DEVELOPMENT OF SINGLE-MOLECULE METHODS FOR RNA SPLICING

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Robert Weinmeister

Department of Biochemistry

University of Leicester

July, 2014

Abstract

Development of Single-Molecule Methods for RNA Splicing

-

Robert Weinmeister

RNA splicing is an important step in the synthesis of most mammalian proteins and understanding the underlying molecular mechanisms is critical for tackling diseases linked to splicing. An important determinant is SRSF1, which has multiple roles in constitutive and alternative splicing. One of these roles is in the recognition and selection of 5′ splice sites due to an interaction with U1 snRNP during formation of complexes E and A. The exact roles and modes of interaction are not clear. Single-molecule methods are a key to understanding these.

In this work, single-molecule methods were developed to investigate the number of bound proteins under different conditions. A home-built microscope utilising objective-based illumination by total internal reflection was used to look at these interactions at a single-molecule level and investigate the number of bound proteins under different conditions.

The results showed that there was a distinctive reduction in the number of bound proteins in complex A, dependent on the availability of ATP. This was linked to the number of functional 5′ splice sites present, the U1 snRNP and phosphorylation. We could not find any evidence that sequences known to mediate stimulation by SRSF1 affect its binding.

Using total internal reflection has inherent limitations, among them the necessary surface attachment and the dilutions required. These limitations could be overcome by the use of isolated microenvironments in the form of tiny droplets. A robust and convenient microfluidic device with a feature size of 3 μm was set up and a suitable surfactant for biological samples was identified. Droplets with a diameter of 1 μm were generated for the first time using flow focussing and single quantum dots and fluorescent proteins where identified within these droplets. The fluorescence intensity time traces from these droplets enabled the number of encapsulated fluorescent particles to be measured.

Acknowledgements

I had the great fortune of having not only one, but two supervisors, Prof. I. C. Eperon and Dr. A. J. Hudson. I would like to thank them for the opportunity to pursue such an interesting and multifaceted PhD project. I am deeply grateful for the patience they showed and the guidance they provided. I would also like to thank my committee members Dr. C. Dominguez and Dr. P. B. Howes for the helpful discussions.

I would like to thank everyone from the labs 3/19 and 3/13 for the friendly atmosphere that made my work there such a pleasant experience. A special thanks goes to Rachel Dumas, Christian Lucas, Paul Smith, Jesvin Samuel, Nicola Portolano, Mark Hodson, Lindsay Smith, Carika Weldon, Diana Marouco, Oksana Gonchar, Li Chen, Miran Rada and Nilesh Mistry for the support, the invaluable discussions and of course the enjoyment they provided.

I also would like to thank everyone else, who I could not name here solely for reasons of space, who accompanied me along the way to make life even more interesting and exciting. I would like to thank the university volleyball club for providing a friendly and much need athletic diversion and I would also like to thank England for the amazing weather and food.

I would like to thank Prof. A. Griffiths and his group for the introduction to microfluidics and for providing an initial mask, wafer and design for testing. I would like to thanks Mr. T. Fulham, Ms. R. N. Tshulu and Dr. A. M. Stuart for synthesising and testing the surfactants. I would like to thank Dr L. P. Eperon, Dr M. J. Hodson, Mr. T. Webber and Ms. B. Hutchinson for providing the various plasmids, Dr. L. P. Eperon and Mr. C. M. Lucas for preparing the nuclear extracts containing said plasmids and Dr. M. J. Hodson and Mr. B. Song for providing several useful pre-mRNA substrates.

Lastly, I would like to thank the University of Leicester for providing the funding for this PhD.

Mein besonderer Dank gilt meiner Familie, welche mich all die Jahre hinweg wundervoll unterstützte, Interesse an meiner Arbeit zeigte, einen erholsamen Rückzugsort für immer wieder benötigte Entspannung bot und wovon 83 % mich sogar persönlich besuchen kamen.

Abbreviations

A	Adenine
AFM	Atomic force microscopy
AMPS	Ammonium persulfate
ATP	Adenosine triphosphate
bp	Base pair(s)
BSA	Bovine serum albumin
C	Cytosine
CCD	Charge-coupled device
CrPi	Creatine phosphate
Cy5	Cyanine 5
DAPI	4′,6-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid

EMCCD	Electron-multiplying charge-coupled device
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
EtBr	Ethidium bromide
FCS	Fluorescence correlation spectroscopy
FRAP	Fluorescence recovery after photobleaching
FRET	Förster resonance energy transfer
G	Guanine
GFP	Green fluorescent protein
HEPES	N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid
HLB	Hydrophile-lipophile balance
hnRNP	Heterogeneous nuclear ribonucleoproteins
ISE	Intronic splicing enhancer
ISS	Intronic splicing silencer
KGlu	Potassium glutamate
LED	Light-emitting diode
mRNA	Messenger RNA
Nonidet P-40	Octylphenoxypolyethoxyethanol
NP-40	Octylphenoxypolyethoxyethanol
nt	Nucleotide(s)

oligo(s)   oligonucleotide(s)

PCA   Protocatechuic acid

PCD   Protocatechuate-3,4-dioxygenase

PCR   Polymerase chain reaction

PDMS   Polydimethylsiloxane

PEG   Polyethylene glycol

Pre-mRNA   Precursor messenger RNA

PTB   Polypyrimidine tract-binding protein

QE   Quantum efficiency

R   Arginine

rATP   ribonucleotide adenosine triphosphate

Re   Reynolds number

RNA   Ribonucleic acid

RRM   RNA recognition motif

RS domain   Arginine/Serine-rich domain

S   Serine

SDS   Sodium dodecyl sulfate

SF1   Splicing factor 1

smFRET   Single molecule FRET

snRNP   Small nuclear ribonucleic particle

SRPK1	Serine/threonine-protein kinase 1
SRSF1	Serine/arginine-rich splicing factor 1
T	Thymine
TEMED	N,N,N′,N′-Tetramethylethylenediamine
TIR	Total internal reflection
TIRF	Total internal reflection fluorescence
TIRFM	Total internal reflection fluorescence microscopy
Tris	Tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
U	Uracil
U2AF	U2 auxillary factor
w/v	Weight per volume

Contents

List of Figures

List of Tables

Contributors Statement

It is in the nature of modern multi-disciplinary research that many people are involved in the success of ones project. The work presented in this thesis and all data shown therein is my own work if not explicitly stated otherwise. This includes the setup of the microscope, the data analysis, the single-molecule experiments and the underlying laboratory work, and the microfluidic work.

Exceptions are the preparation of the plasmids for the fluorescent proteins, the nuclear extracts except for the SRSF1-GFP nuclear extract, the initial pre-mRNA constructs and the initial mask, wafer, and design for the microfluidic work and the synthesis and partial testing of the surfactants.

Chapter 1

Introduction

1.1 Pre-mRNA Splicing and the Spliceosomal Machinery

Machines take me by surprise with great frequency

Alan Turing

𝕿he much simplified view of eukaryotic protein biosynthesis is that the genetic information encoded in the sequence of the deoxyribonucleic acid (DNA), stored in the nucleus of the cell, is transcribed first into a temporary intermediate, messenger ribonucleic acid (mRNA). The mRNA gets transported from the nucleus into the cytoplasm where it is translated at the ribosome into proteins.

Ever since the start of the Human Genome Project, the number of human genes was repeatedly revised down (Pertea and Salzberg, 2010) and was lately put at 20 687 (The ENCODE Project Consortium, 2012). This puts it well below that of other species like rice with a current estimate of 37 544 genes (International Rice Genome Sequencing Project, 2005). These numbers show little relation to perceived measures of complexity, leaving the question of how the genetic information, gene number and size of the genome actually relate to the organism.

The answer to this point lies in the discovery of alternative RNA splicing in 1977 (Berget et al., 1977; Chow et al., 1977). It put an end to the original assumption of one gene encoding one protein. Genes contain non-coding parts that get removed or spliced out, the introns, leaving only parts of the original sequence present in the final RNA, named exons. Intrinsic to this is the possibility of alternative splicing, a mechanism that allows the use of different or altered coding parts, resulting in a wide variety of possible transcripts and therefore proteins from one apparent gene. In fact, this is one of the outstanding differences be-

tween species. It was found that there are significant differences in alternative splicing complexity for different species, with the highest complexity found in primates (Barbosa-Morais et al., 2012).

There are on average 6.3 alternatively spliced transcripts found for each human gene (The ENCODE Project Consortium, 2012). It was found that there are on average 8.8 exons per gene with a mean length for internal exons of 145 bp, much less than the average length for introns of 3365 bp (International Human Genome Sequencing Consortium, 2001). The number of possible transcripts per gene as well as the length discrepancy between the actual protein encoding exons and the introns hint at an exceptional mechanism at the heart of this.

1.1.1 Pre-mRNA Splicing Mechanism

The initial transcription of many genes in eukaryotes results in pre-mRNA containing exons intertwined with introns. An exon is defined as that sequence of the gene that remains within the spliced RNA and leaves the nucleus, whereas an intron is the sequence between exons that gets removed. Thus the final or mature mRNA contains only exons. As pre-mRNA transcripts start and end with an exon, there is always one intron less than the number of exons per gene. Pre-mRNA splicing is then the process that removes the introns and joins the remaining exons together, resulting in mature mRNA.

Two groups of introns exist that can splice themselves and are removed by RNA catalysis. These Group I (Cech, 1990) and Group II (Bonen and Vogel, 2001) introns are found in chloroplasts and mitochondria in plants, fungi and protozoa and in mRNA of bacteria. In contrast, chromosomal genes generally cannot splice themselves. For them splicing is done by a macromolecular machine named the spliceosome, which assembles in an ordered process onto the pre-mRNA and removes the introns in a coordinated chain of RNA-RNA, RNA-protein and protein-

protein interactions.

These introns can further be classified into two categories depending on their splice site sequences: U2-type introns and U12-type introns. U2-type introns represent, at 99.8 %, the majority of existing introns. 99.3 % of these start with the sequence GU and end with the sequence AG (Sheth et al., 2006). All of them are spliced by the major spliceosome. In contrast, the rare U12-type introns are spliced by the similar but slightly different minor spliceosome (Turunen et al., 2013).

The major spliceosome involves at its core five small nuclear ribonucleoprotein particles, namely U1, U2, U4, U5 and U6, and 170 other associated proteins (Wahl et al., 2009). The minor spliceosome contains snRNPs (U11, U12, U4atac and U6atac) that are slightly different but functionally analogous to the snRNPs found in the major spliceosome. Both share the U5 snRNP (Patel and Steitz, 2003). Both the major and the minor spliceosome are confined to the nucleus (Steitz et al., 2008).

Splicing requires four sequences (Figure 1.1a). The junctions of introns and exons are conserved sequences called splice sites and denoted as 5′ splice site and 3′ splice site respectively as seen from the intron. Their consensus sequences are G|GURAGU and YAG| (where A is Alanine, C is Cytosine, G is Guanine, U is Uracil, R is either A or G and Y is either C or T, with the vertical bar denoting the junction of exon and intron, Sheth et al. (2006)). Directly upstream of the 3′ splice site lies the polypyrimidine tract with 10 to 12 pyrimidines. Between this tract and the 5′ splice site is the branch point with a sequence around the important adenine that is weakly conserved but still susceptible to single mutations (Reed and Maniatis, 1988).

Splicing at its heart is characterised by two energy-neutral chemical steps: displacement of the 5′ exon by the branchpoint, and displacement of the intron at the

5′ Splice Site   Polypyrimidine tract   3′ Exon

5′ Exon   Branch Point   3′ Splice Site

G GURAGU — (YNYURAY) —— Y10-12 —— YAG

(a) Pre-m RNA with consensus sequences

(b) Lariat formation

(c) Exon ligation

Figure 1.1 – Pre-mRNA Splicing. **(a)** Pre-mRNA with an intron flanked by a 5′ exon and a 3′ exon. Four specific sites with their consensus sequences are shown. Absolutely required nucleotides are underlined. The 5′ splice site marks the 5′ end of the intron with GU. It is followed by the branch point sequence downstream. The branch point has a weak consensus sequence besides the required adenine. The polypyrimidine tract with 10 to 12 pyrimidines is directly upstream of the 3′ splice site where AG are the 3′ end of the intron. R is a purine, Y a pyrimidine. - **(b)** Displacement of the 5′ exon by the branch point adenine leads to the formation of the intron lariat. - **(c)** Displacement of the intron leads to the joined exons as mature RNA with the remaining intron lariat.

3' splice site by the 5' exon. Via a transesterification reaction the 2'-hydroxyl of the adenosine at the branch point attacks the phosphodiester upstream of the guanine and uracil at the 5' splice site, displacing the exon and attaching the guanine to itself. This results in the free 5' exon and the intron lariat still attached to the 3' exon (Figure 1.1b).

Subsequently, via another transesterification reaction, the 3'-hydroxyl of the 5' exon now attacks the phosphodiester bond following the adenine and guanine at the 3' splice site, separating the exon on this side from the intron and joining the 5' splice site and 3' splice site of the respective exons together (Figure 1.1b). The intron lariat is now fully removed, leaving only the exons remaining in the mature mRNA.

The outstanding feature of this is not the underlying chemistry. The longest human gene, dystrophin, has 79 exons with a combined length of 14 kbp. They are 'hidden' within a 2300 kbp transcript that is cotranscriptionally spliced during its required 16 h of transcription (Tennyson et al., 1995). Thus the challenges are the recognition of the correct splice sites and bringing them close enough together for the two catalytic steps. Despite these enormous tasks, the whole process is highly efficient. This is achieved by the highly dynamic macro-molecular machinery consisting of a large number of subunits, the spliceosome.

1.1.2 The Spliceosome and its Assembly

The spliceosome is by some considered "the most complex macromolecular machine in the cell" (Nilsen, 2003). It has a molecular mass of ~2.7 MDa, is assembled from five snRNPs and protein complexes comprising around 170 proteins (Wahl et al., 2009). Each snRNP in turn is a complex of a small nuclear RNA and further factors. The whole spliceosome has to be assembled before it has its splicing activity. This is done anew for each interaction with the pre-mRNA as afterwards it is
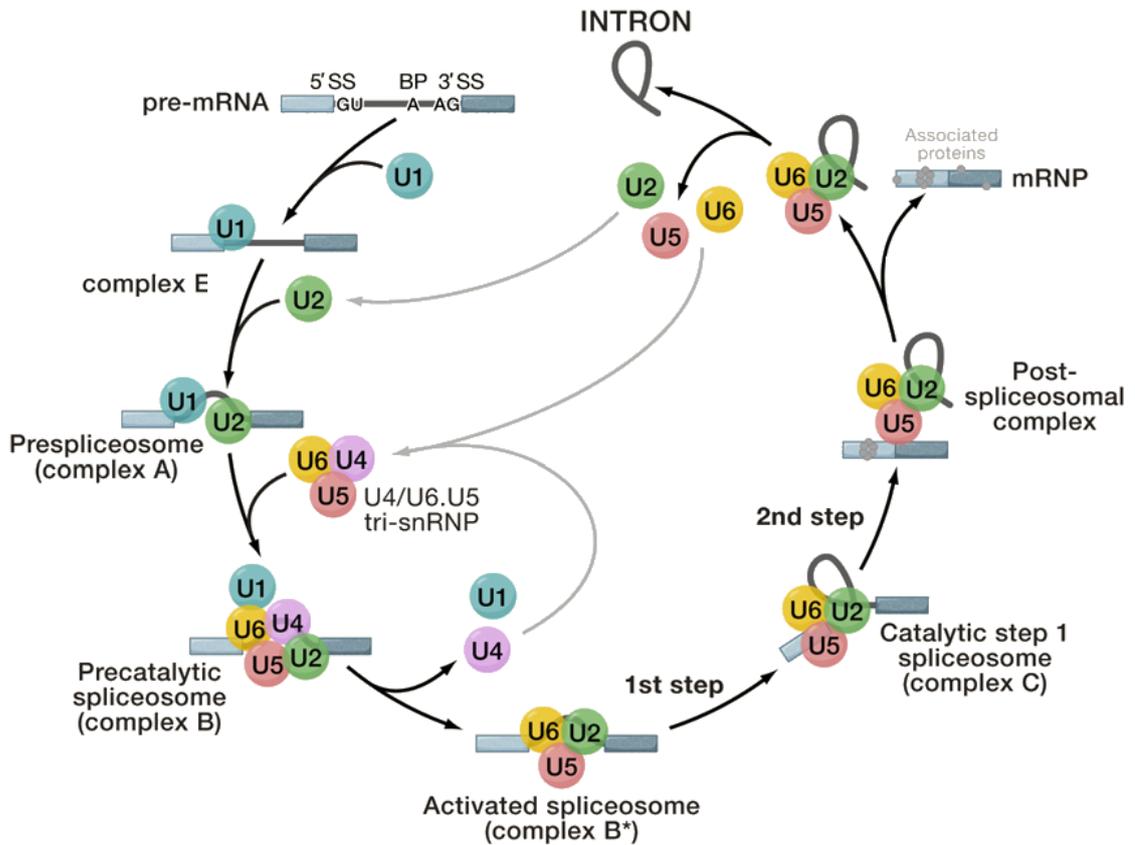
Figure 1.2 – Cross-intron assembly and disassembly of the spliceosome. Spliceosomal complexes form in an ordered way with the two catalytic reactions occurring before and after complex C. This process removes the exon-flanked intron and joins the exons together. Afterwards the spliceosome is disassembled into its building blocks. Modified from Wahl et al. (2009).

recycled to be re-assembled again. The assembly and disassembly of the spliceosome is a highly dynamic process. It occurs in sequential steps that are themselves reversible with possible dead ends (Hoskins and Moore, 2012).

The assembly and splicing process is divided into distinct stages, named as complexes, starting with the complex E, followed by complex A, B, B* (or B^act[1]), C and a post-spliceosomal complex (Figure 1.2). The two important catalytic steps happen before and after complex C. Although these complexes are fixed, the pathways leading to them can differ greatly from each other and there is no underlying assembly mechanism as the recruitment of factors is through the

---

[1]sometimes considered separate complexes

random process of free diffusion (Rino et al., 2007).

It is assumed that the first complex that forms is complex E. It was only found to form in an ATP independent way (Michaud and Reed, 1991) and is characterised by the presence of the U1 snRNP base paired to the GU sequence at the 5′ splice site downstream of the intron through its 5′ terminus (Krämer et al., 1984). This binding is further stabilised by serine-arginine-rich (SR) proteins and proteins of the U1 snRNP. Further interactions are the binding of the splicing factor 1 (SF1) to the branch point and binding of the U2 auxillary factor (U2AF) to both the polypyrimidine tract and the AG sequence at the 3′ splice site. Both these factors interact cooperatively with each other. With these interactions at both the 3′ and the 5′ splice site, the initial recognition of the intron is arranged.

If ATP is available, complex A forms after complex E. It is characterised by the binding of the U2 snRNP to the branch site (Wu and Manley, 1989), which is nonetheless already required for the formation of complex E (Dönmez et al., 2004). It displaces SF1 from the branch point. This binding of the U2 snRNP requires ATP and it is stabilised by other proteins, most notably U2AF65 (Valcárcel et al., 1996).

Formed next is complex B, by the addition of the pre-assembled U4/U6.U5 tri-snRNP. The spliceosome now contains all the necessary components but is not yet active as it requires major rearrangements. The binding of the U1 snRNP to the exon at the 5′ splice site is replaced by base pairing between the 5′ splice site and the U6 snRNP (Staley and Guthrie, 1999). The U1 and U4 snRNP are ultimately expunged. With these rearrangements and further changes in the protein composition, complex B becomes active as the complex B* (Bessonov et al., 2010). Now the first catalytic steps occurs, linking the 5′ end of the intron to the branch point, forming the intron lariat, changing over to complex C. The second catalytic step removes the intron completely and joins the exons, with a pre-spliceosomal

complex remaining. Both intron lariat and exons are subsequently released from the spliceosome which dissociates and is recycled.

1.1.3 5' Splice Site Selection

One important step in identifying the introns lying between the exons is the selection of the 5' splice site. As the average intron is much longer than the average exon (International Human Genome Sequencing Consortium, 2001) and a multitude of 5' splice site sequences exist, a plenitude of possible 5' splice site is present. Roca et al. (2012) found more than 9000 unique human 5' splice site sequences for a 9 nt motif, 3.5 % of all possible sequences, that can be recognised. Excluding the fixed GU is this more than half (58 %) of all possible sequences for the remaining 7 nt. With not only one sequence that indicates a splice site, but an overwhelmingly wide array of them, a daunting task presents itself.

The exon downstream of the intron is displaced from the guanine and uracil that are base paired to the required U1 snRNP. This binding of the U1 snRNP itself plays an important role in selecting the correct splice site. Zhuang and Weiner (1986) demonstrated this with mutations in the RNA part of the U1 snRNP responsible for sequence recognition, leading to different recognised splice site sequences. A highly suitable 5' splice site sequence that provides a high affinity for the U1 snRNP helps in selecting this 5' splice site. Splice sites with such a consensus sequence are termed strong splice sites in comparison to weak splice sites with lower affinity sequences. Accordingly, it was shown that strong splice sites are favoured over weaker splice sites (Eperon et al., 1986; Lear et al., 1990), showing a relationship between the predicted U1 snRNP binding and 5' splice site usage. This holds true for the extreme cases, but breaks down for splice sites with an intermediate affinity between those found in strong and weak splice sites (Lear et al., 1990; Roca et al., 2005).

Despite this, usage of splice sites and their binding by U1 snRNPs does not always correlate (Eperon et al., 1993; Nelson and Green, 1990) and so it is not always the strongest site that gets used (Shapiro and Senapathy, 1987). The relative position of potential 5′ splice sites plays a role, too. Downstream splice sites are generally preferred when strong sites are in competition (Eperon et al., 1993; Reed and Maniatis, 1986). Even inserting a stronger competing 5′ splice site upstream can stimulate usage of the downstream splice site (Hicks et al., 2010), showing a preference for intron definition over exon definition.

The primary sequence of the RNA is not the only factor, its secondary structure can also influence splicing (Buratti and Baralle, 2004). Splice sites themselves must be single stranded (Eperon et al., 1988), it must be possible to bring splicing-relevant sites into close contact and enhancer or silencer sequences show different efficiency if single-stranded (Muro et al., 1999). Instead of modulating splicing directly, it can be done indirectly by changing the secondary structure (McManus and Graveley, 2011; Warf and Berglund, 2010).

While recognition of the correct splice site can also be affected by many other different properties like intron length (Bell et al., 1998) or polymerase processivity (Nogués et al., 2003), the other main contribution arises from the involvement of other sequences besides the ones at the 5′ splice site.

1.1.4 Splicing Enhancers

The 5′ splice site, 3′ splice site, branch point and polypyrimidine tract are all required parts for splicing. Their respective sequences vary considerably in nature. This could lead to unwanted alternative splicing events, inefficient splicing or no splicing at all. It was found that the information within the 5′ splice site sequence is not enough to accurately detect them. Further information is necessary and has to contribute to accurately determine them (Lim and Burge, 2001). This

information can be found in additional sequences.

Often referred to as enhancers, these binding sites are named exonic splicing enhancers (ESE) when inside an exon, or intronic splicing enhancer (ISE) when inside an intron. Other factors can be present that do not only enhance splicing, but suppress it instead. In that case, they are named silencers with the sequences they bind to then termed exonic splicing silencer (ESS) and intronic splicing silencer (ISS) respectively. New research seems to indicate that this classification is not fully applicable if just looking at the sequence. The same sequence can have opposite effects if placed in an intron instead of in an exon (Erkelenz et al., 2013).

These elements primarily function by providing binding sites for regulatory factors that in turn recruit or suppress association of parts of the splicing complex. Enhancer motif sequences can affect the U1 snRNP and the selection of the 5' splice site through factors that bind to them. Eperon et al. (1993) showed that SR proteins like SRSF1 can fulfil this role. Early studies suggested that SRSF1 interacts via its RS domain with the RS motif of U1-70K which is a part of the U1 snRNP. This protein-protein interaction enhances the binding of the U1 snRNP to the pre-mRNA (Kohtz et al., 1994; Wu and Maniatis, 1993). Cho et al. (2011a) later showed that it is actually the RRM of SRSF1 that bridges the RRM of U1-70K to the pre-mRNA and the RS domain of SRSF1 has a modulatory role. The SR proteins influence the early spliceosome assembly. They usually promote exon inclusion (Côté et al., 1999) and can shift the usage of 5' splice sites downstream (Eperon et al., 2000; Krainer et al., 1990b) or upstream (Zhang and Wu, 1996). Due to its importance, considerable effort was made to identify all possible ESE binding sites for SRSF1 (Smith et al., 2006). Other proteins enhancing the stability of the U1 snRNP are RBM25 (Zhou et al., 2008) and TIA-1 (Förch et al., 2002).

On the opposite site, silencer motifs are bound by proteins of the hnRNP family

and can inhibit splicing. hnRNP A1 was shown to antagonise SRSF1 (Mayeda et al., 1993) and so usually promotes exon skipping (Blanchette and Chabot, 1999). It is thought the antagonising effect is through reciprocal blocking of binding sites (Eperon et al., 2000; Zhu et al., 2000). New research points at more complicated underlying patterns as SRSF1 too can promote exon skipping, especially when bound to downstream exons (Pandit et al., 2013).

1.1.5 Serine/arginine-rich Splicing Factor 1 and its Role in 5' Splice Site Selection

The serine/arginine-rich splicing factor 1 (SRSF1, also known as ASF, SF2 and SFSR1) is a human protein encoded by the SRSF1 gene located on chromosome 17 (Bermingham et al., 1995). It is part of the conserved family of SR proteins that encompasses nine, recently conveniently renamed, proteins: SRSF1 to SRSF7, SRSF9, SRSF11. They all share the same structural features of one or two RNA recognition motif(RRM)-type RNA binding domains and a domain rich in arginine and serine dipeptides, the RS domain. The only exception is SRSF7 with a zinc-finger instead of a second RRM (Shepard and Hertel, 2009).

The protein SRSF1 is composed of 248 amino acids with a molecular weight of 28 kDa (*UniProt*). It has a two RRM domains and one RS domain at its C-terminal (Section 8.4). To date, only parts of the structure of SRSF1 are known (Ngo et al., 2008; Tintaru et al., 2007).

SRSF1 is involved in a surprising number of pathways, even referred to as one of the "master regulators of gene expression" (Long and Caceres, 2009). It stimulates transcriptional elongation (Zhong et al., 2009), is involved in microRNA processing (Wu et al., 2010), is present after splicing in the exon junction complex (Singh et al., 2012) and has further roles post-transcriptional (Huang and Steitz, 2005) and even in protein sumoylation (Pelisch et al., 2010). This all besides

its part in consensus and alternative splicing. It leaves one wondering if there is anything in which it is not involved.

SRSF1 plays an extensive role in the regulation of splicing. It was shown that for some pre-mRNA substrates efficient splicing could be recovered by adding SRSF1 to the otherwise splicing insufficient cytoplasmic extract S100, proving its role in constitutive splicing (Krainer et al., 1990a). It participates in alternative splicing (Ge and Manley, 1990), often by binding to exonic splicing enhancers (Sun et al., 1993) and in an antagonistic mechanism together with hnRNP A1 (Cáceres et al., 1994). Functional SELEX screening provided ways to find motifs that can act as ESEs for SRSF1 (Smith et al., 2006).

SRSF1 was usually seen as an enhancer of splicing as it promotes the inclusion of weak exons (Côté et al., 1999), but this might be due to its usual exonic binding and it can have antagonistic effects if binding to introns (Erkelenz et al., 2013; Pandit et al., 2013). If multiple splice sites are present it can shift the usage of 5′ splice sites downstream (Eperon et al., 2000; Krainer et al., 1990b) or upstream (Zhang and Wu, 1996).

It works mainly by influencing the U1 snRNP and increasing its binding affinity (Eperon et al., 1993, 2000). This is either via its RS domain with the RS motif of U1-70K (Wu and Maniatis, 1993) or via its RRM (Cho et al., 2011a). This is not a one-way relationship as the U1 snRNP can complementarily increase the binding affinity of SRSF1 to form a complex together (Jamison et al., 1995; Kohtz et al., 1994). Both seem to be required for the formation of the complex E (Cho et al., 2011a).

SRSF1 is heavily regulated by phosphorylation. Splicing factors are recruited from nuclear speckles to assist in splicing (Misteli et al., 1997) and this recruitment requires serine phosphorylation (Misteli et al., 1998). For SRSF1 this is done by phosphorylation of the RS domain by SR protein kinase 1 (SRPK1) and Clk/Sty

kinases. Hypophosphorylation of the N-terminal part of the RS domain by SRPK1 leads to deposition in nuclear speckles. It is released from these if it is subsequently hyperphosphorylated by Clk/Sty kinases at its C-terminal part of the RS domain (Ngo et al., 2005). SRSF1 has to be phosphorylated to participate in consensus splicing as this is required for the advancement of complex A to complex B (Roscigno and Garcia-Blanco, 1995). This initial phosphorylation also increases its interaction with the U1 snRNP (Xiao and Manley, 1997). It has to be dephosphorylated again before the first catalytic step (Cao et al., 1997). SRSF1 does not need to be dephosphorylated to act as a splicing enhancer (Xiao and Manley, 1998).

Removal of the RS domain of SRSF1 prevents mostly its recovery of constitutive splicing in S100 extract but still allows alternative splicing (Cáceres and Krainer, 1993; Zuo and Manley, 1993). Sometimes it is not required, depending on the polyprimidine tract (Zhu, 2000). Astonishingly, the deletion of SRSF1's first 11 amino acids restores function for these mutants (Shaw et al., 2007). Its RRMs can recognise the 5' splice site (Zuo and Manley, 1994) and removal of the second RRM inactivates it (Cáceres and Krainer, 1993). Mutating only the amino acids at position 162 to 163 in the second RRM does not seem to affect binding greatly but the ability to restore splicing in an S100 extract was lost (Zuo and Manley, 1993).

There is no consensus about the identity of the optimal binding site for SRSF1. Repeated rounds of SELEX with purified versions of SRSF1 lacking the RS domain found the motifs RGAAGAAC, AGGACAGAGC and AGGACGAAGC, where R is a purine (Tacke and Manley, 1995). Functional SELEX, in which repeated rounds of selection were used to identify enhancer sequences that respond best to SRSF1, produced the possible sequence SRSASGA (Liu et al., 1998) and CSScSSR and CSMMSSA, where S is either C or G and M is either A or C, with those binding sites found to be enriched within exons, usually within 200 nt of a 5' splice sites and with a peak occurrence at 20 nt to 40 nt (Smith et al., 2006). The CLIP method

identified UGRWG, where W is either A or T (Sanford et al., 2008). Using a computational model, Wang et al. (2011) predicted that the optimal binding motifs are AGAAGA, AAGAAG or GAAGAA. Cho et al. (2011b) found the decuple sequences of GGCGGAGGAA and GCGGAGGAAG. Cléry et al. (2013) measured a dissociation constant of $0.8\,\mu\text{M}$ for a complex of the sequence UGAAGGAC with the RRM2 domain of SRSF1. RNAcompete experiments found the sequence of GGAGGA (Ray et al., 2013).

These found binding sequences differ greatly between each other. A divide seems to be between sequences found based on the binding of SRSF1 and those found based on function of SRSF1. Indeed, Pandit et al. (2013) found only a weak correlation of binding sites for both SRSF1 and SRSF2 and the observed effects on splicing, but a strong correlation between the binding of SR proteins and intron length. Similarly, Dreumont et al. (2010) identified an inverse correlation between the binding strength of SRSF1 to an enhancer and the enhancer's actual activity. This suggests that the majority of enhancers favours a transient binding of SRSF1 for their activity, with SRSF1 binding more strongly to other sequences (Eperon, personal communication).

1.1.6 Alternative Splicing, Splicing Regulation and Linked Diseases

One important feature of pre-mRNA splicing done by the spliceosome is the capability to produce alternative patterns of splicing from a single pre-mRNA and possibly leading to multiple functional mRNAs and even proteins. Patterns where existing splice sites are used exclusively are considered as constitutive splicing. If there are suppressed or weak splice sites present, different splicing patterns can emerge as alternative splicing. Recent surveys estimate that human pre-mRNAs with multiple exons show alternative splicing in around 95 % of all

cases (Pan et al., 2008).

Alternative splicing can roughly be categorised into four different modes. A whole exon might be skipped, an intron might not be excluded, a different 5′ splice site might be used defining a longer or shorter exon and respectively the same for a 3′ splice site. Crucial to this is the identification of the intron and its defining 3′ and 5′ splice site. Failure to find the correct 5′ splice site leads to alternatively spliced RNA and with it to either nonsense RNA, non-functioning proteins or proteins with a different and sometimes antagonistic function.

Modulation of the splicing pattern in one particular gene can occur by direct mutations of the splice sites but also by mutation in other regulatory sequences. When investigating these disease-causing mutations, the change in the splicing code is often ignored although it is estimated that up to 62 % of these mutations operate by disrupting splicing (López-Bigas et al., 2005). Mutations can also affect the function and assembly of the spliceosome, as for spinal muscular atrophy and retinitis pigmentosa (Wang and Cooper, 2007). For spinal muscular atrophy, a possible enhancer is converted to a silencer, consequently inhibiting the spliceosome assembly (Martins de Araújo et al., 2009).

For tackling these diseases at the root, the possibility to shift splicing accurately and nudge it back on its 'healthy' pathway has great appeal. 5′ splice site sequences vary widely and many deviate far from the consensus sequence and therefore rely on auxiliary sequences for their selection in the face of competitive rivals. These additional contributions can come from regulatory sequences such as exonic splicing enhancers (Section 1.1.4), which greatly influence the actual splicing pattern. Several therapeutic approaches are therefore taken to influence the splicing result; one direct way is with antisense oligonucleotides that can anneal to the enhancer and silencer sequences to block splicing factors from accessing them (Hua et al., 2007), and another is with bifunctional oligonucleotides

that combine a targeting sequence with a functional tail containing an ESE that recruits SR proteins like SRSF1 to increase exon inclusion (Owen et al., 2011).

SRSF1 plays an important part in alternative splicing and it is involved in the modulation of the expression of oncogenes and therefore seen as an proto-oncogene (Karni et al., 2007). It is involved at each stage of splicing (Chew et al., 1999; Shen and Green, 2006; Shen and Green, 2007). SRSF1 seems to have roles in splice site selection that sometime involve clearly defined enhancer, but that is not always the case. It is part of the splicing mechanism in complex B and C where it has to be dephosphorylated and for the second catalytic step in combination with enhancer sequences. A common underlying mechanism that is just mediated by the binding or interaction of SRSF1 with the U1 snRNP seems improbable, especially if SRSF1 can replace the U1 snRNP under certain conditions (Crispino and Sharp, 1995; Tarn and Steitz, 1994). Thus it is not really clear how SRSF1 interacts in the spliceosome assembly, what the roles and functions of its domains are and most of the underlying interactions still elude us.

A deeper understanding of SRSF1 would greatly help not only in consensus splicing but also in alternative splicing and its regulation. Single-molecule methods might provide the right tools to reveal the role played by SRSF1 in splicing.

1.2 Single-Molecule Studies

> If an alien species would make an
> ensemble-averaged observation of the human
> population on Earth, they would conclude that
> each person has one testicle and one ovary!
>
> Steven Chu

$\mathfrak{I}$n 1926, the Nobel Prize in Physics was awarded to Jean Baptiste Perrin for his studies on Brownian motion that confirmed the atomic nature of matter (from *nobelprize.org*). Ever since then, scientists have been on the quest to understand nature at the smallest level. For biological sciences this meant a journey towards single-molecule studies of the building blocks of life.

But, for a long time, the observation of single molecules was far out of reach and only ensembles of molecules could be studied in bulk experiments. Those experiments yield an average measurement for the quantity observed, with statistical fluctuations already indicating the existence of values deviating from the measured mean. These averaged observations can hide crucial information and sometimes even be outright misleading by covering underlying heterogeneities. The conceptually opposite extreme of traditional bulk experiments is represented by single-molecule studies. They present a way to reveal the fundamental hidden information by separately interrogating each molecule and recording the result, exposing the full distribution and highlighting possible hidden heterogeneities. In the end, these individual measurements can be combined to produce the same averaged bulk results.

A practical demonstration of this is given in the two scatter plots provided by Anscombe (1973), in which the datasets (Figure 1.3) are quite different. The first dataset illustrates a linear correlation in which the deviation of individual points
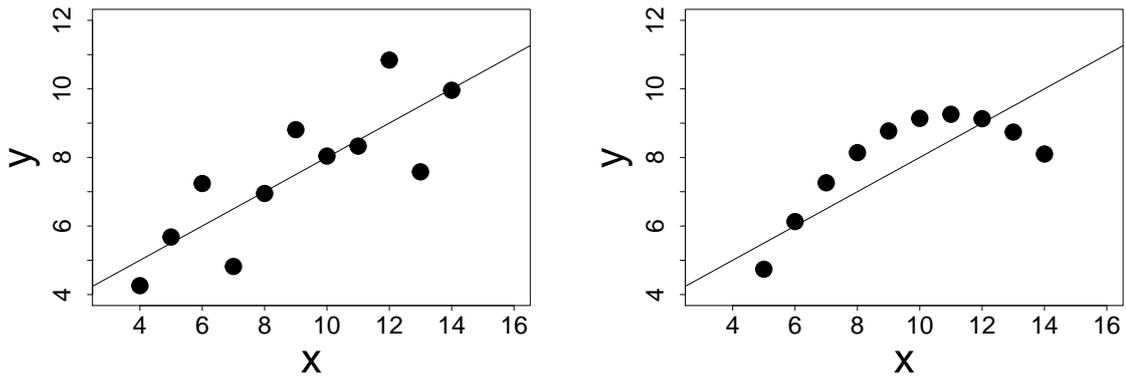
(a) Anscombe's quartet 1   **(b)** Anscombe's quartet 2

Figure 1.3 – The first two scatter plots from Anscombe's quartet. **(a)** Seemingly normal distribution around a linear fit. - **(b)** Non-linear relationship between x and y. - Both (a) and (b) have the same mean and variance for x and y and even the same linear regression. Looking at the underlying values reveals their fundamental difference.

from the line follows a normal distribution (Figure 1.3a). The second dataset shows a clear quadratic correlation (Figure 1.3b). Despite this, they both have the same mean and variance in both x and y and would result in the same linear regression analysis. While it seems obvious that they are both different, this would not be clear by looking at their average values. Only the availability of information for each data point exposes the different underlying relationships.

In the context of pre-mRNA splicing did bulk experiments reveal basic information about if and when splicing factors are present and their potential binding sites. But the actual number of splicing factors present at each stage remained inaccessible. Single-molecule studies would allow to have, for the first time, a glimpse at the underlying properties to not only determine the number of splicing factors present, but to also try to answer formerly inaccessible questions about the behaviour of different subgroups of a splicing factor and how modifications change the nature of individual single splicing factors to give rise to the observed behaviour in bulk measurements.

While the size of molecules can range tremendously over several magnitudes

from just 0.74 Å (DeKock and Gray, 1991) for diatomic hydrogen up to the 5 cm in length of a single human chromosomal DNA molecule (Chen, 1998), they all have in common that the extent in their smallest dimension is below the resolution limit of light microscopes (Lipson et al., 1995). This means that for a long time it was impossible to resolve them directly with light microscopy.

New and powerful methods had to be developed first to allow the investigation of single molecules directly. Only in the last 25 years was the door finally opened to overcome this last step to provide a myriad of new ways to obtain powerful insights beyond the measurement of averages. Due to the advantages of single-molecule studies, they play nowadays an important part in the fields of biology, chemistry and physics (Moerner, 2002).

For a current overview, several excellent in-depth reviews of single-molecule studies exist for the interested reader (Deniz et al., 2008; Greenleaf et al., 2007; Moerner, 2002; Sotomayor and Schulten, 2007). The most widely used technique in single-molecule research is molecular fluorescence.

1.2.1 Fluorescence Microscopy of Individual Molecules

The reflection and absorption of light or electrons from molecules is not enough to directly observe single molecules. A suitable approach is to let the molecules emit photons themselves. The advantage is that it makes the detection of molecules independent of their size and shifts the focus to sensitivity, i. e. the ability to detect this luminescent signal. Distinguishing two molecules from each other would still be subject to the resolution limit.

While there are several types of luminescence (Melhuish, 1984), with some even also used for microscopy (Créton and Jaffe, 2001), fluorescence is the one that "now permeates all of cell and molecular biology" (Lichtman and Conchello, 2005).

(a) Jablonski diagram

Excitation

Internal Conversion and Vibrational Relaxation

Fluorescence

Phosphorescence

Quenching

Intersystem Crossing

Non-Radiative Relaxation

Excited Singlet States

S_2

S_1

S_0

Ground State

Excited Triplet States

T_1
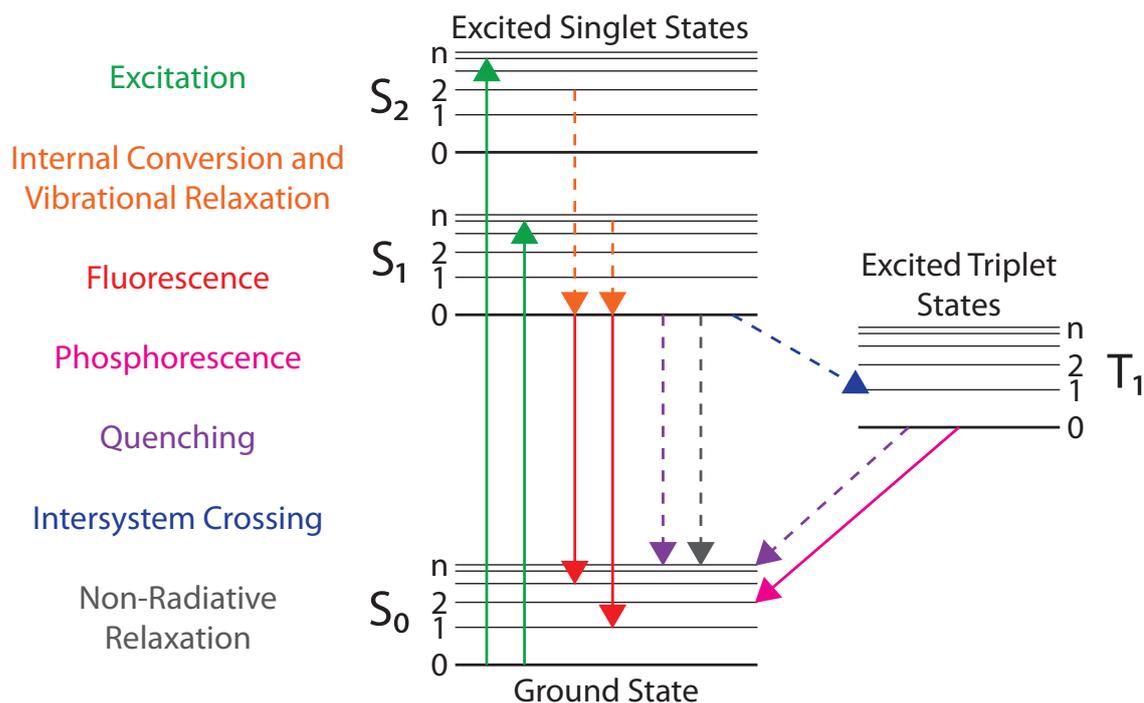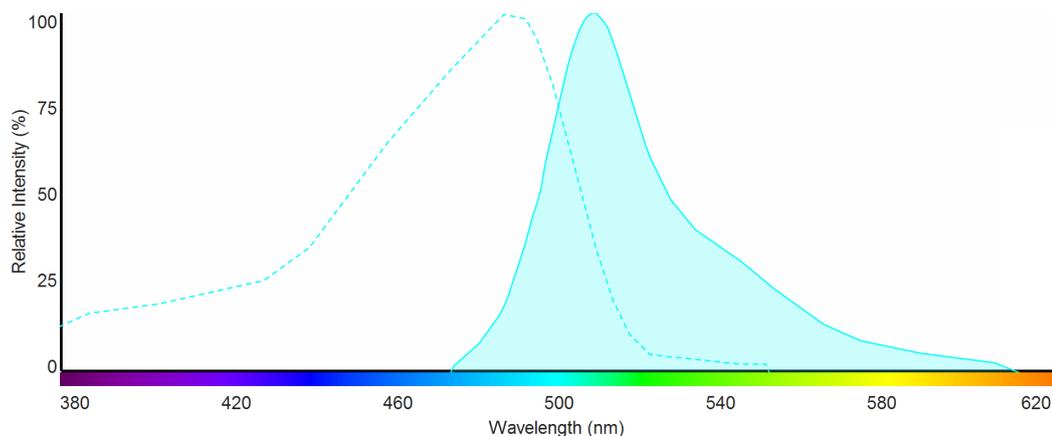
(b) Excitation and emission spectra of GFP

Figure 1.4 – **(a)** Electrons of the molecule are excited from the ground state into an excited singlet state followed by fast internal vibrational and rotational relaxation. Eventually the molecule return to the ground state through an energy transfer mechanism. It is termed fluorescence if this is done by the immediate emission of a photon. Adapted from *Basic Concepts in Fluorescence*. - **(b)** Excitation (dashed) and emission (filled) spectra of GFP. The difference in the wavelengths of the maxima for emission and excitation is the Stokes shift. It is always positive. Created with *Life Technolgies Fluorescence SpectraViewer*.

Fluorescence is the spontaneous emission of light following the absorption of a photon by a molecule, first reported by Stokes (1852). The photophysical pathways can be shown schematically in a Jablonski diagram (Figure 1.4a). A photon of suitable wavelength excites the molecule from the ground state into a higher excited singlet state. Fast internal rotational and vibrational relaxation lower the energy of this excited state before the molecule returns to the ground state by emitting a photon. Other energy transfer pathways are possible and reduce the quantum yield of fluorescence (Lakowicz, 1999). Usually the energy of the emitted photon is lower than the absorbed photon due to internal relaxations. The resulting difference between the wavelength of the absorbed and emitted photon is known as the Stokes shift (Figure 1.4b). This shift leads to spectral separation of excitation and emission.

Some biological structures like mitochondria and lysosomes contain endogenous fluorophores like pyridinic and flavin coenzymes and therefore show autofluorescence (Monici, 2005). The three aromatic amino acids tryptophan, tyrosine and phenylalanine show fluorescence in the near ultraviolet (Teale and Weber, 1957). These are exceptions and most biological molecules do not fluoresce by themselves. For these a fluorescent marker has to be attached to them.

A couple of fluorophores like fluorescein (Sjöback et al., 1995) and DAPI (Kapuscinski, 1995) were already available before the advent of single-molecule fluorescence microscopy. Since then, the pursuit of better and more suitable markers has revealed a variety of other candidates. There are strongly fluorescent proteins that cover a wide range of wavelengths ranging from the near ultraviolet to the near infrared (Shaner et al., 2005); examples of these include the now ubiquitous green fluorescent protein (GFP) and its many derivatives (Day and Davidson, 2009), and mCherry (Graewe et al., 2009). Another option is the labelling with artificial fluorescent dyes which often provide superior photostability, such as the ATTO

dyes (Buschmann et al., 2003), Alexa Fluor (Berlier et al., 2003) and Cyanine (Mujumdar et al., 1993). Recent additions are quantum dots that show unique optical properties that expand the possibilities for labels even further (Resch-Genger and Grabolle, 2008).

While some fluorophores like DAPI bind specifically already, most have to be attached to the molecule of interest. This can be done afterwards by immunofluorescent staining using fluorescent antibodies (Miller and Shakes, 1995) or with fluorophore specific labelling (Haugland, 2002; Wiederschain, 2011). In the case of fluorescent proteins, genetic engineering can create new combined proteins that contain both the protein of interest and the fluorescent protein directly attached. The label is conveniently expressed by the cell alongside the protein (Chalfie et al., 1994). For labelling DNA or RNA, modified anti-sense single stranded nucleotides with attached fluorophores can easily be annealed. Crawford et al. (2008) used a hybrid 2'-O-Me/locked nucleic acid oligo for good specificity and low dissociation rates.

The attachment of the label is not without problems. The complex formed between a fluorescent label and protein has different properties from the protein alone, at least differing in molecular weight, but potentially affecting the folding and function of the protein as well. Sometimes this can lead to widely dissimilar observed behaviours, from false colocalization in *Escherichia coli* cells (Landgraf et al., 2012) due to changes in association profiles of glycan-binding proteins (Fei et al., 2011) to an altered charge of bovine serum albumin (Bingaman et al., 2003). It is therefore of great importance to investigate whether the results are independent of the fluorescent label.

Despite this, fluorescence microscopy is a hugely popular technique and multifaceted applications exist that employ fluorescence for single-molecule experiments. Some of these are described below.

A method to measure distances below 10 nm is Förster resonance energy transfer (FRET), which takes advantage of dipole-dipole interactions between two fluorophores. The energy of one excited donor fluorophore is transferred by this interaction to a second acceptor fluorophore. The efficiency of this transfer is heavily dependent on the distance between those two fluorophores and limited to small distances below 10 nm. It enables either intramolecular distances to be measured or site-specific binding between molecules to be detected. It is a general technique but has been adapted to single molecules (Roy et al., 2008). Single-molecule FRET (smFRET) can be used to obtain kinetic constants for protein folding (Schuler and Eaton, 2008) or DNA unwinding by helicases (Myong et al., 2007). Both variants of FRET will just give relative populations of folded states.

Fast alternating laser-excitation (ALEX) spectroscopy monitors the emission of donor and fluorescence concurrently (Kapanidis et al., 2005). This not only allows FRET efficiency to be obtained, but the quick alternating separate excitation of donor and acceptor also allows simultaneous measurement of the donor-acceptor relative stoichiometry by measuring the acceptor emission directly and indirectly by FRET.

The polarisation of the electric field of light relative to the fluorophore orientation affects excitation efficiency and the polarisation of emitted photons is equally dependent on fluorophore orientation. Fluorescence polarisation studies exploit these properties and allow the determination of orientations of single molecules with temporal resolutions of 20 ms to 40 ms (Mattheyses et al., 2004). Similarly, this can be applied to FRET to investigate the three-dimensional structural dynamics of proteins (Forkey et al., 2003). The time between absorption and emission of a photon contains further useful information about internal properties like rotation and the likelihood of intersystem crossings and external properties like the formation of excimers (Berezin and Achilefu, 2010).

Fluorescence correlation spectroscopy (FCS) covers numbers of molecules that are above a single one but below 100. It interprets spontaneous fluctuations in the fluorescence emission and correlates these to analyse diffusion and binding kinetics (Elson, 2011; Haustein and Schwille, 2007). It allows higher fluorophore concentrations and spatial and temporal measurements at the same time, for example motions within lipid membranes of giant unilamellar vesicles (Ruan et al., 2004). Closely related to this is the method of the photon counting histogram. The histogram of the detected photons over time is used to investigate the amplitude of the fluorescence fluctuations (Chen et al., 1999).

Different additional methods for the illumination of the sample exist that further minimise the illuminated volume and background illumination. The standard confocal illumination can be extended with two-photon excitation (Denk et al., 1990). Even more selective illuminations are done by using total internal reflection (Axelrod et al., 1984) and zero-mode waveguides (Levene et al., 2003), that provide a way to overcome possible dilution limitations. Interestingly, most of these methods can be and often are combined. Examples of these are zero-mode waveguides and FCS (Samiee et al., 2006), two-photon excitation and FRET (Dichtel et al., 2004; Elangovan et al., 2003) and total internal reflection with polarisation measurements (Sund et al., 1999).

1.2.2 Total Internal Reflection Fluorescence Microscopy

Total internal reflection fluorescence (TIRF) microscopy utilises an evanescent wave to illuminate a thin region of a sample located above a surface. As an exciting laser beam passes into the aqueous sample next to an interface, it is refracted according to Snell's law,

$$n_1 \sin(\theta_1) = n_2 \sin(\theta_2), \tag{1.1}$$

(a) Total internal reflection

- Sample
- Evanescent wave
- Immersion oil
- Excitation beam path
- Objective
- Back focal plane
- Emission beam path
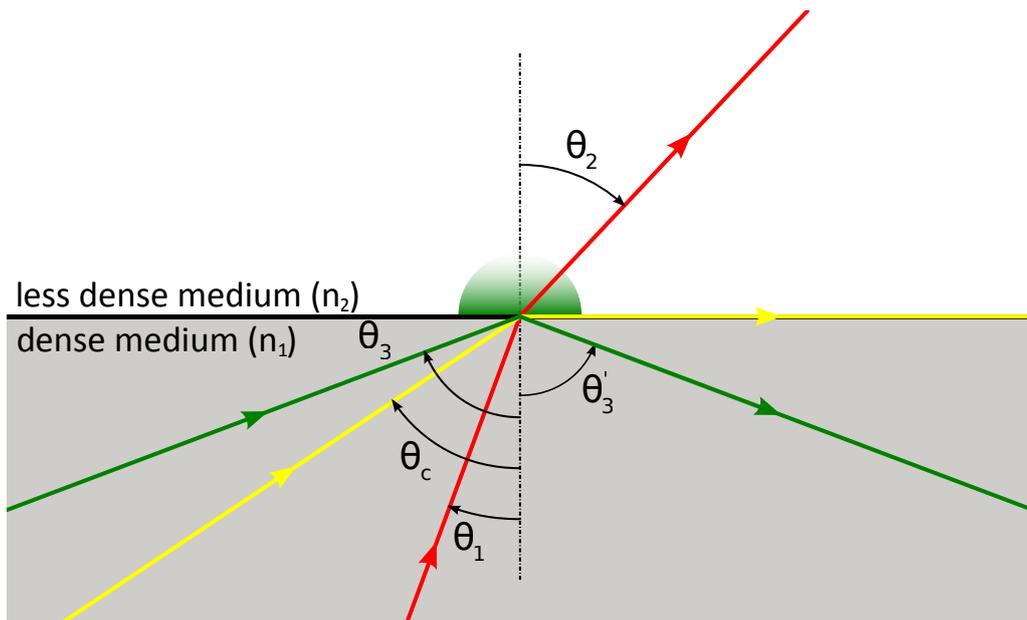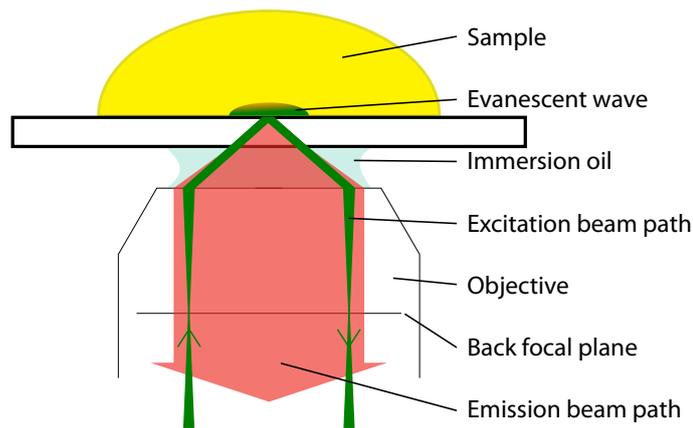
(b) Through-objective setup

Figure 1.5 – **(a)** Refraction at the interface of two different media. Medium with higher refractive index in grey, with lower refractive index in white ($n_1 > n_2$). Light path shown in red is for normal reflection ($\Theta_2 > \Theta_1$) with an incident angle smaller than the critical angle ($\Theta_1 < \Theta_C$), shown in yellow with the reflection parallel to the boundary for an incident angle equal to the critical angle, shown in green for an incident angle greater than the critical angle ($\Theta_3 > \Theta_C$). In this case, an evanescent field is generated in the lower refractive index medium. Adapted from *Wikimedia Commons 1*. - **(b)** Excitation and emission beam path way shown in green and red respectively for a through-objective setup. Adapted from *Wikimedia Commons 2*.

where θ_1 is the angle of incidence, θ_2 the angle of refraction and n_1 and n_2 the corresponding refractive indices of the media. For a laser beam passing from a medium with a higher refractive index, like glass, into a medium with a lower refractive index, like water ($n_1 > n_2$), the angle of refraction is larger than the angle of incidence ($\theta_2 > \theta_1$). For large enough angles, the light is completely reflected at the interface (Figure 1.5a). The minimum angle at which this total internal reflection occurs is called the critical angle,

$$\theta_C = \arcsin\left(\frac{n_2}{n_1}\right), \text{ for } n_2 < n_1. \tag{1.2}$$

Above this angle, the incident light is entirely reflected. An important feature of this total reflection is the generation of an evanescent wave that propagates into the lower refractive index medium. In biological microscopy, this would be an aqueous sample. The generated evanescent field has an exponential decay of the intensity, I, along the optical axis, z:

$$I(z) = I_0 \exp(-\frac{z}{d}) = I_0 \exp(-\kappa z), \tag{1.3}$$

with the penetration depth, d:

$$d = \frac{\lambda}{4\pi\sqrt{n_2^2 \sin^2 \theta - n_1^2}}. \tag{1.4}$$

that depends on the vacuum wavelength λ of the exciting laser, its incident angle θ and the two refractive indices n_1 and n_2. For characteristic angles of incidence in total internal reflection microscopy, this lies in the range of 50 nm to 100 nm, usually excluding the region of the sample that is more than 250 nm above the surface (Axelrod et al., 1984). The advantage of this illumination geometry is the extremely restricted illumination of the sample, greatly reducing a possible background from the rest of the sample.

Various ways exist to utilise total internal reflection as a method of fluorescence microscopy, they use either an alternative high refractive index object, commonly a prism (Axelrod, 2001; Thompson and Steele, 2007), or the microscope objective directly (Thompson and Steele, 2007). A method was proposed to combine these two ways (Burghardt et al., 2010) and another method makes use of a parabolic mirror (Ruckstuhl and Seeger, 2003).

The advantages of prism-based illuminations are the lower background, because there are completely separate beam paths for excitation and emission, and a larger possible illumination area. In contrast, the objective-based illumination makes use of objectives with a higher magnification and a higher numerical aperture, collecting more light and achieving a higher collection efficiency per single molecule (Hassler et al., 2005). Usually the sample is more easily accessible with objective illumination, allowing uncomplicated manipulations.

Due to the advantages stated above, which will be discussed in this thesis in relation to colocalization and bleaching step studies, the objective-based illumination was chosen (Figure 1.5b). Here the objective is positioned below the sample interface and coupled to the cover slip with immersion oil, all three having comparable refractive indices. The incident laser beam is guided into the objective parallel to the optical axis, focussed onto the back focal plane. Moving it outwards results in different incident angles of collimated laser light at the cover slip-sample interface. With a high enough numerical aperture, angles greater than the critical angle can be reached. The incident beam is reflected at the interface, giving rise to the afore-mentioned evanescent wave penetrating into the sample.

With its selective surface-defined illumination, TIRF is especially well suited to investigate membranes (Thompson et al., 1993) and membrane-related events like cell-signalling (Sako et al., 2000). The movement of single molecules that are confined to the membrane can be tracked (Schmidt and Schütz, 1996) or resolved with

such spatial accuracy that single steps of myosin V moving along actin filaments can be seen (Yildiz et al., 2003). Another implementation of it is for colocalization experiments which will be discussed below.

1.2.3 Colocalization

In fluorescence microscopy, colocalization refers to deconvoluting the spatial overlap between spectrally discernible fluorophores. While these methodologies were introduced in cell imaging, they can easily be extended to be used for single-molecule studies to deconvolute the positions of two single molecules in close proximity to each other where the fluorescence emission from their attached fluorophores overlaps.

If the conditions are chosen in such a way that colocalization is unlikely to happen by chance, then the relationship between the two colocalized species can be revealed. They could be bound directly to each other, bound to a common partner or associated through another indirect mechanism.

Two methods for illumination can be used for colocalization studies. Confocal illumination is used for detecting dual labelled complexes in solution that move through the detection volume; this can be done by alternating laser excitation (Kapanidis et al., 2005) or using interleaved ultra-fast laser pulse excitation (Olofsson and Margeat, 2013). While it allows direct observation of complexes in solution, a single complex cannot be followed over time. For this, surface selective illumination is preferred, where the complexes are immobilised on the surface. This illumination can be achieved by total internal reflection (Section 1.2.2) or with zero-mode waveguides (Levene et al., 2003). The work presented here was done with the surface illumination done by total internal reflection.

To assess colocalization, two steps have to be taken (Malkusch et al., 2012). First, the spots that indicate a labelled molecule are identified. From these, the position

(a) Red spots

(b) Green Spots

(c) Superposed spots

(d) Chromatic aberration

(e) Intensity profile
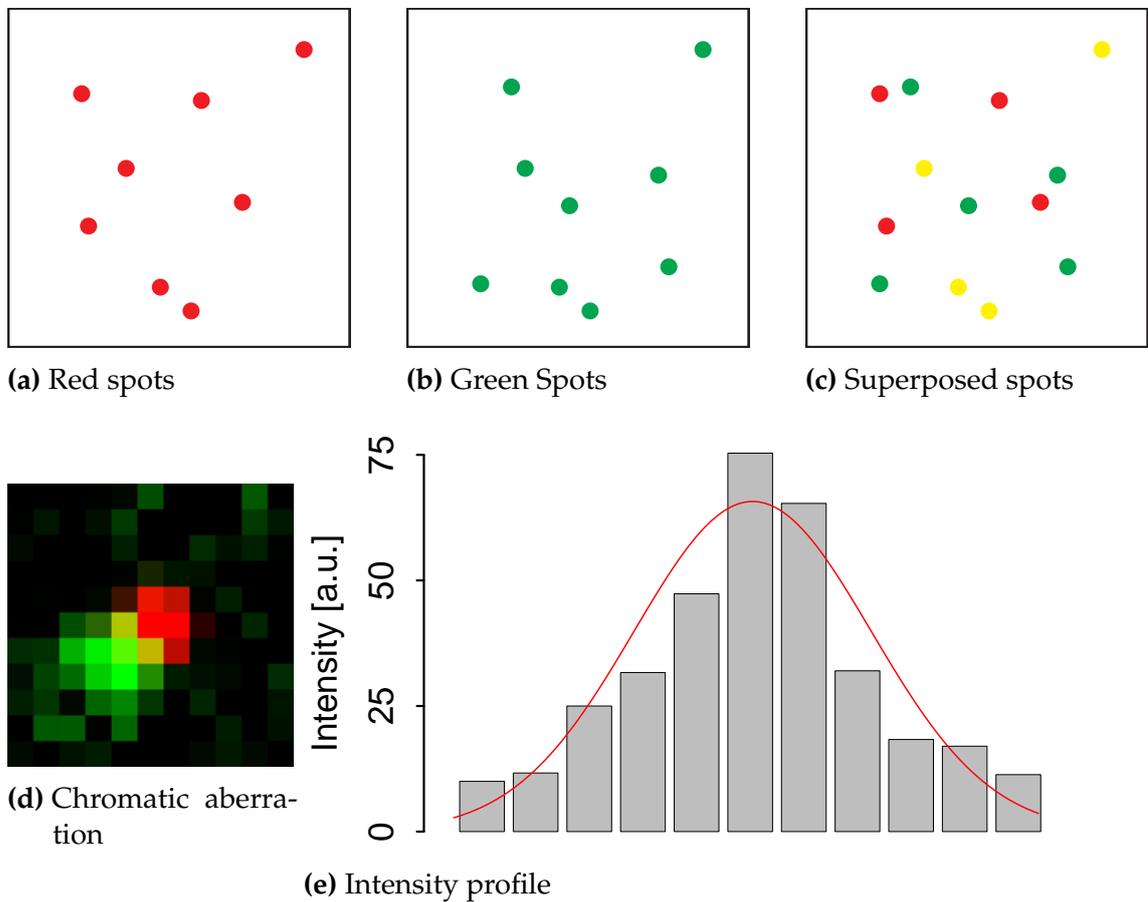
Figure 1.6 – **(a)** Red spots, as obtained for example with Cy5 at 633 nm. - **(b)** Green spots, as obtained for example with GFP at 488 nm. - **(c)** Superposing (a) and (b) shows colocalized spots in yellow, with the not colocalized spots still being visible as red and green. - **(d)** Exemplary superposed image of a red and green image channel. Visible are the pixelated nature of the recorded spots and the chromatic shift of the spot centres. - **(e)** Exemplary intensity profile of a single spot. The maximum intensity does not necessarily correspond to the true centre. A Gaussian intensity fit allows sub-pixel determination of the centre position.

of each fluorophore is obtained and the distance between possibly colocalized molecules is measured. If the distance falls below a certain threshold, labels are considered colocalized. All this can be done quickly by eye or more thoroughly by computer analysis.

