 plasmid pair and a SICLOPPS plasmid (indicated by number). Cells were treated with 100 ng/mL anhydrotetracycline prior to treatment with 1 mM IPTG but no arabinose. Data are mean values (n=2, N=2).

Induction of the sfGFP-RBP fusion protein at 0.125 % arabinose, demonstrated in Figure 44, led to rather an inhomogeneous expression over all samples ranging from around 500 - 1300 units, while the negative control samples, NC and NC+B, show the highest protein expression.

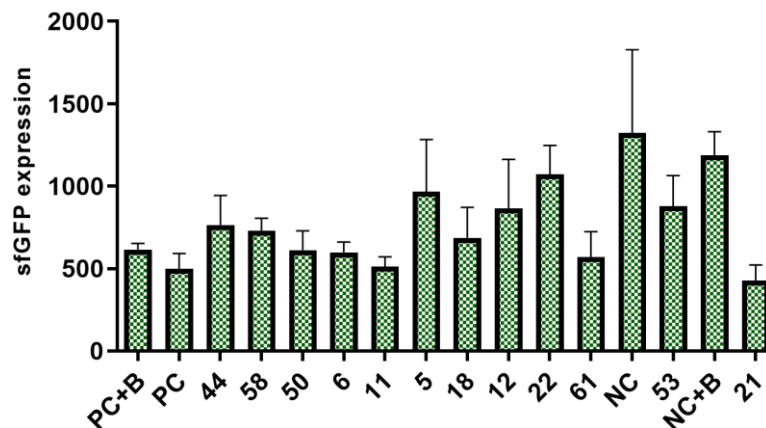


Figure 44: SfGFP expression of cells carrying the sfGFP-SRSF1/S6 plasmid pair and a SICLOPPS plasmid (indicated by number). Cells were treated with 100 ng/mL anhydrotetracycline prior to treatment with 1 mM IPTG and 0.125 % arabinose. Data are mean values (n=2, N=2).

The SRSF1_{mut} construct used as positive controls, PC and PC+B, was known to express less efficiently than SRSF1, which explains reduced sfGFP expression levels (Supplementary figure 4). It is evident that samples expressing SICLOPPS peptides other than the blank peptide, *cyclo*-CAAAA, express the sfGFP-RBP less efficiently. Sample 21 expresses the

4. Results

sfGFP-RBP as the lowest, indicating that this sample may be impaired to some extent. Whether the reason lies in the peptide expression remains unclear but cannot be excluded.

The TRAP evaluation allowed ranking of all SICLOPPS samples according to their tagBFP production rates upon induction. Five of 12 peptides led to a >50 % increase in tagBFP production rate in comparison to the negative control NC+B. Table 11 shows the ranking of all 12 peptides, their amino acid sequences, occurrence in the primary screen and tagBFP production rate upon induction with arabinose at 0.125 %.

Table 11: Peptide sequences, occurrence in sequencing and tagBFP production rate upon induction. The positive control measurement was performed using the sfGFP-SRSF1_{mut} fusion protein construct.

Peptide	Sequence	Occurrence in screen	TagBFP production rate (1/min)
44	C Y Q S Y M	One	149.8 ± 4.6
58	C A S M G R	One	148.5 ± 4.6
50	C E S L C H	One	137.0 ± 22.1
6	C Q I H G W	Five	134.3 ± 17.1
11	C T V S L G	One	124.0 ± 11.5
5	C G D F L Y	Two	112.5 ± 23.4
18	C Q L S N R	One	104.6 ± 20.8
12	C S W F C S	One	98.5 ± 29.5
22	C G K L V M	One	97.9 ± 22.7
61	C L L G A A D	One	96.3 ± 13.2
53	C V R W M W	One	85.0 ± 19.9
21	C R S N F H	Two	76.2 ± 11.2
PC+B	C A A A A A	-	372.9 ± 33.1
PC	-	-	369.0 ± 66.2
NC+B	C A A A A A	-	78.2 ± 19.6
NC	-	-	95.7 ± 21.3

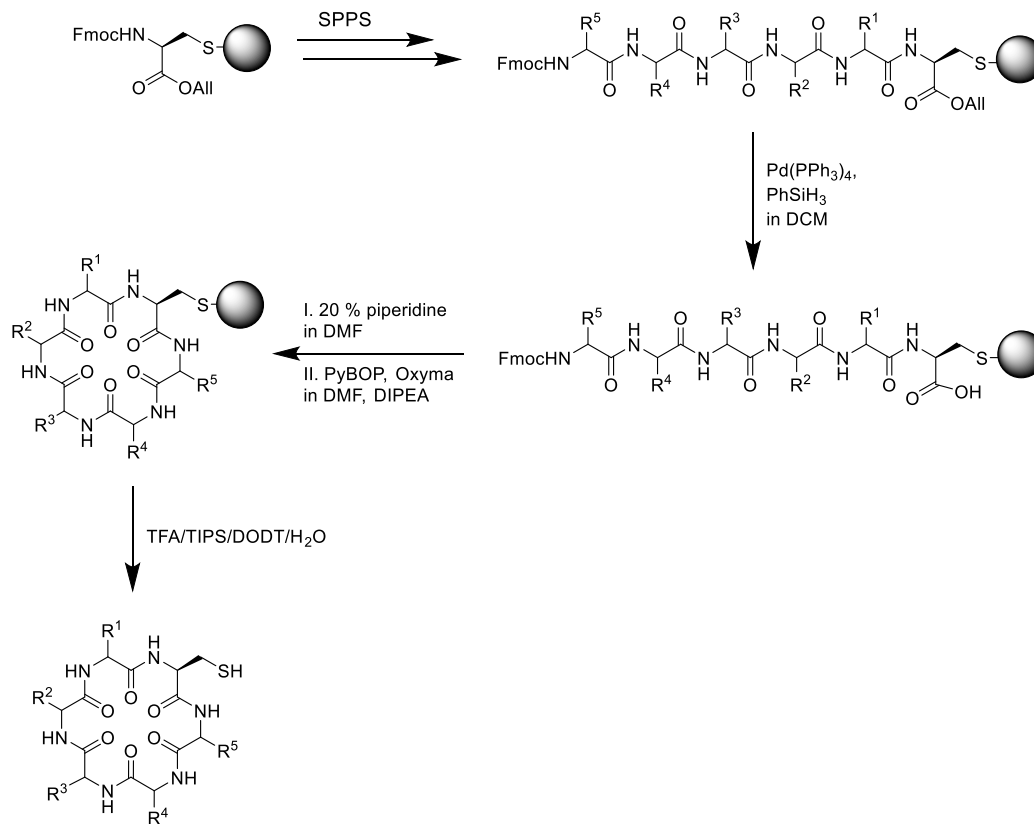
Since peptides 44, 58, 50, 6 and 11 were the most effective in terms of tagBFP modulation, they were chemically synthesized and biophysically tested.

4. Results

4.3.2. Synthesis of hexameric, cyclic peptides and evaluation of peptides by FP

Synthesis protocols used in this chapter are derived from Dr. Peter 't Hart. Synthesis of cyclic peptides was performed in collaboration with Joseph Openy, who synthesised peptide 44.

The synthesis of peptides historically was performed in solution and was advanced to the synthesis on solid phase (solid-phase peptide synthesis, SPPS) introduced by Merrifield in the 1960s.^{113,114} An insoluble polymeric support (resin) is C-terminally covalently linked to the nascent peptide chain, which can be extended stepwise at the N-terminus by additional synthesis cycles with the use of orthogonal protecting groups. After maturation of the peptide chain, deprotection and cleavage from the resin can be performed simultaneously, thus releasing a peptide ready for purification. Solid-phase approaches are also compatible with cyclization techniques, e.g. using the three-dimensional protection scheme of Fmoc/tBu/Allyl protection groups for head-to-tail cyclization.¹¹⁴ In order to perform further characterisation of the hits identified in the SICLOPSS screening, SPPS following an on-resin cyclization was used for the synthesis of the top 5 hits using a 2-chlorotrityl chloride resin that allows connection of the cysteine thiol to the resin. The following illustration shows a general synthesis route for cyclic peptides.



4. Results

Figure 45: Synthesis route of SICLOPPS peptides on solid phase (2CTC resin). The first amino acid (cysteine) is loaded to the resin via its side chain thiol group. Once completion of the sequence is achieved, the peptide is cyclised head-to-tail and then cleaved from the resin.

Out of the 5 hits, 3 peptides (6, 44 and 50) were successfully synthesized. Cyclization of peptides 58 and 11 were unsuccessful with the proposed synthesis route and therefore paused.

To validate these hits, their ability to prevent SRSF1 binding to RNA *in vitro* was investigated by FP measurements (Figure 46). FP of FAM-labelled S3 was measured when combined with SRSF1 in the presence of the peptides or with DMSO as a control. Unlabelled S3, which should have an equal affinity to SRSF1, was included to confirm that a decrease of binding to FAM-S3 could be detected in this assay.

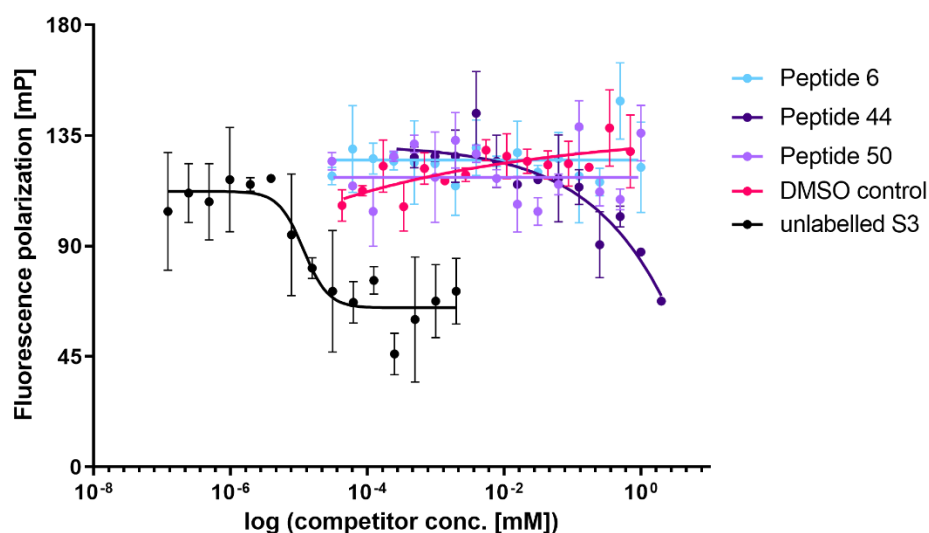


Figure 46: Fluorescence polarisation competition curves of peptides 6, 44, 50, unlabelled S3 RNA and a DMSO control (5 %) with 80 nM SRSF1 and 1 nM FAM-S3. Unlabelled S3 RNA and DMSO were used as controls. Data was measured after 20 minutes incubation and is presented as mean values (N=2).

Fluorescence polarisation measurements showed competition between the unlabelled S3 RNA and FAM-labelled tracer. As the labelled and unlabelled S3 RNAs possess the same nucleotide sequence and only differ in the FAM label, high affinity of the unlabelled S3 RNA to the protein is expected similarly as for FAM-S3. The concentration of the unlabelled RNA at which 50 % of the tracer is competed off (here: IC_{50}) is determined to be 11.11 ± 5.54 nM. Technically, this number is slightly below the detection limit as the IC_{50} value cannot be lower than the used protein concentration, indicating that there might be inaccuracy in detection which could have resulted from inaccurate protein concentration determination. Nevertheless,

4. Results

competition can be measured, and the assay was further used to evaluate the synthesised peptide hits. Peptides 6 and 50, as well as the DMSO control, do not compete with the tracer RNA and therefore do not show binding. Peptide 44 does compete with the tracer at concentrations above 100 nM, however, an IC_{50} cannot be determined as the curve does not reach bottom-value saturation indicating that the starting concentration chosen for testing was not sufficient. As the peptide amounts were limited, repetition of the experiment with an increased concentration could not be performed.

Interestingly, peptide 44 led to the highest restoration in tagBFP production rate and, in contrast to peptides 6 and 50, led to competition with the tracer RNA. To confirm that the signal change indeed results from a true competition event, another FP experiment was performed, including a control measurement with peptide 44 and buffer only (shown in Figure 47). The incubation time was increased to 45 minutes.

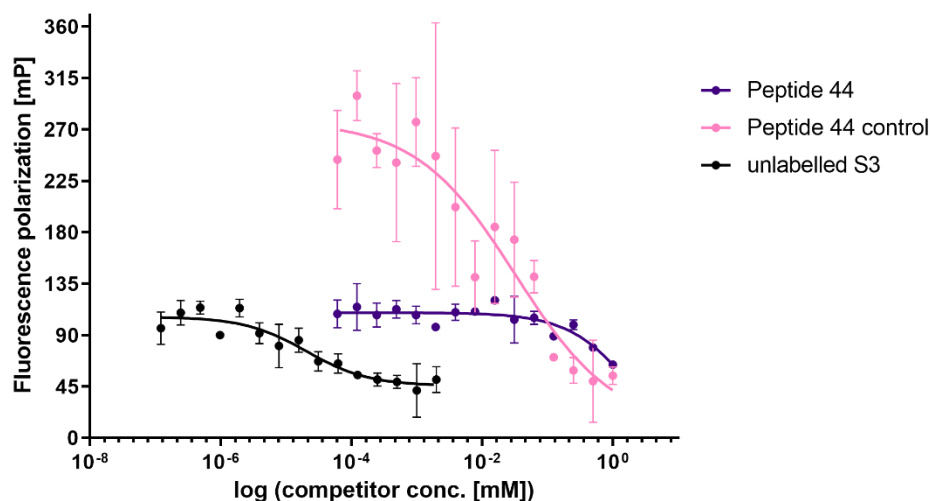


Figure 47: Fluorescence polarisation competition curves of peptide 4 and unlabelled S3 RNA with 80 nM SRSF1 and 1 nM FAM-S3 and Peptide 44 with buffer only. Unlabelled S3 RNS was used as control. Data was measured after 45 minutes incubation and is presented as mean values (N=2).

Surprisingly, the FP measurements of the peptide with buffer alone show a competition-like curve which is not expected as neither protein nor tracer was present in the experiment. The signal change seems to be an effect of the peptide titration, while the peptide influences polarised light at high concentrations. The competitive behaviour of peptide 44 seems not due to the inhibitory properties of the peptide but instead result from false positive signal development. To verify that the signal change is truly originating from the peptide Cy5-labelled S3 RNA as tracer was used in a competition experiment, as can be seen in Figure 48. Using

4. Results

this label allowed the measurement to be done at higher wavelengths where the intrinsic effect of peptide 44 was less likely. The incubation time was increased to 60 minutes.

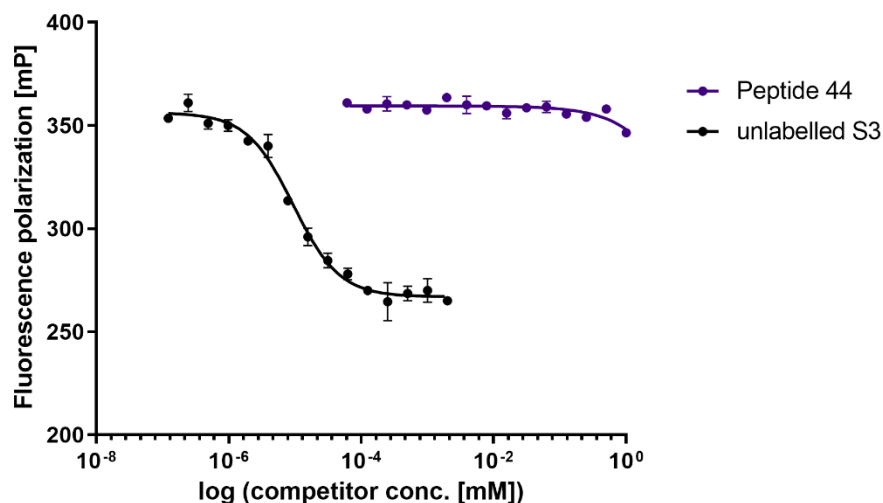


Figure 48: Fluorescence polarisation competition curves of peptide 4 and unlabelled S3 RNA with 80 nM SRSF1 and 1 nM Cy5-S3. Unlabelled S3 RNA was used as a control. Data was measured after 60 minutes incubation and is presented as mean values (N=2).

The experiment shows competition between the Cy5-S3 and the unlabelled S3 RNA with an IC_{50} of 9.36 ± 0.85 nM, which is in similar range as detected before (compare to Figure 46) but no competition between peptide 44 and the tracer. Therefore, the competition response of peptide 44 in the initial FP experiment must be considered as a false positive, at the same time excluding peptide 44 as a binder for SRSF1.

None of the tested hit peptides interfered with SRSF1 binding to RNA, albeit their ability to modulate tagBFP production rate restoration. Peptide 6 appeared 5 times in the screening, indicating enrichment of hit peptide in the screen, however, a binding event could not be measured. Not even the most promising candidate, peptide 44, showed binding properties in FP, which indicated that the current screening approach is not effective and another approach should be chosen.

Considering only 90 surviving colonies of over 3000 sorted cells, a representation of only 2.66 % of all hit peptides is achieved, suggesting that this aspect of the assay setup should be optimized. This could be achieved by reducing the antibiotics used during cell plating after sorting on agar plates from all three selective antibiotics to the selection marker of the SICLOPPS plasmid (AmpR). Furthermore, manual colony picking after sorting is strongly

4. Results

limited practicality and could be replaced using NGS techniques. A screening setup including NGS analysis allows an overview of the entirety of all surviving cells, is independent of the colony number, and also allows a ranking by sequence abundance. In this manner, an optimised SICLOPPS screening was performed and is presented in the following sections.

4.3.3. SICLOPPS screening analysed by Illumina sequencing

The SICLOPPS screening protocol was optimised by changes of a few steps to increase peptide sequence diversity and cell survival after sorting. Cell analysis and sorting were performed as before and the result can be seen in Figure 49.

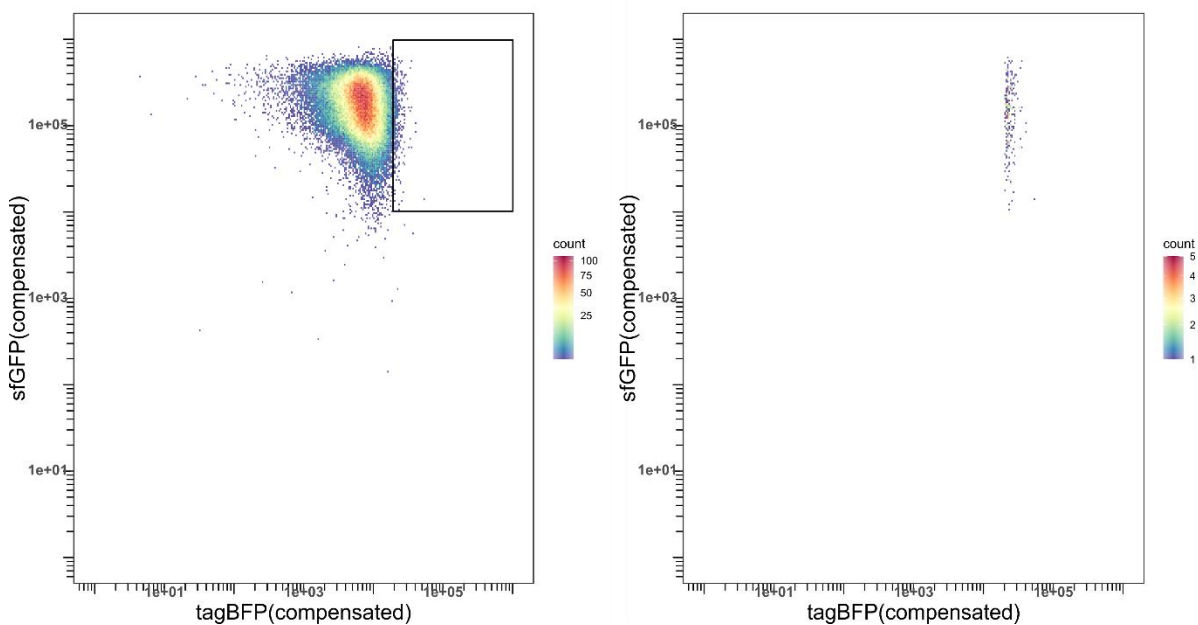


Figure 49: SRSF1 SICLPPS screening with optimized conditions. Cells were treated with 100 ng/mL anhydrotetracycline prior to treatment with 1 mM IPTG and 0.125 % arabinose. Left: Whole population of the cells in the analysis. Cells that produced high sfGFP levels and high tagBFP levels were gated for sorting (black box). Right: Sorted cells with gated fluorescence properties.

The cells were plated onto agar plates with ampicillin only instead of using all three antibiotics to only select for the SICLOPPS plasmid. It is evident that the whole cell population is shifted more towards high tagBFP levels, which indicates a higher count of cells that show disruption of the repressed phenotype in comparison to the screening performed before (compare Figure 40). Of 12×10^6 total events, the gate for sorting was chosen to comprise 0.31 % of total event count (23,732 events). For hit peptide sequences analysis, paired-end Illumina sequencing of the variable peptide insert was performed. 16,424 total unique sequences and 12,823 unique sequences that did not contain a stop codon were identified. The unique sequence count

4. Results

without stop codons is 328-fold higher than in the screening performed before and therefore allows a better overall consideration of the screening. The top 15 hit sequences ranked by their overall abundance are shown in Table 12.

Table 12: SRSF1 SICLOPPS screening hit sequences and sequence abundance. Entries with stop codon-containing sequences (indicated by X) are highlighted in red.

Number	Peptide sequence	Sequence abundance (%)
1	C P S E W S	3.03
2	C L L T E L	2.29
3	C L F D S V	1.97
4	C Y S X S L	1.63
5	C Y F D V L	1.51
6	C S E L Q N	1.38
7	C T F S P M	1.34
8	C V E A V X	1.30
9	C S C V P A	1.29
10	C K K L W L	1.28
11	C G L T K I	1.14
12	C H R X V C	1.10
13	C Y V A T A	1.01
14	C E I Y X X	0.88
15	C S S L C T	0.85

Of 15 hit sequences, stop codon-containing sequences appear 4 times (26.67 %). As the codons for the variable peptide sequence were cloned as NNS constructs, there is a 3.13 % probability for stop codon expression in every variable position of the peptide. The overall probability for a peptide to contain a stop codon at any position is 15.63 %. The appearance of stop codons in the hit list is therefore slightly increased as a result of enrichment and not random. The sequence with the highest abundance (CPSEWS, 3.03 %) is 0.74 % higher than the next entry and possesses the highest difference in % points. This indicates that the sequence is more enriched in comparison to the other entries. To evaluate the hit peptide biophysically, the top 6 hits that do not contain stop codons were chemically synthesized.

4. Results

To explore the effect of sequence enrichment on hit peptide abundance, a sorting experiment over two rounds of sorting was performed. For that, cells were sorted, recultivated, and sorted again. The sorting results after the first sorting are visible in Figure 50.

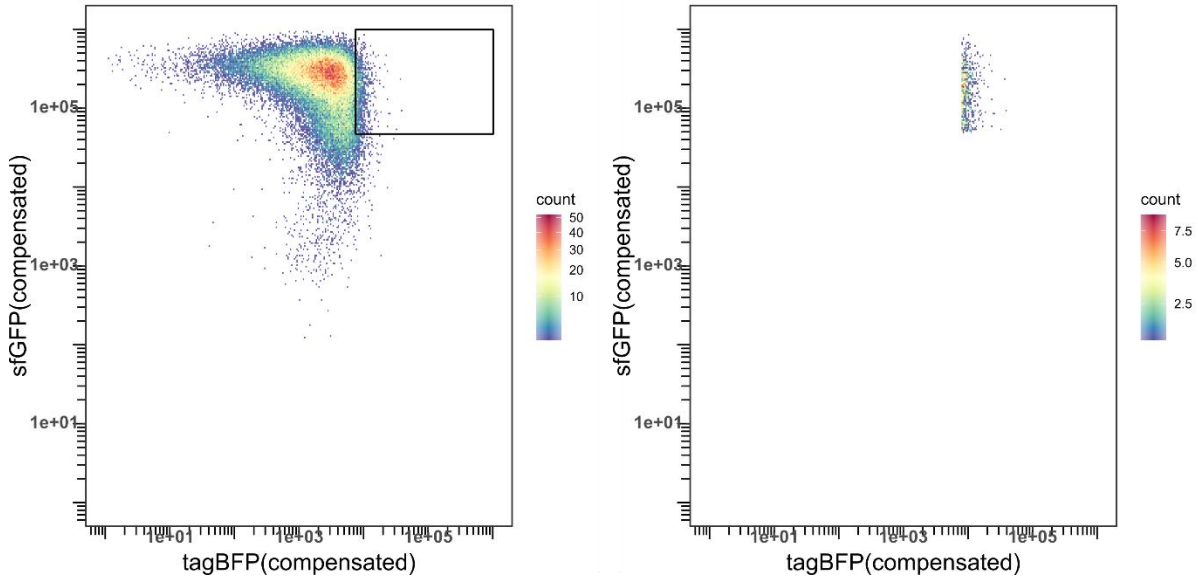


Figure 50: Enriched SRSF1 SICLPPS screening (round 1). Cells were treated with 100 ng/mL anhydrotetracycline prior to treatment with 1 mM IPTG and 0.125 % arabinose. Left: Whole population of the cells in the analysis. Cells that produced high sfGFP levels and high tagBFP levels were gated for sorting (black box). Right: Sorted cells with gated fluorescence properties.

To pursue a screening over multiple rounds, cells have to be recultivated on plates with all three selection markers. In comparison to the SRSF1 SICLOPPS screening with optimised conditions in Figure 49, it is evident that the whole cell population is shifted slightly less towards high tagBFP levels. As a similar shift would have been expected, this could be an indication that fluctuations between experiments are possible. Of $\sim 10.9 \times 10^6$ total events, 0.87 % were gated for sorting (77,063 events). Results of the second sorting can be seen in Figure 51.

4. Results

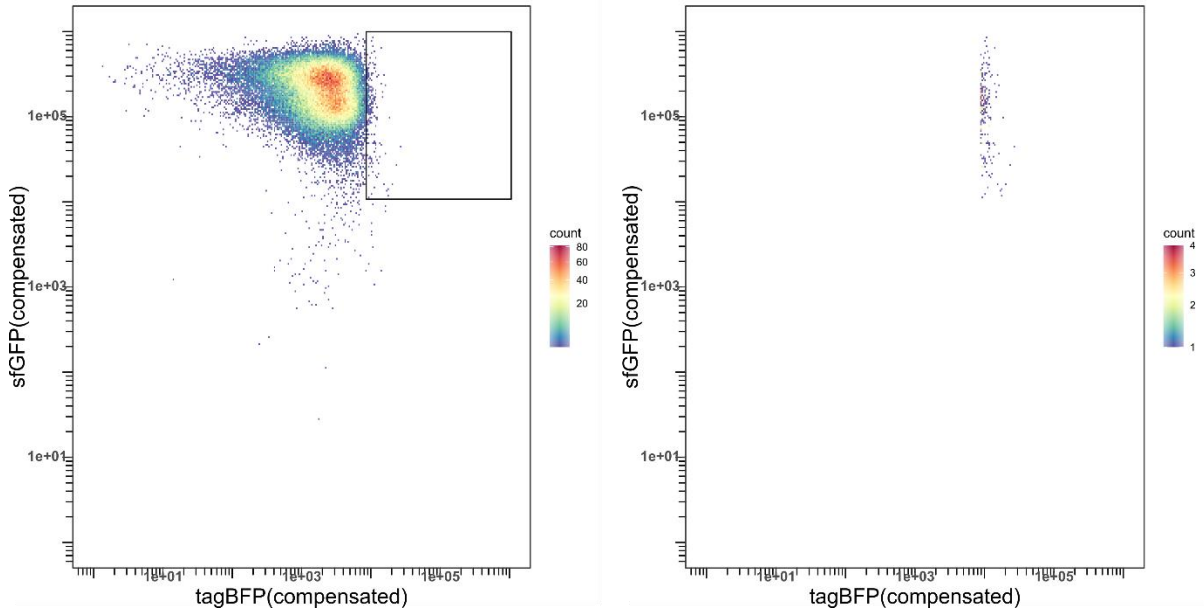


Figure 51: Enriched SRSF1 SICLPPS screening (round 2). Cells were treated with 100 ng/mL anhydrotetracycline prior to treatment with 1 mM IPTG and 0.125 % arabinose. Left: Whole population of the cells in the analysis. Cells that produced high sfGFP levels and high tagBFP levels were gated for sorting (black box). Right: Sorted cells with gated fluorescence properties.

After the second sorting round, an enrichment of cells towards higher tagBFP levels would have been expected but cannot be seen. The gate for sorting was chosen more focused as in the first sorting round and comprised 0.25 % of total events ($\sim 12 \times 10^6$). After the final sorting, cells were plated on agar plates with ampicillin only to prevent cellular stress and increase cell survival. Peptide sequences were analysed by Illumina sequencing and revealed a total of 7259 unique sequences and 2433 unique hit sequences without a stop codon. The latter is approx. one-fifth of the unique hit count of the one-round screening, indicating that a large set of sequences was eliminated. The top 15 hit sequences can be seen in Table 13.

Table 13: Enriched SRSF1 SICLOPPS screening hit sequences and sequence abundance. Entries with stop codon-containing sequences (indicated by X) are highlighted in red.

4. Results

Number	Peptide sequence	Sequence abundance (%)
1	C N T W S X	6.49
2	C X V G I V	4.42
3	C R D N X K	2.92
4	C K X L D L	2.91
5	C N P H X L	2.78
6	C L A L V X	2.75
7	C A W W C X	2.61
8	C S V X I N	2.44
9	C S R L V D	2.35
10	C X C W G V	2.07
11	C X M P L X	2.02
12	C V L X T L	1.91
13	C X M V L V	1.78
14	C V R X I H	1.72
15	C T X G D K	1.69

The two-step sorting led to an enrichment of stop codon-containing sequences. In comparison to the single-step sorting presented in Table 12, there was no enrichment of hit sequences. However, mostly stop codon-containing sequences appear and only one sequence (CSRLVD) is present that can form a cyclic peptide. Cells that carry a variable insert that possesses stop codons do not express the full intein construct and, therefore, might have a growth advantage over cells that express the full-length intein. Only cells without a stop codon in the variable insert are able to perform intein splicing and therefore produce the cyclic peptide. However, it could be possible that cells that do not express the full-length intein are able to express both the sfGFP protein construct and the fluorescent reporter more efficiently as they are less affected by cellular stress. This could have falsified the phenotype of the cells gated for sorting. Nevertheless, the hit peptide CSRLVD was also submitted for chemical synthesis.

4. Results

Similarly, as for the SRSF1 SICLOPPS screening, an hnRNP A2B1 SICLOPPS screening was performed in order to identify inhibitors that target the interaction of hnRNP A2B1 with the H8 RNA. Sorting results can be seen in Figure 52.

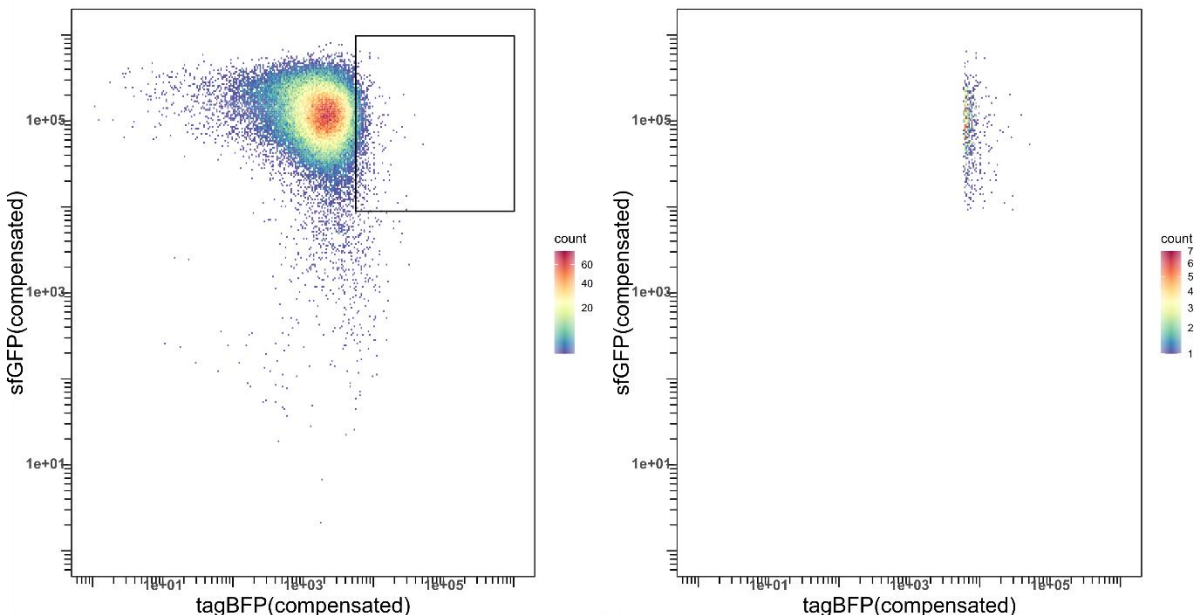


Figure 52: HnRNP A2B1 SICLOPPS screening with optimized conditions. Cells were treated with 100 ng/mL anhydrotetracycline prior to treatment with 1 mM IPTG and 0.125 % arabinose. Left: Whole population of the cells in the analysis. Cells that produced high sfGFP levels and high tagBFP levels were gated for sorting (black box). Right: Sorted cells with gated fluorescence properties.

The cell population is less shifted towards higher tagBFP levels than in the SRSF1 screening, which can be explained due to the fact that the initial tagBFP production rate of H8 in comparison to S3 or S6 is lower. The sorting was performed in two consecutive steps, so one density plot was shown representatively. 0.52 % (28,624 events) of total event count ($\sim 7 \times 10^6$) were gated for the first sorting step, and 0.72 % (28,567 events) of total event count ($\sim 5 \times 10^6$) were gated in the second step. Further sorting statistics can be found in Supplementary table 7. The overall cell count gated for sorting is approx. 2-fold as high as for the SRSF1 sorting. Again, Illumina sequencing was chosen for analysis which revealed 37,207 unique sequences and 30,853 unique sequences without a stop codon. The top 15 hits are shown in Table 14.

Table 14: HnRNP A2B1 SICLOPPS screening hit sequences (hidden) and sequence abundance. Entries with stop codon-containing sequences (indicated by X) are highlighted in red.

Number	Peptide sequence	Sequence abundance (%)
1	- - - - -	1.34

4. Results

2	- - - - - -	1.09
3	- - - - - -	0.53
4	- - - - - -	0.46
5	- - - - - -	0.39
6	- - - X - X	0.35
7	- - - - - -	0.35
8	- - - - - X	0.34
9	- - - - - -	0.31
10	- - - - - X	0.30
11	- - - - - -	0.29
12	- - - - - -	0.27
13	- - X - - -	0.26
14	- X - - - -	0.26
15	- - - - - -	0.25

Of 15 peptide sequences 4 sequences contain stop codons (26.67 %) which is slightly higher than statistically expected, as described above. The same ratio was observed for the SRSF1 SICLOPPS screening. In contrast to the SRSF1 SICLOPPS screening, the most frequent hit sequence is only enriched by 1.34 % (versus 3.03 % for the SRSF1 screening). Also, the difference of the highest-ranked entry to the next entry is lower than in the SRSF1 screening (0.25 % vs. 0.74 %), which suggests that the enrichment is not as strong. The upper 6 hit sequences that did not contain a stop codon were chemically synthesized to evaluate the peptides biophysically.

Sorting over two rounds was not performed for the hnRNP A2B1 screening as an enrichment of stop codon-containing sequences was again expected.

4.3.4. Evaluation of hit peptides by FP

The synthesis of the cyclic peptides was performed by Gulshan Amrahova and Joseph Openy. FP assays were performed by Gulshan Amrahova.

4. Results

4.3.4.1. Evaluation of hit peptides of the SRSF1 SICLOPPS screenings by FP

In order to validate the hits derived from the screenings performed, the hit peptides were chemically synthesised using SPPS. As cyclisation on solid support was rather unsuccessful, a solution-phase cyclisation approach was used. The peptide sequences and molecule numbering, which will be used for further designation, can be seen in the following Table 15 - Table 17.

Table 15: Cyclic peptides derived from the SRSF1 SICLOPPS screening with optimised conditions.

Ranking number	Peptide sequence	Molecule number	Note
1	C P S E W S	SR1-1	1 st isomer
1	C P S E W S	SR1-2	2 nd isomer
2	C L L T E L	SR2-1	1 st isomer
2	C L L T E L	SR2-2	2 nd isomer
3	C L F D S V	SR3	
5	C Y F D V L	SR4	
6	C S E L Q N	SR5-1	1 st isomer
6	C S E L Q N	SR5-2	2 nd isomer
7	C T F S P M	SR6-1	1 st isomer
7	C T F S P M	SR6-2	2 nd isomer

Table 16: Cyclic peptides derived from the enriched SRSF1 SICLOPPS screening with optimised conditions.

Ranking number	Peptide sequence	Molecule number
9	C S R L V D	SR7

During the purification of some of the peptides, two peaks were observed in the HPLC chromatogram with identical mass indicating epimer formation. For these peptides, both isomers were tested once they were obtained after chemical synthesis. FP competition curves for the first tested peptides can be seen in Figure 53.