Spots have to satisfy two general requirements to be considered valid. Their intensity has to be significantly above the background and their shape has to match certain expectations. The threshold intensity has to be sufficiently high to discriminate them from local variations and possible cross-talk from other fluorophores, but it should still allow for the identification of weak spots. If their shape is not symmetrical, the spot might be a contamination or multiple indistinguishable spots next to each other.

Once a spot is identified, its position has to be determined. While fluorescence is emitted from a location much smaller than the diffraction limit, the resulting intensity profile is a two-dimensional diffraction pattern, known as an Airy disk, that can be much wider than the diffraction limit (Pertsinidis et al., 2010). Photon noise, fluorescent background, and, with CCD cameras, readout noise and pixelation make it difficult to identify the centre of the intensity profile and a precise position for the label. To find a good approximation of this centre, one can use the pixel that received the highest intensity, as this would correspond to the centre of the Airy disk. This however would limit one to the resolution of a pixel. To achieve sub-pixel resolution, one can use the whole recorded distribution of intensities by calculating a weighted centroid or trying to reconstruct the Airy pattern to obtain the original position (ibid.).

Cheezum et al. (2001) found that a two-dimensional Gaussian function,

$$G\left(x, y\right) = A \times \exp\left[-\frac{\left(x - x_0\right)^2 + \left(y - y_0\right)^2}{B}\right], \qquad (1.5)$$

with x_0 the centre of the x coordinate, y_0 the centre of the y coordinate and two

constants A and B, is the best fit for the images of a single fluorescent dye. The position of immobilised single organic dyes can be determined with better than 1.5 nm precision (Yildiz and Selvin, 2005). For this method, high photon counts are indispensable, making the evaluation heavily reliant on the labelling. Flexible fluorophores like GFP can move within certain limits and so can the molecule it is attached to, therefore, there is a finite uncertainty in position. While precise position determination is favoured, it only has to be good enough to distinguish between true colocalization and colocalization by chance.

For a random distribution of red and green fluorophores that do not overlap is the number of green spots that will be colocalized with red spots just by chance, C, given as

$$C = \frac{\pi r^2}{A} \times G \times R, \tag{1.6}$$

where r is the distance below which spots are considered colocalized, A, the total observed area, G, the number of green spots and R is the number of red spots. With a recorded field of view of 250 pixels by 250 pixels, 50 green and 50 red spots, even a colocalization distance as large as 2 pixels would result on average in only 0.05 randomly colocalized spots. With a total colocalization percentage of 30 % would this amount to an error of 0.3 %. Low positional accuracy might therefore still be sufficiently precise for the majority of experimental measurements.

1.2.4 Photobleaching and Analysis of Photobleaching Steps

The percentage of colocalized molecule measurements described in the previous chapter is not the only information that can be extracted from the data. Multiple molecules with the same label can be positioned at distances within the resolution limit, for example if they form part of a single macromolecular complex and their
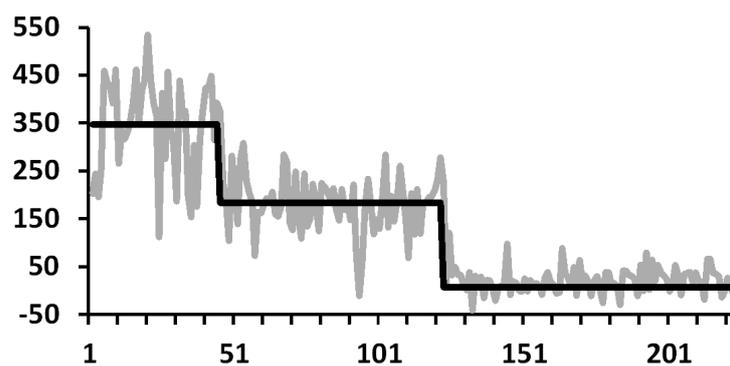
Figure 1.7 – An exemplary trace, taken from an experimental acquisition, of the background-corrected fluorescence intensity over time is shown in grey. Apparent underlying steps are shown in black (Steps were obtained with the algorithm described in Section 3.9).

distance apart is below the resolution limit of the microscope. In this case, the labelled molecules cannot be distinguished from each other; however, the intensity of their fluorescence over time contains additional information about the number of colocalized fluorophores.

Each fluorophore has a certain quantum yield, indicating the ratio of decay occuring by fluorescence and non-fluorescent pathways. For the same laser intensities, different quantum yields result in different numbers of emitted photons (Weber and Teale, 1957). Of concern from the experimental side is the detection efficiency; how many of these emitted photons are collected by the setup. With photons detected ultimately by a camera, the quantum efficiency (QE) of the camera plays a role. QE indicates how collected photons are finally transduced into an electron signal (Zhang et al., 2009). On top of this, the angle of incidence of the light relative to the dipole moment of the molecule (Ha et al., 1999) and the number of fluorophores close by, through fluorescence enhancement or fluorescence quenching by nonradiative energy transfer, influence the overall intensity (Anger et al., 2006). While the wavelength-dependent QE of the camera is known (Magnan, 2003), the relative orientation of the fluorophore is

unknown. The number of the fluorophores is a quantity that might have physical importance and is something that one might want to find out. This can be done directly.

If the presence of an additional fluorophore results in a big change in signal level, it becomes possible to count the fluorophores. If multiple fluorophores suddenly stop the emission of photons, one at a time, simply counting the successive downward steps in the intensity time trace would tell the number of fluorophores present initially. Almost all fluorophores stop emitting photons at some point due to photobleaching (Dempsey et al., 2009a; Widengren and Rigler, 1996). This causes prolonged or even permanent loss of their ability to fluoresce. This process is stochastic for a small number of molecules. The average time until bleaching for a fluorophore depends on the excitation intensities and other environmental factors, e. g. oxygen concentration, temperature and pH. (Dempsey et al., 2009a).

If the average time until bleaching is adjusted by either controlling the laser intensities or the environmental conditions, the separate bleaching events can be distinguished (Casanova et al., 2007; Leake et al., 2006); this must still take place in a reasonable total time period. Each bleaching step corresponds to one fluorophore and each fluorophore is bound to one molecule, so each bleaching step indicates that one labelled molecule is present. However, the factors described above that influence total intensity make the step detection not necessarily straightforward.

In biological systems, most single-molecule fluorophores tend to bleach quickly due to the presence of molecular oxygen, O_2, via photo-oxidation (Renn et al., 2006). The presence of O_2 also reduces triplet-state blinking (ibid.). Bleaching times need to be long enough to accurately discern individual steps. A common solution to increase the average lifetime was to use an enzymatic system consisting of glucose oxidase and catalases that effectively reduces O_2 concentration (Be-

nesch, 1953). Unfortunately, due to the loss of O_2, the lifetime of triplet-states increase. The addition of Trolox counteracts this, and the combination of these components provides stable long-time fluorescence (Rasnik et al., 2006).

Unfortunately, the addition of glucose rapidly depletes ATP and inhibits splicing (Crawford et al., 2008). An alternative system is protocatechuate dioxygenase (PCD) and protocatechuic acid (PCA) which does not inhibit splicing (ibid.). PCD catalyses the reaction of PCA to β-carboxy-*cis,cis*-muconic acid, which results in a loss of O_2 in solution (Patil and Ballou, 2000).

Colocalization studies on single molecules have now been used to look at a wide range of biochemical process like transcription initiation (Friedman and Gelles, 2012), ribosomal translation of mRNAs (Tsai et al., 2012) and spliceosome assembly (Hoskins et al., 2011b).

1.2.5 Analysing Single-Molecule Studies

The detection of single molecules is a challenging experiment. When this has been overcome, the analyses of data and the necessary data reduction to find meaningful and significant results presents a new challenge.

One of the advantages of single-molecule studies lies in exposing heterogeneities by not only looking at an averaged measurement, but having information for each molecule available. This means one has to actually look at and analyse each and every molecule separately. For meaningful statistics, this can be hundreds to thousands of events one has to look at in one single set of experimental data after it was already performed. This can be a time-consuming process.

The sample set of hundreds or thousands of single molecules still represents a tiny fraction of all the events present. One has to take every step to make this subfraction as representative of the whole population as possible. This means to make sure that the selection process is unbiased. This is demonstrated by
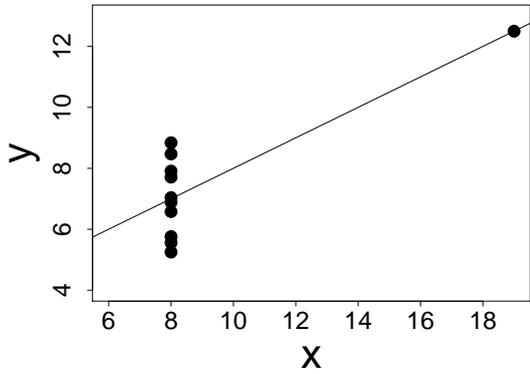
(a) Anscombe's quartet 4
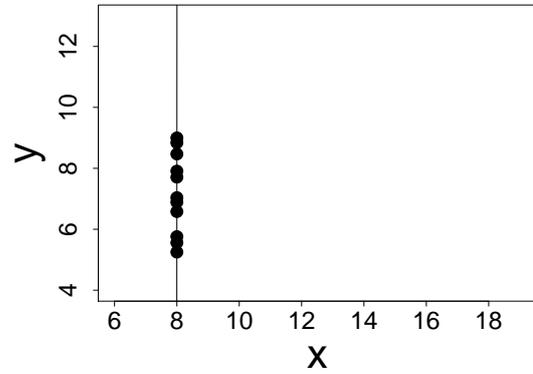
(b) Dismissed outlier
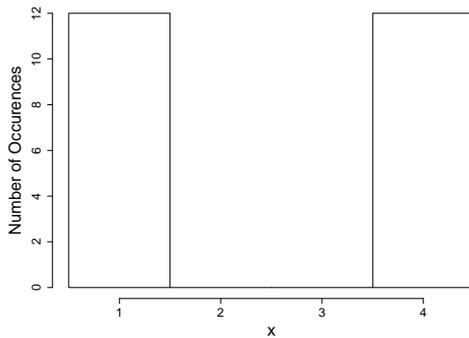
(c) Unbiased distribution
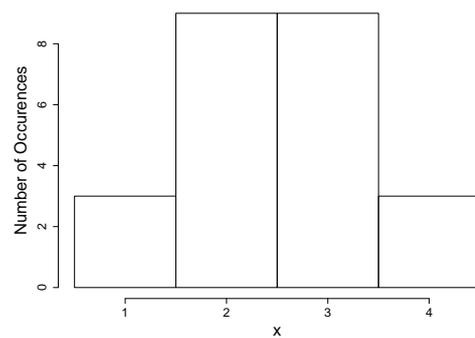
(d) Biased distribution

Figure 1.8 – **(a)** Anscombe's fourth dataset (Anscombe, 1973) with one outlying value. - **(b)** Removal of the outlier changes the distribution significantly both in mean and standard deviation. Careful measures have to be taken to guarantee that this outlier is a true artefact and not an intrinsic property. - **(c)** True underlying distribution for an experiment that groups events into four different categories, showing two clear subgroups. - **(d)** Biased categorising of the distribution in (c) changes the distribution into a seemingly Gaussian distribution.

Anscombe's fourth dataset (Figure 1.8a). The one outlying data point in the top right changes the whole dataset considerably (Figure 1.8b). One might easily either miss this point completely, as one is only 'detecting' values from 4 to 12, or dismiss it as an artefact. While there might be situations where this would be appropriate, it defeats the purpose of actually looking at single molecules. For example, assuming x might be reaction rate, this one outlier could actually contribute much more to the overall reaction than all the other ones combined. Ignoring it would mean missing potentially the most important bit of information.

As single-molecule studies are often performed at the technical limits, the acquired data will not always present a clear picture. One has to carefully balance false positives against false negatives in the selection process. Consequentially, not only the selection, but also the analysis of events has to be unambiguous. Bias here could potentially obscure the information existing on heterogenities. For demonstration, the true underlying distribution for an experiment is shown in Figure 1.8c, where the brightness of events is classified into four categories with 1 being the darkest and 4 being the brightest. A human assessor might be biased to put especially dim ones with a chance of 75 % into the brighter category and likewise especially bright ones into the darker category. Instead of two clear sub-populations, only a normal distribution remains (Figure 1.8d). Similarly, an improperly analysed single-molecule experiment might in the end only give one the same average that was obtainable by bulk experiments.

Even when an unbiased analysis method has been developed, the results should still be investigated for possible hidden biases by checking against other independently known properties.

All this makes the analysis a potentially time-consuming undertaking. Additionally, the raw data from single-molecule experiments often encompasses huge amounts of information wherein the specific data pertaining to the single

molecules is hidden. Fortunately, the development of single-molecule experiments has been accompanied and supported by an outstanding expansion in computing resources which reduce the time needed to analyse the data substantially.

A look back to the early days of particle physics nicely illustrates the benefits of the current computing resources available in research. Then, cloud chambers were used to observe and identify elementary particles. While acquiring one image with these took around 1 s, analysing them without the aid of computers took thousands of specially trained personnel (Galison, 1997).

Computer-based algorithms also eliminate the human bias. They are impartial within their programmed limits and will judge every event by the same criteria. While biases might still be present, depending on the implementation of an analysis program, these are at least fixed and do not vary over time and still do not depend on the judgement of the experimenter. This allows careful analyses to detect, counteract and eliminate them.

1.2.6 Single-Molecule Experiments on RNA Splicing

The first observation of splicing on a single-molecule level in real time for whole cell extracts of yeast was done by Crawford et al. (2008). Pre-mRNA with different fluorophores incorporated into its exons and introns was attached via streptavidin to a PEG-biotin glass surface and observed with TIRFM. Analysing colocalization over time showed a loss of intron-specific fluorescence, indicating the progress of splicing.

Cherny et al. (2010) were the first to assess the stoichiometry of proteins involved in mammalian splicing by counting the number of polypyrimidine tract-binding proteins (PTB) bound to the pre-mRNA in crude (functional) nuclear extract. Its action as an exon repressor is mediated through distant binding sites

in the flanking introns and two competing theories existed, one proposing a PTB-mediated looping, the other the propagation of complexes. Splicing complexes were incubated before dilution and tethered to the surface via their RNA substrate. TIRFM was used for observation. Identifying the number of bound proteins was done by counting bleaching steps and by analysing the total emission intensity. With both methods it was found that the number of PTB molecules is five or six. This was inconsistent with the propagation model, leaving the looping model. Lamichhane et al. (2010) used single-molecule FRET to investigate the same activity of pure PTB on short RNA oligonucleotides. They found that the domains three and four of PTB can form RNA loops that can bring the 3′ and 5′ ends of the RNA together.

Similarly, Hodson et al. (2012) investigated the stoichiometry of U1 snRNPs in the early splicing complexes and their role in splice site selection. For different substrates with either strong or weak 5′ splice sites and different distances between alternative 5′ splice sites, complex formation was stalled in complex E and A. Their results showed that for two possible 5′ splice sites there can be two bound U1 snRNPs in complex E with only one U1 snRNP remaining in complex A. This reconciles the two findings. On one hand, the complex E is already committed to splicing, but on the other hand, the selection of 5′ splice sites is not yet fixed. An ATP-dependent process removes splicing factors at orphaned splice sites and prevents further binding in complex A. Furthermore, the intron-proximal splice site seems to be favoured when a U1 snRNP is bound to it.

Single-molecule FRET studies can provide information about distances during splicing, highlighting conformational changes. Abelson et al. (2010) labelled a 3′ and 5′ exon with FRET compatible fluorophores, attached the 3′ exon to the surface and observed pre-mRNA via TIRFM. They observed time- and ATP-dependent conformational transitions, all of them reversible, pointing at a highly fluctuating

spliceosome operating close to thermal equilibrium.

The first insights into the dynamics of the yeast spliceosome itself were revealed by Hoskins et al. (2011a). They observed its subcomplex dynamics in real time in yeast whole cell extract. RNA substrates were bound to a PEG-modified glass surface and illuminated using TIR with three different wavelengths. The colocalization of pre-mRNA with splicing components, such as U1 snRNP, was monitored over time. To distinguish between bleaching and dissociation of the components, the splicing components were labelled with two fluorophores of the same kind. If the observed intensity dropped by half, it indicated a bleaching event. If the intensity reduced to zero, either both fluorophores had bleached at the same time or, more likely, the component had disassociated. This allowed the authors to follow association and dissociation events over time and calculate reaction constants. They found an association of U1 snRNP only in the presence of a 5′ splice site. Only in the presence of ATP did the subcomplexes U2 snRNP, tri-snRNP and NTC associate, and each showed different rates. They found this to be consistent with an assembly order of U1 → U2 → tri-snRNP → NTC. All associations were reversible and only increased the commitment to splicing; no step guaranteed a proceeding splicing event.

This method was subsequently extended to include FRET (Crawford et al., 2013). This allowed the authors to not only follow the spliceosome assembly over time but also any conformational changes that occurred during the same time. The FRET measurements revealed that the 5′ splice site and branch site remain separate during the spliceosome assembly and only approach each other just before the first catalytic step, after which the process becomes highly dynamic. This was contrary to previous bulk studies that suggested that this conformational change occurs around complexes E or A.

All of these single-molecule studies on RNA splicing have the same common

methodology. To follow the same molecule over a longer time, the pre-mRNA is immobilised on a surface and selectively illuminated using TIR illumination. Analysis is done by colocalization studies and further analysis is done by looking at intensity time traces.

A different approach is fluorescence correlation spectroscopy (FCS). While this is still a method in the single-molecule regime, one cannot follow a specific molecule over time. Instead, FCS is an approach that allows to study living cells. It was used by Rino et al. (2007) in combination with computational methods. They were interested in the intranuclear kinetics of snRNPs in the presence or absence of splicing activity. They found that, against predominant opinion, the spliceosomal components do not need a trigger signal to be recruited from their storage site in nuclear speckles. Rather, the recruitment is done by random diffusion throughout the whole nucleus.

Huranova et al. (2010) used fluorescence recovery after photobleaching (FRAP) and FCS to determine the interaction rates of individual snRNPs with pre-mRNA. They found that U1 and U4/U6 snRNPs are only transiently associated with pre-mRNA, and U2 and U5 snRNPs show longer interaction times in the range of 15 s to 30 s.

1.2.7 Limitations of Colocalization Studies with Total Internal Reflection for Human RNA Splicing

It is evident that the combination of total internal reflection and colocalization measurements can deliver unique insights into the workings of splicing. Still, there are some inherent limitations to this method.

RNA substrates are attached to a surface, severely limiting translational and rotational freedom of the RNA. The surface might further affect the structure of the substrate. The same problems arise for the participating proteins. The

experimental setup can only present a modified version of the real circumstances.

Due to the large number of different proteins involved in splicing and the absence of any minimal system, experiments are always performed in nuclear extracts. Protein concentrations in nuclear extracts are too high to distinguish individual proteins labelled with the same fluorophore from each other. With the current resolution limits of microscopes and necessary protein concentrations, dilutions of the sample are unavoidable to correctly asses colocalization. However, the dilutions of nuclear extracts will disrupt the local environment and many ongoing processes, leaving the preservation of all high affinity interactions. This makes it impossible to follow all changes over time. The dynamics of spliceosomal assembly have been followed with yeast whole cell extract only, and never in human nuclear cell extract (Crawford et al., 2013; Hoskins et al., 2011a). The yeast splicing factors can be purified, labelled and added back at lower concentration because their affinities are high. Mammalian systems suffer from intrinsic high levels of absorption to surfaces, possibly because of the highly charged and unstructured domains on many of their proteins, and the lower affinity of splicing factors. The latter bind in association with regulatory proteins, themselves of low affinity. Thus, high concentrations of labelled proteins are required to support splicing.

Another limitation is the surface attachment. It is necessary to allow longer observation times but it restricts the possible conformations of the RNA substrate. A better option would be, if possible, to confine proteins to a continuously observable space without the need for any artificial attachment.

A totally different approach might solve the problem of observing spliceosome assembly in the complex environment of nuclear extract. If one could put the molecules of interest in isolated microenvironments where the concentrations would be exactly the same as in nuclear extracts, then the current limitations of

single-molecule experiments would be overcome. A potential solution to obtain isolated microenvironments would be to use tiny droplets provided by microfluidic techniques.

1.3 Microfluidics

> There's plenty of room at the bottom
>
> ―――――――――――――――――――――――――――
>
> Richard Feynman

Microfluidics as a field is the study and manipulation of fluid motions on the scale of, as the names suggests, micrometres. Owing to these dimensions, the volumes of fluids manipulated are usually in the range of microlitres to femtolitres (Whitesides, 2006). Fluid properties that could be neglected at larger dimensions will now dominate a system resulting in a unique behaviour. It is a multi-disciplinary field as the manufacturing and understanding of microfluidic devices needs knowledge from the fields of physics, engineering and chemistry. It provides useful applications in medicine, biochemistry and biology.

Miniaturisation is a general trend of development in electronics and engineering, which also includes innovations in nanotechnology; the advantages of similar developments in chemical and biochemical studies are closely related. Smaller volumes provide crucial technical advantages such as an increased sensitivity, lower costs and faster analysis times. Smaller length scales also lead to faster sample processing and response times, all critical values in the pursuit of high throughput analysis and efficiency.

The revolutionary concept of a lab-on-a-chip, as the unique behaviour of liquids in microfluidics allows to manipulate them in order to perform complex laboratory work, is envisaged to replace entire scientific laboratories (Chin et al., 2007; Daw and Finkelstein, 2006; Haeberle and Zengerle, 2007), with promising recent achievements (Greener et al., 2012). By divorcing advanced medical diagnostic systems from expensive and elaborate labs, sophisticated diagnostic tools could be brought to parts of the world that would not have been able to afford them before (Yager et al., 2006).

The possibilities of its application seem almost endless, with a drive to try to replicate everything possible in a lab inside a microfluidic device (Theberge et al., 2010). Besides its diverse applications in inorganic chemistry (Abou-Hassan et al., 2010), analytical chemistry (Choi et al., 2012; Zhu and Fang, 2013), mass spectroscopy (Kirby and Wheeler, 2013), medical diagnostics, in combination with biosensors (Rivet et al., 2011) or acoustics (Friend and Yeo, 2011), and even computing (Fuerstman et al., 2007), the field where it seems to offer the most interesting and far-reaching developments is in the biological sciences (Schneider et al., 2013).

1.3.1 Pressure Driven Laminar Flows

As not all the factors that affect the behaviour of fluids scale similarly, the flow of a liquid constrained in microdimensions can differ quite significantly from the flow of larger length scales. Surface area to volume is proportionally higher, viscous liquid properties dominate and surface tension, fluidic resistance and capillary forces play an ever greater role.

The dimensionless Reynolds number (Re) can be used to describe flow conditions:

$$Re = \frac{v D_H}{\nu} \tag{1.7}$$

with the mean velocity of the fluid v, the hydraulic diameter D_H and the kinematic viscosity ν. For values below 2040, liquid flows are typically laminar (Avila et al., 2011); for values above 5000 liquid flows are turbulent; for intermediate values there is a mixed flow regime. For the micrometer-sized features of microfluidic devices and liquid viscosities generally similar to water ($\approx 1 \times 10^{-6} \, \mathrm{m^2/s}$), even extremely fast flows are confined to the laminar range. Thus the only mixing that

occurs is on a molecular level due to diffusion. While this certainly provides some challenges, it is also one of the advantages that can be exploited.

The propulsion of liquids in microfluidic devices can be achieved by different means (Mark et al., 2010). In the work described in this thesis, the flows are driven by pressure. The necessary pressure for the liquid transport can come from internal or external sources, controlled either by volume as with syringe pumps or directly by compressed gas using a hydrostatic reservoir (Seemann et al., 2012).

In most microfluidic applications, syringe pumps are used as they are readily available. The syringes can be loaded with the necessary solutions and provide a continuous supply. The speed of the plunger can be changed to provide different flow rates or flow pressures. Together with the laminar flows and their predictable velocity profiles one has highly adjustable flow rates for each injected solution (Mark et al., 2010).

The problem with syringe pumps is their indirect way of adjusting the pressure. They primarily apply a force by changing the hold volume of the syringe linearly, resulting in an increase in pressure. The resulting pressure may depend on factors other than the the syringe itself. It might depend on the dimensions of the tubing used and the microsystem characteristics. It can take long times to reach pressure equilibrium, with the possibility of hysteresis, flow rate setting irreproducibility or an equilibrium may not be reached and pressure will continue to increase. They are also are driven by step motors, where tiny periodic changes occur, leading to pulsing. This makes them poorly adapted to microfluidic flows in the nanolitre range.

In such situations, it may be better to control the pressure. With a direct control of the pressure, the responses of the systems are almost instantaneous with fast equilibration times providing highly stable flow rates (see *Fluigent FASTAB*). This alleviates the problems of syringe pumps. A greater control is obtained, up to

47

the possibility of droplet-on-demand generation (Galas et al., 2009; Zagnoni and Cooper, 2010).

1.3.2 Microfluidic Applications in Biochemistry

Numerous reviews of microfluidics methods exist; many of which describe the vast array of possible applications in the life sciences. Reviews cover the manipulation and analysis of biological cells (Yi et al., 2006; Yin and Marshall, 2012), PCR amplification (Khandurina et al., 2000; Schaerli et al., 2009; Zhang et al., 2006), PCR amplification combined with capillary electrophoresis (Woolley et al., 1996), the crystallisation of proteins (Li and Ismagilov, 2010; Zheng et al., 2003), investigating enzyme activity (Baret et al., 2009; He et al., 2005), the expression of proteins (Courtois et al., 2008; Dittrich et al., 2005) and cell sorting (Wang et al., 2005).

1.3.3 Droplet-based Microfluidics and Flow Focussing

A subclass of pressure driven microfluidics is based on droplets. One fluid phase encapsulates a second fluid phase in such a way that mono-disperse droplets are formed by the second phase. Most commonly, water droplets are formed in a continuous oil phase to provide a biologically suitable environment. The droplets provide confined compartments of identical size, allowing large numbers of parallel experiments and potential for high throughput analysis (Agresti et al., 2010; Brouzes et al., 2009; Granieri et al., 2010; Guo et al., 2012; Srisa-Art et al., 2007; Taly et al., 2007). Additionally, the droplets can be manipulated in various ways including mixing, splitting, fusion and sorting (Teh et al., 2008) and offer a wide range of applications (Huebner et al., 2008; Theberge et al., 2010)

Several ways exist to generate water droplets in oil; one can use microlitre-sized chambers driven by piezoelectric actuators that dispense droplets on demand (Xu

(a) Co-flowing streams

(b) Cross-flowing streams
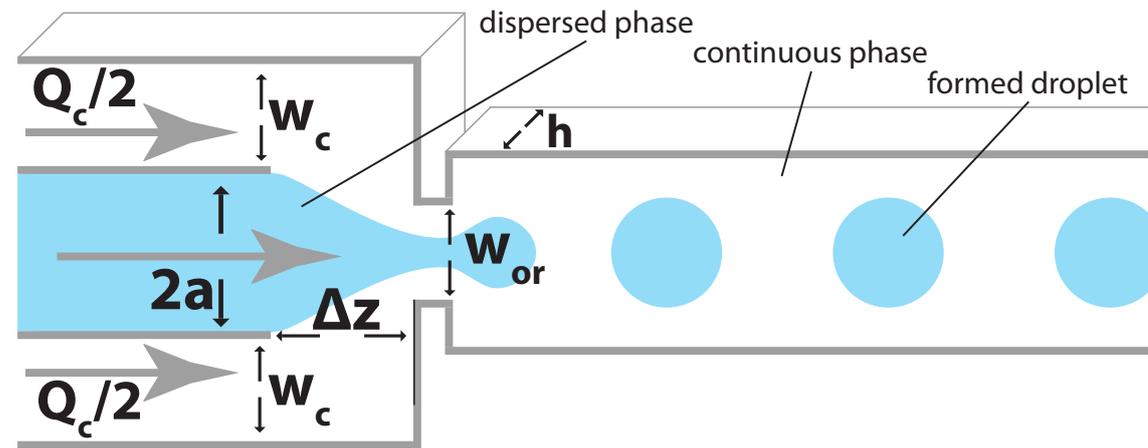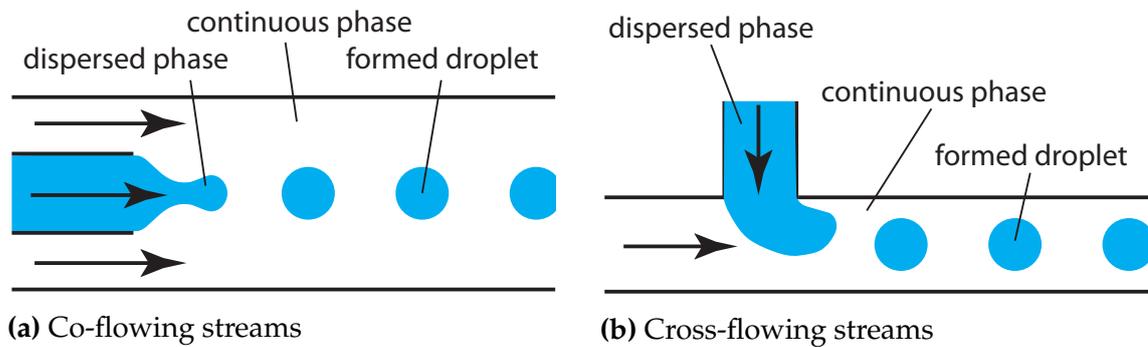
(c) Flow focussing with relevant properties

Figure 1.9 – The three main microfluidic geometries for droplet formation by intersecting streams. All channels are assumed to have a uniform depth. The dispersed phase is shown in light blue, the continuous phase as white. **(a) - (c)** Classes of droplet generating geometries. - **(c)** Relevant dimensions and flow rates in flow focussing.

and Attinger, 2008), or, alternatively, one can generate aqueous monodisperse droplets at the junction of two pressure-driven intersecting streams (Lagus and Edd, 2013). The latter method has become popular in recent years. The two immiscible phases in the co-flowing streams correspond to a continuous phase as the carrier and a dispersed phase as the droplet (Figure 1.9a and Figure 1.9b). These are usually passive methods, relying on the flow field to generate droplets. The pressures in this non-equilibrium system oscillate and the competition between these local fluid pressures drives the droplet formation. Increased pressure in the dispersed phase leads to limited flow of the continuous phase and a droplet forms at the junction. The pressure of the continuous flow increases until the dispersed phase droplet is pinched off. Now the flow of the continuous phase dominates and blocks the dispersed phase. The pressure of the dispersed phase increases again, the process continues *ad infinitum*.

The designs of the junction between the dispersed and continuous phase can be grouped into three categories depending on how the droplets are formed. The design can involve co-flowing streams (Figure 1.9a, Umbanhowar et al., 2000), or cross-flowing streams as in a T-junction (Figure 1.9b, Nisisako et al., 2002); alternatively, the junction can lead to flow focussing (Figure 1.9c, Anna et al., 2003).

In this thesis, flow focussing was used. Its advantage is that it can easily be tuned to produce different droplet sizes over a wide range (Xu et al., 2005). Theoretical work to characterise this method was done by Lee et al. (2009). They found four different droplet breakup regimes: (i) squeezing, (ii) thread formation, (iii) dripping and (iv) jetting. From the observations, the formulae given below concerning droplet size were identified.

A capillary number Ca was defined for planar flow focusing, given as

$$Ca \equiv \frac{\mu_c a \Delta U}{\sigma \Delta z} = \frac{\mu_c Q_c a}{\sigma h \Delta z} \left[\frac{1}{w_o r} - \frac{1}{2 w_c} \right], \tag{1.8}$$

where μ_c is the viscosity, Q_c is the volumetric flow rate of the continuous phase, σ is the equilibrium interfacial tension and the five geometric properties are depth of the channel, h, the width of the upstream channel for the continuous phase, w_c, and for the dispersed phase, $2a$, the distance between inlet channels and the contraction, Δz, and the width of the contraction, w_{or} (Figure 1.9c). Downscaled geometries lead directly to smaller droplet sizes d. This can be corrected for by a characteristic diameter D_h,

$$d \propto D_h \equiv \frac{2 w_{or} h}{w_{or} + h}. \tag{1.9}$$

Lee et al. (2009) show that d / D_h linearly decreases with larger capillary numbers Ca. From this it follows that larger values of μ_c, Q_c, w_c and a as well as smaller values of σ, h, w_{or} and Δz lead to smaller droplets.

Smaller capillary number Ca are associated both with a decrease in droplet size and a change in the regime from squeezing to dripping. In between these two extremes there is a flow condition named threading. This condition has the possibility to generate droplet sizes smaller than those formed under other regimes of flow focussing (Anna and Mayer, 2006; Ward et al., 2010). Critically, the droplet formation is now governed not by the capillary number linked to the upstream flow conditions but by the flow processes downstream of the mixing junction.

1.3.4 Single-Molecule Microfluidics

Biochemistry, in its quest to understand the workings of life, has its focus on the biological molecules as the building blocks of life. As the basic unit of all known

living organisms, the cell, is on the same magnitude as the microfluidic dimensions, biochemistry and microfluidics are well suited to be combined. Droplets, with similar dimension as cells could be used to support biochemical processes in a simplified environment on the same length scales as nature and with much greater control. This could prove especially useful for applications on the single-molecule level. Molecules could be confined to a small volume without restricting their local movement as done for TIRF microscopy. This would still allow for their observation over much longer time periods.

Several applications exist that combine the measurements of single molecules with microfluidics. For example, protein folding kinetics were observed using FRET (Lipman et al., 2003) and double-stranded DNA molecules (Wabuyele et al., 2001) were detected with a single-molecule burst technique. However, none of these employed encapsulation of reagents in droplets.

The first experiments using microfluidics involved assays for measuring the expression of β-galactosidase with single-molecule sensitivity (Cai et al., 2006), although performed in *E.coli* cells instead of droplets. Another early application in which a single molecule was confined to a droplet included a single DNA template trapped inside a picolitre volume. Fluorescent proteins were expressed, and detected, *in vitro*, from the encapsulated DNA molecule (Courtois et al., 2008).

Only recently have measurements of single-enzyme activity inside femtolitre volume droplets been reported (Arayanarakool et al., 2012), with a different group repeating similar measurements at higher droplet generating frequencies (Shim et al., 2013).

1.3.5 Emulsion Stability

A mixture of one fluid dispersed in form of droplets within another immiscible phase is called an emulsion. For an emulsion, the free energy of formation, ΔG_f,

is

$$\Delta G_f = \gamma A - T\Delta S_f, \qquad (1.10)$$

where A is the total surface area A, γ is the interfacial tension, T is the temperature and ΔS_f is the entropy of formation. ΔG_f is larger than zero, as the interfacial energy, γA, outweighs the entropy of formation, ΔS_f. Therefore, emulsions are not thermodynamically stable. The system will attempt to reduce the total free energy through increasing droplet size for a more energetically favourable surface-to-volume ratio. Emulsions, however, may be kinetically stable (i. e. metastable, or kinetically arrested) due to an energy barrier of steric or electrostatic nature (Taylor, 1998).

The two processes that lead to the break down of the emulsion are *Ostwald* ripening and coalescence. *Ostwald* ripening is the process that takes place when two droplets are in contact and molecules diffuse to the larger droplet, despite an intact separating barrier between the droplets (ibid.). Coalescence can happen following rupture of the thin film separating two droplets and one large droplet forms. Both mechanisms lead to an increase in the average size of droplets over time which is not desired and must be hindered to prolong the useful lifetime of the emulsion. This is achieved by the use of surfactants to decrease the surface tension.

1.3.6 Surfactants

Surfactants are surface-active substances that congregate at the water-oil interface, creating an oriented layer and lowering surface tension. This increases the lifetime of emulsions. Surfactants are amphiphilic with a hydrophilic head group and one or more hydrophobic tail groups (Figure 1.10a). They can be grouped

53
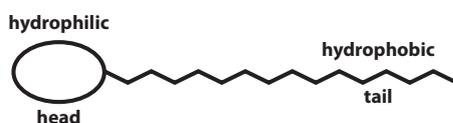
(a) Sketch of a surfactant
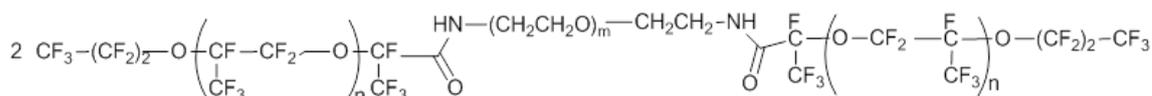
(b) Structural formula of a surfactant
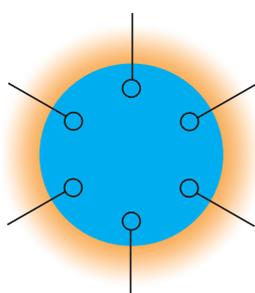
(c) Water-in-oil droplet
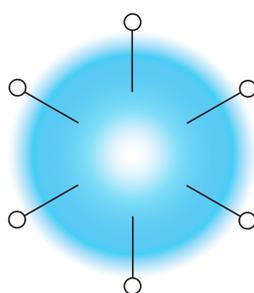
(d) Oil-in-water droplet
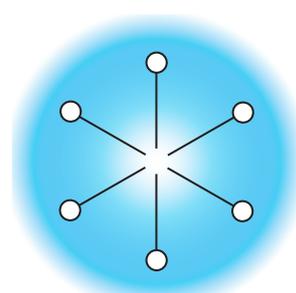
(e) Micelle
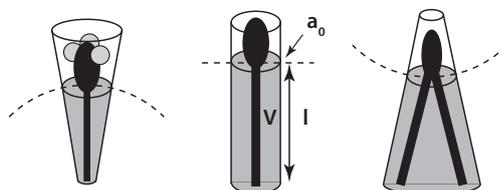
(f) Packing Parameter

Figure 1.10 – **(a)** General structure of surfactants with the head group on the left and the tail group on the right. - **(b)** Structural formula of a non-ionic fluorosurfactant. - **(c)** Water-in-oil droplet stabilised by a surfactant. The hydrophilic head group is inside. - **(d)** Oil-in-water droplet stabilised by a surfactant, with the head group now sticking out. - **(e)** Micelle formed by surfactant molecules that do not surround a droplet. - **(f)** The packing parameter depends on the volume of the hydrophobic chain V, the cross sectional area of the hydrophobic core a_0 and the length of the hydrophobic chain l. For smaller values of p, oil-in-water droplets are favoured, for larger values water-in-oil droplets.

by their head group into the four classes: (i) anionic, (ii) cationic, (iii) non-ionic or (iv) amphoteric. Classification of the tail group is more complex, as many more possibilities exist. Generally the tails are long chains of hydrocarbon or oxygenated or fluorinated hydrocarbons (Cullum, 1994; Myers, 2005).

To correlate surface activity and chemical structure, a qualitative tool named the hydrophilic-lipophilic balance (HLB) relates the molecular composition of a surfactant to its property for studying emulsions as a surfactant:

$$HLB = 20 * M_H/M \qquad (1.11)$$

where M_H is the molecular mass of the hydrophilic portion and M the molecular mass of the molecule. The HLB values can range from 0 to 20. Values below 10 are usually seen as indication for suitable formation of water droplets in oil. In this case, the hydrophilic head group would insert into a water droplet with the tail group(s) sticking out into a continuous phase of oil (Figure 1.10c). Values above 10 indicate that oil droplets will form in a continuous phase of water. The hydrophobic tail groups would insert into the oil droplets and the head groups would stick out into the water phase (Figure 1.10d). The formula is empirical and not all surfactants follow the guidelines for HLB values, but the model offers a first approach to the estimation of surfactant performance.

Another model involves the calculation of packing parameter p,

$$p = \frac{V}{a_0 l}, \qquad (1.12)$$

where V is the volume of the hydrophobic chain, a_0 the cross sectional area of the hydrophobic core and l the length of the hydrophobic chain. This packing parameter relates the geometry of the surfactant to its preferred arrangement. Lager head groups would prefer a positive curvature interface and therefore form oil-

in-water droplets, while smaller head groups would prefer a negative curvature interface and form a water-in-oil emulsion (Figure 1.10f).

Unfortunately, these formulae only hint at the possible behaviour of the surfactant as many other conditions play a role. There is no definitive way to correlate the structure of a surfactant to how it will work, and there is no direct way to design a suitable surfactant for a specific requirement. The relation of surfactant structure to behaviour is an ongoing field of study (Katritzky et al., 2010). While both HLB and packing parameters might provide a good starting point, the design of surfactants is still largely down to trial-and-error. Holt et al. (2010a) and Holt et al. (2010b) made some progress with surfactants specially targeted for microfluidic droplets with biological applications in mind.

In addition to the molecular structure, the surfactant concentration plays an important role as well. If the concentration is above the critical micelle concentration for the surfactant then micelles are formed in the continuous phase. Micelles are thermodynamically stable single or double layer structures of surfactant molecules that do not encapsulate a droplet and, therefore, prevent droplet stabilisation of emulsions (Figure 1.10e). If the concentration of surfactant is low, then not enough surfactant molecules are readily available to form the stable interface around the droplet.

Further things to consider are the compatibility of the surfactant with the experiments to be performed. It would be preferable to add the surfactant to the oil phase to prevent any interaction with the biological molecules inside the water phase. Once the droplet is formed, the surfactant should present an inert surface to the interior of the droplet to minimise interactions and reduce any artefacts (Baret, 2012).
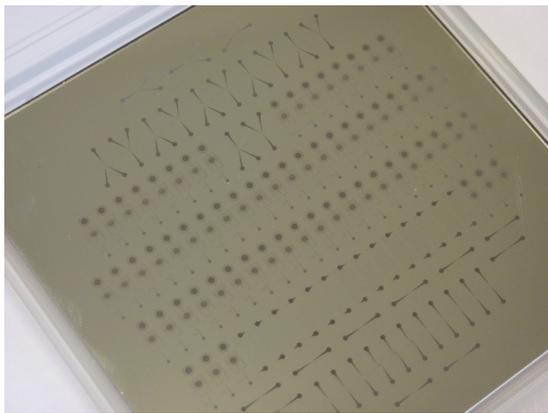
(a) High-resolution chrome mask
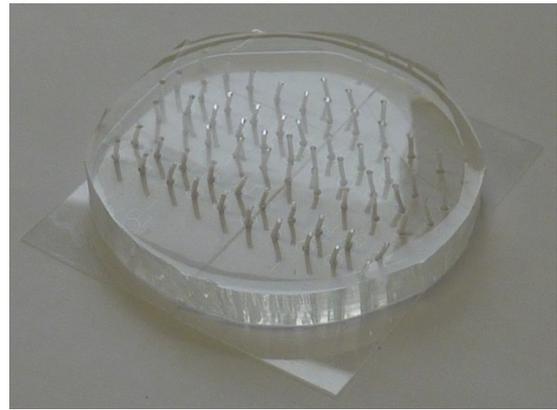
(b) Microfluidic device
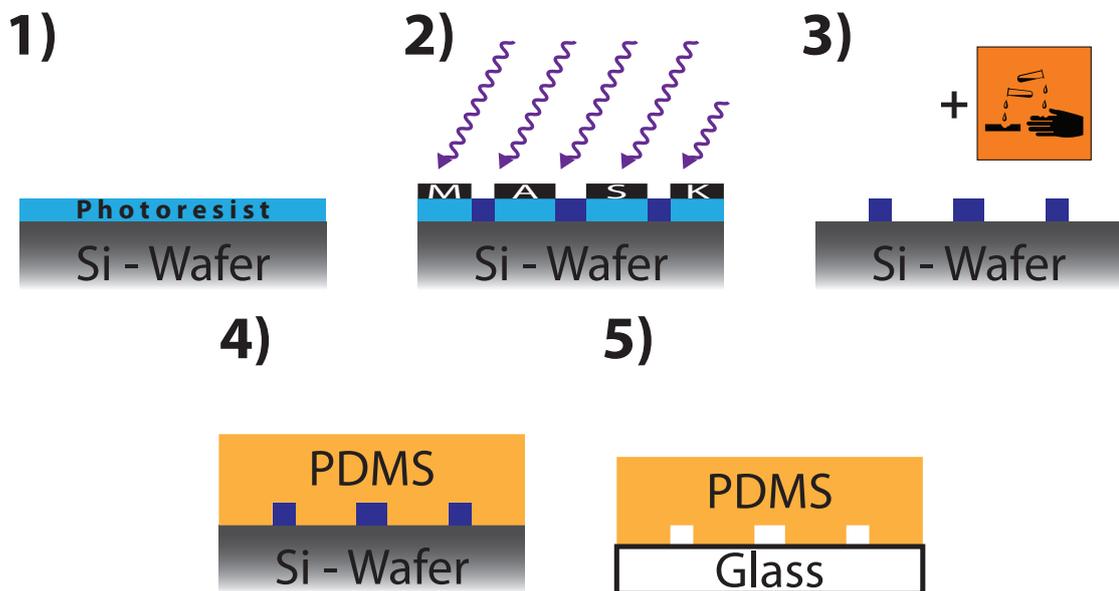
(c) Soft-photolithography

Figure 1.11 – **(a)** A chromium mask used in this work, size $4\,\text{in} \times 4\,\text{in}$, and a pixel reso-
lution of $200\,\text{nm}$. The smallest feature is $3\,\mu\text{m}$ while visible channels are
around $60\,\mu\text{m}$ wide. - **(b)** Final microfluidics device made from the mask
seen in (a). - **(c)** The soft-photolithography process. (1) A silicon wafer
is coated with the photoresist. (2) UV illumination followed by a post-
exposure bake transfers the mask pattern onto the photoresist. Transparent
parts of the mask have hardened photoresist. (3) The developing removes
the non-hardened photoresist. (4) PDMS is poured on top to create a nega-
tive mould of the wafer surface. (5) After peeling off the PDMS and bonding
to glass, the formerly transparent parts of the initial mask translate to chan-
nels in the microfluidic device.

1.3.7 Fabrication of Microfluidic Devices

A fabrication technique that allows fast, cheap and reliable creation of microfluidic devices is soft-photolithography. The method can be used to create many replicas of a microfluidic device. The master is made by photolithography and the replicas are made in Polydimethylsiloxane (PDMS). PDMS is a polymeric organosilicon compound, inert and non-toxic. It is suitable for use in combination with fluorescent microscopy as it is optically transparent with a low autofluorescence (Piruska et al., 2005). As PDMS can replicate feature sizes as small as 2 nm (Hua et al., 2004), the smallest achievable structures are limited by the photolithography process.

The photomask (Figure 1.11a) used for lithography consists of a printed pattern on either glass or an an emulsion film substrate. These options offer different resolutions for patterned structures (see *JD Photo-Tools Website*). The pattern consists of transparent and opaque regions representing the channels in the microfluidic device. The transfer of this pattern onto the silicon wafer by photolithography is done in five essential steps: (i) coating the substrate with a photoresist, (ii) soft bake, (iii) UV light exposure, (iv) post-exposure bake and (v) development (see *SU-8 2000 Datasheet*).

Silicon wafers are commonly used as substrates; they are readily available and provide an extremely uniform surface. First, the wafer has to be coated with a photoresist (Figure 1.11c - 1). The most widely used photoresist for microfluidics is SU8, an epoxy based photoresist available in a range of different viscosities that result in different thicknesses for the photoresist layer (see *SU-8 2000 Datasheet*). SU8 is spin coated onto the wafer. Depending on the viscosity of the photoresist and the set spin speed, the desired film thickness is uniformly achieved. This is followed by a soft bake (or pre-bake) to remove the excess photoresist solvent.

Long wave UV light, between 350 nm to 400 nm, is used to transfer the pattern from the photomask onto the wafer (Figure 1.11c - 2). For a negative photoresist,

such as SU8, illuminated parts form a strong acid while the non-illuminated parts remain unchanged. The following post-exposure bake induces an acid-catalysed cross-linking process, hardening the parts of the photoresist that were exposed to UV light.

The last step is the development. The wafer is immersed in the developer solution. For a negative photoresist, only non-illuminated parts would be dissolved and washed away, leaving behind the previously UV illuminated regions (Figure 1.11c - 3). Altogether, in this combination, transparent parts of the photomask leave features in the photoresist behind. For opaque parts of the mask, the bare substrate surface remains. The pattern of the photomask is transferred onto the wafer which becomes the mould for the subsequent production of microfluidic devices.

PDMS is used to create stamps from this mould (Figure 1.11c - 4). Cured PDMS is flexible and reversibly deformable for ease of handling. It is the base material of a two component adhesive together with a cross-linking or curing agent. Different ratios of base to agent lead to different elasticities in the cured material (Khanafer et al., 2009). Once mixed, the cross-linking process starts and the gel form is converted into a hydrophobic elastomer. This process works at room temperature but can be accelerated by heating. Plasma oxidation of the PDMS surface enables strong covalent bonds to be formed to a glass surface, i. e. Si—O—Si bonds (Bhattacharya et al., 2005). The tensile strength of the adhesive bond to glass is very high.

The unpolymerised PDMS is thoroughly mixed with the curing agent, poured onto the silicon wafer mould and afterwards degassed in a medium vacuum environment. This leaves the PDMS completely clear and transparent. Once the curing process is finished, it can easily be separated from the mould. The resulting PDMS block is now a negative cast of the silicon wafer. Transparent parts of the

initial mask were protruding features on the silicon wafer and are now recessed features in the PDMS. Bonding it to a glass by plasma oxidation creates channels formed by PDMS on three sides and glass on the fourth side (Figure 1.11c - 5).

The final microfluidic device (Figure 1.11b) can be additionally treated by washing it with a silanising agent to create hydrophobic or lipophilic channel surfaces.

1.3.8 Implementation of Microfluidic Devices

The microfluidic device made using soft-photolithography is integrated into a fluid delivery system incorporating pumps to push the carrier phase (oil) and the biological sample in aqueous solution into the microfluidic device. Flow parameters are carefully adjusted to yield the required droplet size. The generated droplets can be either observed directly inside the microfluidic device or collected and stored for later study. Several different methods exist to observe and interrogate the droplets, the most common method being by fluorescence microscopy (Lim et al., 2013; Ryu et al., 2011). The microfluidic devices are interchangeable in the setup and can be rapidly exchanged, greatly enhancing the number of possible applications.

1.4 Summary

In each human cell, many cellular machineries are at work. These can range from relatively simple to the most complicated, such as the spliceosome with hundreds of proteins that could all interact with each other. Parts of its function, such as splice site selection, are still a riddle wrapped up in an enigma even after years of research and there are no means to predict sites accurately that solely rely on the sequence found in the pre-mRNA.

SRSF1 is heavily engaged in splicing, as well as splice site selection. Various findings about its involvement sometimes seem to conflict with each other. The roles of its domains, the nature of its interaction, the number of involved molecules and its sites of phosphorylation and binding are unclear and further insight is desired.

As it is still unclear what parts are needed for consensus splicing, no minimal system exists that could be used to replicate splicing in an artificial system. Experiments are done in nuclear extracts with countless possible interactions of the dauntingly numerous proteins. This provides a challenging system to study.

Especially problematic is that different states of each machinery can be present at any given moment, as they all might progress at slightly different velocities and/or include different paths. All these deviations are easily hidden in ensemble measurements. Fluorescent-microscopy together with TIRF and colocalization provides a tool that can be used for single-molecule studies. These might provide essential insights in directly assessing heterogeneities. So far, no single-molecule studies have been done with SRSF1.

These traditional single-molecule studies have some inherent limitations. An alternative and possibly superior way might involve the use of microfluidics. Allowing the encapsulation of splicing components, on a single-molecule level in a droplet, would allow to overcome the currently necessary dilutions and the sur-

face attachment and allow the observation of splicing in real-time and in almost undiluted nuclear extract, providing a great step forward.

1.5 Aims

The objective of the research thesis was to characterise and optimise an existing TIRF microscope and improve the analysis of the microscopy data to provide a reliable and convenient methodology to study splice site selection. This will be employed to analyse the role of SRSF1 and other closely related proteins especially in the 5′ splice site selection during the early splicing complexes.

Alongside this work, the microfluidic methodology was developed. The goal here is to set up the means to fabricate microfluidic devices with feature sizes down to 1 µm ranges. There will be the need to develop a setup that allows convenient generation of uniformly-sized droplets. The intention is to separate droplets with diameters of 1 µm. These droplets will be collected and imaged by fluorescence microscopy. Fluorescent-labelled proteins are used to check the biological compatibility of water-in-oil emulsions and the surfactant.

At the end lies the possibility of bringing both parts together and observing splicing events in droplets and comparing the results obtained from both techniques.

Materials and Methods

2.1 *In vitro* Techniques

2.1.1 Polymerase Chain Reaction

All constructs were amplified by polymerase chain reaction from plasmids. PCRs were performed with G-Storm Thermal Cycler (G-Storm) and Red Taq DNA Polymerase (Sigma-Aldrich) according to the manufacturer's instructions. Reactions were checked on a 1 % agarose Gel containing 0.01 % EtBr for impurities and the correct size and phenol-chloroform extracted and ethanol precipitated.

2.1.2 RNA and DNA Purification

H_2O was added to the samples up to 100 µl. It was purified with the same amount of Phenol:Chloroform:Isoamyl (Sigma-Aldrich) followed by chloroform (Sigma-Aldrich). Sodium acetate was added to a concentration of 300 mM. The triple volume of ethanol was added for ethanol precipitation and the sample was centrifuged for 15 min at 16 200 g. The supernatant was removed and 190 µl ethanol was added followed by a final centrifugation for 5 min. After removing the supernatant the sample was kept under vacuum for 20 min and dissolved in 10 mM Tris pH 7.5 and 100 µM EDTA.

2.1.3 SDS PAGE Gel Electrophoresis

SDS PAGE gels were made of a resolving gel (370 mM Tris Base pH 8.6, 12 % acrylamide (ProtoGel, National Diagnostics), 0.1 % SDS and polymerised with 0.1 % AMPS and 0.14 % TEMED) and a stacking gel (125 mM Tris Base pH 6.8, 4 % acrylamide, 0.1 % SDS, and polymerised with 0.1 % AMPS and 0.2 % TEMED) on top. They were run in SDS PAGE running buffer (125 mM Tris Base pH 8.3, 0.96 M glycine, 0.1 % SDS) at 80 V until the samples were in the resolving gel

part and at 150 V for the remainder. Precision Plus Protein Kaleidoscope protein marker (Bio-Rad) was used as a marker.

2.1.4 Western Blot Analysis

Samples were incubated with an equal amount of loading buffer (400 mM Tris pH 8.6, 3.6 % SDS, 19.2 % glycerol, bromophenol blue, xylene cyanol, 1 % freshly added 2-mercaptoethanol) for 5 min at 80 °C and run on an SDS PAGE gel (Section 2.1.3). The gel was soaked in transfer buffer (48 mM Tris base, 39 mM glycine, 0.0375 % SDS, 10 % methanol). A stack saturated with transfer buffer consisting, from bottom to top, of 8 pieces of 3MM chromatography paper (Whatman), nitrocellulose (Amersham Hybond - ECL, GE Healthcare), the gel and another 8 pieces of 3MM chromatograhpy paper was run at 10 W for 25 min using a Biometra Fast-Blot B33 (Bio-Rad) to transfer the samples onto the nitrocellulose membrane.

The membrane was incubated overnight at 4 °C in blocking buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1 % Tween-20, 4 % milk (w/v) and spun down for 5 min at 4637 g). Afterwards it was incubated with the relevant primary antibody in blocking buffer for 1 h at 4 °C followed by 3 washes with the blocking buffer for 5 min. This was repeated with the secondary antibody. The membrane was developed using ECL reagent (Pierce) and the resulting chemiluminescence was detected with X-ray film (Fuji).

For fluorescent secondary antibodies, the membrane was washed additionally with 20 mM Tris-HCl pH 7.5, 150 mM NaCl at the end and a small amount of SDS was added to the incubations with the fluorescent antibody. The fluorescence was detected with an Odyssey CLx (LI-COR Biosciences).

Quantification of bands was done using OptiQuant (PerkinElmer)

66

2.1.5 Proteinase K Treatment

The proteinase K treatment of the splicing timepoints was done with $50\,\mu l$ of $0.4\,mg/ml$ proteinase K (Roche) in $100\,m_M$ Tris HCL pH 7.5, $12.5\,m_M$ EDTA pH 8.2, $150\,m_M$ NaCL, 1% SDS at $37\,°C$ for $10\,min$.

2.1.6 *In vitro* Transcription of RNA

A transcription reaction containing $40\,m_M$ Tris HCl pH 7.5, $20\,m_M$ MgCl$_2$, $10\,m_M$ NaCl, $10\,m_M$ DTT, $4\,m_M$ rNTPs, 0.05% Nonidet P-40, 5% T7 polymerase (1:20), $2\,units/\mu l$ RNaseOUT (Invitrogen) and $5\,ng/\mu l$ linear DNA template was set up at room temperature and incubated for $4\,h$ at $37\,°C$. $1\,unit$ DNase (Promega) was added and incubated for $30\,min$ at $37\,°C$ to degrade the remaining DNA. $2\,\mu l$ DNase Stop solution per $1\,unit$ DNase was added and incubated for $10\,min$ at $65\,°C$ to stop the enzymatic activity. The transcript was purified with MicroSpin G-50 columns (GE Healthcare), phenol-chloroform extracted and ethanol precipitated before being dissolved in $1\,m_M$ Tris and $100\,\mu_M$ EDTA.

2.1.7 *In vitro* Transcription of ^{32}P Radiolabelled RNA

The transcription reaction containing $5\,m_M$ DTT, $1\,m_M$ diguanosine triphosphate sodium [G(5′)ppp(5′)G] (GE Healthcare), $0.5\,m_M$ rATP, $0.5\,m_M$ rCTP, $0.5\,m_M$ rUTP, $0.05\,m_M$ rGTP, $40\,m_M$ Tris HCl pH 7.5, $6\,m_M$ MgCl, $10\,m_M$ NaCl, $0.33\,\mu_M$ $[\alpha - ^{32}$ P] rGTP ($10\,mCi/ml$, $3000\,Ci/mmol$, Perkin Elmer), $2\,units/\mu l$ RNase-OUT (Invitrogen) and 5% T7 polymerase (1:20) was set up at room temperature and incubated for $1\,h$ to $2\,h$ at $37\,°C$. The transcripts were gel purified, ethanol precipitated and re-suspended in $1\,m_M$ Tris and $100\,\mu_M$ EDTA.

2.1.8 *In vitro* Splicing

Splicing reactions containing approximately 1 nM of radiolabelled RNA, 20 mM CrPi, 3.2 mM MgCl$_2$, 20 mM HEPES pH 7.5, 1.5 mM rATP, 50 mM KGlu, 3 % RNase-Out (Invitrogen) and 50 % nuclear extract were set up typically in a final volume of 10 μl at 4 °C and incubated for 2 h at 30 °C. At designated timepoints, aliquots of 2 μl were removed and placed in a microtitre plate on dry ice. All aliquots were proteinase K treated. Followed by ethanol precipitation, 10 μl formamide containing 10 mM EDTA pH 8 and 0.2 % xylene cyanol and 0.2 % bromothenol blue was added. The samples were heated to 80 °C for 30 s before being run on a 6 % denaturing polyacrylamide gel. The gel was dried and exposed to a phosphorimage screen. Quantification of mRNA bands was done using OptiQuant (PerkinElmer).

2.1.9 Native Agarose Gel Electrophoresis

Splicing reactions were done as described in Section 2.1.8. Complexes other than complex E were additionally treated with 2 μl of 4 mg/ml heparin and incubated for 30 min at 30 °C. An equal volume of 50 mM Tris, 50 mM glycine and 40 % glycerol was added and the samples were run on a native agarose gel (1 % to 2 % w/v UltraPure low melting point agarose (Invitrogen), 50 mM Tris, 50 mM glycine) for 5 h to 6 h at 4 °C. The gel was compressed between two 3 mm thick paper sheets and a stack of paper towels overnight, then dried and exposed to a phosphorimage screen.

2.2 Cell Culture and Nuclear Extract Preparation

2.2.1 Transfection of HeLa Cells

Two dishes with each an area of $176\,cm^2$ were seeded with 4×10^6 HeLa cells each and grown for 24 h. They were transfected using FuGENE 6 (Roche) according to the manufacturer's instructions and cultured for further 48 h to 72 h before being harvested.

2.2.2 Nuclear Extract Preparation

Cells were washed with chilled PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) once before being harvested with a cell scraper. The cells were centrifuged at 150 g for 8 min. The cell pellet was washed in 1 ml of PBS and centrifuged at 4 °C at 9250 g for 5 min. The washed cell pellet was re-suspended in a volume equal to its packed cell volume (PCV) of 10 mM Hepes pH 8, 1.5 mM $MgCl_2$, 10 mM KCl and 1 mM DTT. It was left for 15 min on ice to swell the cells. Cells were lysed by forcing them 6 times through a 25g needle. The lysed cells were centrifuged at 9250 g at 4 °C for 1 min and the pelleted nuclei were re-suspended in two-thirds of its PCV of 20 mM Hepes pH 8, 25 % glycerol, 420 mM NaCl, 0.2 mM EDTA and 1 mM DTT. They were incubated for 30 min at 4 °C on a magnetic stirrer with an added small magnetic flea. The extract was centrifuged at 9250 g at 4 °C for 5 min. The supernatant was dialysed against 20 mM Hepes pH 8, 10 % glycerol, 100 mM KCl, 0.2 mM EDTA and 1 mM DTT for 2 h at 4 °C.

The dialysed nuclear extract was centrifuged for 5 min at 9250 g and snap-frozen in liquid nitrogen and stored at −80 °C. All nuclear extracts were tested to show a sufficient splicing efficiency.

Plasmids expressing mEGFP fused to SRSF1, U1A and mutants thereof were made by Dr L.P. Eperon, Dr M.J. Hodson, T. Webber and B. Hutchinson. The

nuclear extracts containing the expressed proteins from these plasmids were made

by Dr L.P. Eperon and C.M. Lucas.

2.3 Single Molecule Techniques

2.3.1 Labelling of the pre-mRNA

Pre-mRNA transcripts for single molecule experiments were labelled by annealing an oligonucleotide containing the fluorophore Cy5 (Table 8.8) to them. Reaction were set up in silanised tubes and done in 100 mM NaCl and 10 mM HEPES pH 8 with 1.67 µM pre-mRNA and 1.67 µM of the Cy5 oligo. They were incubated for 5 min at 80 °C in a heat block and then cooled down for 1 h to 40 °C in same switched of heat block. 1 µl of each sample was mixed with 4 µl H_2O and 5 µl formamide containing 10 mM EDTA pH 8 and run on a native 6 % acrylamide gel and imaged by the Typhoon Trio+ (GE Healthcare). Samples had to have less than 2 % of free fluorescent oligos to be used, otherwise the annealing was done anew with adjusted concentrations of pre-mRNA.

2.3.2 Sample Chamber Preparation

Cover slides, 22 mm by 50 mm #1 (Menzel-Gläser), were washed with distilled water, followed by purified water and sonicated in a water bath for 12 min and dried under a nitrogen stream. They were cleaned in an argon plasma (MiniFlecto-PC-MFC, Gala Instrumente) 5 times for 5 min each with pure argon at 0.15 mbar and an applied power of 80 W. Double-sided tape was used to create a 5 mm to 10 mm wide channel parallel to the cover slide, covered by another cover slip, 22 mm by 22 mm #1.5, to form the sample chamber.

The sample chamber was incubated with 20 µg/ml biotin-BSA in PBS for 10 min, washed with buffer A (100 mM NaCl, 50 mM Hepes pH 7.5, 1 µM DTT, 20 units/ml RNase OUT (Invitrogen)), incubated with 10 µg/ml streptavidin (Invitrogen) in PBS and again washed with buffer A.

2.3.3 Preparation of Samples for the Splicing Complexes E, A, and C

All samples were prepared with 50 % nuclear extract, 3.2 mM MgCl$_2$, 50 mM KGlu and 3 units RNase OUT (Invitrogen). This was used for complex E, as it did not contain any ATP. 20 mM CrPi, 1.5 mM rATP and 20 mM Hepes pH 7.5 were added to progress splicing past complex E. To halt splicing in complex A, either 1 µM αU6 or 570 nM anacardic acid was added. For complex C, the Globin C GG pre-mRNA was used where the 3′ splice site sequence was replaced by a random sequence. To block the U1 snRNP, 3.3 µM αU1 oligo was added. Experiments with PhosStop (Roche) had it at a two times final concentration of its suggested stock concentration.

Samples were pre-incubated for 15 min at 30 °C before pre-mRNA was added for a final concentration of 50 nM and incubated for further 15 min at 30 °C (30 min for the formation of complex C).

2.3.4 Single Molecule Sample Preparation

Serial dilutions of the samples were prepared down to a concentration of 5 pM of the pre-mRNA using buffer A (100 mM NaCl, 50 mM HEPES pH 7.5, 1 mM DTT, 0.02 units/µl RNaseOut). The most dilute samples were loaded into the sample chamber and incubated for 5 min. The density of the apparent spots was judged, and, if considered too low, a higher concentrated samples was loaded. Usual final sample concentrations were around 50 pM. The imaging buffer (2 mM PCA, 90 nM PCD, 1 mM DTT, 0.16 units/µl RNaseOut) was then flushed into the sample chamber and incubated for 5 min.

2.4 Microfluidics

2.4.1 Designing the Photolithographic Mask

The photolithographic masks were designed with Adobe Illustrator (Adobe Systems) and converted to the Gerber file format using LinkCAD (Bay Technology). A chrome covered glass mask 4 in by 4 in was fabricated by JD Photo-Tools.

2.4.2 Fabrication of the Wafer

Polished silicon wafers (Microchemicals and Siltronix) with a diameter of 2 in were cleaned with compressed air to remove any potential dust. The permanent epoxy negative photoresist SU-8 2002 (MicroChem) was coated onto the wafer with a spin coater (Model WS-650MZ-23NPP/UD3B, Laurell Technologies Corporation) with an initial speed of 500 rpm for 30 s and a final speed of 1500 rpm for 30 s. It is followed by a soft bake of 1 min at 95 °C on a hotplate (SD160, Stuart, Bibby Scientific). The UV illumination for the exposure was set to 10 % for 36 s. The post bake was done for 6 min at 95 °C. The wafer was developed by placing it in Microposit EC solvent and gently shaking the containing beaker for a total of 10 s before washing it with acetone.

2.4.3 Fabrication of the Microfluidic Device

The patterned wafers were placed in a plastic Petri dish and any dust was removed by blasting it with a strong nitrogen stream. PDMS mixtures with base to curing agent ration of 7:1 to 15:1 were prepared by 2 min of stirring. The PDMS mixture was poured over the wafer to a thickness of 6 μm to 8 μm and degassed in a vacuum chamber until no visible gas bubbles remained. The Petri dish containing the wafer and the PDMS mixture was heated to 65 °C for 2 h or overnight in an

oven (Kelvitron T, Heraeus).

The solid PDMS block was cut along the outline of the final microfluidic device and carefully separated from the wafer. Holes were punched straight through the PDMS block from the side of the imprinted pattern using a biopsy puncher with a diameter of 0.7 mm. The punched-out PDMS cylinders were removed from the other side with tweezers. The patterned side of the PDMS block was cleaned by attaching and removing tape (Scotch Magic, 3M) multiple times.

Multiple cover slides, 22 mm by 50 mm, #1 (Menzel-Gläser), their number depending on the size of the microfluidic device, and the PDMS surface were oxygen plasma activated for 1 min at 0.1 mbar and 28 W of applied power (MiniFlecto-PC-MFC, Gala Instrumente). The PDMS was covered in methanol and the glass slides were firmly pressed against it from above. Trapped bubbles were carefully removed by brushing them sideways. The device was placed in an oven at 65 °C for 1 h.

The channels of the microfluidic device were flushed with a 1:100 mixture of 3-Aminopropyltriethoxysilane and Fluorinert FC-40 (3M) for those with the 5 μm junction. The device was left overnight for the FC-40 to evaporate.

2.4.4 Droplet Generation

The microfluidic device was mounted on a light microscope using epi-illumination and a variable magnification ranging from 0.7 x to 5 x. The dispersed phase was distilled water or 0.5 % NP-40 and 0.01 M $MgCl_2$ with or without fluorescent samples. The continuous phase was 1 % to 10 % of surfactant in FC-40, vortexed for 30 s and sonicated for 30 s. The tubing (Microbore PTFE Tubing, 0.022 in ID x 0.042 in OD, Cole-Parmer) was first connected to the 1.5 ml reservoirs of the pressure-driven flow controller (MFCS, Fluigent) and flushed to remove any air or other residual contaminants. The flushed tubing was connected to the respective

inlets. The pressures to drive the flows were adjusted until the desired droplet size was achieved. The outlet was connected to a vial to collect the generated droplets for the designs without an internal reservoir.

The visualisation was done with a camera (acA2000-340kc, Basler) which was operated using a self-written LabView (National Instruments) program that allowed to change exposure time, electronic enhancement and resolution during run-time and to acquire videos or single images.

2.4.5 Fluorescence Microscopy of Droplets

Fluorescence measurements with droplets were done on an inverted microscope with an 100x/1.25 NA oil immersion objective (Nikon). White bright-field illumination was done with an LED D1 (Thorlabs) diode mounted above, illumination by 488 nm from a diode laser (JDSU) focused by the objective onto its focal point. Images were filtered with a 488 nm notch filter and acquired using an iXon EM-CCD camera (Andor). The acquired data was saved in the Tagged Image File Format (TIFF) and single images and intensity time traces were obtained manually using Fiji (http://fiji.sc).

From the Microscope to the Final Data

The aim was to set up the means to investigate the interaction and binding distribution of proteins and pre-mRNA in splicing by colocalization measurements and the analysis of intensity time traces via total internal reflection microscopy with three spectroscopically different colours.

The performance of the microscope has to be optimised and its characteristic determined. The acquired data has to be analysed in a reliable way to accurately gain the desired information. The whole process from performing the experiment to obtaining the final data should be streamlined to ensure an efficient analysis and being accessible to allow a broad use by non-specialists.

3.1 Three-Colour Fluorescence Microscopy

The use of three different fluorophores at the same time allows to investigate the interaction of three different components in a splicing reaction. The chosen fluorophores are mEGFP (from here on referred to as GFP), mCherry and Cy5 and provide a sufficient spectral separation. The available diode lasers provide the wavelengths of 488 nm, 532 nm, 561 nm and 633 nm. Of those, three wavelengths are used for excitation. The wavelength of 488 nm is used to excite GFP, of 561 nm to excite mCherry and of 633 nm to excite Cy5. The wavelength of 488 nm does not only excite GFP but also, to a lesser extend, mCherry. The same is true for the wavelength of 561 nm that not only excites mCherry but also, again to a lesser extend, Cy5 (Figure 3.1).

It was decided to use a sequential illumination as there is no need for a simultaneous illumination for a static system. The longest wavelength is used for the excitation until all corresponding fluorophores are bleached. Any subsequent illumination with a shorter wavelength cannot excite these fluorophores as they are now inactive. This procedure is followed by the next shorter wavelength until the shortest wavelength is reached. This allows images to be obtained for the different fluorophores that are captured on the same camera using the same beam path without any detectable cross-talk. This reduces the variation of the acquired data and simplifies the later data analysis.

For the three color setup, an excitation with the 633 nm laser is used first to excite Cy5. The excitation wavelength is unable to excite mCherry or GFP and the collected fluorescence signal originates from Cy5 alone. This is done until all Cy5 fluorophores are bleached. The excitation switches to the 561 nm laser to excite mCherry. This excitation wavelength is unable to excite GFP but could also excite Cy5, if it were not inactive at this stage of the experiment. Therefore, visible fluorescence only stems from mCherry. This is done until all the mCherry
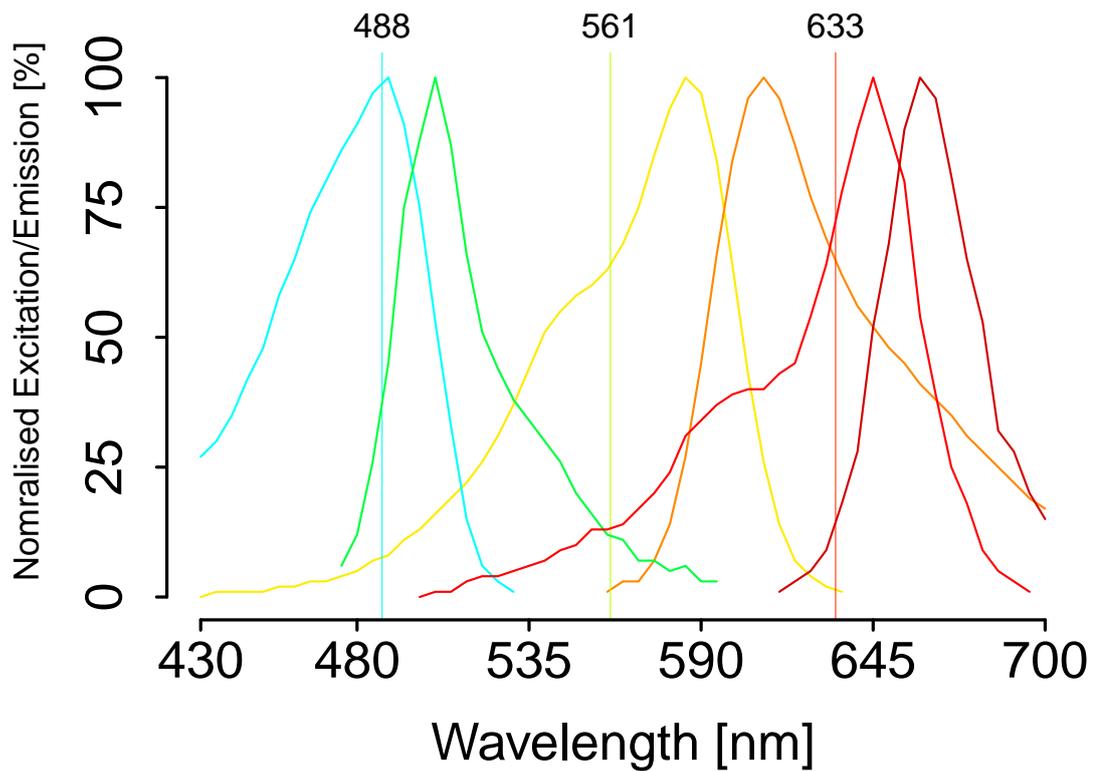
Figure 3.1 – The normalised spectra from left to right for the excitation (▮) and emission (▮) of GFP, for the excitation (▮) and emission (▮) of mCherry and for the excitation (▮) and emission (▮) of Cy5. The three used laser wavelengths to excite the fluorophores are shown as lines at 488 nm, 561 nm and 633 nm. The excitation with 488 nm for GFP can also excite mCherry, and the excitation with 561 nm for mCherry can also excite Cy5. Data obtained from *Life Technolgies Fluorescence SpectraViewer*

fluorophores are bleached and the excitation is changed to the 488 nm laser for the final excitation of GFP. All other fluorophores are bleached and the collected fluorescence originates solely from GFP.

3.2 The Microscopic Setup

The setup is an inverted microscope suitable for an excitation by total internal reflection. The encircled numbers correspond to the numbers in Figure 3.2.

Four diode lasers are available for the wavelengths of 488 nm, 532 nm, 561 nm and 633 nm ①. The different laser beams are fed to the beam combiner that aligns all laser beams along the same beam path ②. An optical fibre connects the output from the beam combiner with the microscope ③.

The rotatable half-wave plate allows the plane of polarisation of the polarised laser light to be rotated to any other desired plane ④. The polarising beam splitter separates the incoming beam into two beams with orthogonal polarisation, of which only the transmitted beam is used ⑤. Together with the rotating half-wave plate allows the polarising beam splitter the intensity of the laser beam to be adjusted. The resulting laser beam has a linear polarisation. The quarter-wave plate changes the polarisation of the laser beam to a circular polarisation ⑥.

The beam extender widens the incoming laser beam from a diameter of less than 1 mm to a diameter of 15 mm ⑦. The diaphragm block all light that does not pass through its aperture of a diameter of 9 mm ⑧. The laser beam has a Gaussian intensity profile before the aperture. Widening the beam and using only its centre part provides a laser beam with a more homogeneous intensity profile.

The total internal reflection lens ⑨ focusses the laser beam onto the back focal plane ⑫ of the objective. This causes the laser beam to be collimated again at the focal point of the objective at the interface to the sample ⑬ so that the laser beam homogeneously illuminates an area with a diameter of around 25 μm. The adjustable mirrors help to guide the laser beam into the objective perpendicular to the axis of the objective ⑩. The first total internal reflection mirror sits directly below the objective and reflects the excitation beam into the objective as far out as possible ⑪. The so guided laser beam reaches the sample interface at an angle

①	Diode lasers		
②	Beam combiner		
③	Multi-mode fibre	⑩	Adjustable mirrors
④	Half-wave plate	⑪	TIR mirror
⑤	Polarizing beam splitter	⑫	Back focal plane
⑥	Quarter-wave plate	⑬	Sample
⑦	Beam extender	⑭	Variable notch filters
⑧	Diaphragm	⑮	EMCCD
⑨	TIR lens	⑯	TIR mirror
		⑰	Quartered photodiode
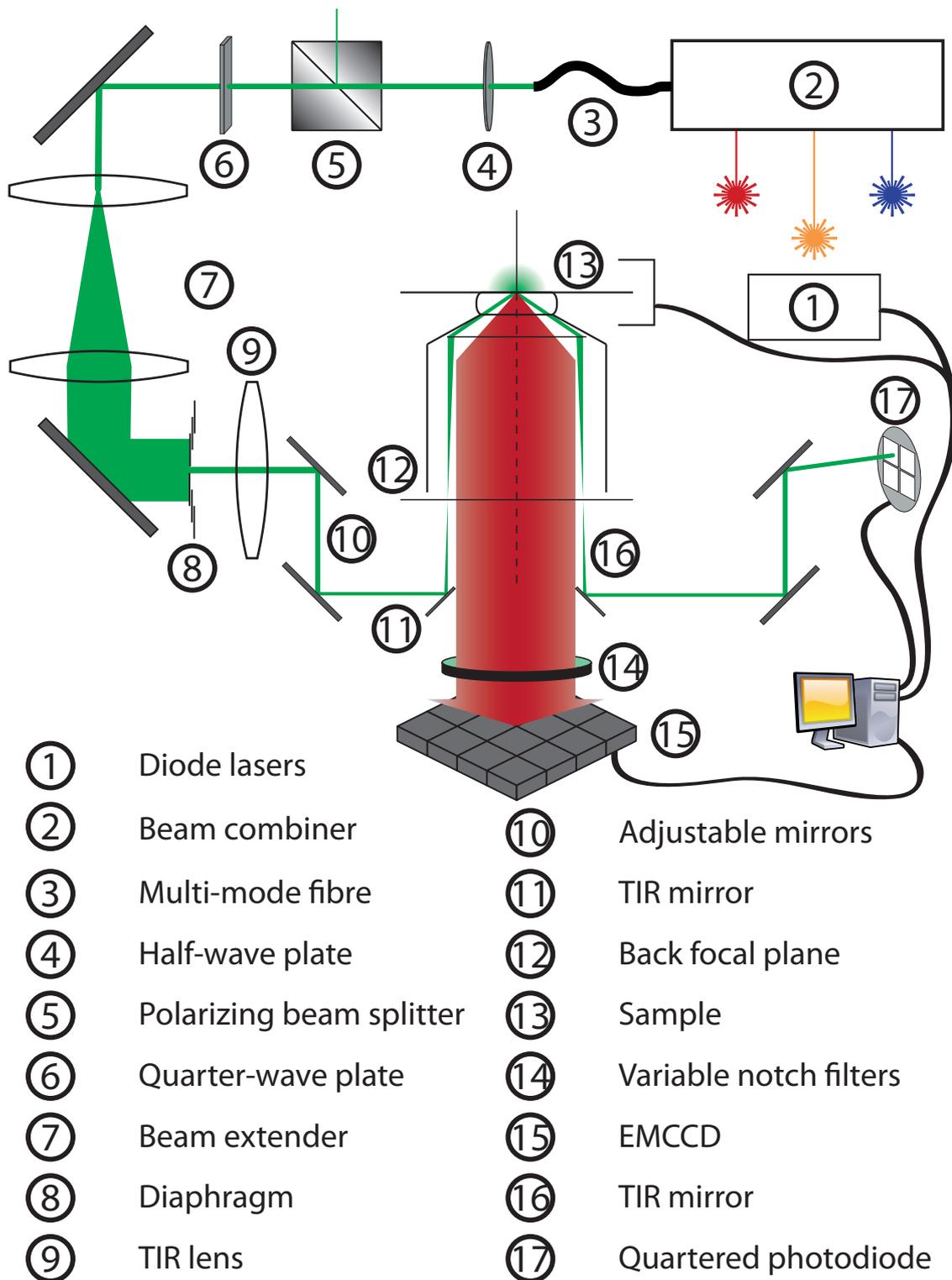
Figure 3.2 – Schematic of the total internal reflection microscope. The excitation pathway is shown in green, the emission pathway is shown in red.

greater than the critical angle and total internal reflection occurs. An evanescent wave penetrates the sample and excites it ⑬.

The emission is collected by the objective and passes through several notch filters suited for the wavelengths of the laser beams used for the excitation ⑭. An EMCCD records the image that corresponds to the interface where the total internal reflection occurrs ⑮.

The total internal reflected laser beam that excites the sample returns through the objective and is separated from the emission by a second small mirror ⑯. The beam is directed onto a quartered photodiode ⑰ and used to keep the sample in focus (see Figure 3.5).

3.3 Background

Optimising the setup both in terms of detection efficiency and towards the lowest possible background and noise levels ensures a more reliable identification of spots, colocalization and bleaching steps. Crucial for this is the camera and its settings. The microscope uses an iXon EMCCD camera from Andor. A feature of EMCCD cameras is the electron multiplying structure that gives an additional gain of intensity for registered photons, greatly increasing the signal. The highest available setting was always used.

An important camera setting is the exposure time. The longer the exposure time, the more photons can be collected per readout. Thus the longest exposure times would be favoured to obtain clearer images. As one task is to identify steps in the intensity time traces obtained from fluorescent spots, the time resolution of the acquisition has to be good enough to allow these events to be resolved. Exposure times of $0.1\,$s were found to provide a good compromise between these two requirements.

Other available camera settings concern the readout of the information from the camera chip. Two bit depths are available, 14 bit and 16 bit. Using a bit depth of 16 bit allows the occurrence of higher photon intensities per pixel to be recorded. Only a bit depth of 14 bit supported a sufficient fast camera readout that allowed exposure times of $0.1\,$s to be used. As photon intensities requiring the 16 bit depth did not occur, 14 bit could be safely used.

The other options are the horizontal and vertical shift speed, the vertical clock voltage and a pre-amplifier gain. Better performing settings compared to the standard settings were found by evaluating sample acquisitions with the camera shutter closed. With these settings, the mean intensity could be brought down to the camera baseline and no significant intensity spikes occurred (Figure 3.3).

Operating the camera with an open shutter under normal experimental condi-

(a) Mean values for the measured background
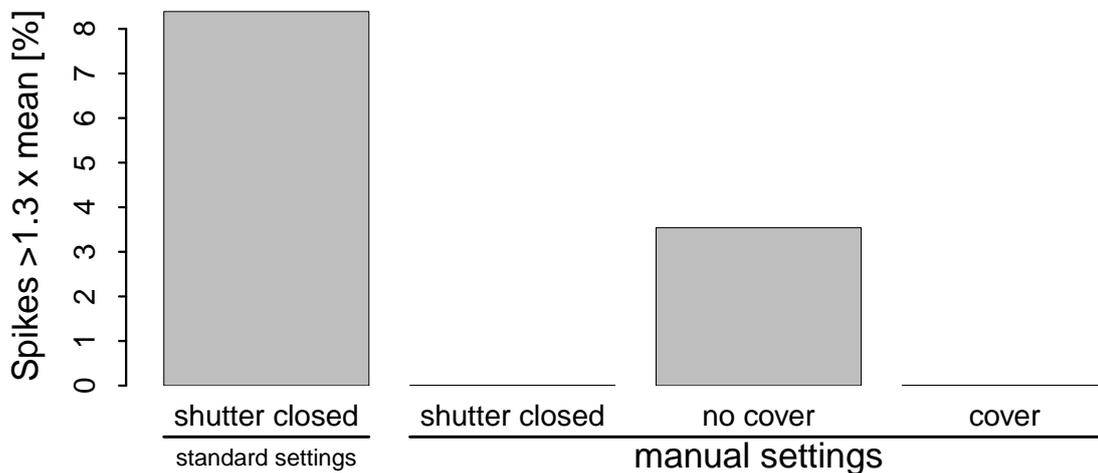
(b) Occurrence of spikes

Figure 3.3 – 50 frames were acquired at $-80\,°C$ with a gain of 300 and analysed. The settings and conditions from left to right are: with the camera shutter closed and standard camera settings, shutter closed and optimised settings, shutter open with no cover over the microscope and optimised settings, and shutter open with the cover and optimised settings. **(a)** Mean values of the pixel intensities. The camera baseline output is 100. Optimised manual camera settings reduce the mean background close to the baseline. Background acquisitions show an increased background intensity that can be obviated by the use of a cover over the microscope. - **(b)** Occurrences of spikes which were defined as pixel intensity values 1.3 times higher than the mean value as a ratio of all evaluated pixels. Both the standard camera settings and the lack of a cover over the microscope result in a high incidence of spikes. Optimised settings and the use of the cover reduce this amount to virtually 0.
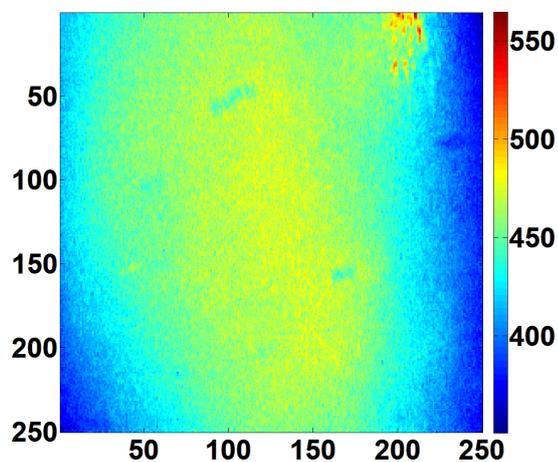
(a) Illumination with 633 nm
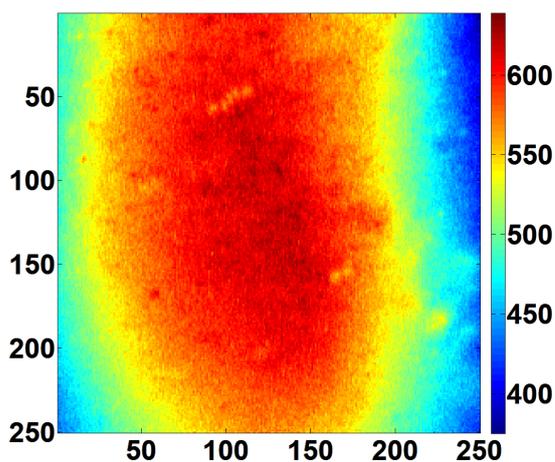
(b) Illumination with 488 nm
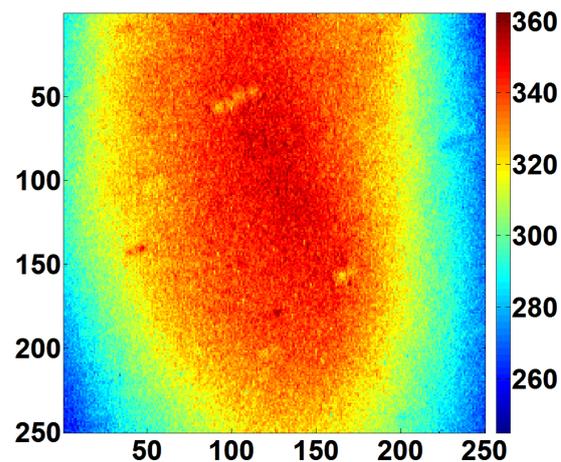
(c) Illumination with 532 nm

Figure 3.4 – 50 frames were acquired at $-80\,°C$ with a gain of 300 and analysed. Shown are the mean pixel intensity values for each pixel. **(a)** Background for the illumination with 633 nm. - **(b)** Background for the illumination with 488 nm. - **(c)** Background for the illumination with 561 nm.

tions showed an increase in both mean intensity and occurrence of spikes. This was traced down to unavoidable ambient light. Placing the microscope inside an opaque cover prevented the ambient light from interfering with the acquisition and led to as good conditions for the acquisition as for the operation with the camera shutter closed (Figure 3.3).

All these optimisations so far were done without any lasers turned on. Illumination by objective-based total internal reflection leads to an characteristic and unavoidable background. Ideally, this background would be homogeneous across the whole field of view. The laser beam has an approximated Gaussian intensity profile, resulting in the similarly shaped intensity profile across the field of view. Using only the central part of the laser beam as described in Section 3.2 lessens this effect but leads to a smaller illuminated area. Due to this, for all acquisitions only the central 250 pixels by 250 pixels part of the 512 pixels by 512 pixels CCD chip is used. The resulting field of view is smaller but with a homogeneous intensity distributions. Figure 3.4 shows exemplary acquisitions for the different lasers.

The background for the illumination with 488 nm and 532 nm differs by only 50 % across the acquired area with a few local variations stemming from the camera itself, albeit they are not prominent enough to disturb the later detection of spots. Only in the corner of the acquired area for the illumination with 633 nm appears a disturbance of the background that can be mistaken for a fluorescent spot. It originates from a non-elastic scattering within the excitation pathway that could not be eliminated entirely. This area was ignored for the later analysis.

3.4 Positional Stability of the Sample

Fluorescence image acquisitions are done over times of up to 3 min. The position of the sample should not change in this time. Lateral changes of the position of the sample could be tracked but would make the data analysis needlessly complicated. Changes of the position in the vertical plane would move the sample out of the focus of the objective and are harder to compensate for. No lateral movements of the sample were detected over the course of 3 min. Movements in the horizontal direction occurred frequently and were sometimes noticeable within a few seconds.

The sample chamber sits on top of a small film of viscous immersion oil on top of the objective. The immersion oil shows slow relaxation times due to its high viscosity and this leads to vertical shifts of the sample chamber out of the focus of the objective over time. Given enough time, the system would find an equilibrium with a stable position. Frequent movements of the sample chamber to image different areas of it make this option infeasible. Additionally, the focus shifts slightly for different excitation wavelengths. The image acquisition for each fluorophore and its corresponding wavelength of excitation should be performed in focus. An automatic correction of the sample position to support this would be preferred.

Any change in the horizontal position of the sample interface is accompanied by a displacement of the same magnitude of the reflected beam (Figure 3.2). The beam returns through the objective and the displacement is magnified. This beam is directed onto a quartered photodiode to measure its deviation (see Figure 3.5). A previous calibration allows to correlate the displacement of the beam on the photodiode to the horizontal shift of the sample interface. Knowing the deviation of the beam from the "in-focus" position allows one to correct for any occurring deviations. The correcting algorithm was implemented in LabView (Section 3.5).
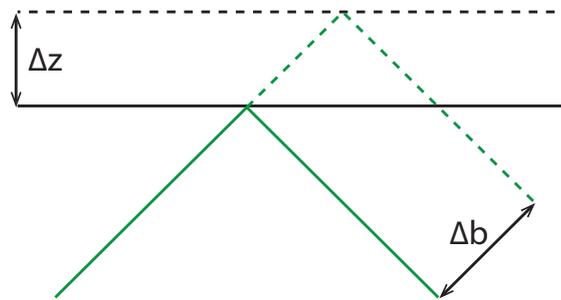
Figure 3.5 – The incoming laser beam is reflected at the interface with the angle of refraction equal to the angle of incidence (——). A horizontal shift Δz of the interface results in a lateral shift Δb of the reflected beam (- - -).

3.5 Automated Data Acquisition with LabView

One experiment at the microscope consists of several sequential acquisitions. For each acquisition, the sample has to be moved to image a new part of it. The first laser is turned on for a certain amount of time before the illumination is switched to a second laser and possibly third one. The sample has to be kept in the correct focus for each laser during the whole acquisition. The acquired data has to be saved in an annotated and recognisable way. The sample is then moved to a new position and the whole process is repeated multiple times until enough data is obtained. Automating and supporting this process via a computer would reduce the time required to perform one experiment and increase the consistency across experiments. A LabView program was written to control the microscope and allow an automatic data acquisition once an initial setup was performed. The encircled numbers correspond to the numbers in Figure 3.6.

The temperature of the CCD chip of the camera can be chosen manually and the cooling process is automatically initiated once the program is started. A coloured rectangle switches from red to green once the target temperature is reached and stabilised ①. Exposure time and EM gain of the camera can be chosen. The camera settings can be either set manually ⑤ or the optimised settings from Section 3.3 can be used ②.

Each laser can be switched on and off individually and its power adjusted. Focus points can be saved to be used in conjunction with the autofocus ③. The camera chip has 512 pixels by 512 pixels. Only parts of it can be used or multiple individual pixels can be binned ④. If an acquisition is in progress, each recorded image is shown in a false colour image ranging from black over blue to white for increasing intensities ⑦. Supplementary information about the current acquisition is shown below ⑧.

The piezoelectric stage holding the sample can be moved in all three dimensions

Figure 3.6 – Interface of the LabView program. Numbers were added for annotation.

with sub-micrometre precision ⑩ . Quick vertical adjustments can also be done with a slide ⑨ . The output of the quartered photodiode ⑫ , (Figure 3.2) is used for the feedback that adjusts the focus automatically once initialised ⑪ (see Section 3.4), using the focus points set at ③ .

Each acquisition can be performed manually but can also be automated if wanted. The change of lasers and the end of an acquisition can be set to occur at certain frames. The stage can be allowed to move within its limits to repeat the acquisition process nine times, automatically saving the data for each one ⑭ . The only actions still needed to be performed manually are the initial mounting of the sample and moving the sample to allow the imaging of a new area for every nine acquisitions.

Acquisitions are saved as multi-page Tagged Image File Format (TIFF) files with 16 bit/pixels. Each file name is automatically complemented with the current date and time to allow a later independent assignment to experiments and to prevent the accidental overriding of data. Each acquisition is accompanied by a text file that describes the experiment. Optional experimental notes ⑬ , the gain of the EMCCD camera, the exposure time, the lasers used and for each acquired frame the laser power used for each laser and the deviation of the reflected laser beam are recorded in this text file.

3.6 Colour Separation and Composite Images

The acquired data consists of sequential intensity images (Figure 3.7a). This data alone does not contain any information about the wavelength used for the current excitation. Additional information is required to assign a wavelength to each image. This information can be gained by inspecting the raw data and knowing the experimental procedure or from the accompanying text file (see Section 3.5). The information from all the images corresponding to one wavelength can now be used to create composite images. This reduces the background noise seen in a single image and highlights individual spots.

Two different ways were used in this work to create these composite images. One can compute the mean or the maximum value for each pixel across all images of a wavelength. The resulting composite images can vary slightly but significantly (Figures 3.7b and 3.7c).

A spot that has a high intensity but only emits fluorescence for a short amount of time compared to the overall acquisition time will be more easily recognisable if the maximum values are computed as its maximum intensity values are much higher than the surrounding background values. It will only be slightly above background if the mean values are used as the short time with a higher intensity at the beginning has a negligible influence on the overall mean. A spot with lower intensity emitting fluorescence over the whole acquisition time will be more easily recognisable if the mean values are computed as its intensity is consistently higher than the surrounding background. This might not be the case if the maximum values are computed. Fluctuations in the surrounding background might be of the same intensity as the spot and it might therefore not be recognisable.

In most cases, spots can be identified in both of the composite images, but some might be missed. To avoid the bias towards one group of fluorescent spots, both ways are used in conjunction to reliably detect spots that are both short-lived with
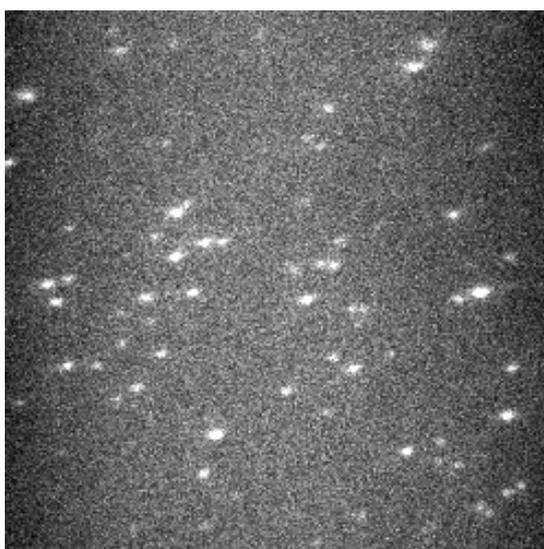
(a) Single image of fluorescence intensity
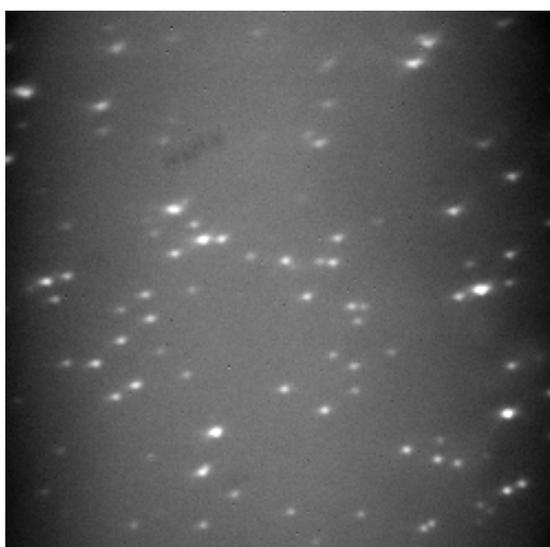
(b) Mean value of pixel intensities

(c) Maximum value of pixel intensities

Figure 3.7 – All images are from the same acquisition. **(a)** Single image of the fluorescence intensity of GFP under illumination by a wavelength of 488 nm. No colour information is saved in the image. - **(b)** Composite image created by calculating the mean vale for each pixel over all images with an illumination by a wavelength of 488 nm. - **(c)** Composite image created by calculating the maximum value for each pixel over all images with an illumination by a wavelength of 488 nm. - **(b), (c)** The spots seen here correspond to the spots seen in (a). Some spots are only recognisable in (b) and not in (c), and *vice versa*. Both (b) and (c) allow the identification of spots not directly visible in (a).

a higher intensity and long-lived with a lower intensity.

3.7 Spot Detection and Identification

The composite images from Section 3.6 are the starting points to detect and identify spots. Spots can either be identified by eye or more reliably and less biased by using an algorithm. The appearance of the composite image can be enhanced to help with the identification by eye. A moving average subtracts the underlying and varying background that emanates from the TIR illumination and a final adjustment of the brightness and contrast yields an enhanced version compared to the original with less variance from one acquisition to the next and a higher visibility of possible spots (Figure 3.8b). The possible spots have to be recognised and either refused or accepted depending on their shape, signal intensity above background and other possible factors. This judgement depends heavily on the person performing this task and can vary greatly. A more reliable way is to use an algorithm. The one used in this work is outlined.

A box of 19 pixels by 19 pixels is moved across the composite image (Figure 3.8a). The first requirement, before the more computational expensive steps of the spot detection are done, is that that the maximum pixel intensity inside this box has to be above a certain threshold compared to the mean pixel intensity of that box. A representative sample of spot was used to calibrate this threshold value accordingly and it was set to 1.119. If two peaks of intensity are detected within this box its size is gradually reduced until one peak remains. The spot is rejected if this box has to be smaller than 9 pixels by 9 pixels to fulfil this requirement. The intensity around the remaining peak is fitted with a 1D Gaussian (Figure 3.10b) for each dimension,

$$G\left(x\right) = B + \frac{1}{\sqrt{2\pi}} \times \exp\left[-\frac{\left(x - \mu\right)^2}{2\sigma^2}\right].$$

(3.1)

The mean μ of the Gaussian fit has to be at least 2.5 pixel away from the border
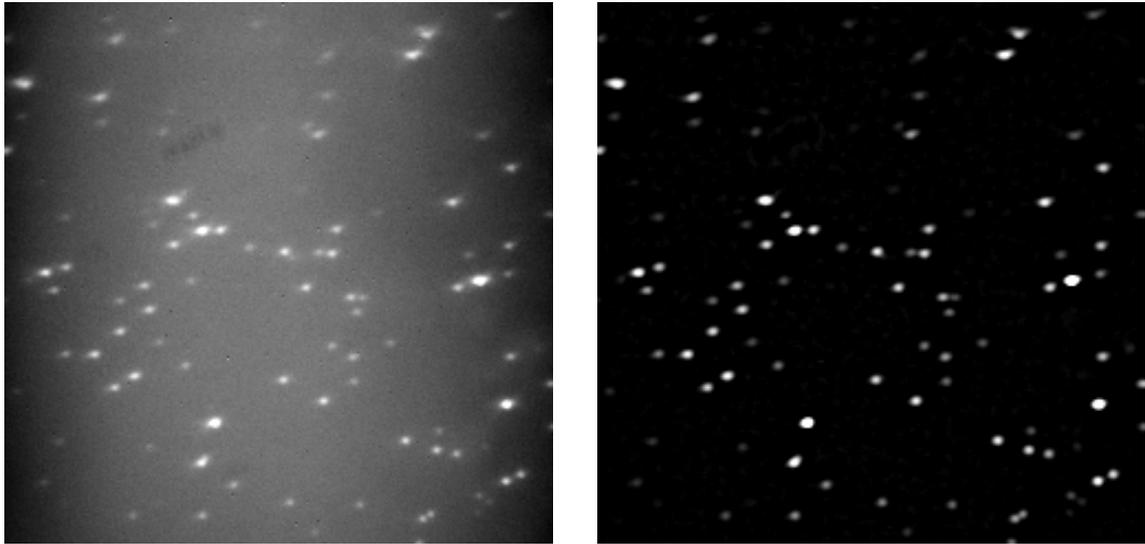
(a) Normal composite image

(b) Enhanced composite image

Figure 3.8 – Both images are from the same acquisition. **(a)** Composite image created by calculating the mean vale for each pixel over all images with an illumination by a wavelength of 488 nm. - **(b)** Enhanced version of the composite image seen in (a). A moving average subtracted the underlying background and the brightness and contrast were adjusted. The spots seen here correspond to the spots seen in (a).

Table 3.1 – Range of the values for the different parameters for a spot to be considered valid for the different used fluorophores.

Fluorophore	σ	B	χ^2
Cy5	0.9 to 3.5	0 to 0.4	$> 4.5 \times 10^{-7}$
GFP	0.3 to 4.7	0 to 0.53	$> 1 \times 10^{-7}$
mCherry	0.3 to 4.7	0 to 0.53	$> 1 \times 10^{-7}$

of the fitted box ($|\mu|$ < size of box - 2.5) to ensure that both flanks are present for an accurate fit. All occurring spots of initial sample acquisitions were used to calibrate the parameters σ and B from the Gaussian fit and its goodness of fit χ^2. It was found that the different used fluorophores require a different set of parameters after which their imaged fluorescent spots are judged by (Table 3.1).

The identification of spots of Cy5 reliably detected all spots present as found by manual inspection without including any false positives out of an representative sample of 200 spots. Only spots too close to each other are sometimes identified as one single spot. Those were manually removed from further analysis.

The identification of spots of mCherry or GFP is not as reliable. The parameters and the final implementation were chosen in a way to rather include false positives than to miss a spot. False positives are excluded at the step assignment of the intensity time traces as no bleaching steps or discernible intensity above the background will be present.

3.8 Assessing Colocalization

Once spots are identified, their colocalization has to be assessed. Spots of one type of fluorophore are used as a marker, for example Cy5 attached to the mRNA, and the amount of colocalization with spots of another type of fluorophore, for example a protein attached to GFP, is determined in respect to the marker spots. Each spot of one type of fluorophore can only be colocalized with one spot from another type of fluorophore. This does not mean that multiple fluorophores cannot be colocalized. They can be colocalized when appearing as part of one spot due to their distance between each other being smaller than the resolution limit of the microscope and therefore most likely being part of the same complex. The amount of colocalization can range from 0 % to 100 % relative to the marker.

Assessing the colocalization can be done by eye. Overlaying the composite images for each fluorophore and using convenient false colours for each of them allows one to identify pairs of spots (Figure 3.9c). Determining by eye if and which spots are colocalized with each other is highly subjective, made more difficult by the occurring chromatic aberration and the requirement to keep track of the colocalized spots. A computer program was employed to help to establish the colocalization between spots in a faster and more reliable and objective way.

A single spot has an intensity profile of a two-dimensional diffraction pattern. The originally continuous intensity profile is binned into pixels (Figure 3.10a). It can be estimated by applying a Gaussian fit (Figure 3.10b).

Different wavelengths for the emissions of different fluorophores result in a shift of their positions relatively to each other depending on their location in the field of view. Fluorescence originating from different types of fluorophores from the same complex can result in spots whose centres are separated by a short distance (Figure 3.10c). The underlying transformation was found to be linear in
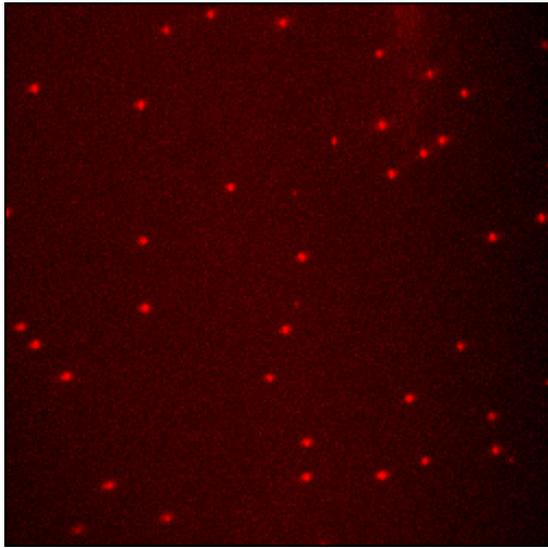
(a) Single image of fluorescence intensity
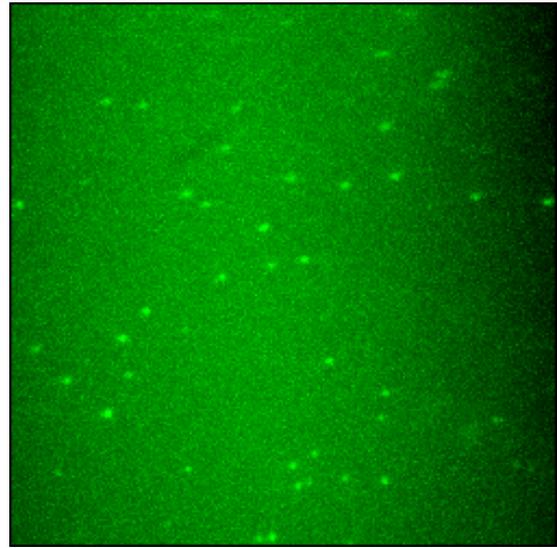
(b) Illumination by 633 nm
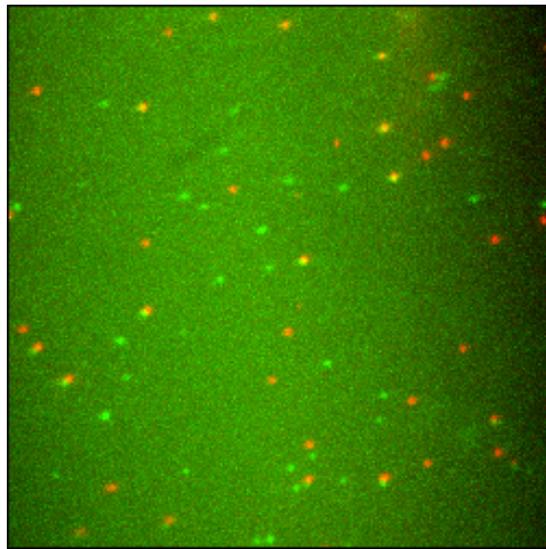
(c) Overlay of the individual illuminations

Figure 3.9 – All images are created from the same acquisition. **(a)** False colour image created by averaging over all images with an illumination by a wavelength of 633 nm, showing spots originating from Cy5. - **(b)** False colour image created by averaging over all images with an illumination by a wavelength of 488 nm, showing spots originating from GFP. - **(d)** The Overlay of the images (a) and (b). Colocalized spots can be see as green and reds spots close to each other. Spots on top of each other show up as yellow due to the used individual false colours. Note that green spots are shifted to the bottom left relatively to their red counter parts the further to the bottom left of the image they are.

(a) Red Spot

(b) Gaussian fit of intensities

(c) Red and green spot

(d) Colour shift
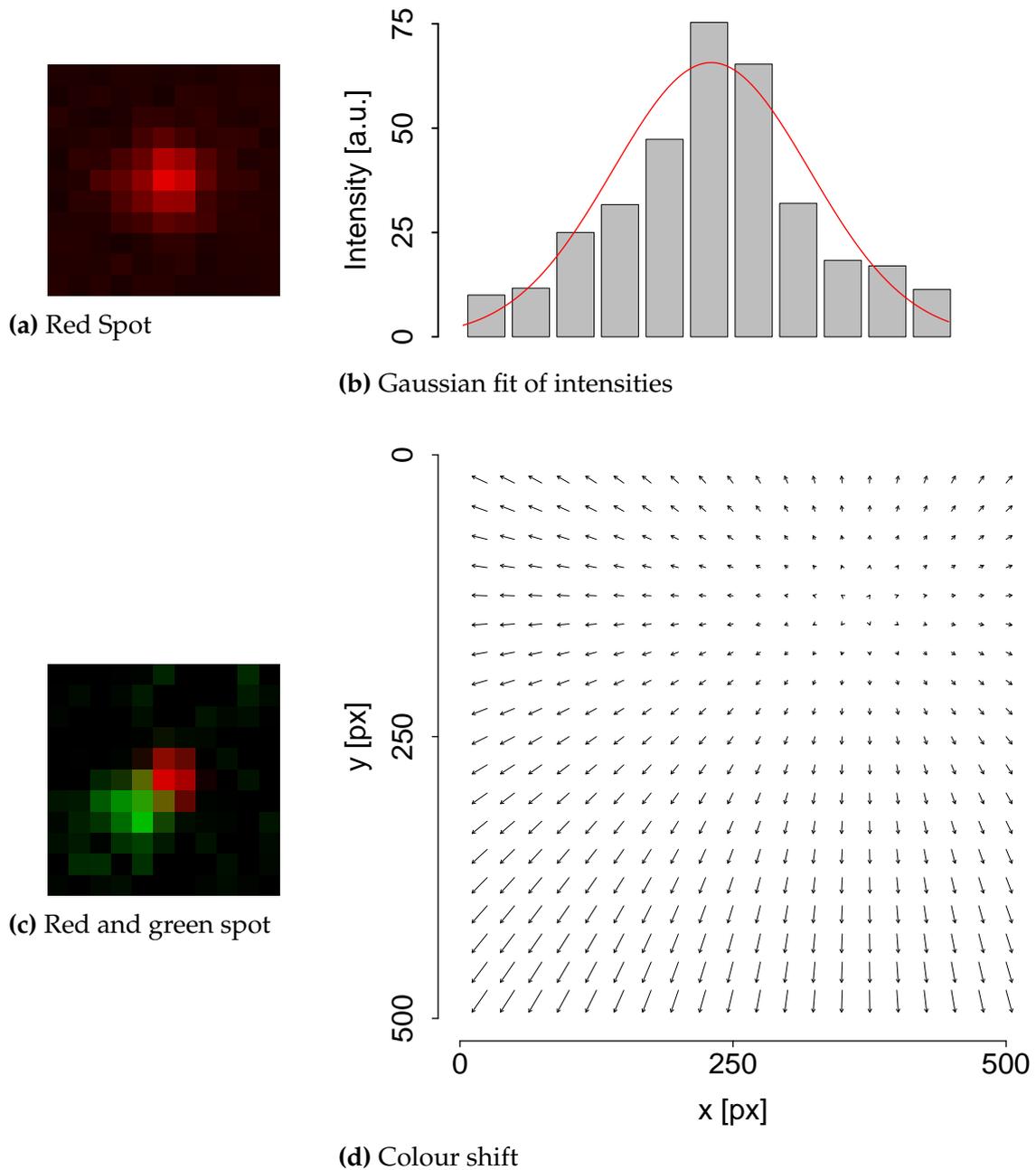
Figure 3.10 – **(a)** Amplified spot from a composite image of the 633 nm illumination. - **(b)** Gaussian fit of the individual pixel intensities in red. Using the centre of the Gaussian fit allows an increased accuracy of the positional estimate. - **(c)** Overlay of the spot from (a) with the corresponding composite image of the 488 nm illumination. The centre of the green spot is is shifted towards the bottom left. - **(d)** Measured colour shift across the field of view for a fluorophore imaged with a wavelength of 488 nm compared to a fluorophore imaged with a wavelength of 633 nm. The length of the shift is magnified by a factor of 5.

101

two dimensions and can be described for the two axes as

$$\Delta x = F \times (x - x_c) / SF_x \qquad (3.2a)$$

and

$$\Delta y = F \times (y - y_c) / SF_y, \qquad (3.2b)$$

where Δx and Δy are the shift for a spot with the coordinates x and y. The parameters for this transformation were obtained using reference spots and are $x_c = 367$, $SF_x = 123$, $y_c = 133$ and $SF_y = 88$ (Figure 3.10d). x_c and y_c are the coordinates of the centre where no shift occurs and SF_x and SF_y are scaling factors to translate the distance from the centre to the occurring shift. F is a weighting parameter to incorporate different fluorophores and their wavelengths and it is set to 0 for Cy5 (as shifts are relative to each other), 0.52 for mCherry and 1 for GFP.

The obtained positions for the spots were corrected using the described transformation and the distances between marker spots and all other spots are calculated. If this distance falls below a certain threshold, spots are considered colocalized. If more than one spot fulfils this criteria, the two closest spots are assigned to each other. The threshold distance can be varied to accommodate different requirement. It was set to 2 pixels for the analysis of all the single-molecule data presented in this work. This ensures an accurate assessment of the colocalization as discussed in Section 1.2.3, with an error of 0.3 % for the colocalization.

Analysing the colocalization gives one the percentage of colocalization as a general measurement of the relationship between the marker spot and the secondary spots. Only the intensity time traces of colocalized spots are later analysed for occurring bleaching steps to ascertain the numbers of fluorophores present.

3.9 Step Assignment

Intensity time traces are used to assess the number of fluorophores within a single complex that cannot be distinguished directly. Irreversible changes of the intensity over time due to the bleaching of fluorophores are analysed. Assigning and counting the steps allows the determination of fluorescent proteins present.

Background corrected intensity time traces are used. It was found that the best way to obtain these is by subtracting the mean intensity of the distant pixels, five pixels away from the centre, from the mean intensity of the pixels found in a central box of 3 pixels by 3 pixels (for a graphical representation see Figure 3.11a). The small central area where the intensity is taken from prevents the contamination of the intensity time trace by other spots close by. The distant pixels for the background correction allow the intensity time trace to not be influenced by changes in the spot intensity itself and is not noticeably disturbed by other spots it could cover.

Steps in fluorescence intensity time traces from single molecules were first observed 20 years ago (Ambrose et al., 1994). Initially, the distinct steps indicating the bleaching of a fluorophore were assigned by eye. Recent advances try to employ the help of algorithms. Carter et al. (2008) compared four algorithms for detecting steps of the kinesin motors and found that some are good enough to detect nearly all steps present. This was facilitated by two properties of these traces, their uni-directionality and a fixed step size. Both of these are not true for the intensity time traces for the surface bound fluorophores and to date no algorithm exists that can fully compete with a trained individual at assigning these steps. Nonetheless, the advantages of using an algorithm are the consistency that it provides between different samples and different assessors and a decreased analysis time.

The algorithm implemented for this work has at its heart a Bayesian step detec-

(a) Pixel positions   **(b)** One step

(c) One step   **(d)** Two steps

(e) Two steps   **(f)** Three steps
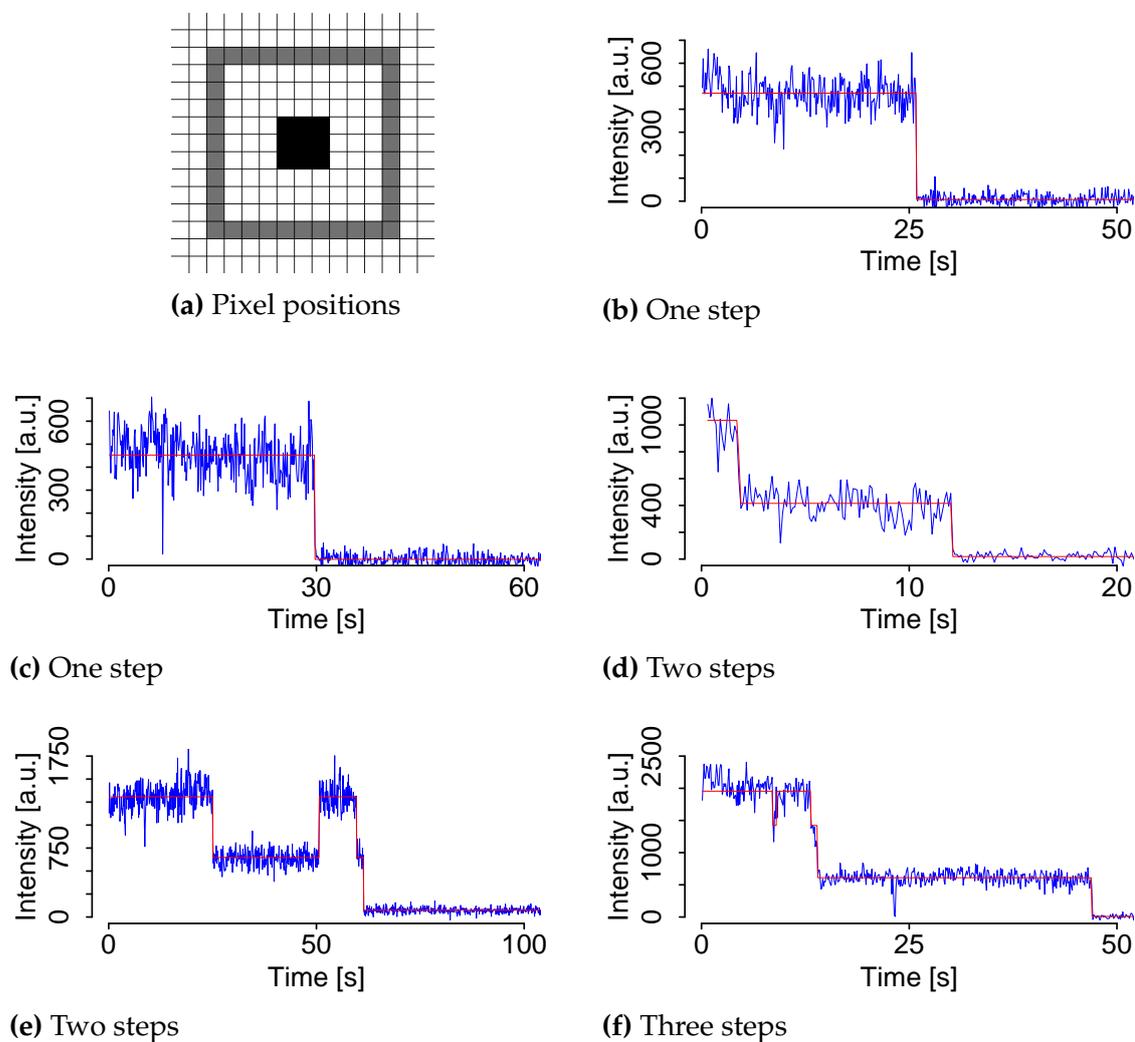
Figure 3.11 – **(a)** The intensity of a spot is taken as the mean pixel intensity of the area in black minus the mean pixel intensity from the area in grey. - **(a)** - **(f)** Representative intensity time traces from the experimental data for Cy5 with a time resolution of 0.1 s. The intensity is shown in blue, the algorithmic fit is shown in red. **(b)**, **(c)** Single-step bleaching. - **(d)** Two-step bleaching. - **(e)** Two-step bleaching with a short re-activation. - **(f)** Three-step bleaching.

(a) One step

(b) One step

(c) Two steps

(d) Two steps

(e) Three steps
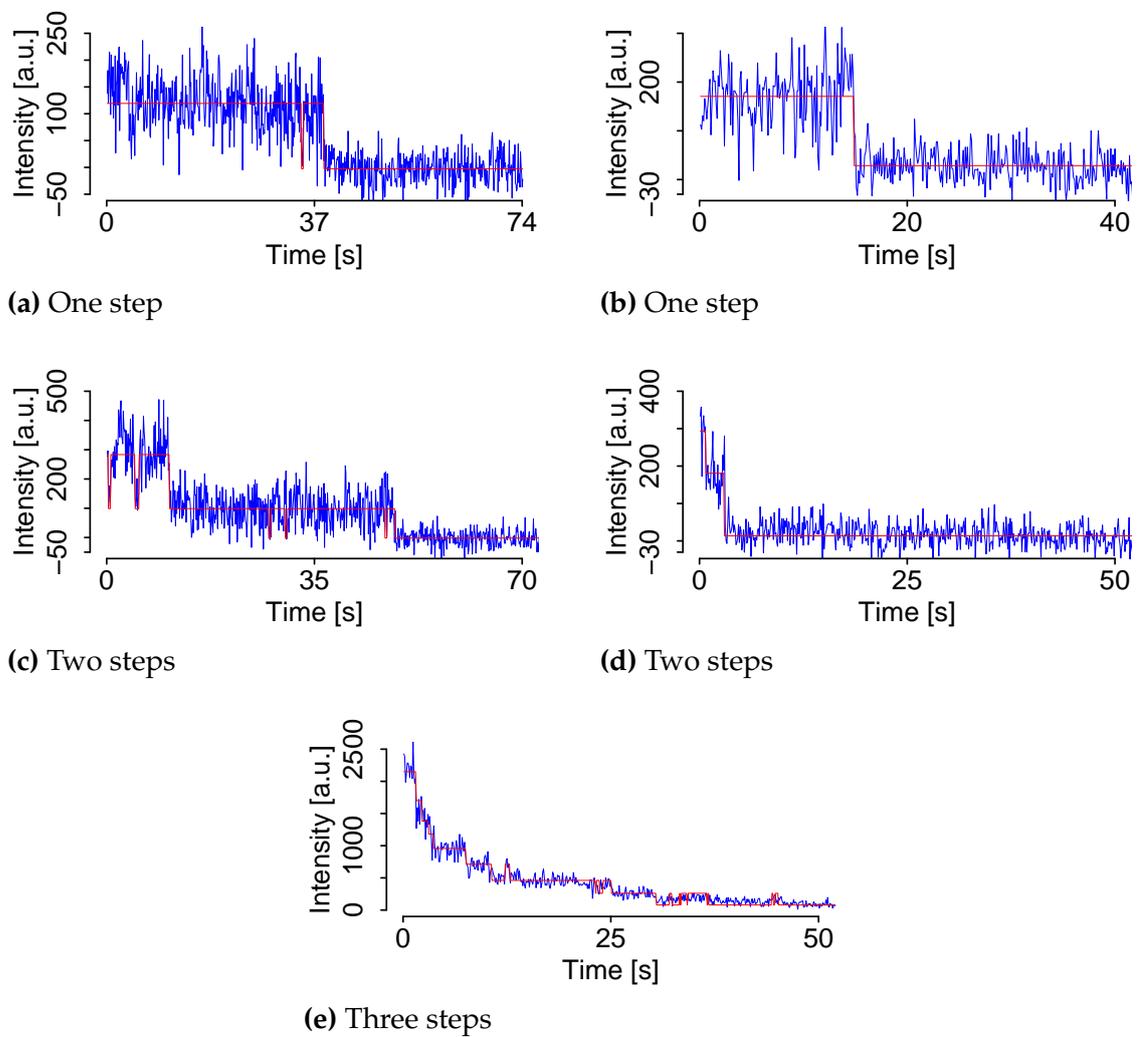
Figure 3.12 – Representative intensity time traces from the experimental data for GFP with a time resolution of 0.1 s. The intensity is shown in blue, the algorithmic fit is shown in red. **(a), (b)** Single-step bleaching. - **(c), (d)** Two-step bleaching. - **(e)** Bleaching in eight steps.

tor. For N data samples d_i of an intensity time trace that are piecewise constant with added Gaussian noise,

$$d_i = \begin{cases} \mu_1 + e_i, & \text{if } i < m \\ \mu_2 + e_i, & \text{otherwise,} \end{cases} \tag{3.3}$$

where μ is the underlying signal and e the added noise, is the likelihood p of a step given by

$$p\left(\{m\} \mid \mathbf{d}\right) \propto \frac{1}{\sqrt{m\left(N - m\right)}} \left[\sum_{i=1}^{N} d_i^2 - \frac{1}{m} S_l^2 - \frac{1}{N-m} S_r^2 \right]^{-\frac{N-2}{2}}, \tag{3.4}$$

with

$$S_l = \sum_{i=1}^{m} d_i, \tag{3.5a}$$

$$S_r = \sum_{i=m+1}^{N} d_i, \tag{3.5b}$$

where $\mathbf{d}$ is the intensity time trace and m the position along the trace (as described in *Numerical Bayesian Methods Applied to Signal Processing*). It detects the most likely point of change. p does not add up to 1 over the whole trace and has to be seen as a proportional representation. If this likelihood exceeds the set threshold, the trace is divided at this point into two sub-traces and the most likely changepoint for these is found again. This is repeated until no sufficiently likely changepoint is found or the examined trace is too short. The result is a list of changepoints for the intensity time trace. These points are further refined by limiting the length of the resulting plateaus and adjusting them in such a way that the resulting plateaus fit the set required minimal distances between them and their minimal heights.

The intensity time traces of Cy5 fluorophores as found in the experiments show

a low noise and clear steps. The algorithm successfully detects one and two bleaching step within an intensity time trace (Figures 3.11b to 3.11d). Re-activated fluorophores result in the correct number of found steps as the final assessment is based not on the position of the changepoint but on the resulting plateaus. Thus parts of the intensity time trace belonging to the same number of concurrently active fluorophores are assigned to the same plateaus (Figure 3.11e). Bleaching steps that occur in close temporal proximity can be resolved by taking account of the expected step heights (Figure 3.11f).

GFP as a fluorophore can undergo blinking, has a less constant emission of photons and a lower brightness compared to Cy5. Its intensity time traces have more noise and smaller step heights for individual bleaching events. The algorithm still detects the bleaching steps accurately (Figures 3.12a and 3.12b). Multiple changepoints that arise from the blinking of GFP might be detected, but the assignment of plateaus gives the correct result for the number of overall bleaching steps (Figure 3.12c). The often rapid bleaching of GFP and the bleaching of steps in quick succession can be detected for temporal separations down to 6 data points (here 0.6 s) (Figure 3.12d). The algorithm allows to identify up to 9 steps per intensity time trace as this was one of its restrains (Figure 3.12d).