4. Results

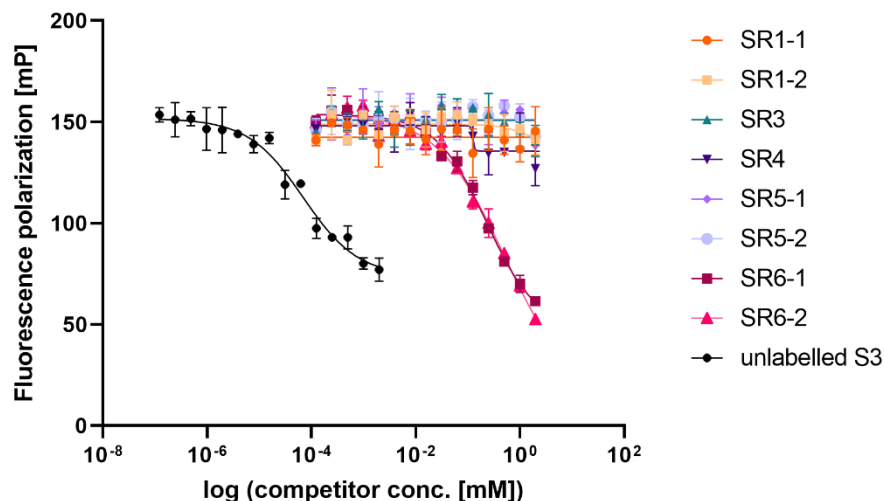
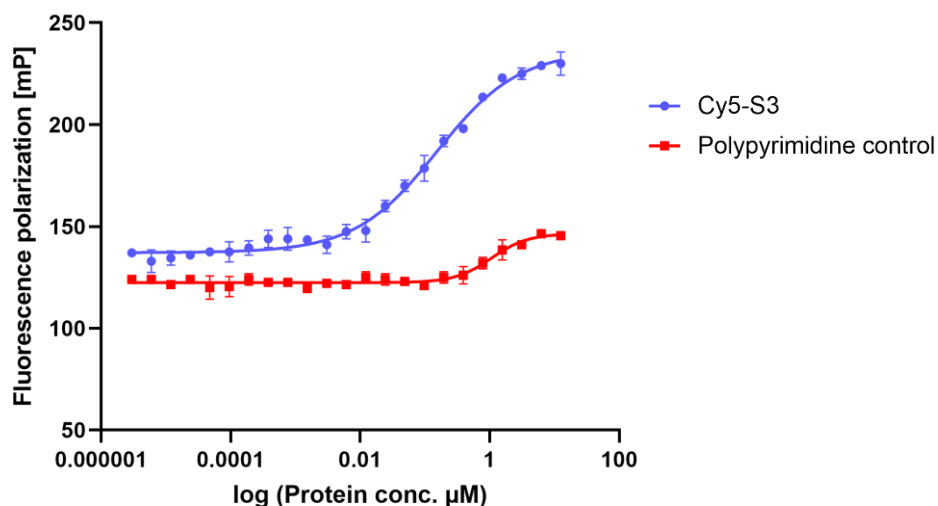


Figure 53: Fluorescence polarisation competition curves of peptides SR1-1, SR1-2, SR3, SR4, SR5-1, SR5-2, SR6-1, SR6-2 and unlabelled S3 RNA with 200 nM SRSF1 and 1 nM FAM-S3. Unlabelled S3 RNA was used as control. Data was measured after 30 minutes incubation and is presented as mean values (N=2).

FP measurements show competition between the unlabelled S3 RNA and the labelled tracer ($IC_{50}=69.42$ nM). Besides that, only the two SR6 isomers show competition-like curves, however, bottom-value saturation is not achieved and an IC_{50} cannot be calculated. The signal reached lower values than the control curve, indicating possible false positives, therefore, the peptides were measured again using Cy5-labelled tracer RNA. First, a direct measurement of the Cy5-labelled S3 RNA was performed to determine optimal competition concentrations of the protein and can be seen in Figure 54.



4. Results

Figure 54: Fluorescence polarisation binding curves of Cy5-S3 and Cy5-Polypyrimidine RNA with SRSF1. Polypyrimidine RNA was used as a negative control. Data was measured after 60 minutes incubation and is presented as mean values (N=2).

The FP measurements provided a K_D value of 152.00 ± 0.28 for the Cy5-labelled S3 RNA, which is approx. 5-fold higher than the K_D of FAM-S3 (compare Table 6). The 5-fold change in affinity seems to be caused by the change of the fluorescent label. As a consequence, the competition concentration for the protein was increased to 400 nM. To verify the competitive behaviour of peptide SR6, another measurement with the Cy5-labelled tracer was performed and can be seen in Figure 55. Representatively, only one isomer was tested and peptides SR3 and SR4 were included.

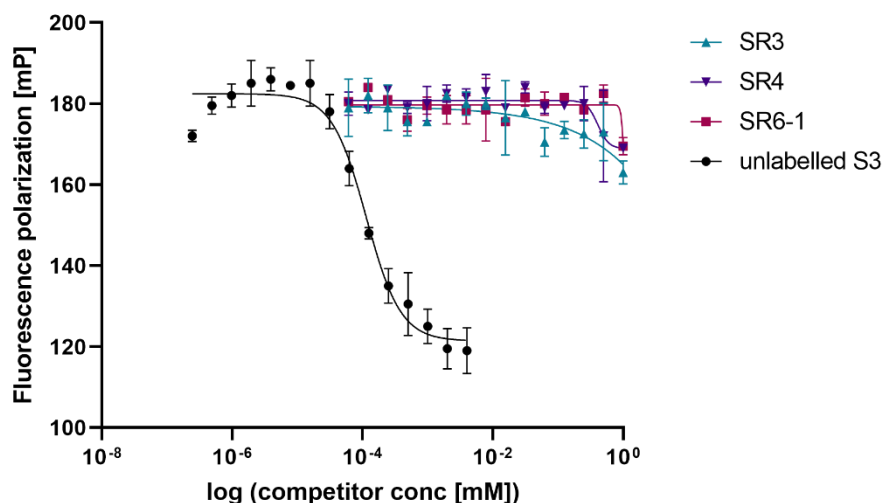


Figure 55: Fluorescence polarisation competition curves of peptides SR3, SR4, SR6-1 and unlabelled S3 RNA with 400 nM SRSF1 and 1 nM Cy5-S3. Unlabelled S3 RNA was used as control. Data was measured after 60 minutes incubation and is presented as mean values (N=2).

The competition-like curve of peptide SR6 observed when using FAM-S3 cannot be reproduced when using the Cy5 labelled RNA as a tracer. The signal change could be an effect of the peptide itself changing polarized light at high concentrations. Peptide SR6 should therefore be excluded as a hit peptide. To exclude false positive signal development caused by the peptide, the Cy5-labelled tracer will be used for the evaluation of SR2 isomers and SR7 (see Figure 56).

4. Results

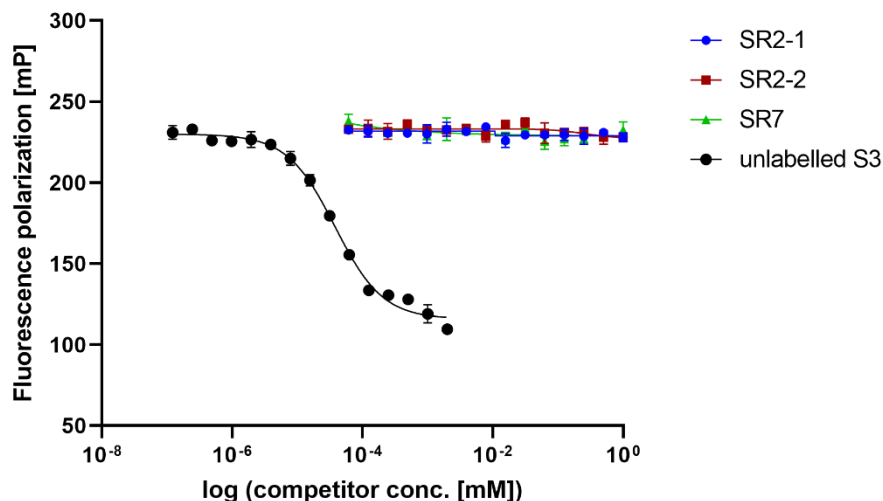


Figure 56: Fluorescence polarisation competition curves of peptides SR2-1, SR2-2, SR7 and unlabelled S3 RNA with 400 nM SRSF1 and 1 nM Cy5-S3. Unlabelled S3 RNA was used as control. Data was measured after 60 minutes incubation and is presented as mean values (N=2).

The FP measurement did not show competition between the peptides with the protein-RNA interaction, while the positive control showed competitive behaviour.

None of the hit peptides SR1-SR7 was able to inhibit the interaction of SRSF1 with S3 RNA, therefore, the SRSF1 SICLOPPS screening was unsuccessful. The interaction chosen as the target for inhibition is of high affinity in the low nanomolar range and also possesses a high repression ratio in context of the TRAP assay that was chosen for screening. In combination with the typically low affinity of primary SICLOPPS hits, it is possible that the interaction of SRSF1 with the S3 RNA is of too high affinity to effectively screen for inhibitors. Potentially an interaction that has a lower repression ratio could lead to a more successful screening outcome. The interaction of SRSF1 with S1 RNA is of low to moderate affinity but has a rather low repression ratio (3.3 ± 0.3) which potentially could be more easily targeted.

4.3.4.2. Evaluation of hit peptides of the hnRNP A2B1 SICLOPPS screenings by FP

To rule out that the repetitive failure of the screenings is a result of the screening setup and rather has to do with the target PRI, the hit peptides for hnRNP A2B1 were evaluated. Peptide sequences and molecule numbers can be seen in Table 17.

Table 17: Cyclic peptides (hidden) derived from the hnRNP A2B1 SICLOPPS screening with optimised conditions.

4. Results

Ranking number	Peptide sequence	Molecule number	Note
1	- - - - -	A1	
2	- - - - -	A2	
3	- - - - -	A3	
4	- - - - -	A4	
5	- - - - -	A5-1	1 st isomer
5	- - - - -	A5-2	2 nd isomer
7	- - - - -	A6	

Equivalent to the PRI pair chosen in the screening, the hnRNP A2B1 interaction with H8 RNA was used as a basis for the FP evaluation. Based on the direct FP measurements (shown in Figure 23), a protein concentration of 25 nM was chosen for competition FP measurements. The measurements were performed on one plate, but for better visibility, were split into three graphs, visible in Figure 57 - Figure 59.

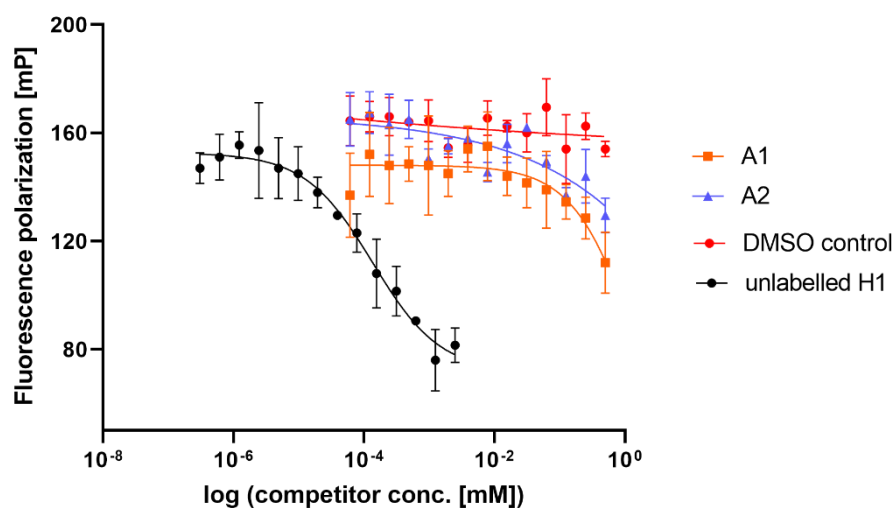


Figure 57: Fluorescence polarisation competition curves of peptides A1, A2 and unlabelled H1 RNA with 25 nM hnRNP A2B1 and 1 nM FAM-H8. Unlabelled H1 RNA and DMSO were used as controls. Data was measured after 60 minutes incubation and is presented as mean values (N=2).

4. Results

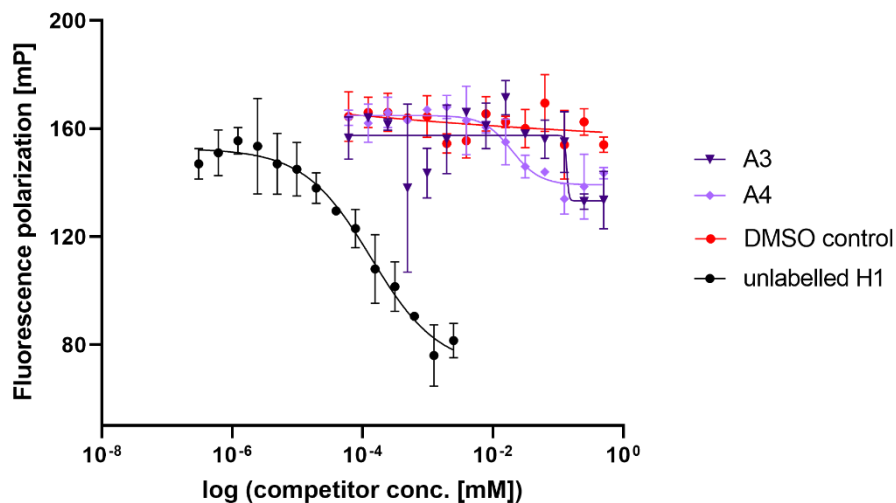


Figure 58: Fluorescence polarisation competition curves of peptides A3, A4 and unlabelled H1 RNA with 25 nM hnRNP A2B1 and 1 nM FAM-H8. Unlabelled H1 RNA and DMSO were used as controls. Peptide A4 was dissolved in 1 % DMSO. Data was measured after 60 minutes incubation and is presented as mean values (N=2).

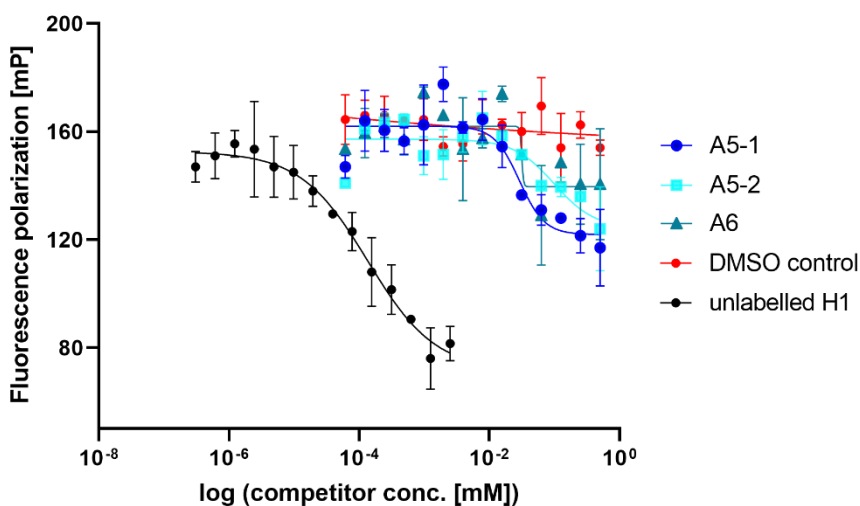


Figure 59: Fluorescence polarisation competition curves of peptides A5-1, A5-2, A6 and unlabelled H1 RNA with 25 nM hnRNP A2B1 and 1 nM FAM-H8. Unlabelled H1 RNA and DMSO were used as controls. Data was measured after 60 minutes incubation and is presented as mean values (N=2).

Unlabelled H1 was chosen as a positive control since unlabelled H8 RNA was not available. Competition between H1 RNA and the tracer is detectable. Of all tested peptides, an IC₅₀ could be determined for three candidates as a competition curve could be fitted. Peptide A4 and both isomers of A5 seem to inhibit the interaction of hnRNP A2B1 and H8 RNA with low

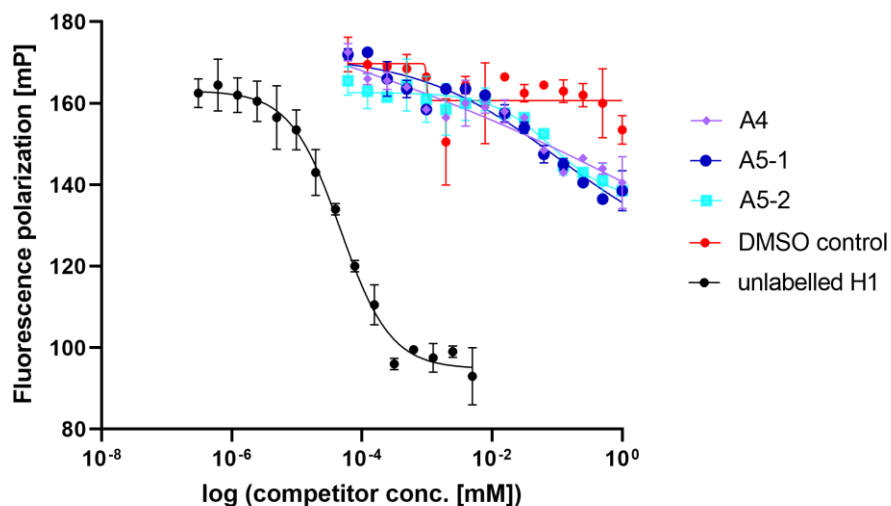
4. Results

to moderate micromolar concentrations (see Table 18). The IC_{50} values derived from the nonlinear regression fit allow calculation of inhibition constants (K_i) according to the reports by Nikolovska-Coleska *et al.*, 2004, which are also reported in Table 18.¹¹⁵

Table 18: FP derived IC_{50} values, Hill slopes and calculated K_i values for the hit peptides tested with hnRNP A2B1 in interaction with FAM-H8 RNA.

Competitor	IC_{50} (μ M)	Hill slope	K_i (μ M)
Unlabelled H1 RNA	0.15 ± 0.10	-0.88 ± 0.25	0.05 ± 0.02
A1	>1000	-	-
A2	>1000	-	-
A3	-	-	-
A4	22.91 ± 15.46	-2.58 ± 1.06	7.83 ± 5.29
A5-1	46.92 ± 29.91	-3.30 ± 3.05	16.04 ± 10.23
A5-2	261.09 ± 312.56	-3.09 ± 3.01	89.26 ± 106.86
A6	>1000	-	-
DMSO control	-	-	-

To confirm that the fitted curves are not a result of false positive signal development competition, experiments with Cy5-labelled H1 RNA were performed. As direct FP measurements were not performed, the competition concentration was slightly increased to 35 nM.



4. Results

Figure 60: Fluorescence polarisation competition curves of peptides A4, A5-1, A5-2 and unlabelled H1 RNA with 35 nM hnRNP A2B1 and 1 nM FAM-H1. Unlabelled H1 RNA and DMSO were used as controls. Data was measured after 60 minutes incubation and is presented as mean values (N=2).

The FP measurement with the Cy-5 labelled RNA shows competition with the unlabelled H1 RNA with the tracer. Competition can also be detected for the peptides A4, A5-1 and A5-2, however, the curve fitting was challenging due to variations in the data points. As a consequence, the determined IC₅₀ values are higher when using the H1 RNA as a tracer, while for peptide A4, the IC₅₀ could not be determined (see Table 19). K_i values could not be calculated as direct binding experiments were not performed.

Table 19: FP derived IC₅₀ values and Hill for the hit peptides tested with hnRNP A2B1 in interaction with Cy5-H1 RNA.

Competitor	IC ₅₀ (µM)	Hill slope
Unlabelled H1 RNA	0.05 ± 0.01	-1.12 ± 0.16
A4	>1000	-
A5-1	161.44 ± 144.48	-0.39 ± 0.03
A5-2	216.36 ± 187.86	-0.86 ± 0.67
DMSO control	-	-

IC₅₀ determination is highly dependent on the assay setup, including the chosen protein concentration and the tracer RNA affinities and concentrations.¹¹⁶ Consequently, IC₅₀ values vary if assay parameters are changed. Nevertheless, both measurements, independent from the tracer RNA, show inhibition of the interaction of hnRNP A2B1 with RNA for the peptides A5-1 and A5-2, albeit at micromolar concentrations, while the measurement performed before also identified A4 as an inhibitor with low micromolar concentrations (see Figure 58 and Table 18). It is remarkable that both FP measurements led to inhibition curves that only cover half of the dynamic range of the unlabelled RNA inhibition curve. Wu *et al.*, 2018 reported that both domains of hnRNP A2B1 are able to bind an individual copy of the 10mer RNA (H1 RNA).²⁷ One explanation could be that the peptide only binds to one of the protein domains resulting in a decrease in fluorescence polarisation to half of the dynamic range, while the other domain is still RNA-bound. Therefore, further characterisation of the most promising candidates using complementary assays should also be performed to gain better insights into protein-peptide binding.

5. Discussion

5.1 A translational reporter assay for the analysis of RNA-binding protein consensus sites

5.1.1. Development of the TRAP assay for SRSF1

The TRAP assay was successfully applied for the human splicing factor SRSF1 in interaction with a synthetic, linear 8 nt RNA construct (AGAAGAAC) derived from a previously performed SELEX experiment,⁸⁹ and variations of it. Beforehand, the assay was presented to work with RBPs mostly of phage or viral origin that interacted with biologically relevant RNA hairpins.^{64,78,80–82,117} The exception of the only human core splicing component that was tested before was the high-affinity interaction of the U1A protein with its cognate interaction partner U1 snRNA hairpin II that has a K_D of ~40 pM.^{81,118} With the results presented here, the TRAP assay was applied to linear inserts with K_D s of up to 92.3 nM, therefore broadening the application spectrum of the assay.

For successful assay development, the design of the protein and RNA reporter constructs should be performed properly. For the protein construct, only the RNA-binding moieties of the SRSF1, RRM1 and 2 were chosen. Full-length constructs were not tested in the TRAP assay as the RS domain possesses disordered structures that could lead to expression difficulties.^{119,120} The attachment of the sfGFP tag can be considered from each terminus but should be chosen rationally to not interfere with RNA binding. In this work, a C-terminal tagged SRSF1 construct was not tested since the success of the N-terminal tagged version was sufficient. RBP expression can be monitored by tag fluorescence and differences in RBP expression are possible, as observed for the RBP controls (PTBP1 and sfGFP only) in comparison to sfGFP-SRSF1, which can be explained by the increased competition for protein synthesis resources. The same holds true for reporter expression that may be altered when different RBP constructs are expressed, as observed for the constructs S1 and S1^b (see Figure 13). It is therefore important to compare RBP constructs of similar expression levels.

Successful development of the assay is also dependent on the reporter expression that, ideally, should be high in basal tagBFP production rate providing a sufficient dynamic range to observe changes in tagBFP production rate accurately. This can be controlled by comparing the basal tagBFP production rate of the unmodified reporter (no insert) to insert-containing reporters. Introduction of an insert should not lower the basal tagBFP production rate

5. Discussion

drastically, while achieving good repression ratios. Both could be influenced by the placement of the insert, which should be close enough to the S/D sequence so translational repression can occur. Boundaries in insert length can be assessed by exploring the influence of a linker on tagBFP production rates, as performed in Figure 14. In comparison to the best binding construct S6, it is evident that long linkers (S6-7 -S6-10) lower the basal tagBFP production rate while also lowering repression ratios which could be explained by RBP binding that is too far from the S/D sequence and, therefore, cannot sterically hinder ribosome binding. Constructs with shorter linkers (S6-4 - S6-6) have a slightly lower tagBFP production rate, yet repression ratios are significantly lower, which indicates that the basal tagBFP production rate is not the only determinant for optimal reporter design. In the case of reporter S6, introduction of an AUA-linker lead to optimal basal tagBFP production rate and repression ratio and is favoured over no linker containing construct S3. As SRSF1 binds purine-rich sequences, it may be possible that the AUA-linker is also bound by the protein and therefore contributes to increasing repression ratios.

Multiplication of the binding motif led to increases in repression ratios that are higher than the sum of the individual motifs suggesting an avidity effect. In literature, it was found that multiplication of splice sites, identical or non-identical, led to additive effects on splicing.¹²¹⁻¹²³ Moreover, SRSF1 binding sites were enriched near 5' or 3' splice sites which is also in line with the observation made with the TRAP assay.

Introduction of point mutations in one or all repeats identified that the motif placed the closest to the S/D sequence is the most important in terms of the effect on repression, further indicating that the binding motif should not be too far upstream of the S/D sequence. The cloning techniques used for the manipulation of the assay plasmids are simple and straightforward, which makes mutation studies easy to perform.

The success of the S-series constructs in TRAP assay context followed rational considerations and allowed to gradually explore design rules and boundaries of the assay.

When checking for *tagBFP* mRNA expression by qPCR after treatment of cells under assay conditions, we did not detect a decrease in mRNA levels (see Figure 15), which proved that the observed effects are truly a result of translational repression and not via unexpected mRNA degradation. Instead, increases in *tagBFP* mRNA are detectable, which is even more pronounced for *kanR* mRNA. Both genes are encoded on the same plasmid and should not be influenced by increasing arabinose concentrations. In addition, the *E. coli* strain used for

5. Discussion

the TRAP assays (Top10 F') contains an araD139 mutation and a $\Delta(\text{ara-leu})$ 7697 deletion and is therefore deficient in the arabinose metabolism.^{124–126} The influence of arabinose on the transcription performance of the cell can therefore be excluded. It can only be speculated that the changes could be caused by an increase in plasmid stability or plasmid replication.

5.1.2. Development of the TRAP assay for hnRNP A2B1

Experiments performed in this section were performed by Mateo Malenica.

In the case of hnRNP A2B1, the TRAP assay could also be adopted, however, bearing more obstacles than the SRSF1 interaction. The protein was tested in interaction with a high affinity, 10 nt sequence, AAGGACUAGC, that itself also appears in linear conformation (see Supplementary figure 19). As before, an N-terminally tagged hnRNP A2B1 construct that comprised the RRM1 and 2 was used. A C-terminal tagged construct was tested as well but showed a reduced repression of 2.0 ± 0.4 ratio to reporter H1 (see Figure 61) in comparison to 5.3 ± 0.4 for the N-terminally tagged version (compare Supplementary figure 5).

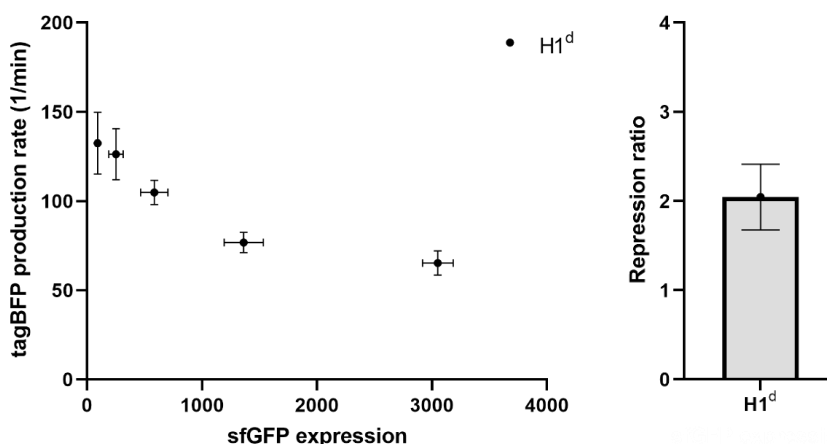


Figure 61: TRAP assay data of reporter construct H1 with hnRNP A2B1-sfGFP (indicated by d). Left: Repression curve. Right: Repression ratio. Data are mean values (n=2, N=2).

In contrast to the S-series reporters, the H-series reporters were less rationally manipulable. Multiplication of binding sequence did not increase repression ratios in an additive or synergistic manner. In addition, basal tagBFP production rates were, on average, much lower than for the S-series. One explanation could be that basal tagBFP production rates are not only dependent on the length of the construct or the distancing between the binding sequence and the S/D sequence but also on the sequence tested. A parameter that cannot be controlled with the TRAP assay is the association of other proteins apart from the sfGFP-RBP fusion to

5. Discussion

the target RNA. It is possible that endogenous host proteins are able to bind specific RNA sequences. Therefore, it is wise, if available, to test multiple binding sequences to find a construct with a good dynamic range and binding capacity.

Another design element that led to increased repression ratios was the application of a spacer. Like SRSF1, hnRNP A2B1 binds purine-rich sequences, so the insertion of a GGG spacer was plausible in terms of better binding and positioning of the motifs. However, insertion of a spacer did not increase basal tagBFP production rates.

Mutation studies with a literature reported mutation that resulted in an 8-fold reduction in affinity was not in line with the TRAP results where only minimal effects were detectable. The affinity measurements were based on ITC experiments with short RNA sequences, where the point mutation had strong effects on the overall binding. In context of the TRAP assay, where the RNA sequence is embedded in the whole mRNA transcript, the protein may also bind to surrounding nucleotides, so a point mutation is probably not enough to reduce overall binding affinity. Also, in contrast to the SRSF1 examples, the construct H11, which carries the mutation near the S/D sequence, led to smaller reductions in the repression ratio in comparison to when the mutations were more upstream of the S/D sequence. A different observation is made for the construct H13-H16, where the mutation in the binding sequence near the S/D (H15) and the most upstream of the S/D sequence (H13) have equally strong effects on repression ratio reduction, while the mutation in the binding sequence placed in the centre had almost no effect on the repression ratio. The latter, however, led to an increase in tagBFP production rate. The construct that had point mutations in all three binding sequences (H16) led to similar decreases in repression ratios as observed for H13 and H15. Especially interesting is the strong increase in basal tagBFP production rate that is the highest observed for the H-series constructs.

5.1.3. Analysis of the effect of secondary structures in the 5'UTR of the reporter mRNA

The above-discussed mutation studies suggest that there are other design elements interfering with successful construct performance. When performing predictions of the secondary structure of the 5'UTR region up until the S/D sequence formation of hairpins in the reporters becomes evident.

Throughout the reporters of the S-series whose binding sequences are not far upstream of the S/D sequence, the construct S1, S4 and S10 are rather low in tagBFP production rate. Relating this to the secondary structure predictions, all three constructs possess hairpins that

5. Discussion

are very close to the S/D sequence (4, 7 and 3 nt distance), while the distancing for other constructs is larger. Indeed, when plotting the basal tagBFP production rates versus nucleotide distancing, trends are visible where longer distancing is favoured, while a maximum in basal tagBFP production rate is reached at a certain distance length. Following this logic, the reporter S0/H0 should initially be very high, which holds true for the H0 reporter while the S0 reporter is slightly decreased. As they are the same sequence-wise, differences in basal tagBFP production rates are not expected. The error for the S0 constructs was rather large, which, taken together with the findings before, suggests that its actual basal tagBFP production rate should be higher.

For the H-series constructs, the overall tagBFP production rate, except for construct H16, is severely low. Apart from the association of endogenous proteins, secondary structure formation also explains low basal tagBFP production rates. Constructs that possess a secondary structure that is distanced >8 nucleotides from the S/D sequence have a higher basal tagBFP production rate, with the exception of construct H10.

Interestingly, inversion of one of the motifs in construct H12 led to an increase in repression ratio and basal tagBFP production rate in comparison to H8. Inverted repeats are a feature in pre-mRNA transcripts and can contribute to both increased or decreased splice site selection as a consequence of structural changes of the mRNA transcript that enhances or represses RBP binding.¹²⁷⁻¹²⁹ Inversion of the binding sequence close to the S/D sequence changes the type of secondary structure formed, as well as the number of distancing nucleotides before the S/D sequence. Both changes are favoured for the construct performance and could be predicted by adopting structural analysis. The construct variation of H8 in which the binding sequence close to the S/D sequence was kept the same, but the one further upstream was inverted was not tested. Structure predictions for this construct, which is complementary to H12 and will be named H17, can be seen in Figure 62.

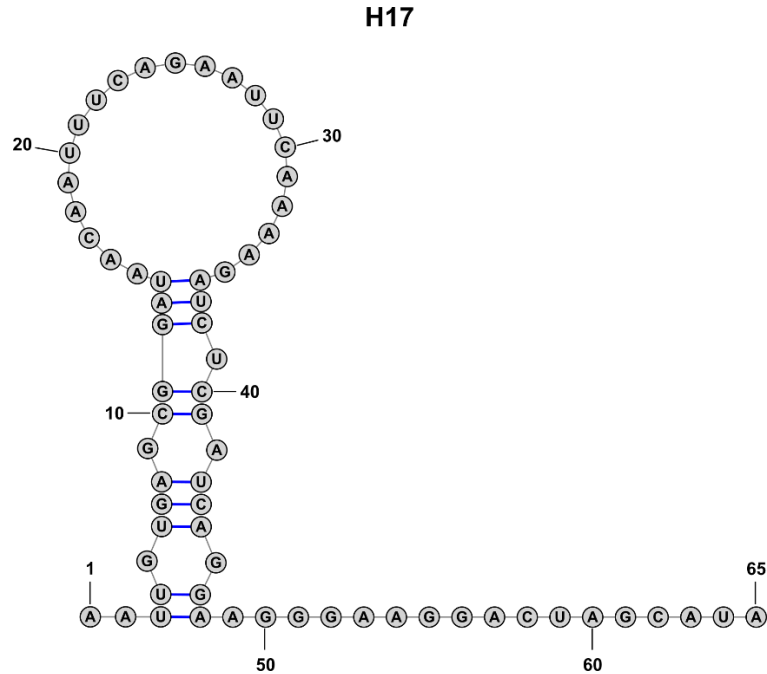


Figure 62: Secondary structure analysis for reporter H17 using RNAfold.

Structural analysis predicts a secondary structure that differs substantially from those made before (compare Supplementary figure 13 - Supplementary figure 17 and Figure 19). The hairpin structure in the 5'UTR is exchanged by a larger secondary structure bearing internal loops and a bulge, while secondary structure formation near the S/D sequence is completely abolished. Through this, the number of distancing nucleotides increases to 17 nucleotides in comparison to reporter H12. According to the correlation plot in Figure 20, this change should lead to peak basal tagBFP production rates and possibly also high repression ratios. In literature, it is proposed that hnRNP A2B1 binds AGGACU sequences, while the domains RRM1 and 2 bind AGG and UAG motifs, respectively.²⁷ Moreover, it is stated that hnRNP A2B1 binds a UAGG or UAGGG motifs and, in general, AG-rich sequences.^{28,130,131} The structure proposed for H17 shows the presence of two AGG and one UAG motif in a non-base paired, linear form in close proximity to the S/D sequence, which is not interrupted by a hairpin. All of these conditions are unique among the H constructs tested and could potentially form the optimal H reporter, but they need to be further tested.

The prediction performed in this study was conducted for the 5'UTR until the S/D sequence, however, there may be influences of the S/D sequence itself or even beyond that could change predictions. Secondary structure predictions are a powerful and inevitable tool for the design of reporter constructs. They mostly function by assuming the structure with the lowest free energy, but the conformation of the RNA is rather determined by the folding kinetics

5. Discussion

instead of the thermodynamic free energy. ¹³² Structure predictions should be therefore considered with caution, yet, in the context of TRAP assay, construct planning was exceptionally helpful.

Successful assay development is a result of proper reporter design that is achieved considering high basal tagBFP production rate, and sufficient dynamic range, which both are influenced by structural elements and, ultimately, a high-affinity binding sequence.

5.1.4. Correlation of translational repression with binding affinity

Measurements of absolute binding affinities by FP confirmed the assumptions about the binding affinity proposed by repression ratios. FP measurements for the SRSF1/RNA series provided more information about the binding ratio of RNA to protein. Multiplication of binding sequences led to increased repression ratios which could have been a result of multiple proteins binding, therefore leading to better translational repression. Binding of multiple copies of the protein to a single RNA can be observed in FP measurements by changes in the Hill slopes (see Table 6). The FAM-S1ext RNA contains the AGAAGAAC motif once, and the Hill slope determined is 1.33 ± 0.32 , while the Hill slopes of the FAM-S2 and FAM-S3 RNA, which contain the motif two and three times, are approx. halved. Binding ratios of 1:1 for FAM-S1ext, 1:2 for FAM-S2 and 1:3 for FAM-S3 would have been expected, and this consideration seems to be true for FAM-S1ext and FAM-S2 according to their Hill slopes. For FAM-S3, a lower Hill slope would have been expected to fit the assumptions. High error bars are detected for the Hill slopes and the K_D s, also, the FP curves of the RNAs are relatively error-prone, which could be a reason that Hill slopes are not exact. K_D values, instead, are almost precisely proportional to one another, where the presence of three motifs led to a ~three-fold reduction in K_D .

A different situation is observed for hnRNP A2B1. While the interaction with FAM-H1 leads to high-affinity binding, duplication of the binding sequence leads to a decrease in binding affinity and triplication of the sequence again leads to high-affinity binding (see Table 6). The drop in affinity for FAM-H1 in comparison to FAM-H2 could be explained through secondary structure predictions (compare Supplementary figure 19), where FAM-H1 is present in a linear conformation and FAM-H2 possesses a hairpin. The FAM-H1 RNA has both motifs, AGG and UAG, in a non-base paired form which are recognized by RRM1 and 2, respectively. This suggests, that binding sequence is bound by both RRMs of a single protein, whereas FAM-H2 can only be bound through its non-base-paired AGG motif. FAM-H3 is also present with a

5. Discussion

secondary structure, in whose terminal loop the AGGACU motif is present in a non-base-paired form. Wu *et al.*, 2018, propose that the AGGACU motif is sufficient to be bound by both RRM1 of the protein, which could be an explanation for why the binding affinity of FAM-H3 is higher than that of FAM-H2.²⁷ However, these approaches do not explain why the binding of FAM-H3 is of higher affinity than FAM-H1.

In the report of Wu *et al.*, 2018, the authors also present the possibility that two proteins bind to two RNA strands in an antiparallel conformation, as depicted in Figure 63.

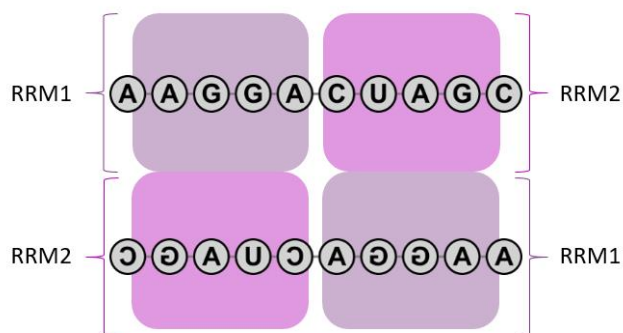


Figure 63: Schematic representation of two hnRNP A2B1 proteins (RRM1 and 2, represented by rectangles in purple tones) in complex with two AGGACUAGC RNAs. The proteins bind the RNAs in an antiparallel conformation. Illustration adapted from Wu *et al.*, 2018.²⁷

A binding mode of this kind could explain why the Hill-slope of H1 is approx. half of the Hill-slope of S1ext, where binding of one protein copy is assumed. However, this would suggest that also two protein copies bind to RNAs H2, H3, H8 and H12. Considering the available binding sites, a case like this is most likely not possible for H3 and H8 RNA as only one binding site is present in a non-base-paired form. Technically, there are two non-base-paired adenine and guanine nucleotides, each in the internal loop of RNA structures H3 and H8 (see Supplementary figure 19 and Supplementary figure 20). It was shown that hnRNP A2B1's structure analogue, hnRNP A1, is able to bind to a loop's base of a G-Quadruplex.¹³³ However, this is little evidence that hnRNP A2B1 is also able to bind in a similar manner.

The K_D measured for FAM-H8 is approx. half of the K_D of FAM-H3. Although both RNAs are similar in the secondary structure they form, the RNAs differ in sequence. The terminal loop of both RNAs exposes a different RNA sequence in a non-base-paired form (AGGACUA vs. AGCGGGA) that may differ in binding affinity to the protein.

The structure of FAM-H12 is, in comparison to all others of the H-series, unique as it contains the AGGACUACG sequence twice and three additional guanine bases, all in a linear

5. Discussion

conformation. Theoretically, this could allow binding of multiple proteins. It is indeed the construct with the lowest K_D , however, the Hill slope does not reflect binding of multiple protein copies or by more proteins than for the other constructs.