The step detection algorithm was tested on the acquired experimental data and compared to the manual assignments. Simulated data was not used for comparison as the underlying conditions are complex with a multitude of possible parameters. This out it outside the scope of this work.

An uncertainty of the measurement might arise from traces that show multiple steps where the assignment of steps might differ from the underlying number of labelled proteins present. This effect is the more pronounced the more proteins are present. Assuming that the number of steps can err equally in both directions, the observed distribution should then not differ significantly as the effects of

neighbouring categories cancel each other out.

3.10 Automation of Analysing Data

The data obtained by the microscope are sequential intensity images accompanied by a text file with additional information. Several individual steps have to be performed to obtain the desired information. Composite images have to be created for the different laser excitations. Fluorescent spots have to be identified and their positions to be determined accurately. This information is used to assess the colocalization. Intensity time traces from colocalized spots are extracted from the raw data and for each steps are assigned. The number of these steps and their distribution together with the amount of overall colocalization is the desired information. This and all intermediate steps have to be documented.

Each step of this process was implemented in a MATLAB program to both simplify the process and reduce the time necessary for it (Figure 3.13). A graphical user interface guides the user through all steps and all relevant data can be exported into an excel file. The program employs the algorithms described in this chapter. The user can interfere at any stage with the program and correct the decisions made by the used algorithms. This allows to correct any erroneous results from a single algorithm and to enhance the overall accuracy.

Analysing a single data set can now be done within hours instead of days as before. It is more reliable and reproducible with a greater accessibility for new users and automatically provides an accurate documentation.
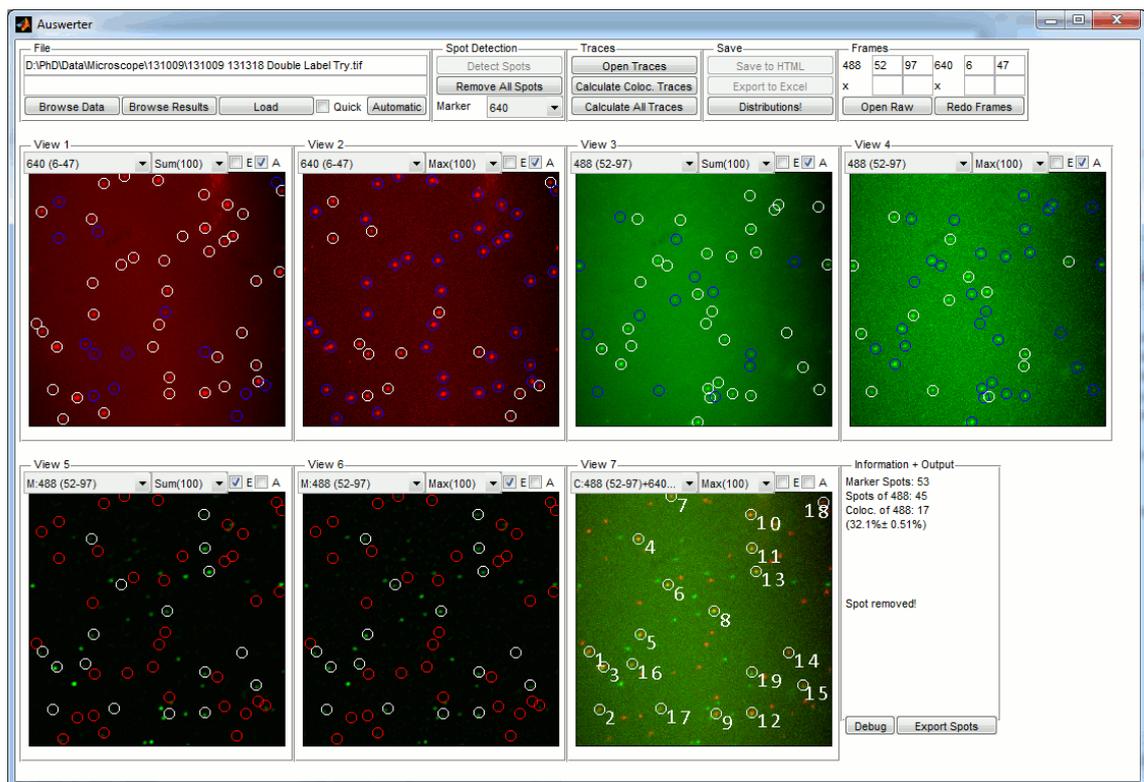
Figure 3.13 – Interface of the MATLAB Program.

3.11 Presentation of Data

The results of the single-molecule experiments will be presented as bar charts. Most results are from the two-colour experiments which are shown as seen in Figure 3.14a. On top of the bar chart is the total number of identified marker spots, the number of spots colocalized with these spots and the resulting colocalization percentage. The bar chart shows the distribution of the number of steps found in the intensity time traces of the colocalized spots only. Separated to the right in a darker grey is the bar that shows the accumulated percentage for all spots that had more identified steps than individually shown on the bars to the left. The last bar in white represents identified colocalized spots for which no certain number of bleaching steps could be identified. This can be due to a fluorophore that did not bleach over the course of one acquisition or where the trace did not allow a clear assignment of the number of bleaching steps. These cases still count as colocalized. All individual categories add up to 100 %.

Three-colour experiments, with one colour for the marker fluorophore and two colours left for the colocalized fluorophores, result in two individual distributions for each individual colocalized fluorophore as shown in Figure 3.14a. An assessment of their correlation between each other requires a representation in two dimensions. The distribution for the number of steps found for the two colocalized fluorophores has the same individual categories as for the normal analysis. Occurrences are now given as absolute numbers as both groups have different references values. This representation allows to assess if the presence of one factor depends on the presence or absence of another factor (Figure 3.14b).

111
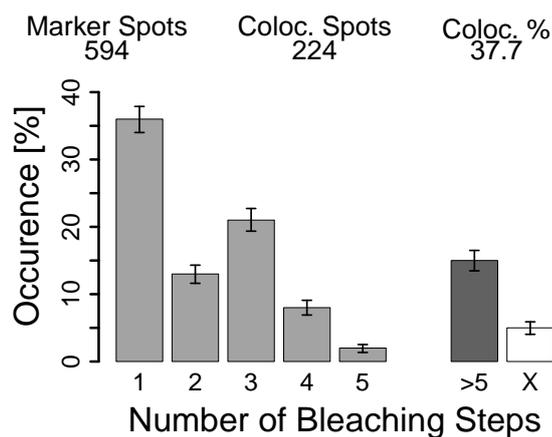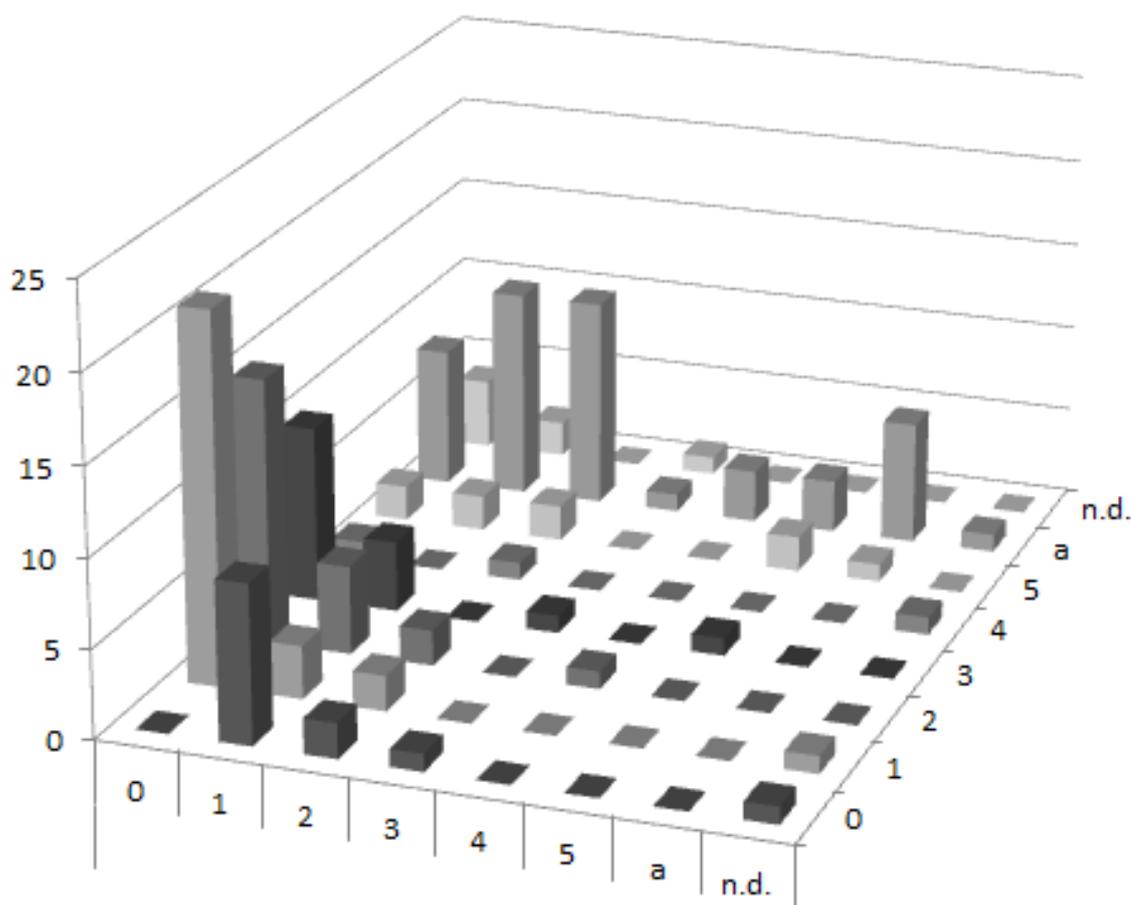
(a) Two colours

(b) Three colours

Figure 3.14 – Exemplary results of the single-molecule experiments. **(a)** With two colours, one for the marker fluorophores and one for the colocalized fluorophores, a one-dimensional diagram suffices. Error bars are the standard deviations of the underlying binomial distribution. - **(b)** Three colours, one for the marker fluorophores and two for the colocalized fluorophores, require a two-dimensional representation.

Chapter 4

Preparation for Single Molecule Experiments

Before anything else, preparation is the key to success

Alexander Graham Bell

Several steps have to be taken as preparation before the single-molecule experiments can be performed. The proteins of interest have to be labelled with fluorescent markers to make them visible for fluorescence microscopy and nuclear extracts have to be prepared from the cells expressing these modified proteins. Likewise, the pre-mRNA constructs have to be labelled. The surface of the sample chamber plays an important role in the microscopy based on total internal reflection. The molecules of interest have to be specifically confined to the surface to allow investigation over longer time ranges while simultaneously preventing the non-specific binding.

4.1 Labelling of SRSF1 in Nuclear Extract

Initially prepared nuclear extracts with labelled SRSF1 made from 293T cells contained excessive amounts of over-expressed SRSF1-GFP. It could not used in single-molecule experiments as the amount of fluorescence originating from SRSF1-GFP overshaded everything else at viable levels of labelled pre-mRNA. Therefore, nuclear extracts were made from HeLa cells where more favourable expression levels were present. Fluorescent Western blots were done on the nuclear extract to assess the labelling efficiency and compare the individual nuclear extract with the commercially available nuclear extract.

The labelling efficiency of the nuclear extract with SRSF1-mCherry is 93 % and SRSF1 is 2.6 times over-expressed compared to the commercial nuclear extract (Figure 4.1a).

The nuclear extract with SRSF1-GFP has a labelling efficiency of 90 %. SRSF1 is 2.3 times over-expressed if directly compared to the commercial nuclear extract. U1A was used as a relevant comparison of protein levels as it is closely linked to the splicing of RNA. U1A is 3 times as abundant in the commercial nuclear extract as compared to the SRSF1-GFP nuclear extract. This is in agreement with the splicing efficiency found in the SRSF1-GFP nuclear extract. All single-molecule experiments with SRSF1 were performed with this nuclear extracts unless stated otherwise (Figure 4.1b).

The labelling and mutational efficiency for the nuclear extract with SRSF1-ΔRS-GFP is 57 %. SRSF1-ΔRRM2-GFP could not be detected in a Western blot albeit having comparable levels of fluorescence. SRSF1 is 1.3 times over-expressed in the SRSF1-ΔRS-GFP nuclear extract compared to the commercial nuclear extract. As SRSF1-ΔRRM2-GFP could not be detected, no comparison was possible for this nuclear extract. U1A is 1.5 times more abundant in the commercial nuclear extract (and 1.2 in the HeLa nuclear extract) compared to both SRSF1 nuclear ex-

(a) Western blot of SRSF1

(b) Western blot of SRSF1 and U1A

(c) Western blot of SRSF1 and U1A
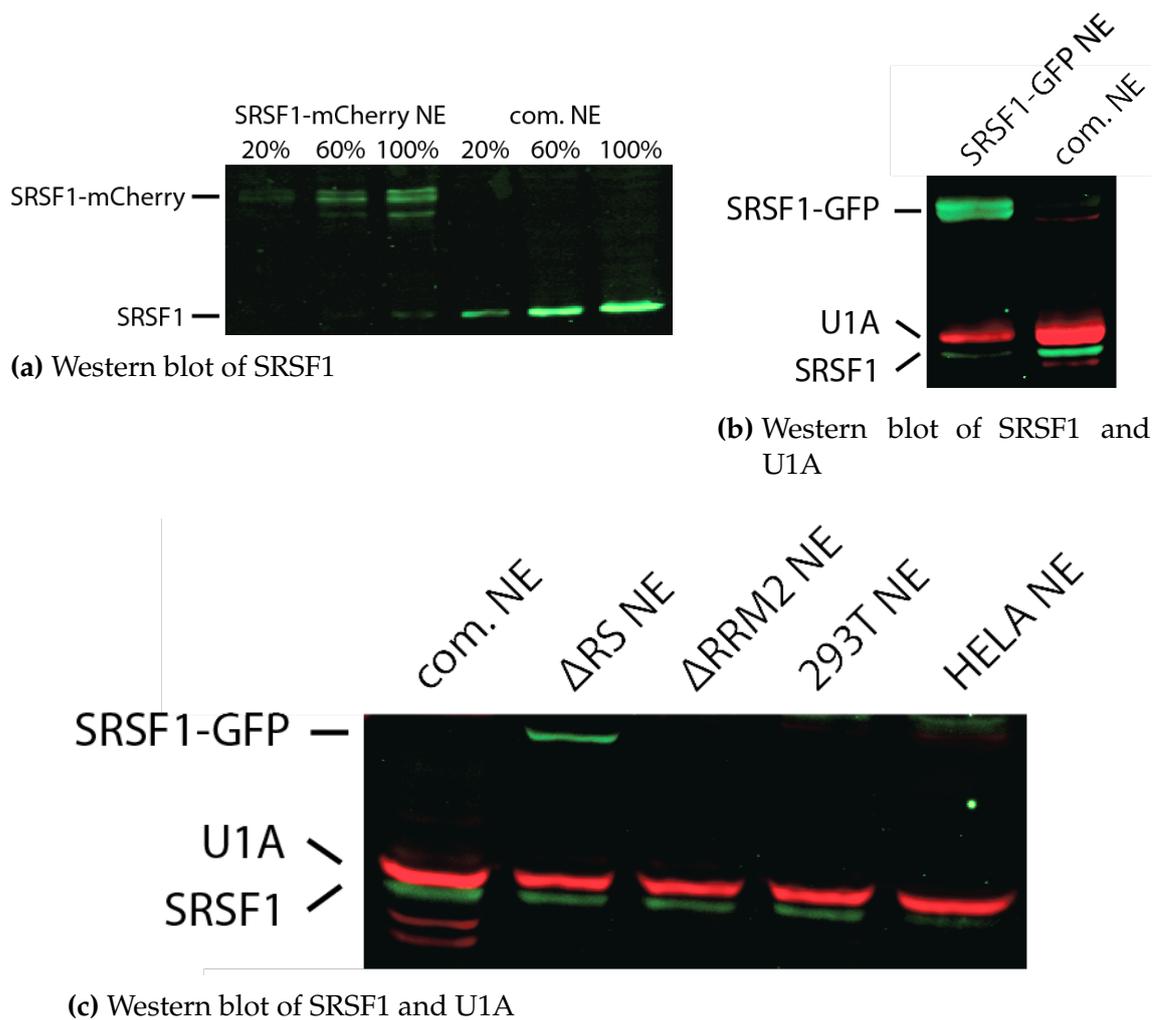
Figure 4.1 – **(a)** Fluorescent Western blot of different concentrations of the SRSF1-mCherry nuclear extract and the commercial nuclear extract. It was probed with an antibody for SRSF1, shown in green. - **(b)** Fluorescent Western blot of the SRSF1-GFP nuclear extract and the commercial nuclear extract. It was probed with an antibody for SRSF1, shown in green, and an antibody for U1A, shown in red. - **(c)** Fluorescent Western blot of the commercial nuclear extract, the ΔRS-SRSF1-GFP and ΔRRM2-SRSF1-GFP nuclear extracts and the unlabelled nuclear extracts from 293T and HeLa cells. It was probed with an antibody for SRSF1, shown in green, and an antibody for U1A, shown in red.

tracts that have similar levels of U1A. These two nuclear extracts were used for the single-molecule experiments investigating the role of the domains of SRSF1 (Figure 4.1b).

4.2 Labelling of the Pre-mRNA

A fluorescent probe has to be attached to the single molecules of the pre-mRNA to allow its visualisation under a fluorescent microscope. The two oligos used in this work, β-globin 5′ Cy5 and Ad Cy5, have both Cy5 at one end and a biotin-group at the other end (Table 8.8). They can be annealed to β-globin and adenovirus based constructs, respectively, and both simultaneously label the pre-mRNA constructs and provide one part of the surface attachment of the biotin-streptavidin interaction.

These oligos were designed to not contain any known binding sequence for SRSF1. Nonetheless, it was found that SRSF1 associates with the β-globin 5′ Cy5 oligo under conditions permitting formation of both complexes E and A (Figures 4.2a and 4.2b, for exemplary data see Figure 4.3a and Figure 4.4). Once the oligo is annealed to its complementary sequence, the colocalization with SRSF1 is drastically reduced (Figure 4.2c, for exemplary data see Figure 4.3b and Figure 4.4). It is therefore crucial to ensure that each oligo used for labelling the pre-mRNA is annealed so that no free oligos remain. This can be achieved by optimising the annealing conditions and using an excess of pre-mRNA compared to the amount of oligo.

The annealing of the labelling oligo with the pre-mRNA is done with increasing concentrations of pre-mRNA. The RNA is run on a 6 % denaturing polyacrylamide gel and its fluorescence imaged (Figure 4.5a). The amount of free Cy5 is assessed. It decreases with increasing amounts of pre-mRNA (Figure 4.5b). The lowest amount of free Cy5 that could be achieved is around 2 %. Higher amounts of pre-mRNA do not allow for less free Cy5. This corresponds to the colocalization found for the labelling oligo annealed to its complementary sequence (Figure 4.2c).

(a) Pure β-globin 5′ Cy5 in complex E

(b) Pure β-globin 5′ Cy5 in complex A

(c) Annealed β-globin 5′ Cy5 in complex E
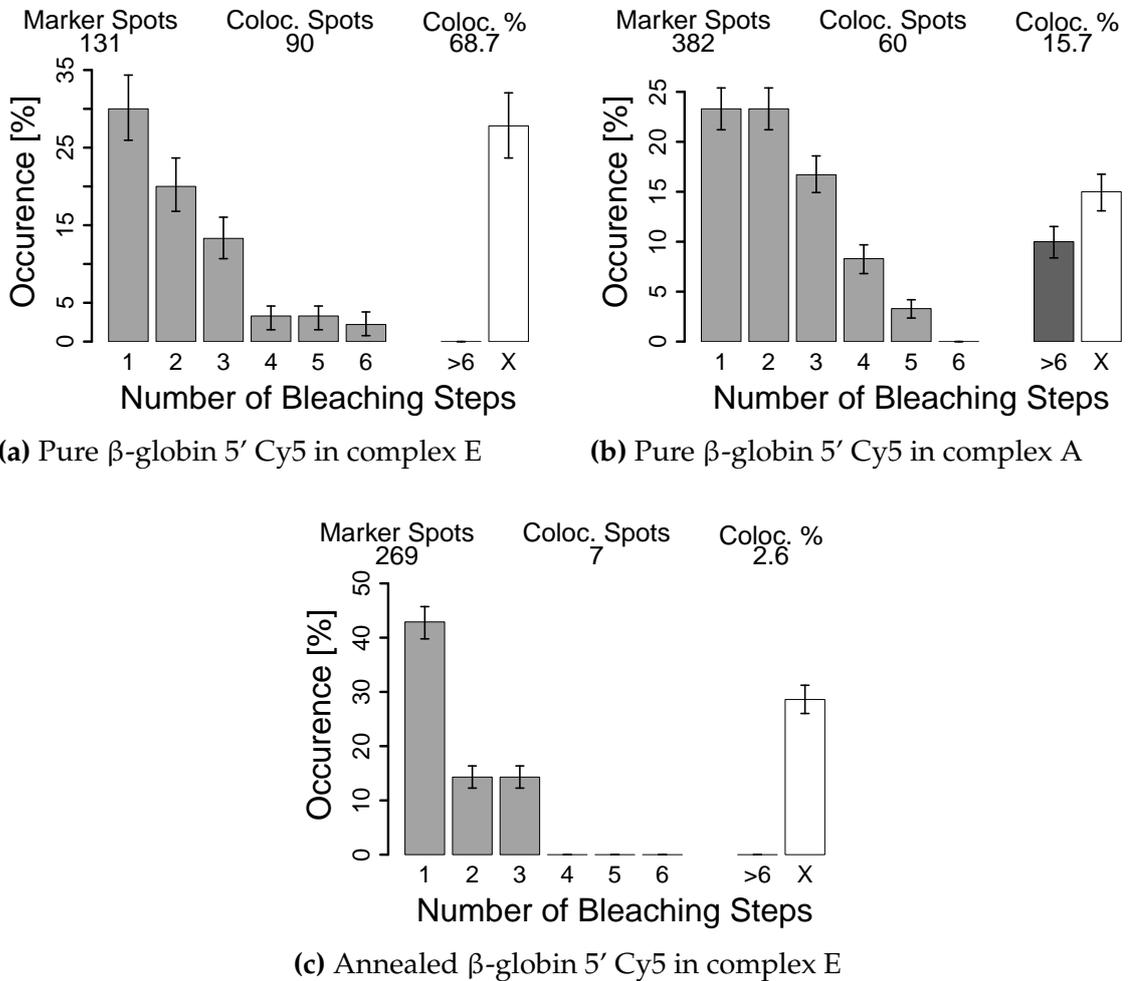
Figure 4.2 – (a), **(b)** The β-globin 5′ Cy5 oligo incubated in nuclear extract without any other pre-mRNA present under the conditions for (a) complex E and (b) complex A shows the presence of SRSF1 along the pre-mRNA. - **(c)** The β-globin 5′ Cy5 oligo annealed to its complimentary sequence and incubated with nuclear extract under the conditions for complex E shows minimal colocalization with SRSF1.
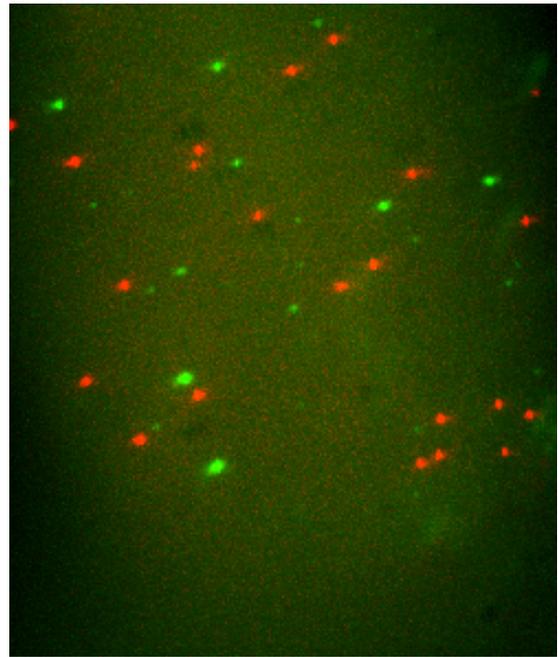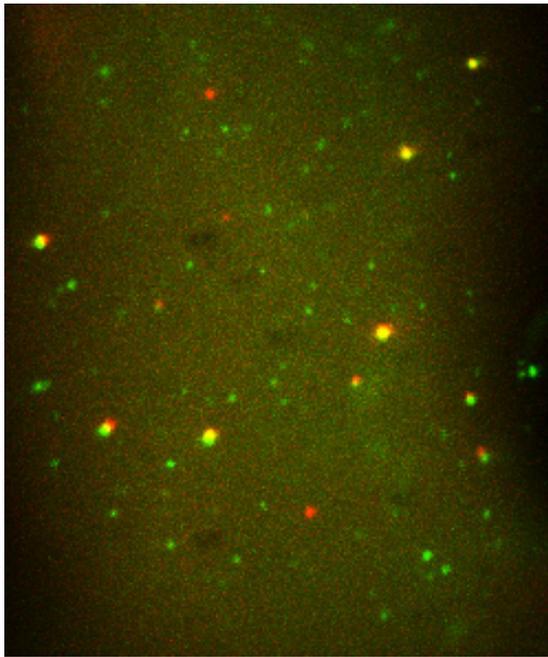
(a) Pure β-globin 5′ Cy5 in complex E   **(b)** Annealed β-globin 5′ Cy5 in complex E

Figure 4.3 – Overlays of the composite images for both Cy5 (in red) and GFP (in green), calculated by using the maximum value for each pixel, as described in Section 3.6. **(a)** Exemplary image corresponding to the experimental results shown in Figure 4.2a. 10 out the 13 red spots are visibly colocalized with an adjacent green spot, illustrating the high colocalization percentage of 68.7 %. - **(b)** Exemplary image for the results shown in Figure 4.2c. Not one of the 21 red spots is colocalized with any green spot, showcasing the low colocalization of 2.6 %.

(a) Raw trace, one step

(b) Background corrected trace, one step

(c) Raw trace, one step

(d) Background corrected trace, one step

(e) Raw trace, two steps

(f) Background corrected trace, two steps

(g) Raw trace, two steps

(h) Background corrected trace, two steps
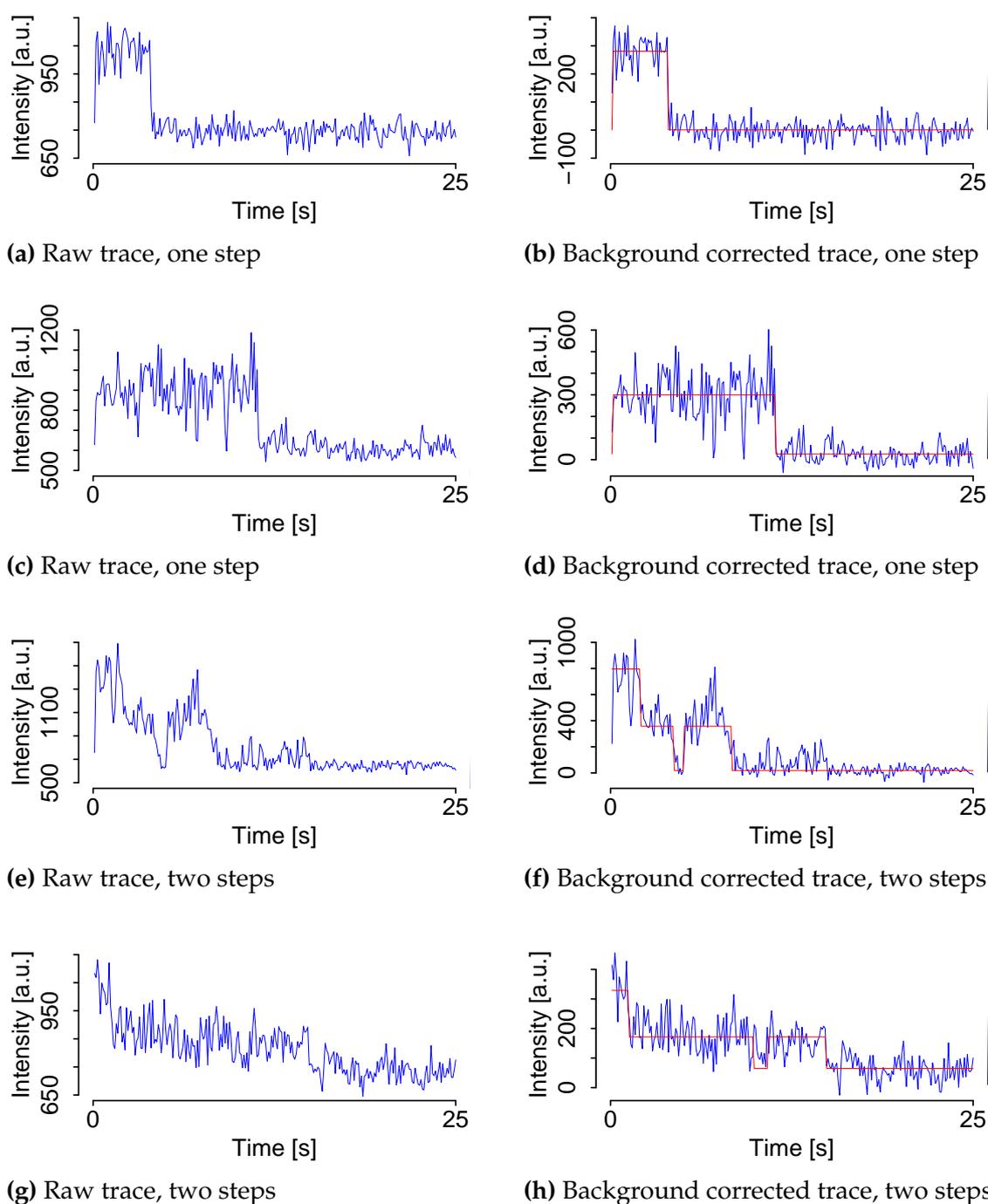
Figure 4.4 – Exemplary traces of the experimental results shown in Figure 4.2. On the left hand are the raw traces obtained from the central 3 by 3 pixels around the spot position, and on the right hand are the background corrected traces, both shown in blue (see Figure 3.11a). The algorithmic fit (described in Section 3.9) is shown in red for the background corrected trace. **(a) - (d)** Single step bleaching. - **(e) - (h)** Two step bleaching.
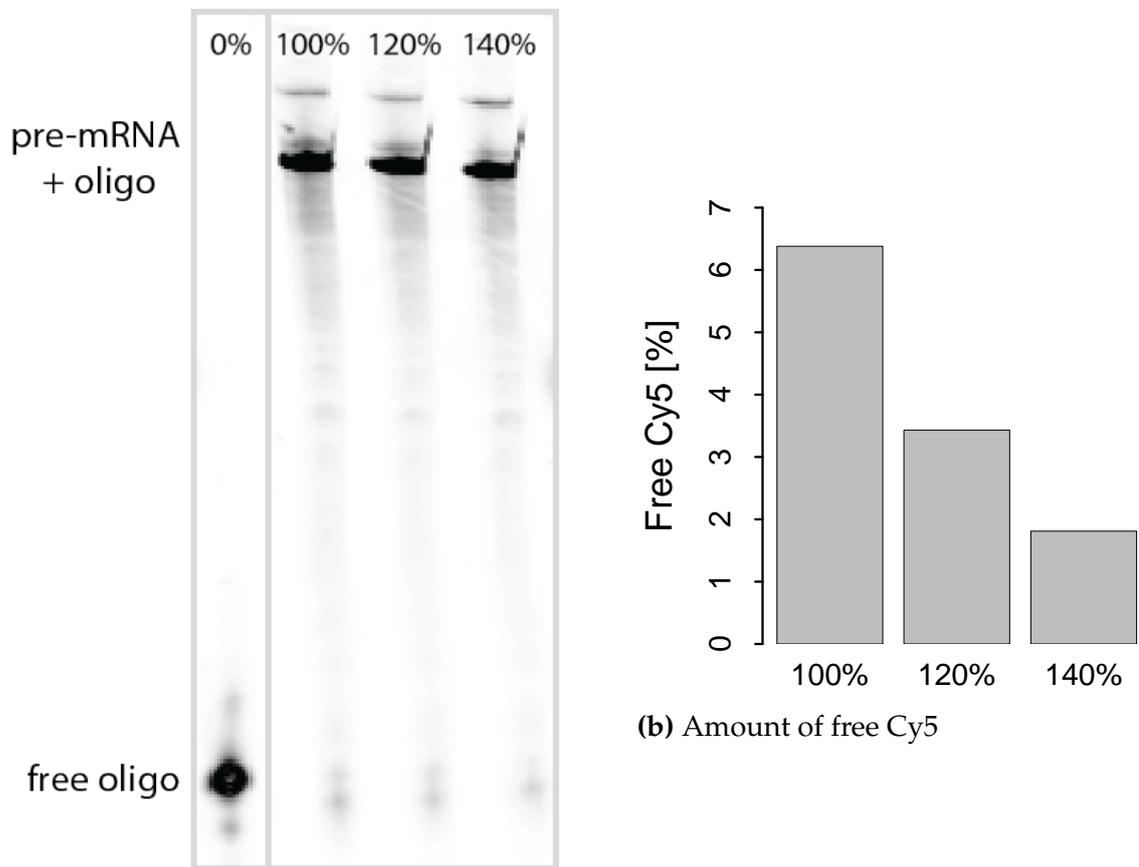
(a) Denaturing polyacrylamide gel

Figure 4.5 – **(a)** Increasing amounts of pre-mRNA compared to the labelling oligo are run on a 6 % denaturing polyacrylamide gel and its fluorescence imaged. The amount of the pre-mRNA compared to the labelling oligo is noted, 100 % means the same number of pre-mRNA molecules as molecules of the labelling oligo. - **(b)** The amount of free Cy5 dye for the gel shown on the left. The amount of free Cy5 decreases with increasing amounts of pre-mRNA.

4.3 Surface Tethering

The fluorescent probes have to be confined to the surface to allow their observation under TIR illumination. A commonly used system employs the strong specific binding of biotin and streptavidin. The surface is incubated with BSA-biotin, where the BSA part binds non-specifically to the surface. The added streptavidin, as it has four binding sites for the biotin, can bind to the biotin attached to the BSA on the surface and provides further available binding sites for the biotin of the labelling oligo (Table 8.8 and Section 4.2). This chain of interactions confines the pre-mRNA to the surface. At the same time, the usage of BSA was intended to prevent further non-specific binding of other proteins to the surface.

Different amounts of streptavidin and BSA-biotin (biotin covalently attached to BSA) were used to optimise the number of pre-mRNA and SRSF1 spots, especially to reduce the number of SRSF1 that is non-specifically bound to the surface and not a part of the splicing complex. Different amounts of streptavidin and BSA-biotin were used and a normal splicing reaction was observed. The amount of identified pre-mRNA and SRSF1 spots was assessed in triplicates. Both the amount of BSA-biotin and the amount of streptavidin have no influence on the amount of identified surface confined pre-mRNA spots. Pre-mRNA can be found on the surface even in the absence of the BSA-botin and streptavidin (Figure 4.6a). An increased amount of BSA-biotin reduced the number of SRSF1 spots by up to 28 % (Figure 4.6b).

It was decided to stay consistent with other performed experiments and still use the streptavidin-biotin system. Its influence seems to be negligible but was not tested for all potentially occurring conditions and pre-mRNA constructs.
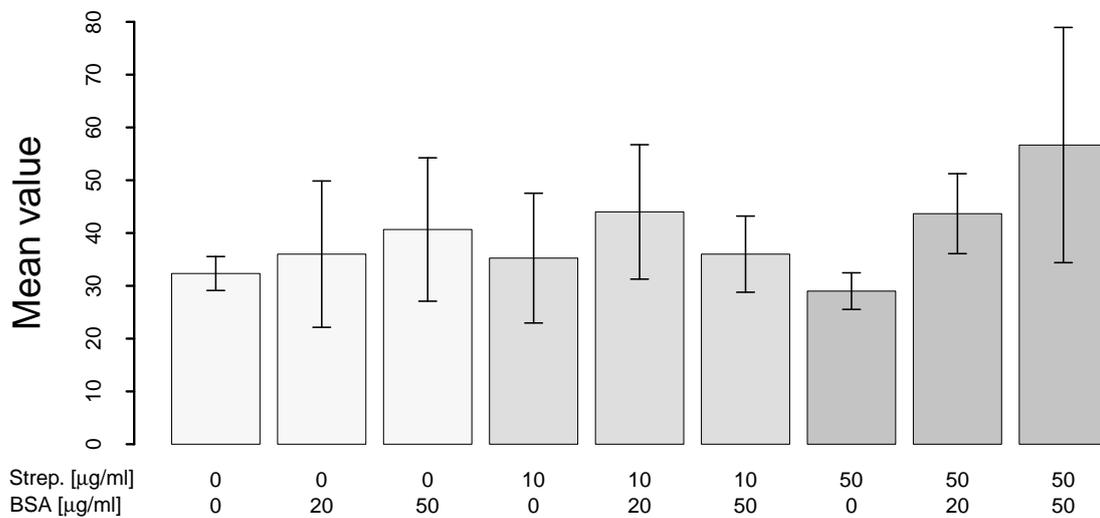
(a) Amount of identified pre-mRNA spots
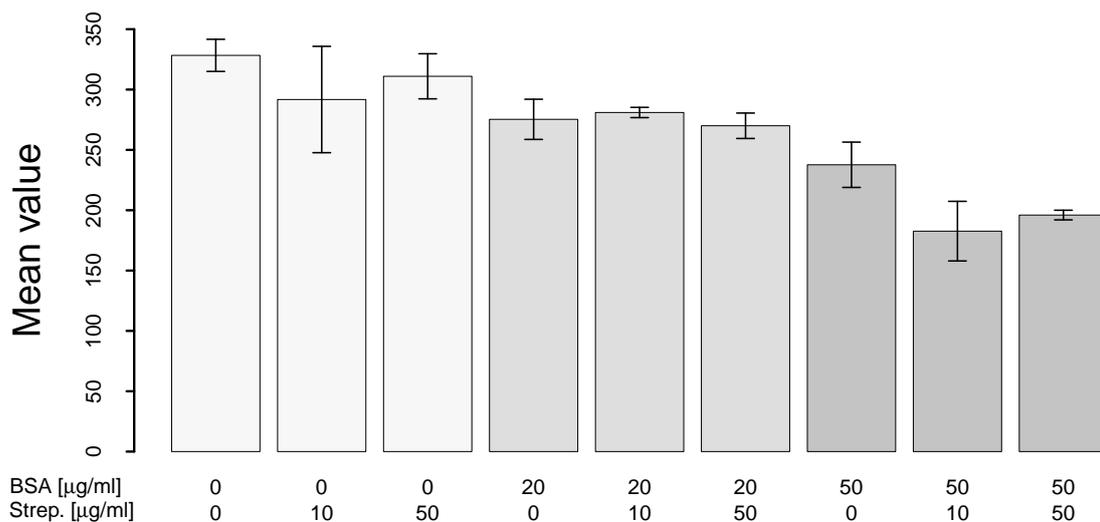
(b) Amount of identified SRSF1 spots

Figure 4.6 – Experiments for each condition were done in triplicates and the error bars are standard deviations. **(a)** Amounts of identified pre-mRNA spots are independent of the used amounts of streptavidin and BSA-biotin. - **(b)** Amounts of identified SRSF1 spots are independent of the used amounts of streptavidin and decrease with increasing amounts of BSA-biotin.

4.4 Surface Modifications

The amount of non-specific binding of SRSF1-GFP to the surface impedes the evaluation of the single-molecule data. The splicing reactions have to be severely diluted to allow an analysis on a single-molecule level and the limiting factor is the amount of labelled protein found on the surface. The amount of pre-mRNA confined to the surface can be regulated by adjusting the amount of pre-mRNA added to the splicing reaction and the amount of proteins specifically binding to the pre-mRNA would be of the same magnitude. The amount of non-specific protein binding is therefore the true limiting factor. Throughout the course of this work it was tried to reduce this non-specific binding by various modifications of the surface.

One idea on how to achieve this was to cover the surface with non-fluorescent proteins. SRSF1 does not bind non-specifically to other proteins and it cannot bind to the surface if it is covered by other proteins. This can be done with BSA, which did not reduce the non-specific binding of SRSF1 enough (Section 4.3). Nuclear extract was used to wash the sample chambers and achieved similar non-sufficient results. It was suggested to use a mixture of arginine and glutamic acid to cover the surface. This did not reduce the non-specific binding but resulted in a more homogeneous regional distribution of surface bound SRSF1.

Another way is to make the surface hydrophilic. Proteins would be repelled from the surface and could not bind to it non-specifically. Six different silanes, RepelCote (VWR), dimethyldimethoxysilane, 3-Aminopropyltriethoxysilane, [3-(Diethylamino)propyl]trimethoxysilane, perfluorooctyltriethoxysilane and dichlorodimethylsilane were used. They were diluted in either methanol or toluene, applied by spin coating or pipetting, washed with acetone, methanol, chloroform, water or combinations thereof and dried at different temperatures. All of these succeeded in the reduction of the non-specific binding but none of them were

suitable for single-molecule investigations as the surface was contaminated with spots identical to the ones originating from the fluorophores GFP and mCherry, making a further analysis infeasible.

Another method is to coat slides with PEG. Various different protocols were tried but could never be implemented in a consistent manner. The prepared sample chambers had contaminations of the background that were indistinguishable from the spots produced by the GFP attached to SRSF1, making the analysis impossible.

None of these methods showed any satisfactory results. Either contaminations from chemicals not suitable for single-molecule fluorescence measurements prevented an accurate later detection of spots or prevention of non-specific binding was to little. The last fact might have been due to a failure in consistently and sufficiently covering the glass surface, especially with BSA. One possibility not pursued was the addition of BSA to all used buffers to resupply potentially washed off BSA.

In the end, no additional modifications of the surface were used besides the prolonged plasma cleaning.

5

The Role of SRSF1 in the Early Splicing Complexes

Silently one by one, in the infinite meadows of
heaven, Blossomed the lovely stars, the
forget-me-nots of the angels

_____

Henry Wadsworth Longfellow

So far it was not known how many proteins of SRSF1 bind to the pre-mRNA during splicing. The binding of SRSF1 during the complexes E, A and C was studied and it was investigated how the number of bound SRSF1 is influenced by the presence of the U1 snRNP, the 5′ splice site and the number of 5′ splice sites, phosphorylation, the presence of an ESE, and what the role of the domains of SRSF1 is.

5.1 Stalling Splicing in the Complexes E, A and C

Splicing is a highly dynamical process characterised by the formation of several distinct complexes. The complexes investigated in this work are the complex E, A, and C. Each pre-mRNA and its splicing might progress along the same path but at different paces. The splicing progress can be stalled to obtain homogeneous samples representing a chosen complex. The depletion of ATP stalls splicing in complex E as ATP is needed for the progression to complex A (Michaud and Reed, 1991). The addition of anacardic acid or of α-U6 oligos (Kuhn et al. (2009), see Table 8.8) prevents the progression of splicing to complex B, effectively stalling it in complex A. Modifying the 3' splice site of the pre-mRNA (Globin C GG) leads to a stall in complex C (Jurica et al., 2002). This was confirmed by native gel electrophoresis and worked as described (Figures 5.1 and 5.2a). All conditions allowed an efficient stalling of the complex formation.

The depletion of ATP is accompanied by the formation of complexes that are too big to leave the wells of the gel. This cannot be seen for any other complex or for the 5' splice site mutation. It is aggravated by more severe depletions of ATP (Figure 5.2b).

For all of the following, referring to a certain complex means that the progression of splicing was stalled in this complex as outlined in this section.
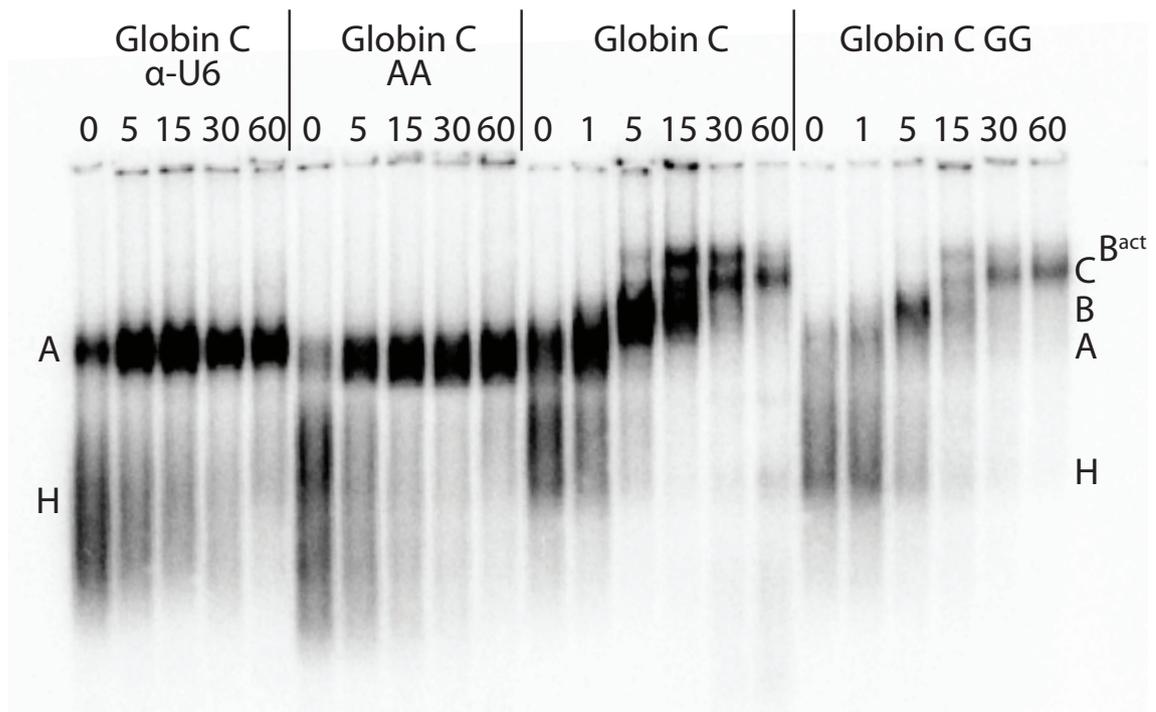
Figure 5.1 – Native agarose gel electrophoresis of samples treated with heparin and run on a 2 % agarose gel. Timepoints were taken from 0 min to 60 min. The normal progression of splicing through the complexes can be seen for Globin C. Complex H is followed by A, B, B^{act} and C. Both the addition of the α-U6 oligo and anacardic acid stall the progression in complex A. The mutation of the 3' splice site in Globin C GG stalls complex formation in complex C.

(a) Complex E
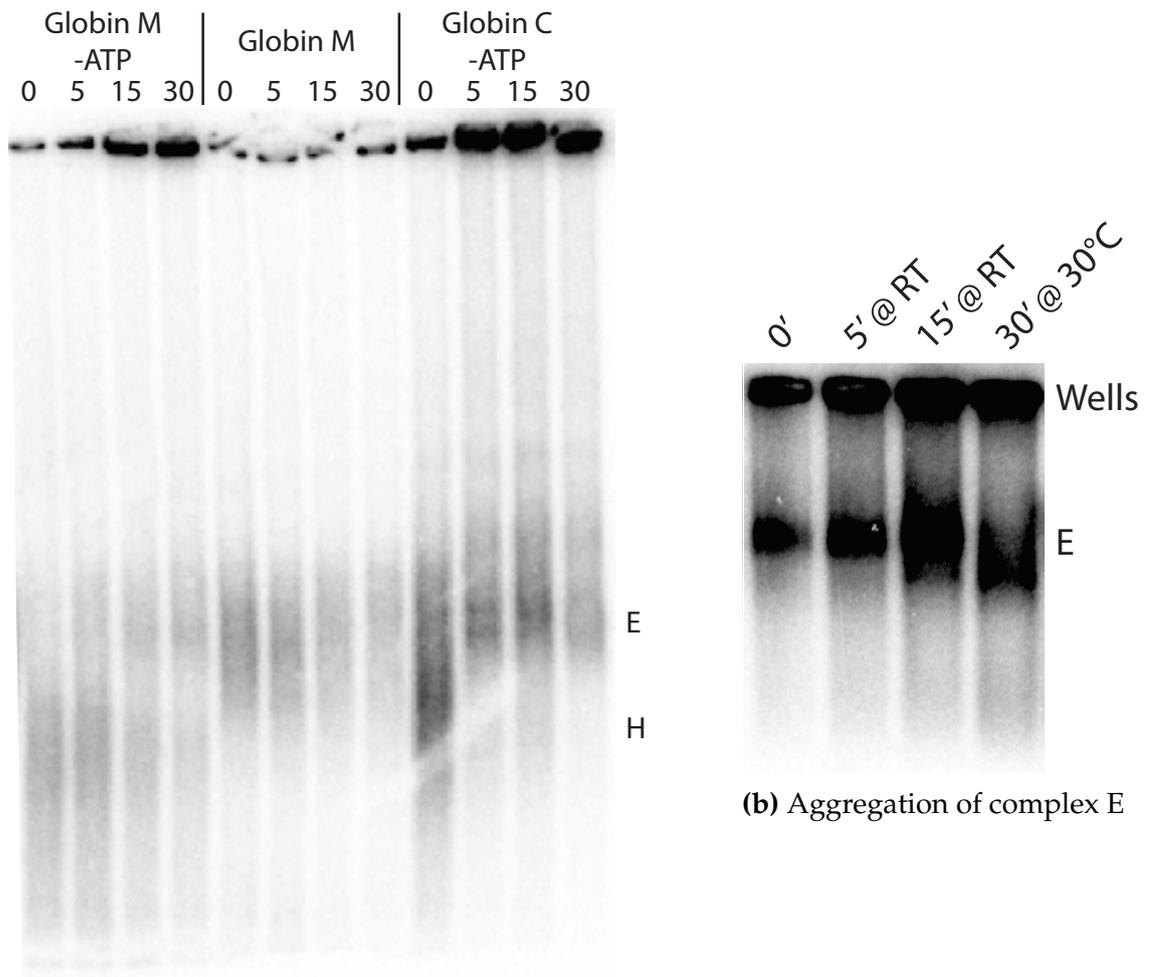
(b) Aggregation of complex E

Figure 5.2 – **(a)** Native agarose gel electrophoresis of samples on a 1 % agarose gel. The depletion of ATP leads to a stalling of the complex formation in complex E on Globin C where a consensus 5′ splice site is present, accompanied by aggregations that do not leave the wells of the gel. Globin M has a weak cryptic 5′ splice site. The formation of complexes and aggregations occurs much slower than for for Globin C. No distinct complex can be seen for Globin M under normal splicing conditions. - **(b)** Native agarose gel electrophoresis on a 2 % agarose gel of Globin C. The reaction was incubated for 0 min, 5 min and 15 min at room temperature and for 30 min at 30 °C to deplete ATP before the pre-mRNA was added to stall the splicing in complex E. The longer the incubation, the more pre-mRNA was part of aggregated complexes and would not migrate out of the wells of the gel.

5.2 SRSF1 in the Complexes E, A and C

The progression of the splicing machinery was stalled in the complexes E, A and C to investigate the involvement of SRSF1 within these splicing complexes.

In complex E, the pre-mRNA is equally either bound by a single SRSF1, two SRSF1 or by a multitude of them visible as a long tail in the distribution that ranges from 3 proteins to over 20 (mean value of SRSF1 present at each observed pre-mRNA $= 1.19 \pm 0.2$, Figure 5.3b). In the absence of ATP contain the biggest aggregations of SRSF1 multiple labelled pre-mRNA. Additionally, the aggregations of SRSF1 occur even in the absence of pre-mRNA.

A radical change occurs in complex A (p-value for the observed distribution, here in complex A, originating from the same underlying distribution, here complex E, < 0.0001). The binding of SRSF1 is almost exclusively reduced to a single remaining protein. This is also accompanied by a decrease in the colocalization (mean $= 0.32 \pm 0.06$, Figure 5.3b).

This change is reversed once complex C is reached ($p < 0.0001$). Both the colocalization as well as the actual number of SRSF1 present is increased over complex A (mean $= 1.17 \pm 0.17$). The distribution of the number of bound proteins differs from complex E ($p < 0.0001$, Figure 5.3c).

(a) Globin C in complex E

(b) Globin C in complex A

(c) Globin C GG in complex C
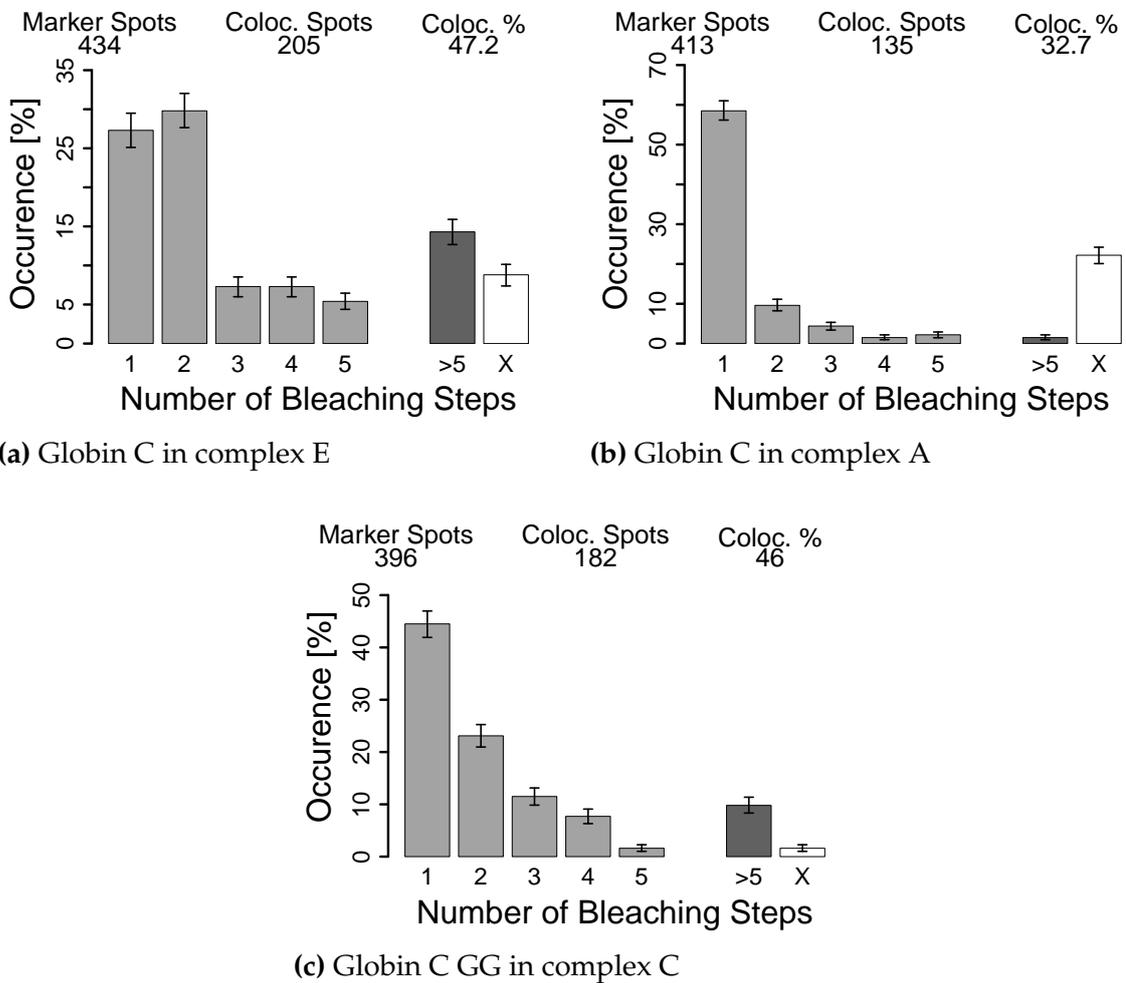
Figure 5.3 – The binding of SRSF1 to Globin C and Globin C GG. **(a)** Globin C in complex E is predominantly either bound by one or two SRSF1 or a multitude of them. - **(b)** The number of bound SRSF1 is strikingly reduced to a remaining single SRSF1 in complex A with a reduced colocalization percentage. - **(c)** In complex C, the number of bound SRSF1 and also the colocalization percentage is increased compared to complex A. The resulting distribution is different from both complex E and complex A.

5.3 The Dependence on the Number of 5' Splice Sites

The number of SRSF1 present in complex A might be linked to the number of 5' splice sites present. The adenovirus based pre-mRNA with two consensus splice sites showed a clear shift to two bound SRSF1 (Figure 5.4a), accompanied by 1 or 2 U1 snRNP (Figure 5.4b). Disabling the second 5' splice site reverted the binding distribution back to a single bound SRSF1 (Figure 5.4c). The same effect could be seen for a Globin C based construct where two identical exons with identical 5' splice site were used, although not as pronounced (Figure 5.4d). The same effect could not be observed for a third Globin C based construct with two 5' splice sites (Figure 5.4e).

(a) Ad1 CC in complex A

(b) U1A-GFP and Ad1 CC in complex A

(c) Ad1 CM in complex A

(d) CEC in complex A

(e) C174C in complex A
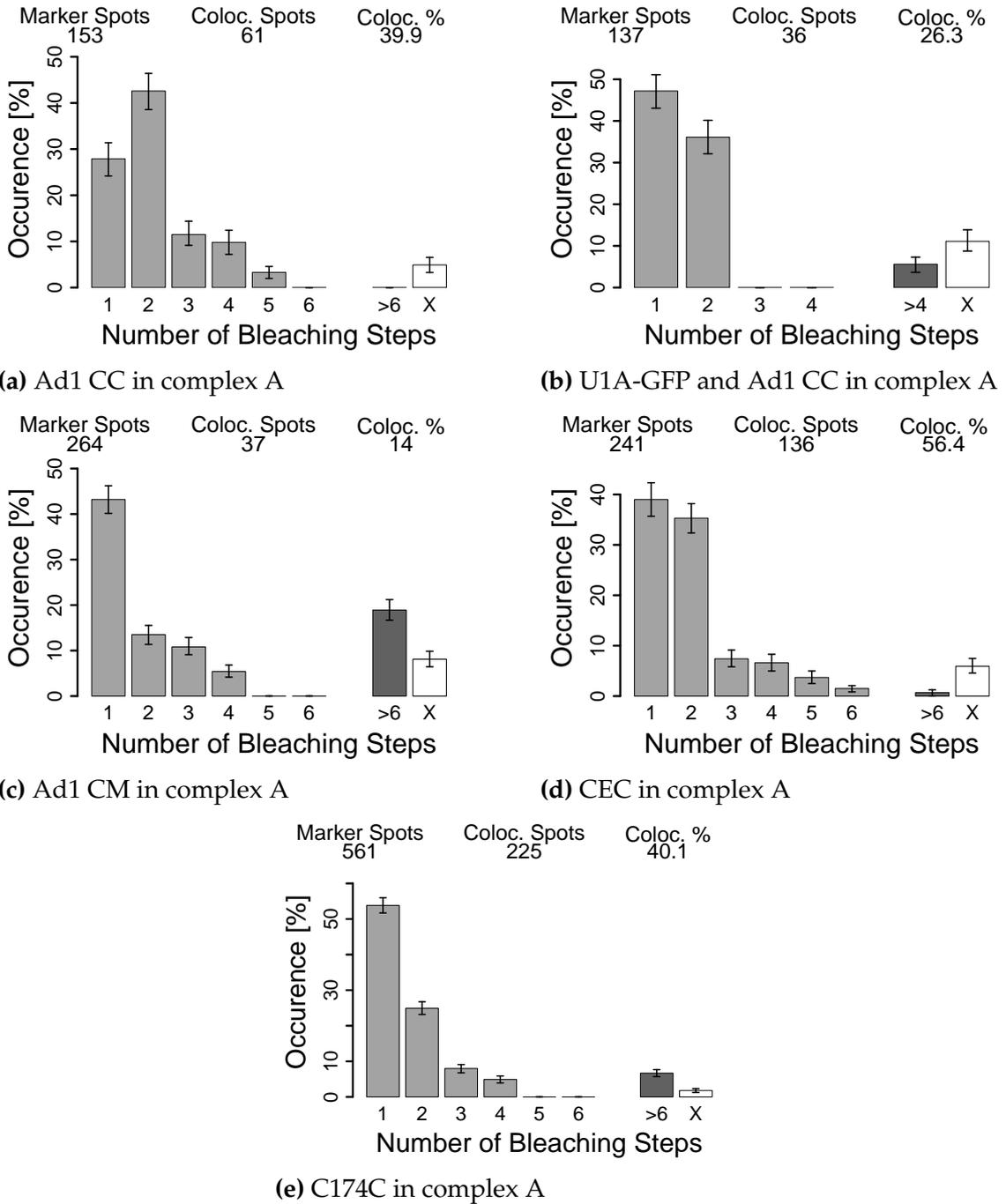
Figure 5.4 – The binding of SRSF1 in complex A. **(a)** The adenovirus pre-mRNA construct with two consensus 5′ splice sites shows a shift of the binding of SRSF1 to two bound proteins with **(b)** 1 or 2 U1 snRNP present. **(c)** The adenovirus based pre-mRNA construct with only one consensus 5′ splice site has predominantly only one bound SRSF1, although it retains bigger accumulations. - **(d)** The Globin C based pre-mRNA construct with a second identical 5′ splice site shows a shift towards two bound SRSF1. - **(e)** The Globin C based pre-mRNA construct has two consensus 5′ splice sites with mainly one bound SRSF1.

5.4 The Interplay of SRSF1 with the U1 snRNP and the 5' Splice Site

Globin M is a pre-mRNA construct derived from the the Globin C pre-mRNA by changing the consensus 5' splice site sequence to a non-splice site sequence. No binding of U1 snRNP should occur without a 5' splice site present. Additionally, an incubation with the αU1 oligo disables the RNA-binding ability of the U1 snRNP. Under conditions for complex E does this not affect the colocalization, but no clear binding pattern can be seen for SRSF1. It is dominated by big aggregations of it that form with the pre-mRNA (Figure 5.5a). Once ATP is added for complex A conditions, even though no nominal complex A can be formed, is the number of bound SRSF1 reduced again. The distribution now has a seemingly stochastic distribution (Figure 5.5b).

The reduction of bound SRSF1 is closely linked to the absence of the U1 snRNP, but not to the missing splice site. Without the additional blocking of the U1 snRNP was no similarly striking reduction of bound SRSF1 observed as seen with a consensus 5' splice site, albeit with reduced colocalization (Figure 5.5c).
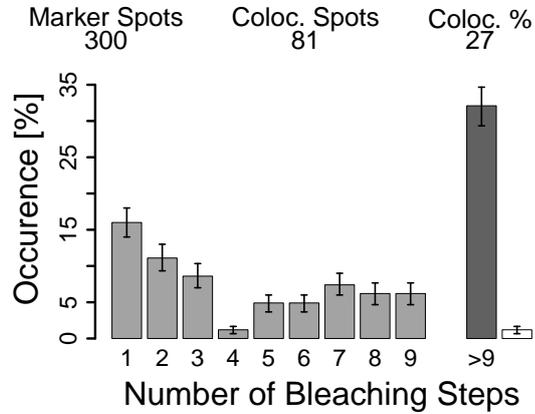
(a) Globin M without ATP with αU1
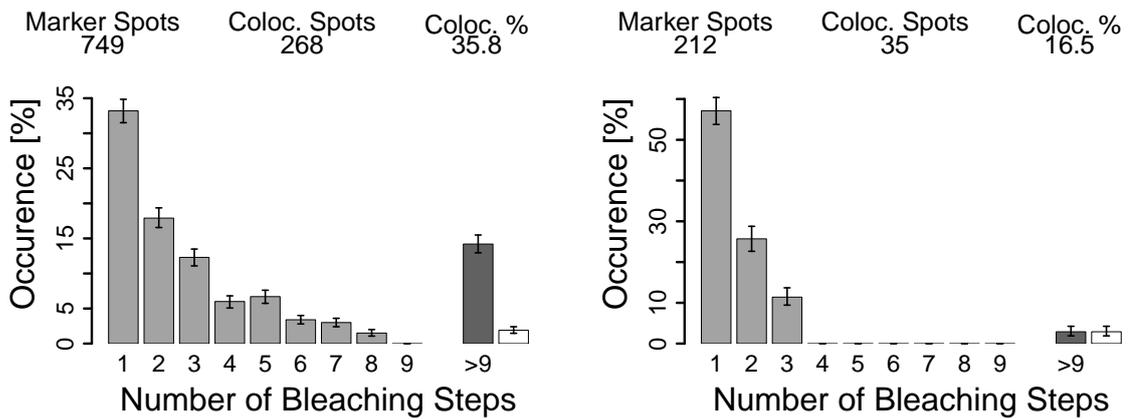
(b) Globin M with ATP and with αU1

(c) Globin M with ATP

Figure 5.5 – Globin M is derived from the the Globin C pre-mRNA by changing the consensus 5' splice site sequence to a non-splice site sequence. **(a)** Without the consensus 5' splice site and without the U1 snRNP the binding distribution of SRSF1 has no clear definition and is dominated by big accumulations of SRSF1 at the pre-mRNA. - **(b)** Blocking the U1 snRNP prevents the shift seen in (a), even though no 5' splice site is present in the pre-mRNA. - **(c)** Although Globin M does not form complex A due to the lack of a 5' splice site, incubation with ATP as for complex A does lead to the same striking shift to mostly one remaining bound SRSF1.

5.5 The Dependence of the Binding of SRSF1 on Phosphorylation

The phosphorylation of the domains of SRSF1 plays an important role in splicing. Pre-incubation of the nuclear extracts with PhosStop was done to prevent the general dephosphorylation of proteins to investigate the influence in changes of SRSF1's phosphorylation on its binding.

For Globin C in complex E with the added PhosStop does this result in a binding distribution that is similar to the one in complex A (Figure 5.6a and Figure 5.3b). This effect is clearly linked to the U1 snRNP as blocking the snRNP changes the occurring binding of SRSF1 to its opposite with large aggregations being present (Figure 5.6b). Using Globin M as a pre-mRNA construct without a consensus 5' splice site and a blocked U1 snRNP still shows an effect of PhosStop. The occurring binding distribution is split into two different, dominant parts, a single bound SRSF1 and aggregations (Figure 5.6c). An effect of the addition of PhosStop on Globin CC GG in complex C can be seen as the binding of SRSF1 is shifted towards two bound molecules of SRSF1 (Figure 5.6d).

Probing SRSF1 with mAb 104, an antibody that recognises phosphoepitopes present in SR proteins (Roth et al., 1990), showed no detectable effect of PhosStop on the phosphorylation of SRSF1 (C.M. Lucas, personal communication).
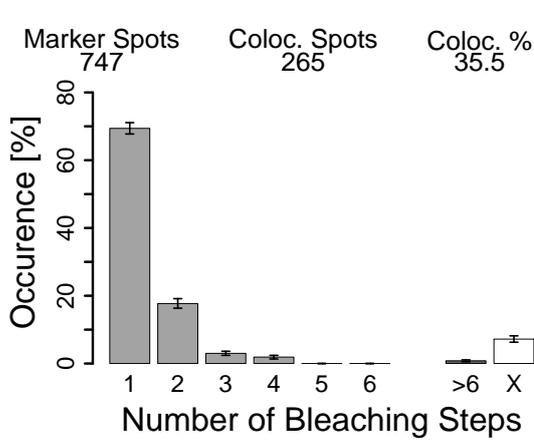
(a) Globin C in complex E with PhosStop
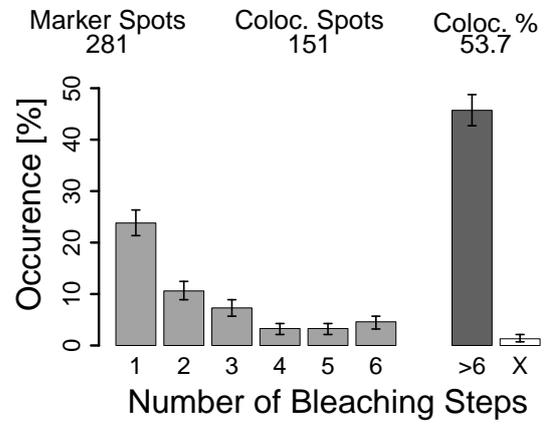
(b) Globin C in complex E with PhosStop and αU1
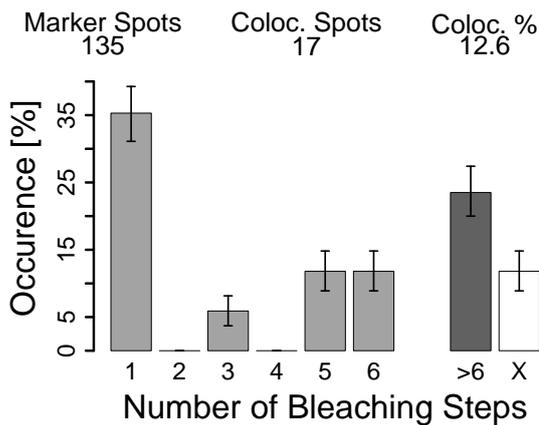
(c) Globin M in complex E with PhosStop and αU1
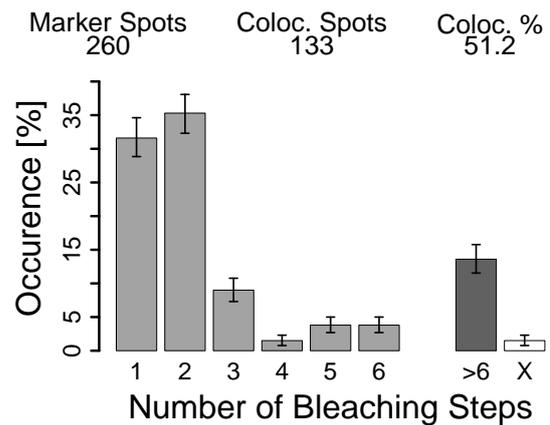
(d) Globin C GG in complex C with PhosStop

Figure 5.6 – **(a)** Using PhosStop in complex E results in the identical binding as seen for complex A without PhosStop. - **(b)** The combination of PhosStop and αU1 leads to a majority of aggregations at the pre-mRNA. - **(c)** The combination of Globin M and αU1 with PhosStop splits the distribution almost evenly into a single bound SRSF1 or aggregations of it. - **(d)** The binding of SRSF1 in complex C is dominated by one or two bound SRSF1.

5.6 The Role of the Domains of SRSF1

Two mutants of SRSF1 were made. SRSF1-ΔRS is missing its RS domain and SRSF1-ΔRRM2 is missing its RRM2 domain (Section 8.4). Both mutations showed a reduced colocalization percentage. The observed binding distributions for both mutations differ in both complexes from the non-mutated SRSF1 (each $p < 0.0001$, Figures 5.7a to 5.7d). No significant change occurs from complex E to complex A for SRSF1-ΔRS ($p = 0.09$) and both mutants have a similar binding distribution in complex A ($p = 0.41$). Only SRSF1-ΔRRM2 in complex E with its higher colocalization percentage differs significantly ($p < 0.0001$) from the rest. All four distributions conform highly to a Poisson distribution (ΔRS in complex E: $r^2 = 0.9991$, $\lambda = 0.08$; ΔRS in complex A: $r^2 = 0.9997$, $\lambda = 0.06$; ΔRRM2 in complex E: $r^2 = 0.9965$, $\lambda = 0.13$, ΔRRM2 in complex E: $r^2 = 0.9991$, $\lambda = 0.08$) with each having a reduced mean from complex E to complex A.

A remarkable feature not visible in the shown binding distributions was the greatly reduced non-specific binding of these SRSF1 mutants to the surface of the sample chamber. Non-mutated versions of SRSF1 were still present but only the mutations were labelled with GFP and visible to obtain their binding distribution.
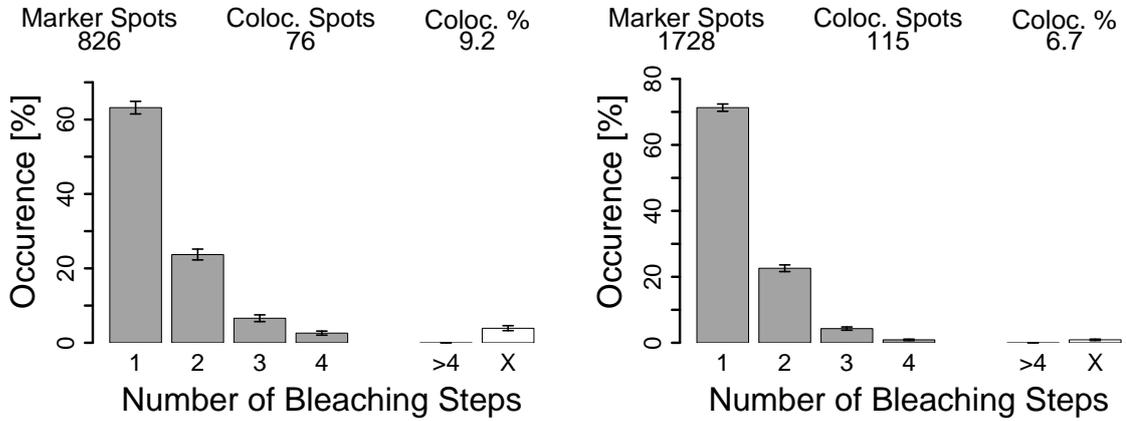
(a) Globin C with SRSF1-ΔRS in complex E

(b) Globin C with SRSF1-ΔRS in complex A
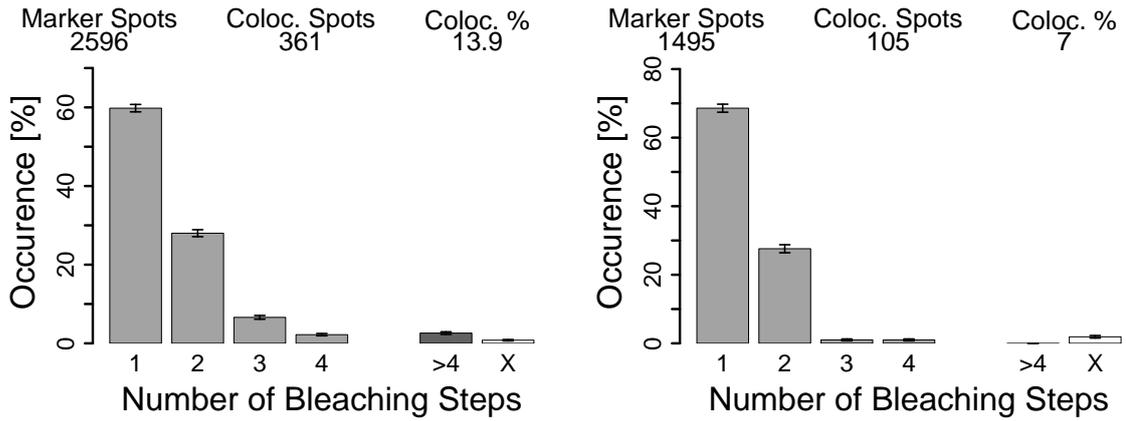
(c) Globin C with SRSF1-ΔRRM2 in complex E

(d) Globin C with SRSF1-ΔRRM2 in complex A

Figure 5.7 – The binding of the mutants of SRSF1 to the Globin C pre-mRNA. All mutants show a reduced binding as a reduced colocalization percentage and a reduced number of bound SRSF1. - SRSF1-ΔRS in **(a)** complex E and in **(b)** complex A. - SRSF1-ΔRRM2 in **(c)** complex E and in **(d)** in complex A.

5.7 The Interplay Between SRSF1 and hnRNP A1

Three colour fluorescence colocalization experiments allowed to investigate the interplay between hnRNP A1 and SRSF1 during the early splicing complexes E and A. HnRNP A1 was labelled with GFP and SRSF1 was labelled with mCherry.

In both complexes can SRSF1 and hnRNP A1 be present on the same pre-mRNA, in complex E even as part of big accumulations of both proteins. The number of bound SRSF1 is on average larger than hnRNP A1 but not necessarily for individual pre-mRNAs. The distribution found for complex A differs from the one for complex E with a reduced number of bound SRSF1 as seen before. The number of hnRNP A1 is similarly reduced but not as strikingly (Figures 5.8a and 5.8b).

A strong correlation of SRSF1 to hnRNP A1 was found in both complex E and complex A. The linear regression of SRSF1 to hnRNP A1 ($r^2 = 0.85$) showed that for each additional hnRNP A1 are on average 0.7 more SRSF1 present in complex E and similarly in complex A ($r^2 = 0.76$) with 0.7 more SRSF1 per additional hnRNP A1. Under the assumption that only lower numbers of these proteins compete for binding sites, and those with more than 2 are due to other interaction, is the correlation changed. The number of SRSF1 present decreases now from 1.42 ± 0.01 for no hnRNP A1 present to 0.76 ± 0.05 for a single hnRNP A1 present.

The average number of hnRNP A1 on pre-mRNA in complex E without any detected SRSF1 present is 1.33 ± 0.1. The average number of hnRNP A1 present decreases to 0.25 ± 0.06 for a single SRSF1 present and increases for every additional SRSF1 but only reaches a higher averaged occupancy for 5 or more SRSF1 present. The same is true in complex A, where the shift from no SRSF1 present to a single SRSF1 present is from 2.5 ± 0.3 to 1.1 ± 0.4 hnRNP A1 bound.
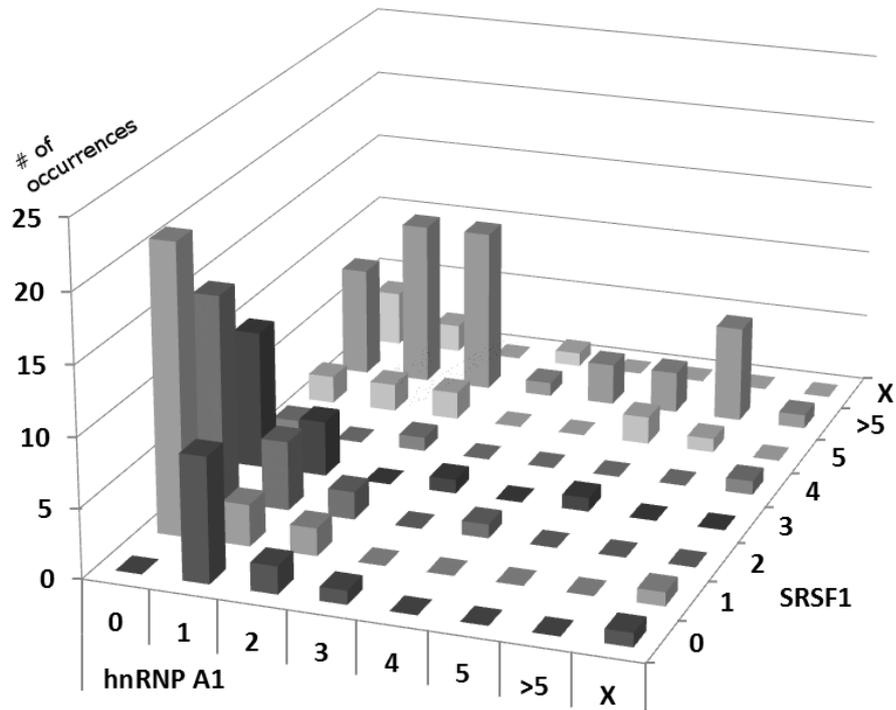
(a) Binding of hnRNP A1 and SRSF1 in complex E
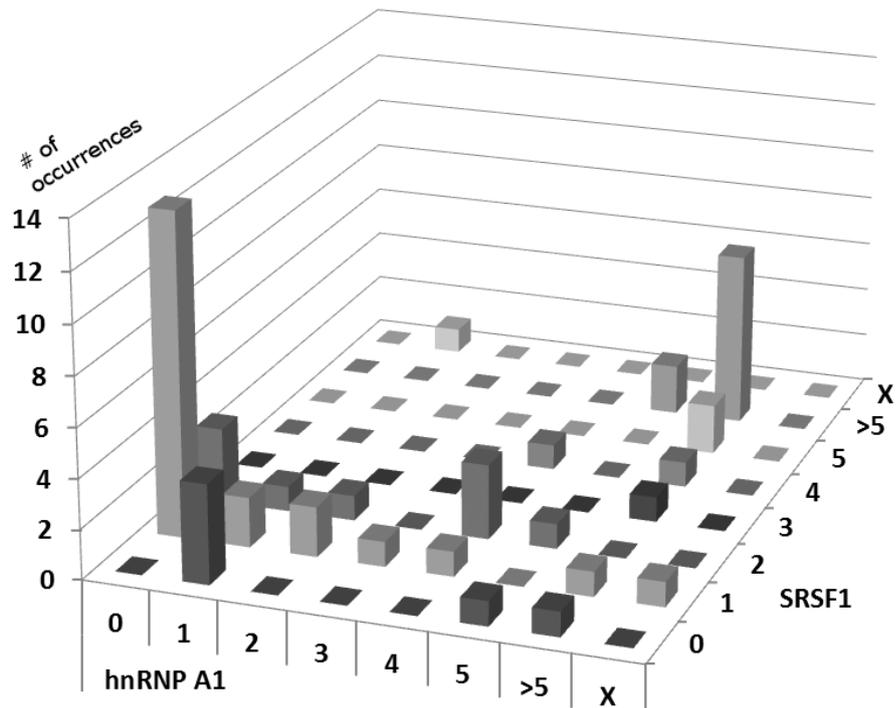
(b) Binding of hnRNP A1 and SRSF1 in complex A

Figure 5.8 – Cross-correlation of the binding of hnRNP A1-GFP and SRSF1-mCherry to the Globin C pre-mRNA in **(a)** complex E and **(b)** complex A.

5.8 Control Experiments

Splicing is an elaborate system influenced by a multitude of factors. Several control experiments were performed to test the validity of the previous findings.

The binding distributions could be affected by the way splicing is stalled in the different complexes. The same experiments were done with timepoints that correspond to the investigated complexes without actually stalling the complex formation through the absence of ATP, the addition of interfering oligos or mutations. The so obtained results showed different binding distributions from their stalled counterparts (each $p < 0.0001$) and were similar to each other (10 s compared to 5 min $p = 0.08$, 5 min compared to 30 min $p = 0.15$). The mean of the SRSF1 present does decrease slightly at 5 min and increases again at 30 min (1.04 ± 0.21, 0.94 ± 0.1 and 1.12 ± 0.13). All three experiments showed a remarkably higher and similar colocalization (Figures 5.9a to 5.9c).

The amount of SRSF1 present in the nuclear extracts may play a role influencing the number of bound SRSF1. The observed aggregations in complex E might be a result of the relative abundance of SRSF1. Experiments with an abundance of yeast RNA were done to lower the ratio of SRSF1 to RNA. The colocalization is lowered but otherwise the same distribution remains ($p = 0.24$, Figure 5.10a). An even greater abundance of SRSF1 as found in nuclear extract made from 293T cell for an higher ratio of SRSF1 to RNA led to a high colocalization with binding dominated by large accumulations of SRSF1 (Figures 5.10b and 5.10b).

The observed non-specific binding of SRSF1 to the glass surface of the sample chamber might interfere with the observed results, especially with the occurring confinement of the pre-mRNA to the surface even without any specific binding mechanism present. Experiments on top of a glass surface covered in PEG were performed and obtained similar binding distributions with a reduced colocalization percentage (Figures 5.10c and 5.10c).

(a) Globin C after an incubation of 10 s

(b) Globin C after an incubation of 5 min

(c) Globin C after an incubation of 30 min
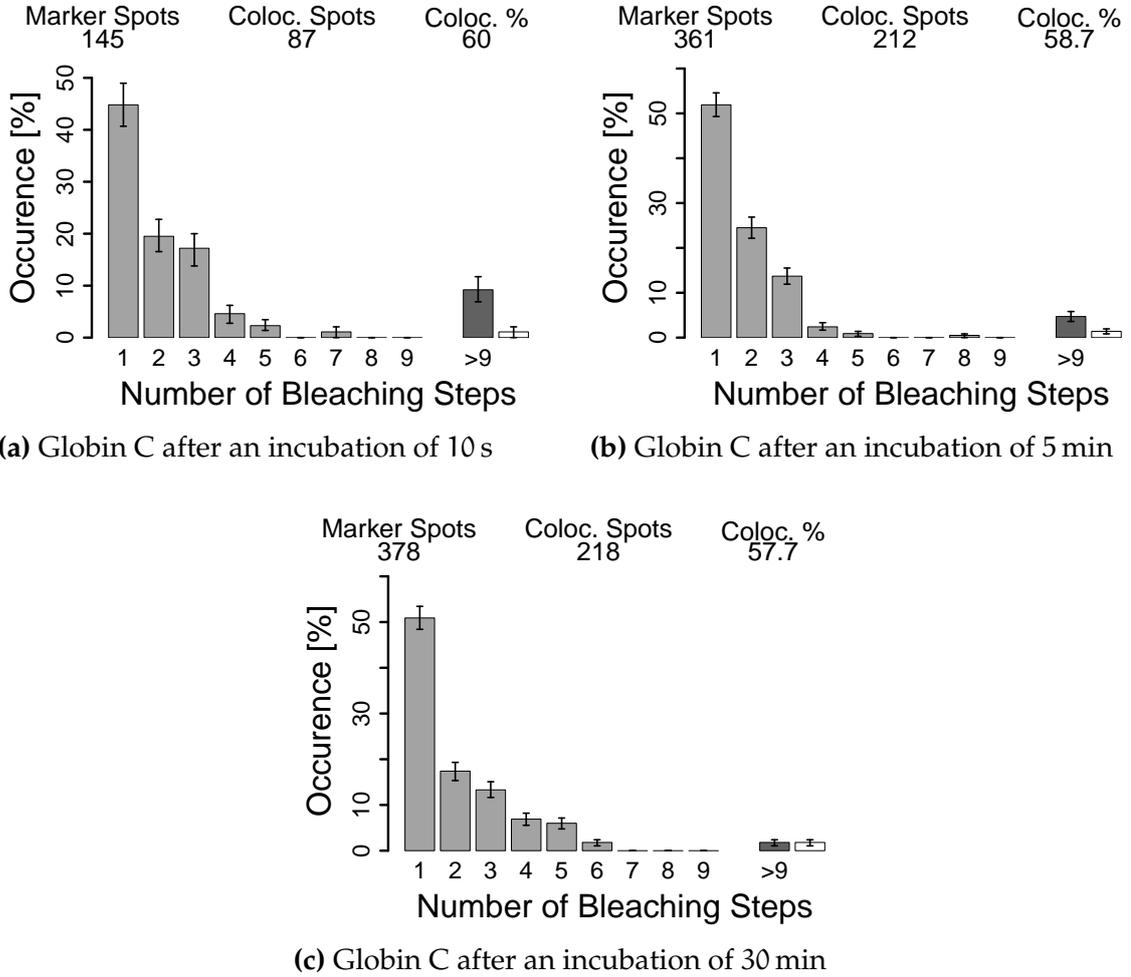
Figure 5.9 – The binding of SRSF1 at different timepoints that correspond to the complexes E, A and C without actually stalling the complex formation. **(a)** The timepoint of 10 s corresponds to complex E, see Figure 5.3a. - **(b)** The timepoint of 5 min corresponds to complex A, see Figure 5.3b. - **(c)** The timepoint of 30 min corresponds to complex C, see Figure 5.3c.
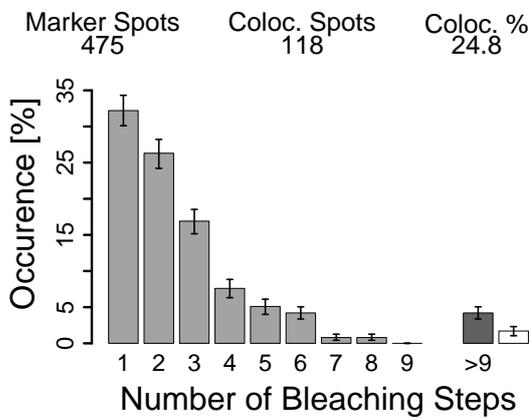
(a) Globin C in complex E with an abundance of yeast RNA
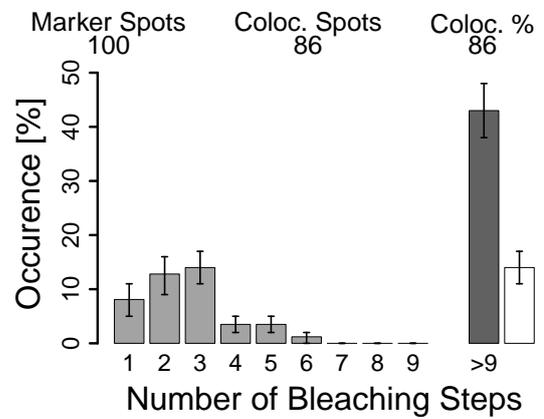
(b) Globin C in complex A in nuclear extract from 293T cells
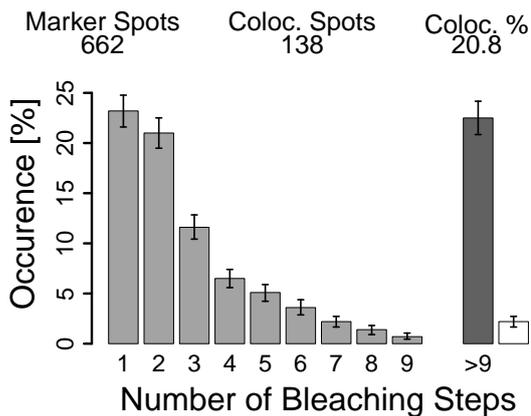
(c) Globin C in complex E on a PEG surface
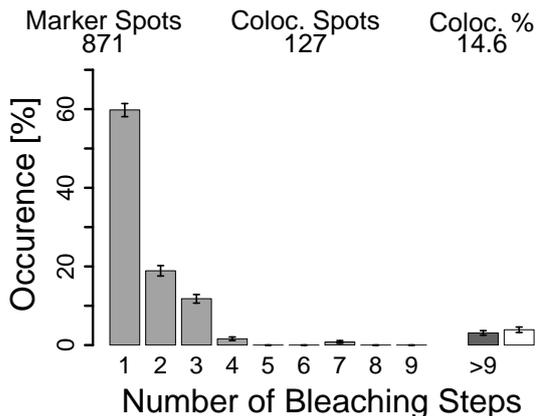
(d) Globin C in complex A on a PEG surface

Figure 5.10 – **(a)** Globin C pre-mRNA was incubated together with an ten times abundance of yeast RNA and stalled in complex E. - **(b)** Nuclear extract from 293T cells had exceptional high levels of SRSF1 and resulted in a remarkable colocalization and binding of SRSF1. - Single-molecule experiments with Globin C performed on top of a PEG coated surface stalled in **(c)** complex E and **(d)** complex A.

5.9 SRSF1 as a Splicing Enhancer

SRSF1 is generally seen as a splicing enhancer, acting through its binding to an ESE. Two constructs based on Globin C were used to investigate if an effect of the ESE can be seen on the observed binding pattern of SRSF1. The first construct, Short Exon Globin C, has a much shorter leading exon and an optional ESE on its 3′ end. The second construct, 2xMut Globin C, has one mutation at its branch point and one at the polypyrimidine tract and an optional ESE at its 3′ end.

The complex formation for both constructs showed only slight differences with or without the ESE present. 2xMut Globin C did not form any stable complexes besides complex H. Blocking the U1 snRNP greatly reduced the amount of formed complexes, but not as much for Short Exon Globin C (Figure 5.11).

2xMut Globin C does not splice, with or without its ESE, in agreement with its poor formation of complexes (data not shown). The enhancing effect of the ESE can be seen on Short Exon Globin C. Without the ESE does it splice with only 6 % efficiency after 120 min, compared to 28 % if it has the optional ESE at its 3′ end. Its splicing efficiency is much less reduced compared to Globin C if the U1 snRNP is blocked, again in agreement with the observed complex formation (Figure 5.12).

It was further looked at the binding of SRSF1 in complex E and A to investigate the relationship between splicing efficiency, complex formation and the observed binding distribution. The binding distribution for 2xMut Globin C in conditions for complex E, despite not forming any complex past complex H, is similar to normal Globin C. Complex A conditions lead to the loss of aggregations, but not clear shift can be seen for the number of SRSF1 present that are not part of a big accumulation. The presence of the ESE made no significant difference (Figure 5.13).

Short Exon Globin C does form all complexes comparable to Globin C. It shows a much reduced colocalization in complex E with a much reduced number of bound SRSF1. The colocalization increases for complex A, accompanied with a

147
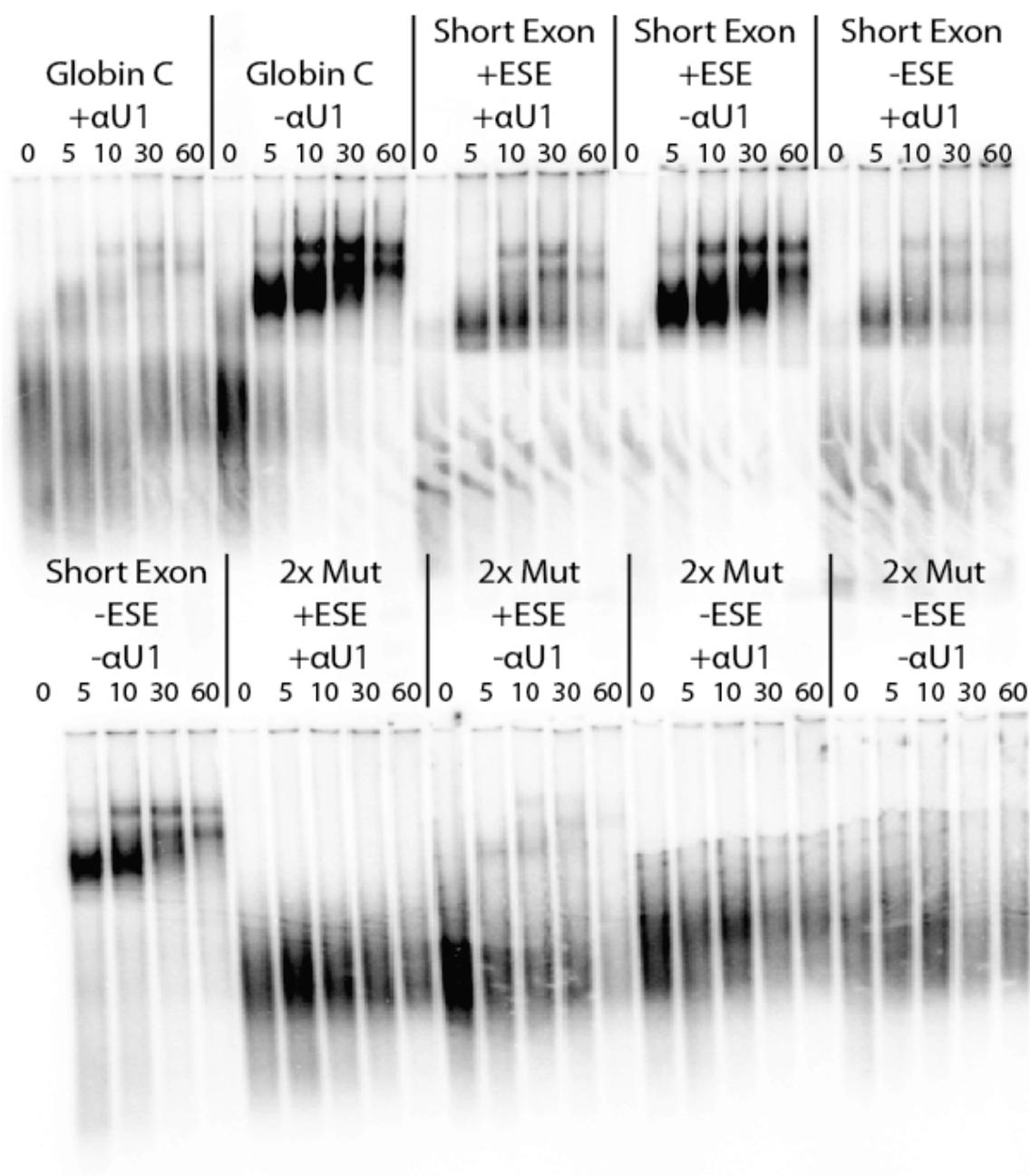
Figure 5.11 – Influence of an ESE and the U1 snRNP on complex formation. Native gel electrophoresis of samples treated with heparin and run on a 2 % agarose gel. The constructs are based on Globin C with and without an ESE and were incubated with and without the αU1 oligo. Timepoints were taken from 0 min to 60 min. The 2xMut construct does not form any recognisable complexes beyond complex H. The shortened Globin construct with and without the ESE does form complexes even in the absence of the U1 snRNP, albeit at lower levels.
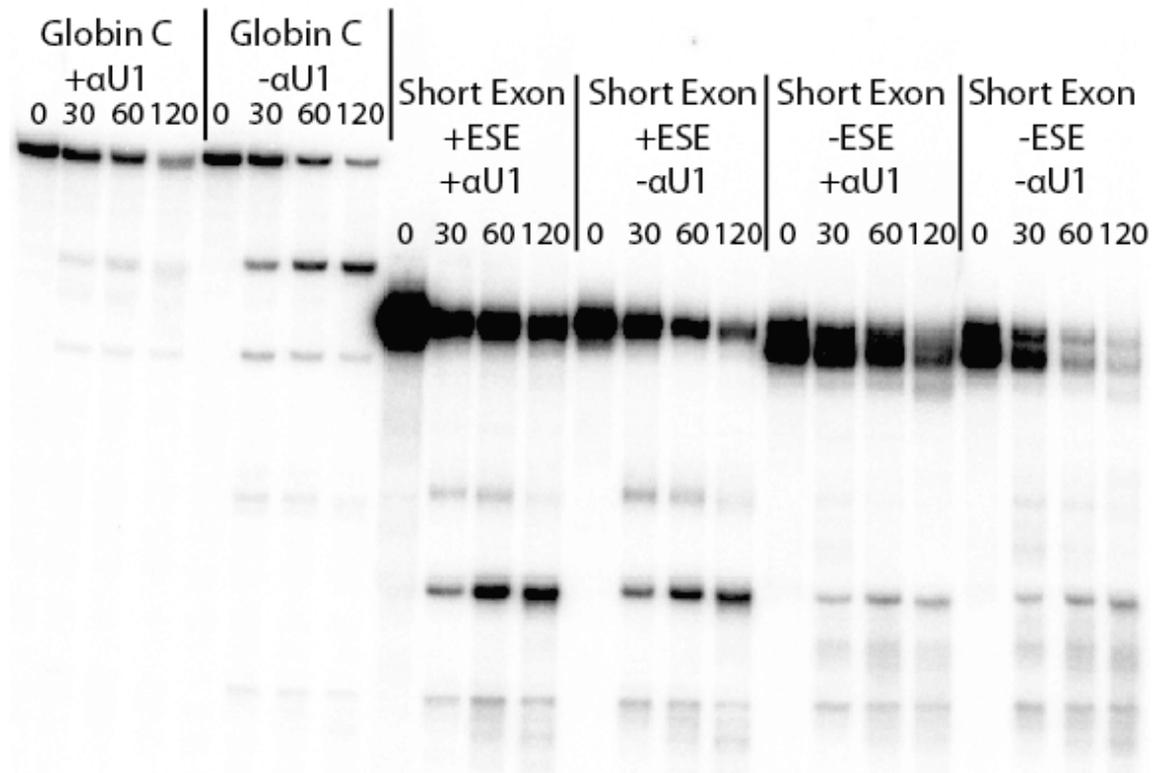
(a) Splicing
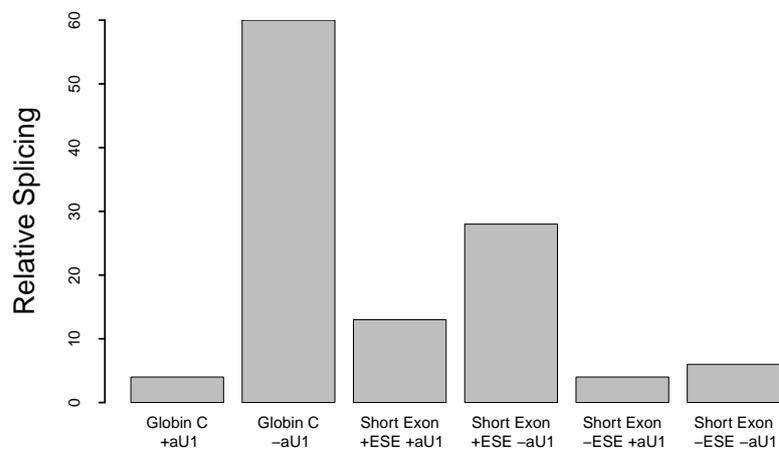
(b) Relative splicing efficiency

Figure 5.12 – Influence of an ESE and the U1 snRNP on Splicing. - **(a)** Gel electrophoresis of samples run on a 7 % polyacrylamide gel. Short Exon is Globin C with shortened exons with and without an ESE, incubated with and without the αU1 oligo. Timepoints were taken from 0 min to 120 min. - **(b)** Relative splicing efficiency after 120 min from (a). The Short Exon construct splices better with an ESE present by a factor of 5, but shows not an as pronounced difference with or without the U1 snRNP as Globin C.
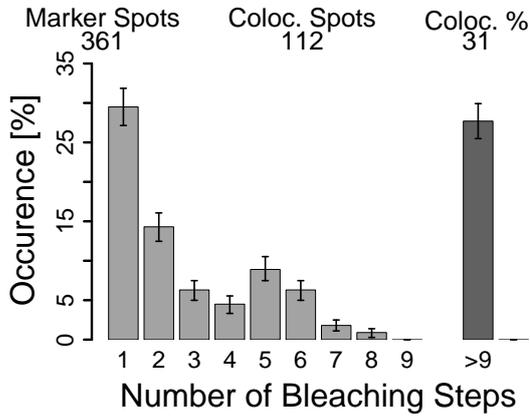
(a) 2xMut+ESE in complex E
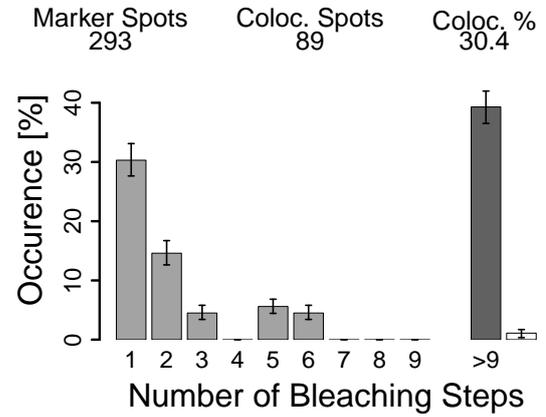
(b) 2xMut-ESE in complex E
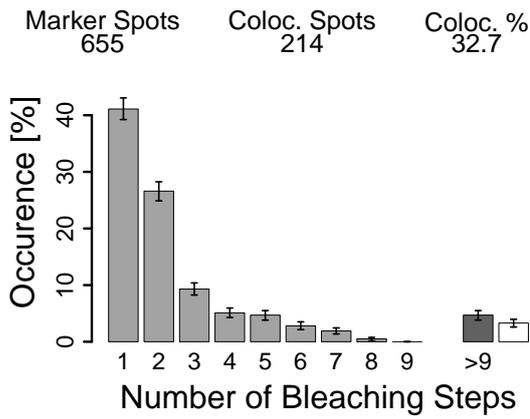
(c) 2xMut+ESE in complex A
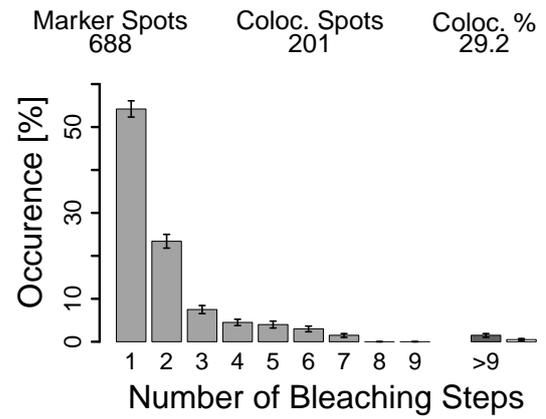
(d) 2xMut-ESE in complex A

Figure 5.13 – **(a) - (d)** 2xMut with and without ESE in Complex E and A. No significant difference between +ESE and -ESE can be found but the same shift towards lesser bound SRSF1 in complex A occurs for both variants.

(a) Short Globin C+ESE in complex E

(b) Short Globin C-ESE in complex E

(c) Short Globin C+ESE in complex A

(d) Short Globin C-ESE in complex A
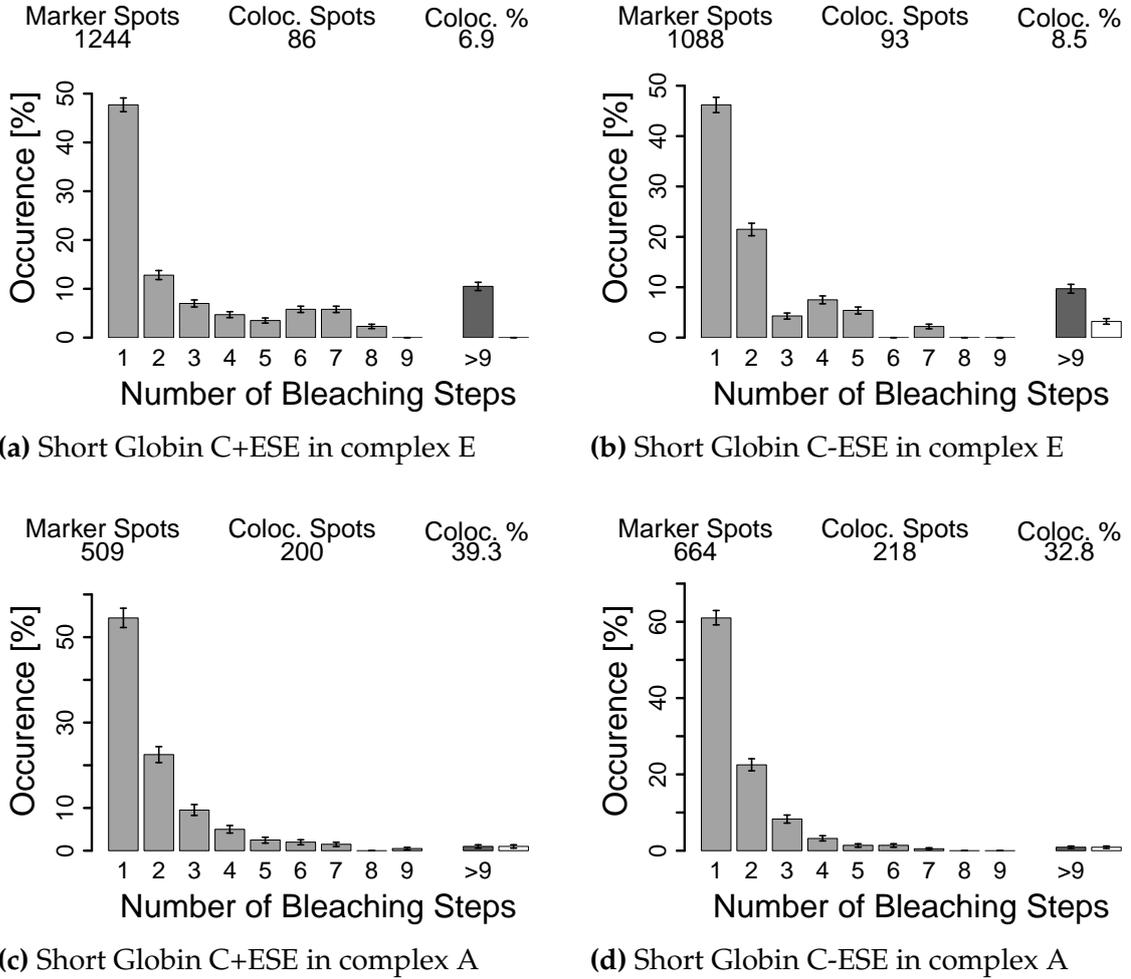
Figure 5.14 – **(a) - (d)** Short Exon Globin C with and without ESE in Complex E and A. No significant difference between +ESE and -ESE can be found but the same shift towards lesser bound SRSF1 in complex A occurs for both variants. The colocalization in complex E is much lower than for complex A, a unique occurrence among the tested constructs.

reduced number of bound SRSF1 for colocalized pre-mRNA. The presence of the ESE made no significant difference (Figure 5.14).

Generation of Fluorescent Droplets by Using Microfluidics

𝕿he ultimate goal of this part of this work was to observe fluorescence originating from single biological molecules that are encapsulated inside droplets generated inside microfluidic devices by flow focussing.

The first part of the chapter describes the setup of a robust and convenient system to create functioning microfluidic devices and characterise this process and the devices produced. A suitable surfactant that is both compatible with the encapsulated biological system and allows the generation of stable aqueous droplets in oil has to be identified. The size of the generated droplets has to be decreased enough to allow the sensitive detection of single-molecule fluorescence inside. A suitable method to collect the generated droplets and imaging the contents by fluorescence microscopy is required.

6.1 The Design of the Photomask

The starting point of every photolithographic process for the production of a microfluidic device is the photomask. An initial design for a mixer with a 5 μm junction was obtained from the group of Andrew Griffiths as an emulsion film mask. The design was recreated as a vector graphic using Adobe Illustrator (Figure 6.1a). The left circle is the inlet for the continuous phase and the middle circle is the inlet for the dispersed phase. The smaller circles within those inlets act both as pillars to prevent the sagging of the ceiling as well as a first rough filter. They are enclosed by two rings of blocks that form a system of channels with the widths of the channels decreasing from 11 μm down to 8 μm to act as a filter (Figure 6.1e). Any particles larger than 8 μm are trapped and prevented from blocking a fluidic channel downstream. A filter system is not necessary for the outlet on the right. The connecting channels have a width of 170 μm and intersect at a junction where the width is reduced to 5 μm for a short distance towards the inlet of the dispersed phase and for a short distance towards the outlet (Figure 6.1c).

This initial design was improved for the requirements of this thesis (Figure 6.1b). The amount of pillars in the inlets and outlet was increased as it was found that they provided a visual aid to identify the inlet and outlet positions when punching the holes. The junction was uniformly reduced in size to have a width of 3 μm (Figure 6.1d). Equivalently, the channel widths at the filter system at the two inlets now range from 8 μm down to 5 μm. The width of the connecting channels was reduced to 60 μm and the distance from the junction to the outlet was reduced as well to improve the ease of handling and to reduce the required minimal amount of fluids. Chrome covering a soda lime glass, rather than an acetate sheet, was used as a mask as the smallest feature size of this improved design was below the resolution limit for emulsion film masks (JD Photo-Tools).
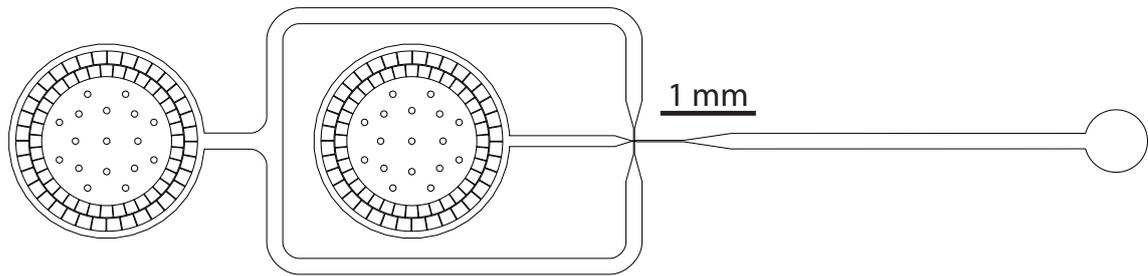
(a) Initial design for a mixer with a 5 µm junction
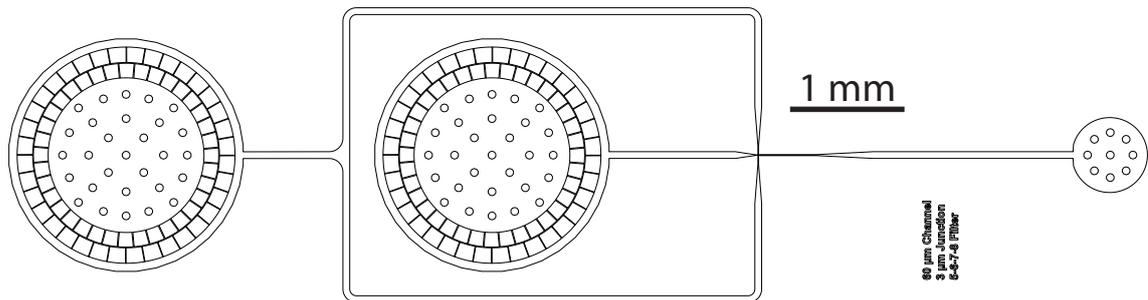
(b) Improved design for a mixer with a 3 µm junction
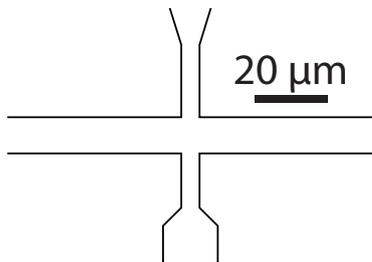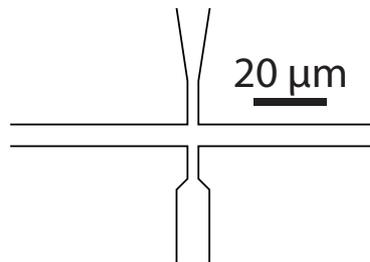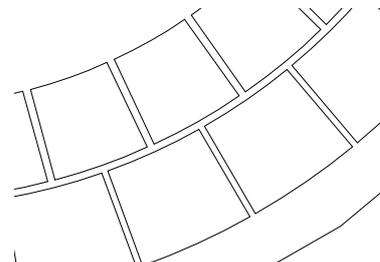
(c) 5 µm junction   **(d)** 3 µm junction   **(e)** Filter system

Figure 6.1 – **(a)** Initial design for the mixer with a 5 µm junction. The channels are 170 µm wide and the smallest width for a channel in the filter system is 8 µm. The inlet for the continuous phase is on the left, the inlet for the dispersed phase is in the middle, and the outlet is on the right. - **(b)** Improved design for the mixer with a 3 µm junction. The channels are 60 µm wide and the smallest width for a channel in the filter size is 5 µm. - **(c)** Blow up of the junction corresponding to the 5 µm design shown in (a). The design is turned clockwise by 90°. - **(d)** Blow up of the junction corresponding to the 3 µm design shown in (b). The design is turned clockwise by 90°. All lengths are reduced compared to (c). - **(e)** Blow up of the filter. The width of the channels between the filter blocks decreases from 11 µm down to 8 µm.

6.2 The UV Light Exposure

The exposure of the photoresist with UV light is a crucial step during the fabrication process. A homogeneous and perpendicular illumination over the whole width of the wafer is necessary for an accurate transfer of the photomask pattern onto the wafer. The exposure is commonly done with the UV attachment of a commercially available mask aligner (for examples see Clausell-Tormos et al., 2010; Rogers et al., 1998). These provide an exceptional quality of illumination and alignment of wafer and photomask with a precision on the micrometre scale, but the price of the equipment can be a significant entry barrier for research with microfluidics.

As part of setting up the production process, it was explored whether a cheaper, more compact and easier to use home-made solution could be employed for less stringent requirements. The homogeneity of the illumination must be sustained over the diameter of the 2 in wafers used in this work. The illumination can slightly deviate from being perfectly uniform in contact-mode photolithography for features that are only 2 µm deep. Alignment of the wafer and photomask is not necessary as there is only a single illumination step with a one-layered design.

Commercial mask aligners employ mercury lamps that need a time of >20 min to warm up, a long cool down period and are inflexible for alignment. Nowadays light-emitting diodes (LEDs) are able to provide both a similar power output and the necessary wavelengths while requiring less than a few seconds to warm up, no external cooling and are more compact.

Suitable LED devices were required for both the flexible emulsion film mask and the rigid chrome covered mask (Section 6.1). The built device for the emulsion film mask consists of six LEDs arranged in a hexagon around one central LED above two lenses that collimate the light onto a sandwich comprising the glass plate, the mask and the wafer (Figure 6.2a). The final device is less than 30 cm
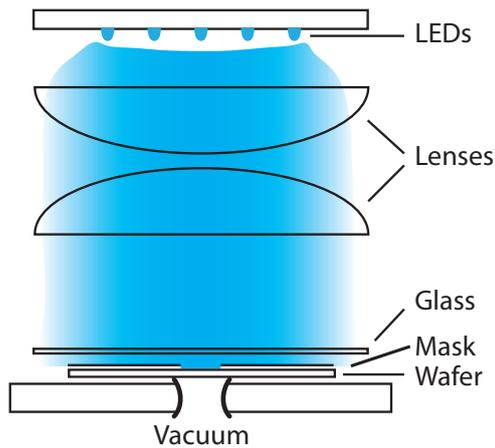
(a) Schematic
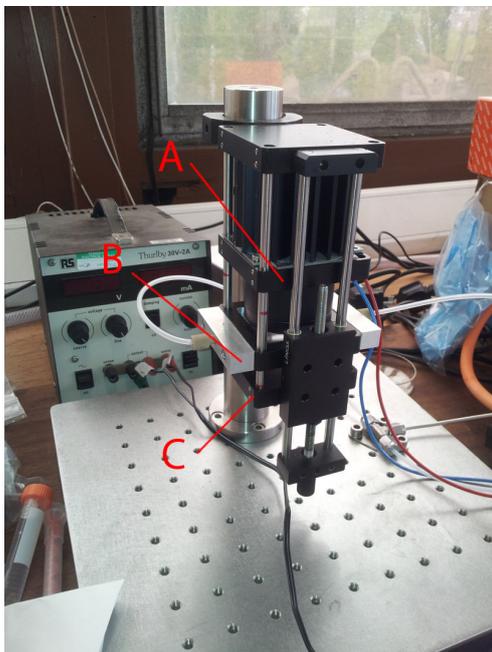
(b) Illumination device with self-assembled light source
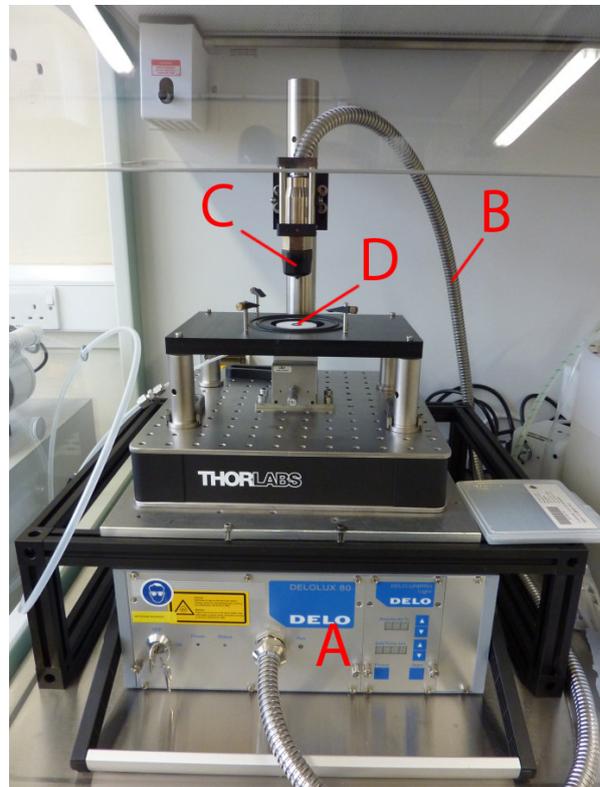
(c) Illumination device with commercial light source

Figure 6.2 – Devices for the UV light exposure during the photolithographic process. **(a)** Schematic for the device shown in (b). Two lenses collimate the light from seven UV LEDs at the top to provide a homogeneous and perpendicular illumination of the mask. The mask is sandwiched between the wafer at the bottom and a glass plate at the top. - **(b)** UV light exposure device for emulsion film substrates and 2 in wafers. The LEDs and lenses (A) are located above the glass plate (B). The mount for the wafer (C) can be moved vertically. - **(c)** UV light exposure device for glass substrate masks and 2 in wafers. The mount for the mask (D) sits on top of the commercial light source (A). The UV light is guided to the top of the mount (C) by a fibre system (B).

high and fits on a small bread board (Figure 6.2b). The wafer lies at the bottom part of the device on a stage. The glass plate is situated between the two lenses and the wafer and can be moved vertically. The emulsion film mask is put below this glass plate and held there by an applied vacuum. The glass plate with the mask attached is lowered to bring the mask and the wafer into close contact. The LEDs are switched on for the required time before the glass plate and the mask are elevated again, leaving the exposed wafer lying on the bottom stage.

The second instrument for the chrome mask uses a commercially available LED light source (DELOLUX 80, DELO Industrial Adhesives, Figure 6.2c). The wafer is held by vacuum on top of the central mount and can be moved vertically. Above, the mask is fixed by three clamps on the central table. The wafer is raised until it is in contact with the mask. A separate glass plate is not necessary as the chrome mask is already rigid compared to the flexible emulsion mask. The UV light is guided to the top of the device and switched on for the required time before the UV illuminated wafer is lowered again.

The homogeneity of illumination has been determined across the wafer for both devices. The UV light intensity was measured at the level of the wafer in 1 mm intervals across the entire illuminated area. Over this area, the device with the self-assembled light source shows a variation of its intensity of $\pm 6\,\%$ compared to the device with the commercial light source where the intensity varies by $\pm 10\,\%$. For this device, a larger area can be illuminated, although the UV light intensity might vary by up to $\pm 22\,\%$. The achieved intensity variations are twice the values reported in the specifications of commercially available mask aligners (see *MA/BA Product Brochure*). However, the homogeneity of the UV light illumination in the devices shown in Figure 6.2 is sufficient for the application needed for this work (see Section 6.3).

The incident angle of the light was not measured. The requirements for this
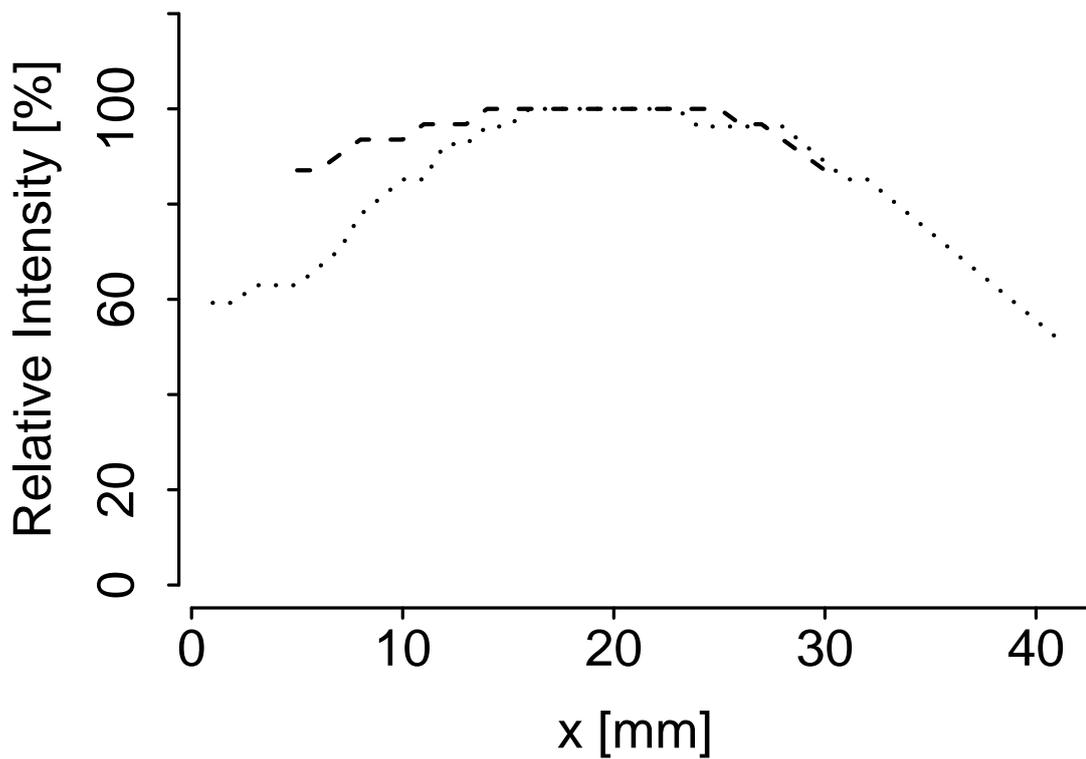
Figure 6.3 – Relative intensities across the wafer. The profile was measured once for each device. The relative intensity profile for the device in Figure 6.2b is shown as a dashed line, and for the device in Figure 6.2c it is shown as a dotted line. The relative intensity of 100 % is the maximum intensity for each device and x is the position across the illuminated area. The smaller device shows a more homogeneous illumination but covers only a smaller area.

are less stringent due to the close contact of mask and wafer and desired final feature heights of about 2 μm. Surface profile measurements of the final wafers showed that a sufficiently steep vertical definition of the protruding features was achieved (Figure 6.4).

6.3 The Photolithography Process

As presented in detail in Section 1.3.7, the photolithography process has five essential steps: (i) coating the substrate with a photoresist, (ii) soft bake, (iii) UV light exposure, (iv) post-exposure bake and (v) development. The photoresist used in this work is the permanent epoxy negative photoresist SU-8 2002 from Michrochem. For a target height of the features of 2 μm, the processing guidelines (see *SU-8 2000 Datasheet*) recommend to coat the wafer by spin coating at 2000 rpm followed by a soft bake for 1 min at 95 °C, an exposure energy of 60 mJ/cm² to 80 mJ/cm², a post bake for 1 min to 2 min at 95 °C and a development time of 1 min.

Initial trials with these parameters resulted in no visible transfer of the pattern from the photomask onto the wafer. The parameters could not be optimised for each step of the photolithography process individually as their influence on the final quality of the process could only be judged in conjunction with all other necessary steps. All parameters were varied until a visible transfer of the pattern onto the wafer was achieved.

The crucial parameter was found to be the exposure energy of the UV light source. Slight variations from the optimal value resulted in the transfer of features that were not firmly attached to the wafer surface. The adjusted protocol for the photolithography process was spin coating at 1500 rpm followed by a soft bake for 90 s at 95 °C and an exposure for 36 s at 18 mW/cm² (648 mJ/cm²) for the device with the commercial light source and for 3 min at 3.1 mW/cm² (558 mJ/cm²) for the device with the self-assembled light source. The exposure energies used are much higher than the energies suggested in the processing guidelines for SU-8 2002. The post exposure bake is for 6 min at 95 °C and a development time of 10 s.