The Hill slopes for all constructs are below 1, which in general, suggests that more than one binding protein is present.⁹⁵ Considering all Hill slopes of the H-series, no plausible binding ratios can be calculated, as it was possible for the SRSF1 example, therefore, no further assumptions can be made.

Repression ratios are a good parameter for assessing binding affinities in the context of the TRAP assay and correlate well with absolute binding affinities generated by FP. For the SRSF1 experimental series, repression ratios correlate very well with binding affinities leading to an R^2 value of 0.995. For hnRNP A2B1, however, the correlation is less precise ($R^2=0.775$), which could have to do with the last data point derived from the H12 measurement that is below its theoretical limit. In contrast to the other measurements, the measurement for H12 was performed only once and should therefore be considered with caution.

5.1.5. Analysis of translational repression by flow cytometry

The TRAP assay was also successfully used for flow cytometry analyses as an end-point measurement. The results were presented as histograms of two populations (sample and control) and in comparison, population shifts were well visible for cell populations, which produced high repression ratios but were still present even with lower repression ratios. This holds especially true for the S-series constructs and was less pronounced for the H-series constructs but still significant. Population shifts were small for some samples (as seen in, e.g., Supplementary figure 26), especially when repression ratios were initially also low, indicating that flow cytometry measurements can only successfully be performed when repression ratios are high, i.e. when construct design is well performed. This was achieved when replicating recognition motifs and optimizing reporter design by choosing sufficient distance between secondary structures and S/D sequence and paying attention to secondary structure formation that all contributed to a clearer read-out in both TRAP and flow cytometry.

Flow cytometry analyses, as performed in this work, were valued as successful when population shifts were nicely visible. A parameter that describes population shifts would be a better solution to assess and compare population shifts. Comparing the difference in extrema of the histograms could be a possible parameter to do so.

5. Discussion

The results presented in this chapter prove how the TRAP assay can be useful to study a variety more proteins and provide design rules to successfully apply the assay that can be used in plate format or for flow cytometry.

5.2 Using the TRAP assay to screen for RNA consensus sequences for RBPs

5.2.1. RBP consensus sequence screening for SRSF1 and hnRNPA2B1

Using the TRAP assay as a basis to screen for RBP binding motifs has not been performed before. The simplicity of the TRAP assay allows an easy combination of the protein plasmid with an RNA plasmid library.

In the first screening approach, a fraction of cells that fell in the gate that represented a “repressed phenotype” (high sfGFP and low tagBFP levels) were sorted. As described above, some features of a reporter construct, such as strong secondary structure formation or the association with endogenous proteins to the RNA target site, will lead to low basal tagBFP levels. This fraction of reporter constructs was termed the “autorepressors”, as they intrinsically display the repressed phenotype. The autorepressors were sorted and analysed to serve as a dataset that can be subtracted from any sorting performed. The non-induced sfGFP-SRSF1 plasmid was cotransformed with the induced 10mer library and was used as a model population to generate the dataset.

First, an SRSF1 screening was performed and the autorepressor dataset was subtracted. The upper 10 hits, according to the abundance of hit sequences in the datasets (non-subtracted and subtracted), were cloned and tested using the TRAP assay. The autorepressor removal seemed to be successful as the basal tagBFP production rate of the autorepressor corrected samples was moderate to high, with two exceptions, where almost no tagBFP was present. However, no binding sequences among all 20 tested clones were identified.

Secondary structure prediction of the two autorepressor corrected sequences that barely produced tagBFP shows the presence of bulky structures directly or close to the S/D sequence, which can be seen in Figure 64.

5. Discussion

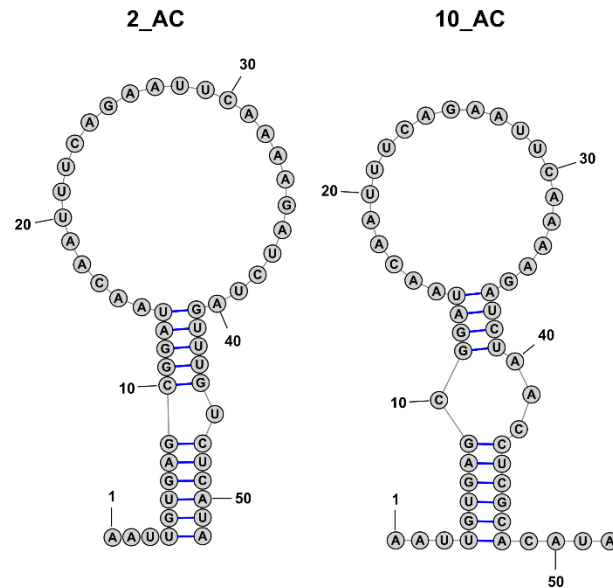


Figure 64: S Secondary structure analysis for reporters 2_AC and 10_AC derived from the SRSF1 RBP consensus sequence screening structures using RNAfold.

According to these analyses, the autorepressor subtraction was only partly successful. This indicates that the gate chosen for the autorepressor selection should have been chosen larger, more towards higher tagBFP levels, to capture more autorepressors. Nevertheless, the screening was unsuccessful even when disregarding the incomplete autorepressor correction.

As reasons for the unsuccessful SRSF1 consensus sequence screening could not be found, an hnRNP A2B1 screening was performed to exclude that the failure was due to SRSF1. The hnRNP A2B1 screening identified two hit sequences that showed binding in the TRAP assay, sequence 3 and 10_AC. Although both sequences do not generate particularly high repression ratios, the screening was considered a partial success. With a closer look at the hit sequences (UCAUUUAGUU and AGCCAGGCGC), the binding motifs UAG and AGG can be found in both sequences, respectively, which strongly suggests hnRNP A2B1 binding. The repression ratios suggest that the AGG motif is slightly favoured over the UAG motif, however, this may also have to do with the binding mode of hnRNP A2B1 and how strongly it affects the steric hindrance of the ribosome.

Interestingly, the autorepressor correction did not work for the hnRNP A2B1 screening, as the basal tagBFP production rate, on average, gets even lower for the autorepressor subtracted samples. This could potentially be explained by the model that was chosen to generate the autorepressor dataset, in which SRSF1 was present. Although the protein being present

5. Discussion

should not have an influence since it is not induced, the protein plasmid is the only parameter that differed.

The success of the screening is only low. Optimally the screening should identify moderate to high binding sequences, while basal tagBFP levels ideally should also be high. Therefore, an optimised screening approach was tested.

5.2.2. Autorepressor presorted RBP consensus sequence screening

Contrary to the screening approach before, where autorepressor sequences were removed by subtraction of datasets, a novel approach was tested for SRSF1, in which autorepressor sequences were eliminated experimentally rather than computationally. At first, the cell population that did not produce a repressed phenotype was presorted, which eliminated ~16 % of total event count. The presorted population was then screened, choosing the same gate as before to focus on cells that produce a strongly repressed phenotype.

The screening identified two hit sequences that were ranked among the upper five hits, which produced moderate repression ratios (4.6 ± 0.5 for Spre3 and 2.5 ± 0.2 for SPre5). The hit sequences ACCGACGGAC (Spre3) and UAGAUAGUC (SPre5) were analysed with the web-based resource ESEfinder 3.0 to identify ESEs for SRSF1. ¹³⁴ For Spre3 indeed, two motifs, CCGACGG and CGACGGA, were identified that had a score above the threshold, therefore indicating at least one binding site for SRSF1. For SPre5, the ESE prediction was unsuccessful. Considering the variable insert length chosen for library construction and the sequence length for the SRSF1 constructs, reporter S4 is the most similar in length and overall design. Reporter S4 produced a repression ratio of 4.8 ± 0.1 , and FP measurements revealed a binding affinity of 92.3 ± 44.1 nM, which suggests that hit sample Spre3 could have a similar affinity. Hit sample Spre3 probably represents the upper limit in binding affinity of a sequence that can be reached with the chosen library design. Hit sample Spre 5 is ranked lower and also has a lower repression ratio suggesting a lower binding affinity.

Surprisingly, only two out of 15 hit sequences were identified as binding sequences. The upper two sequences did not show binding, which was not expected as they were ranked the highest, indicating other reasons why they were falsely positive.

With regard to the autorepressor removal, the screening was not fully successful. Although two hit sequences were found, five of 15 sequences were strongly autorepressed. Technically, autorepressed samples should not be present in the screening population anymore as they were excluded in the presorting. However, they still appear, suggesting that the presorting

5. Discussion

step was not fully efficient and some cells were mistakenly sorted. According to the manufacturer's manual, the sorting mode "purity", chosen in this screening causes stringent sorting of high purity but is not of highest purity possible (>98 %).¹³⁵ To increase the stringency, the sorting mode "Ultra purity" or "Single cell" can be chosen, however, sorting times will increase remarkably. Alternatively, to more accurately remove autorepressors, several rounds of autorepressor elimination can be performed after a screening in an alternating manner, as proposed in Figure 65.

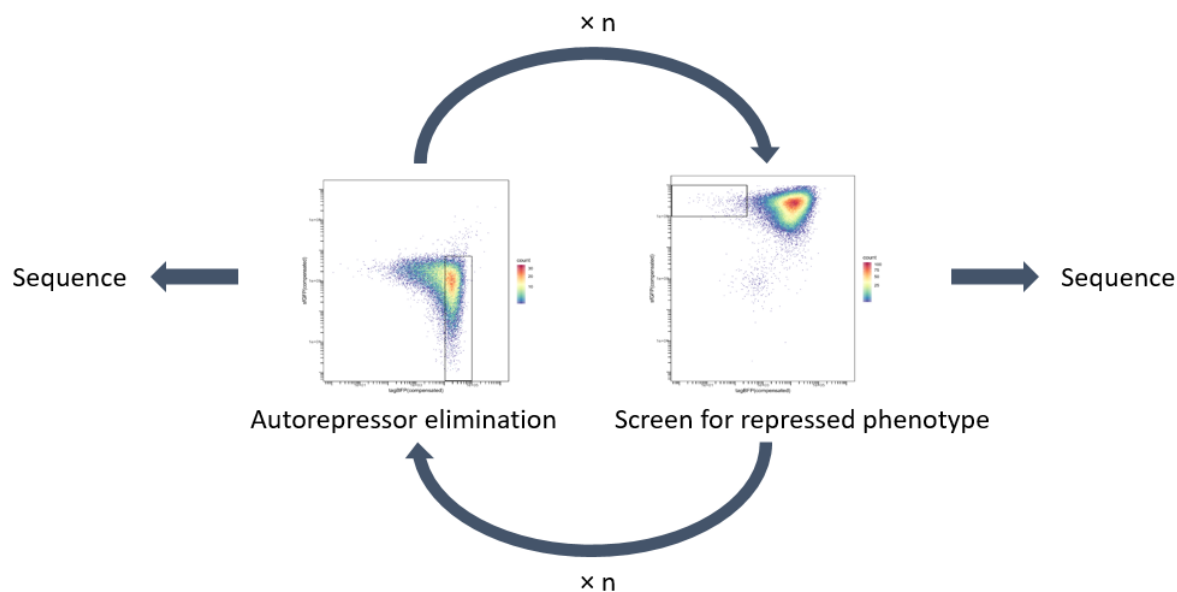


Figure 65: Optimised autorepressor presorted screening, including several rounds of alternating sorting.

After the screening as performed before, another autorepressor elimination step following another sorting step could be included to maximise autorepressor removal. At the same time, an enrichment of hit sequences with high binding affinities is likely. Sequencing can be performed at any step to track the success of the enrichment.

Alternatively, an optimised RNA plasmid design could facilitate the elimination of autorepressor sequences using a life/death selection system. Figure 66 shows an alternative plasmid design for the RNA plasmid, which makes use of a reporter gene fusion to an antibiotic resistance gene.

5. Discussion

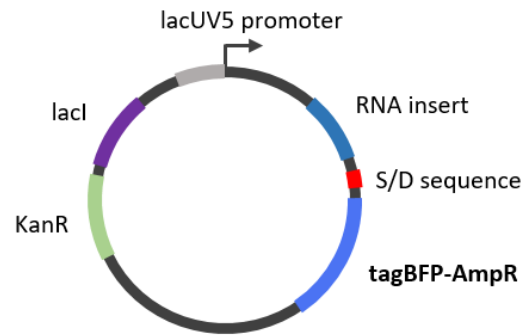


Figure 66: Schematic representation of the optimised RNA plasmids for the RBP consensus sequence screening. The plasmid contains a tagBFP-AmpR fusion gene that allows life/death selection for secondary structures in the 5'UTR.

If secondary structure formation occurs after transcription of the 5'UTR, which leads to autorepression of the reporter gene, the AmpR will not be expressed and cells are not able to grow on selective LB agar. In a second step, cell sorting can be performed as described before while leaving out ampicillin in the growing medium.

The RBP consensus sequence screening is technically possible for proteins of all kinds as long they express well as a fluorescent protein fusion. The screening was successfully performed for two AS factors, though optimisation needs to be considered to increase screening success. Once the RNA plasmid library was available, the screening was performed in a minimal amount of time without the need to express a protein or commercially purchase an RNA library, suggesting a good technique to replace SELEX. Further on, a screening system of the opposite kind with a randomized protein library would allow screen for RBPs for a certain RNA sequence.

5.3 Identification of hexameric peptide inhibitors for the splicing factors SRSF1 and hnRNP A2B1 using SICLOPPS

5.3.1. SICLOPPS screening analysed by Sanger sequencing

Although the use of SICLOPPS in combination with cell-based assay systems such as reverse-two hybrid models was used before, the combination of SICLOPPS with the TRAP assay is a novel approach. The SICLOPPS screening using Sanger sequences analysis for the sorted colonies was unsuccessful. The low survival rate of the cells after sorting (2.66 %) could have led to the elimination of hit sequences. The reason for the low survival can potentially be found in the plating conditions, where three antibiotics were used. That, in

5. Discussion

combination with the sorting procedure, could have led to a low survival rate because of high cellular stress. For sequence analysis of the hit peptides, retention of the SICLOPPS plasmid is needed, while the other plasmids are not useful anymore. Plating should therefore be carried out with only one selection marker. Sorting of cells into nutrient-rich medium such as 2YT or SOC could rescue the influences of cellular stress, however, cell division is rather unwanted as it would increase hit count but not the number of unique hits. This issue is circumvented when using the Illumina sequencing technique, as relative plasmid quantities would not change. Alternatively, it could be considered to incorporate the TRAP assay onto the host genome to circumvent the use of selection markers for retaining assay plasmids over the whole screening.

Another aspect to consider is the ability to handle a larger number of colonies with regard to Sanger sequencing sample preparation. While 90 colonies are rather undemanding to handle, handling a colony count of 3000 is extremely laborious and cost-intensive.

Sequence analysis revealed the presence of 52 sequences, five stop codon-containing sequences and 16 sequences that carried deletions in the backbone. During sequencing sample preparation, the liquid culture preparation for two colonies was unsuccessful, therefore, the cultures were lost. Sample preparation is rather error-prone, as every single colony has to be picked and transferred for analysis and storage.

The SICLOPPS cyclisation mechanism is driven by the enzymatic activity of the split intein moieties.⁵⁸ The report by Tavassoli and Benkovic, 2007, shows an approach to monitor cyclisation by pulling down the intein domains with the CBD tag at the N-intein. Gel-electrophoretic analysis shows that all fusion fragments, the I_C-peptide-I_N, the peptide-I_N, the I_N, the I_C-peptide and the I_C fragment, can be detected, except the cyclic peptide itself due to its small size. The I_N and I_C domains serve as markers for cyclisation.⁵⁹ An experiment using the *E. coli* DHFR gene as the extein moiety shows similar results, where also the linear and cyclic DHFR protein can be detected by gel-electrophoresis and the cyclic product also by mass spectrometry analysis.⁵⁸ Cyclisation analysis, as described above, was not performed for the hit peptides in this work. It can only be speculated whether cyclisation occurred efficiently or not. The *synechocystis sp. (Ssp)* strain PCC6803 derived Intein domains used in this work are less efficient in their splicing activity than the equivalent domains of *Nostoc punctiforme (Npu)*, so incomplete cyclisation is possible.^{136,137} For optimal splicing activity, the Intein domains should be exchanged for the more efficient equivalents of *Npu*.

5. Discussion

To eliminate stop codon-containing cells, an alternative SICLOPPS plasmid design can be considered that uses a fluorescent tag in addition to or as a replacement for the CBD domain. If a stop codon is expressed, the downstream fluorescent tag will not be expressed. FACS analysis and sorting allow to gate for a third fluorescent channel and could therefore be used to eliminate stop codon-containing cells. Instead, an antibiotic resistance gene could be fused to the I_N domain that causes survival of cells only if the full transcript is translated. An approach as such, however, should only be applied if the TRAP assay system is incorporated into the genome to prevent extra cellular stress.

The ranking of hit peptides was performed using the TRAP assay in a modified form, where an extra induction step for the SICLOPPS plasmid was included. A protocol as such has not been proposed in literature. Although ranking with regard to the peptides' ability to restore tagBFP production rate was possible, it is unclear if the restoration truly results from PRI inhibition. The peptides, except peptide 21, do not seem to interfere with tagBFP translation, as can be seen in Figure 42. The inhomogeneous expression of the sfGFP-RBP fusion proteins (shown in Figure 44) indicated that peptide formation might have an influence on sfGFP-RBP expression and, therefore, could lead to reduced translational repression that is reflected as increased tagBFP production rates. The use of SRSF1_{mut} as the positive control protein in a comparative manner might also confound the results as it expresses less efficient than the wild-type.

5.3.2. Synthesis of hexameric, cyclic peptides and evaluation of peptides by FP

Out of six hit peptides, the synthesis of three candidates was successful using the proposed synthesis route, which used SPPS followed by an on-resin cyclisation approach. As the C-terminal cysteine side chain of the peptide is attached to a solid support, low resin substitution was used to reduce dimer formation. The N- and C-terminus of the peptide were deprotected in order to cyclise and dimer formation was potentially possible if the proximity of the linear peptides is too close. In addition, solid-phase cyclisation allows removal of unreacted reagents by repetitive washing steps and also repetition of cyclisation reaction to increase yield.¹¹⁴

However, the synthesis of the remaining three peptides was unsuccessful at the cyclisation step and repetition was equally unsuccessful. Alternatively, a solution-phase cyclisation approach could have been considered. In-solution cyclisation approaches allow multimer formation, which can be minimised by increasing the reaction volume.¹³⁸ This can be

5. Discussion

problematic when DMF is used as a reaction solvent as removal of larger volumes of high boiling point solvents can be challenging.¹³⁹

Evaluation of hit peptides with FP initially revealed peptide 44 as an inhibitor. However, when changing the fluorescent label from FAM to Cy5, this effect could not be reproduced, identifying peptide 44 as a false positive. The signal changes must have resulted from the peptide itself. Peptide 44 (*cyclo*-CYQSYM) carries two tyrosine amino acids, and tyrosine possesses fluorescence properties at low wavelengths also when incorporated into a peptide.¹⁴⁰ However, the excitation and emission wavelengths used to detect tracer fluorescence ((ex/em) 490 nm/520 nm) do not correspond with tyrosine fluorescence properties. In addition, the tyrosine fluorescence emission spectrum is rather not sensitive when changing the local environment.¹⁴¹ Therefore, it is unlikely that peptide 44 interferes with the fluorescence polarisation measurements, and an explanation of why signal changes are visible cannot be found.

5.3.3. SICLOPPS screening analysed by Illumina sequencing

The switch from Sanger sequencing to Illumina sequencing for the analysis of hit peptides led to a 328-fold higher hit count of unique sequences that do not contain a stop codon. With the Sanger sequencing approach, 5 out of 75 hit sequences (6.67 %) contained a stop codon, while with the Illumina sequencing approach, 3,601 out of 16,424 hit sequences (21.93 %) contained a stop codon. The overall chance for stop codon expression in the library design chosen here is 15.63 % which indicates a slight increase in stop codon expression for the Illumina sequencing approach and could be a result of the growth advantage of cells expressing incomplete intein constructs.

None of the hit sequences demonstrated in Table 10 are present in the top hit list of the Illumina-based approach (compare Table 12), which is an indicator for the screening setup analysed by Sanger sequencing to have failed.

Sorting over multiple rounds to enrich hit sequences were unsuccessful, as mostly stop codon-containing sequences were enriched. The only hit sequence to appear in the top 15 list CSRLVD did not inhibit the PRI of SRSF1 with S3 RNA, as visible in Figure 56.

As stop codon appearance is a limitation of the enrichment sorting, the approach described in 5.3.1. should be applied to better compare sorting approaches and determine if enrichment of peptide sequences can truly be achieved.

5. Discussion

The hnRNP A2B1 screening approach, instead, was successful as inhibitory peptides could be found. ~2.3 times as many events were gated for the hnRNP A2B1 screening in comparison to the SRSF1 screening and identified 30,853 unique sequences without a stop codon, which resembles ~2.4 times the count of the SRSF1 screening. The same sequencing depth, 2.22×10^6 reads for hnRNP A2B1 screening vs 2.40×10^6 reads for the SRSF1 screening, was chosen, so the increased hit count seems to result from the increased event gate that was set. Out of 37,207 total unique hit sequences, 6,354 (17.01 %) stop codon-containing sequences appeared in the screening. The percent number is slightly decreased in comparison to the SRSF1 screening (21.93 %) and resembles more the theoretical probability (15.63 %). This indicates that there was minor to no stop codon enrichment in the screening. The reason could have lied in the screening basis. The interaction of hnRNP A2B1 with the H8 reporter was of lower affinity and resulted in a lower repression ratio than the interaction of SRSF1 with the S3 reporter. The screening basis for the SRSF1 SICLOPPS screening was probably not suitable and should be exchanged.

5.3.4. Evaluation of hit peptides by FP

The SRSF1 SICLOPPS screening did not lead to the discovery of inhibiting peptides. As described above, it is most likely that the screening setup was wrongly designed and the changes mentioned should be considered.

For hnRNP A2B1, two peptide candidates, A4 and A5-1, with low micromolar K_i values of $7.83 \pm 5.29 \mu\text{M}$ and $16.04 \pm 10.23 \mu\text{M}$, respectively, were identified as PRI inhibitors by FP competition measurements.

In the work of Tavassoli and Benkovic, 2005, the authors discovered the peptide cyclo-CRYFNV, derived from a SICLOPPS screening, as an inhibitor of the AICAR transformylase homodimerization. Using an *in vitro* AICAR Tfase assay, an initial K_i of $17 \pm 4.2 \mu\text{M}$ was measured, which is in similar range as the K_i measured for peptide A5-1.¹⁴² Based on these results, Spurr *et al.*, 2012, used cyclo-CRYFNV for an alanine scan to decipher the most important amino acids for inhibition and discovered the RY motif as the active fragment. Further SAR studies performed on the RY dipeptide revealed a small molecule with a 25-fold improved K_i ($685 \pm 35 \text{ nM}$) over the parental SICLOPPS hit.¹⁴³ The successful story of cyclo-CRYFNV provides a good outlook for the hit peptides found in this work and similar proceedings should be undertaken.

6. Summary and Outlook

It has to be mentioned that peptide A4 is rather insoluble and 1 % DMSO was needed indicating that further experiments could be challenging.

6. Summary and Outlook

The TRAP assay was successfully applied to two splicing factors, SRSF1 and hnRNP A2B1, in interaction with RNA that harboured consensus sequences derived from *in vitro* experiments or carried known binding motifs. A truncated protein construct with both RRMs but lacking the unstructured protein domain was C-terminally fused to a fluorescent tag that allowed monitoring of relative protein content in the assay. The RNA constructs were cloned directly upstream of the S/D sequence and led to moderate binding given in repression ratios. Reporter RNA optimisation through multiplication of consensus motif, incorporation of linkers and, optionally, insertion of spacers increased repression ratios that correlated well with absolute binding affinities. Secondary structure formation was reflected by low basal tagBFP production rates and could be confirmed by predictions with the online-tool RNAfold. The TRAP assay was also used as an end-point assay by using flow cytometry as a read-out suggesting the applicability of the assay for screening purposes. The utility of the TRAP assay could be expanded to PRIs other than splicing regulators and their cognate RNA interaction partners. The TRAP assay allows to investigate PRIs in cellular context and is a good tool for *in vivo* assessment of binding affinities given in repression ratios.

The use of the TRAP assay was further expanded to work as a screening tool for RBP consensus sequence motifs, albeit with limited success. A plasmid library of 10mer RNA sequences was cloned and screened against the two model proteins, SRSF1 and hnRNP A2B1, in order to identify sequences that are recognized by the RBPs. With the screening design of the 1st generation, no binding sequences could be found for SRSF1, while two binding sequences were found for hnRNP A2B1. The improved design of the assay also led to the identification of two binding sequences for SRSF1. Both screening approaches were contaminated with autorepressor sequences that most likely disturbed assay readout even though autorepressor elimination was attempted. As an improvement, autorepressor removal should be performed with the above-mentioned proposals. The RBP consensus sequence screening platform could be useful for finding binding sequences for any RBP, also non-canonical RBPs, and even for those proteins which have not yet been identified as RBPs.

The TRAP assay was used for the first time as a screening platform for the identification of macrocyclic peptide PRI inhibitors using the SICLOPPS technology. The SRSF1 SICLOPPS

6. Summary and Outlook

screening was unsuccessful, as no peptide inhibitors were identified. As described above, the failure could have been in the choice of the RNA reporter that resulted in a repression ratio and, therefore, binding affinity that was too high. As an improved assay design, a reporter RNA with a lower repression ratio, e.g. the S4 reporter, can be chosen, as it would be more comparable to the screening parameters used for the hnRNP A2B1 SICLOPPS screening, which was successful. Optimisation on the SICLOPPS library plasmid could also be performed by exchanging the split-intein domains with the *Npu* equivalents. Furthermore, it could also be considered to increase the variable insert size to generate larger macrocycles that allow to target larger surface areas.

In the hnRNP A2B1 SICLOPPS screening, two promising peptide candidates were found that inhibit the PRI of hnRNPA2B1 with the H8 RNA at a low micromolar concentration. To verify the binding of both peptide candidates, orthogonal assays such as differential scanning fluorimetry or microscale thermophoresis can be performed, which would also give better insights into protein-peptide binding. Next, an alanine scan of the most promising candidate can be performed to dissect crucial amino acids for peptide binding. Substitutions of less important amino acids by other or unnatural amino acids, or even small molecule fragments, can be considered to increase binding affinity to the protein target.

If a molecule of sufficient affinity is found, first attempts to prove splicing modulation activity should be undertaken by performing, e.g. mini-gene splicing assays with the ultimate goal to find molecules that inhibit the splicing regulatory activity of the protein targets.

7. Literature

- (1) Crick, F. Central Dogma of Molecular Biology. *Nature* **1970**, 227 (5258), 561–563. <https://doi.org/10.1038/227561a0>.
- (2) Ren, X.; Zhang, K.; Deng, R.; Li, J. RNA Splicing Analysis: From In Vitro Testing to Single-Cell Imaging. *Chem* **2019**, 5 (10), 2571–2592. <https://doi.org/10.1016/j.chempr.2019.05.027>.
- (3) Fong, N.; Bentley, D. L. Capping, Splicing, and 3' Processing Are Independently Stimulated by RNA Polymerase II: Different Functions for Different Segments of the CTD. *Genes Dev.* **2001**, 15 (14), 1783–1795. <https://doi.org/10.1101/gad.889101>.
- (4) Gehring, N. H.; Roignant, J.-Y. Anything but Ordinary – Emerging Splicing Mechanisms in Eukaryotic Gene Regulation. *Trends Genet.* **2021**, 37 (4), 355–372. <https://doi.org/10.1016/j.tig.2020.10.008>.
- (5) Zhan, X.; Yan, C.; Zhang, X.; Lei, J.; Shi, Y. Structures of the Human Pre-Catalytic Spliceosome and Its Precursor Spliceosome. *Cell Res.* **2018**, 28 (12), 1129–1140. <https://doi.org/10.1038/s41422-018-0094-7>.
- (6) Ding, F.; Elowitz, M. B. Constitutive Splicing and Economies of Scale in Gene Expression. *Nat. Struct. Mol. Biol.* **2019**, 26 (6), 424–432. <https://doi.org/10.1038/s41594-019-0226-x>.
- (7) David, C. J.; Manley, J. L. Alternative Pre-mRNA Splicing Regulation in Cancer: Pathways and Programs Unhinged. *Genes Dev.* **2010**, 24 (21), 2343–2364. <https://doi.org/10.1101/gad.1973010>.
- (8) Chen, M.; Manley, J. L. Mechanisms of Alternative Splicing Regulation: Insights from Molecular and Genomics Approaches. *Nat. Rev. Mol. Cell Biol.* **2009**, 10 (11), 741–754. <https://doi.org/10.1038/nrm2777>.
- (9) Bradley, R. K.; Anczuków, O. RNA Splicing Dysregulation and the Hallmarks of Cancer. *Nat. Rev. Cancer* **2023**, 23 (3), 135–155. <https://doi.org/10.1038/s41568-022-00541-7>.
- (10) Cerasuolo, A.; Buonaguro, L.; Buonaguro, F. M.; Tornesello, M. L. The Role of RNA Splicing Factors in Cancer: Regulation of Viral and Human Gene Expression in Human Papillomavirus-Related Cervical Cancer. *Front. Cell Dev. Biol.* **2020**, 8.
- (11) Stanley, R. F.; Abdel-Wahab, O. Dysregulation and Therapeutic Targeting of RNA Splicing in Cancer. *Nat. Cancer* **2022**, 3 (5), 536–546. <https://doi.org/10.1038/s43018-022-00384-z>.
- (12) Rooke, N.; Markovtsov, V.; Cagavi, E.; Black, D. L. Roles for SR Proteins and HnRNP A1 in the Regulation of C-Src Exon N1. *Mol. Cell. Biol.* **2003**, 23 (6), 1874–1884. <https://doi.org/10.1128/MCB.23.6.1874-1884.2003>.
- (13) Bielli, P.; Bordi, M.; Biasio, V. D.; Sette, C. Regulation of BCL-X Splicing Reveals a Role for the Polypyrimidine Tract Binding Protein (PTBP1/HnRNP I) in Alternative 5' Splice Site Selection. *Nucleic Acids Res.* **2014**, 42 (19), 12070–12081. <https://doi.org/10.1093/nar/gku922>.
- (14) Vaux, D. L.; Korsmeyer, S. J. Cell Death in Development. *Cell* **1999**, 96 (2), 245–254. [https://doi.org/10.1016/S0092-8674\(00\)80564-4](https://doi.org/10.1016/S0092-8674(00)80564-4).
- (15) Oltean, S.; Bates, D. O. Hallmarks of Alternative Splicing in Cancer. *Oncogene* **2014**, 33 (46), 5311–5318. <https://doi.org/10.1038/onc.2013.533>.
- (16) Anczuków, O.; Rosenberg, A. Z.; Akerman, M.; Das, S.; Zhan, L.; Karni, R.; Muthuswamy, S. K.; Krainer, A. R. The Splicing Factor SRSF1 Regulates Apoptosis and Proliferation to Promote Mammary Epithelial Cell Transformation. *Nat. Struct. Mol. Biol.* **2012**, 19 (2), 220–228. <https://doi.org/10.1038/nsmb.2207>.
- (17) Lo Giudice, A.; Asmundo, M. G.; Broggi, G.; Cimino, S.; Morgia, G.; Di Trapani, E.; Luzzago, S.; Musi, G.; Ferro, M.; de Cobelli, O.; Russo, G. I. The Clinical Role of

7. Literature

- SRSF1 Expression in Cancer: A Review of the Current Literature. *Appl. Sci.* **2022**, *12* (5), 2268. <https://doi.org/10.3390/app12052268>.
- (18) Das, S.; Krainer, A. R. Emerging Functions of SRSF1, Splicing Factor and Oncoprotein, in RNA Metabolism and Cancer. *Mol. Cancer Res.* **2014**, *12* (9), 1195–1204. <https://doi.org/10.1158/1541-7786.MCR-14-0131>.
- (19) Zuo, P.; Manley, J. I. Functional Domains of the Human Splicing Factor ASF/SF2. *EMBO J.* **1993**, *12* (12), 4727–4737. <https://doi.org/10.1002/j.1460-2075.1993.tb06161.x>.
- (20) Cáceres, J. F.; Misteli, T.; Sreaton, G. R.; Spector, D. L.; Krainer, A. R. Role of the Modular Domains of SR Proteins in Subnuclear Localization and Alternative Splicing Specificity. *J. Cell Biol.* **1997**, *138* (2), 225–238. <https://doi.org/10.1083/jcb.138.2.225>.
- (21) Cho, S.; Hoang, A.; Sinha, R.; Zhong, X.-Y.; Fu, X.-D.; Krainer, A. R.; Ghosh, G. Interaction between the RNA Binding Domains of Ser-Arg Splicing Factor 1 and U1-70K SnRNP Protein Determines Early Spliceosome Assembly. *Proc. Natl. Acad. Sci.* **2011**, *108* (20), 8233–8238. <https://doi.org/10.1073/pnas.1017700108>.
- (22) Cáceres, J. F.; Krainer, A. R. Functional Analysis of Pre-mRNA Splicing Factor SF2/ASF Structural Domains. *EMBO J.* **1993**, *12* (12), 4715–4726. <https://doi.org/10.1002/j.1460-2075.1993.tb06160.x>.
- (23) Golan-Gerstl, R.; Cohen, M.; Shilo, A.; Suh, S.-S.; Bakács, A.; Coppola, L.; Karni, R. Splicing Factor HnRNP A2/B1 Regulates Tumor Suppressor Gene Splicing and Is an Oncogenic Driver in Glioblastoma. *Cancer Res.* **2011**, *71* (13), 4464–4472. <https://doi.org/10.1158/0008-5472.CAN-10-4410>.
- (24) Stockley, J.; Villasevil, M. E. M.; Nixon, C.; Ahmad, I.; Leung, H. Y.; Rajan, P. The RNA-Binding Protein HnRNPA2 Regulates β -Catenin Protein Expression and Is Overexpressed in Prostate Cancer. *RNA Biol.* **2014**, *11* (6), 755–765. <https://doi.org/10.4161/rna.28800>.
- (25) Harrison, A. F.; Shorter, J. RNA-Binding Proteins with Prion-like Domains in Health and Disease. *Biochem. J.* **2017**, *474* (8), 1417–1438. <https://doi.org/10.1042/BCJ20160499>.
- (26) Liu, Y.; Shi, S.-L. The Roles of HnRNP A2/B1 in RNA Biology and Disease. *WIREs RNA* **2021**, *12* (2), e1612. <https://doi.org/10.1002/wrna.1612>.
- (27) Wu, B.; Su, S.; Patil, D. P.; Liu, H.; Gan, J.; Jaffrey, S. R.; Ma, J. Molecular Basis for the Specific and Multivalent Recognitions of RNA Substrates by Human HnRNP A2/B1. *Nat. Commun.* **2018**, *9* (1), 420. <https://doi.org/10.1038/s41467-017-02770-z>.
- (28) Yin, M.; Cheng, M.; Liu, C.; Wu, K.; Xiong, W.; Fang, J.; Li, Y.; Zhang, B. HNRNPA2B1 as a Trigger of RNA Switch Modulates the miRNA-Mediated Regulation of CDK6. *iScience* **2021**, *24* (11), 103345. <https://doi.org/10.1016/j.isci.2021.103345>.
- (29) Du, J.-X.; Luo, Y.-H.; Zhang, S.-J.; Wang, B.; Chen, C.; Zhu, G.-Q.; Zhu, P.; Cai, C.-Z.; Wan, J.-L.; Cai, J.-L.; Chen, S.-P.; Dai, Z.; Zhu, W. Splicing Factor SRSF1 Promotes Breast Cancer Progression via Oncogenic Splice Switching of PTPMT1. *J. Exp. Clin. Cancer Res.* **2021**, *40* (1), 171. <https://doi.org/10.1186/s13046-021-01978-8>.
- (30) Shi, X.; Ran, L.; Liu, Y.; Zhong, S.-H.; Zhou, P.-P.; Liao, M.-X.; Fang, W. Knockdown of HnRNP A2/B1 Inhibits Cell Proliferation, Invasion and Cell Cycle Triggering Apoptosis in Cervical Cancer via PI3K/AKT Signaling Pathway. *Oncol. Rep.* **2018**, *39* (3), 939–950. <https://doi.org/10.3892/or.2018.6195>.
- (31) Lakhani, S. A.; Masud, A.; Kuida, K.; Porter, G. A.; Booth, C. J.; Mehal, W. Z.; Inayat, I.; Flavell, R. A. Caspases 3 and 7: Key Mediators of Mitochondrial Events of Apoptosis. *Science* **2006**, *311* (5762), 847–851. <https://doi.org/10.1126/science.1115035>.
- (32) Jiang, L.; Huang, J.; Higgs, B. W.; Hu, Z.; Xiao, Z.; Yao, X.; Conley, S.; Zhong, H.; Liu, Z.; Brohawn, P.; Shen, D.; Wu, S.; Ge, X.; Jiang, Y.; Zhao, Y.; Lou, Y.; Morehouse, C.; Zhu, W.; Sebastian, Y.; Czapiga, M.; Oganessian, V.; Fu, H.; Niu, Y.; Zhang, W.;

7. Literature

- Streicher, K.; Tice, D.; Zhao, H.; Zhu, M.; Xu, L.; Herbst, R.; Su, X.; Gu, Y.; Li, S.; Huang, L.; Gu, J.; Han, B.; Jallal, B.; Shen, H.; Yao, Y. Genomic Landscape Survey Identifies SRSF1 as a Key Oncodriver in Small Cell Lung Cancer. *PLOS Genet.* **2016**, *12* (4), e1005895. <https://doi.org/10.1371/journal.pgen.1005895>.
- (33) Song, S.; Zhang, W.; Li, Q.; Wang, Z.; Zhang, X.; Li, B.; Zhuang, W. *Dysregulation of Alternative Splicing Contributes to Multiple Myeloma Pathogenesis*; preprint; In Review, 2022. <https://doi.org/10.21203/rs.3.rs-1649660/v1>.
- (34) Hua, Y.; Sahashi, K.; Hung, G.; Rigo, F.; Passini, M. A.; Bennett, C. F.; Krainer, A. R. Antisense Correction of SMN2 Splicing in the CNS Rescues Necrosis in a Type III SMA Mouse Model. *Genes Dev.* **2010**, *24* (15), 1634–1644. <https://doi.org/10.1101/gad.1941310>.
- (35) Commissioner, O. of the. *FDA approves first drug for spinal muscular atrophy*. FDA. <https://www.fda.gov/news-events/press-announcements/fda-approves-first-drug-spinal-muscular-atrophy> (accessed 2023-03-29).
- (36) Berciano, M. T.; Puente-Bedia, A.; Medina-Samamé, A.; Rodríguez-Rey, J. C.; Calderó, J.; Lafarga, M.; Tapia, O. Nusinersen Ameliorates Motor Function and Prevents Motoneuron Cajal Body Disassembly and Abnormal Poly(A) RNA Distribution in a SMA Mouse Model. *Sci. Rep.* **2020**, *10*. <https://doi.org/10.1038/s41598-020-67569-3>.
- (37) Hua, Y.; Sahashi, K.; Rigo, F.; Hung, G.; Horev, G.; Bennett, C. F.; Krainer, A. R. Peripheral SMN Restoration Is Essential for Long-Term Rescue of a Severe Spinal Muscular Atrophy Mouse Model. *Nature* **2011**, *478* (7367), 123–126. <https://doi.org/10.1038/nature10485>.
- (38) Singh, N. N.; O’Leary, C. A.; Eich, T.; Moss, W. N.; Singh, R. N. Structural Context of a Critical Exon of Spinal Muscular Atrophy Gene. *Front. Mol. Biosci.* **2022**, *9*.
- (39) Ulanova, M.; Schreiber, A. D.; Befus, A. D. The Future of Antisense Oligonucleotides in the Treatment of Respiratory Diseases. *Biodrugs* **2006**, *20* (1), 1–11. <https://doi.org/10.2165/00063030-200620010-00001>.
- (40) Ratni, H.; Scalco, R. S.; Stephan, A. H. Risdiplam, the First Approved Small Molecule Splicing Modifier Drug as a Blueprint for Future Transformative Medicines. *ACS Med. Chem. Lett.* **2021**, *12* (6), 874–877. <https://doi.org/10.1021/acsmchemlett.0c00659>.
- (41) Yang, X.; Childs-Disney, J. L.; Disney, M. D. A Meditation on Accelerating the Development of Small Molecule Medicines Targeting RNA. *Expert Opin. Drug Discov.* **2023**, *18* (2), 115–117. <https://doi.org/10.1080/17460441.2022.2084528>.
- (42) Childs-Disney, J. L.; Yang, X.; Gibaut, Q. M. R.; Tong, Y.; Batey, R. T.; Disney, M. D. Targeting RNA Structures with Small Molecules. *Nat. Rev. Drug Discov.* **2022**, *21* (10), 736–762. <https://doi.org/10.1038/s41573-022-00521-4>.
- (43) Effenberger, K. A.; Urabe, V. K.; Jurica, M. S. Modulating Splicing with Small Molecular Inhibitors of the Spliceosome: Modulating Splicing with Small Molecular Inhibitors. *Wiley Interdiscip. Rev. RNA* **2017**, *8* (2), e1381. <https://doi.org/10.1002/wrna.1381>.
- (44) Bartling, C. R. O.; Alexopoulou, F.; Kuschert, S.; Chin, Y. K.-Y.; Jia, X.; Sereikaite, V.; Özcelik, D.; Jensen, T. M.; Jain, P.; Nygaard, M. M.; Harpsøe, K.; Gloriam, D. E.; Mobli, M.; Strømgaard, K. Comprehensive Peptide Cyclization Examination Yields Optimized APP Scaffolds with Improved Affinity toward Mint2. *J. Med. Chem.* **2023**, *66* (4), 3045–3057. <https://doi.org/10.1021/acs.jmedchem.2c02017>.
- (45) Hu, L.; Liu, S.; Yao, H.; Hu, Y.; Wang, Y.; Jiang, J.; Li, X.; Fu, F.; Yin, Q.; Wang, H. Identification of a Novel Heterogeneous Nuclear Ribonucleoprotein A2B1 (HnRNPA2B1) Ligand That Disrupts HnRNPA2B1/Nucleic Acid Interactions to Inhibit the MDMX-P53 Axis in Gastric Cancer. *Pharmacol. Res.* **2023**, *189*, 106696. <https://doi.org/10.1016/j.phrs.2023.106696>.