It was found that it is not necessary to implement the photolithography process in a clean room facility. Instead, a laminar flow cabinet (Esco) was found to be sufficient.
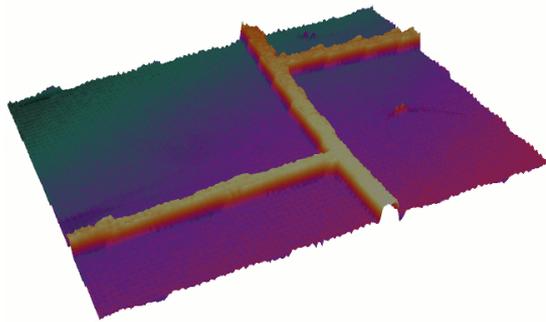
(a) 2D surface profile of the inlet filter
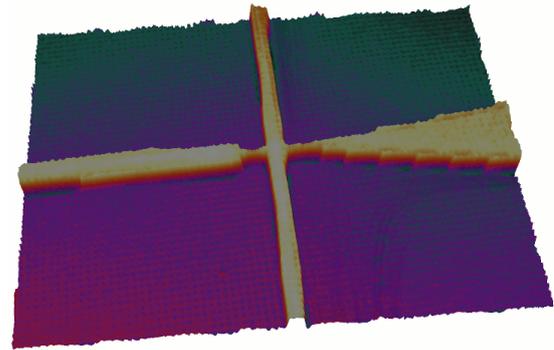
(b) 2D surface profile of the 3 µm junction

30 µm
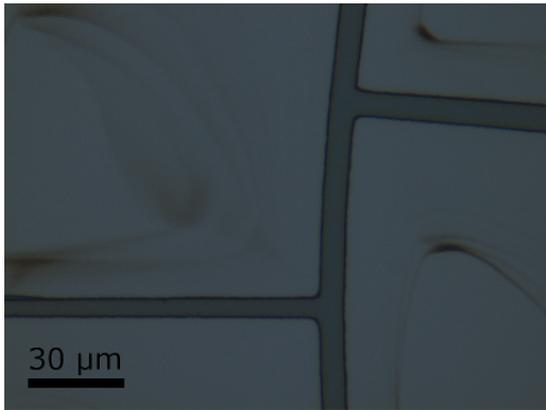
(c) Top view of the inlet filter

30 µm
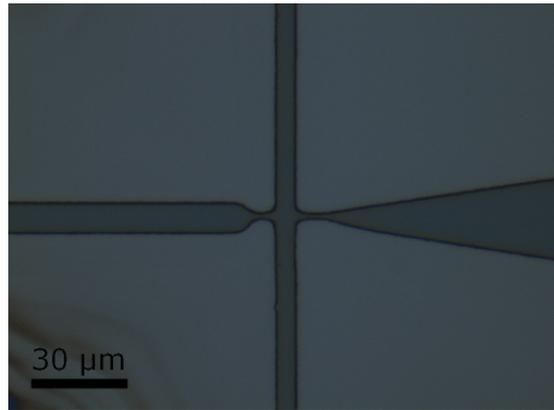
(d) Top view of the 3 µm junction
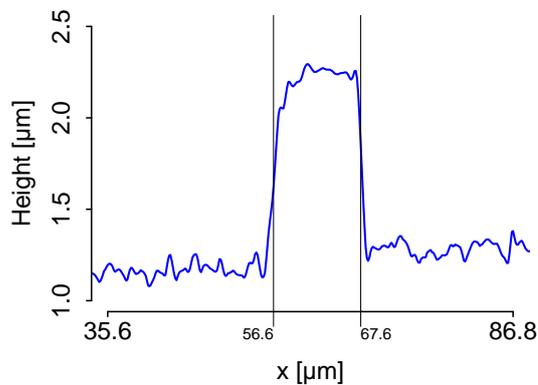
(e) Surface profile
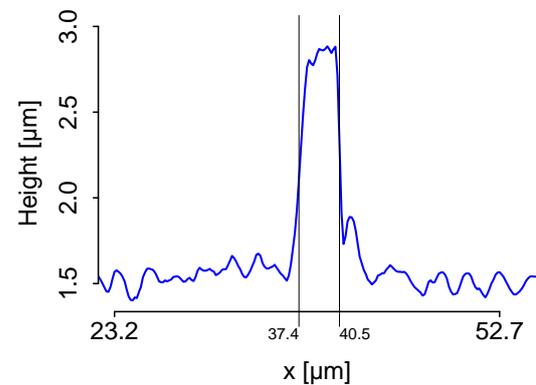
(f) Surface profile

Figure 6.4 – Data acquired with Zeta-20 by interferometry. **(a)** The 2D surface profile of the inlet filter system has 9 µm wide channels between the filter blocks. - **(b)** The 2D surface profile shows the 3 µm junction in the middle. The water flows from the right, the oil from the top and left. The generated droplets flow to the left to the outlet. The stepped channel sides for the water channel are due to the limitations of the resolution of the surface profiler. - **(c), (d)** Top views of the 2D surface profiles shown in (a) and (b). - **(e), (f)** Surface profiles corresponding to (a) and (b). - **(e)** shows the width of the channel as 9 µm with a height of 1.2 µm. - **(f)** shows the width of the orifice as 3.1 µm with a height of 1.4 µm.

163

Two-dimensional surface profiles of the final wafers were acquired with a 3D optical profiler by interferometry (Zeta-20, Zeta, Figure 6.4). An exemplary surface profile for a wafer with a pattern transfer from an emulsion film mask with the corresponding illumination device (Figure 6.2b) is shown in Figure 6.4a. The area is part of the inlet filter system from the design seen in Figure 6.1a. The three visible channels were designed with a width from left to right of 9 μm, 10 μm and 11 μm. All three channels stand out clearly. The visible steps along the channel edges are the result of the resolution limit of the surface profiler. The corresponding top view shows a clear separating edge between the deposited photoresist that forms the channels and the surrounding region in which the photoresist has been removed (Figure 6.4c). The profile measured across the right channel shows the measured width of the channel as 11 μm, consistent with the original feature on the mask. The height of the feature is approximately 1.1 μm (Figure 6.4e).

For smaller feature widths, the chrome covered glass mask and the corresponding illumination device were used (Figure 6.2c). The smallest designed width was the droplet-generating junction with a width of the orifice of 3 μm (Figure 6.1d). The 2D surface profile of the junction shows that the desired pattern could be transferred accurately onto the wafer (Figure 6.4b and Figure 6.4d). The measured width of the junction is 3.1 μm (a value of 3 μm was printed on the photomask) and an approximate height of 1.3 μm (Figure 6.4f).

The optimised parameters for the photolithography process and both illumination devices allow the designed pattern to be transferred accurately from the photo mask onto the wafer. This includes, for the chrome covered glass mask, feature sizes down to only 3 μm. The height of the photoresist layer was consistently less then expected for the spin speeds used. For the profiles shown, this is 1.1 μm to 1.3 μm for spin speeds of 1500 rpm compared to the value of 2.6 μm given in the photoresist processing guidelines. As the reduced dimension was beneficial

for this project, it was not further investigated.

6.4 Assembling the Microfluidic Device

Pouring the PDMS mixture on top of the wafer and curing it, results in a solid PDMS block with the design imprinted. The PDSM block is separated from the wafer and holes are punched through it at the inlets and outlets to connect the tubing through the PDMS to the microfluidic channels. The sealed microfluidic channels are formed by the depressed features in the PDMS together with the glass support after bonding together.

Plasma oxygen treatment of both the PDMS block and the glass support creates active silane groups on the surfaces necessary for a strong permanent bond between the PDMS and the glass. The independent parameters for plasma bonding are the chamber pressure, the applied power, the time of exposure. The temperature for the post-bonding oven treatment is also important. A strong bond will only be formed for a narrow range of parameters as a thin layer of undamaged oxide with active silane groups is needed. With too little plasma treatment, not enough silane groups are activated, and with too much plasma treatment, the surface is damaged.

Bhattacharya et al. (2005) systematically analysed the conditions of the plasma treatment and found that generally a high pressure up to 1 mbar, an applied power of 30 W and a time of exposure of 20 s to 30 s provides the strongest adhesive bond. Tang et al. (2006) reported a lower pressure of 0.04 mbar and an exposure of 10 s with 20 W to 60 W as optimal and found that the post-bonding temperature had no influence on the adhesive strength.

Both of these parameter sets did not provide a sufficient bonding between the PDMS and the glass in trials using the equipment described here. A set of experiments was performed to find parameters that result in strong adhesive bonding. The available method of plasma generation only showed a visible oxygen plasma below pressures of 0.3 mbar and above powers of 20 W. The parameters for the
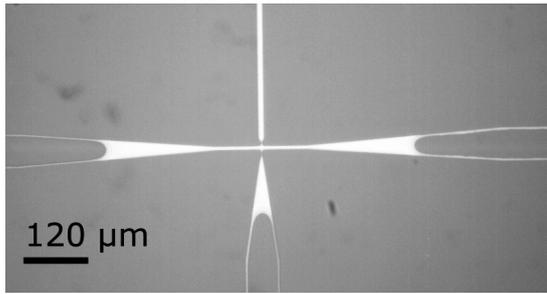
(a) Sagging of the PDMS ceiling
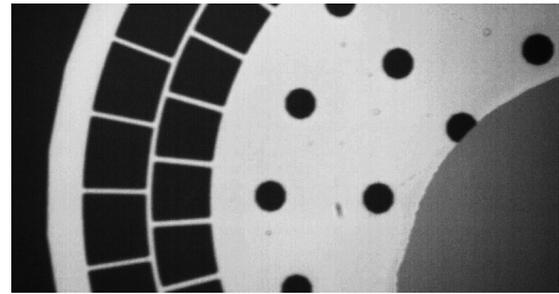
(b) No sagging of the PDMS ceiling

Figure 6.5 – PDMS is seen as darker against the brighter air-filled channels. **(a)** Sagging of the ceiling causes the channels to be blocked. - **(b)** Bonding with methanol as a separating layer protects even larger free areas of the microfluidic device from the sagging of the ceiling.

oxygen plasma activation were varied within an exposure time of 15 s to 60 s, an applied power of 40 W to 80 W and a chamber pressure of 0.1 mbar to 0.2 mbar. All results were compared after a post-bonding treatment of 5 min at 65 °C by qualitatively assessing the shear force necessary to separate the PDMS from the glass. It was found that longer oxygen plasma treatment with less applied power showed the strongest adhesive bonding. The best results were obtained with a 1 min treatment at 0.1 mbar and 28 W. Under these conditions, the PDMS would break before separating form the glass. It was found separately that, for the initial bonding, placing the glass support on top of the PDMS block resulted in stronger adhesive bonding than *vice versa*. These settings and procedures were used for all subsequent experiments.

A force is applied to bring the PDMS and the glass support into contact for bonding. The elastomeric character of PDMS leads to a deformation and might cause shallow microfluidic channels with high aspect ratios (i. e. width to height) to collapse. The ceiling of the channels will flex downwards and bond with the glass to block the channels. Xia and Whitesides (1998) report this for widths 20 times larger than the height of the feature.

In practice, the collapse of channels was no problem for feature depths of 15 µm

and a maximum aspect ratio of around 10. Every feature was clearly visible and no incident of a collapsed channel was found. This was not the case with features that were less than 2 μm deep (Figure 6.5a). The ceiling sagged inwards for aspect ratios above 15. This channel collapse occurred frequently, rendering the device unusable. Experiments were done with a lower ratio of PDMS base to curing agent, as this leads to lower elasticity of the final PDMS block (Khanafer et al., 2009), which could lessen the problem of the sagging ceiling of the channels. Ratios as high as 5:1 were tested but no significant difference could be observed.

It has been previously reported that immediately covering the surfaces of both the PDMS and the glass with a solvent such as methanol as a separating layer can prevent the instantaneous bonding of the two layers when brought into contact (Orabona, 2011). The solvent evaporates when heated, allowing the base surfaces to bond when no further force is applied to the device. This technique has never before been used in microfluidic devices for the generation of droplets. When applied in the production of the microfluidic devices, no incidents of collapsed channels were found for aspect ratios up to 30; this is the largest occurring aspect ratio in the microfluidic devices (Figure 6.5b).

6.5 Finding a Suitable Surfactant

A surfactant suitable for the encapsulation of RNA and proteins in µm-sized droplets has to fulfil several requirements: (i) it must dissolve in the oil phase to not interfere with the preparation of the biological samples, (ii) it must present an inert surface for the droplet contents both to prevent the denaturing of proteins and to prevent biological components from adsorbing on the inner droplet surface, (iii) it must stabilise droplets of the desired size and prevent coalescence and (iv) it must prevent the leakage of the droplet contents into the surrounding continuous phase.

Initial trials were done with the commercially available surfactant Krytox at concentrations of $2\,{}^{w}/v\%$ to $8\,{}^{w}/v\%$. No stable droplets could be generated (Figure 6.6a). It was decided to design and optimise a custom perfluorosurfactant, building on the previous work by Holtze et al. (2008). Based on the general structure (Figure 6.7a), different combinations of molecular weights for PEG to PFPE (Table 6.1) were used to synthesise triblock copolymers by Tom Fulham and René N. Tshulu under the supervision of Dr. Alison M. Stuart (Department of Chemistry, University of Leicester).

All designed surfactants were tested for their ability to prevent the coalescence of generated droplets at the two concentrations of $3\,{}^{w}/v\%$ and $6\,{}^{w}/v\%$. The surfactant variants 300/2500 and 400/2500, corresponding to the molecular weight for PEG and PFPE, respectively, were both viable over a large range of droplet sizes. They allowed the generation of the smallest possible droplets and prevented coalescence. These surfactants were used for all subsequent experiments.

Droplet Side of the channel of PDMS Channel filled with oil
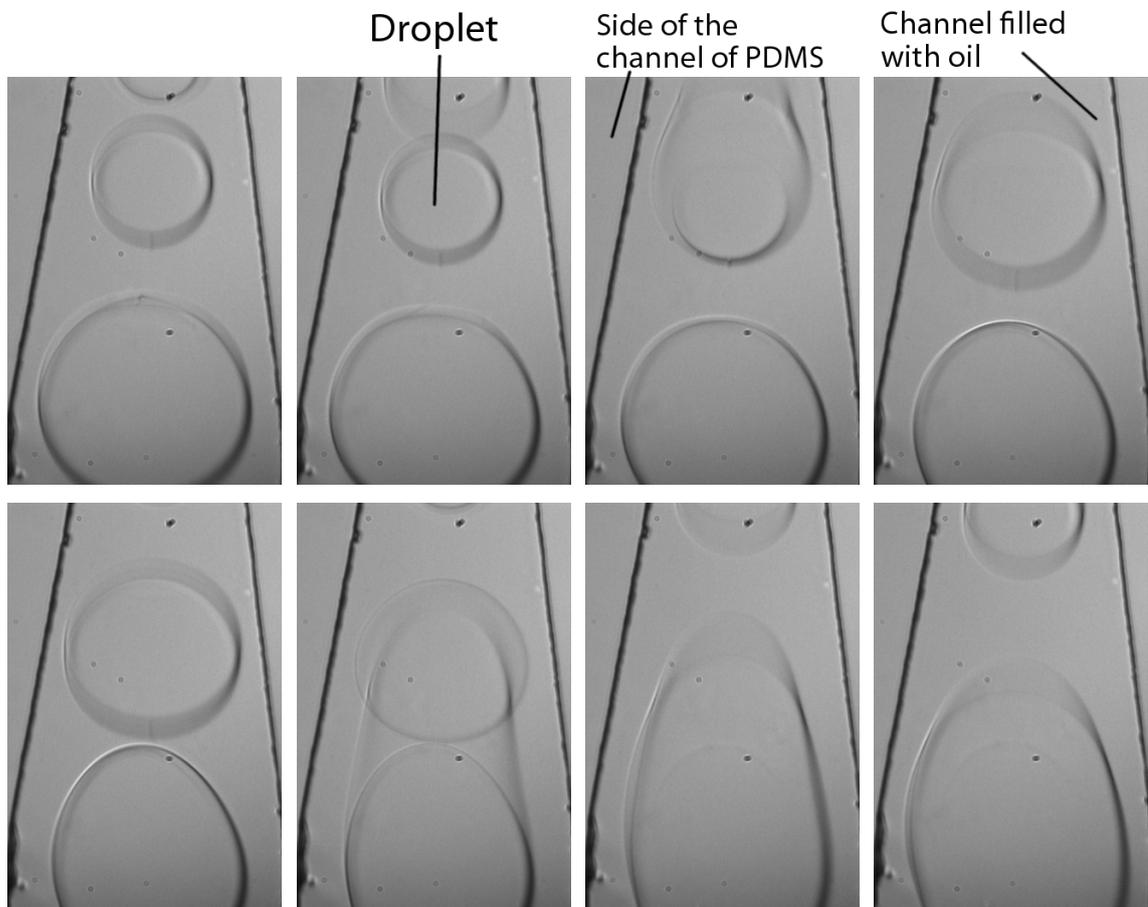
(a) Fusion of droplets
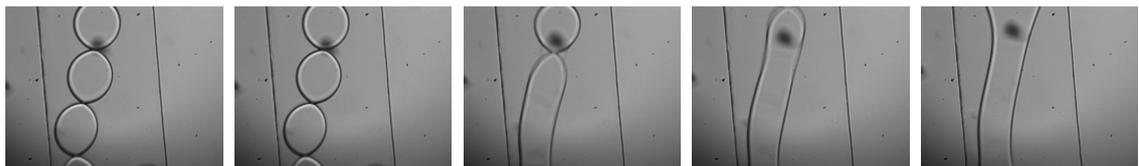
(b) Fusion of droplets

Figure 6.6 – **(a)** Immediate coalescence of droplets with Krytox as the surfactant. As soon as the generated droplets come into contact they fuse. - **(b)** Coalescence of droplets with the the 400/7000 variant of the surfactant. The droplets do not fuse immediately, but coalesce within the outlet channel.

(a) Surfactant

150 μm
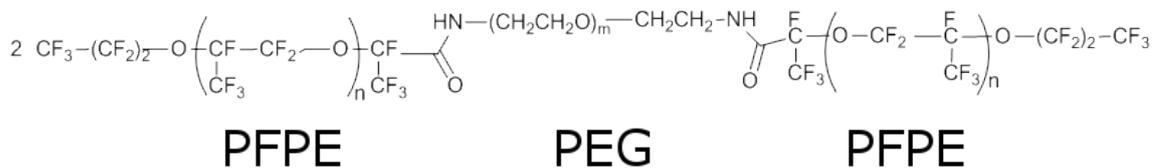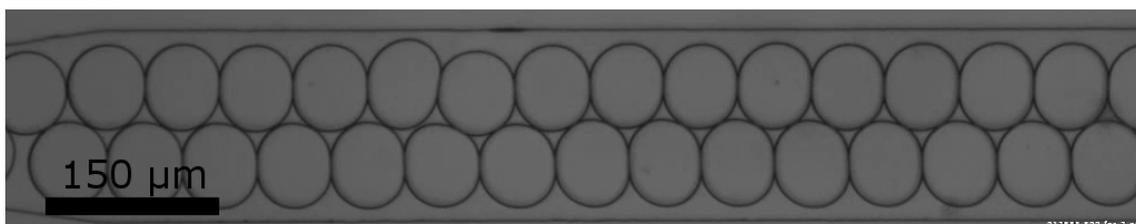
(b) Stable droplets

Figure 6.7 – **(a)** Underlying chemical structure for the different designs of the perfluoro-surfactants. - **(b)** Droplets stabilised by the 400/2500 variant of the surfactant do not coalesce even when squeezed against each other.

Table 6.1 – Combinations of designed and synthesised perfluorosurfactants and their calculated Hydrophilic-lipophilic balance (HLB) values.

PEG (g/mol)	PFPE (g/mol)	HLB
300	2500	1.13
400	2500	1.48
600	2500	2.14
400	3750	1.01
600	3750	1.48
400	7000	0.56
600	7000	0.82

6.6 Fast and Reliable Control over the Size of Droplets

The pressure-driven flow controller provided positive pressures for both the dispersed and the continuous phase in the range of 1 mbar to 1024 mbar. Pressures of 900 mbar to 1024 mbar could occasionally break the cover slides at the bottom of the microfluidic device and were avoided. This range of pressures can be used to manipulate the flow rates of both the dispersed and the continuous phase over a wide range with a response time of less than a second. It was important to avoid the flow rates to become negative for large pressure differences, as the channels containing separate phases are connected. The maximum flow rates depended on the setup but could reach up to 500 µl/h. It was possible to halt the flow of one phase, but not to stop the flow of both phases at the same time without back flow into the channel of the other phase.

It was possible to change the flow in the outlet channel from only containing the continuous phase to almost exclusively containing the dispersed phase, steadily increasing the size and density of the generated droplets (Figure 6.8).
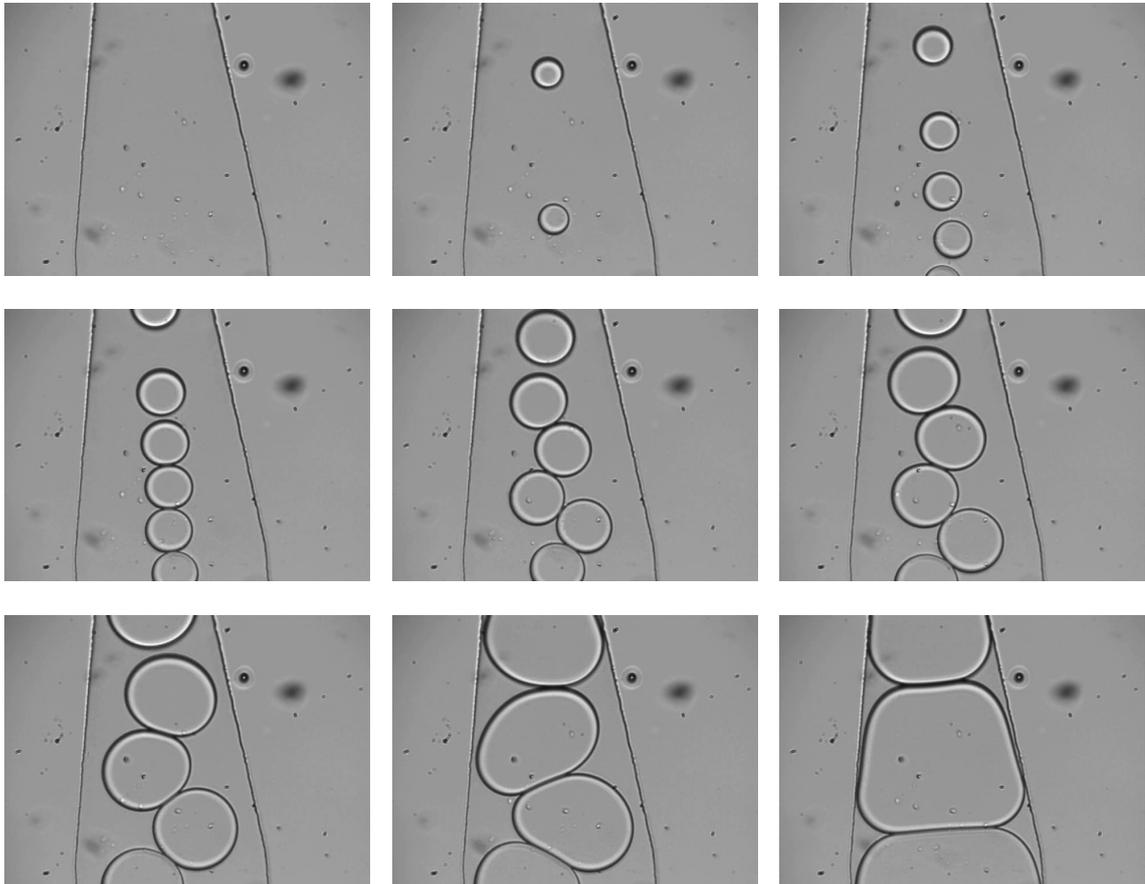
Figure 6.8 – The pressure applied to the dispersed phase was continuously increased. Each image is roughly 2 s apart. At first, no droplets are generated as the pressure applied to the dispersed phase is not enough for the dispersed phase to break through the continuous phase. Once the pressure is high enough, the first separated droplets are formed. The size of the droplets increases as the ratio of the flow of the dispersed phase to the continuous phase increases until there is an almost continuous stream of of the dispersed phase.

6.7 The Generation of Droplets

The initial surface of the PDMS is hydrophobic. The plasma oxidation treatment necessary for bonding to the glass cover slides forming the bottom renders the surface hydrophilic (Hillborg et al., 2000). Under these conditions, flow focussing does not occur with water as the dispersed phase and oil as the continuous phase and oil droplets within a continuous phase of water are generated instead (Figure 6.9a). By increasing the flow of the oil phase this would eventually result in oil droplets embedded in larger water droplets (Figure 6.9b).

As this was not within the scope of this work, the microfluidic channels were flushed with 3-Aminopropyltriethoxysilane to create hydrophobic channel walls. Water as the dispersed phase could still be pushed through the hydrophobic channels. Flow focussing worked as intended now.

Four different droplet breakup regimes for the generation of droplets with flow focussing exist (Section 1.3.3). These regimes cover a wide range of droplet sizes. In the squeezing regime, a droplet with a size larger than the orifice completely blocks the flow of the continuous phase. The pressure builds up for the continuous phase until it pinches off the droplet (Figure 6.10a). Lowering the pressure applied to the continuous phase extends the time the droplet has to form, resulting in larger droplets. Droplets can be formed of increasing size until the pressure never reaches the point where it is able to break the dispersed phase.

Increasing the ratio of the flow of the continuous phase to the dispersed phase, i. e. increasing the pressure for the continuous phase, leads to a faster separation of the droplets. In this case, as soon as the flow of the continuous phase is blocked, the aqueous droplet is pinched off (Figure 6.10b). An even higher flow ratio leads to the dripping regime. In this case, generated droplets are smaller than the orifice and never block it completely. Instead, the much faster flow of the continuous phase deforms the protruding dispersed phase and pinches the droplet off the
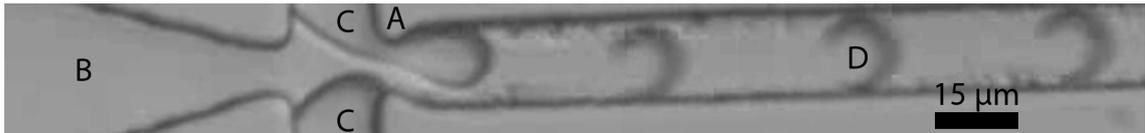
(a) Generation of oil in water droplets
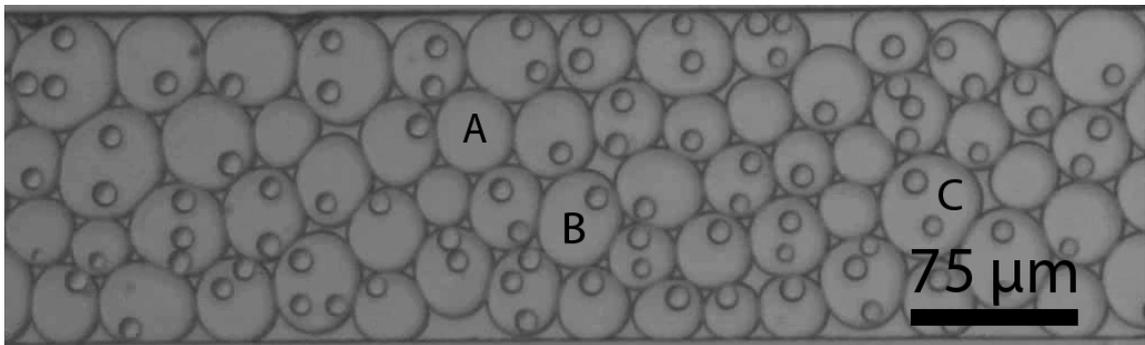
(b) Oil in water in oil droplets

Figure 6.9 – **(a)** Microfluidic channels untreated with silane are too hydrophilic and the junction (A) does not allow flow focussing. An alternative mode of droplet generation is used where the intended continuous phase is encapsulated instead. Water coming from the left (B) should be periodically cut off by the two oil flows (C), but instead the oil flows are cut off to form oil in water droplets (D). - **(b)** Oil in water droplets as seen in (a) are not a favoured condition and eventually water in oil droplets (A) are formed that then itself can contain one (B) or more (C) oil droplets.
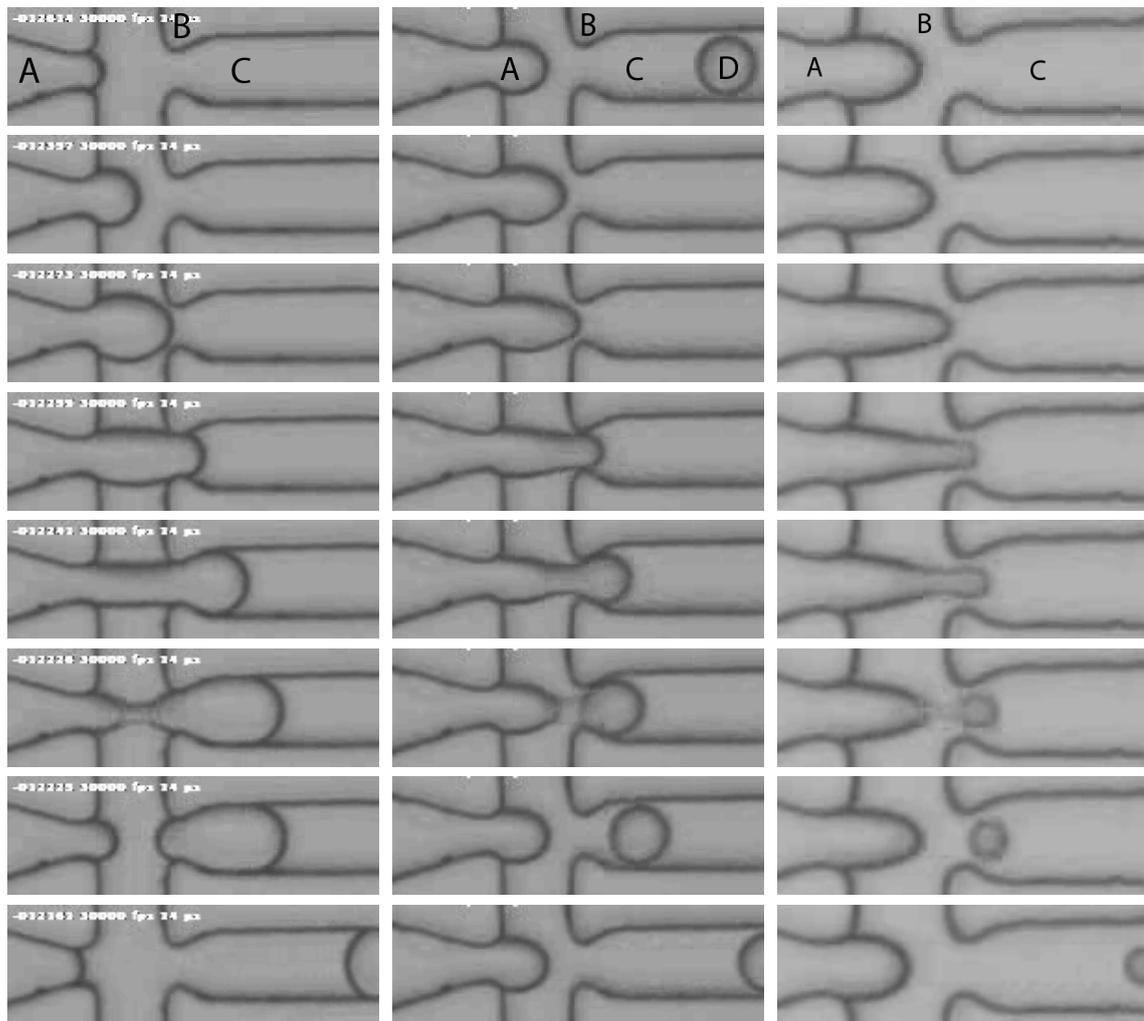
(a) Large droplet   **(b)** Medium droplet   **(c)** Small droplet

Figure 6.10 – Droplet generation with the 5 µm junction. The width of the channel to the right is 9 µm. The dispersed water phase (A) comes from the left, the continuous oil phase (C) from the top and the bottom, confined by the walls of PDMS (B). The generated droplets (D) flow through the outlet to the right. - **(a)** A droplet with an approximate diameter of 20 µm is generated in the squeezing regime. The water flow cuts off the oil flow completely, which then break the water flow to form the droplet. - **(b)** Conditions between the squeezing and the dripping regime generate a droplet of 14 µm diameter. - **(c)** A droplet with a diameter of 5 µm is generated in the dripping regime. The water flow does not cut off the oil flow, instead the much faster oil flows rip a water droplet off the tip of the water flow.

end (Figure 6.10b).

Thread formation could not be observed and the jetting regime could not be seen with the 5 μm junction, see Section 6.10.

6.8 Dependence of the Droplet Size on Flow Rates and the Composition of the Dispersed Phase

The generation of droplets depends, besides on the geometry of the microfluidic flow focussing junction, on the flow rates of the continuous and the dispersed phase as well as the viscosity of the continuous phase (Section 1.3.3). This relationship was investigated by generating droplets using either pure water or adding 0.5 % NP-40 to the aqueous phase in order to lower the fluid viscosity of the dispersed phase over the accessible range of pressures provided by the microfluidic flow controller.

First, the diameter of droplets generated with different flow rates was compared (Table 6.2). It was found that, as a general trend and as predicted, the higher the ratio of continuous phase flow rates to dispersed phase flow rates, the smaller the diameter of the generated droplets. Increasing the flow rates of both phases has the same result. For similar flow rate ratios, higher individual flow rates also generate smaller droplet sizes. In conclusion, for smallest possible droplet sizes, the highest achievable flow rates and the highest flow rate ratio, must be used. This cannot be exploited indefinitely as droplet generation breaks down at the lower end of droplet sizes. Additionally, only certain flow rates within a certain ratio can generate droplets.

For similar flow rates, the addition of 0.5 % NP-40 to the dispersed phase generated smaller droplets than pure water in line with the predictions. The use of 0.5 % NP-40 also allowed a wider range of possible flow rates to be covered for which droplets could be successfully generated.

The individual flow rates are manipulated by changing the applied pressure for each phase, but not independently for each phase, as both phases interact. The available flow rates are therefore limited by the available pressures. On the lower

Table 6.2 – Droplet diameters for combinations of dispersed and continuous phase flow rates. Flow rates are in nl/min and were binned in 100 nl/min intervals, droplet diameters are in µm. Results are shown in red for water and in blue for 0.5 % NP-40 as the dispersed phase. The 5 µm junction was used. For the standard deviation of the droplet diameters, see Table 8.6.

disp. \ cont.	100	300	400	500	600	700	800	900	1000	1100	1200	1300	1400
100	29.0	23.3 / 25.2	25.0 / 15.8	21.7 / 19.3	19.7	19.5	15.2 / 13.9	16.2 / 14.6	14.6 / 12.6		no droplets		
200		19.6	27.7	24.0 / 15.0		17.9	19.6 / 19.6		15.9	15.5			
300		34.4		23.5 / 23.2		20.4	19.4		18.5	16.8		16.0	15.8
400		31.4		23.4		24.0							
500		29.9		21.2 / 20.6				21.2	19.0				
600		31.4 / 22.4		32.3			25.4			17.9		17.3	
800									21.9		18.3		
900								25.7			18.6		
1000	only dispersed phase								22.7				
1200										22.8		18.7	
1600											24.1	burst	

179

Table 6.3 – Flow rates of the dispersed phase for combinations of dispersed and continuous phase pressures. Pressures are in mbar, flow rates are in nl/min. Results are shown in red for water and in blue for 0.5 % NP-40 as the dispersed phase. The 5 µm junction was used. The flow rates are accumulative over the time of the measurement.

cont. / disp.	200	300	400	500	600	700	800	900	1000
150	1.2 / 16.5								
175	6.1 / 12.5								
200	8.6 / 26.9								
225		3.1 / 8.7							
250		6.4 / 16.1							
275		7.5 / 19.5							
300		11.1 / 23.8							
325			5.6 / 13.1						
350			9.1 / 17.4						
375			11.4 / 23.6	3.7 / 13.4					
400			14.0 / 27.8	7.2 / 14.9					
425				9.7 / 19.7					
450				11.1 / 25.2	1.2 / 9.0				
475				13.7 / 28.7	6.5 / 15.3				
500				16.9 / 33.9	8.7 / 20.3				
525					11.3 / 24.6	2.5 / 11.7			
550					13.0 / 31.0	4.4 / 18.3			
575					17.8 / 33.0				
600					19.4 / 36.8	11.3 / 27.6			
625							5.3		
650						16.7 / 35.3	8.5 / 20.7		
700						22.0 / 42.8	13.8 / 31.6	5.2	
725								8.7	
750							19.5 / 38.5	11.3 / 26.6	
775									2.6
800							25.2 / 44.7	17.6 / 34.3	6.5 / 18.9
850								21.9	13.1
900								27.5	19.8 / 38.6
925								31.1	
950									25.1 / 45.5
1000									30.4

Table 6.4 – Droplet diameters for combinations of dispersed and continuous phase pressures. Pressures are in mbar, droplet diameters are in µm. Results are shown in red for water and in blue for 0.5 % NP-40 as the dispersed phase. The 5 µm junction was used. For the standard deviation of the droplet diameters, see Table 8.7.

disp. \ cont.	200	300	400	500	600	700	800	900	1000
150	16.7 / 15.0								
175	21.1 / 19.3								
200	28.6 / 22.4								
225		14.3 / 14.6							
250		18.0 / 17.9							
275		22.5 / 23.2							
300		30.5	12.6 / 12.9						
325			17.2 / 16.6						
350			21.4 / 20.3						
375			25.1 / 24.0	9.6 / 12.6					
400			32.3 / 32.3	16.2 / 15.9					
425				19.5 / 18.5					
450				23.3 / 21.2	8.7				
475				26.6 / 25.4	14.9 / 15.5				
500				33.8	18.2 / 16.8				
525					21.1 / 19.0	6.8			
550					23.7 / 21.9	13.4			
575					27.7 / 25.7				
600					34.4	19.4 / 17.9	6.0		
625							13.3		
650						24.0 / 22.7	15.2 / 16.0		
700						31.4	19.6 / 18.3	8.4	
725								14.2	
750							23.5 / 22.8	16.6 / 17.3	
775									6.4
800							29.9	19.6 / 18.6	9.4 / 15.8
850								23.4	16.9
900								31.4	19.4 / 18.7
925								45.0	
950									21.2 / 24.1
1000									28.2

end, flow rates could be brought to a stop by reducing the pressure sufficiently to a level where it counteracts any internal pressure. For the higher end, pressures were limited to a maximum of 1032 mbar as the limit of the microfluidic flow controller. This coincided with the pressure limit for maintaining the adhesive bond to the cover glass that formed the bottom of the microfluidic device as well as the structural strength of the glass itself.

Flow rates for the dispersed phase that resulted in a successful droplet generation were measured over the range of available pressures (Table 6.3). The lower viscosity on addition of 0.5 % NP-40 to the dispersed phase resulted on average in 2.8 times higher flow rates compared to water at the same pressure. However, with pure water, it was possible to generate droplets over a wider range of pressures.

The final relationship of pressure to droplet size is shown in Table 6.4. While the same pressures result in different flow rates depending on the composition of the dispersed phase, this does not change the size of the droplets significantly. The ability of a dispersed phase of pure water to generate droplets at ratios beyond the boundaries possible for the 0.5 % NP-40 enables smaller sizes. The advantage of faster flow rates is the increased rate of droplet generation and a higher density of droplets.

Section 8.2 contains the underlying data for the droplet diameters used for the binned tables.

6.9 Limitations of the $5\,\mu m$ Junction

The direct control over the pressure for delivery of each phase allowed the size of the generated droplets in a continuous stream through the $5\,\mu m$ junction to be decreased to a diameter of around $7\,\mu m$ (Figure 6.11a). Smaller droplet sizes down to $5\,\mu m$ were achievable intermittently (Figure 6.10c). The problem was that the range of pressures that sustained conditions for generating droplets of less than $5\,\mu m$ was too narrow and the droplet generation was sporadic and unreliable where the stream of generated droplets would regularly break down, resulting in a heterogeneous distribution of droplet sizes (Figures 6.11b and 6.11c).

Smaller droplets sizes could theoretically be obtained with higher flow rates, but those were not possible due to the maximum viable pressure of $900\,mbar$. Decreasing the viscosity of the dispersed phase for higher flow rates did not ultimately help (Section 6.8).

This left only the geometry parameters of the flow focussing junction as an option to reach smaller droplet sizes. With a limit for the minimal size of microfluidic structures that could be fabricated of $3\,\mu m$, a junction with a reduced orifice of $3\,\mu m$ compared to the conventional one with $5\,\mu m$ was designed and fabricated. The dimensions of the surrounding elements were reduced accordingly (compare Figure 6.1d with Figure 6.1c).
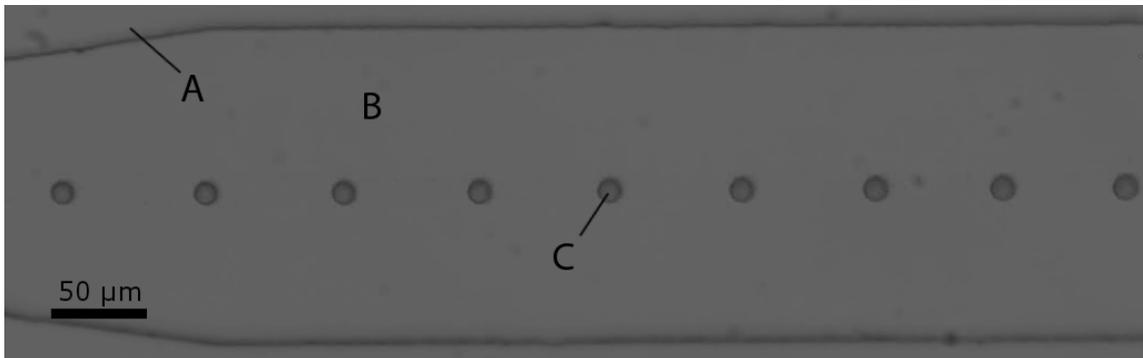
(a) Stable droplet generation
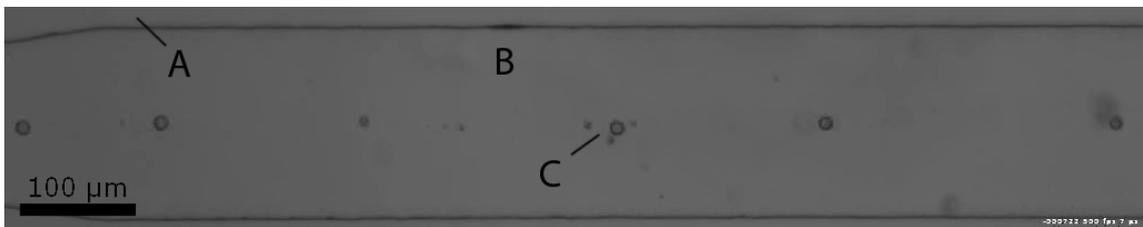
(b) Breakdown of droplet generation
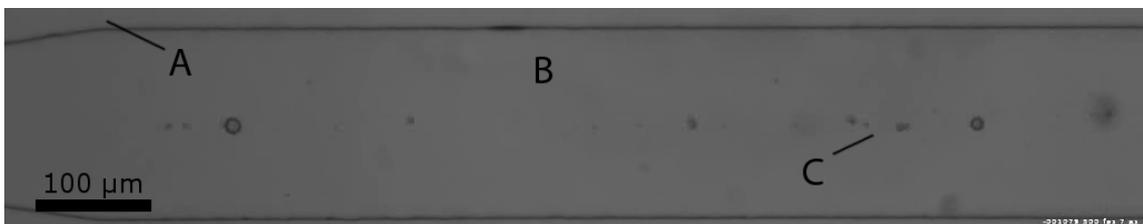
(c) Breakdown of droplet generation

Figure 6.11 – Droplets were formed using the 5 µm junction. The outlet channel formed by the PDMS mould (A) is filled with the continuous phase (B), here FC-40. **(a)** The smallest possible droplets that can be reliably generated with a homogeneous size have a diameter of 5 µm (C). - **(b), (c)** Changing the flow parameters to achieve even smaller droplets results in an unstable flow regime. The generation of droplets breaks down intermittently (C) which results in the differently sized droplets (b), until it almost completely breaks down (c).

6.10 Generating Droplets with a Diameter of 1 µm

The transition from the 5 µm junction to the 3 µm junction changed drastically the behaviour of the flows. Water as the continuous phase could not be pushed through the microfluidic channels after the silane treatment rendered the surface hydrophobic. The silane treatment step was eliminated from the protocol and the surfaces were allowed to remain hydrophilic after the plasma oxidation. Although the water phase could now be pushed through the microfluidic channels, it was not possible to get it to flow through the 3 µm constriction in conjunction with oil as the competing continuous phase. Only with the addition of NP-40 to the dispersed phase in order to reduce viscosity, could flow focussing be achieved. For all further experiments, a dispersed phase of 0.5 % of NP-40 was used. While the results from Section 6.8 showed that this would result in similar droplet sizes under the same conditions used with pure water as the dispersed phase, the high flow ratios cannot be reached. Fortunately, a high ratio of flow rates was not needed to reach droplet diameters in the 1 µm range.

The acquisition of images showing the formation of droplets with a diameter of 1 µm was difficult. The use of NP-40 in the dispersed phase leads to faster flow rates compared to water (Table 6.3). Together with the small dimensions, the generation of a droplet takes place in time intervals below 1 ms. The lowest possible flow rates for the dispersed phase were used and exposure times of 5 µs were necessary to illustrate the process of droplet generation (Figure 6.12).

The formation of a droplet starts with the dispersed phase protruding into the junction, then a droplet is ripped off and the dispersed phase retracts. Droplet generation events are separated by a longer phase wherein the pressure for the dispersed phase builds up and no movement of the dispersed phase can be observed. The droplets with a diameter of 1 µm were generated within the dripping regime. Tip-streaming, as used by Shim et al. (2013), was not necessary for these

185
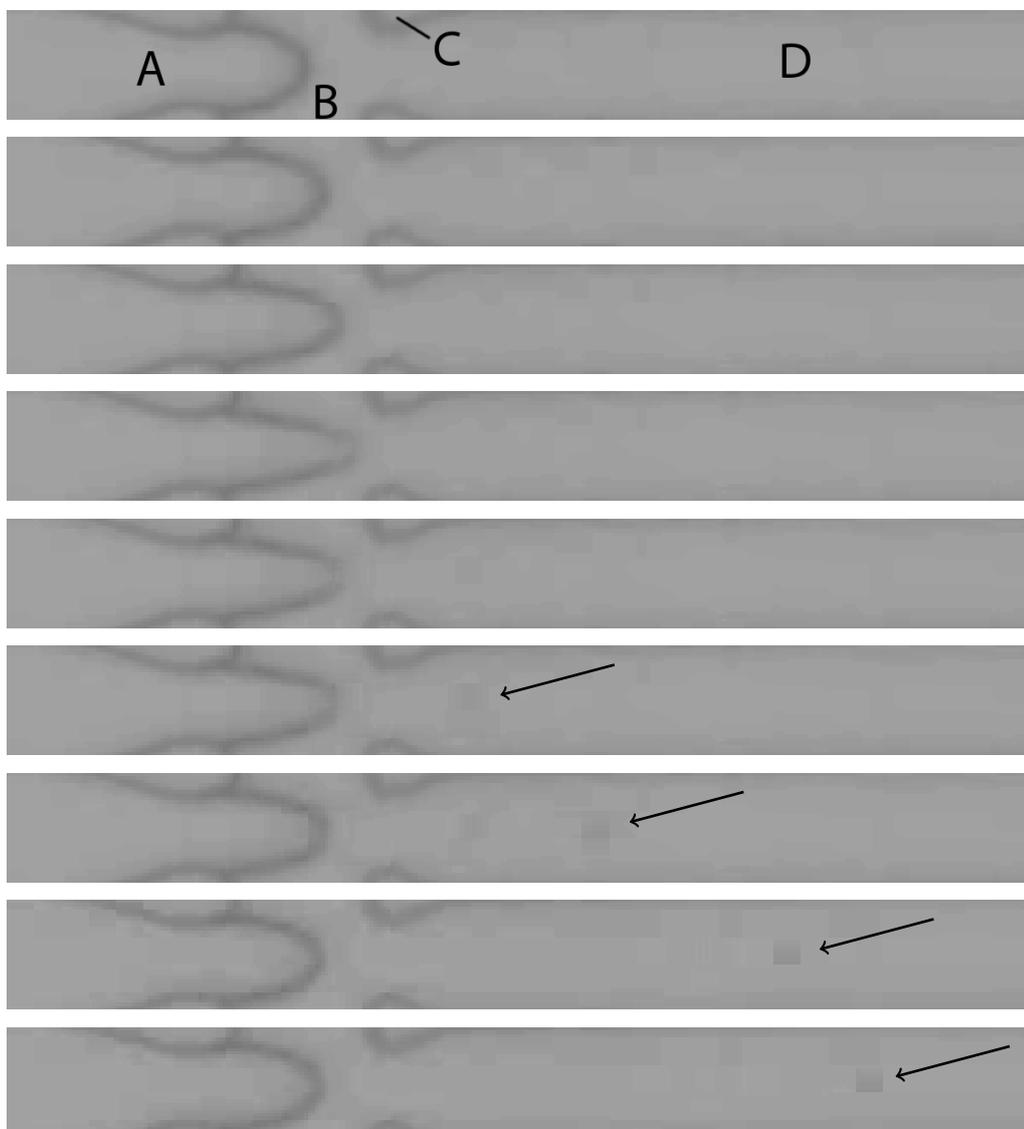
Figure 6.12 – Generation of one droplet with a diameter of below 1 µm. The dispersed phase protrudes from the left (A) into the junction, formed by the PDMS (C). The continuous phase (B) comes from the top and bottom. Both phases flow towards the outlet (D). The orifice has a width of 3 µm. The generation of the droplet occurs within the dripping regime. The generated droplet can be seen in the last four images as it moves in the outlet channel from left to right.

sizes.

Faster flow rates for the dispersed phase led to a higher density of generated droplets, but the accompanying faster droplet generation could not be visualised for individual droplets. Low magnification imaging of the stream of generated droplets allowed the use of shorter exposure times to obtain sharper images of the droplets. Fast flow rates shifted the regime of the flow focussing from the dripping regime into the jetting regime. No withdrawing of the tip of the dispersed phase could be observed and droplets were rapidly produced (Figures 6.13a and 6.13b). The droplets slow down on entering the wider outlet channel and move closer to each other until they are loosely grouped towards the end of the outlet channel (Figures 6.13c and 6.13d). The flow rate ratios that corresponded to droplets with a diameter of 1 μm could be reached within a few seconds. Slight variations in the applied pressures did not perturb the droplet generation and it stayed stable over up to 4 hours.
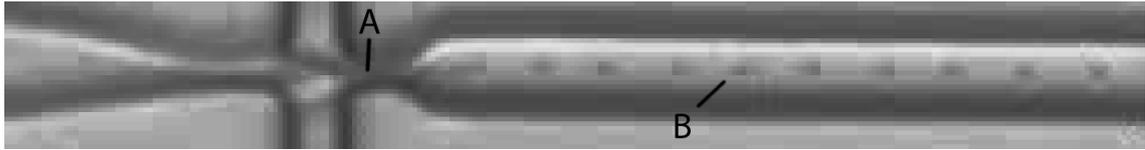
(a) Generation of 1 μm droplets
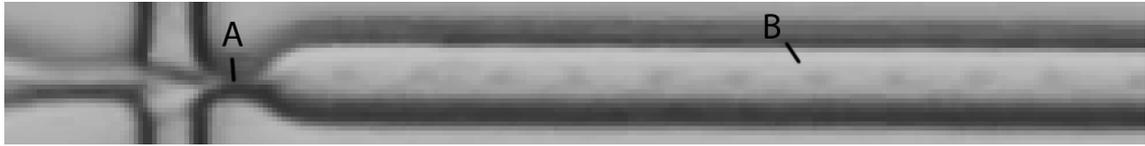
(b) Generation of 1 μm droplets
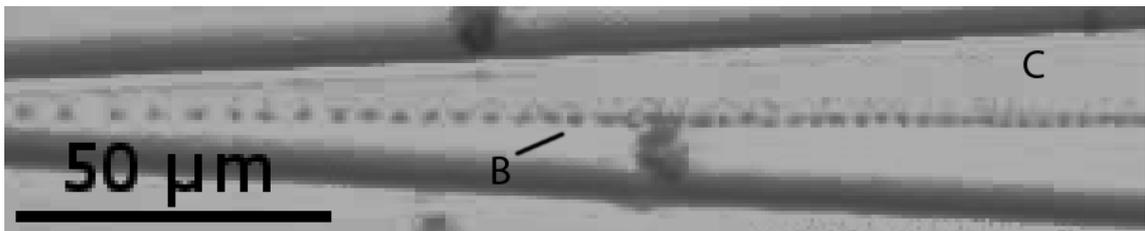
(c) Line of droplets
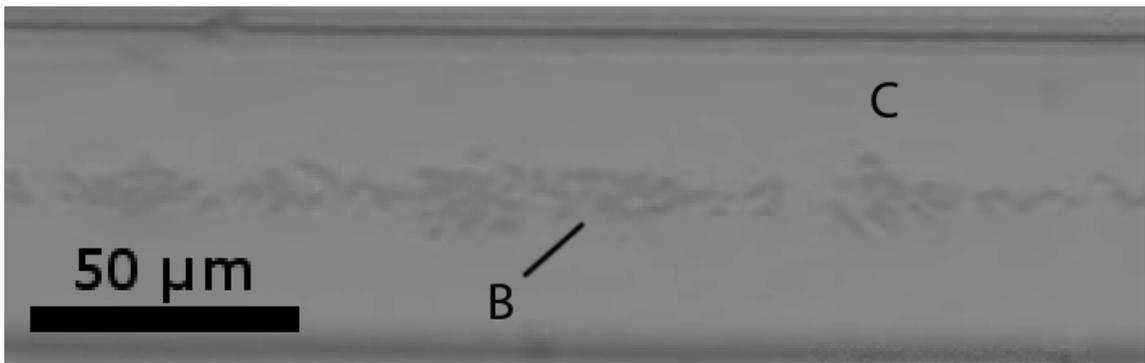
(d) Dispersing of droplets

Figure 6.13 – All droplets were generated using the 3 μm junction and 0.5 % NP-40 as the dispersed phase. The flow rate for the dispersed phase is around 0.2 nl/min and 5 nl/min for the continuous phase. 1000 droplets are produced per second and flow with a velocity of around 9500 μm/s. **(a), (b)** Generation of 1 μm droplets in the dripping regime. The homogeneously sizes droplets (B) are generated in a stable way at the junction (A). - **(c)** As the outlet channel widens, the droplets (B) slow down within the flow of the continuous phase (C), here FC-40, and get closer to each other. The line shows the same size for each generated droplet. - **(d)** Further along the channel, the flows start to mix and the droplets flow by in different densities (B) within the flow of the continuous phase (C), here FC-40. No fusion of droplets can be observed. Images **(a)** - **(c)** had their intensities non-linearly enhanced to highlight the generated droplets.

6.11 Re-injection of Generated Droplets

Initially, the outlet was connected via tubing to an Eppendorf sample vial and the generated droplets were collected over 3 h to 4 h. They were stored at 4 °C. As the density of the continuous phase (FC-40, $1855 \, \text{kg/m}^3$) is almost twice as much as the density for the dispersed phase (H_2O, $998 \, \text{kg/m}^3$), the generated droplets tend to float and migrate to the top of the collected emulsion where they aggregate (Figure 6.14a). These droplets were stable over one week if kept at 4 °C; they did not shrink or change from their homogeneous size distribution.

When the collected droplets were put on cover slides to image them with an inverted microscope, they would quickly rise to the top of the emulsion. There they were constantly moving, making it impossible to image them over time. Sample chambers out of two cover slides separated by double sided sticky tape were fabricated. This allowed limiting the movement of the droplets enough to visualise them. The aggregation of the droplets below the top cover slide made it impossible to probe any encapsulated fluorescent probes. The laser light could not be sufficiently focused on a single droplet due to the occurring refraction from other droplets close by.

A second microfluidic device consisting of a big chamber with hourglass-shaped pillars and an inlet and outlet at opposite ends was fabricated. The generated droplets were then re-injected through the inlet into the chamber. The ceiling of this chamber was 2 µm high, preventing any additional aggregation of droplets. The droplets could still form two layers if the aggregated droplets from the top of the collecting vial were directly re-injected, as the concentration of droplets was high (Figure 6.14b). Diluting the collected droplets eliminated the occurrence of multiple layers, resulting in sufficiently isolated droplets that could be probed separately (Figure 6.14c).

While the generated droplets were stable for at least one week if kept at 4 °C, the
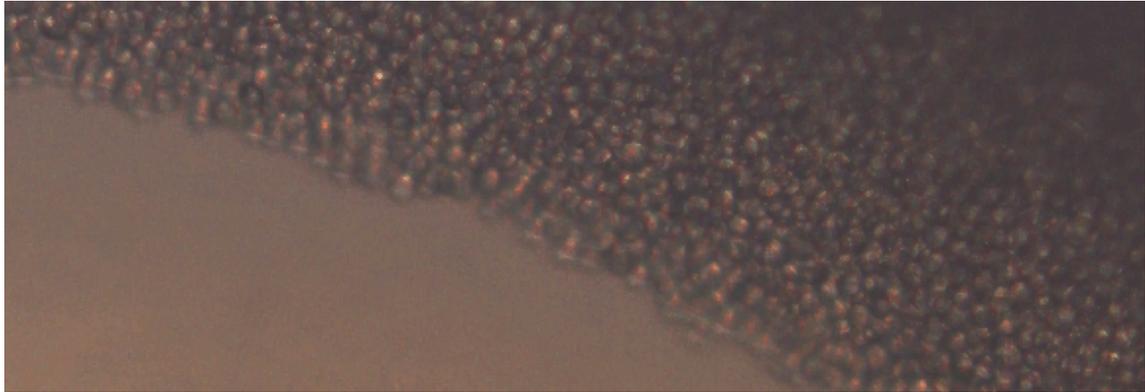
(a) Aggregated, floating droplets

(b) Different densities of re-injected droplets
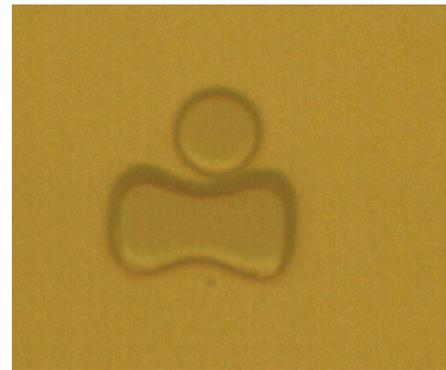
(c) Single, isolated droplet

Figure 6.14 – **(a)** The generated and collected droplets with a diameter of 15 μm aggregate by floating to the top. - **(b), (c)** The collected droplets of 15 μm were re-injected into a second microfluidic device. **(b)** Different densities of droplets can be found. Droplet densities can range from separated droplets as seen on the bottom left, over a single layer of droplets in the middle, up to multiple droplets stacked on top of each other on the top right. - **(c)** A single, isolated droplet sitting next to an hourglass shaped pillar.

process of re-injection seemed to degrade the stability of droplets with a diameter of less than 5 μm severely. The droplets began to shrink immediately and survived for less than 10 min. Exposure to a laser enhanced the rate of shrinkage and droplets also seemed to leak their encapsulated contents. Fluorescent intensity time traces could only occasionally be obtained and never in a consistent manner.

6.12 Collecting Droplets on the Same Microfluidic Device They are Generated on

As droplets only degraded after re-injecting them, it was decided to not collect droplets externally. Instead a similar compartment as used in Section 6.11 to image droplets was added after the junction. Directly after being generated at the junction, droplets would be collected in the imaging compartment which was connected at the far end to an outlet to maintain the flow of the oil/water mixture. A repeating pattern of four squares with sides of length 20 µm acted as pillars to prevent the ceiling from collapsing and to trap the generated droplets.

Flow conditions had to be maintained for a maximum of 1 min to yield a sufficient number of trapped droplets in the imaging chamber. The desired target diameter of the droplets was achieved within seconds by using the pressure-driven flow control. The flows were abruptly stopped after a sufficient number of droplets were trapped, resulting in a slight back flow of either the dispersed or the continuous phase, preventing an overflow and contamination of the collection chamber. The number of collected droplets was usually in the range of 200 to 1000. The generated droplets could be immediately imaged, providing a much faster detection method. The collected droplets showed no sign of degradation over the course of 4 h.
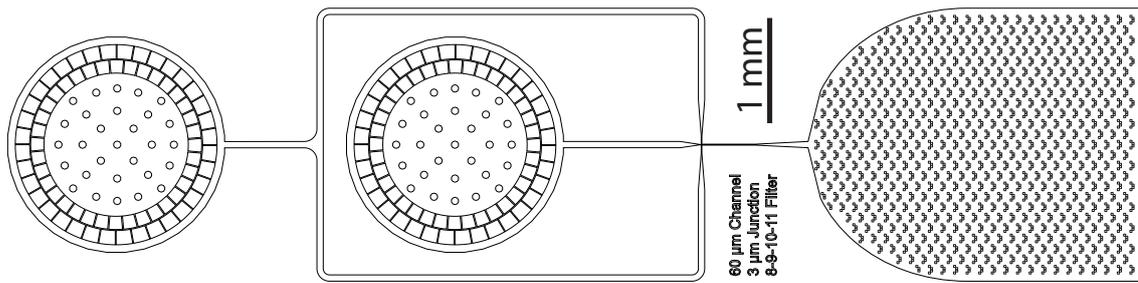
(a) 3 µm Mixer with Reservoir
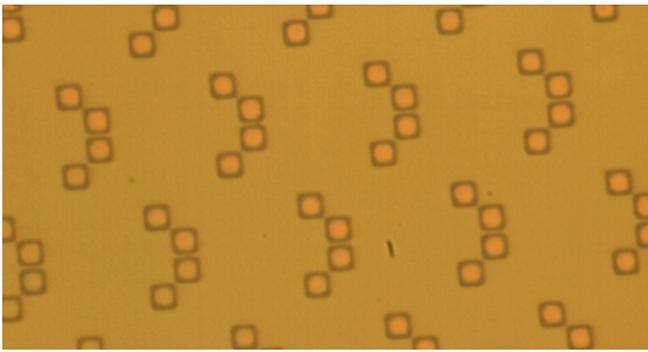
(b) Collector
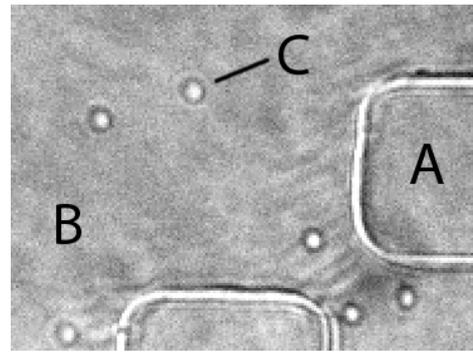
(c) 1 µm droplets next to collector

Figure 6.15 – **(a)** Design for a mixer with a 3 µm junction and an attached reservoir at the right, instead of the outlet. - **(b)** The length of the sides of the squares is 20 µm. The four squares act simultaneously as pillars to hold the ceiling and as filters to slow down, disperse and catch the generated droplets. The flows are from the left to the right. - **(c)** Droplets (C) with a diameter of 1 µm are located next to two squares of the collector (A), surrounded by oil (B).

6.13 Encapsulated Fluorescent Quantum Dots

Fluorescent quantum dots were used as a first test subject to see if fluorescent probes could be permanently encapsulated inside droplets and, if possible, their number be deduced from the acquired fluorescent intensity time traces.

The collected droplets within the attached imaging compartment were first observed using bright field illumination. Single droplets were moved into the focus of the laser beam by translation of the microscopic stage. Acquisition was started with the bright field illumination initially to later identify the position of the observed droplet. Once the bright field illumination was turned off, the fluorescence laser was switched on and the emission from the droplet was imaged until no discernible fluorescence could be observed.