7. Literature

- (46) Verdine, G. L.; Walensky, L. D. The Challenge of Drugging Undruggable Targets in Cancer: Lessons Learned from Targeting BCL-2 Family Members. *Clin. Cancer Res.* **2007**, *13* (24), 7264–7270. <https://doi.org/10.1158/1078-0432.CCR-07-2184>.
- (47) Mukherjee, S.; Bahadur, R. P. An Account of Solvent Accessibility in Protein-RNA Recognition. *Sci. Rep.* **2018**, *8* (1), 10546. <https://doi.org/10.1038/s41598-018-28373-2>.
- (48) Wu, Z.; Hu, G.; Yang, J.; Peng, Z.; Uversky, V. N.; Kurgan, L. In Various Protein Complexes, Disordered Protomers Have Large per-Residue Surface Areas and Area of Protein-, DNA- and RNA-Binding Interfaces. *FEBS Lett.* **2015**, *589* (19, Part A), 2561–2569. <https://doi.org/10.1016/j.febslet.2015.08.014>.
- (49) Wang, L.; Wang, N.; Zhang, W.; Cheng, X.; Yan, Z.; Shao, G.; Wang, X.; Wang, R.; Fu, C. Therapeutic Peptides: Current Applications and Future Directions. *Signal Transduct. Target. Ther.* **2022**, *7* (1), 1–27. <https://doi.org/10.1038/s41392-022-00904-4>.
- (50) Pal, S.; 't Hart, P. RNA-Binding Macrocyclic Peptides. *Front. Mol. Biosci.* **2022**, *9*.
- (51) Tavassoli, A. SICLOPPS Cyclic Peptide Libraries in Drug Discovery. *Curr. Opin. Chem. Biol.* **2017**, *38*, 30–35. <https://doi.org/10.1016/j.cbpa.2017.02.016>.
- (52) Smith, G. P. Phage Display: Simple Evolution in a Petri Dish (Nobel Lecture). *Angew. Chem. Int. Ed.* **2019**, *58* (41), 14428–14437. <https://doi.org/10.1002/anie.201908308>.
- (53) Smith, G. P.; Petrenko, V. A. Phage Display. *Chem. Rev.* **1997**, *97* (2), 391–410. <https://doi.org/10.1021/cr960065d>.
- (54) Kamalinia, G.; Grindel, B. J.; Takahashi, T. T.; Millward, S. W.; Roberts, R. W. Directing Evolution of Novel Ligands by mRNA Display. *Chem. Soc. Rev.* **2021**, *50* (16), 9055–9103. <https://doi.org/10.1039/D1CS00160D>.
- (55) Blanco, C.; Verbanic, S.; Seelig, B.; A. Chen, I. High Throughput Sequencing of in Vitro Selections of mRNA-Displayed Peptides: Data Analysis and Applications. *Phys. Chem. Chem. Phys.* **2020**, *22* (12), 6492–6506. <https://doi.org/10.1039/C9CP05912A>.
- (56) Lin, H.; Cornish, V. W. Screening and Selection Methods for Large-Scale Analysis of Protein Function. *Angew. Chem. Int. Ed.* **2002**, *41* (23), 4402–4425. [https://doi.org/10.1002/1521-3773\(20021202\)41:23<4402::AID-ANIE4402>3.0.CO;2-H](https://doi.org/10.1002/1521-3773(20021202)41:23<4402::AID-ANIE4402>3.0.CO;2-H).
- (57) Wang, H.; Liu, R. Advantages of mRNA Display Selections over Other Selection Techniques for Investigation of Protein–Protein Interactions. *Expert Rev. Proteomics* **2014**, *8* (3), 335–346. <https://doi.org/10.1586/epr.11.15>.
- (58) Scott, C. P.; Abel-Santos, E.; Wall, M.; Wahnou, D. C.; Benkovic, S. J. Production of Cyclic Peptides and Proteins in Vivo. *Proc. Natl. Acad. Sci.* **1999**, *96* (24), 13638–13643. <https://doi.org/10.1073/pnas.96.24.13638>.
- (59) Tavassoli, A.; Benkovic, S. J. Split-Intein Mediated Circular Ligation Used in the Synthesis of Cyclic Peptide Libraries in E. Coli. *Nat. Protoc.* **2007**, *2* (5), 1126–1133. <https://doi.org/10.1038/nprot.2007.152>.
- (60) Miranda, E.; Nordgren, I. K.; Male, A. L.; Lawrence, C. E.; Hoakwie, F.; Cuda, F.; Court, W.; Fox, K. R.; Townsend, P. A.; Packham, G. K.; Eccles, S. A.; Tavassoli, A. A Cyclic Peptide Inhibitor of HIF-1 Heterodimerization That Inhibits Hypoxia Signaling in Cancer Cells. *J. Am. Chem. Soc.* **2013**, *135* (28), 10418–10425. <https://doi.org/10.1021/ja402993u>.
- (61) Cheng, L.; Naumann, T. A.; Horswill, A. R.; Hong, S.-J.; Venters, B. J.; Tomsho, J. W.; Benkovic, S. J.; Keiler, K. C. Discovery of Antibacterial Cyclic Peptides That Inhibit the ClpXP Protease. *Protein Sci.* **2007**, *16* (8), 1535–1542. <https://doi.org/10.1110/ps.072933007>.
- (62) Li, M. Z.; Elledge, S. J. SLIC: A Method for Sequence- and Ligation-Independent Cloning. In *Gene Synthesis: Methods and Protocols*; Peccoud, J., Ed.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2012; pp 51–59. https://doi.org/10.1007/978-1-61779-564-0_5.

7. Literature

- (63) Katz, N.; Cohen, R.; Atar, O.; Goldberg, S.; Amit, R. An Assay for Quantifying Protein-RNA Binding in Bacteria. *J. Vis. Exp.* **2019**, No. 148, 59611. <https://doi.org/10.3791/59611>.
- (64) Katz, N.; Cohen, R.; Solomon, O.; Kaufmann, B.; Atar, O.; Yakhini, Z.; Goldberg, S.; Amit, R. Synthetic 5' UTRs Can Either Up- or Downregulate Expression upon RNA-Binding Protein Binding. *Cell Syst.* **2019**, 9 (1), 93-106.e8. <https://doi.org/10.1016/j.cels.2019.04.007>.
- (65) Livak, K. J.; Schmittgen, T. D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* **2001**, 25 (4), 402–408. <https://doi.org/10.1006/meth.2001.1262>.
- (66) The Galaxy Community. The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses: 2022 Update. *Nucleic Acids Res.* **2022**, 50 (W1), W345–W351. <https://doi.org/10.1093/nar/gkac247>.
- (67) Nowacki, J.; Malenica, M.; Schmeing, S.; Schiller, D.; Buchmuller, B.; Amrahova, G.; Hart, P. A Translational Repression Reporter Assay for the Analysis of RNA-Binding Protein Consensus Sites. *RNA Biol.* **2023**, 20 (1), 85–94. <https://doi.org/10.1080/15476286.2023.2192553>.
- (68) Cléry, A.; Sinha, R.; Anczuków, O.; Corrionero, A.; Moursy, A.; Daubner, G. M.; Valcárcel, J.; Krainer, A. R.; Allain, F. H.-T. Isolated Pseudo-RNA-Recognition Motifs of SR Proteins Can Regulate Splicing Using a Noncanonical Mode of RNA Recognition. *Proc. Natl. Acad. Sci.* **2013**, 110 (30). <https://doi.org/10.1073/pnas.1303445110>.
- (69) Cléry, A.; Jayne, S.; Benderska, N.; Dominguez, C.; Stamm, S.; Allain, F. H.-T. Molecular Basis of Purine-Rich RNA Recognition by the Human SR-like Protein Tra2-B1. *Nat. Struct. Mol. Biol.* **2011**, 18 (4), 443–450. <https://doi.org/10.1038/nsmb.2001>.
- (70) Djordjevic, M. SELEX Experiments: New Prospects, Applications and Data Analysis in Inferring Regulatory Pathways. *Biomol. Eng.* **2007**, 24 (2), 179–189. <https://doi.org/10.1016/j.bioeng.2007.03.001>.
- (71) Cozzolino, F.; Iacobucci, I.; Monaco, V.; Monti, M. Protein–DNA/RNA Interactions: An Overview of Investigation Methods in the -Omics Era. *J. Proteome Res.* **2021**, 20 (6), 3018–3030. <https://doi.org/10.1021/acs.jproteome.1c00074>.
- (72) Eunen, K. van; Kiewiet, J. A. L.; Westerhoff, H. V.; Bakker, B. M. Testing Biochemistry Revisited: How In Vivo Metabolism Can Be Understood from In Vitro Enzyme Kinetics. *PLOS Comput. Biol.* **2012**, 8 (4), e1002483. <https://doi.org/10.1371/journal.pcbi.1002483>.
- (73) Tan, S. C.; Yiap, B. C. DNA, RNA, and Protein Extraction: The Past and The Present. *J. Biomed. Biotechnol.* **2009**, 2009, 574398. <https://doi.org/10.1155/2009/574398>.
- (74) Stockert, O. M.; Gravel, C. M.; Berry, K. E. A Bacterial Three-Hybrid Assay for Forward and Reverse Genetic Analysis of RNA–Protein Interactions. *Nat. Protoc.* **2022**, 17 (4), 941–961. <https://doi.org/10.1038/s41596-021-00657-4>.
- (75) Peled-Zehavi, H.; Horiya, S.; Das, C.; Harada, K.; Frankel, A. D. Selection of RRE RNA Binding Peptides Using a Kanamycin Antitermination Assay. *RNA* **2003**, 9 (2), 252–261. <https://doi.org/10.1261/rna.2152303>.
- (76) Wilhelm, J. E.; Vale, R. D. A One-Hybrid System for Detecting RNA-Protein Interactions. *Genes Cells* **1996**, 1 (3), 317–323. <https://doi.org/10.1046/j.1365-2443.1996.25026.x>.
- (77) Harada, K.; Martin, S. S.; Frankel, A. D. Selection of RNA-Binding Peptides in Vivo. *Nature* **1996**, 380 (6570), 175–179. <https://doi.org/10.1038/380175a0>.
- (78) Jain, C.; Belasco, J. G. A Structural Model for the HIV-1 Rev–RRE Complex Deduced from Altered-Specificity Rev Variants Isolated by a Rapid Genetic Strategy. *Cell* **1996**, 87 (1), 115–125. [https://doi.org/10.1016/S0092-8674\(00\)81328-8](https://doi.org/10.1016/S0092-8674(00)81328-8).

7. Literature

- (79) HOOK, B.; BERNSTEIN, D.; ZHANG, B.; WICKENS, M. RNA–Protein Interactions in the Yeast Three-Hybrid System: Affinity, Sensitivity, and Enhanced Library Screening. *RNA* **2005**, *11* (2), 227–233. <https://doi.org/10.1261/rna.7202705>.
- (80) Jain, C. An Escherichia Coli-Based Genetic Strategy for Characterizing RNA Binding Proteins. In *RNA-Protein Interaction Protocols*; Humana Press: New Jersey, 1999; Vol. 118, pp 161–175. <https://doi.org/10.1385/1-59259-676-2:161>.
- (81) Paraskeva, E.; Atzberger, A.; Hentze, M. W. A Translational Repression Assay Procedure (TRAP) for RNA–Protein Interactions *in Vivo*. *Proc. Natl. Acad. Sci.* **1998**, *95* (3), 951–956. <https://doi.org/10.1073/pnas.95.3.951>.
- (82) Katz, N.; Cohen, R.; Solomon, O.; Kaufmann, B.; Atar, O.; Yakhini, Z.; Goldberg, S.; Amit, R. An *in Vivo* Binding Assay for RNA-Binding Proteins Based on Repression of a Reporter Gene. *ACS Synth. Biol.* **2018**, *7* (12), 2765–2774. <https://doi.org/10.1021/acssynbio.8b00378>.
- (83) Anczuków, O.; Akerman, M.; Cléry, A.; Wu, J.; Shen, C.; Shirole, N. H.; Raimer, A.; Sun, S.; Jensen, M. A.; Hua, Y.; Allain, F. H.-T.; Krainer, A. R. SRSF1-Regulated Alternative Splicing in Breast Cancer. *Mol. Cell* **2015**, *60* (1), 105–117. <https://doi.org/10.1016/j.molcel.2015.09.005>.
- (84) Cléry, A.; Krepl, M.; Nguyen, C. K. X.; Moursy, A.; Jorjani, H.; Katsantoni, M.; Okoniewski, M.; Mittal, N.; Zavolan, M.; Sponer, J.; Allain, F. H.-T. Structure of SRSF1 RRM1 Bound to RNA Reveals an Unexpected Bimodal Mode of Interaction and Explains Its Involvement in SMN1 Exon7 Splicing. *Nat. Commun.* **2021**, *12* (1), 428. <https://doi.org/10.1038/s41467-020-20481-w>.
- (85) Nie, M.; Htun, H. Different Modes and Potencies of Translational Repression by Sequence-Specific RNA-Protein Interaction at the 5'-UTR. *Nucleic Acids Res.* **2006**, *34* (19), 5528–5540. <https://doi.org/10.1093/nar/gkl584>.
- (86) Ono, H.; Kawasaki, S.; Saito, H. Orthogonal Protein-Responsive mRNA Switches for Mammalian Synthetic Biology. *ACS Synth. Biol.* **2020**, *9* (1), 169–174. <https://doi.org/10.1021/acssynbio.9b00343>.
- (87) Han, A.; Stoilov, P.; Linares, A. J.; Zhou, Y.; Fu, X.-D.; Black, D. L. De Novo Prediction of PTBP1 Binding and Splicing Targets Reveals Unexpected Features of Its RNA Recognition and Function. *PLOS Comput. Biol.* **2014**, *10* (1), e1003442. <https://doi.org/10.1371/journal.pcbi.1003442>.
- (88) Denichenko, P.; Mogilevsky, M.; Cléry, A.; Welte, T.; Biran, J.; Shimshon, O.; Barnabas, G. D.; Danan-Gotthold, M.; Kumar, S.; Yavin, E.; Levanon, E. Y.; Allain, F. H.; Geiger, T.; Levkowitz, G.; Karni, R. Specific Inhibition of Splicing Factor Activity by Decoy RNA Oligonucleotides. *Nat. Commun.* **2019**, *10* (1), 1590. <https://doi.org/10.1038/s41467-019-09523-0>.
- (89) Tacke, R.; Manley, J. L. The Human Splicing Factors ASF/SF2 and SC35 Possess Distinct, Functionally Significant RNA Binding Specificities. *EMBO J.* **1995**, *14* (14), 3540–3551. <https://doi.org/10.1002/j.1460-2075.1995.tb07360.x>.
- (90) Brenner, S.; Jacob, F.; Meselson, M. An Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis. *Nature* **1961**, *190* (4776), 576–581. <https://doi.org/10.1038/190576a0>.
- (91) Jain, C. Degradation of mRNA in Escherichia Coli. *IUBMB Life* **2002**, *54* (6), 315–321. <https://doi.org/10.1080/15216540216036>.
- (92) de Smit, M. H.; van Duin, J. Secondary Structure of the Ribosome Binding Site Determines Translational Efficiency: A Quantitative Analysis. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87* (19), 7668–7672. <https://doi.org/10.1073/pnas.87.19.7668>.
- (93) Gold, L. Posttranscriptional Regulatory Mechanisms in Escherichia Coli. *Annu. Rev. Biochem.* **1988**, *57*, 199–233. <https://doi.org/10.1146/annurev.bi.57.070188.001215>.

7. Literature

- (94) Gruber, A. R.; Lorenz, R.; Bernhart, S. H.; Neuböck, R.; Hofacker, I. L. The Vienna RNA Websuite. *Nucleic Acids Res.* **2008**, *36* (Web Server issue), W70-74. <https://doi.org/10.1093/nar/gkn188>.
- (95) Prinz, H. Hill Coefficients, Dose–Response Curves and Allosteric Mechanisms. *J. Chem. Biol.* **2009**, *3* (1), 37–44. <https://doi.org/10.1007/s12154-009-0029-3>.
- (96) Jolma, A.; Zhang, J.; Mondragón, E.; Morgunova, E.; Kivioja, T.; Laverly, K. U.; Yin, Y.; Zhu, F.; Bourenkov, G.; Morris, Q.; Hughes, T. R.; Maher, L. J.; Taipale, J. Binding Specificities of Human RNA-Binding Proteins toward Structured and Linear RNA Sequences. *Genome Res.* **2020**, *30* (7), 962–973. <https://doi.org/10.1101/gr.258848.119>.
- (97) Hentze, M. W.; Castello, A.; Schwarzl, T.; Preiss, T. A Brave New World of RNA-Binding Proteins. *Nat. Rev. Mol. Cell Biol.* **2018**, *19* (5), 327–341. <https://doi.org/10.1038/nrm.2017.130>.
- (98) Perez-Perri, J. I.; Rogell, B.; Schwarzl, T.; Stein, F.; Zhou, Y.; Rettel, M.; Brosig, A.; Hentze, M. W. Discovery of RNA-Binding Proteins and Characterization of Their Dynamic Responses by Enhanced RNA Interactome Capture. *Nat. Commun.* **2018**, *9* (1), 4408. <https://doi.org/10.1038/s41467-018-06557-8>.
- (99) Quattrone, A.; Dassi, E. The Architecture of the Human RNA-Binding Protein Regulatory Network. *iScience* **2019**, *21*, 706–719. <https://doi.org/10.1016/j.isci.2019.10.058>.
- (100) Lee, F. C. Y.; Ule, J. Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Mol. Cell* **2018**, *69* (3), 354–369. <https://doi.org/10.1016/j.molcel.2018.01.005>.
- (101) Munteanu, A.; Mukherjee, N.; Ohler, U. SSMART: Sequence-Structure Motif Identification for RNA-Binding Proteins. *Bioinformatics* **2018**, *34* (23), 3990–3998. <https://doi.org/10.1093/bioinformatics/bty404>.
- (102) Xue, Y.; Zhou, Y.; Wu, T.; Zhu, T.; Ji, X.; Kwon, Y.-S.; Zhang, C.; Yeo, G.; Black, D. L.; Sun, H.; Xiang-Dong, F.; Zhang, Y. Genome-Wide Analysis of PTB-RNA Interactions Reveals a Strategy Used by the General Splicing Repressor to Modulate Exon Inclusion or Skipping. *Mol. Cell* **2009**, *36* (6), 996–1006. <https://doi.org/10.1016/j.molcel.2009.12.003>.
- (103) Änkö, M.-L.; Müller-McNicoll, M.; Brandl, H.; Curk, T.; Gorup, C.; Henry, I.; Ule, J.; Neugebauer, K. M. The RNA-Binding Landscapes of Two SR Proteins Reveal Unique Functions and Binding to Diverse RNA Classes. *Genome Biol.* **2012**, *13* (3), R17. <https://doi.org/10.1186/gb-2012-13-3-r17>.
- (104) Van Nostrand, E. L.; Freese, P.; Pratt, G. A.; Wang, X.; Wei, X.; Xiao, R.; Blue, S. M.; Chen, J.-Y.; Cody, N. A. L.; Dominguez, D.; Olson, S.; Sundararaman, B.; Zhan, L.; Bazile, C.; Bouvrette, L. P. B.; Bergalet, J.; Duff, M. O.; Garcia, K. E.; Gelboin-Burkhart, C.; Hochman, M.; Lambert, N. J.; Li, H.; McGurk, M. P.; Nguyen, T. B.; Palden, T.; Rabano, I.; Sathe, S.; Stanton, R.; Su, A.; Wang, R.; Yee, B. A.; Zhou, B.; Louie, A. L.; Aigner, S.; Fu, X.-D.; Lécuyer, E.; Burge, C. B.; Graveley, B. R.; Yeo, G. W. A Large-Scale Binding and Functional Map of Human RNA-Binding Proteins. *Nature* **2020**, *583* (7818), 711–719. <https://doi.org/10.1038/s41586-020-2077-3>.
- (105) Glanville, J.; D'Angelo, S.; Khan, T. A.; Reddy, S. T.; Naranjo, L.; Ferrara, F.; Bradbury, A. R. M. Deep Sequencing in Library Selection Projects: What Insight Does It Bring? *Curr. Opin. Struct. Biol.* **2015**, *33*, 146–160. <https://doi.org/10.1016/j.sbi.2015.09.001>.
- (106) Julio, A. R.; Backus, K. M. New Approaches to Target RNA Binding Proteins. *Curr. Opin. Chem. Biol.* **2021**, *62*, 13–23. <https://doi.org/10.1016/j.cbpa.2020.12.006>.

7. Literature

- (107) Sunkari, Y. K.; Siripuram, V. K.; Nguyen, T.-L.; Flajolet, M. High-Power Screening (HPS) Empowered by DNA-Encoded Libraries. *Trends Pharmacol. Sci.* **2022**, *43* (1), 4–15. <https://doi.org/10.1016/j.tips.2021.10.008>.
- (108) Castillo, F.; Tavassoli, A. Genetic Selections with SICLOPPS Libraries: Toward the Identification of Novel Protein–Protein Interaction Inhibitors and Chemical Tools. In *Cyclic Peptide Design*; Goetz, G., Ed.; Methods in Molecular Biology; Springer New York: New York, NY, 2019; Vol. 2001, pp 317–328. https://doi.org/10.1007/978-1-4939-9504-2_15.
- (109) Kille, S.; Acevedo-Rocha, C. G.; Parra, L. P.; Zhang, Z.-G.; Opperman, D. J.; Reetz, M. T.; Acevedo, J. P. Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synth. Biol.* **2013**, *2* (2), 83–92. <https://doi.org/10.1021/sb300037w>.
- (110) Patrick, W. M.; Firth, A. E.; Blackburn, J. M. User-friendly Algorithms for Estimating Completeness and Diversity in Randomized Protein-encoding Libraries Wayne M. Patrick and Andrew E. Firth Contributed Equally to This Work. *Protein Eng. Des. Sel.* **2003**, *16* (6), 451–457. <https://doi.org/10.1093/protein/gzg057>.
- (111) Cunningham, B. C.; Wells, J. A. High-Resolution Epitope Mapping of HGH-Receptor Interactions by Alanine-Scanning Mutagenesis. *Science* **1989**, *244* (4908), 1081–1085. <https://doi.org/10.1126/science.2471267>.
- (112) Morrison, K. L.; Weiss, G. A. Combinatorial Alanine-Scanning. *Curr. Opin. Chem. Biol.* **2001**, *5* (3), 302–307. [https://doi.org/10.1016/S1367-5931\(00\)00206-4](https://doi.org/10.1016/S1367-5931(00)00206-4).
- (113) Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85* (14), 2149–2154. <https://doi.org/10.1021/ja00897a025>.
- (114) Stawikowski, M.; Fields, G. B. Introduction to Peptide Synthesis. *Curr. Protoc. Protein Sci. Editor. Board John E Coligan Al* **2002**, CHAPTER, Unit-18.1. <https://doi.org/10.1002/0471140864.ps1801s26>.
- (115) Nikolovska-Coleska, Z.; Wang, R.; Fang, X.; Pan, H.; Tomita, Y.; Li, P.; Roller, P. P.; Krajewski, K.; Saito, N. G.; Stuckey, J. A.; Wang, S. Development and Optimization of a Binding Assay for the XIAP BIR3 Domain Using Fluorescence Polarization. *Anal. Biochem.* **2004**, *332* (2), 261–273. <https://doi.org/10.1016/j.ab.2004.05.055>.
- (116) Huang, X. Fluorescence Polarization Competition Assay: The Range of Resolvable Inhibitor Potency Is Limited by the Affinity of the Fluorescent Ligand. *J. Biomol. Screen.* **2003**, *8* (1), 34–38. <https://doi.org/10.1177/1087057102239666>.
- (117) Lim, F.; Downey, T. P.; Peabody, D. S. Translational Repression and Specific RNA Binding by the Coat Protein of the Pseudomonas Phage PP7. *J. Biol. Chem.* **2001**, *276* (25), 22507–22513. <https://doi.org/10.1074/jbc.M102411200>.
- (118) Crawford, D. W.; Blakeley, B. D.; Chen, P.-H.; Sherpa, C.; Grice, S. F. J. L.; Laird-Offringa, I. A.; McNaughton, B. R. An Evolved RNA Recognition Motif That Suppresses HIV-1 Tat/TAR-Dependent Transcription. *ACS Chem. Biol.* **2016**, *11* (8), 2206. <https://doi.org/10.1021/acschembio.6b00145>.
- (119) Nikolakaki, E.; Drosou, V.; Sanidas, I.; Peidis, P.; Papamarcaki, T.; Iakoucheva, L. M.; Giannakouros, T. RNA Association or Phosphorylation of the RS Domain Prevents Aggregation of RS Domain-Containing Proteins. *Biochim. Biophys. Acta* **2008**, *1780* (2), 214–225. <https://doi.org/10.1016/j.bbagen.2007.10.014>.
- (120) Linding, R.; Jensen, L. J.; Diella, F.; Bork, P.; Gibson, T. J.; Russell, R. B. Protein Disorder Prediction: Implications for Structural Proteomics. *Structure* **2003**, *11* (11), 1453–1459. <https://doi.org/10.1016/j.str.2003.10.002>.
- (121) Hertel, K. J.; Maniatis, T. The Function of Multisite Splicing Enhancers. *Mol. Cell* **1998**, *1* (3), 449–455. [https://doi.org/10.1016/s1097-2765\(00\)80045-3](https://doi.org/10.1016/s1097-2765(00)80045-3).

7. Literature

- (122) Wang, J.; Smith, P. J.; Krainer, A. R.; Zhang, M. Q. Distribution of SR Protein Exonic Splicing Enhancer Motifs in Human Protein-Coding Genes. *Nucleic Acids Res.* **2005**, *33* (16), 5053–5062. <https://doi.org/10.1093/nar/gki810>.
- (123) Jobbins, A. M.; Reichenbach, L. F.; Lucas, C. M.; Hudson, A. J.; Burley, G. A.; Eperon, I. C. The Mechanisms of a Mammalian Splicing Enhancer. *Nucleic Acids Res.* **2018**, *46* (5), 2145–2158. <https://doi.org/10.1093/nar/gky056>.
- (124) Englesberg, E.; Anderson, R. L.; Weinberg, R.; Lee, N.; Hoffee, P.; Huttenhauer, G.; Boyer, H. L-Arabinose-Sensitive, L-Ribulose 5-Phosphate 4-Epimerase-Deficient Mutants of Escherichia Coli. *J. Bacteriol.* **1962**, *84* (1), 137–146. <https://doi.org/10.1128/jb.84.1.137-146.1962>.
- (125) *TOP10F' Electrocomp™ Kit*. <https://www.thermofisher.com/order/catalog/product/de/de/C66511> (accessed 2023-06-13).
- (126) Kessler, D. P.; Englesberg, E. Arabinose-Leucine Deletion Mutants of Escherichia Coli B/r. *J. Bacteriol.* **1969**, *98* (3), 1159–1169.
- (127) Nasim, F.-U. H.; Hutchison, S.; Cordeau, M.; Chabot, B. High-Affinity HnRNP A1 Binding Sites and Duplex-Forming Inverted Repeats Have Similar Effects on 5' Splice Site Selection in Support of a Common Looping out and Repression Mechanism. *RNA* **2002**, *8* (8), 1078–1089. <https://doi.org/10.1017/S1355838202024056>.
- (128) Fitzpatrick, T.; Huang, S. 3'-UTR-Located Inverted Alu Repeats Facilitate mRNA Translational Repression and Stress Granule Accumulation. *Nucleus* **2012**, *3* (4), 359–369. <https://doi.org/10.4161/nucl.20827>.
- (129) Martinez-Contreras, R.; Fiset, J.-F.; Nasim, F. H.; Madden, R.; Cordeau, M.; Chabot, B. Intronic Binding Sites for HnRNP A/B and HnRNP F/H Proteins Stimulate Pre-mRNA Splicing. *PLoS Biol.* **2006**, *4* (2), e21. <https://doi.org/10.1371/journal.pbio.0040021>.
- (130) Nguyen, E. D.; Balas, M. M.; Griffin, A. M.; Roberts, J. T.; Johnson, A. M. Global Profiling of HnRNP A2/B1-RNA Binding on Chromatin Highlights LncRNA Interactions. *RNA Biol.* **2018**, *15* (7), 901–913. <https://doi.org/10.1080/15476286.2018.1474072>.
- (131) Han, S. P.; Tang, Y. H.; Smith, R. Functional Diversity of the HnRNPs: Past, Present and Perspectives. *Biochem. J.* **2010**, *430* (3), 379–392. <https://doi.org/10.1042/BJ20100396>.
- (132) Schroeder, S. J. Advances in RNA Structure Prediction from Sequence: New Tools for Generating Hypotheses about Viral RNA Structure-Function Relationships. *J. Virol.* **2009**, *83* (13), 6326–6334. <https://doi.org/10.1128/JVI.00251-09>.
- (133) Liu, X.; Xu, Y. HnRNPA1 Specifically Recognizes the Base of Nucleotide at the Loop of RNA G-Quadruplex. *Molecules* **2018**, *23* (1), 237. <https://doi.org/10.3390/molecules23010237>.
- (134) Cartegni, L.; Wang, J.; Zhu, Z.; Zhang, M. Q.; Krainer, A. R. ESEfinder: A Web Resource to Identify Exonic Splicing Enhancers. *Nucleic Acids Res.* **2003**, *31* (13), 3568–3571. <https://doi.org/10.1093/nar/gkg616>.
- (135) *Specifications*. Sony Biotechnology. <https://www.sonybiotechnology.com/us/instruments/sh800s-cell-sorter/specifications/> (accessed 2023-06-19).
- (136) Iwai, H.; Züger, S.; Jin, J.; Tam, P.-H. Highly Efficient Protein Trans-Splicing by a Naturally Split DnaE Intein from Nostoc Punctiforme. *FEBS Lett.* **2006**, *580* (7), 1853–1858. <https://doi.org/10.1016/j.febslet.2006.02.045>.
- (137) Züger, S.; Iwai, H. Intein-Based Biosynthetic Incorporation of Unlabeled Protein Tags into Isotopically Labeled Proteins for NMR Studies. *Nat. Biotechnol.* **2005**, *23* (6), 736–740. <https://doi.org/10.1038/nbt1097>.

7. Literature

- (138) C. Hayes, H.; P. Luk, L. Y.; Tsai, Y.-H. Approaches for Peptide and Protein Cyclisation. *Org. Biomol. Chem.* **2021**, *19* (18), 3983–4001. <https://doi.org/10.1039/D1OB00411E>.
- (139) Kolesnichenko, I. V.; Goloverda, G. Z.; Kolesnichenko, V. L. A Versatile Method of Ambient-Temperature Solvent Removal. *Org. Process Res. Dev.* **2020**, *24* (1), 25–31. <https://doi.org/10.1021/acs.oprd.9b00368>.
- (140) Poveda, J. A.; Prieto, M.; Encinar, J. A.; González-Ros, J. M.; Mateo, C. R. Intrinsic Tyrosine Fluorescence as a Tool To Study the Interaction of the Shaker B “Ball” Peptide with Anionic Membranes. *Biochemistry* **2003**, *42* (23), 7124–7132. <https://doi.org/10.1021/bi027183h>.
- (141) Protein Fluorescence. In *Principles of Fluorescence Spectroscopy*; Lakowicz, J. R., Ed.; Springer US: Boston, MA, 2006; pp 529–575. https://doi.org/10.1007/978-0-387-46312-4_16.
- (142) Tavassoli, A.; Benkovic, S. J. Genetically Selected Cyclic-Peptide Inhibitors of AICAR Transformylase Homodimerization. *Angew. Chem. Int. Ed.* **2005**, *44* (18), 2760–2763. <https://doi.org/10.1002/anie.200500417>.
- (143) Spurr, I. B.; Birts, C. N.; Cuda, F.; Benkovic, S. J.; Blaydes, J. P.; Tavassoli, A. Targeting Tumour Proliferation with a Small-Molecule Inhibitor of AICAR Transformylase Homodimerization. *ChemBioChem* **2012**, *13* (11), 1628–1634. <https://doi.org/10.1002/cbic.201200279>.

8. Appendix

8.1. Tables and Figures

8.1.1. Table of figures

Figure 1: Splicing mechanism performed by the spliceosome. The U1 snRNP binds to the 5'SS, while U2 binds to the BP to form the A complex. U4/U6.U5 (tri-snRNP) joins the A complex, and together they form the B complex. The spliceosome is activated upon the removal of U1 and U4. Once the first splicing reaction is completed, formation of the C complex occurs. After the second splicing reaction, the spliced mRNA is released from the P complex. Adapted from Gehring and Roignant, 2014. ⁴..... 17

Figure 2: Alternative splicing mechanisms. Blue boxes represent exons, while straight lines indicate introns. Angled lines show the junction path of pre-mRNA to form the spliced mRNA. Adapted from Ren et al., 2019. ^{2,9}..... 18

Figure 3: Role of cis and trans-acting factors SR and hnRNP in alternative splicing of pre-mRNAs. Binding of SR proteins to ESE and ISE (blue boxes) promotes exon inclusion (light grey boxes), while hnRNP binding to ESS and ISS elements (purple boxes) promotes exon skipping. ^{8,10}..... 19

Figure 4: Role of SRSF1 in cellular transformation. Overexpression of SRSF1 generates proliferative and anti-apoptotic isoforms that are unable to interact with pro-apoptotic factors such as MYC. Adapted from Anczuków et al., 2012. ¹⁶..... 21

Figure 5: Splicing factors SRSF1 (blue tones) and hnRNP A2B1 (purple tones). Schematic representation of the domains (top, adapted from Anczuków *et al.*, 2012 and Wu *et al.*, 2018 ^{16,27}) and crystal or NMR structures of RRM1 and RRM2 (bottom). PDB: SRSF1 RRM1: 1X4A, RRM2: 2M8D; hnRNP A2B1 RRM1 and 2: 5HO4..... 22

Figure 6: SICLOPPS mechanism (schematic). The two intein domains (I_N and I_C) of the SICLOPPS intein fusion protein fold and generate an active intein. An N-to-S acyl shift at the I_N -junction forms a thioester. In a transesterification reaction with a side chain nucleophile (here: cysteine) at the I_C -junction, a lariat intermediate is produced. An asparagine side chain (I_C -junction) promotes lactone formation, which then generates the thermodynamically favoured lactam product via an S-to-N acyl

8. Appendix

shift (<i>in vivo</i>). The hexameric cyclic peptide contains five randomized amino acids (each indicated by X). Adapted from Tavassoli and Benkovic 2007 ^{58,59} , created with Biorender.com.	26
Figure 7: Concept of the SICLOPPS screening using TRAP. The SICLOPPS plasmids are transformed into <i>E. coli</i> cells carrying an assays system that allows intracellular screening with FACS. Isolated cells can be sequenced to identify the macrocyclic peptide that caused a phenotypic change. Adapted from Tavassoli, 2017. ⁵¹	28
Figure 8: Schematic representation of the analysis steps of the RBP consensus sequence screening dataset using the Galaxy web platform. ⁶⁶	46
Figure 9: Schematic representation of the analysis steps of the SICLOPPS screening dataset using the Galaxy web platform. ⁶⁶	50
Figure 10: Principle of TRAP. Left: Reporter RNA in the absence of an RBP. Right: Binding of an RBP to its RNA target results in translational repression. Translation is initiated by the recruitment of ribosomes to the ribosomal binding site (RBS). As an improvement of the procedure, Katz <i>et al.</i> added a fluorescent tag to the RBP to monitor its expression during the measurement. ^{63,82}	56
Figure 11: Schematic representation of the two assay plasmids used in TRAP. Left: The “RNA plasmid” contains the genetic information for the RNA insert, the S/D sequence, tagBFP which are under control of the lacUV5 promoter and the kanamycin resistance (KanR). Right: The “protein plasmid” contains the genetic information for the sfGFP-RBP fusion protein which is controlled by the araBAD promoter and the chloramphenicol resistance gene (CamR).	56
Figure 12: TRAP assay plasmid constructs. The RNA plasmid contains the RNA insert (here: AGAAGAAC), the S/D sequence and the reporter gene <i>tagBFP</i> . The protein plasmid carries the gene for the sfGFP-SRSF1 fusion protein.....	57
Figure 13: TRAP assay with the reporter constructs S0 and S1. Left: Repression curves for sfGFP-SRSF1 (blue and black), and sfGFP-PTBP1 (green and grey, indicated by b). Right: Repression ratios for sfGFP-SRSF1 (blue and black) and sfGFP-PTBP1 (green and grey, indicated by b). Data are mean values (n=2, N=2).	58

8. Appendix

Figure 14: TRAP assay data of reporter constructs S6-4 – S6-10. Left: Repression curves with sfGFP-SRSF1. Right: Repression ratios 10 with sfGFP-SRSF1 (blue tones). Data are mean values (n=2, N=2).	60
Figure 15: RT-qPCR analysis of <i>E. coli</i> cells cotransformed with the sfGFP-SRSF1/S6 plasmid pair. Gene expression levels were normalized to <i>gapA</i> . Data are mean values (n=2, N=2).	62
Figure 16: TRAP assay plasmid constructs. The RNA plasmid contains the RNA insert (here: AAGGACUAGC), the S/D sequence and the reporter gene <i>tagBFP</i> . The protein plasmid carries the gene for the sfGFP-hnRNP A2B1 fusion protein.	63
Figure 17: TRAP assay data of reporter constructs H1-H3. Left: Repression curves for sfGFP-A2B1 (pink tones) and sfGFP-PTBP1 (green tones, indicated by b). Right: Repression ratios for sfGFP-A2B1 (pink tones) and sfGFP-PTBP1 (green tones, indicated by b). Data are mean values (n=2, N=2).	63
Figure 18: Secondary structure analysis for reporters S0/H0 using RNAfold.	66
Figure 19: Secondary structure analysis for reporter H12 using RNAfold.	66
Figure 20: Correlation between basal tagBFP production rate (1/min) of all reporters for SRSF1 (left, blue) and hnRNP A2B1 (right, purple) vs the distance of nucleotides between the first occurring secondary structure and the S/D sequence.	67
Figure 21: Fluorescence polarisation binding curves of FAM-S1, FAM-S1ext, FAM-S2, FAM-S3 and a Polypyrimidine RNA with SRSF1. Polypyrimidine RNA was used as a negative control. Data are mean values (n=3, N=2).	68
Figure 22: Fluorescence polarisation binding curves of FAM-H1, FAM-H2, FAM-H3, and Poly C RNA with MBP-hnRNP A2B1. Poly C RNA was used as a negative control. Data are mean values (n=2, N=2).	68
Figure 23: Fluorescence polarisation binding curves of FAM-H8, FAM-H12 and Poly C RNA with MBP-hnRNP A2B1. Poly C RNA was used as a negative control. Data are mean values (for FAM-H8: n=2, N=2; for FAM-H12: n=1, N=2).	70
Figure 24: Repression ratios measured with the TRAP assay correlate with binding affinities generated by FP. Left: Correlation graph for SRSF1; Right: Correlation graph for hnRNP A2B1.	71

8. Appendix

Figure 25: Flow cytometry results for sfGFP-SRSF1 (blue) or sfGFP-PTBP1 (orange) in combination with reporter S4 (A), reporter S5 (B) or reporter S6 (C). As well as sfGFP-A2B1 (blue) or sfGFP-PTBP1 (orange) in combination with reporter H4 (D), reporter H6 (E), or reporter H8 (F). Histograms were produced by analysing all events $>1 \times 10^5$ in the sfGFP channel.	74
Figure 26: Concept of the RBP consensus sequence screening. A plasmid library with a randomized 10mer insert is transformed into <i>E. coli</i> cells carrying the sfGFP-RBP fusion plasmid and screened using flow cytometry. Cells with the repressed phenotype are sorted and sequenced. Illustration was created with Biorender.com.	75
Figure 27: SRSF1 consensus sequence screening. Cells carrying the sfGFP-SRSF1 plasmid were transformed with the 10mer plasmid library and were treated with 1 mM IPTG and 0.125 % arabinose. Left: Whole population of the cells in the analysis. Cells that produced high sfGFP levels and low levels of tagBFP were gated for sorting (black box). Right: Sorted cells with gated fluorescence properties.	76
Figure 28: Autorepressor sequence selection using FACS. Cells carrying the sfGFP-SRSF1 plasmid were transformed with the 10mer plasmid library and were treated with 1 mM IPTG but no arabinose. Left: Whole population of the cells in the analysis. Cells that produced low sfGFP and tagBFP levels were gated for sorting (black box). Right: Sorted cells with gated fluorescence properties.	77
Figure 29: Repression ratios of the hit sequences 1-10 and the autorepressor corrected samples 1_AC-10_AC for the SRSF1 consensus sequence screening. Data are mean values (n=2, N=2).	79
Figure 30: TagBFP production rate [1/min] of the hit sequences 1-10 and the autorepressor corrected samples 1_AC-10_AC upon arabinose induction at 0 or 1 % for the SRSF1 consensus sequence screening. Data are mean values (n=2, N=2).	80
Figure 31: SfGFP expression of the hit sequences 1-10 and the autorepressor corrected samples 1_AC-10_AC upon arabinose induction at 0 or 1 % for the SRSF1 consensus sequence screening. Data are mean values (n=2, N=2).	80
Figure 32: HnRNP A2B1 consensus sequence screening. Cells carrying the sfGFP-hnRNP A2B1 plasmid were transformed with the 10mer plasmid library and were treated with 1 mM IPTG and 0.125 % arabinose. Left: Whole population of the cells in the analysis. Cells that produced high sfGFP levels	

8. Appendix

and low levels of tagBFP were gated for sorting (black box). Right: Sorted cells with gated fluorescence properties.....	81
Figure 33: Repression ratios of the hit sequences 1-10 and the autorepressor corrected samples 1_AC-10_AC for the hnRNP A2B1 consensus sequence screening. Data are mean values (n=2, N=2).	83
Figure 34: Concept of the autorepressor presorted RBP consensus sequence screening. <i>E. coli</i> cells carrying the sfGFP-RBP fusion plasmid are transformed with the 10mer library, induced with IPTG but without arabinose and presorted to remove autorepressor sequences. Cells that do not show a repressed phenotype are collected, while cells that show a repressed phenotype are discarded. In a second sorting round, cells are induced with IPTG and arabinose, and those with the repressed phenotype are sorted and sequenced. Illustration was created with Biorender.com.	84
Figure 35: Presorting of the SRSF1 consensus sequencing screening. Cells carrying the sfGFP-SRSF1 plasmid were transformed with the 10mer plasmid library and were treated with 1 mM IPTG but no arabinose. Left: Whole population of the cells in the analysis. Cells that produced low sfGFP levels and high tagBFP levels were gated for sorting (black box). Right: Sorted cells with gated fluorescence properties.....	85
Figure 36: Presorted SRSF1 consensus sequencing screening. Presorted cells were treated with 1 mM IPTG and 0.125 % arabinose. Left: Whole population of the cells in the analysis. Cells that produced high sfGFP levels and low tagBFP levels were gated for sorting (black box). Right: Sorted cells with gated fluorescence properties.	86
Figure 37: Histograms for the presorted SRSF1 consensus sequence screening (blue) and the SRSF1 consensus sequence screening without presorting (orange).	87
Figure 38: Repression ratios of the hit sequences 1-15 for the autorepressor presorted SRSF1 consensus sequence screening. Data are mean values (n=2, N=2).	88
Figure 39: The TRAP assay as a screening platform for SICLOPPS. Left: Reporter RNA/RBP interaction results in translational repression and low reporter production. Right: Inhibition of the RNA/RBP interaction with a peptide inhibitor, highlighted in red, allows ribosome binding and restores reporter translation, leading to high tagBFP production rates.	90

8. Appendix

Figure 40: SRSF1 SICLOPPS screening. Cells were treated with 100 ng/mL anhydrotetracycline prior to treatment with 1 mM IPTG and 0.125 % arabinose. Left: Whole population of the cells in the analysis. Cells that produced high sfGFP levels and high tagBFP levels were gated for sorting (black box). Right: Sorted cells with gated fluorescence properties.....	91
Figure 41: TagBFP production rate of cells carrying the sfGFP-SRSF1/S6 plasmid pair and a SICLOPPS plasmid (indicated by number). Cells were treated with 100 ng/mL anhydrotetracycline prior to treatment with 1 mM IPTG and 0.125 % arabinose. Data are mean values (n=2, N=2).....	93
Figure 42: TagBFP production rate of cells carrying the sfGFP-SRSF1/S6 plasmid pair and a SICLOPPS plasmid (indicated by number). Cells were treated with 100 ng/mL anhydrotetracycline prior to treatment with 1 mM IPTG but no arabinose. Data are mean values (n=2, N=2).	94
Figure 43: SfGFP expression of cells carrying the sfGFP-SRSF1/S6 plasmid pair and a SICLOPPS plasmid (indicated by number). Cells were treated with 100 ng/mL anhydrotetracycline prior to treatment with 1 mM IPTG but no arabinose. Data are mean values (n=2, N=2).	95
Figure 44: SfGFP expression of cells carrying the sfGFP-SRSF1/S6 plasmid pair and a SICLOPPS plasmid (indicated by number). Cells were treated with 100 ng/mL anhydrotetracycline prior to treatment with 1 mM IPTG and 0.125 % arabinose. Data are mean values (n=2, N=2).	95
Figure 45: Synthesis route of SICLOPPS peptides on solid phase (2CTC resin). The first amino acid (cysteine) is loaded to the resin via its side chain thiol group. Once completion of the sequence is achieved, the peptide is cyclised head-to-tail and then cleaved from the resin.	98
Figure 46: Fluorescence polarisation competition curves of peptides 6, 44, 50, unlabelled S3 RNA and a DMSO control (5 %) with 80 nM SRSF1 and 1 nM FAM-S3. Unlabelled S3 RNA and DMSO were used as controls. Data was measured after 20 minutes incubation and is presented as mean values (N=2).	98
Figure 47: Fluorescence polarisation competition curves of peptide 4 and unlabelled S3 RNA with 80 nM SRSF1 and 1 nM FAM-S3 and Peptide 44 with buffer only. Unlabelled S3 RNS was used as control. Data was measured after 45 minutes incubation and is presented as mean values (N=2).	99

8. Appendix

Figure 48: Fluorescence polarisation competition curves of peptide 4 and unlabelled S3 RNA with 80 nM SRSF1 and 1 nM Cy5-S3. Unlabelled S3 RNA was used as a control. Data was measured after 60 minutes incubation and is presented as mean values (N=2).	100
Figure 49: SRSF1 SICLPPS screening with optimized conditions. Cells were treated with 100 ng/mL anhydrotetracycline prior to treatment with 1 mM IPTG and 0.125 % arabinose. Left: Whole population of the cells in the analysis. Cells that produced high sfGFP levels and high tagBFP levels were gated for sorting (black box). Right: Sorted cells with gated fluorescence properties.	101
Figure 50: Enriched SRSF1 SICLPPS screening (round 1). Cells were treated with 100 ng/mL anhydrotetracycline prior to treatment with 1 mM IPTG and 0.125 % arabinose. Left: Whole population of the cells in the analysis. Cells that produced high sfGFP levels and high tagBFP levels were gated for sorting (black box). Right: Sorted cells with gated fluorescence properties.	103
Figure 51: Enriched SRSF1 SICLPPS screening (round 2). Cells were treated with 100 ng/mL anhydrotetracycline prior to treatment with 1 mM IPTG and 0.125 % arabinose. Left: Whole population of the cells in the analysis. Cells that produced high sfGFP levels and high tagBFP levels were gated for sorting (black box). Right: Sorted cells with gated fluorescence properties.	104
Figure 52: HnRNP A2B1 SICLPPS screening with optimized conditions. Cells were treated with 100 ng/mL anhydrotetracycline prior to treatment with 1 mM IPTG and 0.125 % arabinose. Left: Whole population of the cells in the analysis. Cells that produced high sfGFP levels and high tagBFP levels were gated for sorting (black box). Right: Sorted cells with gated fluorescence properties.....	106
Figure 53: Fluorescence polarisation competition curves of peptides SR1-1, SR1-2, SR3, SR4, SR5-1, SR5-2 SR6-1, SR6-2 and unlabelled S3 RNA with 200 nM SRSF1 and 1 nM FAM-S3. Unlabelled S3 RNA was used as control. Data was measured after 30 minutes incubation and is presented as mean values (N=2).	109
Figure 54: Fluorescence polarisation binding curves of Cy5-S3 and Cy5-Polypyrimidine RNA with SRSF1. Polypyrimidine RNA was used as a negative control. Data was measured after 60 minutes incubation and is presented as mean values (N=2).	110

8. Appendix

Figure 55: Fluorescence polarisation competition curves of peptides SR3, SR4, SR6-1 and unlabelled S3 RNA with 400 nM SRSF1 and 1 nM Cy5-S3. Unlabelled S3 RNA was used as control. Data was measured after 60 minutes incubation and is presented as mean values (N=2).	110
Figure 56: Fluorescence polarisation competition curves of peptides SR2-1, SR2-2, SR7 and unlabelled S3 RNA with 400 nM SRSF1 and 1 nM Cy5-S3. Unlabelled S3 RNA was used as control. Data was measured after 60 minutes incubation and is presented as mean values (N=2).	111
Figure 57: Fluorescence polarisation competition curves of peptides A1, A2 and unlabelled H1 RNA with 25 nM hnRNP A2B1 and 1 nM FAM-H8. Unlabelled H1 RNA and DMSO were used as controls. Data was measured after 60 minutes incubation and is presented as mean values (N=2).	112
Figure 58: Fluorescence polarisation competition curves of peptides A3, A4 and unlabelled H1 RNA with 25 nM hnRNP A2B1 and 1 nM FAM-H8. Unlabelled H1 RNA and DMSO were used as controls. Peptide A4 was dissolved in 1 % DMSO. Data was measured after 60 minutes incubation and is presented as mean values (N=2).	113
Figure 59: Fluorescence polarisation competition curves of peptides A5-1, A5-2, A6 and unlabelled H1 RNA with 25 nM hnRNP A2B1 and 1 nM FAM-H8. Unlabelled H1 RNA and DMSO were used as controls. Data was measured after 60 minutes incubation and is presented as mean values (N=2).	113
Figure 60: Fluorescence polarisation competition curves of peptides A4, A5-1, A5-2 and unlabelled H1 RNA with 35 nM hnRNP A2B1 and 1 nM FAM-H1. Unlabelled H1 RNA and DMSO were used as controls. Data was measured after 60 minutes incubation and is presented as mean values (N=2).	115
Figure 61: TRAP assay data of reporter construct H1 with hnRNP A2B1-sfGFP (indicated by d). Left: Repression curve. Right: Repression ratio. Data are mean values (n=2, N=2).	118
Figure 62: Secondary structure analysis for reporter H17 using RNAfold.	121
Figure 63: Schematic representation of two hnRNP A2B1 proteins (RRM1 and 2, represented by rectangles in purple tones) in complex with two AGGACUAGC RNAs. The proteins bind the RNAs in an antiparallel conformation. Illustration adapted from Wu <i>et al.</i> , 2018. ²⁷	123
Figure 64: S Secondary structure analysis for reporters 2_AC and 10_AC derived from the SRSF1 RBP consensus sequence screening structures using RNAfold.	126

8. Appendix

Figure 65: Optimised autorepressor presorted screening, including several rounds of alternating sorting.....	128
Figure 66: Schematic representation of the optimised RNA plasmids for the RBP consensus sequence screening. The plasmid contains a tagBFP-AmpR fusion gene that allows life/death selection for secondary structures in the 5'UTR.	129

8.1.2. Table of Tables

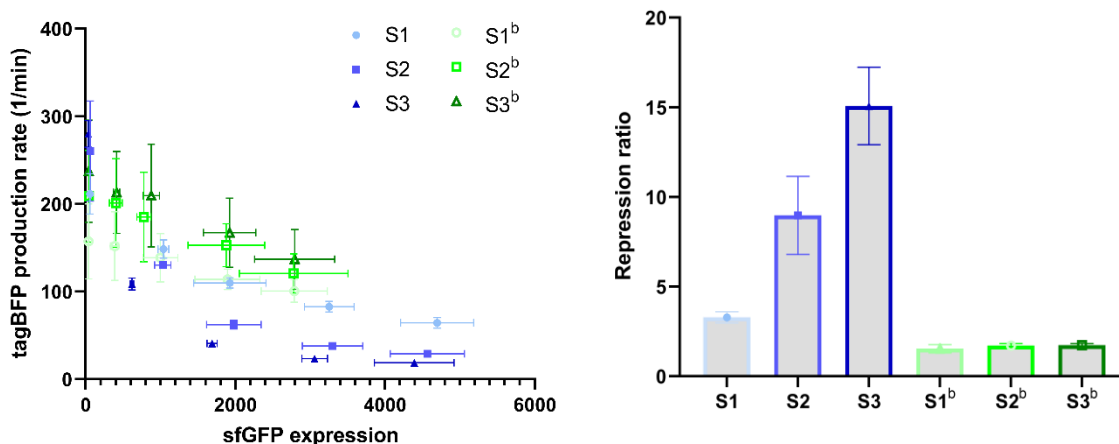
Table 1: Examples of target genes spliced by SRSF1, their splicing change and induced property. Adapted from Das and Krainer, 2014 ¹⁸	20
Table 2: Thermocycling conditions for Phusion PCR.	36
Table 3: Thermocycling conditions for the CPEC reaction.....	38
Table 4: Reporter constructs used in TRAP and the resulting repression ratios with SRSF1.	59
Table 5: Reporter constructs used in TRAP and the resulting repression ratios with hnRNP A2B1. ..	63
Table 6: Affinities of SRSF1 and hnRNP A2B1 for representative RNA sequences measured by fluorescence polarisation.....	70
Table 7: Hit sequence constructs of the SRSF1 consensus sequence screening used in TRAP and the resulting repression ratios. The negative control measurement was performed with the sfGFP-PTBP1 fusion protein.	78
Table 8: Hit sequence constructs of the hnRNP A2B1 consensus sequence screening used in TRAP and the resulting repression ratios. The negative control measurement was performed with the sfGFP-PTBP1 fusion protein.....	82
Table 9: Hit sequence constructs of the Autorepressor presorted SRSF1 consensus sequence screening used in TRAP and the resulting repression ratios. The negative control measurement was performed with the sfGFP-PTBP1 fusion protein.	87
Table 10: Hit-peptides, their sequences and occurrence in the screen.	92
Table 11: Peptide sequences, occurrence in sequencing and tagBFP production rate upon induction. The positive control measurement was performed using the sfGFP-SRSF1 _{mut} fusion protein construct.	96

8. Appendix

Table 12: SRSF1 SICLOPPS screening hit sequences and sequence abundance. Entries with stop codon-containing sequences (indicated by X) are highlighted in red.	102
Table 13: Enriched SRSF1 SICLOPPS screening hit sequences and sequence abundance. Entries with stop codon-containing sequences (indicated by X) are highlighted in red.	104
Table 14: hnRNP A2B1 SICLOPPS screening hit sequences (hidden) and sequence abundance. Entries with stop codon-containing sequences (indicated by X) are highlighted in red.	106
Table 15: Cyclic peptides derived from the SRSF1 SICLOPPS screening with optimised conditions.	108
Table 16: Cyclic peptides derived from the enriched SRSF1 SICLOPPS screening with optimised conditions.	108
Table 17: Cyclic peptides (hidden) derived from the hnRNP A2B1 SICLOPPS screening with optimised conditions.	111
Table 18: FP derived IC ₅₀ values, Hill slopes and calculated KI values for the hit peptides tested with hnRNP A2B1 in interaction with FAM-H8 RNA.	114
Table 19: FP derived IC ₅₀ values and Hill for the hit peptides tested with hnRNP A2B1 in interaction with Cy5-H1 RNA.	115

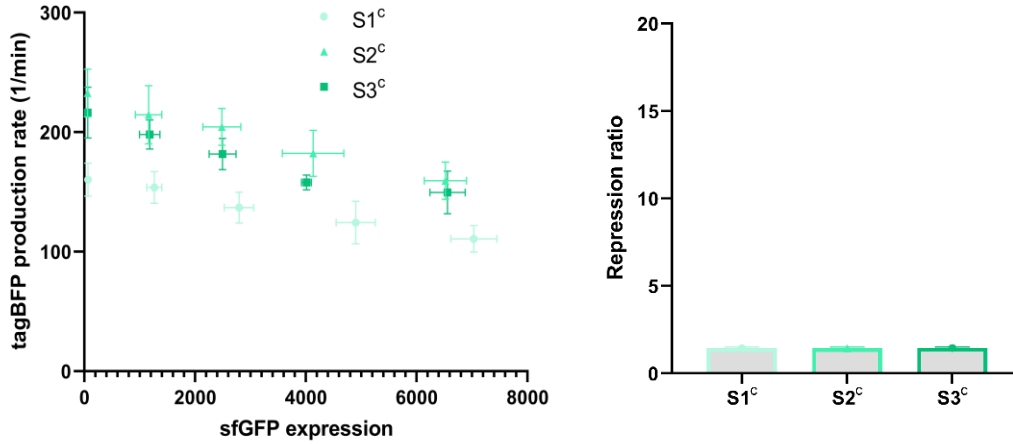
8.2. Supplementary figures

8.2.1. Supplementary figures: TRAP figures

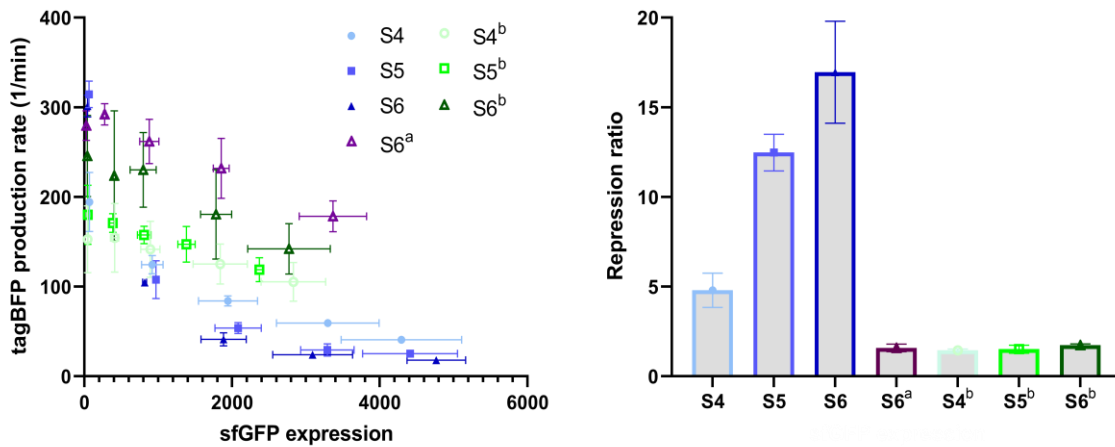


8. Appendix

Supplementary figure 1: TRAP assay data of reporter constructs S1-S3. Left: Repression curves for sfGFP-SRSF1 (blue tones) and sfGFP-PTBP1 (green tones, indicated by b). Right: Repression ratios for sfGFP-SRSF1 (blue tones) and sfGFP-PTBP1 (green tones, indicated by b). Data are mean values (n=2, N=2).

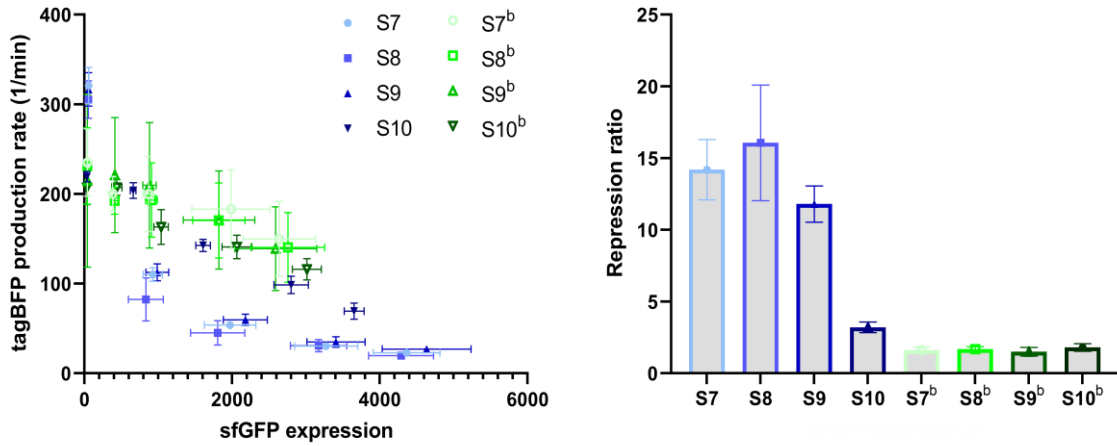


Supplementary figure 2: TRAP assay data of reporter constructs S1-S3. Left: Repression curves with sfGFP (green tones, indicated by c). Right: Repression ratios with sfGFP (green tones, indicated by c). Data are mean values (n=2, N=2).

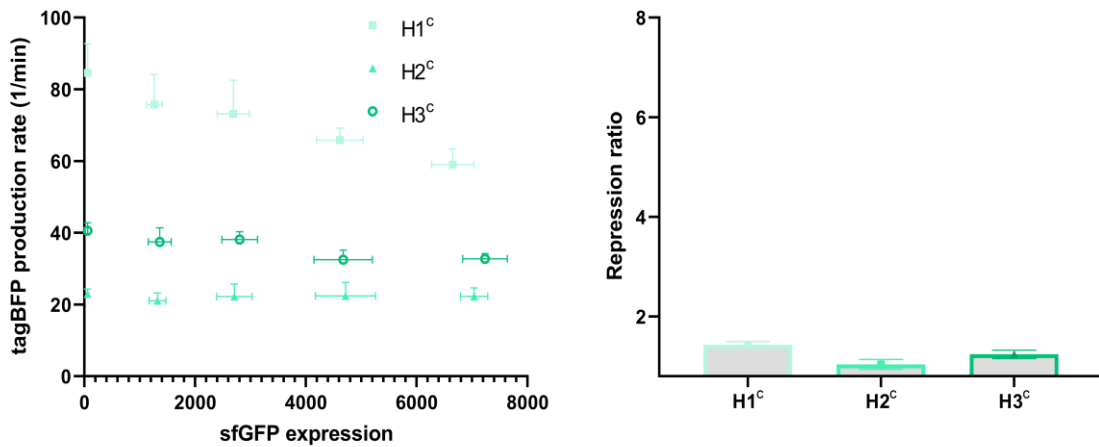


Supplementary figure 3: TRAP assay data of reporter constructs S4-S6. Left: Repression curves for sfGFP-SRSF1 (blue tones), sfGFP-SRSF1_{mut} (purple, indicated by a) and sfGFP-PTBP1 (green tones, indicated by b). Right: Repression ratios for sfGFP-SRSF1 (blue tones), sfGFP-SRSF1_{mut} (purple, indicated by a) and sfGFP-PTBP1 (green tones, indicated by b). Data are mean values (n=2, N=2).

8. Appendix

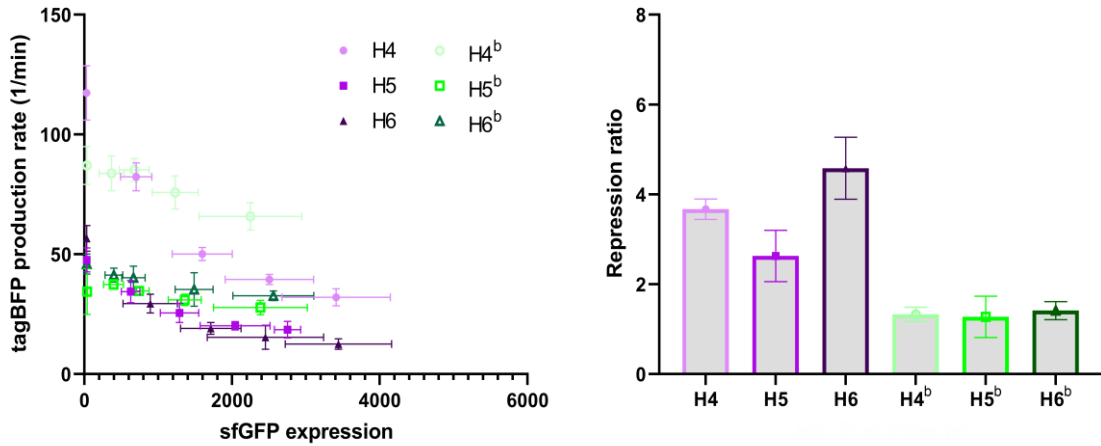


Supplementary figure 4: TRAP assay data of reporter constructs S7-S10. Left: Repression curves with sfGFP-SRSF1 (blue tones) and sfGFP-PTBP1 (green tones). Right: Repression ratios with sfGFP-SRSF1 (blue tones) and sfGFP-PTBP1 (green tones). Data are mean values (n=2, N=2).

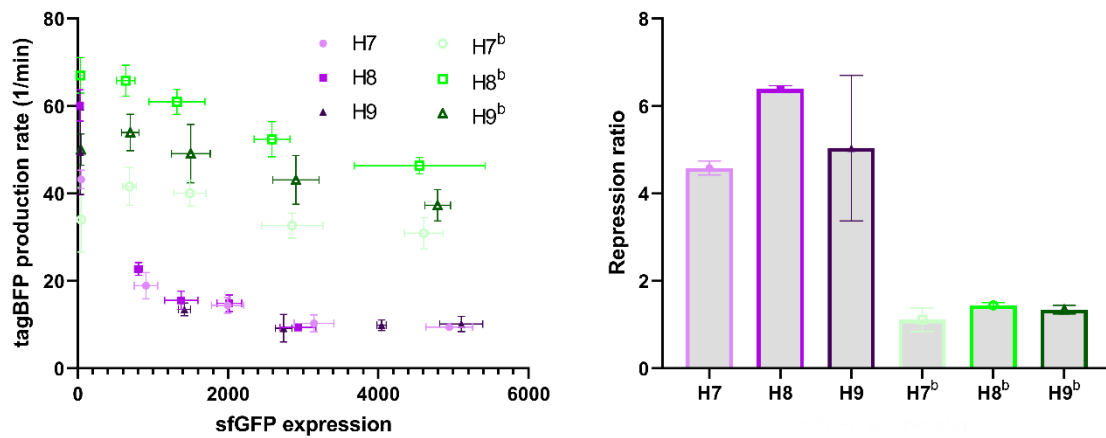


Supplementary figure 5: TRAP assay data of reporter constructs H1-H3. Left: Repression curves with sfGFP (green tones, indicated by c). Right: Repression ratios with sfGFP (green tones, indicated by c). Data are mean values (n=2, N=2).

8. Appendix

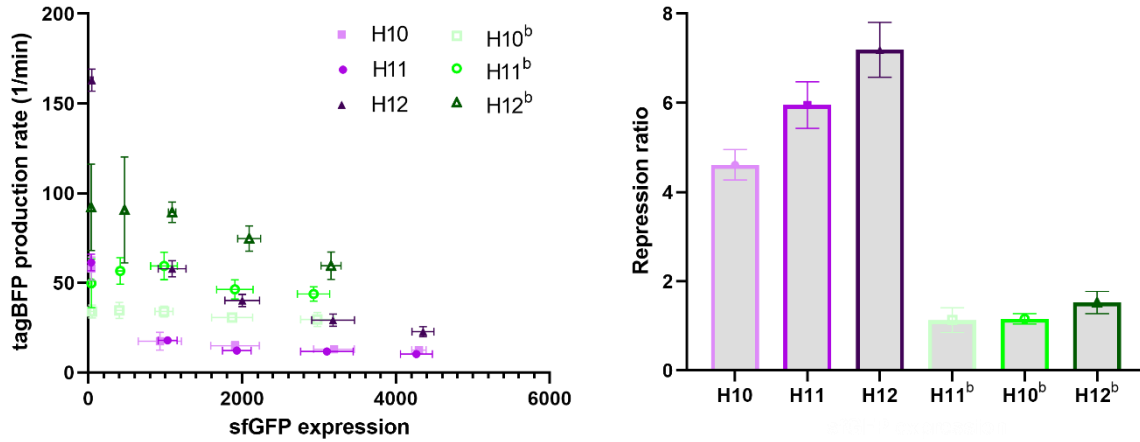


Supplementary figure 6: TRAP assay data of reporter constructs H4-H6. Left: Repression curves for sfGFP-A2B1 (pink tones) and sfGFP-PTBP1 (green tones, indicated by b). Right: Repression ratios for sfGFP-A2B1 (pink tones) and sfGFP-PTBP1 (green tones, indicated by b). Data are mean values (n=2, N=2).

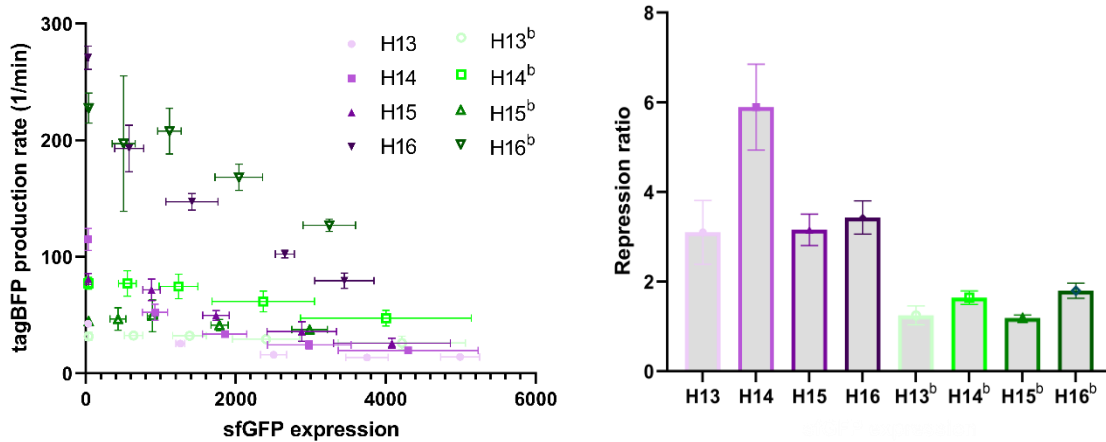


Supplementary figure 7: TRAP assay data of reporter constructs H7-H9. Left: Repression curves for sfGFP-A2B1 (pink tones) and sfGFP-PTBP1 (green tones, indicated by b). Right: Repression ratios for sfGFP-A2B1 (pink tones) and sfGFP-PTBP1 (green tones, indicated by b). Data are mean values (n=2, N=2).

8. Appendix

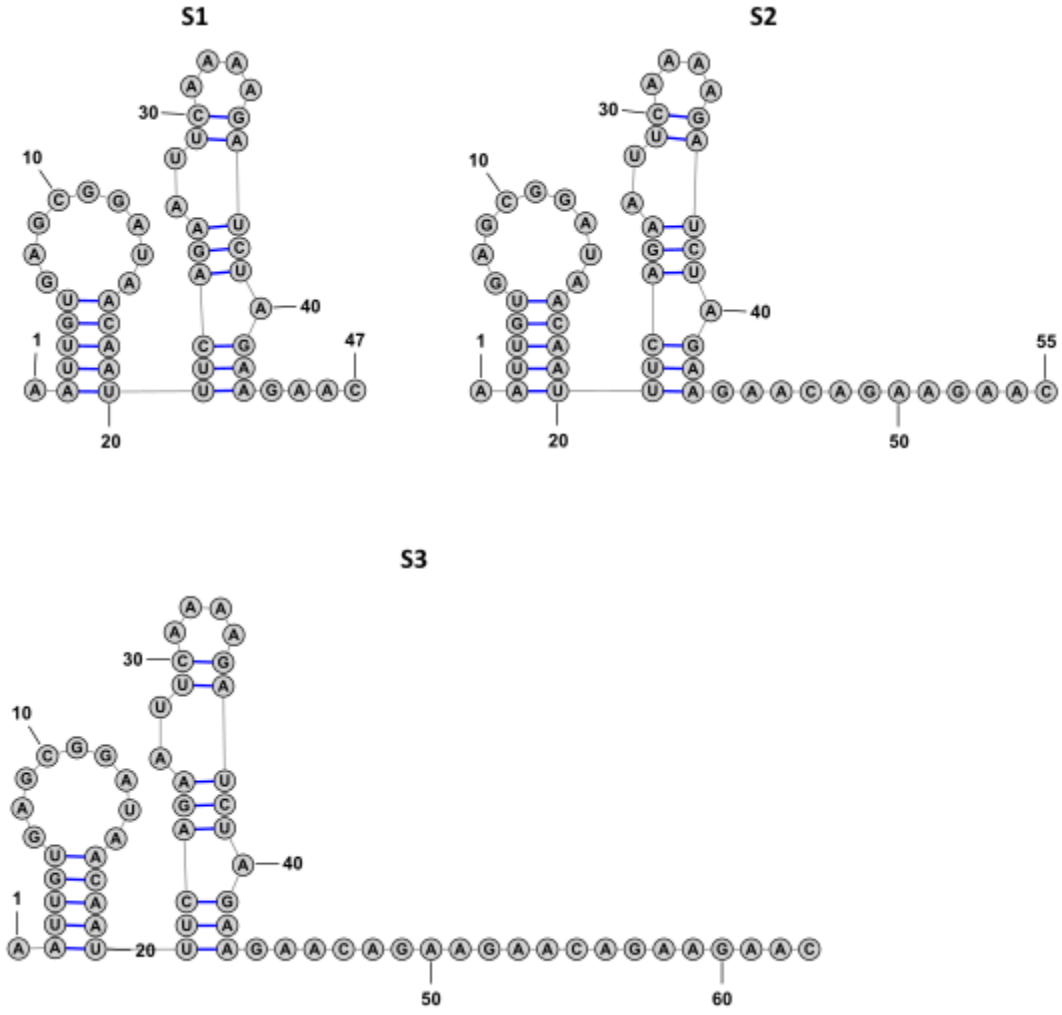


Supplementary figure 8: TRAP assay data of reporter constructs H10-H12. Left: Repression curves for sfGFP-A2B1 (pink tones) and sfGFP-PTBP1 (green tones, indicated by b). Right: Repression ratios for sfGFP-A2B1 (pink tones) and sfGFP-PTBP1 (green tones, indicated by b). Data are mean values (n=2, N=2).