A concentration of $8\,\text{nM}$ of fluorescent quantum dots was added to the dispersed aqueous phase containing $0.5\,\%$ NP-40 and $0.1\,\text{M}$ $MgCl_2$ and would in theory result on average in 161 quantum dots within a droplet with a diameter of $4\,\mu\text{m}$. No fluorescence could be detected from the continuous phase, showing that the quantum dots were confined to the dispersed phase and did not leak out of the droplets. The generated droplets were also stable for up to $4\,\text{h}$; the quantum dots did not negatively interfere with the surfactant or the stability of the droplets.

An illuminated droplet shows intense fluorescence, clearly confined to the exact same space as the droplet. Under continuous illumination, individual quantum dots bleach over time until single quantum dots could be discerned (Figure 6.16a). The corresponding intensity time trace shows an initial phase with an increase in intensity before it reaches its peak. This is most likely due to the high density of quantum dots that either quench each other or prevent the laser from efficiently exciting all available quantum dots. This first phase is followed by a linear decrease of intensity before the fluorescent intensity decays exponentially over time as expected (Figure 6.16b).

(a) 5 μm droplet

(b) Intensity time trace for the droplet seen to the left

(c) 3 μm droplets

(d) Intensity time trace for the droplet seen to the left

(e) A droplet with a diameter of 3 μm encapsulate a single fluorescent quantum dot

Figure 6.16 – Droplets are confined vertically within less than 2 μm and droplets with a diameter larger than 2 μm are therefore squished and their visible diameter appears larger than it the would be for an undisturbed sphere. **(a)** A droplet with a diameter of 5 μm filled with fluorescent quantum dots is first shown in bright field illumination and then its fluorescence. At the beginning, the whole droplet lights up until individual quantum dots can be discerned. - **(b)** Intensity time trace corresponding to the droplet in (a). After an initial phase, the fluorescence intensity decays exponentially (fit in red). - **(c)** A droplet with a diameter of 3 μm containing several fluorescent quantum dots is first shown in bright field illumination and then its fluorescence. Individual quantum dots can be distinguished. As they stop emitting fluorescent, there number decreases visibly. - **(d)** Intensity time trace (blue) corresponding to the droplet in (c). The fit in red highlights seven bleaching steps. - **(e)** A droplet with a diameter of 3 μm encapsulates a single fluorescent quantum dot. The time between each image is 3 s.

The concentration of quantum dots was then lowered to 800 pM, which will result in 2 to 7 nanoparticles in aqueous droplets with a diameter of 2 μm to 3 μm. Single-digit numbers of quantum dots were encapsulated and could be visually identified (Figure 6.16c). The intensity time traces showed a combination of blinking and bleaching steps and allowed the number of quantum dots contained in a droplet to be obtained (Figure 6.16d).

For even lower concentrations of quantum dots of around 80 pM, single quantum dots could be encapsulated and imaged. The confined movements of these single quantum dots could be observed over time and shows no visible additional interaction with the droplets interface that restricts its movement. The observed bleaching of the quantum dots could be due to interactions with the interface (Figure 6.16e).

6.14 Encapsulated Fluorescent Green Proteins

Experiments similar to the one described in Section 6.13 were performed with GFP instead of quantum dots. Concentrations of 50 nM GFP in aqueous solution with 0.5 % NP-40 and 0.1 M MgCl$_2$ as the dispersed phase had to be used to successfully encapsulate the fluorescent proteins. For droplets with a diameter of 2 µm to 3 µm there should be an average of 100 to 400 protein molecules encapsulated, whereas considerably less where actually detected. Investigations showed that GFP stuck to the inner surface of the tubing leading to the inlet of the microfluidic device and to the microfluidic channels leading to the junction. No such behaviour could be observed once GFP was encapsulated within the droplets.

Detection was done as described in Section 6.13. The brightness of GFP compared to the quantum dots is several orders of magnitude lower and the fluorescent proteins bleach quickly. The background signal originating from the continuous phase and the droplet surface was comparable to the detected fluorescent signal from the GFP. The acquired data had to be background-corrected to yield meaningful intensity time traces.

GFP could be encapsulated within droplets and the fluorescence detected. Individual fluorescent proteins could be distinguished in the fluorescent images of droplets (Figure 6.17a). The droplets were observed over a time of 3 h and the number of encapsulated proteins did not change. This indicates no apparent leakage of proteins from the droplets and that the fluorescent activity was still preserved.

The detected fluorescence signal is confined to the droplet and decreases over time (Figure 6.17b). Bleaching steps can be seen for the background corrected intensity time traces (Figure 6.17c)

It was possible to encapsulate and detect a single fluorescent protein within a droplet (Figure 6.17d). The corresponding intensity time trace shows a clear, single
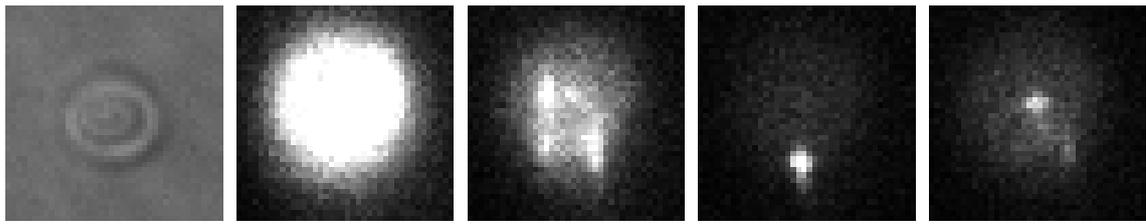
(a) Bleaching of droplet-confined GFP over time
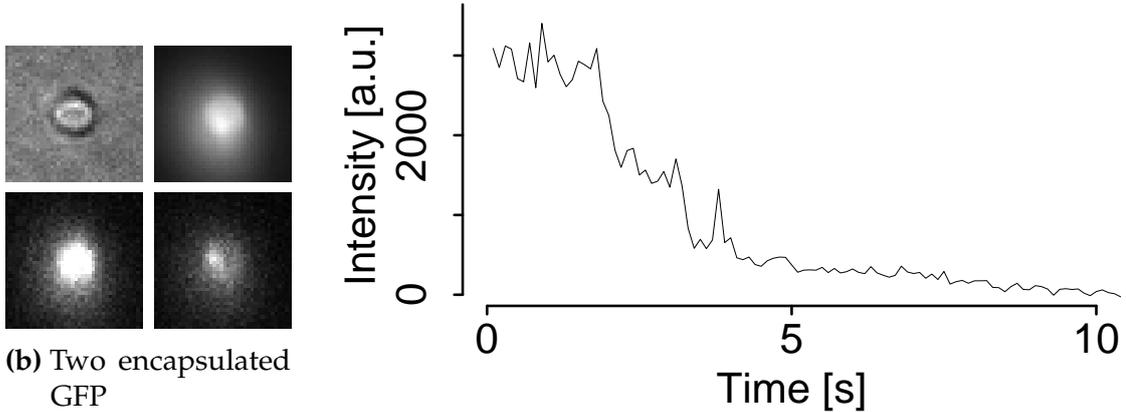
(b) Two encapsulated GFP

(c) Intensity time trace for the droplet seen to the left
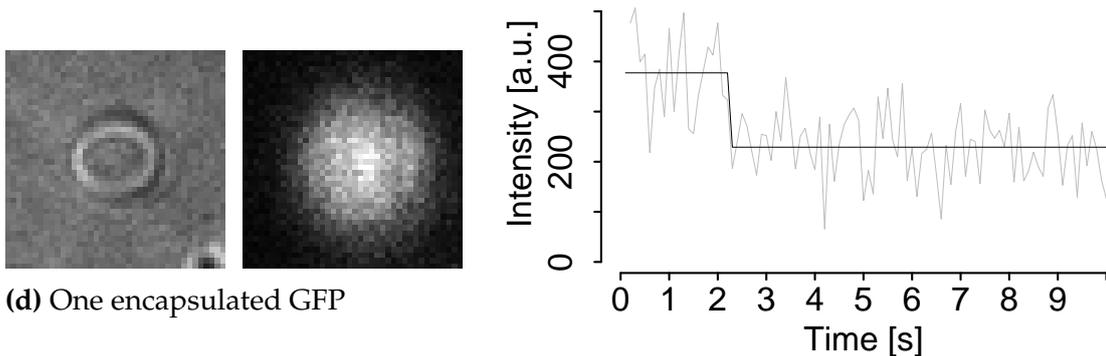
(d) One encapsulated GFP

(e) Intensity time trace for the droplet seen to the left

Figure 6.17 – **(a)** A droplet encapsulating several GFP. The first image shows the droplet under bright-field illumination, subsequent images show detected fluorescence and the bleaching of individual GFP over time. - **(b)** A droplet encapsulating two GFP. The top left image shows the droplet under bright-field illumination. The top right image shows an average over all acquired frames to highlight the confined fluorescence originating from the droplet that is brighter than the surrounding background. The bottom images correspond to an intensity indicating two, resp. one GFP. - **(c)** Intensity time trace corresponding to the droplet in (b). Two bleaching steps can be seen. - **(d)** A droplet encapsulating a single GFP. The top image shows the droplet under bright-field illumination, the bottom image shows a central single GFP. - **(e)** Intensity time trace in grey corresponding to the droplet in (d). The Algorithmic fit (Section 3.9) in black shows a single bleaching step.

bleaching step, equivalent to the single GFP molecule seen in the corresponding fluorescence image of the droplet (Figure 6.17e). The observed ratio of the step size to the standard deviation of the noise (2.3 ± 0.27) is comparable to the ratio found for surface bound GFP (3 ± 1), for examples see Figure 3.12)

Chapter 7

Discussion

Reasoning draws a conclusion, but does not
make the conclusion certain, unless the
mind discovers it by the path of experience

Roger Bacon

7.1 The Microscope

Several methods exists to investigate the binding of proteins to RNA. The advantage of colocalization measurements combined with the analysis of intensity time traces is their ability to obtain quantitative information about the non-averaged number of bound proteins. Total internal reflection allows a selective illumination of surface-bound molecules over longer time ranges and is therefore perfectly suited for these measurements.

Using three spectroscopically different fluorophores allowed to employ two labelled proteins together with the labelled pre-mRNA to not only look at the relationship between the mRNA and each protein but also at protein-protein relations. The chosen fluorophores are GFP, mCherry and Cy5. Cy5 can easily be attached to mRNA molecules and GFP and mCherry are available in existing vectors that allow to clone them into the proteins of interest.

The wavelengths of 488 nm, 561 nm and 633 nm do not exclusively excite a single fluorophore each. The wavelength of 488 nm excites GFP, but also, to a lesser extent, mCherry. The same is true for the wavelength of 561 nm that not only excites mCherry, but also, again to a lesser extent, Cy5. This makes it necessary to separate the resulting fluorescent emission. The separation can be done by a combination of various beam splitters, filters and multiple cameras for a simultaneous acquisition. Disadvantages of this are the remaining cross-talk between different channels, reduced signal-to-noise ratios, different beam paths and differences between cameras.

The diluted splicing samples that were ultimately looked at are in a stable, fixed state. It is not required to detect the fluorophores simultaneously, instead it can be done sequentially. This removes the necessity of a complicated detection system. A single detection pathway can be used with multiple notch filters to block the different excitation wavelengths. This guarantees the highest possible signal to

noise ratios, crucial for the later identification of steps in the recorded intensity time traces. Only one kind of fluorophore should be excited at any time as no filters are used to separate the emission. This is not necessarily guaranteed as the wavelengths used for excitation can, even if with much less efficiency, excite a second fluorophore. This can be avoided by carefully choosing the order of excitation. An initial excitation with the longest wavelength (and lowest energy) excites only a single fluorophore that can be excited at these low energy levels. The illumination is continued until this initial fluorophore is bleached and not able to fluoresce any more. A subsequent excitation with the next shorter wavelength for the next fluorophore requiring a higher energy for its excitation is now restricted to this second fluorophore. The bleaching process is then repeated and followed by the final excitation of the last fluorophore that requires the hight energy for its excitation. This sequence of excitation guarantees a spectrally separated detection from the three employed fluorophores.

It is possible to have an emission with a shorter wavelength than the wavelength used for excitation if multiple photons concurrently excite a single fluorophore. Not only are high photon densities required but also do wavelengths of the photons for the excitation have to match, making this an unlikely case. Another possibility is the occurrence of a reactivation of a formerly bleached fluorophore as reported for Cy5 (Bates et al., 2007; Dempsey et al., 2009b). The laser intensities used for reactivation are generally a few magnitudes higher than the ones used in the setup used in this work and a triplet quencher is required which was also not used in this work. Reactivation events could not be observed and even if they would occur, they would not register as an event in the later analysis as it would not stand out enough above the background. Similarly, the presence of other proteins could lead to a reactivation or FRET-like energy transfer. No indication for this could be seen in the experiments and no references for it were found in the

literature that would apply to the system used here.

Optimising the available microscopic setup both in terms of detection efficiency and towards the lowest possible background and noise levels ensures a more reliable identification of spots, colocalization and bleaching steps. The setup employs a minimal set of narrow notch filters to ensure the highest detection efficiency. Following the experimental procedure described, no disadvantages of this detection setup were observed.

One way to increase the signal is to increase the intensity of the excitation which leads directly to an increased emission. The downside is a shortened time before the fluorophores bleach. This could possibly lead to intervals between individual bleaching steps that cannot be resolved and unobserved bleaching during the initial frames. Conversely, decreasing the intensity of the excitation leads to a longer active time of the fluorophores before bleaching, increasing the average time between bleaching events for an easier detection of steps in the intensity time traces. The lowered intensity also results in a lower signal, complicating the accurate detection of fluorescent spots and the analysis of the intensity time traces.

The longer bleaching times also increase the necessary acquisition times, as each species of fluorophores has to be bleached completely. Acquisition times should be short enough to ensure a consistent state of the sample over the whole set of acquisitions. Additionally, the amount of recorded data increases with longer acquisition times. This is not a crucial factor, but it makes data handling, storage and analysis more difficult.

The other side is the background noise. One source is the ambient light in the room. Its interference with the acquisition could be obviated by covering the whole microscopic setup with an opaque cover.

An important setting of the camera is the exposure time. Longer exposure times lead to an increased signal-to-noise ratio as the readout noise is proportionally

reduced and the background noise is averaged over a longer time period. The drawback of a longer exposure time is the worse time resolution. Less time points in the intensity time traces complicate the correct identification of subsequent bleaching step. A good compromise was found with an exposure time of 0.1 s.

Further camera settings concerning the actual readout of the information from the camera chip were optimised. Not only did this bring the mean values of the background down to the lowest possible levels but also did it severely reduce the occurrence of spikes. Their initial incidence of over 8 % dotted the recorded images, obscuring spots and leading to false positives. Their significance at the identification of spots is reduced by the use of averaged images but not for the intensity time traces. The repeated occurrence of spikes affects the generated intensity time traces more severely. The optimised settings of the camera decreased the incidence of these spikes to virtually zero.

The performance of the setup with these various improvements resulted in a good quality of the acquisitions. With the exception of the occurring scattering that is visible with an illumination at 633 nm, the variation of the background is a magnitude less then the collected signal and does not influence the identification of spot in a negative way. In the intensity time traces its variations over time are less than the Gaussian noise and therefore negligible.

As acquisitions are done over up to 3 min, the positional stability of the sample over time has to be guaranteed. A lateral movement of the sample did not occur. The implementation of an auto-focus compensated for the occurring vertical changes. The ability to automatically focus in correspondence to the currently used wavelength provides the additional advantage of a quick and precise switch between different wavelengths without skipping the valuable first frames.

An important feature of the microscope is its close dovetailing with the computer. The lasers, the stage, the quartered photodiode and the camera are all con-

nected with the computer. Not only does this increase the ease of handling and allows the implementation of the automatic focus, but it also allows the acquisition to be automated. This automation brings several advantages. It greatly reduces the time needed to perform one set of experiments, it minimises the potential for human errors during the acquisitional process and ensures a high consistency between different acquisitions. The implementation was done in LabView. The provided user interface simplifies the interaction with the microscope and allows a quick familiarisation for new users.

Due to the fact that operating the microscope and recording the data is managed by the LabView program, annotations accompanying the recorded data can be created automatically. This is helpful as a documentation of the performed experiments and for the later analysis of the data.

7.2 The Analysis

The analysis of the recorded data consists of several individual steps. The individually acquired images have to be assigned to a wavelength, spots have to be identified, their positions determined and colocalized spots have to be assigned to each other. The intensity time traces of the colocalized spots have to be generated and analysed for the number of bleaching steps. Finally, the resulting information has to be documented and presented in an accessible way.

Single images of the acquisition contain the combined noise from the background, the emitted fluorescence and the camera readout. The intensity of fluorescence varies over time. A single image is not representative and does not allow an accurate detection of spots. Due to the high positional stability of the sample, multiple individual images of each wavelength can be combined. The two described ways of generating these images, either using the maximum or the mean pixel value, complement each other. It is important to not exclude any spot that might be missed by one of these methods as this would introduce a bias into the analysis.

The identification of spots, the determination of the colocalization and the assignment of steps for the intensity time traces can each be done by eye. This is time-consuming and prone to an undocumented bias. Additionally, it requires time for training and practice for people new to the analysis of single-molecule data who have to familiarise themselves with the procedure to obtain reliable and comparable results. Employing algorithms for these steps facilitates these steps and opens the door to further automation.

Several algorithms and comparisons of them exist for both the detection of spots and the analysis of intensity time traces (Carter et al., 2008; Ruusuvuori et al., 2010; Smal et al., 2010; Xue et al., 2011). These algorithms are often tailored to the specific task at hand. Often no usable implementation is presented, making a

judgement of their usefulness for the different data difficult and time-consuming. Various theoretical descriptions exist especially for the analysis of time traces without any actual implementations or not enough validation against actual data. The performance of several algorithms was assessed and found less than satisfactory for the analysis of the data of this work. It was decided to implement algorithms tailored to the specific task at hand.

The main obstacle for the identification of fluorescent spots was the widespread of representations of them in the recorded images. The identification of the fluorophore Cy5 was uncomplicated. The strong and consistent emission of single fluorophores of Cy5 make them stand out. As Cy5 fluorophores used as marker spots were identified first, they could be excluded without any bias to the final results as their visual representation should be independent of the number of proteins bound to them. This was done for spots both too close to each other as well as for those too close to the border of the acquired image.

The main challenge was the identification of spots from the fluorophores of GFP and mCherry. Spots could contain anything from a single fluorophore to up to 30 of them. The individual intensities of the emission from these fluorophores also varied drastically. This resulted in a diverse representation of these fluorophores in the composite images. A set of parameters that allows fluorescent spots of GFP, mCherry and Cy5 to be identified reliably without including false positives could not be found and two different parameter sets were identified. As missing out on the hard-to-detect spots leads to a bias towards easier-to-detect spots which are often the brighter one with multiple fluorophores present, it was decided to accept the inclusion of false positives at that stage. They are then excluded in the next step at the assignment of steps to their intensity time traces. Local fluctuations might be seen as spots in the composite images, but an additional requirement is the visibility of bleaching steps and a consistent intensity above background

levels. This ensures an ultimately unbiased identification of spots.

Multiple fluorophores within a single complex cannot be distinguished directly, as their distance between each other is below the resolution limit of the microscope. Further information about the number of fluorophores located within one visible spot can be obtained from the intensity of fluorescence from this spot. The method used identifies irreversible changes due to bleaching in the intensity over time and counts those steps. Background corrected intensity time traces are used as a starting point for this.

Ideally, the background corrected traces include only fluorescence from the spot under investigation and corrections for the background are done without any influence originating from other spots. In the recorded data, the complete intensity profile of a single spot can easily reach across a distance much larger than its full width half maximum. This way of calculating the background corrected traces ensures, while it does not completely exclude the influence of other spots, that their influence on the intensity time trace is kept well below the intensity levels of the actual spot.

As said for algorithms before, the ultimate algorithm for the assignment of steps does not exist. The lack of *a priori* information further impedes the analysis in the case of the intensity time traces. Neither the number of plateaus nor the height of them or the time between bleaching steps is known. Information about their distribution exists but does not help in an individual case. It is further hindered by the possibility of anti-bleaching events where the intensity increases again due to re-activation or blinking.

A good starting point was found in the Bayesian step detector. It provided an unbiased way to assessing the likelihood for an step to have occurred at a certain point within the intensity time trace without any *a priori* information besides the assumption of an occurring step and Gaussian noise. The algorithm performs

well but the underlyingly assumed step does not have to occur or can occur multiple times. Recursively applying the algorithm to the intensity time traces and carefully choosing the threshold parameter for a step lessens this impact. In the end, the output from the Bayesian step detector had to be refined to enhance the quality of the step assignment.

The implementation of the data analysis with these algorithms was done in Matlab as it is well suited to analyse larger amounts of data, especially for data that fits into the structure of an array, for its ability for rapid prototyping of algorithms and differing implementations and for its easily accessible user interface.

The presentation of the data is done as a bar chart to quickly convey the information extracted from the recorded data. The number of marker spots and the number of colocalized spots is shown to give an indication of the statistical relevance of the results shown. The distribution of the number of fluorescent proteins present can range from 1 to sometimes up to 30. It was decided to merge occurrences for numbers of steps above a certain threshold into one category to allow a more concise presentation. The category for uncertain assignments of bleaching steps, represented as X, was included to avoid the exclusion of colocalization events where the intensity time traces are of poor quality. It also allows certain spots to be safely excluded so as to not force a biased decision. A two-dimensional bar chart is used for experiments with three different fluorophores used simultaneously.

7.3 The Preparation

Initial preparations of nuclear extract from 293T cells with SRSF1-GFP had an enormous over-expression of labelled SRSF1. It could not be used for single-molecule experiments. It was decided to prepare nuclear extracts from HeLa cells instead where the expression levels were more comparable to nuclear extracts without a labelled protein.

The labelling efficiency for SRSF1-GFP nuclear extract from HeLa cells was 89 %. This high efficiency means that the number of fluorescent SRSF1 bound to the pre-mRNA are almost one to one the number of overall bound SRSF1, both labelled and unlabelled, allowing a direct comparison. The same is true for SRSF1-mCherry with a labelling efficiency of 93 %. Even with HeLa cells, both labelled variants of SRSF1 are over-expressed compared to the commercial nuclear extract. The levels of U1A as an essential splicing factor were assessed to allow a comparison of splicing related protein levels. Taking the abundance of U1A in the commercial nuclear extract into account, SRSF1 is 7 times over-expressed compared to U1A in the SRSF1-GFP nuclear extract.

The labelling efficiency of SRSF1-ΔRRM2 could not be assessed. No band was found in the Western blot that corresponded to SRSF1-ΔRRM2-GFP. An alternative probing with an antibody for GFP could have given an approximate estimate if the effectiveness of both antibodies would be comparable. A more suitable antibody that is now available, recognising one of the remaining domains, would allow to detect the labelling efficiency directly. The single-molecule experiments with its nuclear extract showed comparable levels of labelled proteins as for the SRSF1-ΔRS nuclear extract. The most likely reason for its absence in the Western blot is that the removal of the RRM2 domain also removed the part recognised by the SRSF1 antibody. As both the nuclear extracts for SRSF1-ΔRRM2 and SRSF1-ΔRS were prepared simultaneously and showed equal amounts of unlabelled

SRSF1 as well as comparable amounts of fluorescence on a single-molecule levels, it was assumed that the labelling efficiency of both nuclear extracts is equal at 50 %. It has to be remarked that the labelling efficiency of the mutated versions of SRSF1 is 100 %. Labelled versions of non-mutated SRSF1 were not transfected.

Optimised annealing conditions and an excess of pre-mRNA allow for less than 2 % of free tethering oligo with Cy5 as seen on a denaturing gel. No free Cy5 could not be achieved. Reasons for this can be a non-perfect initial annealing with a later dissociation. The amount of free oligos is comparable to the amount of colocalization found for the experiments with a complementary oligo annealed to the tethering oligo. This remaining colocalization is most likely the result of the binding of SRSF1 to the remaining free Cy5 even under optimal conditions. This colocalization percentage is much smaller than the colocalization percentage encountered for all other single-molecule experiments and does not influence the results significantly (Figures 4.2 and 4.5). Any remaining amount of free pre-mRNA without any fluorescent label does not affect the results as it is unlabelled and not visible and does not interfere with the splicing progress of labelled pre-mRNA.

The attachment of the pre-mRNA to the surface showed an unexpected behaviour. Labelled pre-mRNA stuck to the surface without any streptavidin present that provides binding sites for the biotin end of the tethering oligo. The presence of streptavidin did not even increase the amount of pre-mRNA confined to the surface. This can be explained if the biotin end is not accessible to the streptavidin. This can be due to a conformation of the pre-mRNA that hides the biotin internally. This seems unlikely as the biotin group is attached at the end of the pre-mRNA and this effect could be observed for a wide range of different constructs of pre-mRNAs. Another explanation is that the biotin is smothered by the protein complex forming around the pre-mRNA. Several SRSF1 are found in a

complex with the short tethering oligo alone and, in complex E, big aggregations of SRSF1 can be seen around the pre-mRNA. Other non-fluorescently labelled proteins will most likely behave in a similar way, especially from the SR family of proteins. This would also explain how the pre-mRNA can nonetheless be found at the surface. SRSF1 was found to be extremely likely to non-specifically bind to the surface and the same is true for other proteins. With a pre-mRNA wholly covered by proteins there exists a multitude of these non-specific binding possibilities that will eventually confine the RNA molecule to the surface.

This goes in hand with the poor performance of BSA-coated slides. BSA is able to alleviate the non-specific binding of labelled SRSF1, but only slightly. If SRSF1 is always a part of big accumulations of proteins, a myriad of possibilities exist to ultimately attach it to the surface. The only modifications of the slides that could mitigate this entirely were those that created a strong hydrophobic surface. Almost no binding of labelled SRSF1 to the surface could be observed for concentrations up to 100 times higher than normally used. This effect was spoiled by spots identical in shape and behaviour to those originating from SRSF1-GFP, making the analysis of the acquired data infeasible. The spots most likely stem from impurities of the used chemicals, as they are not designed for single-molecule work. Using PEG for coating the slides showed some good initial results but these results could not be reliably replicated. Using PEG or hydrophilic surfaces would require more work to be able to usefully apply them to the experiments at hand. The poor control over the amount of proteins and pre-mRNA present on the surface restricted the range of experiments that could be performed.

7.4 The Protein

Although the binding sites of SRSF1 on pre-mRNA are mapped and its role during the splicing process are known, the actual mode of interaction remains uncertain. It binds to regulatory motifs containing a GGA motif that stimulate splicing, yet its actual presence on an ESE that stimulates splicing is undetectable, while SRSF1 still binds to the pre-mRNA. It is involved in constitutive splicing even with short exons that contain no binding motif for SRSF1, where it is required, so its sole role is not just an interaction with the U1 snRNP or U2AF. SRSF1 needs to be dephosphorylated for constitutive splicing, but not to act as an enhancer in alternative splicing. For all cases there is nothing known about the actual number of bound SRSF1 during splicing. It was investigated how many SRSF1 are concurrently present and possible factors that influence their numbers to gain a deeper understanding of the role of SRSF1 and to try to reconcile the previous findings.

Splicing is a highly dynamic process and complex formation of individual pre-mRNA molecules can occur at different rates. The formation of splicing was stalled at complex E, A and C to accumulate species of predominantly one complex. All conditions allowed an efficient stalling of the complex formation.

Stalling the splicing progress in complex E by a depletion of ATP led to big aggregations that did not leave the wells of the native gel due to their immense size. It was aggravated by more severe depletions of ATP. This could not be seen for Globin M which did not form complex E.

The number of SRSF1 present in complex E is split into equal proportions of one, two, or more than two proteins present. The prominent occurrence of a large number of SRSF1 present conforms to the results from the native gel electrophoresis. Aggregations seen for single molecules are accompanied by aggregations in the native gels. The formed aggregations contain a multitude of proteins, among

them SRSF1.

The addition of ATP and with it the advancement to complex A leads to an almost complete breakup of aggregations and a striking reduction to almost exclusively one remaining SRSF1, accompanied by a slightly reduced colocalization. This is contrasted by the results from complex C, where the amount of colocalization and the individual numbers of bound SRSF1 are increased again. The increased amount of SRSF1 present can be due to the beginning formation of the exon junction complex, where it was found that multiple SRSF1 are present. The numbers of bound SRSF1 are in agreement with the numbers found by Singh et al. (2012).

The amount of bound SRSF1 in complex A stands out. Both the complex before, complex E, and a later complex, complex C, contain more bound SRSF1. This alludes to an existing mechanism behind this reduction. SRSF1 is linked to the selection of 5′ splice sites and the single bound SRSF1 coincides with the number of 5′ splice sites present in the construct Globin C. The adenovirus construct Ad1 CC with two alternative consensus 5′ splice sites does indeed show a peak around two bound SRSF1 in complex A, with the number of bound SRSF1 coinciding with the number of possible 5′ splice sites. The number of U1 snRNP present is similarly with one or two present, contradicting the results from Hodson et al. (2012). Mutating one of the alternative 5′ splice sites brings the number of bound SRSF1 back to one, showing a link between the number of SRSF1 and the number of 5′ splice sites. Contradicting evidence came from construct C174C, where only a single SRSF1 was present in complex A even with two 5′ splice sites. It might be that the choice of a 5′ splice site was already made at this stage, hence only one remaining SRSF1. The distance between the splice sites is 174 nt for C174C compared to the 93 nt from Ad1 CC. Different separating lengths between splice sites influence the choice of splice site, this might also affect the mechanism of

splice site selection. For CEC with three exons and therefore two separate 5′ splice sites are one or two SRSF1 present. This fits the expectation as both splice sites are ultimately used but do not have to be spliced simultaneously.

The U1 snRNP binds to the 5′ splice site and does the selection of the 5′ splice site in conjunction with SRSF1. A random amount of SRSF1 is bound to the pre-mRNA without a strong consensus 5′ splice site present in Globin M and simultaneously blocking of the U1 snRNP with a large proportion of aggregations. The addition of ATP lowers the amount of aggregations and the number of bound proteins slightly, but the aggregations still occur on levels equal to Globin C in complex E without the blocked U1 snRNP, hinting at an E-like complex that can form even in the presence of ATP. Only without the blocked U1 snRNP is the number of bound SRSF1 again reduced. The striking change in complex A is clearly linked to the presence of the U1 snRNP, seemingly independent of the presence of a strong consensus 5′ splice site.

The RS domain of SRSF1 can be heavily phosphorylated and phosphorylational changes of SRSF1 play an important role in splicing. Incubation with PhosStop shifts the equilibrium present towards phosphorylated proteins. This change of the proteins of the nuclear extract leads to a change of the binding distribution of SRSF1 under otherwise same conditions. Already in complex E is only one SRSF1 bound to the pre-mRNA, contrarily to normal conditions where this same distribution can only be seen in complex A. The change that is introduced by the addition of ATP is seemingly linked to a change in phosphorylation. This could not be found in the absence of the U1 snRNP. The effect that PhosStop had on the binding of SRSF1 in complex E cannot be seen without the U1 snRNP present. The additional removal of the consensus 5′ splice site leads to similar results albeit with poor colocalization. The binding distribution for conditions of complex C with the addition of PhosStop is different from the one found for normal conditions. This

is in agreement with the requirement of a dephosphorylated SRSF1 to perform a first catalytic step (Cao et al., 1997).

No direct effect of PhosStop on SRSF1 could be shown. On the one hand, it can be that PhosStop and the effect of it seen by including it in the incubation is not mediated directly by SRSF1. On the other hand, the experiments performed were not that conclusive. The antibody used might not recognise all occurring changes and a positive control of definitely phosphorylated SRSF1 was not available.

Assuming that the phosphorylation of SRSF1 was not changed suggests that either the phosphorylation of SRSF1 itself does not play a role at this stage and it is moderated through phosphorylational changes of a secondary factor. The actual phosphorylation of the domains of SRSF1 is then free to be used for other controlling mechanisms, hinting at how SRSF1 is able to fulfil its multiple roles in splicing.

The mutations of SRSF1 with its missing domains RS or RRM2 resulted in similar binding patterns for both complex E and complex A and all conformed to Poisson distributions with each having a reduced mean from complex E to complex A. This suggests that the mutants do not take part in splicing, as shown by Cáceres and Krainer (1993) for the RRM2 domain and by Zuo and Manley (1993) for the RS domain for the here present constitutive splicing. They are rather only present in the big aggregations that can form at the pre-mRNA, and this only in small numbers, leading to the observed Poisson distribution. Their low numbers in even those suggests that they are there rather by chance than actively taking part. The effect of the missing RS or RRM2 domain might therefore be already the missing recruitment to these aggregations, limiting the later participation of the so mutated SRSF1 in splicing.

The antagonising effects of hnRNP A1 towards SRSF1 was thought to be through the reciprocal blocking of binding sites (Eperon et al., 2000; Zhu et al., 2000). A

strong correlation of SRSF1 to hnRNP A1 was found in both complex E and complex A across the whole range of bound protein number, but not for numbers of bound proteins of up to 3, where a clear antagonistic effect is seen. The antagonistic effect of both proteins and their competition for binding sites would not be expected for high protein number as they indicate a presence of proteins not due to them binding to the pre-mRNA but due to other interactions.

Samples taken at three timepoints during splicing showed all binding patterns similar to each other that did not reflect the binding seen for the stalled complexes. A reduction of bound SRSF1 could be observed as the number of bound SRSF1 at 5 min compared to the timepoints at 10 s and 30 min is reduced by 20 %. The similar binding patterns suggest that an equilibrium of SRSF1 binding was observed as a result of the observation of a wide mixture of formed complexes present. Complex A can form within seconds and a pure complex E is unlikely to be caught even after 10 s without stalling the splicing progression. Each step that is leading to the formation of complex C is kinetically independent, resulting in different rates of progression, creating a wide mixture of other splicing complexes or completely spliced pre-mRNAs.

SRSF1 in the used nuclear extract is overexpressed compared to untransfected nuclear extracts. It was found that an extremely overexpression of SRSF1 as found in nuclear extracts from 293T cells shifts the binding pattern towards an extremely high colocalization with a distribution dominated by big aggregations. The severe overexpression of SRSF1 smothers the pre-mRNA even in complex A where usually one SRSF1 remains. This effect could not be seen in the nuclear extract. used in this work. The formation of aggregations can also be seen under the same conditions in native gels that are done with the commercially available nuclear extract. The aggregations are therefore not due to the overexpression of SRSF1. Using an abundance of RNA to decrease the ratio of SRSF1 to pre-mRNA showed

no significantly different binding pattern in complex E as compared to normal conditions. From these results it was concluded that the levels of expression of SRSF1 did not affect the observed binding patterns.

Experimental results from covering the surface with PEG suggest that the uncovered surface stabilises the confined complexes. A reduced colocalization was found without this stabilisation. Paradoxical to this seems the high proportion of aggregations found in complex E. This is explained by the fact that these aggregations are more likely to bind to the surface even if it is covered in PEG, and are then stabilised, showing up as a higher proportion than usual. The used uncovered glass surface interferes with the formed splicing complexes in a favourable way and without changing the observed binding patterns.

The construct 2xMut Globin C does not splice and forms no stable complexes independent of the presence of the ESE at its 3' end. The binding patterns of SRSF1 with and without the ESE show no difference, despite the mutations in potential SRSF1 binding sites. The loss of aggregations is the only observed difference from complex E to A, and this is due to the addition of ATP. The observed behaviour of SRSF1 is not linked to the stable formation of complexes and does not rely on the ability of the construct to splice.

For the construct Short Exon Globin C, the presence of the ESE increased the splicing efficiency by a factor of five, although no major difference could be seen in its formation of splicing complexes. No difference of the observed binding of SRSF1 could be observed that accompanied the changes in the splicing efficiency. Instead the same binding distributions was seen as observed for Globin C, albeit with an extremely reduced colocalization in complex E. Presumably, the low colocalization is due to the depletion of ATP that prevents helicases from unwinding inhibitory secondary structures of the pre-mRNA, which is not the case any more in the conditions for complex A. The role the ESE plays in relation to SRSF1 has

no influence in the early stages of splicing. From the formation of complexes up to complex C that show no difference with or without the ESE do can be concluded that the enhancing effect of the ESE plays only a role after the formation of complex C.

The addition of the ESE also allowed the Short Exon Globin C construct to be spliced with a blocked U1 snRNP. Globin C has a greatly reduced formation of complexes and splicing efficiency with the U1 snRNP blocked, as expected. For Short Exon Globin C is this much less pronounced, hinting at an U1-independent splicing pathway. The possibility of U1-independent splicing was first reported by Crispino et al. (1994) for substrates based on β-Globin in nuclear extracts that were complemented with an excess of SR proteins. Crispino et al. (1996) identified crucial sequences at the branch point and around the 5′ splice site that allowed a U1-independent splicing even without an excess of SR proteins. They suggested that either these sequences promote the binding of the U2 snRNP by providing an optimal sequence for SR proteins or other factor, or that the structure of the pre-mRNA may in fact be the decisive factor. The independence of splicing of the U1 snRNP for the Short Exon Globin C construct was achieved by shortening the flanking exons, removing any regulatory sequences that could positively contribute to the U1-independent pathway. Either inhibitory sequences were removed or the resulting structure changed in a way to remove the requirement for the U1 snRNP.

It was concluded that big aggregations of proteins are naturally present as it was shown that part of these aggregations are large numbers of SRSF1 and hnRNP A1 and, as the aggregations seen in native gels are much larger than the ones observed on a single-molecule level, it can be concluded that a vast amount of other splicing related proteins is present. A process that is mediated by a change in the phosphorylation of a secondary factor or of SRSF1 itself removes the aggregated

proteins, independent of an 5' splice site present. Xiang et al. (2013) reported that the dephosphorylated RS domain of SRSF1 is in a fully disordered state and changes into an "arch-like structure" once phosphorylated. Han et al. (2012) and Kato et al. (2012) showed that disordered domains play an important role in forming RNA granules, which have a droplet-like behaviour (Brangwynne et al., 2009). This matches the findings of Misteli et al. (1997) that splicing factors like SRSF1 are recruited to nuclear speckles. We propose that the observed aggregations are of the same nature, formed by the presence of the dephosphorylated, disordered RS domain. The phosphorylation of the RS domain introduces an ordered structure, abolishing the ability of SRSF1 to partake in the formation of the aggregates.

Only in the presence of a 5' splice site does one SRSF1 remain, most likely as part of a more stable complex with the U1 snRNP. This removal process is greatly facilitated by the presence of the U1 snRNP, albeit it is not necessary to bind to the pre-mRNA directly. The mechanism works independently of the stable formation of complexes.

SRSF1 needs both its RS and RRM2 domain to participate at splicing at this stage. This is in agreement with Cho et al. (2011a) and Wu and Maniatis (1993), but the findings from the single-molecule experiment suggest that most likely is this not because it lacks means of binding to the pre-mRNA or interacting with the U1 snRNP, but because it already cannot hold on to the formed aggregations.

Cho et al. (2011a) found that SRSF1 and the U1 snRNP might be required for the formation of complex E. We found that U1 snRNP is at least not required for the complex E like binding of SRSF1 and that splicing of the Short Exon Globin C construct can occur without the U1 snRNP, suggesting that the binding of SRSF1 is more relevant in constitutive splicing of certain substrates.

Eperon et al. (1993) and Kohtz et al. (1994) suggested that SRSF1 binds independently of the U1 snRNP and interacts with it to stabilise the binding of the U1

snRNP, whereas Jamison et al. (1995) argued that the binding of the U1 snRNP has to precede the binding of SRSF1. Our findings strongly support the first view. SRSF1 is present without the U1 snRNP. The removal of excess SRSF1 is facilitated by an interaction with the U1 snRNP that simultaneously stabilises one SRSF1 linked to a 5′ splice sites.

We found no effect of an ESE on SRSF1 at this stage. We did see an antagonistic effect of SRSF1 and hnRNP A1 for low numbers of proteins present on a single-molecule level as suggested before (Cáceres et al., 1994; Eperon et al., 2000; Zhu et al., 1994).

Another process independent of the selection of the 5′ splice site leads to increasing numbers of SRSF1 present in complex C, after the formation of the lariat. We conclude that this is the forming exon junction complex (Singh et al., 2012).

7.5 The Droplets

The ultimate goal of the microfluidic part of this work was to obtain a fluorescent signal from a single protein at biological relevant concentrations encapsulated in a water droplet. The sizes of the generated droplets would not have to be that small as one could simply adjust the dilution of the sample accordingly to the size of the droplets to encapsulate a single fluorescent protein. Only the added requirement of biological relevant sample concentrations demands that the size of the droplets is in the µm range.

Several methods like sonication and extruding exist to generate droplets of that size. Their disadvantages are the heterogeneous size of droplets they produce and the often too harsh conditions during the preparation for biological samples. Whilst microfluidics combined with flow focussing is primarily seen as a method to enable a high-throughout approach for a variety of experiments, it has other favourable properties. We chose it for its ability to both gently encapsulate the contents of the dispersed phase in a predictable way without disrupting any biological processes and for its potential to reach ever so small droplet sizes.

We iterated on the initial design and improved key features to allow and accommodate the generation of 1 µm droplets. The integrated filter system at the inlets performed well. The size of its inlying channels was reduced from 11 µm down to 8 µm which showed to be sufficient for an efficient running of the microfluidic device. A blockage of the junction that rendered the mixer unusable occurred only on rare occasions and never did the filter system itself prevent the flow of either phase.

Additional columns in the centre of the inlets and the outlet prevent the sagging of the ceiling. The addition of them to the outlet served another important role as the punching of the holes in the PDMS, especially for the outlets, showed to be difficult for feature sizes with heights less than 2 µm for lack of contrast.

The added columns increased the visibility of the outlet for an uncomplicated production of the final microfluidic device.

The necessary volume that has to be injected before the generation of droplets starts was minimised by reducing the width and the length of the connecting channels down to the limits imposed by the requirement of a sufficient distance between the inlets and the outlets. The distance has to be large enough to accommodate the punched holes without them interfering with each other.

The initial 5 µm junction allowed the generation of droplets down to the size of 5 µm. For smaller droplet sizes did we decrease the size of the orifice of the droplet-generating junction and the dimensions of all other surrounding structures. The smallest reliably achievable feature width was 3 µm, the size we chose for the orifice.

The final design had to be accurately transferred from the mask onto the wafer. It is not known what are the minimal requirements of the illumination both for its homogeneity and perpendicularity to ensure this transfer. The quality achieved by our two self-build devices was sufficient enough for features down to a width of 3 µm and a height of around 1.5 µm.

The parameters from the guideline for the photolithography process for the epoxy based photoresist SU-8 2000 did not give optimal results. The applied exposure energy had to be significantly changed. This could be due to wrong estimates about the final exposure energy that the wafer receives from our illumination devices as we had to estimate the loss imposed by the mask. Further modifications were necessary to ensure a sufficient attachment of the photoresist to the wafer. A perfect transfer of the design could often be achieved but did not stick to the wafer. This problem could be eliminated by using a longer time for the post bake and a shorter and more gentle final development of the wafer and photoresist. The final, optimised quality of the transferred pattern on the wafer was sufficient for

the reliable creation of high quality features down to a width of 3 μm, allowing the fabrication of microfluidic devices incorporating the desired 3 μm junction.

The diameter of the generated droplets depends on the geometry of the flow-focussing junction, the interfacial tension and the flow rates of the individual phases (Section 1.3.3). The effects of all three have to be combined to achieve the generation of droplets with a diameter of 1 μm. The geometry of the flow-focussing junction was scaled down from the original design as far as our implementation of the photolithographic process allowed.

Changing the flow rates of each phase is done by changing the pressure that is applied to them. Increasing the applied pressure increases the flow rate, and *vice versa*. This would be unconditionally true for our microfluidic system if the two different phases were not connected with each other. It was found that the two flows from the inlets of the dispersed and the continuous phase compete for the single outlet and this results in a delicate interplay. Increasing the pressure and flow rate of one phase decreases the flow rate of the competing second phase, even if this phase maintains the same applied pressure. Thus, for one phase, different applied pressures can result in different flow rates. Together with dynamic changes of the microfluidic device, reaching and maintaining the wanted droplet size can be difficult. This was true for our initial setup using syringe pumps. Reaching the desired equilibrium took up to 30 min. The switch to a pressure-driven flow controller allowed to reliably reach the desired conditions within seconds and keep them over the time needed to generate a sufficient number of droplets. This degree of control was crucial for the delicate ratios that had to be kept for droplets of 1 μm.

As predicted, higher flow rates are favoured for smaller droplets. Therefore a more viscous, dispersed phase that allows higher flow rates would be preferred. However, a less viscous phase like water as the dispersed phase allowed to often

reach the critical higher ratios of continuous phase flow rate to dispersed phase flow rate for smaller droplet sizes. Yet we had to reduce the viscosity of the water phase to allow the generation of droplets with the 3 µm junction as the interfacial tension at the orifice of the 3 µm junction prevented the water phase from breaking through the oil phase.

The highest available flow rates are limited by the applied pressure, which was found to be around 900 mbar for a reliable usage of the microfluidic device. Thicker glass substrates could be used to allow even higher pressures. This would negatively influence imaging the generated droplets as we ultimately imaged them on the same device. We decided to run experiments slightly below the critical parameters which limits the available droplet sizes but allows a more reliable operation. As described, only a certain set of ratios could be used. Of those, the highest available ones resulted in the smallest droplets. The disadvantage of these is the occasional breakdown of the droplet generation due to tiny environmental changes. These ratios were not used to ensure a reliable droplet generation.

Staying within these limits allowed nonetheless to generate droplets with a diameter of 1 µm within a stable flow-focussing regime. The droplet sizes were reached without thread-formation. Compared to other methods did this allow the rapid generation of extremely homogeneous, bio-compatible droplets.

We tested several different surfactants. One surfactant fulfilled all our requirements. This surfactant stabilised droplets over the full range of sizes that we could generate with our microfluidic devices and the provided stability for the droplets over time was satisfactory. We could image droplets collected within the same microfluidic device for up to 4 h.

Initially, droplets were generated on one microfluidic device and imaged on a second one. The rapid degradation of droplets after their re-injection impeded the imaging of these droplets. We decided to enhance the current microfluidic

225

design to incorporate an area on the same device to image droplets without the otherwise necessary re-injection. The rapid degradation of the droplets could not be seen in this case. The conditions during the re-injection might have caused either the degradation of the surfactant, leading to a reduced ability of stabilising the droplets, or removed the number of surfactant molecules at the interface of the droplets below the critical amount.

The achievement of droplets with a diameter of 1 μm was not an absolute requirement for the fundamental objective to image single fluorescent proteins. It was found that slightly larger droplets allowed an easier handling and imaging. With these droplet of a size in the range of 2 μm to 3 μm could both quantum dots and GFP be encapsulated and imaged. The resulting intensity time traces could be used to gain the number of fluorescent objects.

The generation of droplets with a size of 1 μm and the detection of single-molecule fluorescence was done previously by ultrasonication (Reiner et al., 2006). A more gentle approach for biological samples is the generation by flow-focussing as done by Shim et al. (2013), who reached droplet diameters of 2.5 μm and encapsulated single enzymes. To our knowledge are our results the first demonstration of the generation of droplets with a diameter of 1 μm and the encapsulation of single proteins by flow focussing.

The implementation of the whole process did not require a clean room facility. We solely used a laminar flow hood. Our self-build illuminating devices were sufficient for our requirements, expansive commercial mask aligners were not needed. The requirements for both space and money were less demanding as usually assumed and implemented. The final design for our device allows one to generate and image droplets in one, carrying out experiment in a short amount of time. This opens the doors for a wider usage of microfluidics especially in interdisciplinary research.

Appendix

8.1 Single Molecule Data

Table 8.1 – Single Molecule Data. # M is the the number of identified marker spots, # C is the number of colocalized spots and Coloc. is the colocalization percentage.

Experiment	# M	# C	Coloc.	1	2	3	4	5	6	7	8	9	>9	n.d.
GloC/SRSF1/E	434	205	47.2	27.3	29.8	7.3	7.3	5.4	3.9	1.5	4.9	1.5	1	8.8
GloC/SRSF1/A	413	135	32.7	58.5	9.6	4.4	1.5	2	0	0	0	0	1.48	22.2
GloM/SRSF1/A	212	35	16.5	57.1	25.7	11.4	0	0	0	0	0	0	0	2.9
C174C/SRSF1/A	561	225	40.1	53.8	24.9	8	4.9	0	0	0	0	0	0	1.8
GloC/SRSF1/E/PhosStop	747	265	35.5	69.4	17.7	3	1.9	0	0	0	0	0	0	7.2
Tether/E	131	90	68.7	30	20	13.3	3.3	3.3	2.2	0	0	0	0	27.8
Tether+comp. oligo	269	7	2.6	42.9	14.3	14.3	0	0	0	0	0	0	0	28.6
GloM/SRSF1/E/αU1	300	81	27	16	11.1	8.6	1.2	4.9	4.9	7.4	6.2	6.2	7.4	1.2
GloC/SRSF1/293T/A	100	86	86	8.1	12.8	14	3.5	3.5	1.2	0	0	0	0	14
GloM/SRSF1/E/αU1/PhosStop	135	17	12.6	35.3	0	5.9	0	11.8	11.8	0	5.9	0	0	11.8
Tether/A	382	60	15.7	23.3	23.3	16.7	8.3	3.3	0	0	0	0	0	15
AdCC/SRSF1/A	153	61	39.9	27.9	42.6	11.5	9.8	3.3	0	0	0	0	0	4.9
CEC/SRSF1/A	241	136	56.4	39	35.3	7.4	6.6	3.7	1.5	0	0	0	0	5.9

Table 8.1 – *Continued from previous page*

Experiment	# M	# C	Coloc.	1	2	3	4	5	6	7	8	9	>9	n.d.
AdCM/SRSF1/A	264	37	14	43.2	13.5	10.8	5.4	0	0	0	0	0	0	8.1
AdCC/SRSF1/U1A/A	137	36	26.3	47.2	36.1	0	0	0	0	0	0	0	0	11.1
AdCC/mCherry/U1A/A	479	90	18.8	52.2	25.6	11.1	0	0	0	0	0	0	0	11.1
GloCGG/SRSF1/C	396	182	46	44.5	23.1	11.5	7.7	1.6	1.6	0.5	1.1	0	0	1.6
GloM/SRSF1/E/PhosStop	656	105	16	35.2	21.9	7.6	5.7	3.8	0	0	0	0	0	4.8
GloC/SRSF1/A/PEG	871	127	14.6	59.8	18.9	11.8	1.6	0	0	0.8	0	0	0	3.9
GloC/SRSF1/E/PEG	662	138	20.8	23.2	21	11.6	6.5	5.1	3.6	2.2	1.4	0.7	0	2.2
SECN/SRSF1/A/PEG	1457	134	9.2	56	20.1	9	5.2	0	1.5	0.7	0	0	0	0
GloC/SRSF1-ΔRS/E	826	76	9.2	63.2	23.7	6.6	2.6	0	0	0	0	0	0	3.9
GloC/SRSF1-ΔRS/A	1728	115	6.7	71.3	22.6	4.3	0.9	0	0	0	0	0	0	0.9
GloC/SRSF1-ΔRRM2/E	2596	361	13.9	59.8	28	6.6	2.2	1.7	0.6	0.3	0	0	0	0.8
GloC/SRSF1-ΔRRM2/A	1495	105	7	68.6	27.6	1	1	0	0	0	0	0	0	1.9
GloM/SRSF1/A/αU1	749	268	35.8	33.2	17.9	12.3	6	6.7	3.4	3	1.5	0	0	1.9
GloCGG/SRSF1/C/PhosStop	260	133	51.2	31.6	35.3	9	1.5	3.8	3.8	0.8	0.8	0	0	1.5

Table 8.1 – *Continued from previous page*

Experiment	# M	# C	Coloc.	1	2	3	4	5	6	7	8	9	>9	n.d.
GloC/SRSF1/E/αU1	669	173	25.9	31.2	7.5	5.2	4	9.8	5.8	4.6	1.2	0	0	0
GloC/SRSF1/E/αU1/PhosStop	281	151	53.7	23.8	10.6	7.3	3.3	3.3	4.6	4	2	0	0	1.3
GloC/SRSF1/5 min	361	212	58.7	51.9	24.5	13.7	2.4	0.9	0	0	0.5	0	0	1.4
GloC/SRSF1/10 s	145	87	60	44.8	19.5	17.2	4.6	2.3	0	1.1	0	0	0	1.1
GloC/SRSF1/30 min	378	218	57.7	50.9	17.4	13.3	6.9	6	1.8	0	0	0	0	1.8
GloC/2xMut/+ESE/SRSF1/A	655	214	32.7	41.1	26.6	9.3	5.1	4.7	2.8	1.9	0.5	0	0	3.3
GloC/2xMut/-ESE/SRSF1/A	688	201	29.2	54.2	23.4	7.5	4.5	4	3	1.5	0	0	0	0.5
GloC/Short/+ESE/SRSF1/A	509	200	39.3	54.5	22.5	9.5	5	2.5	2	1.5	0	0.5	0	1
GloC/Short/-ESE/SRSF1/A	664	218	32.8	61	22.5	8.3	3.2	1.4	1.4	0.5	0	0	0	0.9
GloC/2xMut/+ESE/SRSF1/E	361	112	31	29.5	14.3	6.3	4.5	8.9	6.3	1.8	0.9	0	0	0
GloC/2xMut/-ESE/SRSF1/E	293	89	30.4	30.3	14.6	4.5	0	5.6	4.5	0	0	0	0	1.1
GloC/Short/+ESE/SRSF1/E	1244	86	6.9	47.7	12.8	7	4.7	3.5	5.8	5.8	2.3	0	0	0
GloC/Short/-ESE/SRSF1/E	1088	93	8.5	46.2	21.5	4.3	7.5	5.4	0	2.2	0	0	0	3.2
GloC/10xYeast/SRSF1/E	475	118	24.8	32.2	26.3	16.9	7.6	5.1	4.2	0.8	0.8	0	0	1.7

Table 8.2 – Number of occurrences for each number of bound SRSF1-mCherry to number of bound hnRNP A1-GFP on Globin C premRNA in complex E. X denotes colocalization events where the number of bound proteins could not be assessed.

SRSF1 \ hnRNP A1	0	1	2	3	4	5	6	7	8	9	>9	X
0	0	9	2	1	0	0	0	0	0	0	0	1
1	21	3	2	0	0	0	0	0	0	0	0	1
2	15	5	2	0	1	0	0	0	0	0	0	0
3	10	4	0	1	0	1	0	0	0	0	0	0
4	1	0	1	0	0	0	0	0	0	0	0	1
5	2	2	2	0	0	2	0	0	1	0	0	0
6	1	2	3	1	1	0	0	0	0	0	0	0
7	1	4	0	0	0	0	0	0	0	0	0	0
8	0	1	0	0	1	0	0	0	0	0	0	0
9	0	0	0	0	0	1	0	0	0	0	0	0
>9	6	5	9	0	1	2	2	1	2	1	1	1
4	2	0	1	0	0	0	0	0	0	0	0	0

Table 8.3 – Number of occurrences for each number of bound SRSF1-mCHerry to number of bound hnRNP A1-GFP on Globin C premRNA in complex A. X denotes colocalization events where the number of bound proteins could not be assessed.

SRSF1 \ hnRNP A1	0	1	2	3	4	5	6	7	8	9	>9	X
0	0	4	0	0	0	1	1	0	0	0	0	0
1	13	2	2	1	1	0	0	0	0	1	0	1
2	3	1	1	0	3	1	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	1	0
4	0	0	0	0	1	0	0	0	0	0	1	0
5	0	0	0	0	0	0	1	0	0	0	1	0
6	0	0	0	0	0	1	0	1	0	0	0	0
7	0	0	0	0	0	0	0	0	1	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	1	0
>9	0	0	0	0	0	1	0	1	2	0	1	0
X	0	1	0	0	0	0	0	0	0	0	0	0

8.2 Droplet Diameters

The underlying data for the droplet diameters used for the binned tables in Section 6.8.

Table 8.4 – Droplet diameters for combinations of dispersed and continuous phase pressured. Pressures are in mbar, droplet diameters are in µm. H_2O was used as the dispersed phase. The 5 µm junction was used. The shown standard deviation is the deviation of measured droplet sizes, not of the droplets itself.

disp. / cont.	0	100	150	200	225	250	275	300	400	500	600	700	800	900	1000
75	16.1±0.1														
150		25.9±0.5	16.7±0.5												
175			21.1±0.3	16.6±0.4											
200			28.6±0.5	21.0±0.4	16.8±0.5	12.9±0.1									
225				28.0±0.6	21.1±0.4	17.6±0.4	14.3±0.5								
250					28.8±0.5	22.0±0.4	18.0±0.7								
275						29.4±0.6	22.5±0.4								
300							30.5±0.3	12.6±0.7							
325								17.2±0.3							
350								21.4±0.6							
375								25.1±0.4	9.6±0.6						
400								32.3±0.6	16.2±0.4						
425									19.5±0.3						
450									23.3±0.7	8.7±0.1					
475									26.6±0.4	14.9±0.5					
500									33.8±0.9	18.2±0.5					
525										21.1±0.6	6.8±0.2				
550										23.7±0.6	13.4±0.4				
575										27.7±0.5					
600										34.4±0.6	19.4±0.5	6.0±0.1			
625												13.3±0.3			
650											24.0±0.6	15.2±0.4			
700											31.4±0.3	19.6±0.5	8.4±0.3		
725													14.2±0.3		
750												23.5±0.5	16.6±0.4		
775														6.4±0.2	
800												29.9±0.7	19.6±0.5	9.4±0.3	
850													23.4±0.8	16.9±0.7	
900													31.4±0.1	19.4±0.7	
925													45.0±1.0		
950															21.2±0.6
1000															28.2±0.7

Table 8.5 – Droplet diameters for combinations of dispersed and continuous phase pressured. Pressures are in mbar, droplet diameters are in µm. 0.5 % NP-40 was used as the dispersed phase. The 5 µm junction was used. The shown standard deviation is the deviation of measured droplet sizes, not of the droplets itself.

Oil / Water	150	200	300	350	400	500	600	700	800	900	1000
110	15.8±0.1										
130	19.6±0.5										
140	25.2±0.7										
150		15.0±0.7									
175		19.3±0.5									
200		22.4±3.2									
210			11.0±0.1								
225			14.6±0.4								
250			17.9±0.2	12.4±0.1							
275			23.2±0.3	14.8±0.4							
300				19.0±0.3	12.9±0.1						
325				20.4±0.4	16.6±0.4						
350				20.6±0.2	20.3±0.6						
375					24.0±0.5	12.6±0.1					
400					32.3±0.1	15.9±0.3					
425						18.5±0.5					
450						21.2±0.6					
475						25.4±0.1	15.5±0.3				
500							16.8±0.8				
525							19.0±0.7				
550							21.9±0.4				
575							25.7±0.1				
600								17.9±0.4			
650								22.7±0.1	16.0±0.1		
700									18.3±0.1		
750									22.8±0.9	17.3±0.1	
800										18.6±0.5	15.8±0.1
900											18.7±1.0
950											24.1±0.1

Table 8.6 – Droplet diameters for combinations of dispersed and continuous phase flow rates. Flow rates are in nl/min and were binned in 100 nl/min intervals, droplet diameters are in µm. Results are shown in red for water and in blue for 0.5 % NP-40 as the dispersed phase. The 5 µm junction was used. The shown standard deviation is the deviation of measured droplet sizes, not of the droplets itself.

cont. \ disp.	100	300	400	500	600	700	800	900	1000	1100	1200	1300	1400
100	29.0±0.1	23.3±0.7 / 25.2±0.7	25.0±0.1 / 15.8±0.1	21.7±0.5 / 19.3±0.5	19.7±0.1	19.5±0.1	15.2±0.3 / 13.9±0.3	16.2±0.2 / 14.6±0.2	14.6±0.1 / 12.6±0.1		no droplets		
200		19.6±0.5	27.7±0.1	24.0±0.4 / 15.0±0.7		17.9±0.2	19.6±0.4 / 19.6±0.4		15.9±0.3	15.5±0.3			
300		34.4±0.1			23.5±0.3 / 23.2±0.3	20.4±0.4	19.4±0.1		18.5±0.5	16.8±0.8		16.0±0.1	15.8±0.1
400		31.4±0.1			23.4±0.1	24.0±0.5							
500		29.9±0.1			21.2±0.2 / 20.6±0.2			21.2±0.6	19.0±0.7				
600		31.4±2.3 / 22.4±3.2			32.3±0.1		25.4±0.1			17.9±0.4		17.3±0.1	
800									21.9±0.4		18.3±0.1		
900								25.7±0.1			18.6±0.5		
1000		only dispersed phase							22.7±0.1				
1200										22.8±0.9		18.7±1.0	
1600											24.1±0.1	burst	

Table 8.7 – Droplet diameters for combinations of dispersed and continuous phase pressures. Pressures are in mbar, droplet diameters are in µm. Results are shown in red for water and in blue for 0.5 % NP-40 as the dispersed phase. The 5 µm junction was used. The shown standard deviation is the deviation of measured droplet sizes, not of the droplets itself.

Values below are listed as *water (red) / NP-40 (blue)* where both are present.

Water \ Oil	0	200	300	400	500	600	700	800	900	1000
150		16.7±0.5 / 15.0±0.5								
175		21.1±0.3 / 19.3±0.3								
200		28.6±0.5 / 22.4±0.6								
225			14.3±0.5 / 14.6±0.5							
250			18.0±0.7 / 17.9±0.7							
275			22.5±0.4 / 23.2±0.4							
300		30.5±0.3	12.6±0.7 / 12.9±0.7							
325			17.2±0.3 / 16.6±0.4							
350			21.4±0.6 / 20.3±0.7							
375			25.1±0.4 / 24.0±0.4	9.6±0.6 / 12.6±0.4						
400			32.3±0.6 / 32.3±0.6	16.2±0.4 / 15.9±0.4						
425				19.5±0.3 / 18.5±0.4						
450				23.3±0.7 / 21.2±0.7	8.7±0.1					
475				26.6±0.4 / 25.4±0.4	14.9±0.5 / 15.5±0.5					
500				33.8±0.9	18.2±0.5 / 16.8±0.5					
525					21.1±0.6 / 19.0±0.7	6.8±0.2				
550					23.7±0.6 / 21.9±0.7	13.4±0.4				
575					27.7±0.5 / 25.7±0.6					
600					34.4±0.6	19.4±0.5 / 17.9±0.6	6.0±0.1			
625							13.3±0.3			
650						24.0±0.6 / 22.7±0.6	15.2±0.4 / 16.0±0.4			
700						31.4±0.3	19.6±0.5 / 18.3±0.5	8.4±0.3		
725								14.2±0.3		
750							23.5±0.5 / 22.8±0.5	16.6±0.4 / 17.3±0.4		
775										6.4±0.2
800							29.9±0.7	19.6±0.5 / 18.6±0.5	9.4±0.3 / 15.8±0.2	
850								23.4±0.8	16.9±0.7	
900								31.4±0.1	19.4±0.7 / 18.7±0.7	
925								45.0±1.0		
950									21.2±0.6 / 24.1±0.5	
1000									28.2±0.7	

8.3 Sequences of Oligonucleotides

Table 8.8 – Sequences of the used oligonucleotides. 2'O Methyl modified nucleotides are shown in bold, LNA modified nucleotides are shown non-bold, Cy5 denotes a Cyanine-5-Dye and biotin denotes a biotin group.

Name	Sequence
β-globin 5' Cy5	Cy5 - **UAG** ACA **ACC** AGC **AGC CC** - biotin
Ad Cy5	Cy5 - **ACC UGC AGG CAU GCA** - biotin
α-U1	**GCC AGG UAA GUA U** - biotin
α-U6	**CUG UGU AUC GUU CCA AUU UU**

8.4 Sequence of SRSF1

```
         10         20         30         40         50         60
MSGGGVIRGP AGNNDCRIYV GNLPPDIRTK DIEDVFYKYG AIRDIDLKNR RGGPPPFAFVE

         70         80         90        100        110        120
FEDPRDAEDA VYGRDGYDYD GYRLRVEFPR SGRGTGRGGG GGGGGGAPRG RYGPPSRRSE

        130        140        150        160        170        180
NRVVVSGLPP SGSWQDLKDH MREAGDVCYA DVYRDGTGVV EFVRKEDMTY AVRKLDNTKF

        190        200        210        220        230        240
RSHEGETAYI RVKVDGPRSP SYGRSRSRSR SRSRSRSRSN SRSRSYSPRR SRGSPRYSPR


HSRSRSRT
```

The RRMs 1 and 2 are underlined in red and cyan respectively, the RS domain is underlined in cyan, as from Ge et al. (1991) and Zuo and Manley (1993). SRSF1-ΔRS does not contain the RS domain and SRSF1-ΔRRM2 does not contain the RRM2 domain. Both versions were made by B. Hutchinson.

8.5 Sequences of Pre-mRNA Constructs

Exons are shown in cyan (and yellow for alternative splicing), introns are shown in red. Constructs with mutated splice sites show the same exons as if those were consensus splice sites for comparison. The annealing region for the labelling oligos are underlined (Table 8.8).

The constructs Globin C, Globin M, Globin C GG, C174C and CEC were made by Dr. M.J. Hodson. Ad1 CC and Ad1 CM were made by Dr. L. O'Mullane. Short Globin ±ESE and 2xMut Globin ±ESE were made by Mr. B. Song.

8.5.1 Globin C

Globin-based construct with one consensus 5′ splice site.

```
  1 GGGCTGCTGG TTGTCTACCC ATGGACCCAG AGGTTCTTCG AGTCCTTTGG
 51 GGACCTGTCC TCTGCAAATG CTGTTATGAA CAATCCTAAG GTGAAGGCTC
101 ATGGCAAGAA GGTGCTGGCT GCCTTCAGTG AGGGTCTGAG TCACCTGGAC
151 AACCTCAAAG GCACCTTTGC TAAGCTGAGT GAACTGCACT GTGACAAGCT
201 GCACGTGGAT CCTGAGAACT TCAGGTAAGT TTGGGGACCC TTGATTGTTC
251 TTTCTTTTTC GCTATTGTAA AAATTCATGT TATATGGTCG ACTCTGCTAA
301 CCATGTTCAT GCCTTCTTCT TTTTCCTACA GCTCCTGGGC AACGTGCTGG
351 TTATTGTGCT GTCTCATCAT TTTGG
```

8.5.2 Globin M

The consensus 5′ splice site of Globin C was changed to a non-splice site sequence.

```
  1 GGGCTGCTGG TTGTCTACCC ATGGACCCAG AGGTTCTTCG AGTCCTTTGG
 51 GGACCTGTCC TCTGCAAATG CTGTTATGAA CAATCCTAAG GTGAAGGCTC
101 ATGGCAAGAA GGTGCTGGCT GCCTTCAGTG AGGGTCTGAG TCACCTGGAC
151 AACCTCAAAG GCACCTTTGC TAAGCTGAGT GAACTGCACT GTGACAAGCT
```

```
201  GCACGTGGAT  CCTGAGAACT  TCAGGATTGG  GGACCCTTGA  TTGTTCTTTC
251  TTTTTCGCTA  TTGTAAAAAT  TCATGTTATA  TGGTCGACTC  TGCTAACCAT
301  GTTCATGCCT  TCTTCTTTTT  CCTACAGCTC  CTGGGCAACG  TGCTGGTTAT
351  TGTGCTGTCT  CATCATTTTG  G
```

8.5.3 Globin C GG

The 3′ splice site of Globin C was changed to a non-splice site sequence.

```
  1  GGGCTGCTGG  TTGTCTACCC  ATGGACCCAG  AGGTTCTTCG  AGTCCTTTGG
 51  GGACCTGTCC  TCTGCAAATG  CTGTTATGAA  CAATCCTAAG  GTGAAGGCTC
101  ATGGCAAGAA  GGTGCTGGCT  GCCTTCAGTG  AGGGTCTGAG  TCACCTGGAC
151  AACCTCAAAG  GCACCTTTGC  TAAGCTGAGT  GAACTGCACT  GTGACAAGCT
201  GCACGTGGAT  CCTGAGAACT  TCAGGTAAGT  TTGGGGACCC  TTGATTGTTC
251  TTTCTTTTTC  GCTATTGTAA  AAATTCATGT  TATATGGTCG  ACTCTGCTAA
301  CCATGTTCAT  GCCTTCTTCT  TTTTCCTACG  GCTCCTGGGC  AACGTGCTGG
351  TTATTGTGCT  GTCTCATCAT  TTTGG
```

8.5.4 Ad1 CC

Adenovirus-based construct with two alternative consensus 5′ splice sites.

```
  1  GGGAACAAAA  GCTTGCATGC  CTGCAGGTCG  AGGGGGGGCG  CGTTCGTCCT
 51  CACTCTCTTC  CGCATCGCTG  TCTGCGAGGG  CCAGCTGTTC  AGGTAAGTAC
101  TCCCTCTCAA  AAGCGGGCAT  GACTTCTGCG  CTAAGATTGT  CAGTTTCCAA
151  AAACGAGGAG  GATTTGATAT  TCACCAGCTG  TTCAGGTAAG  TCCTTTGAGG
201  GTGGCCGCGT  CCATCTGGTC  AGAAAAGACA  ATCTTTTTGT  TGTCAAGCTT
251  GCTGCACGTC  TAGGGCGCAG  TAGTCCAGGG  TTTCCTTGAT  GATGTCATAC
301  TTATCCTGTC  CCTTTTTTTT  CCACAGCTCG  CGGTTGAGGA  CAAACTCTTC
351  GCGGTCTTTC  CAGTACTCTT  GGATC
```

241

8.5.5 Ad1 CM

The second consensus 5′ splice site of Ad1 CC was changed to a non-splice site sequence.

```
  1  GGGAACAAAA  GCTTGCATGC  CTGCAGGTCG  AGGGGGGGCG  CGTTCGTCCT
 51  CACTCTCTTC  CGCATCGCTG  TCTGCGAGGG  CCAGCTGTTC  AGGTAAGTAC
101  TCCCTCTCAA  AAGCGGGCAT  GACTTCTGCG  CTAAGATTGT  CAGTTTCCAA
151  AAACGAGGAG  GATTTGATAT  TCACCAGCTG  TTGGCGAATT  CCCTTTGAGG
201  GTGGCCGCGT  CCATCTGGTC  AGAAAAGACA  ATCTTTTTGT  TGTCAAGCTT
251  GCTGCACGTC  TAGGGCGCAG  TAGTCCAGGG  TTTCCTTGAT  GATGTCATAC
301  TTATCCTGTC  CCTTTTTTTT  CCACAGCTCG  CGGTTGAGGA  CAAACTCTTC
351  GCGGTCTTTC  CAGTACTCTT  GGATC
```

8.5.6 C174C

Two alternative 5′ splice sites are separated by 174 nt.

```
  1  GGGCTGCTGG  TTGTCTACCC  ATGGACCCAG  AGGTTCTTCG  AGTCCTTTGG
 51  GGACCTGTCC  TCTGCAAATG  CTGTTATGAA  CAATCCTAAG  GTGAAGGCTC
101  ATGGCAAGAA  GGTGCTGGCT  GCCTTCAGTG  AGGGTCTGAG  TCACCTGGAC
151  AACCTCAAAG  GCACCTTTGC  TAAGCTGAGT  GAACTGCACT  GTGACAAGCT
201  GCACGTGGAT  CCCCAGGTAA  GTCGATCTTC  GGTACACCAC  GTTTTCAGCA
251  TGAGAAATTG  CCTCCGTGGG  CCAGGTGGTA  AGCACTTCCT  GCTCAATCTG
301  TTTAACGGCT  TCATTCATGG  ACGGCATCTG  CGTCTTCTGC  GGATGGTTAT
351  GCCACAGGGA  CATCGCCACC  AGAAACGCGC  CGAACTTCAG  GTAAGTTTGG
401  GGACCCTTGA  TTGTTCTTTC  TTTTTCGCTA  TTGTAAAATT  CATGTTATAT
451  GGAGGGGCAA  AGTTTGCTAA  CCATGTTCAT  GCCTTCTTCT  TTTTCCTACA
501  GCTCCTGGGC  AACGTGCTGG  AATTCCAGCA  CG
```

8.5.7 CEC

The exon-intron-exon unit of Globin C is duplicated, resulting in a total of three exons and two 5′ splice sites.

```
  1 GGGCTGCTGG TTGTCTAGGG CTGCTGGTTG TCTACCCATG GACCCAGAGG
 51 TTCTTCGAGT CCTTTGGGGA CCTGTCCTCT GCAAATGCAG TTATGAACAA
101 TCCTAAGGTG AAGGCTCATG GCAAGAAGGT GCTGGCTGCC TTCAGTGAGG
151 GTCTGAGTCA CCTGGACAAC CTCAAAGGCA CCTTTGCTAA GCTGAGTGAA
201 CTGCACTGTG ACAAGCTGCA CGTGGATCCT GAGAACTTCA GGTAAGTTTG
251 GGGACCCTTG ATTGTTCTTT CTTTTTCGCT ATTGTAAAAA TTCATGTTAT
301 ATGGTCGACT CTGCTAACCA TGTTCATGCC TTCTTCTTTT TCCTACAGCT
351 CCTGGGCAAC GTGCTGGTTA GGCCCAGAGG TTCTTCGAGT CCTTTGGGGA
401 CCTGTCCTCT GCAAATGCTG TTATGAACAA TCCTAAGGTG AAGGCTCATG
451 GCAAGAAGGT GCTGGCTGCC TTCAGTGAGG GTCTGAGTCA CCTGGACAAC
501 CTCAAAGGCA CCTTTGCTAA GCTGAGTGAA CTGCACTGTG ACAAGCTGCA
551 CGTGGATCCT GAGAACTTCA GGTAAGTTTG GGGACCCTTG ATTGTTCTTT
601 CTTTTTCGCT ATTGTAAAAA TTCATGTTAT ATGGTCGACT CTGCTAACCA
651 TGTTCATGCC TTCTTCTTTT TCCTACAGCT CCTGGGCAAC GTGCTGGTTA
701 TTGTGC
```

8.5.8 Short Globin -ESE

Shortened Exon 2.

```
  1 GGGCTGCTGG TTGTCTACTC AAAGGCACCT TTGCTAAGCT GAGTGAACTG
 51 CACTGTGACA AGCTGCACGT GGATCCTGAG AACTTCAGGT AAGTTTGGGG
101 ACCCTTGATT GTTCTTTCTT TTTCGCTATT GTAAAAATTC ATGTTATATG
151 GTCGACTCTG CTAACCATGT TCATGCCTTC TTCTTTTTCC TACAGCTCCT
201 GGGCAACGTG CTGGTTATTG TGCTGTCTCA TCATTTTGG
```

8.5.9 Short Globin +ESE

Shortened Exon 2. ESE in orange.

```
  1  GGGCTGCTGG  TTGTCTACTC  AAAGGCACCT  TTGCTAAGCT  GAGTGAACTG
 51  CACTGTGACA  AGCTGCACGT  GGATCCTGAG  AACTTCAGGT  AAGTTTGGGG
101  ACCCTTGATT  GTTCTTTCTT  TTTCGCTATT  GTAAAAATTC  ATGTTATATG
151  GTCGACTCTG  CTAACCATGT  TCATGCCTTC  TTCTTTTTCC  TACAGCTCCT
201  GGGCAACGTG  CTGGTTATTG  TCGAAGGAGG  CGGATTTTGG  CAAGGCGGAG
251  GAAGC
```

8.5.10 2xMut Globin -ESE

One mutation at the branch point, one at the polypyrimidine tract, both highlighted in green.

```
  1  GGGCTGCTGG  TTGTCTACCC  ATGGACCCAG  AGGTTCTTCG  AGTCCTTTGG
 51  GGACCTGTCC  TCTGCAAATG  CTGTTATGAA  CAATCCTAAG  GTGAAGGCTC
101  ATGGCAAGAA  GGTGCTGGCT  GCCTTCAGTG  AGGGTCTGAG  TCACCTGGAC
151  AACCTCAAAG  GCACCTTTGC  TAAGCTGAGT  GAACTGCACT  GTGACAAGCT
201  GCACGTGGAT  CCTGAGAACT  TCAGGTAAGT  TTGGGGACCC  TTGATTGTTC
251  TTTCTTTTTC  GCTATTGTAA  AAATTCATGT  TATATGGTCG  ACTCTGCTTA
301  CCATGTTCAT  GCCTTCTTCT  AAATCCTACA  GCTCCTGGGC  AACGTGCTGG
351  TTATTGTGCT  GTCTCATCAT  TTTGG
```

8.5.11 2xMut Globin +ESE

One mutation at the branch point, one at the polypyrimidine tract, both highlighted in green. ESE in orange.

```
  1  GGGCTGCTGG  TTGTCTACCC  ATGGACCCAG  AGGTTCTTCG  AGTCCTTTGG
 51  GGACCTGTCC  TCTGCAAATG  CTGTTATGAA  CAATCCTAAG  GTGAAGGCTC
```

244

```
101  ATGGCAAGAA  GGTGCTGGCT  GCCTTCAGTG  AGGGTCTGAG  TCACCTGGAC

151  AACCTCAAAG  GCACCTTTGC  TAAGCTGAGT  GAACTGCACT  GTGACAAGCT

201  GCACGTGGAT  CCTGAGAACT  TCAGGTAAGT  TTGGGGACCC  TTGATTGTTC

251  TTTCTTTTTC  GCTATTGTAA  AAATTCATGT  TATATGGTCG  ACTCTGCTTA

301  CCATGTTCAT  GCCTTCTTCT  AAATCCTACA  GCTCCTGGGC  AACGTGCTGG

351  TTATTGTGCT  GTCTCATCAT  TTTGGCAAGG  CGGAGGAAGC
```

245

Publications in Preparation

The first publication in preparation encompasses the obtained results from the single-molecule experiments concerning SRSF1 and the supporting experiments. The last experiments concerning the phosphorylation of SRSF1 to show a definite influence of its phosphorylation on its binding are currently being done.

The second publication in preparation covers the microfluidic work, including the manufacturing of microfluidic devices with a feature size of 3 µm, the development of a suitable surfactant for biological samples, the first demonstrated generation of droplets with a diameter of 1 µm by flow focussing and the first detection of biological samples encapsulated therein on a single-molecule level. Concluding experiments are done to highlight potential biological applications.

Bibliography

Abelson, John et al. (2010). "Conformational dynamics of single pre-mRNA molecules during in vitro splicing". *Nat Struct Mol Biol* 17, pp. 504–512.

Abou-Hassan, Ali, Olivier Sandre, and Valérie Cabuil (2010). "Microfluidics in inorganic chemistry." *Angew. Chem. Int. Ed. Engl.* 49, pp. 6268–6286.

Agresti, Jeremy J et al. (2010). "Ultrahigh-throughput screening in drop-based microfluidics for directed evolution." *Proc. Natl. Acad. Sci. U. S. A.* 107, pp. 4004–4009.

Ambrose, WP et al. (1994). "Single molecule detection and photochemistry on a surface using near-field optical excitation". *Phys. Rev. Lett.* 72, pp. 160–164.

Anger, Pascal, Palash Bharadwaj, and Lukas Novotny (2006). "Enhancement and Quenching of Single-Molecule Fluorescence". *Phys. Rev. Lett.* 96, p. 113002.

Anna, Shelley L. and Hans C. Mayer (2006). "Microscale tipstreaming in a microfluidic flow focusing device". *Phys. Fluids* 18, p. 121512.

Anna, Shelley L., Nathalie Bontoux, and Howard a. Stone (2003). "Formation of dispersions using "flow focusing" in microchannels". *Appl. Phys. Lett.* 82, p. 364.

Anscombe, FJ (1973). "Graphs in statistical analysis". *Am. Stat.* 27, pp. 17–21.

Arayanarakool, R. et al. (2012). "Single-molecule enzymatic analysis in a droplet-based microfluidic system". *MicroTAS 2012, 16th Int. Conf. Miniaturized Syst. Chem. Life Sci.* Pp. 1492–1494.

Avila, Kerstin et al. (2011). "The onset of turbulence in pipe flow." *Science* 333, pp. 192–196.

Axelrod, Daniel (2001). "Total internal reflection fluorescence microscopy in cell biology". *Int. J. Intracell. Transp.* 2, pp. 767–774.

Axelrod, Daniel, San Francisco, and Nancy L Thompson (1984). "TOTAL INTERNAL REFLECTION FLUORESCENCE". *Annu. Rev. Biophys. Bioeng.* Pp. 247–268.

Barbosa-Morais, Nuno L et al. (2012). "The evolutionary landscape of alternative splicing in vertebrate species." *Science* 338, pp. 1587–93.

Baret, Jean-Christophe (2012). "Surfactants in droplet-based microfluidics". *Lab Chip* 12, pp. 422–433.

Baret, Jean-Christophe et al. (2009). "Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity." *Lab Chip* 9, pp. 1850–8.

Bates, M et al. (2007). "Multicolor super-resolution imaging with photo-switchable fluorescent probes". *Science (80-.).* 317, pp. 1749–1753.

Bell, M V et al. (1998). "Influence of intron length on alternative splicing of CD44." *Mol. Cell. Biol.* 18, pp. 5930–41.

Benesch, Reinhold (1953). "Enzymatic removal of oxygen for polarography and related methods". *Science (80-.).* 118, pp. 447–448.

Berezin, Mikhail Y and Samuel Achilefu (2010). "Fluorescence lifetime measurements and biological imaging." *Chem. Rev.* 110, pp. 2641–2684.

Berget, S M, C Moore, and P A Sharp (1977). "Spliced segments at the 5' terminus of adenovirus 2 late mRNA". *Proc. Natl. Acad. Sci.* 74, pp. 3171–3175.

Berlier, J. E. et al. (2003). "Quantitative Comparison of Long-wavelength Alexa Fluor Dyes to Cy Dyes: Fluorescence of the Dyes and Their Bioconjugates". *J. Histochem. Cytochem.* 51, pp. 1699–1712.

Bermingham, JR, KC Arden, and AK Naumova (1995). "Chromosomal localization of mouse and human genes encoding the splicing factors ASF/SF2 (SFRS1) and SC-35 (SFRS2)". *Genomics* 29, pp. 70–79.

Bessonov, Sergey et al. (2010). "Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis." *RNA* 16, pp. 2384–2403.

Bhattacharya, S. et al. (2005). "Studies on surface wettability of poly(dimethyl) siloxane (PDMS) and glass under oxygen-plasma treatment and correlation with bond strength". *J. Microelectromechanical Syst.* 14, pp. 590–597.

Bingaman, Susan, Virginia H Huxley, and Rolando E Rumbaut (2003). "Fluorescent dyes modify properties of proteins used in microvascular research." *Microcirculation* 10, pp. 221–231.

Blanchette, M and B Chabot (1999). "Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization." *EMBO J.* 18, pp. 1939–1952.

Bonen, L and J Vogel (2001). "The ins and outs of group II introns." *Trends Genet.* 17, pp. 322–331.

Brangwynne, Clifford P et al. (2009). "Germline P granules are liquid droplets that localize by controlled dissolution/condensation." *Science* 324, pp. 1729–1732.

Brouzes, Eric et al. (2009). "Droplet microfluidic technology for single-cell high-throughput screening." *Proc. Natl. Acad. Sci. U. S. A.* 106, pp. 14195–14200.

Buratti, Emanuele and FE Baralle (2004). "Influence of RNA secondary structure on the pre-mRNA splicing process". *Mol. Cell. Biol.* 24, pp. 10505–10514.

Burghardt, TP, AD Hipp, and Katalin Ajtai (2010). "Around-the-Objective Total Internal Reflection Fluorescence Microscopy". *Biophys. J.* 48, pp. 6120–6131.

Buschmann, Volker, Kenneth D Weston, and Markus Sauer (2003). "Spectroscopic study and evaluation of red-absorbing fluorescent dyes." *Bioconjug. Chem.* 14, pp. 195–204.

Cáceres, J F and a R Krainer (1993). "Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains." *EMBO J.* 12, pp. 4715–4726.

Cáceres, J F et al. (1994). "Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors." *Science* 265, pp. 1706–1709.

Cai, Long, Nir Friedman, and X Sunney Xie (2006). "Stochastic protein expression in individual cells at the single molecule level." *Nature* 440, pp. 358–362.

Cao, W, S F Jamison, and M a Garcia-Blanco (1997). "Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing in vitro." *RNA* 3, pp. 1456–1467.

Carter, Brian C, Michael Vershinin, and Steven P Gross (2008). "A comparison of step-detection methods: how well can you do?" *Biophys. J.* 94, pp. 306–319.

Casanova, Didier et al. (2007). "Counting the number of proteins coupled to single nanoparticles." *J. Am. Chem. Soc.* 129, pp. 12592–12593.

Cech, TR (1990). "Self-splicing of group I introns". *Annu. Rev. Biochem.* 59, pp. 543–568.

Chalfie, M et al. (1994). "Green fluorescent protein as a marker for gene expression". *Science (80-.).* 263, pp. 802–805.

Cheezum, M K, W F Walker, and W H Guilford (2001). "Quantitative comparison of algorithms for tracking single fluorescent particles." *Biophys. J.* 81, pp. 2378–2388.

Chen, Y et al. (1999). "The photon counting histogram in fluorescence fluctuation spectroscopy." *Biophys. J.* 77, pp. 553–567.

Cherny, Dmitry et al. (2010). "Stoichiometry of a regulatory splicing complex revealed by single-molecule analyses." *EMBO J.* 29, pp. 2161–2172.

Chew, S L et al. (1999). "Evidence for the function of an exonic splicing enhancer after the first catalytic step of pre-mRNA splicing." *Proc. Natl. Acad. Sci. U. S. A.* 96, pp. 10655–10660.

Chin, Curtis D, Vincent Linder, and Samuel K Sia (2007). "Lab-on-a-chip devices for global health: past studies and future opportunities." *Lab Chip* 7, pp. 41–57.

Cho, Suhyung et al. (2011a). "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. U. S. A.* 108, pp. 8233–8238.

Cho, Suhyung et al. (2011b). "The SRSF1 linker induces semi-conservative ESE binding by cooperating with the RRMs." *Nucleic Acids Res.* 39, pp. 9413–9421.

Choi, Kihwan et al. (2012). "Digital Microfluidics". *Annu. Rev. Anal. Chem.* 5, pp. 413–440.

Chow, Louise T, Richard E Gelinas, and Richard J Roberts (1977). "An Amazing Sequence Arrangement of Adenovirus RNA at the 5' Ends". *Cell* 12, pp. 1–8.

Clausell-Tormos, Jenifer, Andrew D Griffiths, and Christoph a Merten (2010). "An automated two-phase microfluidic system for kinetic analyses and the screening of compound libraries." *Lab Chip* 10, pp. 1302–1307.

Cléry, Antoine et al. (2013). "Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition." *Proc. Natl. Acad. Sci. U. S. A.* 110, E2802–E2811.

Côté, J, M J Simard, and B Chabot (1999). "An element in the 5' common exon of the NCAM alternative splicing unit interacts with SR proteins and modulates 5' splice site selection." *Nucleic Acids Res.* 27, pp. 2529–2937.

Courtois, Fabienne et al. (2008). "An integrated device for monitoring time-dependent in vitro expression from single genes in picolitre droplets." *Chembiochem* 9, pp. 439–446.

Crawford, Daniel J et al. (2008). "Visualizing the splicing of single pre-mRNA molecules in whole cell extract." *RNA* 14, pp. 170–179.

Crawford, Daniel J et al. (2013). "Single-molecule colocalization FRET evidence that spliceosome activation precedes stable approach of 5' splice site and branch site." *Proc. Natl. Acad. Sci. U. S. A.* 110, pp. 6783–6788.

Créton, Robbert and Lionel F Jaffe (2001). "Review Chemiluminescence Microscopy as a Tool in Biomedical Research". *Biotechniques* 31, pp. 1098–1105.

Crispino, J D and P a Sharp (1995). "A U6 snRNA:pre-mRNA interaction can be rate-limiting for U1-independent splicing." *Genes Dev.* 9, pp. 2314–2323.

Crispino, J D, B J Blencowe, and P a Sharp (1994). "Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP." *Science* 265, pp. 1866–1869.

Crispino, JD et al. (1996). "Cis-acting elements distinct from the 5'splice site promote U1-independent pre-mRNA splicing." *Rna* 2, pp. 664–673.

Cullum, DC (1994). *Introduction to Surfactant Analysis*. 1994th ed. New York: Springer. Chap. 2.

Daw, Rosamund and Joshua Finkelstein (2006). "Lab on a chip". *Nature* 442, pp. 367–367.

Day, Richard N and Michael W Davidson (2009). "The fluorescent protein palette: tools for cellular imaging." *Chem. Soc. Rev.* 38, pp. 2887–2921.

DeKock, Roger L. and Harry B. Gray (1991). *Chemical Structure and Bonding*. Second. Sausalito: University Science Books.

Dempsey, Graham T, Wenqin Wang, and Xiaowei Zhuang (2009a). *Handbook of Single-Molecule Biophysics*. Vol. 478, pp. 95–127.

Dempsey, Graham T et al. (2009b). "Photoswitching mechanism of cyanine dyes." *J. Am. Chem. Soc.* 131, pp. 18192–18193.

Deniz, Ashok a, Samrat Mukhopadhyay, and Edward a Lemke (2008). "Single-molecule biophysics: at the interface of biology, physics and chemistry." *J. R. Soc. Interface* 5, pp. 15–45.

Denk, Winfried, James H Strickler, and Watt W Webb (1990). "Two-Photon Laser Scanning Fluorescence Microscopy". *Science (80-.).* 248, pp. 73–76.

Dichtel, William R et al. (2004). "Singlet oxygen generation via two-photon excited FRET." *J. Am. Chem. Soc.* 126, pp. 5380–5381.

Dittrich, Petra S, Michael Jahnz, and Petra Schwille (2005). "A new embedded process for compartmentalized cell-free protein expression and on-line detection in microfluidic devices." *Chembiochem* 6, pp. 811–814.

Dönmez, G, K Hartmuth, and R Lührmann (2004). "Modified nucleotides at the 5' end of human U2 snRNA are required for spliceosomal E-complex formation". *RNA* 10, pp. 1925–1933.

Dreumont, Natacha et al. (2010). "Antagonistic factors control the unproductive splicing of SC35 terminal intron." *Nucleic Acids Res.* 38, pp. 1353–1566.

Elangovan, Masilamani et al. (2003). "Characterization of one- and two-photon excitation fluorescence resonance energy transfer microscopy." *Methods* 29, pp. 58–73.

Elson, ElliotÂăL. (2011). "Fluorescence Correlation Spectroscopy: Past, Present, Future". *Biophys. J.* 101, pp. 2855–2870.

Eperon, I C et al. (1993). "Pathways for selection of 5' splice sites by U1 snRNPs and SF2/ASF." *EMBO J.* 12, pp. 3607–3617.

Eperon, I C et al. (2000). "Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1." *Mol. Cell. Biol.* 20, pp. 8303–8318.

Eperon, L P, J P Estibeiro, and I C Eperon (1986). "The role of nucleotide sequences in splice site selection in eukaryotic pre-messenger RNA". *Nature* 324, pp. 280–282.

Eperon, L P et al. (1988). "Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase?" *Cell* 54, pp. 393–401.

Erkelenz, Steffen et al. (2013). "Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms." *RNA* 19, pp. 96–102.

Fei, Y et al. (2011). "Fluorescent labeling agents change binding profiles of glycan-binding proteins". *Mol. Biosyst.* 7, pp. 3343–3352.

Förch, Patrik et al. (2002). "The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites." *EMBO J.* 21, pp. 6882–6892.

Forkey, Joseph N et al. (2003). "Three-dimensional structural dynamics of myosin V by single-molecule fluorescence polarization." *Nature* 422, pp. 399–404.

Friedman, Larry J and Jeff Gelles (2012). "Mechanism of transcription initiation at an activator-dependent promoter defined by single-molecule observation." *Cell* 148, pp. 6796–89.

Friend, James and Leslie Y. Yeo (2011). "Microscale acoustofluidics: Microfluidics driven via acoustics and ultrasonics". *Rev. Mod. Phys.* 83, pp. 647–704.

Fuerstman, Michael J, Piotr Garstecki, and George M Whitesides (2007). "Coding/decoding and reversibility of droplet trains in microfluidic networks." *Science* 315, pp. 828–832.

Galas, Jean-Christophe, Denis Bartolo, and Vincent Studer (2009). "Active connectors for microfluidic drops on demand". *New J. Phys.* 11, p. 075027.

Galison, Peter (1997). *Image and logic: A material culture of microphysics*. Chigago: University of Chicago Press. Chap. 5.

Ge, H and J L Manley (1990). "A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro." *Cell* 62, pp. 25–34.

Ge, H, P Zuo, and J L Manley (1991). "Primary structure of the human splicing factor ASF reveals similarities with Drosophila regulators." *Cell* 66, pp. 373–382.

Graewe, Stefanie et al. (2009). "Going live: a comparative analysis of the suitability of the RFP derivatives RedStar, mCherry and tdTomato for intravital and in vitro live imaging of Plasmodium parasites." *Biotechnol. J.* 4, pp. 895–902.

Granieri, Lucia et al. (2010). "High-throughput screening of enzymes by retroviral display using droplet-based microfluidics." *Chem. Biol.* 17, pp. 229–235.

Greener, Jesse et al. (2012). "Development and applications of a microfluidic reactor with multiple analytical probes." *Analyst* 137, pp. 444–450.

Greenleaf, William J, Michael T Woodside, and Steven M Block (2007). "High-resolution, single-molecule measurements of biomolecular motion". *Annu. Rev. Biophys. Biomol. Struct.* 36, pp. 171–190.

Guo, Mira T et al. (2012). "Droplet microfluidics for high-throughput biological assays." *Lab Chip* 12, pp. 2146–2155.

Ha, Taekjip et al. (1999). "Polarization Spectroscopy of Single Fluorescent Molecules". *J. Phys. Chem. B* 103, pp. 6839–6850.

Haeberle, Stefan and Roland Zengerle (2007). "Microfluidic platforms for lab-on-a-chip applications." *Lab Chip* 7, pp. 1094–110.

Han, TinaÂăW. et al. (2012). "Cell-free Formation of RNA Granules: Bound RNAs Identify Features and Components of Cellular Assemblies". *Cell* 149, pp. 768–779.

Hassler, Kai et al. (2005). "High count rates with total internal reflection fluorescence correlation spectroscopy." *Biophys. J.* 88, pp. L01–L03.

Haugland, Richard P (2002). *Handbook of fluorescent probes and research products*. Ninth. Molecular Probes.

Haustein, Elke and Petra Schwille (2007). "Fluorescence correlation spectroscopy: novel variations of an established technique." *Annu. Rev. Biophys. Biomol. Struct.* 36, pp. 151–169.

He, Mingyan et al. (2005). "Selective encapsulation of single cells and subcellular organelles into picoliter- and femtoliter-volume droplets." *Anal. Chem.* 77, pp. 1539–1544.

Hicks, Martin J et al. (2010). "Competing upstream 5' splice sites enhance the rate of proximal splicing." *Mol. Cell. Biol.* 30, pp. 1878–1886.

Hillborg, H, JF Ankner, and UW Gedde (2000). "Crosslinked polydimethylsiloxane exposed to oxygen plasma studied by neutron reflectometry and other surface specific techniques". *Polymer (Guildf).* 41, pp. 6851–6863.

Hodson, Mark J et al. (2012). "The transition in spliceosome assembly from complex E to complex A purges surplus U1 snRNPs from alternative splice sites." *Nucleic Acids Res.* Pp. 1–13.

Holt, Daniel J et al. (2010a). "Fluorosurfactants for microdroplets: interfacial tension analysis." *J. Colloid Interface Sci.* 350, pp. 205–211.

Holt, Daniel J., Richard J. Payne, and Chris Abell (2010b). "Synthesis of novel fluorous surfactants for microdroplet stabilisation in fluorous oil streams". *J. Fluor. Chem.* 131, pp. 398–407.

Holtze, C et al. (2008). "Biocompatible surfactants for water-in-fluorocarbon emulsions." *Lab Chip* 8, pp. 1632–1639.

Hoskins, a. a. et al. (2011a). "Ordered and Dynamic Assembly of Single Spliceosomes". *Science (80-.).* 331, pp. 1289–1295.

Hoskins, Aaron a and Melissa J Moore (2012). "The spliceosome: a flexible, reversible macromolecular machine." *Trends Biochem. Sci.* 37, pp. 179–188.

Hoskins, Aaron a, Jeff Gelles, and Melissa J Moore (2011b). "New insights into the spliceosome by single molecule fluorescence microscopy." *Curr. Opin. Chem. Biol.* 15, pp. 864–870.

Hua, Feng et al. (2004). "Polymer Imprint Lithography with Molecular-Scale Resolution". *Nano Lett.* 4, pp. 2467–2471.

Hua, Yimin et al. (2007). "Enhancement of SMN2 exon 7 inclusion by antisense oligonucleotides targeting the exon." *PLoS Biol.* 5, e73.

Huang, Yingqun and Joan a Steitz (2005). "SRprises along a messenger's journey." *Mol. Cell* 17, pp. 613–615.

Huebner, Ansgar et al. (2008). "Microdroplets: a sea of applications?" *Lab Chip* 8, pp. 1244–1254.

Huranova, M. et al. (2010). "The differential interaction of snRNPs with pre-mRNA reveals splicing kinetics in living cells". *J. Cell Biol.* 191, pp. 75–86.

International Human Genome Sequencing Consortium (2001). "Initial sequencing and analysis of the human genome." *Nature* 409, pp. 860–921.

International Rice Genome Sequencing Project (2005). "The map-based sequence of the rice genome." *Nature* 436, pp. 793–800.

Jamison, S F et al. (1995). "U1 snRNP-ASF/SF2 interaction and 5' splice site recognition: characterization of required elements." *Nucleic Acids Res.* 23, pp. 3260–3267.

Jurica, Melissa S et al. (2002). "Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis". *Rna* 8, pp. 426–439.

Kapanidis, AN et al. (2005). "Alternating-laser excitation of single molecules". *Acc. Chem. Res* 38, pp. 523–533.

Kapuscinski, J (1995). "DAPI: a DNA-specific fluorescent probe." *Biotech. Histochem.* 70, pp. 220–233.

Karni, Rotem et al. (2007). "The gene encoding the splicing factor SF2/ASF is a proto-oncogene." *Nat. Struct. Mol. Biol.* 14, pp. 185–193.

Kato, Masato et al. (2012). "Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels". *Cell* 149, pp. 753–767.

Katritzky, Alan R et al. (2010). "Quantitative correlation of physical and chemical properties with chemical structure: utility for prediction." *Chem. Rev.* 110, pp. 5714–5789.

Khanafer, Khalil et al. (2009). "Effects of strain rate, mixing ratio, and stress-strain definition on the mechanical behavior of the polydimethylsiloxane (PDMS) material as related to its biological applications." *Biomed. Microdevices* 11, pp. 503–508.

Khandurina, J et al. (2000). "Integrated system for rapid PCR-based DNA analysis in microfluidic devices." *Anal. Chem.* 72, pp. 2995–3000.

Kirby, AE and AR Wheeler (2013). "Digital Microfluidics: An Emerging Sample Preparation Platform for Mass Spectrometry". *Anal. Chem.* 85, pp. 6178–6184.

Kohtz, J.D. et al. (1994). "Protein–protein interactions and 5'-splice-site recognition in mammalian mRNA precursors". *Nature* 368, pp. 119–124.

Krainer, a R, G C Conway, and D Kozak (1990a). "Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells." *Genes Dev.* 4, pp. 1158–1171.

Krainer, Adrian R, Greg C Conway, and Diane Kozak (1990b). "The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites." *Cell* 62, pp. 35–42.

Krämer, a et al. (1984). "The 5' terminus of the RNA moiety of U1 small nuclear ribonucleoprotein particles is required for the splicing of messenger RNA precursors." *Cell* 38, pp. 299–307.

Kuhn, AN, MA Van Santen, and Andreas Schwienhorst (2009). "Stalling of spliceosome assembly at distinct stages by small-molecule inhibitors of protein acetylation and deacetylation". *RNA*, pp. 153–175.

Lagus, Todd P and Jon F Edd (2013). "A review of the theory, methods and recent applications of high-throughput single-cell droplet microfluidics". *J. Phys. D. Appl. Phys.* 46, p. 114005.

Lakowicz, Joseph R. (1999). *Principles of Fluorescence Spectroscopy*. 2nd editio. Springer, p. 725.

Lamichhane, Rajan et al. (2010). "RNA looping by PTB: Evidence using FRET and NMR spectroscopy for a role in splicing repression." *Proc. Natl. Acad. Sci. U. S. A.* 107, pp. 4105–4110.

Landgraf, Dirk et al. (2012). "Segregation of molecules at cell division reveals native protein localization." *Nat. Methods* 9, pp. 480–482.

Leake, Mark C et al. (2006). "Stoichiometry and turnover in single, functioning membrane protein complexes." *Nature* 443, pp. 355–358.

Lear, A L et al. (1990). "Hierarchy for 5' splice site preference determined in vivo." *J. Mol. Biol.* 211, pp. 103–115.

Lee, Wingki, Lynn M. Walker, and Shelley L. Anna (2009). "Role of geometry and fluid properties in droplet and thread formation processes in planar flow focusing". *Phys. Fluids* 21, p. 032103.

Levene, M J et al. (2003). "Zero-mode waveguides for single-molecule analysis at high concentrations." *Science* 299, pp. 682–686.

Li, Liang and Rustem F Ismagilov (2010). "Protein Crystallization Using Microfluidic Technologies Based on Valves, Droplets, and SlipChip". *Annu. Rev. Biophys.* 39, pp. 139–158.

Lichtman, Jeff W and José-angel Conchello (2005). "Fluorescence microscopy". *Nat. Methods* 2, pp. 910–915.

Lim, Jiseok et al. (2013). "Micro-optical lens array for fluorescence detection in droplet-based microfluidics." *Lab Chip* 13, pp. 1472–1475.

Lim, L P and C B Burge (2001). "A computational analysis of sequence features involved in recognition of short introns." *Proc. Natl. Acad. Sci. U. S. A.* 98, pp. 11193–11198.

Lipman, Everett a et al. (2003). "Single-molecule measurement of protein folding kinetics." *Science* 301, pp. 1233–1235.

Lipson, Ariel, Stephen G Lipson, and Henry Lipson (1995). *Optical Physics*. Vol. 55.

Liu, H X, M Zhang, and a R Krainer (1998). "Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins." *Genes Dev.* 12, pp. 1998–2012.

Long, Jennifer C and Javier F Caceres (2009). "The SR protein family of splicing factors: master regulators of gene expression." *Biochem. J.* 417, pp. 15–27.

López-Bigas, Núria et al. (2005). "Are splicing mutations the most frequent cause of hereditary disease?" *FEBS Lett.* 579, pp. 1900–1903.

Magnan, Pierre (2003). "Detection of visible photons in CCD and CMOS: A comparative view". *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.* 504, pp. 199–212.

Malkusch, Sebastian et al. (2012). "Coordinate-based colocalization analysis of single-molecule localization microscopy data." *Histochem. Cell Biol.* 137, pp. 1–10.

Mark, Daniel et al. (2010). "Microfluidic lab-on-a-chip platforms: requirements, characteristics and applications." *Chem. Soc. Rev.* 39, pp. 1153–1182.

Martins de Araújo, Mafalda et al. (2009). "Differential 3' splice site recognition of SMN1 and SMN2 transcripts by U2AF and U2 snRNP". *RNA* 15, pp. 515–523.

Mattheyses, Alexa L, Adam D Hoppe, and Daniel Axelrod (2004). "Polarized fluorescence resonance energy transfer microscopy." *Biophys. J.* 87, pp. 2787–2797.

Mayeda, Akila, DM Helfman, and AR Krainer (1993). "Modulation of Exon Skipping and Inclusion by Heterogeneous Nuclear Ribonucleoprotein Al and Pre-mRNA Splicing Factor SF2 / ASF". *Mol. Cell. Biol.* 13, pp. 2993–3001.

McManus, C Joel and Brenton R Graveley (2011). "RNA structure and the mechanisms of alternative splicing." *Curr. Opin. Genet. Dev.* 21, pp. 373–379.

Melhuish, WH (1984). "Molecular fluorescence spectroscopy". *Pure Appl. Chem.* 56, pp. 231–245.

Michaud, S and R Reed (1991). "An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway." *Genes Dev.* 5, pp. 2534–2546.

Miller, David M and Diane C Shakes (1995). "Immunofluorescence microscopy". *Methods Cell Biol.* 48, pp. 365–394.

Misteli, T, JF Cáceres, and DL Spector (1997). "The dynamics of a pre-mRNA splicing factor in living cells". *Nature* 387, pp. 523–527.

Misteli, T et al. (1998). "Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo." *J. Cell Biol.* 143, pp. 297–307.

Moerner, WE (2002). "A dozen years of single-molecule spectroscopy in physics, chemistry, and biophysics". *J. Phys. Chem. B*, pp. 910–927.

Monici, Monica (2005). "Cell and tissue autofluorescence research and diagnostic applications." *Biotechnol. Annu. Rev.* 11, pp. 227–256.

Mujumdar, R B et al. (1993). "Cyanine dye labeling reagents: sulfoindocyanine succinimidyl esters." *Bioconjug. Chem.* 4, pp. 105–111.

Muro, AF, Massimo Caputi, and Rajalakshmi Pariyarath (1999). "Regulation of fibronectin EDA exon alternative splicing: possible role of RNA secondary structure for enhancer display". *Mol. Cell. Biol.* 19, pp. 2657–2671.

Myers, Drew (2005). *Surfactant Science and Technology*. Third. Hoboken: Wiley. Chap. 2.

Myong, Sua et al. (2007). "Spring-loaded mechanism of DNA unwinding by hepatitis C virus NS3 helicase." *Science* 317, pp. 513–516.

Nelson, K K and M R Green (1990). "Mechanism for cryptic splice site activation during pre-mRNA splicing." *Proc. Natl. Acad. Sci. U. S. A.* 87, pp. 6253–6257.

Ngo, Jacky Chi Ki et al. (2005). "Interplay between SRPK and Clk/Sty kinases in phosphorylation of the splicing factor ASF/SF2 is regulated by a docking motif in ASF/SF2." *Mol. Cell* 20, pp. 77–89.

Ngo, Jacky Chi Ki et al. (2008). "A sliding docking interaction is essential for sequential and processive phosphorylation of an SR protein by SRPK1." *Mol. Cell* 29, pp. 563–576.

Nilsen, Timothy W (2003). "The spliceosome: the most complex macromolecular machine in the cell?" *Bioessays* 25, pp. 1147–1149.

Nisisako, Takasi, Toru Torii, and Toshiro Higuchi (2002). "Droplet formation in a microchannel network." *Lab Chip* 2, pp. 24–26.

Nogués, Guadalupe, Manuel J Muñoz, and Alberto R Kornblihtt (2003). "Influence of polymerase II processivity on alternative splicing depends on splice site strength." *J. Biol. Chem.* 278, pp. 52166–52171.

O Ruanaidh, Joseph J.K. and William J. Fitzgerald (1996). *Numerical Bayesian Methods Applied to Signal Processing*. 1996th ed. New York: Springer. Chap. 5.

Olofsson, Linnea and Emmanuel Margeat (2013). "Pulsed interleaved excitation fluorescence spectroscopy with a supercontinuum source." *Opt. Express* 21, pp. 3370–3378.

Orabona, Emanuele (2011). "Microfluidics assisted platforms for biotechnological applications". PhD thesis. Università di Napoli Federico II, p. 44.

Owen, Nicholas et al. (2011). "Design principles for bifunctional targeted oligonucleotide enhancers of splicing." *Nucleic Acids Res.* 39, pp. 7194–7208.

Pan, Qun et al. (2008). "Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing." *Nat. Genet.* 40, pp. 1413–1415.

Pandit, Shatakshi et al. (2013). "Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing". *Mol. Cell* 50, pp. 223–235.

Patel, Abhijit a and Joan a Steitz (2003). "Splicing double: insights from the second spliceosome." *Nat. Rev. Mol. Cell Biol.* 4, pp. 960–970.

Patil, P V and D P Ballou (2000). "The use of protocatechuate dioxygenase for maintaining anaerobic conditions in biochemical experiments." *Anal. Biochem.* 286, pp. 187–192.

Pelisch, Federico et al. (2010). "The serine/arginine-rich protein SF2/ASF regulates protein sumoylation." *Proc. Natl. Acad. Sci. U. S. A.* 107, pp. 16119–16124.

Pertea, Mihaela and Steven L Salzberg (2010). "Between a chicken and a grape: estimating the number of human genes." *Genome Biol.* 11, p. 206.

Pertsinidis, Alexandros, Yunxiang Zhang, and Steven Chu (2010). "Subnanometre single-molecule localization, registration and distance measurements." *Nature* 466, pp. 647–651.

Piruska, Aigars et al. (2005). "The autofluorescence of plastic materials and chips measured under laser irradiation." *Lab Chip* 5, pp. 1348–1354.

Rasnik, Ivan, SA McKinney, and Taekjip Ha (2006). "Nonblinking and long-lasting single-molecule fluorescence imaging". *Nat. Methods* 3, pp. 891–893.

Ray, Debashish et al. (2013). "A compendium of RNA-binding motifs for decoding gene regulation." *Nature* 499, pp. 172–177.

Reed, R and T Maniatis (1986). "A role for exon sequences and splice-site proximity in splice-site selection." *Cell* 46, pp. 681–690.

Reed, Robin and Tom Maniatis (1988). "The role of the mammalian branchpoint sequence in pre-mRNA splicing." *Genes Dev.* 2, pp. 1268–1276.

Reiner, J. E. et al. (2006). "Optically trapped aqueous droplets for single molecule studies". *Appl. Phys. Lett.* 89, p. 013904.

Renn, Alois, Johannes Seelig, and Vahid Sandoghdar (2006). "Oxygen-dependent photochemistry of fluorescent dyes studied at the single molecule level". *Mol. Phys.* 104, pp. 409–414.

Resch-Genger, U and Markus Grabolle (2008). "Quantum dots versus organic dyes as fluorescent labels". *Nat. Methods* 5, pp. 763–775.

Rino, José et al. (2007). "A stochastic view of spliceosome assembly and recycling in the nucleus." *PLoS Comput. Biol.* 3, pp. 2019–2031.

Rivet, Catherine et al. (2011). "Microfluidics for medical diagnostics and biosensors". *Chem. Eng. Sci.* 66, pp. 1490–1507.

Robert (1984). "How to find something hidden". *On Everything*, pp. 2011–2014.

Roca, X, R Sachidanandam, and AR Krainer (2005). "Determinants of the inherent strength of human 5′ splice sites". *Rna* 11, pp. 683–698.

Roca, Xavier, Martin Akerman, and Hans Gaus (2012). "Widespread recognition of 5′ splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides". *Genes Dev.* 26, pp. 1098–1109.

Rogers, JA et al. (1998). "Generating 90 nanometer features using near-field contact-mode photolithography with an elastomeric phase mask". *J. Vac. Sci. Technol. B* 16, pp. 59–68.

Roscigno, RF and MA Garcia-Blanco (1995). "SR proteins escort the U4/U6. U5 tri-snRNP to the spliceosome." *Rna* 1, pp. 692–706.

Roth, MB, C Murphy, and JG Gall (1990). "A monoclonal antibody that recognizes a phosphorylated epitope stains lampbrush chromosome loops and small granules in the amphibian germinal vesicle." *J. Cell Biol.* 111, pp. 2217–2223.

Roy, Rahul, Sungchul Hohng, and Taekjip Ha (2008). "A practical guide to single-molecule FRET". *Nat. Methods* 5, pp. 507–516.

Ruan, Qiaoqiao et al. (2004). "Spatial-temporal studies of membrane dynamics: scanning fluorescence correlation spectroscopy (SFCS)." *Biophys. J.* 87, pp. 1260–1267.

Ruckstuhl, Thomas and Stefan Seeger (2003). "Confocal total-internal-reflection fluorescence microscopy with a high-aperture parabolic mirror lens." *Appl. Opt.* 42, pp. 3277–3283.

Ruusuvuori, Pekka et al. (2010). "Evaluation of methods for detection of fluorescence labeled subcellular objects in microscope images." *BMC Bioinformatics* 11, p. 248.

Ryu, Gihan et al. (2011). "Highly sensitive fluorescence detection system for microfluidic lab-on-a-chip." *Lab Chip* 11, pp. 1664–1670.

Sako, Y, S Minoghchi, and T Yanagida (2000). "Single-molecule imaging of EGFR signalling on the surface of living cells." *Nat. Cell Biol.* 2, pp. 168–172.

Samiee, K T et al. (2006). "Zero mode waveguides for single-molecule spectroscopy on lipid membranes." *Biophys. J.* 90, pp. 3288–3299.

Sanford, Jeremy R et al. (2008). "Identification of nuclear and cytoplasmic mRNA targets for the shuttling protein SF2/ASF." *PLoS One* 3, e3369.

Schaerli, Yolanda et al. (2009). "Continuous-flow polymerase chain reaction of single-copy DNA in microfluidic microdroplets." *Anal. Chem.* 81, pp. 302–306.

Schmidt, T and GJ Schütz (1996). "Imaging of single molecule diffusion". *Proc. Natl. Acad. Sci. USA* 93, pp. 2926–2929.

Schneider, Thomas, Jason Kreutz, and DT Chiu (2013). "The Potential Impact of Droplet Microfluidics in Biology". *Anal. Chem.* 85, pp. 3476–3482.

Schuler, Benjamin and William a Eaton (2008). "Protein folding studied by single-molecule FRET." *Curr. Opin. Struct. Biol.* 18, pp. 16–26.

Seemann, Ralf et al. (2012). "Droplet based microfluidics." *Rep. Prog. Phys.* 75, p. 016601.

Shaner, NC, PA Steinbach, and RY Tsien (2005). "A guide to choosing fluorescent proteins". *Nat. Methods* 2, pp. 905–909.

Shapiro, MB and P Senapathy (1987). "RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression". *Nucleic Acids Res.* 15, pp. 7155–7174.

Shaw, Stephanie D et al. (2007). "Deletion of the N-terminus of SF2/ASF permits RS-domain-independent pre-mRNA splicing." *PLoS One* 2, e854.

Shen, Haihong and Michael R Green (2006). "RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans." *Genes Dev.* 20, pp. 1755–1765.

Shen, Haihong and Michael R. Green (2007). "RS domain-splicing signal interactions in splicing of U12-type and U2-type introns." *Nat. Struct. Mol. Biol.* 14, pp. 597–603.

Shepard, PJ and KJ Hertel (2009). "The SR protein family". *Genome Biol* 10, p. 242.

Sheth, Nihar et al. (2006). "Comprehensive splice-site analysis using comparative genomics." *Nucleic Acids Res.* 34, pp. 3955–3967.

Shim, Jung-uk et al. (2013). "Ultrarapid Generation of Femtoliter Microfluidic Droplets for Single-Molecule-Counting Immunoassays". *ACS Nano* 7, pp. 5955–5964.

Singh, Guramrit et al. (2012). "The Cellular EJC Interactome Reveals Higher-Order mRNP Structure and an EJC-SR Protein Nexus." *Cell* 151, pp. 1–15.

Sjöback, R, Jan Nygren, and Mikael Kubista (1995). "Absorption and fluorescence properties of fluorescein". *Spectrochim. Acta Part A* 51, pp. 7–21.

Smal, Ihor et al. (2010). "Quantitative comparison of spot detection methods in fluorescence microscopy." *IEEE Trans. Med. Imaging* 29, pp. 282–301.

Smith, Philip J et al. (2006). "An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers." *Hum. Mol. Genet.* 15, pp. 2490–2508.

Sotomayor, Marcos and Klaus Schulten (2007). "Single-molecule experiments in vitro and in silico." *Science* 316, pp. 1144–1148.

Srisa-Art, Monpichar, Andrew J DeMello, and Joshua B Edel (2007). "High-throughput DNA droplet assays using picoliter reactor volumes." *Anal. Chem.* 79, pp. 6682–6689.

Staley, J P and C Guthrie (1999). "An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p." *Mol. Cell* 3, pp. 55–64.

Steitz, Joan a et al. (2008). "Where in the cell is the minor spliceosome?" *Proc. Natl. Acad. Sci. U. S. A.* 105, pp. 8485–8486.

Stokes, GG (1852). "On the change of refrangibility of light". *Philos. Trans. R. Soc. London* 143, pp. 385–396.

Sun, Q et al. (1993). "General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer." *Genes Dev.* 7, pp. 2598–2608.

Sund, S E, J a Swanson, and D Axelrod (1999). "Cell membrane orientation visualized by polarized total internal reflection fluorescence." *Biophys. J.* 77, pp. 2266–2283.

Tacke, R and J L Manley (1995). "The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities." *EMBO J.* 14, pp. 3540–3551.

Taly, Valerie, Bernard T Kelly, and Andrew D Griffiths (2007). "Droplets as microreactors for high-throughput biology." *Chembiochem* 8, pp. 263–272.

Tang, K C et al. (2006). "Evaluation of bonding between oxygen plasma treated polydimethyl siloxane and passivated silicon". *J. Phys. Conf. Ser.* 34, pp. 155–161.

Tarn, W Y and J a Steitz (1994). "SR proteins can compensate for the loss of U1 snRNP functions in vitro." *Genes Dev.* 8, pp. 2704–2717.

Taylor, P. (1998). "Ostwald ripening in emulsions". *Adv. Colloid Interface Sci.* 75, pp. 107–163.

Teale, F W and G Weber (1957). "Ultraviolet fluorescence of the aromatic amino acids". *Biochem. J.* 65, pp. 476–482.

Teh, Shia-Yen et al. (2008). "Droplet microfluidics." *Lab Chip* 8, pp. 198–220.

Tennyson, CN, HJ Klamut, and RG Worton (1995). "The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced". *Nat. Genet.* 9, pp. 184–190.

The ENCODE Project Consortium (2012). "An integrated encyclopedia of DNA elements in the human genome". *Nature* 489, pp. 57–74.

Theberge, Ashleigh B et al. (2010). "Microdroplets in microfluidics: an evolving platform for discoveries in chemistry and biology." *Angew. Chem. Int. Ed. Engl.* 49, pp. 5846–5868.

Thompson, N L, K H Pearce, and H V Hsieh (1993). "Total internal reflection fluorescence microscopy: application to substrate-supported planar membranes." *Eur. Biophys. J.* 22, pp. 367–378.

Thompson, Nancy L and Bridgett L Steele (2007). "Total internal reflection with fluorescence correlation spectroscopy." *Nat. Protoc.* 2, pp. 878–890.

Tintaru, Aura M et al. (2007). "Structural and functional analysis of RNA and TAP binding to SF2/ASF." *EMBO Rep.* 8, pp. 756–762.

Tsai, Albert et al. (2012). "Heterogeneous pathways and timing of factor departure during translation initiation." *Nature* 487, pp. 390–393.

Turunen, Janne J et al. (2013). "The significant other: splicing by the minor spliceosome." *Wiley Interdiscip. Rev. RNA* 4, pp. 61–76.

Umbanhowar, P. B., V. Prasad, and D. a. Weitz (2000). "Monodisperse Emulsion Generation via Drop Break Off in a Coflowing Stream". *Langmuir* 16, pp. 347–351.

Valcárcel, J et al. (1996). "Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA". *Science* 273, pp. 1706–1709.

Wabuyele, M B et al. (2001). "Single molecule detection of double-stranded DNA in poly(methylmethacrylate) and polycarbonate microfluidic devices." *Electrophoresis* 22, pp. 3939–3948.

Wahl, Markus C, Cindy L Will, and Reinhard Lührmann (2009). "The spliceosome: design principles of a dynamic RNP machine." *Cell* 136, pp. 701–718.

Wang, Guey-Shin and Thomas a Cooper (2007). "Splicing in disease: disruption of the splicing code and the decoding machinery." *Nat. Rev. Genet.* 8, pp. 749–761.

Wang, Mark M et al. (2005). "Microfluidic sorting of mammalian cells by optical force switching." *Nat. Biotechnol.* 23, pp. 83–87.

Wang, X et al. (2011). "Predicting sequence and structural specificities of RNA binding regions recognized by splicing factor SRSF1". *BMC Genomics* 12, S8.

Ward, Thomas, Magalie Faivre, and Howard a Stone (2010). "Drop production and tip-streaming phenomenon in a microfluidic flow-focusing device via an interfacial chemical reaction." *Langmuir* 26, pp. 9233–9239.

Warf, M Bryan and J Andrew Berglund (2010). "Role of RNA structure in regulating pre-mRNA splicing." *Trends Biochem. Sci.* 35, pp. 169–178.

Weber, G and FWJ Teale (1957). "Determination of the absolute quantum yield of fluorescent solutions". *Trans. Faraday Soc.* 53, pp. 646–655.

Whitesides, George M (2006). "The origins and the future of microfluidics." *Nature* 442, pp. 368–373.

Widengren, Jerker and Rudolf Rigler (1996). "Mechanisms of photobleaching investigated by fluorescence correlation spectroscopy". *Bioimaging* 4, pp. 149–157.

Wiederschain, G. Ya. (2011). "The Molecular Probes handbook. A guide to fluorescent probes and labeling technologies". *Biochem.* 76, pp. 1276–1276.

Woolley, AT, Dean Hadley, and Phoebe Landre (1996). "Functional integration of PCR amplification and capillary electrophoresis in a microfabricated DNA analysis device". *Anal. Chem.* 68, pp. 4081–4086.

Wu, Han et al. (2010). "A splicing-independent function of SF2/ASF in microRNA processing." *Mol. Cell* 38, pp. 67–77.

Wu, J and J L Manley (1989). "Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing." *Genes Dev.* 3, pp. 1553–1561.

Wu, J Y and T Maniatis (1993). "Specific interactions between proteins implicated in splice site selection and regulated alternative splicing." *Cell* 75, pp. 1061–1070.

Xia, Younan and GM Whitesides (1998). "Soft lithography". *Annu. Rev. Mater. Sci.* 28, pp. 153–185.

Xiang, Shengqi et al. (2013). "Phosphorylation drives a dynamic switch in serine/arginine-rich proteins." *Structure* 21, pp. 2162–2174.

Xiao, S H and J L Manley (1997). "Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing." *Genes Dev.* 11, pp. 334–344.

Xiao, Shou-hua and JL Manley (1998). "Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2". *EMBO J.* 17, pp. 6359–6367.

Xu, Jie and Daniel Attinger (2008). "Drop on demand in a microfluidic chip". *J. Micromechanics Microengineering* 18, p. 065020.

Xu, Shengqing et al. (2005). "Generation of Monodisperse Particles by Using Microfluidics: Control over Size, Shape, and Composition". *Angew. Chemie* 117, pp. 734–738.

Xue, Q, O Harriman, and M C Leake (2011). "Using bespoke fluorescence microscopy to study the soft condensed matter of living cells at the single molecule level". *J. Phys. Conf. Ser.* 286, p. 012001.

Yager, Paul et al. (2006). "Microfluidic diagnostic technologies for global public health." *Nature* 442, pp. 412–418.

Yi, Changqing et al. (2006). "Microfluidics technology for manipulation and analysis of biological cells". *Anal. Chim. Acta* 560, pp. 1–23.

Yildiz, Ahmet and Paul R Selvin (2005). "Fluorescence imaging with one nanometer accuracy: application to molecular motors." *Acc. Chem. Res.* 38, pp. 574–582.

Yildiz, Ahmet et al. (2003). "Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization." *Science* 300, pp. 2061–2065.

Yin, Huabing and Damian Marshall (2012). "Microfluidics for single cell analysis." *Curr. Opin. Biotechnol.* 23, pp. 110–119.

Zagnoni, Michele and Jonathan M Cooper (2010). "A microdroplet-based shift register." *Lab Chip* 10, pp. 3069–3073.

Zhang, Chunsun et al. (2006). "PCR microfluidic devices for DNA amplification." *Biotechnol. Adv.* 24, pp. 243–284.

Zhang, Lijian et al. (2009). "A characterization of the single-photon sensitivity of an electron multiplying charge-coupled device". *J. Phys. B At. Mol. Opt. Phys.* 42, p. 114011.

Zhang, WJ and JY Wu (1996). "Functional properties of p54, a novel SR protein active in constitutive and alternative splicing." *Mol. Cell. Biol.* 16, pp. 5400–5048.

Zheng, Bo, L Spencer Roach, and Rustem F Ismagilov (2003). "Screening of protein crystallization conditions on a microfluidic chip using nanoliter-size droplets." *J. Am. Chem. Soc.* 125, pp. 11170–1171.

Zhong, Xiang-Yang et al. (2009). "SR proteins in vertical integration of gene expression from transcription to RNA processing to translation." *Mol. Cell* 35, pp. 1–10.

Zhou, Anyu et al. (2008). "Novel splicing factor RBM25 modulates Bcl-x pre-mRNA 5' splice site selection." *Mol. Cell. Biol.* 28, pp. 5924–5936.

Zhu, J. (2000). "Pre-mRNA splicing in the absence of an SR protein RS domain". *Genes Dev.* 14, pp. 3166–3178.

Zhu, Jun, Akila Mayeda, and AR Krainer. "Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins". *Mol. Cell* 8s, pp. 1351–1361.

Zhu, Ying and Qun Fang (2013). "Analytical detection techniques for droplet microfluidics– a review." *Anal. Chim. Acta* 787, pp. 24–35.

Zhuang, Y and AM Weiner (1986). "A compensatory base change in U1 snRNA suppresses a 5′ splice site mutation". *Cell* 46, pp. 827–835.

Zuo, P and J L Manley (1993). "Functional domains of the human splicing factor ASF/SF2." *EMBO J.* 12, pp. 4727–4737.

Zuo, Ping and James L Manley (1994). "The human splicing factor ASF/SF2 can specifically recognize pre-mRNA 5′ splice sites." *Proc. Natl. Acad. Sci. U. S. A.* 91, pp. 3363–3367.

Web References

Chen, Steven (1998). `http://hypertextbook.com/facts/1998/StevenChen.shtml`.

Commons, Wikimedia. *Wikimedia Commons 1*. `http://commons.wikimedia.org/w/index.php?title=File:Interne_Reflexion_(Schema).svg&oldid=43499083`.

– *Wikimedia Commons 2*. `http://commons.wikimedia.org/w/index.php?title=File:Tirfm.svg&oldid=95657766`.

Davidson, Michael W. (2009). *Basic Concepts in Fluorescence*. `http://micro.magnet.fsu.edu/primer/techniques/fluorescence/fluorescenceintro.html`.

Fluigent FASTAB. `http://www.fluigent.com/section/microfluidic-flow-control-technology/microfluidic-presentation-fastab/`.

MA/BA Product Brochure. `http://www.suss.com/fileadmin/user_upload/brochures/BR_MABA6_2012_V1.pdf`.

Microchem. *SU-8 2000 Datasheet*. `http://www.microchem.com/pdf/SU-82000DataSheet2000_5thru2015Ver4.pdf`.

nobelprize.org. `http://www.nobelprize.org/nobel_prizes/physics/laureates/1926/`.

Photo-Tools, JD. *JD Photo-Tools Website*. `http://www.jdphoto.co.uk/page-masks.html`.

Technolgies, Life. *Life Technolgies Fluorescence SpectraViewer*. `http://www.lifetechnologies.com/uk/en/home/life-science/cell-analysis/labeling-chemistry/fluorescence-spectraviewer.html`.

UniProt. *UniProt*.

Finishing a good PhD is like leaving a good friend

freely adapted from William Feather