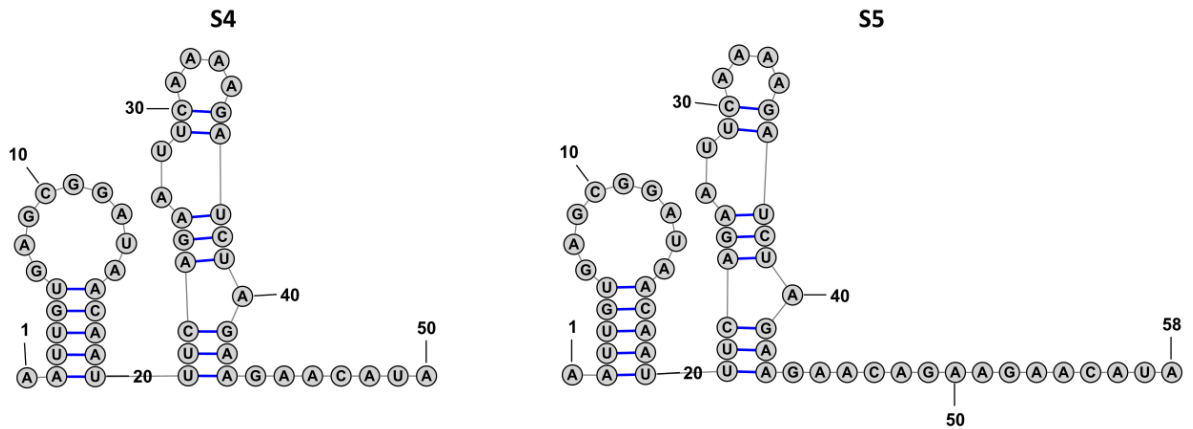


Supplementary figure 9: TRAP assay data of reporter constructs H13-H16. Left: Repression curves for sfGFP-A2B1 (pink tones) and sfGFP-PTBP1 (green tones, indicated by b). Right: Repression ratios for sfGFP-A2B1 (pink tones) and sfGFP-PTBP1 (green tones, indicated by b). Data are mean values (n=2, N=2).

8.2.2. Supplementary figures: Secondary structure analysis

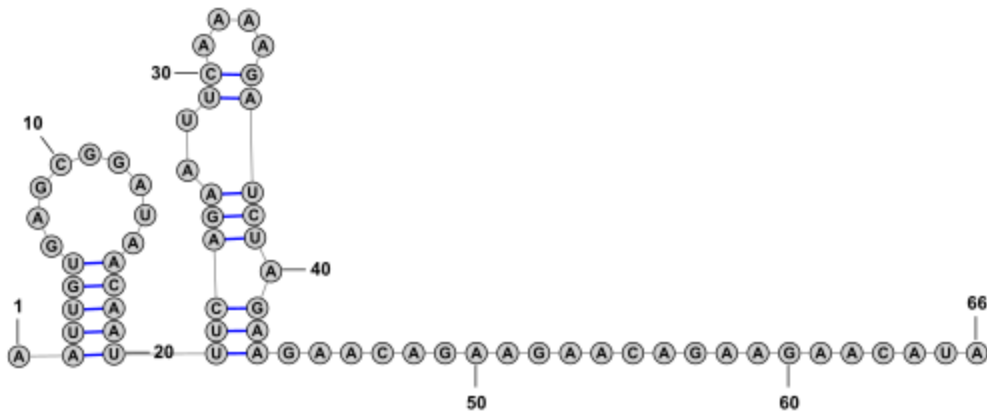


Supplementary figure 10: Secondary structure analysis for reporters S1- S3 with hairpin structures using RNAfold.



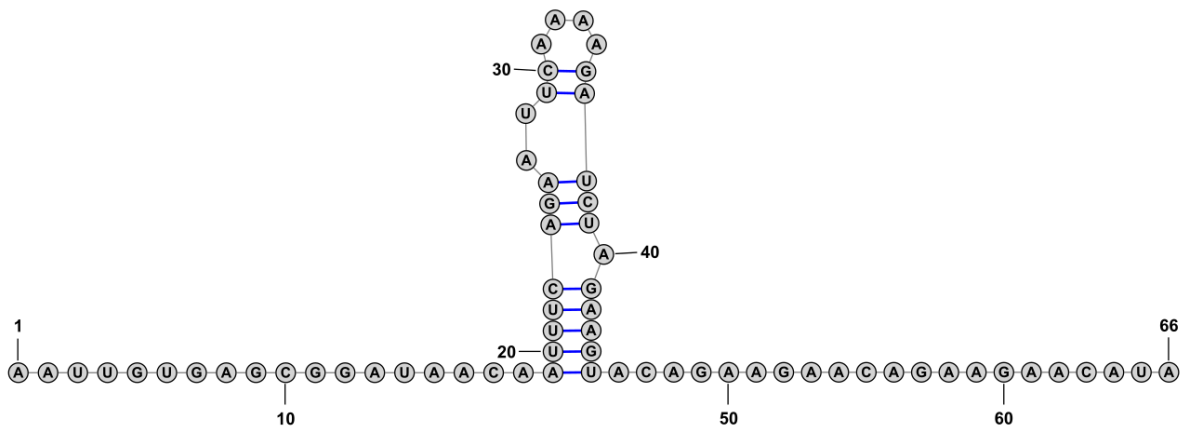
8. Appendix

S6

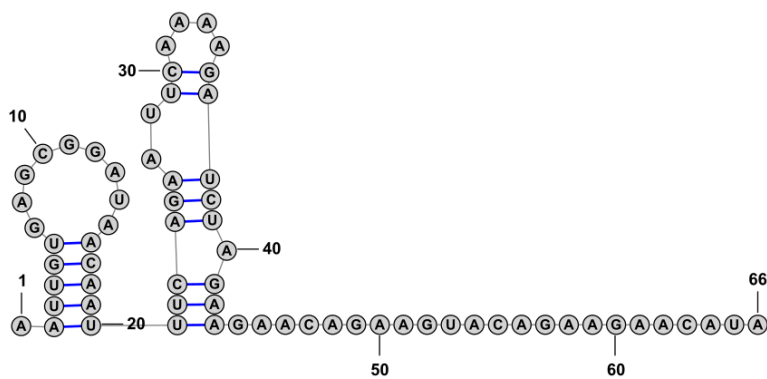


Supplementary figure 11: Secondary structure analysis for reporters S4- S6 with hairpin structures using RNAfold.

S7

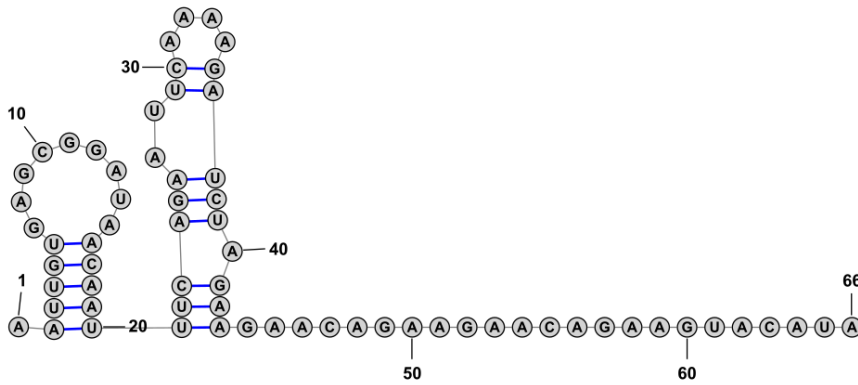


S8

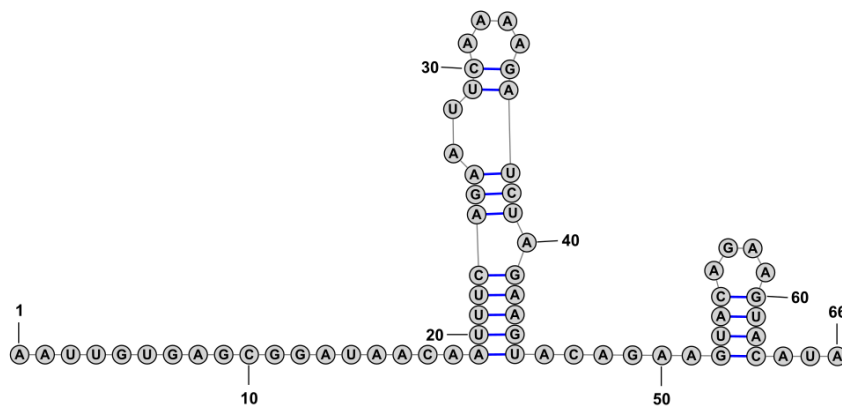


8. Appendix

S9

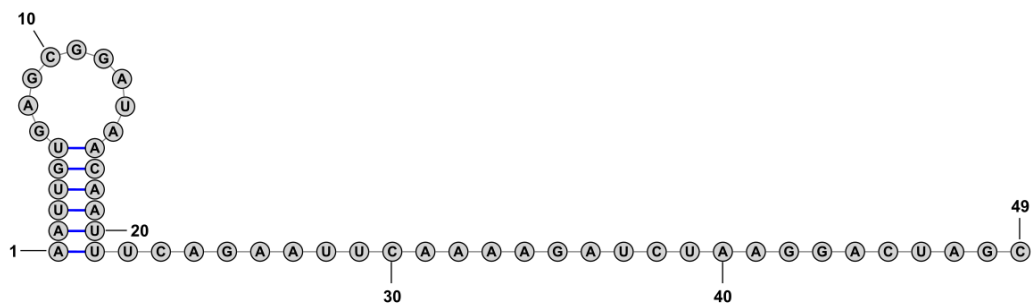


S10



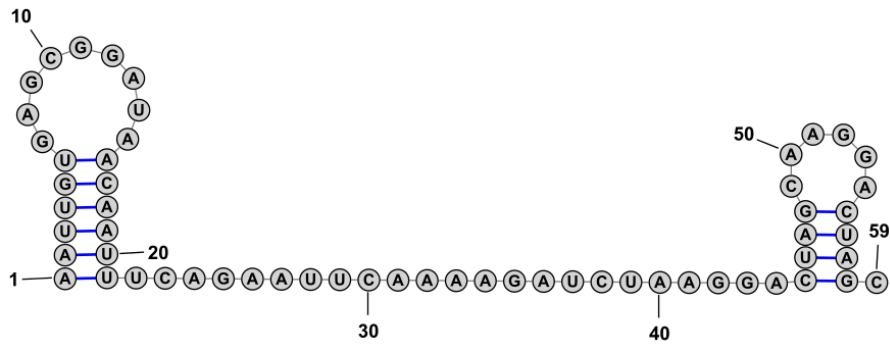
Supplementary figure 12: Secondary structure analysis for reporters S7-S10 with hairpin structures using RNAfold.

H1

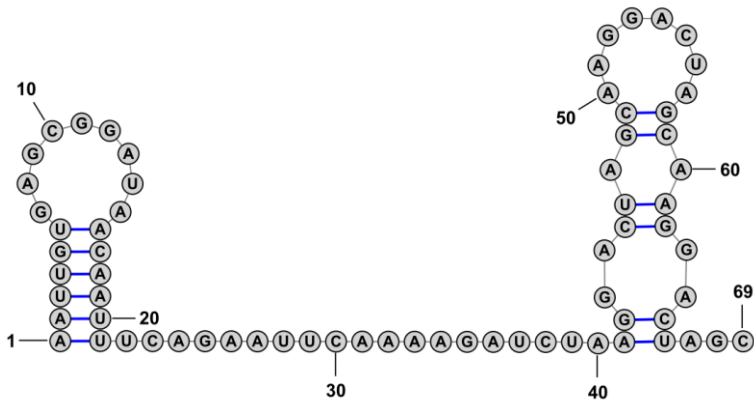


8. Appendix

H2

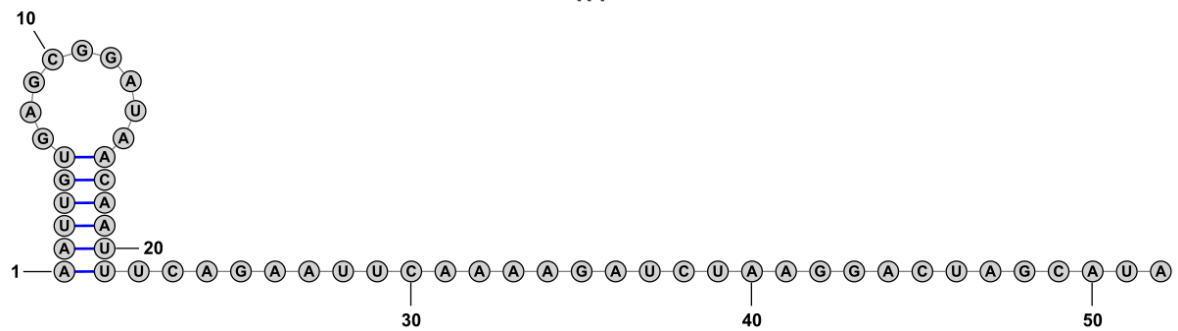


H3



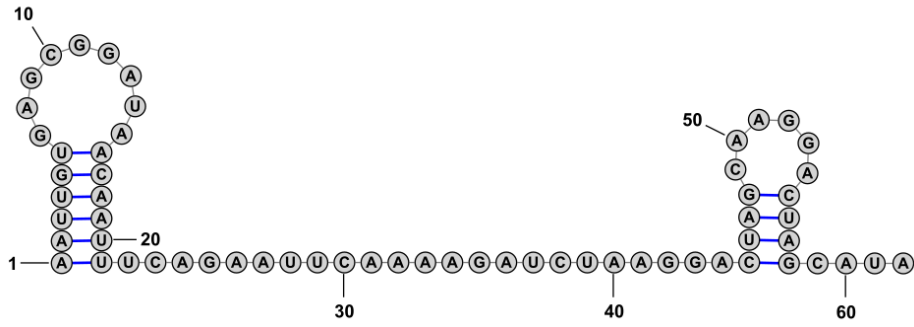
Supplementary figure 13: Secondary structure analysis for reporters H1-H3 with hairpin structures using RNAfold.

H4

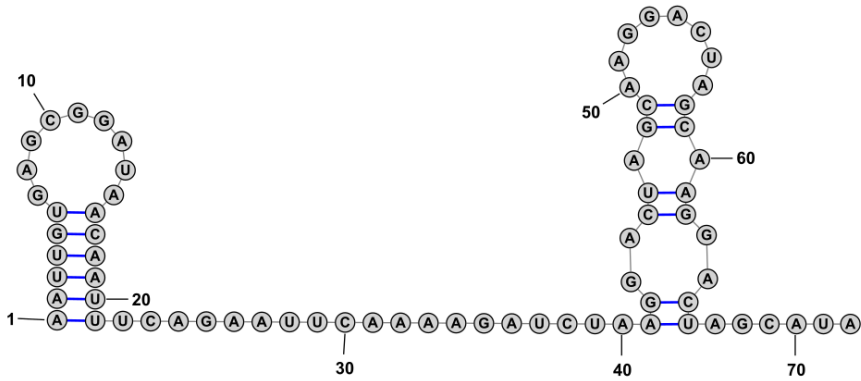


8. Appendix

H5

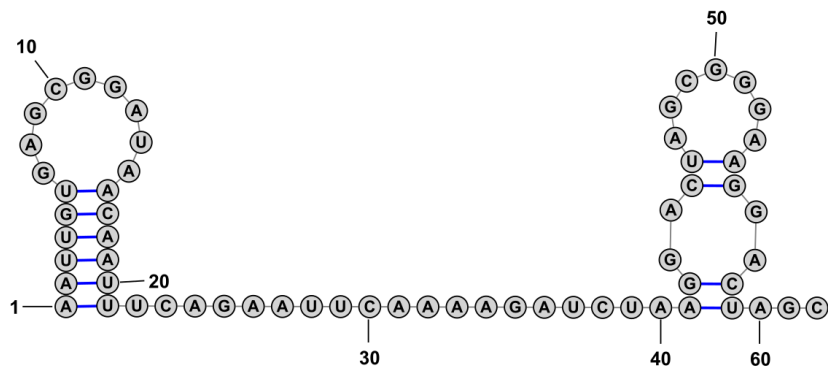


H6

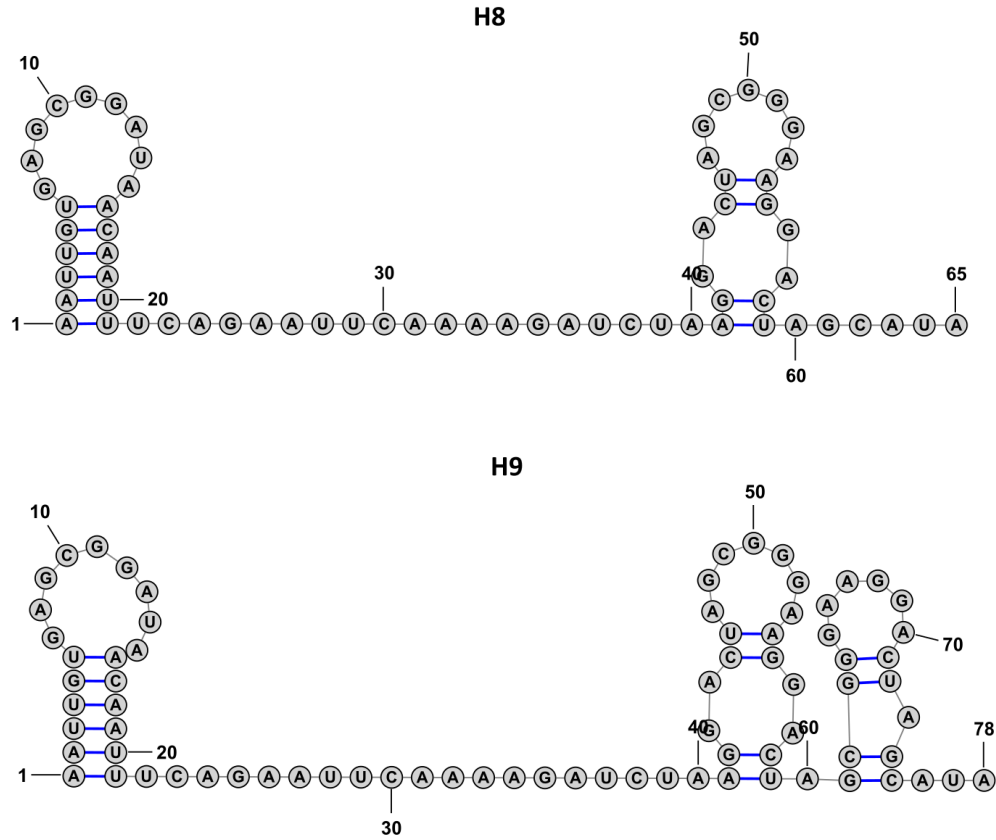


Supplementary figure 14: Secondary structure analysis for reporters H4-H6 with hairpin structures using RNAfold.

H7



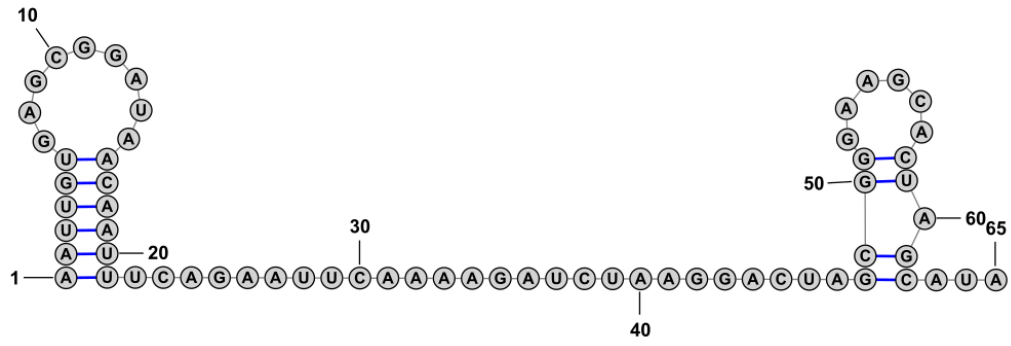
8. Appendix



Supplementary figure 15: Secondary structure analysis for reporters H7-H9 with hairpin structures using RNAfold.

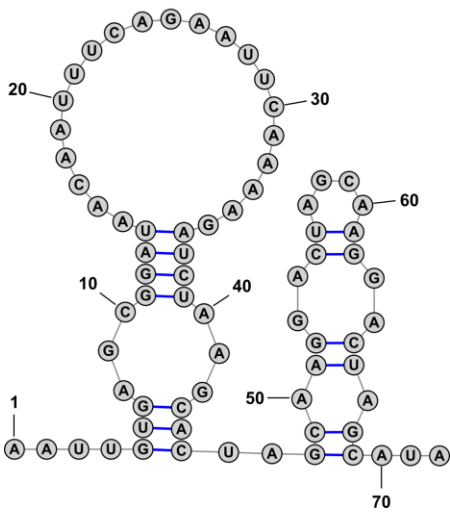
8. Appendix

H11

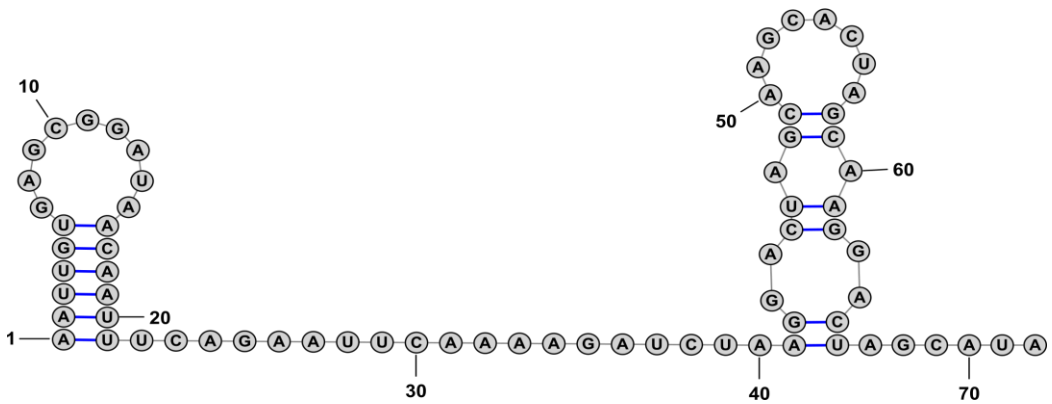


Supplementary figure 16: Secondary structure analysis for reporters H10 and H11 with hairpin structures using RNAfold.

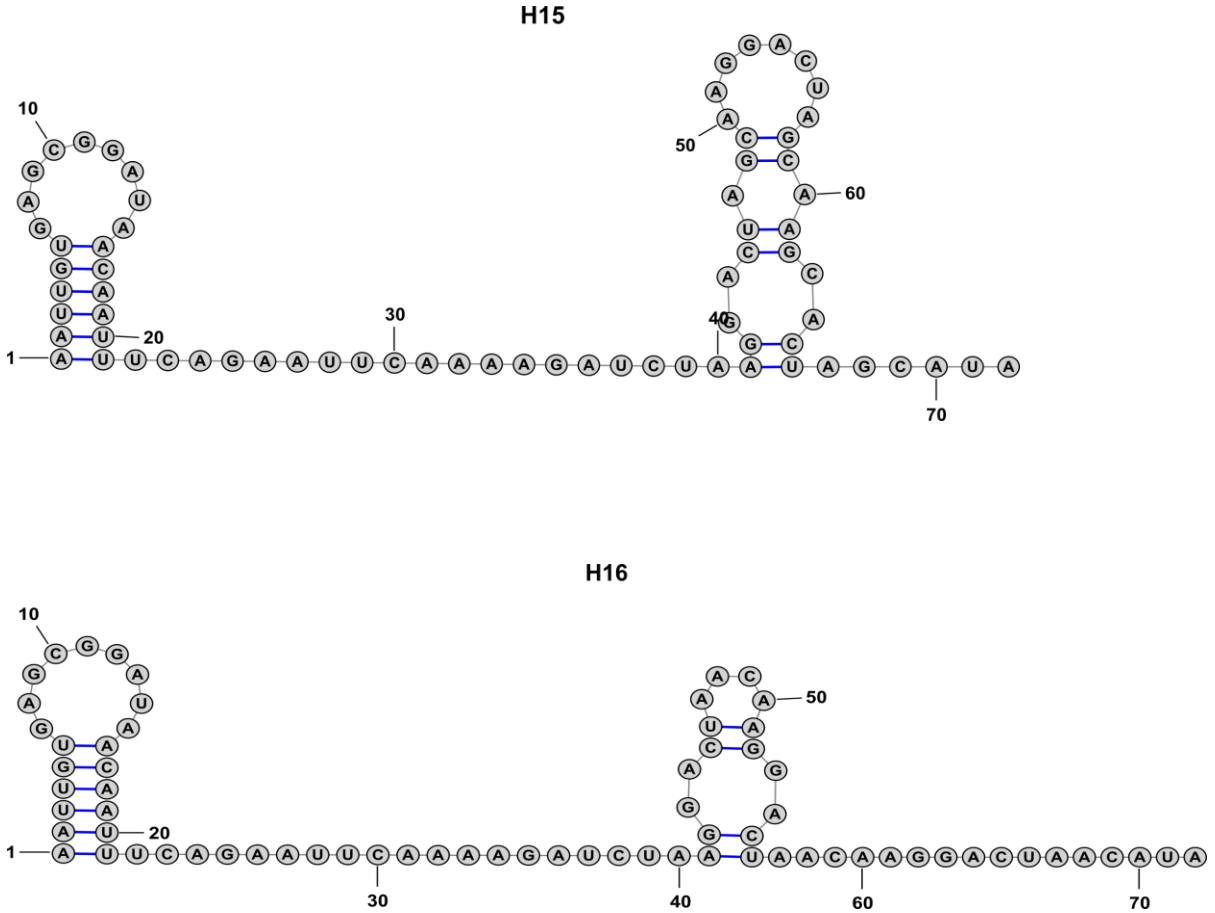
H13



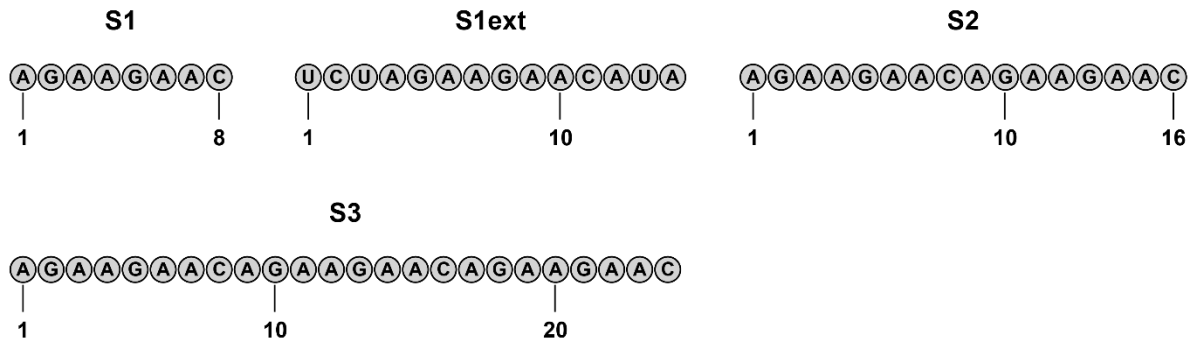
H14



8. Appendix

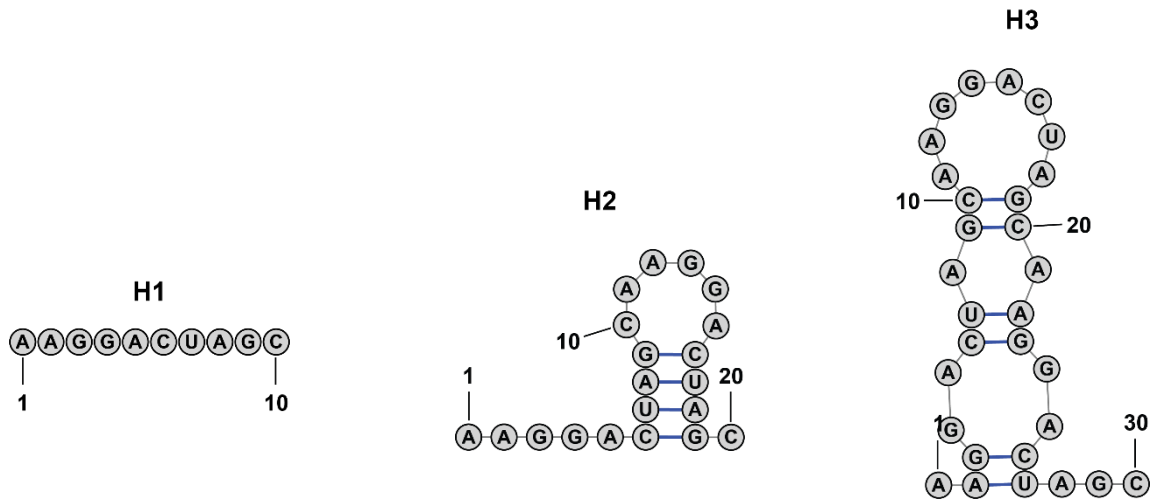


Supplementary figure 17: Secondary structure analysis for reporters H13-H16 with hairpin structures using RNAfold.

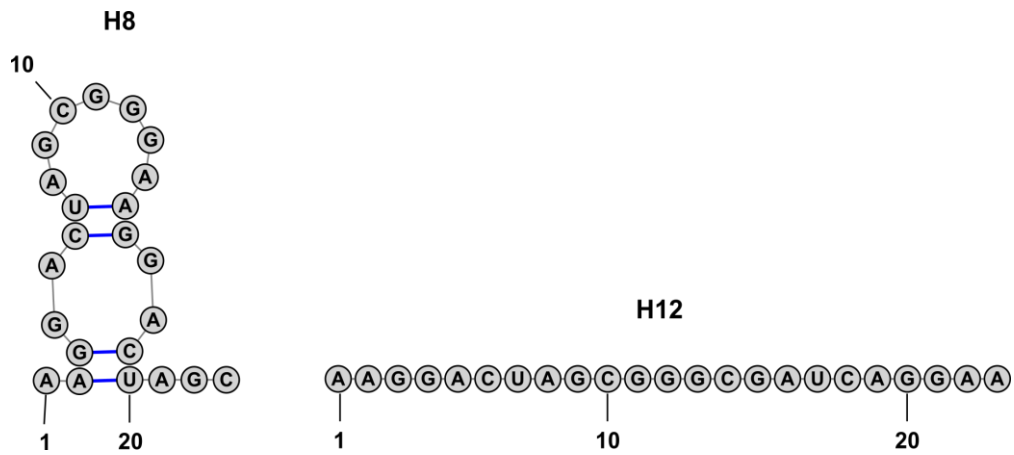


Supplementary figure 18: Secondary structure analysis for RNA sequences FAM-S1, FAM-S1ext, FAM-S2, and FAM-S3 used in fluorescence polarisation assays using RNAfold.

8. Appendix



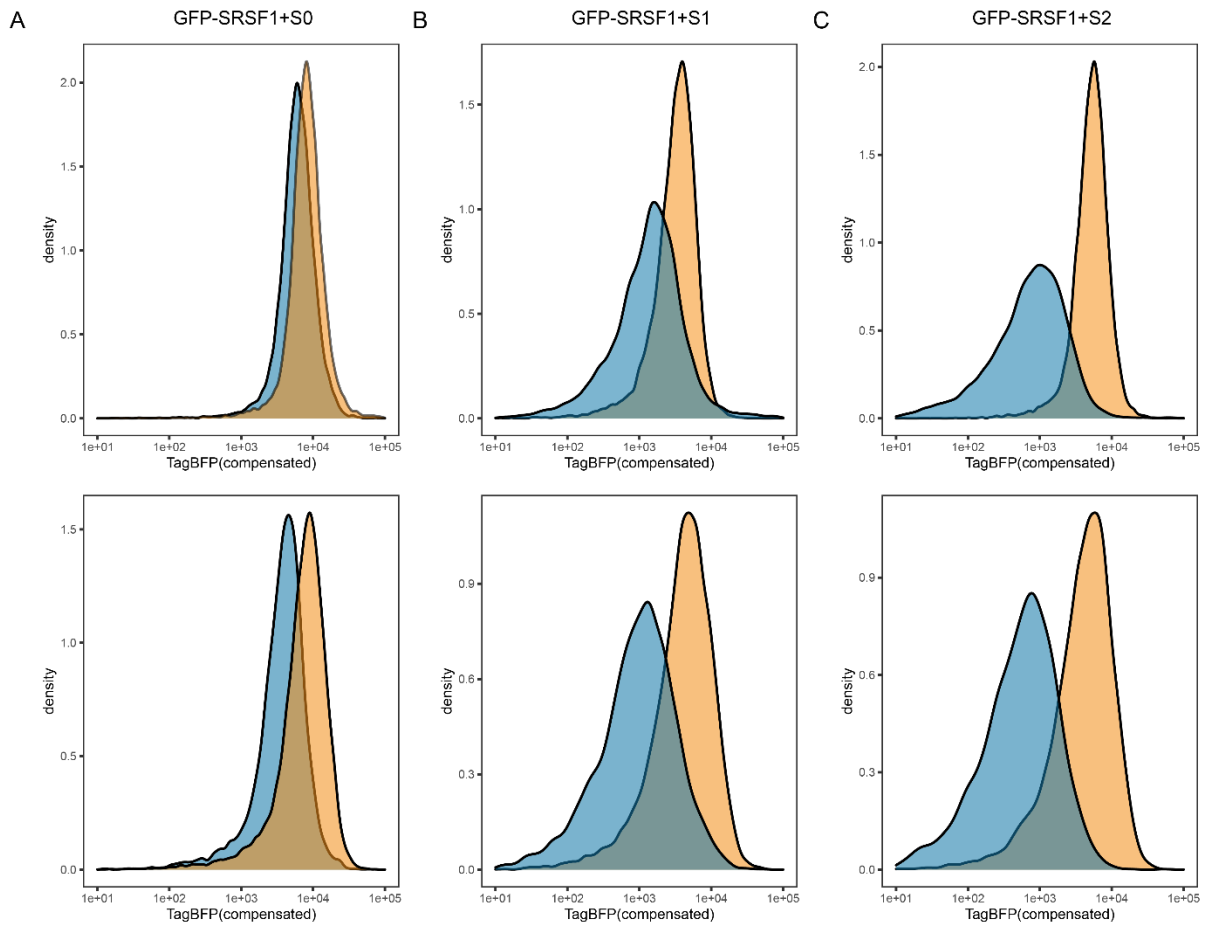
Supplementary figure 19: Secondary structure analysis for RNA sequences FAM-H1-FAM-H3 used in fluorescence polarisation assays using RNAfold.



Supplementary figure 20: Secondary structure analysis for RNA sequences FAM-H8 and FAM-H12 used in fluorescence polarisation assays using RNAfold.

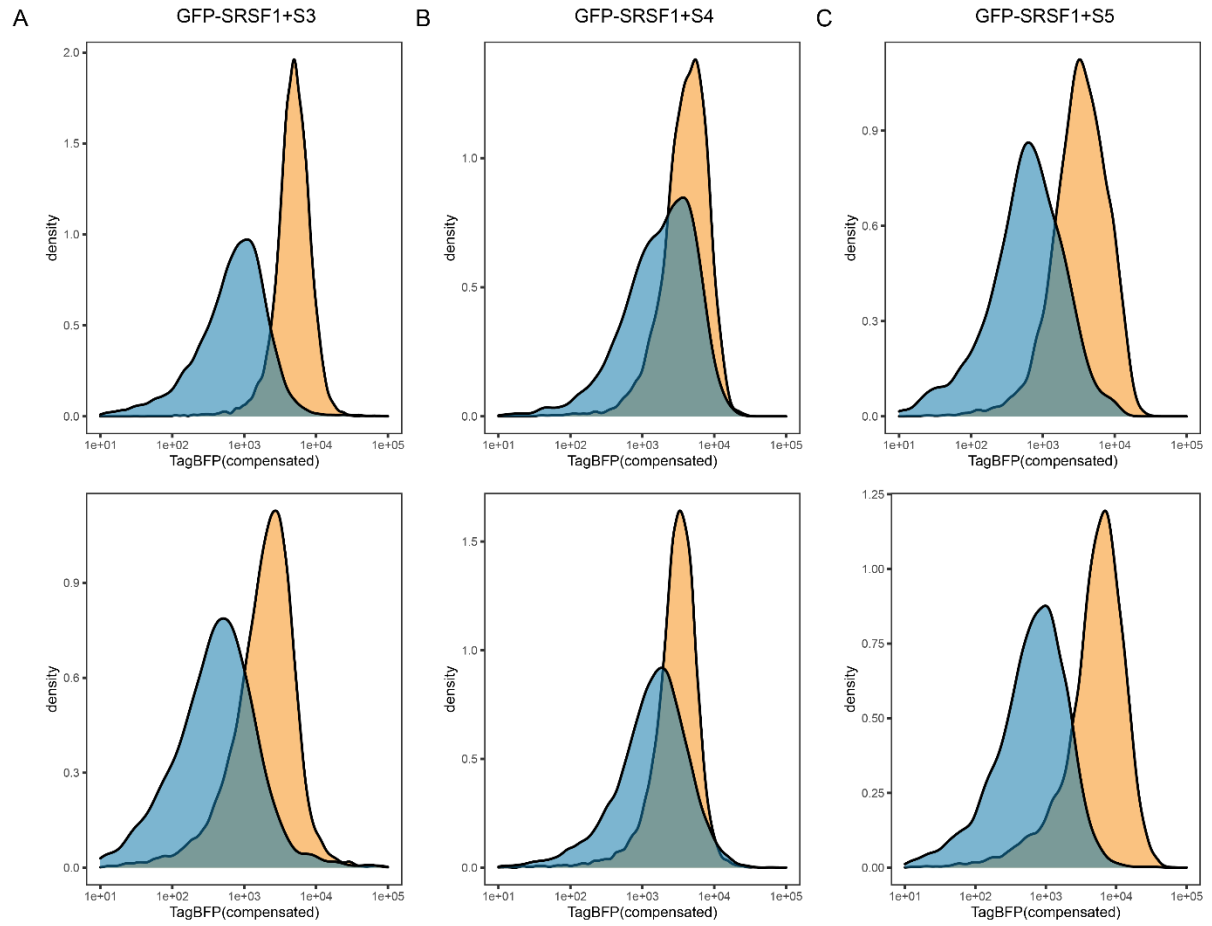
8. Appendix

8.2.3. Supplementary figures: Flow cytometry analysis



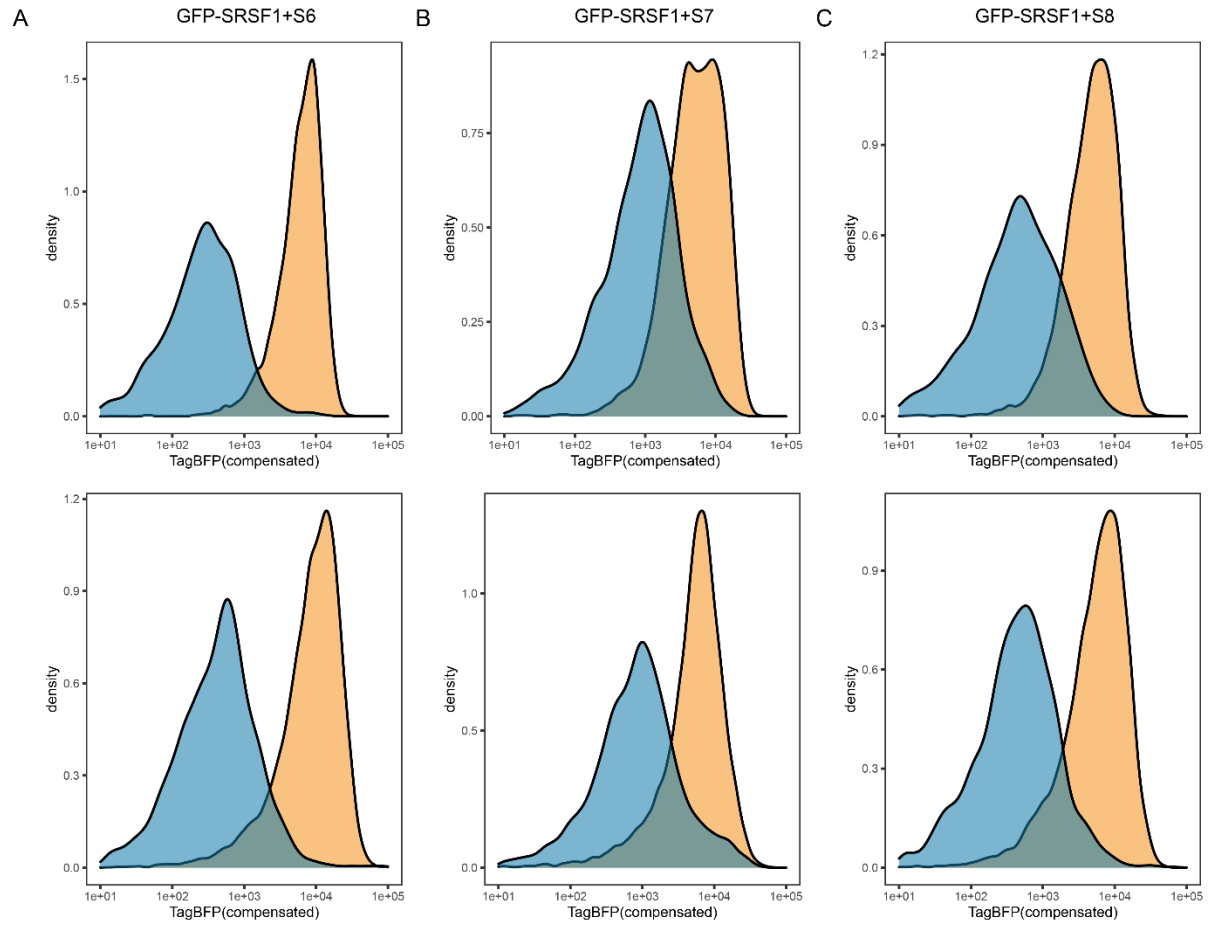
Supplementary figure 21: Histograms for reporters S0 (A), S1 (B), and S2 (C) in the presence of sfGFP-SRSF1 (blue) or sfGFP-PTBP1 (orange). Two replicates were performed for each measurement.

8. Appendix



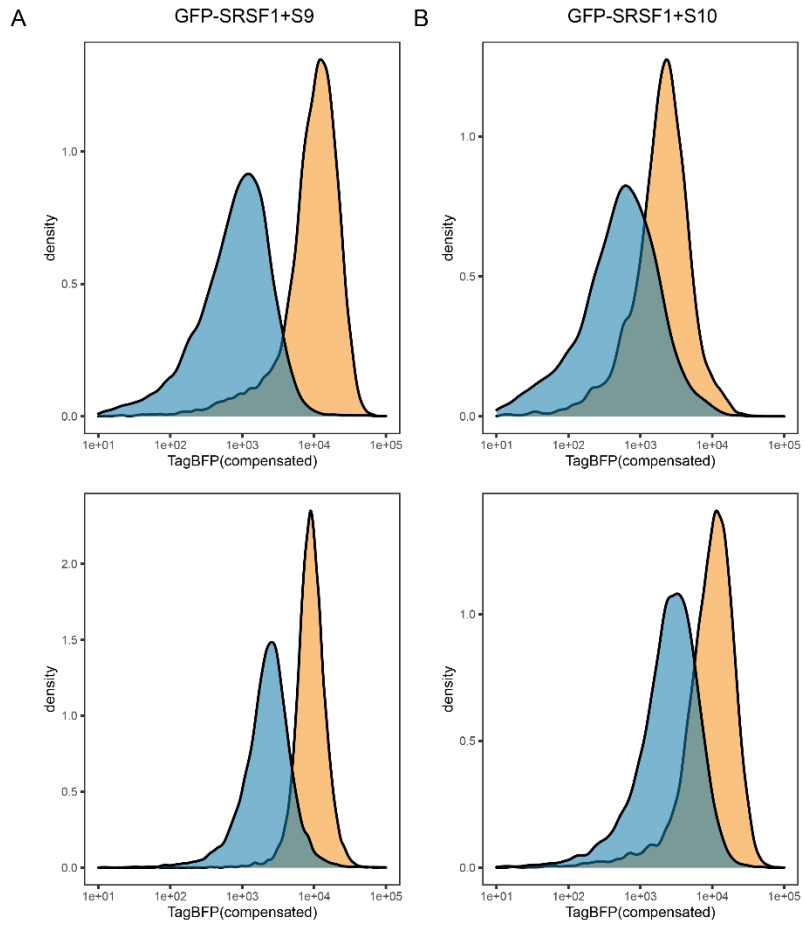
Supplementary figure 22: Histograms for reporters S3 (A), S4 (B), and S5 (C) in the presence of sfGFP-SRSF1 (blue) or sfGFP-PTBP1 (orange). Two replicates were performed for each measurement.

8. Appendix



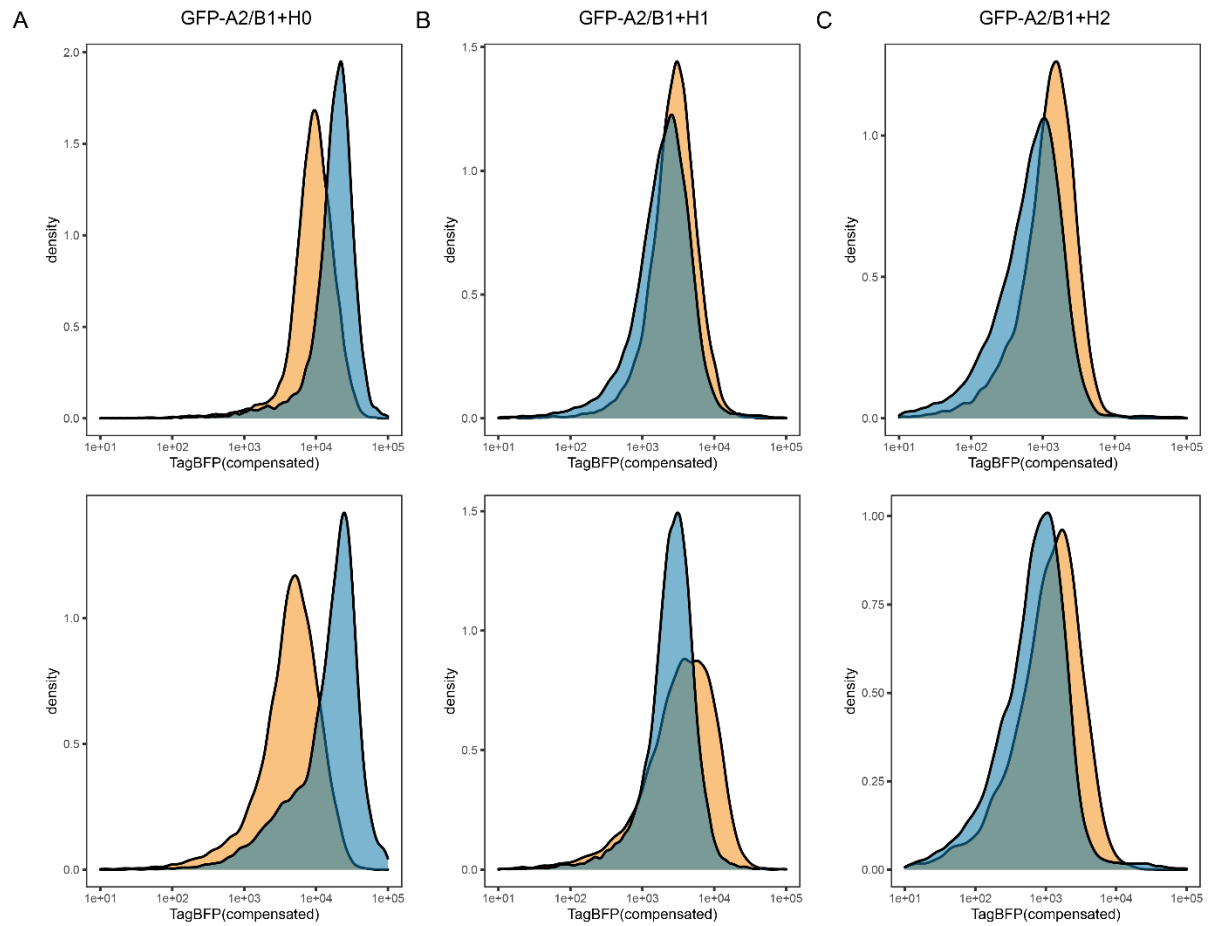
Supplementary figure 23: Histograms for reporters S6 (A), S7 (B), and S8 (C) in the presence of sfGFP-SRSF1 (blue) or sfGFP-PTBP1 (orange). Two replicates were performed for each measurement.

8. Appendix



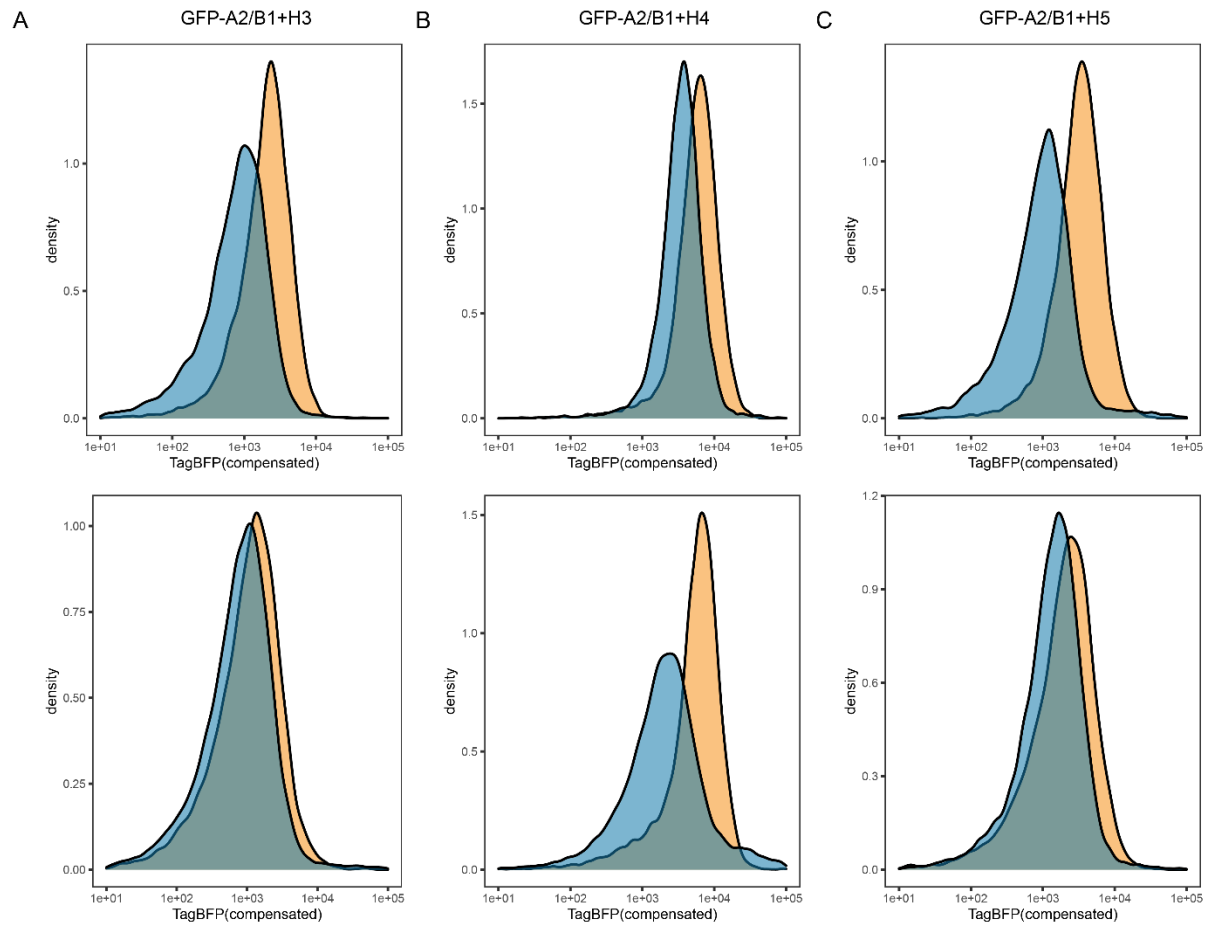
Supplementary figure 24: Histograms for reporters S9 (A) and S10 (B) in the presence of sfGFP-SRSF1 (blue) or sfGFP-PTBP1 (orange). Two replicates were performed for each measurement.

8. Appendix



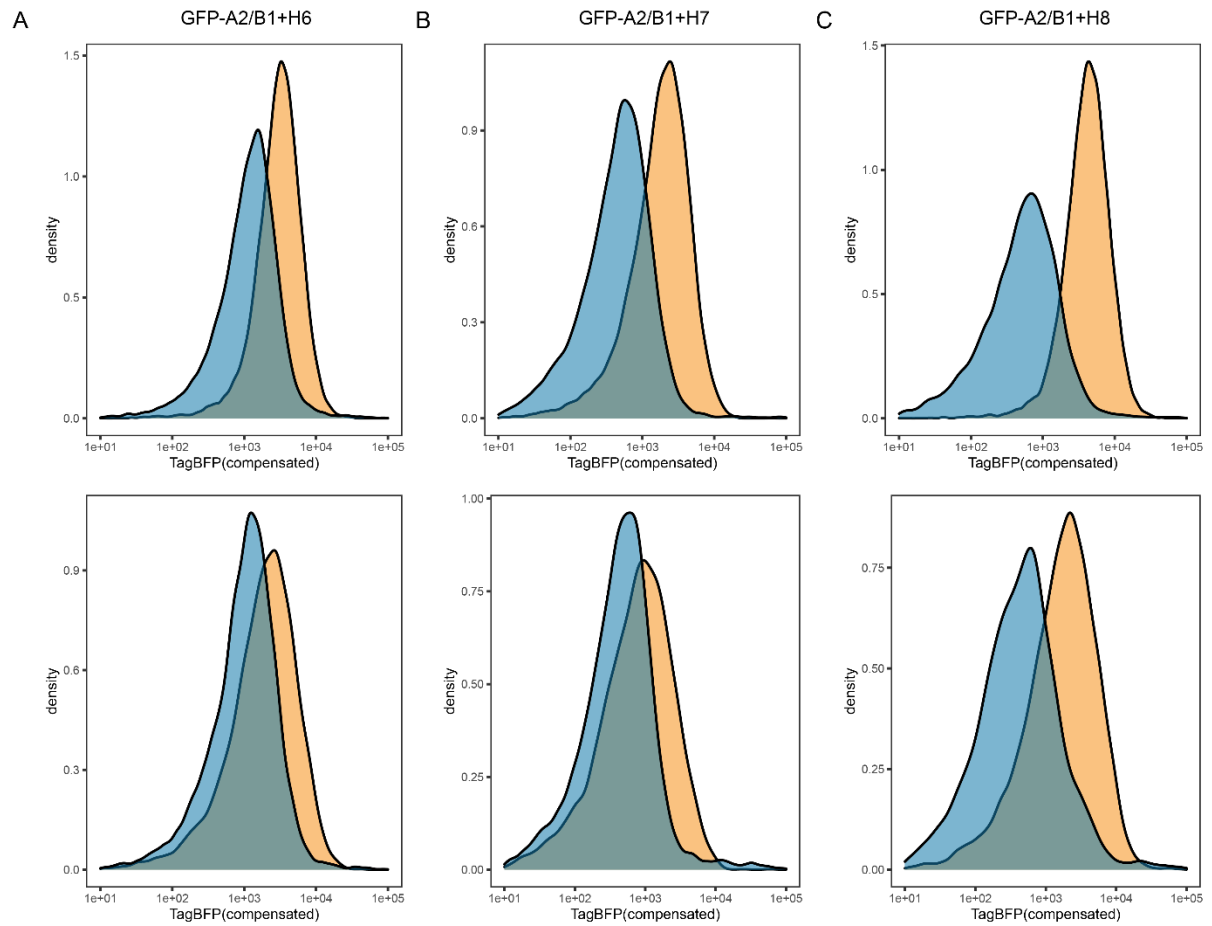
Supplementary figure 25: Histograms for reporters H0 (A), H1 (B), and H2 (C) in the presence of sfGFP-SRSF1 (blue) or sfGFP-PTBP1 (orange). Two replicates were performed for each measurement.

8. Appendix



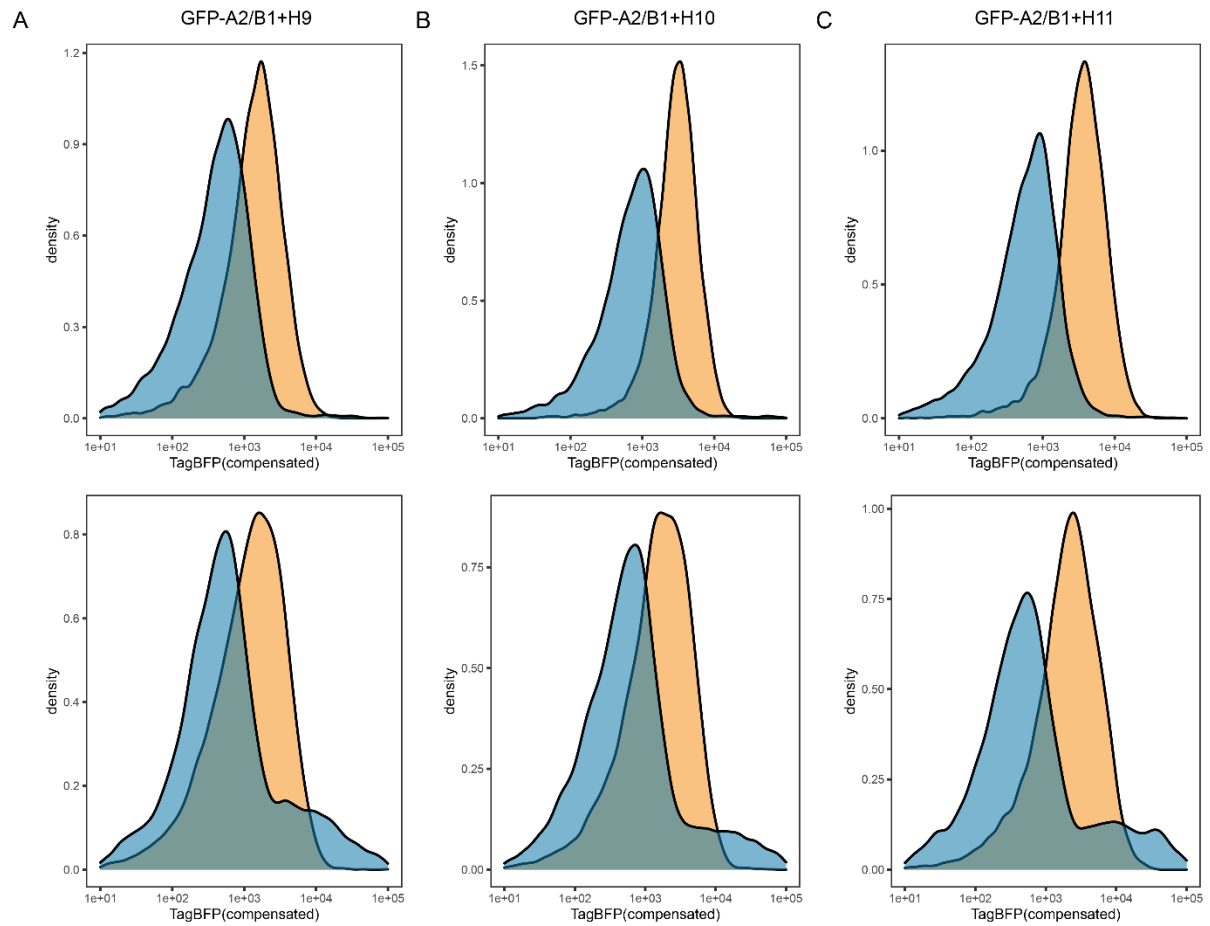
Supplementary figure 26: Histograms for reporters H3 (A), H4 (B), and H5 (C) in the presence of sfGFP-hnRNP A2B1 (blue) or sfGFP-PTBP1 (orange). Two replicates were performed for each measurement.

8. Appendix

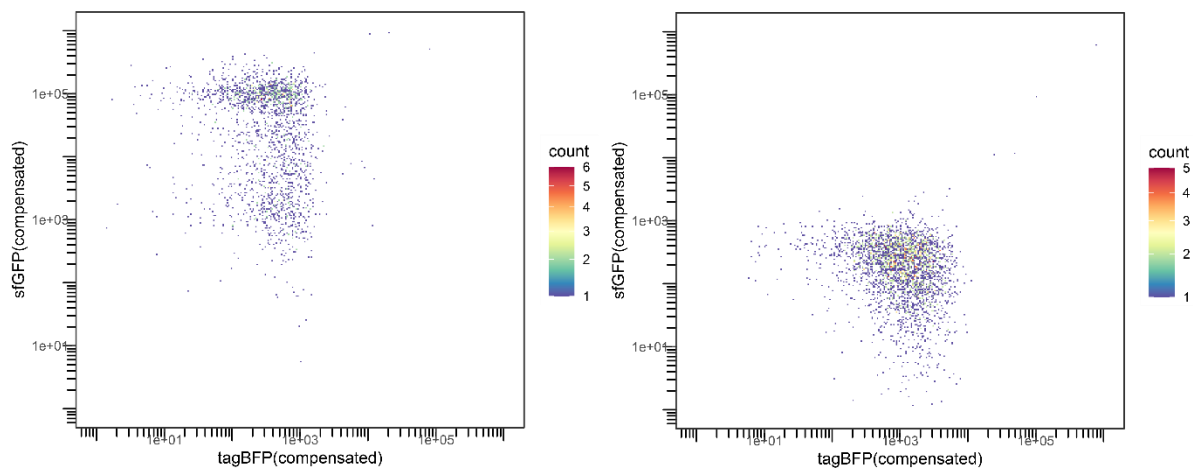


Supplementary figure 27: Histograms for reporters H6 (A), H7 (B), and H8 (C) in the presence of sfGFP-hnRNP A2B1 (blue) or sfGFP-PTBP1 (orange). Two replicates were performed for each measurement.

8. Appendix



Supplementary figure 28: Histograms for reporters H9 (A), H10 (B), and H11 (C) in the presence of sfGFP-hnRNP A2B1 (blue) or sfGFP-PTBP1 (orange). Two replicates were performed for each measurement.

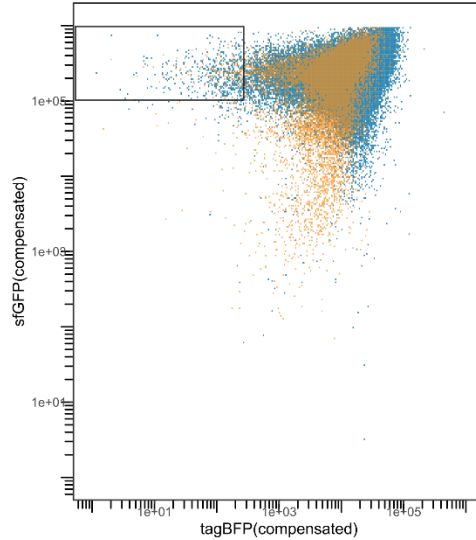


Supplementary figure 29: Flow cytometry results for the One-color controls “sfGFP only” and “reporter S0/H0”. Top10F⁺ cells were transformed with either the protein - or the reporter plasmid and grown in

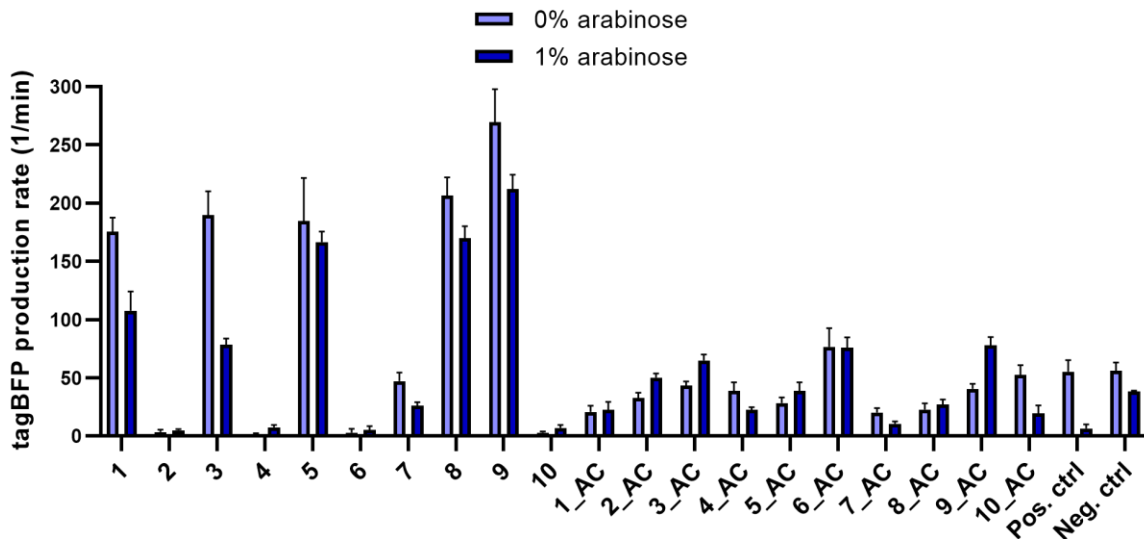
8. Appendix

medium with the respective antibiotic (chloramphenicol or kanamycin) and induced with either arabinose or IPTG.

8.2.4. Supplementary figures: RBP consensus screening

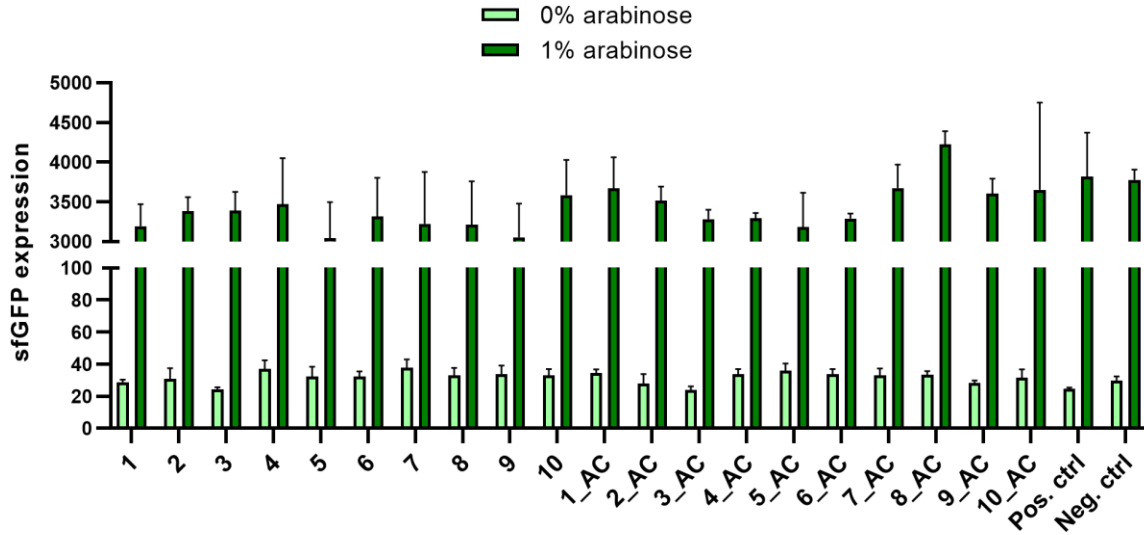


Supplementary figure 30: Histograms of the SRSF1 consensus sequence screening population (blue) with the negative control population (orange) carrying the sfGFP-SRSF1/S0 plasmid pair. Cells that produced high sfGFP levels and low levels of tagBFP were gated for sorting (black box).

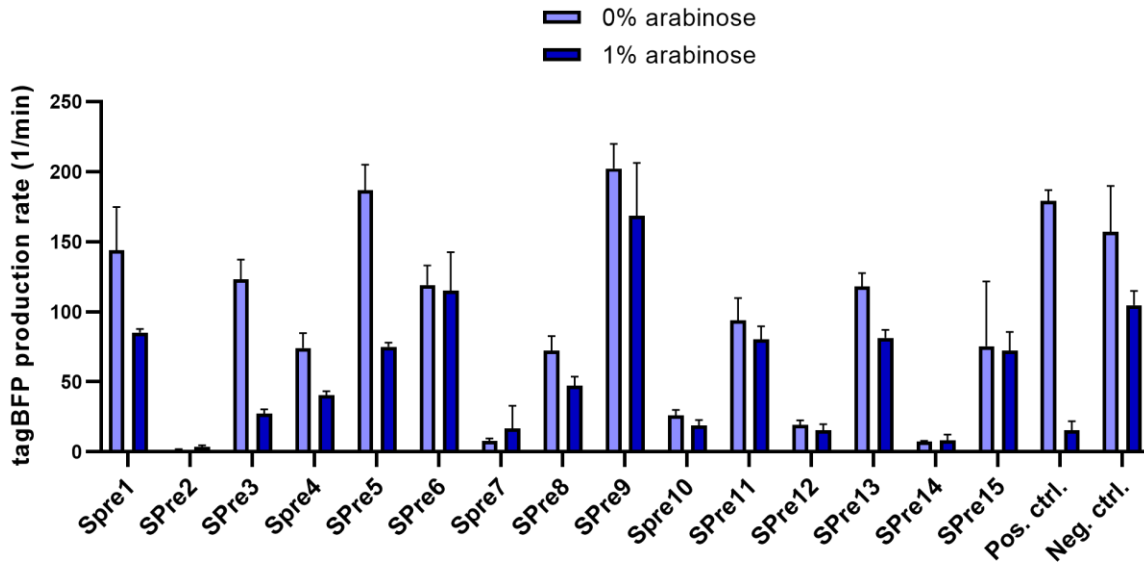


Supplementary figure 31: TagBFP production rate [1/min] of the hit sequences 1-10 and the autorepressor corrected samples 1_AC-10_AC upon arabinose induction at 0 or 1 % for the hnRNP A2B1 consensus sequence screening. Data are mean values (n=2, N=2).

8. Appendix

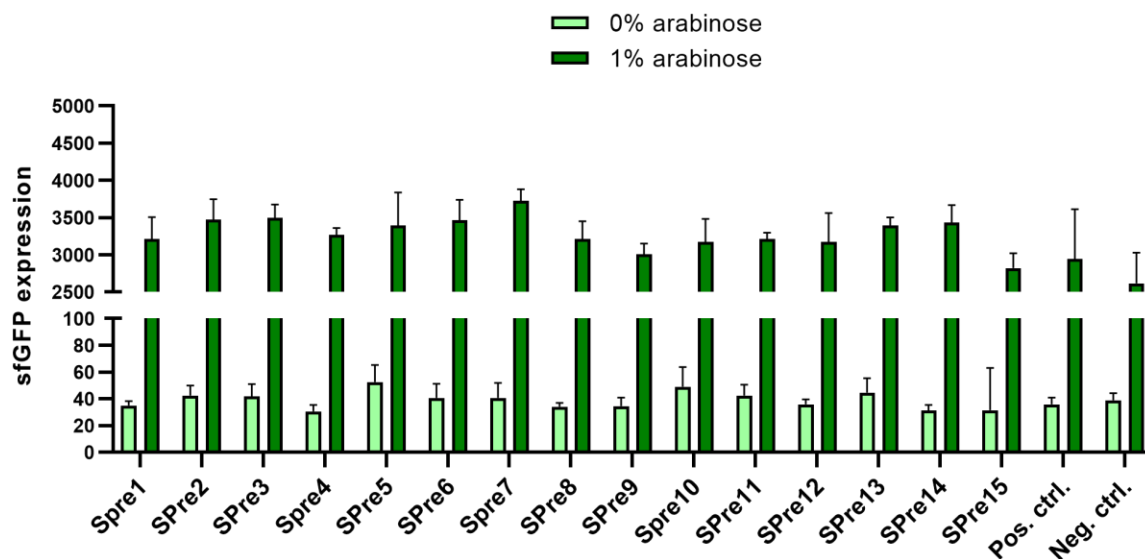


Supplementary figure 32: SfGFP expression of the hit sequences 1-10 and the autorepressor corrected samples 1-10 upon arabinose induction at 0 or 1 % for the hnRNP A2B1 consensus sequence screening. Data are mean values (n=2, N=2).



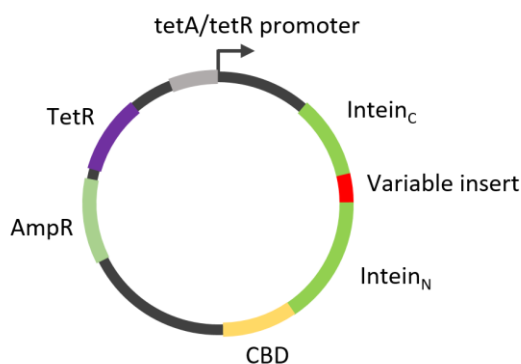
Supplementary figure 33: TagBFP production rate [1/min] of the hit sequences 1-15 upon arabinose induction at 0 or 1 % for the autorepressor presorted SRSF1 consensus sequence screening. Data are mean values (n=2, N=2).

8. Appendix



Supplementary figure 34: SfGFP expression of the hit sequences 1-15 upon arabinose induction at 0 or 1 % for the autorepressor presorted SRSF1 consensus sequence screening. Data are mean values (n=2, N=2).

8.2.5. Supplementary figures: SICLOPPS screening



Supplementary figure 35: Schematic representation of a SICLOPPS plasmid as used in the screenings. The plasmid contains the genetic information for the two Intein domains (Intein_C and Intein_N), the variable insert and a chitin-binding domain (CBD) which are under control of the tetA/tetR promoter and the ampicillin resistance (AmpR).

8.3. Supplementary tables

Supplementary table 1: Used primers for cloning. Homology regions are highlighted in green tones.

Primer	Sequence
1	TTTGTACAGTTCATCCATACCA

8. Appendix

```

2      GGATCCAAACTCGAGTAAGGATC
3      TGGTATGGATGAACTGTACAAAATGTCGGGAGGTGGTGTGA
4      CTTACTCGAGTTTGGATCCTTAATCAACTTTAACCCGGAT
5      TAAGGATCCAAACTCGAGTAAGG
6      AGATCTTTTGAATTCTGAAATTG
7      ATTTCAGAATTCAAAAGATCTNNNNNNNNNATATTTAAGAAGGAGATATAC
8      GATCCTTACTCGAGTTTGGATCC

```

Supplementary table 2: RT-qPCR primer sequences, amplicon sizes and efficiencies.

Gene	Primers	Amplicon size [bp]	Primer efficiency
<i>gapA</i>	Fw: CGGCATCATCGAAGGTCTGA Rec: GCTTTAGCAGCACCGGTAGA	144	98%
<i>16s rRNA</i>	Fw: CAAGACCAAAGAGGGGGACC Rev: GGACCGTGTCTCAGTTCCAG	148	101%
<i>KanR</i>	Fw: ATGCGATGTTTCGCTTGGTG Rev: TTGTCACTGAAGCGGGAAGG	175	94%
<i>tagBFP</i> (synthetic construct, GenBank ID: AZQ25074.1)	Fw: CACCGTGGACAACCATCACT Rev: TGAAGGTCTTGCTGCCGTAG	158	108%
<i>sfgfp</i>	Fw: GGGTGAAGGTGACGCAACTA Rev: CGAGCAAAGCACTGAACACC	123	99%

Supplementary table 3: Repression ratios for the sfGFP-fusion proteins SRSF1, hnRNP A2B1, PTBP1 and sfGFP alone in combination with reporter constructs

Reporter construct	Repression ratio sfGFP-SRSF1	Repression ratio sfGFP-SRSF1 _{mut}	Repression ratio sfGFP-A2B1	Repression ratio sfGFP-PTBP1	Repression ratio sfGFP
S0	1.2 ± 0.0	-	-	1.4 ± 0.1	-
S1	3.3 ± 0.3	-	-	1.5 ± 0.2	1.5 ± 0.1
S2	9.0 ± 2.2	-	-	1.7 ± 0.1	1.4 ± 0.1
S3	15.1 ± 2.2	-	-	1.7 ± 0.1	1.5 ± 0.1
S4	4.8 ± 1.0	-	-	1.4 ± 0.1	-
S5	12.5 ± 1.0	-	-	1.5 ± 0.2	-
S6	17.0 ± 0.4	1.6 ± 0.2	-	1.7 ± 0.1	-
S6-4	14.5 ± 1.6	-	-	-	-
S6-5	13.5 ± 1.1	-	-	-	-
S6-6	11.9 ± 2.3	-	-	-	-
S6-7	11.1 ± 2.7	-	-	-	-

8. Appendix

S6-8	9.2 ± 1.6	-	-	-	-
S6-9	8.4 ± 3.1	-	-	-	-
S6-10	6.5 ± 0.8	-	-	-	-
S8	16.1 ± 4.0	-	-	1.7 ± 0.2	-
S9	11.8 ± 1.3	-	-	1.5 ± 0.3	-
S10	3.2 ± 0.4	-	-	1.8 ± 0.3	-
H0	-	-	1.3 ± 0.2	-	-
H1	-	-	5.3 ± 0.4	1.6 ± 0.1	1.4 ± 0.1
H2	-	-	2.7 ± 0.4	1.3 ± 0.5	1.0 ± 0.1
H3	-	-	5.0 ± 1.1	1.3 ± 0.1	1.2 ± 0.1
H4	-	-	3.7 ± 0.2	1.3 ± 0.2	-
H5	-	-	2.6 ± 0.6	1.3 ± 0.5	-
H6	-	-	4.6 ± 0.7	1.4 ± 0.2	-
H7	-	-	4.6 ± 0.2	1.1 ± 0.3	-
H8	-	-	6.4 ± 0.1	1.4 ± 0.1	-
H9	-	-	5.0 ± 1.7	1.3 ± 0.1	-
H10	-	-	4.6 ± 0.3	1.2 ± 0.1	-
H11	-	-	6.0 ± 0.5	1.1 ± 0.3	-
H12	-	-	7.2 ± 0.6	1.5 ± 0.3	-
H13	-	-	3.1 ± 0.7	1.2 ± 0.2	-
H14	-	-	5.9 ± 1.0	1.6 ± 0.1	-
H15	-	-	3.2 ± 0.4	1.2 ± 0.1	-
H16	-	-	3.4 ± 0.4	1.8 ± 0.2	-

- = not tested

Supplementary table 4: Number of distancing nucleotides between a secondary structure and the S/D sequence of all reporters.

Reporter construct	Nr. of distancing nucleotides	Reporter construct	Nr. of distancing nucleotides
S0	18	H0	18
S1	4	H1	28
S2	12	H2	1
S3	20	H3	3
S4	7	H4	31
S5	15	H5	4
S6	23	H6	6
S7	21	H7	3
S8	23	H8	6
S9	23	H9	3
S10	3	H10	12
S6-4	24	H11	3
S6-5	25	H12	8
S6-6	26	H13	3
S6-7	27	H14	6
S6-8	28	H15	6

8. Appendix

S6-9	29	H16	16
S6-10	30		

Supplementary table 5: List of used RNAs

Name	Sequence	Modification
FAM-S1	AGAAGAAC	5'(6-FAM)
FAM-S1ext	UCUAGAAGAACAUA	5'(6-FAM)
FAM-S2	AGAAGAACAGAAGAAC	5'(6-FAM)
FAM-S3	AGAAGAACAGAAGAACAGAAGAAC	5'(6-FAM)
S3	AGAAGAACAGAAGAACAGAAGAAC	-
Polyuridine	AUUUUUCCAUCUUUGUAUC	5'(6-FAM)
FAM-H1	AAGGACUAGC	5'(6-FAM)
FAM-H2	AAGGACUAGCAAGGACUAGC	5'(6-FAM)
FAM-H3	AAGGACUAGCAAGGACUAGCAAGGACUAGC	5'(6-FAM)
FAM-H8	AAGGACUAGCGGGAAGGACUAGC	5'(6-FAM)
FAM-H12	AAGGACUAGCGGGCGAUCAGGAA	5'(6-FAM)
Poly C	CCCCCCCC	5'(6-FAM)
Cy5-S3	AGAAGAACAGAAGAACAGAAGAAC	5'(Cy5)
Cy5-H1	AAGGACUAGC	3'(Cy5)
H1	AAGGACUAGC	-
Cy5-H8	AAGGACUAGCGGGAAGGACUAGC	3'(Cy5)
Cy5-Polyuridine	AUUUUUCCAUCUUUGUAUC	3'(Cy5)

Supplementary table 6: List of used proteins

Name	Protein	Tags	Residues
SRSF1 RRM1+2	SRSF1	-	1-195
MBP-A2B1	hnRNP A2B1	MBP	1-251

Supplementary table 7: Sorting statistics.

Experiment/sorting	Total events	Gated event count (Regular cells)	Sorted count in gate	Sorting efficiency	Sorting mode

8. Appendix

SRSF1 consensus sequence screening	18,001,701	10,000,000	214,146	79.33 %	Normal
hnRNP A2B1 consensus sequence screening	13,063,274	10,000,000	209,709	88.50 %	Normal
Autorepressor selection	18,137,637	10,000,000	48,820	80.61 %	Normal
Autorepressor Presorting 1/2	20,130,098	10,000,000	6,290,965	85.84 %	Purity
Autorepressor Presorting 2/2	2,819,981	1,000,000	667,116	82.66 %	Purity
Autorepressor presorted SRSF1 consensus sequence screening	6,692,713	5,000,000	43,641	76.12 %	Purity
SRSF1 SICLOPPS screening (Sanger)	10,212,230	-	3389	92.62 %	Normal
SRSF1 SICLOPPS screening (Illumina)	12,072,859	-	23,732	64.30 %	Purity
SRSF1 SICLOPPS screening over 2 rounds - round 1 (Illumina)	10,900,767	-	77,063	80.97 %	Normal
SRSF1 SICLOPPS screening over 2 rounds - round 2 (Illumina)	12,103,564	-	19,719	66.08 %	Purity
hnRNP A2B1 SICLOPPS screening 1/2 (Illumina)	7,043,906	-	28,624	78.05 %	Purity
hnRNP A2B1 SICLOPPS 2/2 (Illumina)	5,048,433	-	28,567	78.90 %	Purity

8.4. Supplementary material

8.4.1. Plasmid sequences

S0/H0 plasmid

GACGTCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCG
CTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGG
GAGAGGCGGTTTTCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCACTGAGACGGG
CAACAGCTGATTGCCCTTACCCGCTTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCAC
GCTGGTTTGGCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAA
CATGAGCTGTCTTCGGTATCGTTCGTATCCCACTACCGAGATGTCCGCACCAACGCGCA
GCCCCGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCA
GCATCGCAGTGGGAACGATGCCCTCATTGAGCATTGTCATGGTTTGTGAAAACCGGA
CATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGA
TATTTATGCCAGCCAGCCAGACGCGAGACGCGCCGAGACAGAACTTAATGGGCCCGCTA
ACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCAGTCGCGTAC
CGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAAT
AACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCG
GATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTT
ACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGA
TCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTG
GAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGG
TTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTTCGCAGA
AACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATA
CTCTGCGACATCGTATAACGTTACTGTTTTACATTCACCACCCTGAATTGACTCTCTTC
CGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATC
TCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGTTGAGG
CCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGT
CCCCCGGCCACGGGGCCTGCCACCATACCACGCCGAAACAAGCGCTCATGAGCCCG
AAGTGGCGAGCCCGATCTTCCCATCGGTGATGTGCGCGATATAGGCGCCAGCAACC
GCACCTGTGGCGCCGGTGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATC
GTTTAGGCACCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGAATTGT
GAGCGGATAACAATTTGAGAATTCAAAAGATCTTTTAAGAAGGAGATATACATATGAGC
GAGCTGATTAAGGAGAACATGCACATGAAGCTGTACATGGAGGGCACCGTGGACAACC
ATCACTTCAAGTGCACATCCGAGGGCGAAGGCAAGCCCTACGAGGGCACCCAGACCA

8. Appendix

TGAGAATCAAGGTGGTCGAGGGCGGCCCTCTCCCCTTCGCCTTCGACATCCTGGCTAC
TAGCTTCCTCTACGGCAGCAAGACCTTCATCAACCACACCCAGGGCATCCCCGACTTC
TTCAAGCAGTCCTTCCCTGAGGGCTTCACATGGGAGAGAGTCACCACATACGAAGACG
GGGGCGTGCTGACCGCTACCCAGGACACCAGCCTCCAGGACGGCTGCCTCATCTACA
ACGTCAAGATCAGAGGGGTGAACTTCACATCCAACGGCCCTGTGATGCAGAAGAAAAC
ACTCGGCTGGGAGGCCTTACCCGAGACGCTGTACCCCGCTGACGGCGGCCTGGAAG
GCAGAAACGACATGGCCCTGAAGCTCGTGGGCGGGAGCCATCTGATCGCAAACATCA
AGACCACATATAGATCCAAGAAACCCGCTAAGAACCTCAAGATGCCTGGCGTCTACTAT
GTGGACTACAGACTGGAAAGAATCAAGGAGGCCAACAACGAGACCTACGTCGAGCAG
CACGAGGTGGCAGTGGCCAGATACTGCGACCTCCCTAGCAAACCTGGGGCACAAGCTT
AACTAAGGATCCAAACTCGAGTAAGGATCTCCAGGCATCAAATAAAACGAAAGGCTCAG
TCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTGGTGAACGCTCTCTACTAGAG
TCACACTGGCTCACCTTCGGGTGGGCCTTCTGCGTTTATACCTAGGGATATATTCCGC
TTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTGACTGCGGCGAGCGGAAATGGCT
TACGAACGGGGCGGAGATTTCTGGAAGATGCCAGGAAGATACTTAACAGGGGAAGTGA
GAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCCCCTGACAAGCATCACGA
AATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCG
TTTTCCCCTGGCGGCTCCCTCGTGCCTCTCCTGTTCTCCTGCCTTTCGGTTTACCGGTG
TCATTCCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTA
GGCAGTTCGCTCCAAGCTGGACTGTATGCACGAACCCCCGTTGAGTCCGACCGCTG
CGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAAGACATGCAAAAGCACCA
CTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGT
TAAGGCTAAACTGAAAGGACAAGTTTTGGTACTGCGCTCCTCCAAGCCAGTTACCTC
GGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAACCGCCCTGCAAGGCGGTTTT
TTCGTTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTA
TTAATCAGATAAAATATTTCTAGATTTTCAGTGCAATTTATCTCTTCAAATGTAGCACCTGA
AGTCAGCCCCATACGATATAAGTTGTTACTAGTGCTTGGATTCTCACCAATAAAAAACG
CCCGGCGGCAACCGAGCGTTCTGAACAAATCCAGATGGAGTTCTGAGGTCATTAAGTGG
ATCTATCAACAGGAGTCCAAGCGAGCTCTCGAACCCAGAGTCCCGCTCAGAAGAACT
CGTCAAGAAGGCGATAGAAGGCGATGCGCTGCGAATCGGGAGCGGCGATACCGTAA
GCACGAGGAAGCGGTCAGCCCATTGCGCCCAAGCTCTTCAGCAATATCACGGGTAG
CCAACGCTATGTCCTGATAGCGGTCCGCCACACCCAGCCGGCCACAGTCGATGAATCC
AGAAAAGCGGCCATTTTCCACCATGATATTCGGCAAGCAGGCATCGCCATGGGTCACG
ACGAGATCCTCGCCGTCGGGCATGCGCGCCTTGAGCCTGGCGAACAGTTCGGCTGGC

8. Appendix

GCGAGCCCCTGATGCTCTTCGTCCAGATCATCCTGATCGACAAGACCGGCTTCCATCC
GAGTACGTGCTCGCTCGATGCGATGTTTCGCTTGGTGGTGAATGGGCAGGTAGCCG
GATCAAGCGTATGCAGCCGCCGATTGCATCAGCCATGATGGATACTTTCTCGGCAGG
AGCAAGGTGAGATGACAGGAGATCCTGCCCGGCACTTCGCCCAATAGCAGCCAGTC
CCTTCCCGCTTCAGTGACAACGTGAGCACAGCTGCGCAAGGAACGCCCGTCGTGGC
CAGCCACGATAGCCGCGCTGCCTCGTCCTGCAGTTCATTAGGGCACCGGACAGGTC
GGTCTTGACAAAAAGAACCGGGCGCCCCTGCGCTGACAGCCGGAACACGGCGGCATC
AGAGCAGCCGATTGTCTGTTGTGCCAGTCATAGCCGAATAGCCTCTCCACCCAAGCG
GCCGGAGAACCTGCGTGCAATCCATCTTGTTCAATCATGCGAAACGATCCTCATCCTGT
CTCTTGATCAGATCATGATCCCCTGCGCCATCAGATCCTTGGCGGCAAGAAAGCCATC
CAGTTTACTTTGCAGGGCTTCCCAACCTTACCAGAGGGCGCCCCAGCTGGCAATTCC

SfGFP-SRSF1 fusion plasmid:

GACGTCTTATGACAACTTGACGGCTACATCATTCACTTTTTCTTACAACCGGCACGGA
ACTCGCTCGGGCTGGCCCCGGTGCATTTTTTAAATACCCGCGAGAAATAGAGTTGATC
GTCAAACCAACATTGCGACCGACGGTGGCGATAGGCATCCGGGTGGTGCTCAAAG
CAGCTTCGCCTGGCTGATACGTTGGTCCTCGCGCCAGCTTAAGACGCTAATCCCTAAC
TGCTGGCGGAAAAGATGTGACAGACGCGACGGCGACAAGCAAACATGCTGTGCGACG
CTGGCGATATCAAATTGCTGTCTGCCAGGTGATCGCTGATGTACTGACAAGCCTCGC
GTACCCGATTATCCATCGGTGGATGGAGCGACTCGTTAATCGCTTCCATGCGCCGCAG
TAACAATTGCTCAAGCAGATTTATCGCCAGCAGCTCCGAATAGCGCCCTTCCCCTTGCC
CGGCGTTAATGATTTGCCAAACAGGTGCGCTGAAATGCGGCTGGTGCCTTCATCCGG
GCGAAAGAACCCCGTATTGGCAAATATTGACGGCCAGTTAAGCCATTGATGCCAGTAG
GCGCGCGGACGAAAGTAAACCCACTGGTGATACCATTGCGGAGCCTCCGGATGACGA
CCGTAGTGATGAATCTCTCCTGGCGGGAACAGCAAATATCACCCGGTCGGCAAACAA
ATTCTCGTCCCTGATTTTTACCCACCCCTGACCGCGAATGGTGAGATTGAGAATATAA
CCTTTCATTCCCAGCGGTGCGTGCATAAAAAAATCGAGATAACCGTTGGCCTCAATCGG
CGTTAAACCCGCCACCAGATGGGCATTAAACGAGTATCCCGGCAGCAGGGGATCATTT
TGCGCTTCAGCCATACTTTTCATACTCCCGCCATTGAGAGAAGAAACCAATTGTCCATA
TTGCATCAGACATTGCCGTCCTGCGTCTTTTACTGGCTCTTCTCGCTAACCAAACCGG
TAACCCCGCTTATTAAGCATTCTGTAACAAAGCGGGACCAAAGCCATGACAAAAACG
CGTAACAAAAGTGTCTATAATCACGGCAGAAAAGTCCACATTGATTATTTGCACGGCGT
CACACTTTGCTATGCCATAGCATTTTTATCCATAAGATTAGCGGATTCTACCTGACGCTT

8. Appendix

TTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTTGGGAATTCAAAGATCTTTT
AAGAAGGAGATATACATATGCGTAAAGGCCGAAGAGCTGTTCACTGGTGTGCGTCCCTATT
CTGGTGGAAGCTGGATGGTGTGATGTCAACGGTCATAAGTTTTCCGTGCGTGGCGAGGGTG
AAGGTGACGCAACTAATGGTAAACTGACGCTGAAGTTCATCTGTACTACTGGTAAACTG
CCGGTACCTTGGCCGACTCTGGTAACGACGCTGACTTATGGTGTTCAGTGCTTTGCTC
GTTATCCGGACCATATGAAGCAGCATGACTTCTTCAAGTCCGCCATGCCGGAAGGCTA
TGTGCAGGAACGCACGATTTCTTTAAGGATGACGGCACGTACAAAACGCGTGCGGAA
GTGAAATTTGAAGGCGATACCCTGGTAAACCGCATTGAGCTGAAAGGCATTGACTTTAA
AGAAGACGGCAATATCCTGGGCCATAAGCTGGAATACAATTTTAAACAGCCACAATGTTT
ACATCACCGCCGATAAACAAAAAATGGCATTAAAGCGAATTTTAAAATTCGCCACAAC
GTGGAGGATGGCAGCGTGCAGCTGGCTGATCACTACCAGCAAACACTCCAATCCGGT
GATGGTCCTGTTCTGCTGCCAGACAATCACTATCTGAGCACGCAAAGCGTTCTGTCTAA
AGATCCGAACGAGAAACGCGATCATATGGTTCTGCTGGAGTTCGTAACCGCAGCGGGC
ATCACGCATGGTATGGATGAACTGTACAAAATGTCGGGAGGTGGTGTGATTTCGTGGCC
CCGCAGGGAACAACGATTGCCGCATCTACGTGGGTAACCTTACCTCCAGACATCCGAAC
CAAGGACATTGAGGACGTGTTCTACAAATACGGCGCTATCCGCGACATCGACCTCAAG
AATCGCCGCGGGGACCGCCCTTCGCCTTCGTTGAGTTCGAGGACCCGCGAGACGCG
GAAGACGCGGTGTATGGTCGCGACGGCTATGATTACGATGGGTACCGTCTGCGGGTG
GAGTTTCCTCGAAGCGGCCGTGGAACAGGCCGAGGCCGCGCGGGGGTGGAGGTGG
CGGAGCTCCCCGAGGTGCTATGGCCCCCATCCAGGCGGTCTGAAAACAGAGTGGT
TGTCTCTGGACTGCCTCCAAGTGAAGTTGGCAGGATTTAAAGGATCACATGCGTGAA
GCAGGTGATGTATGTTATGCTGATGTTTACCGAGATGGCACTGGTGTGCTGGAGTTTGT
ACGGAAAGAAGATATGACCTATGCAGTTCGAAAACCTGGATAACACTAAGTTTAGATCTC
ATGAGGGAGAACTGCCTACATCCGGGTTAAAGTTGATTAAGGATCCAACTCGAGTAA
GGATCTCCAGGCATCAAATAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTTCGTTTT
ATCTGTTGTTTGTGCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGG
GCCTTTCTGCGTTTATACCTAGGGCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCA
AAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAG
CAAAGGCCAGCAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCA
TAGGCTCCGCCCCCTGACGAGCATCAAAAATCGACGCTCAAGTCAGAGGTGGCG
AAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGC
TCTCCTGTTCCGACCCTGCCGTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAA
GCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTTCG
CTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTTCAGCCCGACCGCTGCGCCTTATC

8. Appendix

CGGTA ACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCA
GCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGA
AGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCT
GAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACC
GCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGAT
CTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACCTCA
CGTTAAGGGATTTTGGTCATGACTAGTGCTTGGATTCTCACCAATAAAAAACGCCCGGC
GGCAACCGAGCGTTCTGAACAAATCCAGATGGAGTTCTGAGGTCATTACTGGATCTATC
AACAGGAGTCCAAGCGAGCTCGATATCAAATTACGCCCCGCCCTGCCACTCATCGCAG
TACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATG
AACCTGAATCGCCAGCGGCATCAGCACCTTGTGCGCCTTGCGTATAATATTTGCCCATG
GTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGA
AACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCTTTAGGGAAA
TAGGCCAGGTTTTACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCG
GAAATCGTCGTGGTATTCCTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAA
CGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATA
CGAAATCCGGATGAGCATTATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAA
ACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCT
GGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATT
GGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTC
CTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTTATTATGGTGAA
AGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTTCGCCAGATATC

RBP consensus sequence 10mer library plasmid

GACGTCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCG
CTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGG
GAGAGGCGGTTTTCGTATTGGGCGCCAGGGTGGTTTTTTCTTTTACCAGTGAGACGGG
CAACAGCTGATTGCCCTTACCAGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCAC
GCTGGTTTGGCCAGCAGGCGAAAATCCTGTTTGTGGTGGTTAACGGCGGGATATAA
CATGAGCTGTCTTCGGTATCGTTCGTATCCCACTACCGAGATGTCCGCACCAACGCGCA
GCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCA

8. Appendix

GCATCGCAGTGGGAACGATGCCCTCATTGAGCATTGTCATGGTTTGTGAAAACCGGA
CATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGA
TATTTATGCCAGCCAGCCAGACGCGAGACGCGCCGAGACAGAACTTAATGGGCCCCTA
ACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCAGTCGCGTAC
CGTCTTCATGGGAGAAAATAACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAAT
AACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCG
GATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTT
ACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGA
TCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTG
GAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGG
TTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCCGAGA
AACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATA
CTCTGCGACATCGTATAACGTTACTGTTTTACATTCACCACCCTGAATTGACTCTCTTC
CGGGCGCTATCATGCCATACCGCGAAAGTTTTGCGCCATTCGATGGTGTCCGGGATC
TCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCAGTAGTAGTTGAGG
CCGTTGAGCACCGCCGCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGT
CCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCG
AAGTGGCGAGCCCGATCTTCCCATCGGTGATGTGCGCGATATAGGCGCCAGCAACC
GCACCTGTGGCGCCGGTGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATC
GTTTAGGCACCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGAATTGT
GAGCGGATAACAATTTGAGAATTCAAAAGATCTNNNNNNNNNNATATTTAAGAAGGAGA
TATACATATGAGCGAGCTGATTAAGGAGAACATGCACATGAAGCTGTACATGGAGGGC
ACCGTGGACAACCATCACTTCAAGTGCACATCCGAGGGCGAAGGCAAGCCCTACGAG
GGCACCCAGACCATGAGAATCAAGGTGGTTCGAGGGCGGCCCTCTCCCCTTCGCCTTC
GACATCCTGGCTACTAGCTTCCTCTACGGCAGCAAGACCTTCATCAACCACACCCAGG
GCATCCCCGACTTCTTCAAGCAGTCCTTCCCTGAGGGCTTCACATGGGAGAGAGTCAC
CACATACGAAGACGGGGGCGTGCTGACCGCTACCCAGGACACCAGCCTCCAGGACGG
CTGCCTCATCTACAACGTCAAGATCAGAGGGGTGAACTTCACATCCAACGGCCCTGTG
ATGCAGAAGAAAACACTCGGCTGGGAGGCCTTACCCGAGACGCTGTACCCCGCTGAC
GGCGGCCTGGAAGGCAGAAACGACATGGCCCTGAAGCTCGTGGGCGGGAGCCATCT
GATCGCAAACATCAAGACCACATATAGATCCAAGAAACCCGCTAAGAACCTCAAGATGC
CTGGCGTCTACTATGTGGACTACAGACTGGAAAGAATCAAGGAGGCCAACAACGAGAC
CTACGTGAGCAGCACGAGGTGGCAGTGGCCAGATACTGCGACCTCCCTAGCAAACCT
GGGGCACAAGCTTAACTAAGGATCCAACTCGAGTAAGGATCTCCAGGCATCAAATAA

8. Appendix

AACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTTCGGTGAAC
GCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTCTGCGTTTATACCTA
GGGATATATTCCGCTTCCCTCGCTCACTGACTCGCTACGCTCGGTTCGTTGACTGCGGC
GAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCTGGAAGATGCCAGGAAGATAC
TTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCCCCC
TGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTA
TAAAGATAACCAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCTGCCT
TTCGGTTTACCGGTGTCATTCCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGAC
ACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCACGAACCCCCCGTT
CAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAAGAC
ATGCAAAAGCACCCTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAA
GTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTACTGCGCTCCTCCA
AGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAACCGCCCT
GCAAGGCGGTTTTTTCGTTTTTCAGAGCAAGAGATTACGCGCAGACCAAACGATCTCAA
GAAGATCATCTTATTAATCAGATAAAATATTTCTAGATTTTCAGTGCAATTTATCTCTTCAA
ATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTTACTAGTGCTTGGATTCTCA
CCAATAAAAAACGCCCGGCGGCAACCGAGCGTTCTGAACAAATCCAGATGGAGTTCTG
AGGTCATTACTGGATCTATCAACAGGAGTCCAAGCGAGCTCTCGAACCCAGAGTCCC
GCTCAGAAGAACTCGTCAAGAAGGCGATAGAAGGCGATGCGCTGCGAATCGGGAGCG
GCGATACCGTAAAGCACGAGGAAGCGGTCAGCCCATTCCGCCCAAGCTCTTCAGCA
ATATCACGGGTAGCCAACGCTATGTCCTGATAGCGGTCCGCCACACCCAGCCGGCCA
CAGTCGATGAATCCAGAAAAGCGGCCATTTTCCACCATGATATTCGGCAAGCAGGCAT
CGCCATGGGTCACGACGAGATCCTCGCCGTCGGGCATGCGCGCCTTGAGCCTGGCGA
ACAGTTCGGCTGGCGCGAGCCCCTGATGCTCTTCGTCCAGATCATCCTGATCGACAAG
ACCGGCTTCCATCCGAGTACGTGCTCGCTCGATGCGATGTTTCGCTTGGTGGTGAAT
GGGCAGGTAGCCGGATCAAGCGTATGCAGCCGCCGATTGCATCAGCCATGATGGAT
ACTTCTCGGCAGGAGCAAGGTGAGATGACAGGAGATCCTGCCCCGGCACTTCGCC
AATAGCAGCCAGTCCCTTCCCGCTTCAAGTACAAACGTCGAGCACAGCTGCGCAAGGAA
CGCCCGTCGTGGCCAGCCACGATAGCCGCGCTGCCTCGTCCTGCAGTTCATTCAGGG
CACCGGACAGGTTCGGTCTTGACAAAAAGAACCAGGGCGCCCTGCGCTGACAGCCGGA
ACACGGCGGCATCAGAGCAGCCGATTGTCTGTTGTGCCAGTCATAGCCGAATAGCCT
CTCCACCCAAGCGGCCGGAGAACCCTGCGTGCAATCCATCTTGTTCAATCATGCGAAAC
GATCCTCATCCTGTCTCTTGATCAGATCATGATCCCCTGCGCCATCAGATCCTTGGCGG

8. Appendix

CAAGAAAGCCATCCAGTTTACTTTGCAGGGCTTCCCAACCTTACCAGAGGGCGCCCCA
GCTGGCAATTCC

SICLOPPS library plasmids

GACGTCTTAAGACCCACTTTTCACATTTAAGTTGTTTTTCTAATCCGCATATGATCAATTC
AAGGCCGAATAAGAAGGCTGGCTCTGCACCTTGGTGATCAAATAATTTCGATAGCTTGTC
GTAATAATGGCGGCATACTATCAGTAGTAGGTGTTTCCCTTTCTTCTTTAGCGACTTGAT
GCTCTTGATCTTCCAATACGCAACCTAAAGTAAATGCCCCACAGCGCTGAGTGCATAT
AATGCATTCTCTAGTGAAAAACCTTGTTGGCATAAAAAGGCTAATTGATTTTCGAGAGTT
TCATACTGTTTTTCTGTAGGCCGTGTACCTAAATGTACTTTTGTCTCCATCGCGATGACTT
AGTAAAGCACATCTAAAACTTTTAGCGTTATTACGTAAAAAATCTTGCCAGCTTTCCCCT
TCTAAAGGGCAAAAAGTGAGTATGGTGCCTATCTAACATCTCAATGGCTAAGGCGTCTGA
GCAAAGCCCCTTATTTTTTACATGCCAATAAATGTAGGCTGCTCTACACCTAGCTTCT
GGGCGAGTTTACGGGTTGTTAAACCTTCGATTCCGACCTCATTAAAGCAGCTCTAATGCG
CTGTTAATCACTTTACTTTTATCTAATCTAGACATCATTAAATTCCTAATTTTTGTTGACT
CTATCGTTGATAGAGTTATTTTACCACTCCCTATCAGTGATAGAGAAAAGAATTCAAAAG
ATCTTTTAAGAAGGAGATATACATATGGTTAAAGTTATCGGTTCGTCGTTCCCTCGGAGT
GCAAAGAATATTTGATATTGGTCTTCCCAAGACCATAATTTTCTGCTAGCCAATGGGG
CGATCGCCGCCAATTGCNNSNNSNNSNNSNSTGCTTAAGTTTTGGCACCGAAATTTTA
ACCGTTGAGTACGGCCATTGCCATTGGCAAATTTGTGAGTGAAGAAATTAATTGTTT
TGTGTACAGTGTTGATCCAGAAGGGAGAGTTTACACCCAGGCGATCGCCCAATGGCAT
GACCGGGGAGAGCAGGAAGTATTGGAATATGAATTGGAAGATGGTTCAGTAATCCGAG
CTACCTCTGACCACCGCTTTTTAACCACCGATTATCAACTGTTGGCGATCGAAGAAATT
TTTGCTAGGCAACTGGACTTGTTGACTTTAGAAAATATTAAGCAAACCTGAAGAAGCTCTT
GACAACCATCGTCTTCCCTTTCCATTACTTGACGCTGGGACAATTAACGACAAATCC
TGGTGTATCCGCTTGGCAGGTCAACACAGCTTATACTGCGGGACAGTTGGTCACATAT
AACGGCAAGACGTATAAATGTTTGCAGCCCCACACCTCCTTGGCAGGATGGGAACCAT
CCAACGTTCCCTGCCTTGTTGGCAGCTTCAATGACTCGAGTAAGGATCTCCAGGCATCAA
ATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTGTTTTATCTGTTGTTTGTCTGGT
GAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATA
CCTAGGCTACAGCCGATAGTCTGGAACAGCGCACTTACGGGTTGCTGCGCAACCCAAG
TGCTACCGGCGCGGCAGCGTGACCCGTGTCGGCGGCTCCAACGGCTCGCCATCGTCC
AGAAAACACGGCTCATCGGGCATCGGCAGGCGCTGCTGCCCGCGCCGTTCCCATTCC

8. Appendix

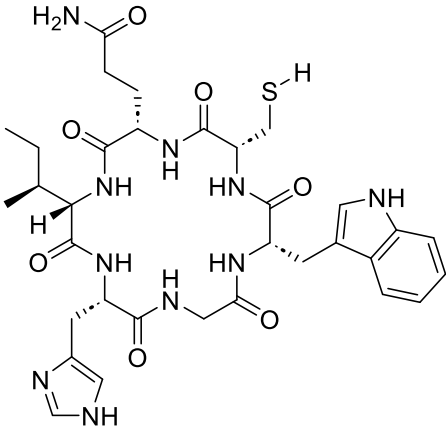
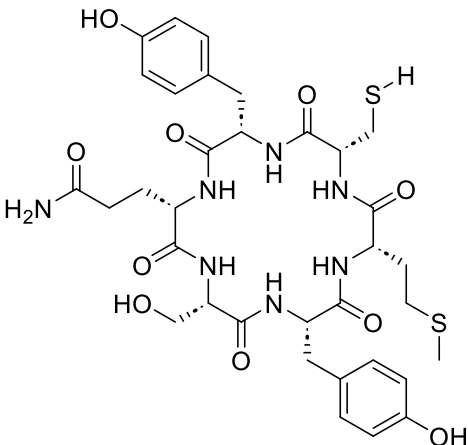
TCCGTTTCGGTCAAGGCTGGCAGGTCTGGTTCCATGCCCGGAATGCCGGGCTGGCTG
GGCGGCTCCTCGCCGGGGCCGGTCGGTAGTTGCTGCTCGCCCGGATACAGGGTCGG
GATGCGGGCGCAGGTCGCCATGCCCAACAGCGATTTCGTCCTGGTCGTGATCAAC
CACCACGGCGGCACTGAACACCGACAGGCGCAACTGGTCGCGGGGCTGGCCCCACG
CCACGCGGTCAATTGACCACGTAGGCCAACACGGTGCCGGGGCCGTTGAGCTTCACGA
CGGAGATCCAGCGCTCGGCCACCAAGTCCTTGACTGCGTATTGGACCGTCCGCAAAG
AACGTCCGATGAGCTTGAAAAGTGTCTTCTGGCTGACCACCACGGCGTTCTGGTGGCC
CATCTGCGCCACGAGGTGATGCAGCAGCATTGCCGCCGTGGGTTTCCTCGCAATAAGC
CCGGCCCACGCCTCATGCGCTTTGCGTTCCGTTTGCACCCAGTGACCGGGCTTGTTCT
TGGCTTGAATGCCGATTTCTCTGGACTGCGTGGCCATGCTTATCTCCATGCGGTAGGG
GTGCCGCACGGTTGCGGCACCATGCGCAATCAGCTGCAACTTTTCGGCAGCGCGACA
ACAATTATGCGTTGCGTAAAAGTGGCAGTCAATTACAGATTTTCTTTAACCTACGCAATG
AGCTATTGCGGGGGGTGCCGCAATGAGCTGTTGCGTACCCCCCTTTTTTAAGTTGTTG
ATTTTTAAGTCTTTCGCATTTGCCCCTATATCTAGTTCTTTGGTGCCCAAAGAAGGGCAC
CCCTGCGGGGTTCCCCACGCCTTCGGCGCGGCTCCCCCTCCGGCAAAAAGTGGCCC
CTCCGGGGCTTGTTGATCGACTGCGCGGCCTTCGGCCTTGCCCAAGGTGGCGCTGCC
CCCTTGGAACCCCCGCACTCGCCGCCGTGAGGCTCGGGGGGCAGGCGGGCGGGCTT
CGCCCTTCGACTGCCCCCACTCGCATAGGCTTGGGTGCTTCCAGGCGCGTCAAGGCC
AAGCCGCTGCGCGGTGCTGCGCGAGCCTTGACCCGCCTTCCACTTGGTGTCCAACC
GGCAAGCGAAGCGCGCAGGCCGCGAGGCCGGAGGCACTAGTGCTTGGATTCTCACCAA
TAAAAACGCCCGGCGGCAACCGAGCGTTCTGAACAAATCCAGATGGAGTTCTGAGGT
CATTACTGGATCTATCAACAGGAGTCCAAGCGAGCTCGTAAACTTGGTCTGACAGTTAC
CAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTT
GCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCA
GTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAA
CCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCAT
CCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGC
GCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGC
TTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCA
AAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTGAGAAGTAAGTTGGCCGCAGT
GTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAA
GATGCTTTTCTGTGACTGGTGAAGTCAACCAAGTCATTCTGAGAATAGTGTATGCGG
CGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAA
CTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACCTCAAGGATCTTA

8. Appendix

CCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAAGTATCTTCAGCATC
 TTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAA
 AGGGAATAAGGGCGACACGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATT
 GAAGCATTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAA
 ATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCT

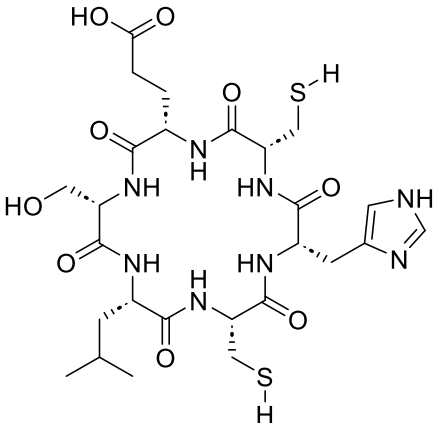
8.4.2. SICLOPPS peptides structures, HRMS analyses and yields.

Supplementary table 8: SICLOPPS peptide structures, masses (calculated and determined) and obtained yields.

Peptide name and structure	m/z calculated [M+H] ⁺	m/z found [M+H] ⁺	Yield
SICLOPPS peptide 6 	725.3149	725.3160	1.75 %
SICLOPPS peptide 44 	776.2703	776.2756	2.31 %

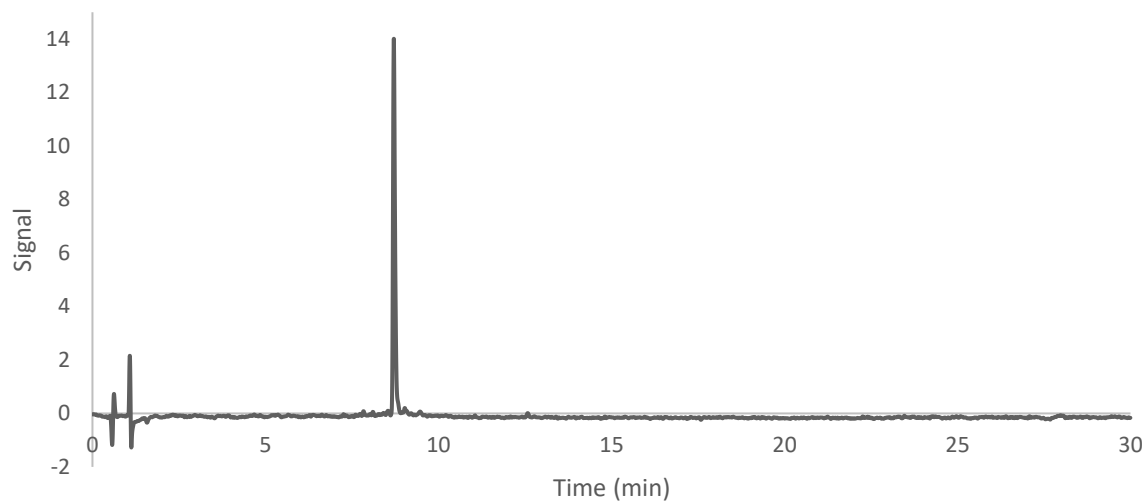
8. Appendix

SICLOPPS peptide 50	673.2393	673.2435	0.65 %
---------------------	----------	----------	--------

CC(C)C[C@@H](C(=O)N[C@@H](CO)C(=O)N[C@@H](CC(=O)O)C(=O)N[C@@H](CS)C(=O)N[C@@H](Cc1c[nH]cn1)C(=O)O

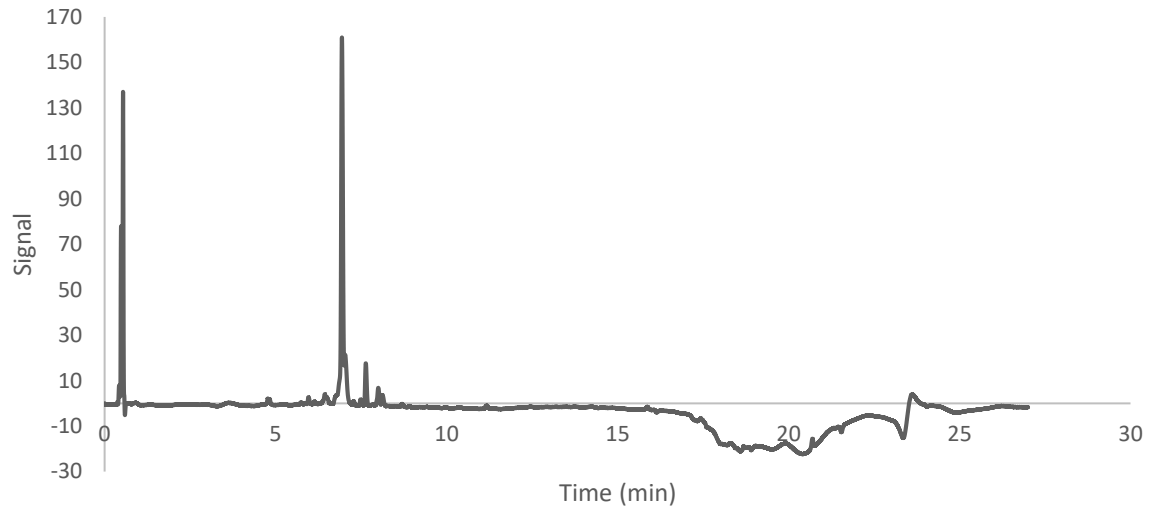
8.4.2. Peptide Purity HPLC Chromatograms

SICLOPPS peptide 6



8. Appendix

SICLOPPS peptide 44



SICLOPPS peptide 50

