THE UNIVERSITY *of* EDINBURGH

THE UNIVERSITY
of EDINBURGH

Biochemical Characterisation of ZFR, a Regulator of Splicing and RNA Editing

Alexander Will

A thesis presented for the degree of
Doctor of Philosophy

Wellcome Centre for Cell Biology
School of Biological Sciences
University of Edinburgh
April 2022

Declaration

I declare that this thesis is an original report of my research, has been written by me and has not been submitted for any previous degree. The experimental work is almost entirely my own work; the collaborative contributions have been indicated clearly and acknowledged. Due references have been provided on all supporting literatures and resources.

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification.

Signed,

Alexander Will

Acknowledgements

I want to thank my supervisor, Dr Atlanta Cook, for her constant support and patience during my PhD and for her feedback in helping me become a well-rounded scientist. I would also like to thank Dr Sander Granneman, my secondary supervisor, and extend my gratitude to my thesis committee members, Dr Tony Li and Dr Julia Richardson, for constructive feedback.

I would also like to thank our collaborators from the Hogg lab, Nazmul and Bobby, for their cooperation with the ZFR project and sharing their RBNS data crucial for investigating ZFR. Also, I would like to thank Anne Sapiro from the Li group for sharing the RNAi dataset before publication that was used for further examination of RNA binding sites.

My deepest gratitude goes to the Darwin Trust of Edinburgh for funding my doctoral research and providing me with this opportunity to conduct my research in Atlanta's group.

I am grateful to the past and present members of the Cook lab for their continued support and help. I want to thank Marc Nieveler, an MSc student, who helped with the RNA end labelling experiments.

I want to thank all the Edinburgh Protein Purification Facility members, Liz, Matt and Martin, for the help, advice, and training over the last few years.

Abstract

Eukaryotes use alternative splicing of pre-mRNA as a crucial mechanism to regulate their gene expression and expand the protein isoform repertoire. The essential human protein ZFR regulates alternative splicing and A-to-I editing of mRNA since many editing sites show lower editing levels upon knockdown of the ZFR homolog, Zn72D. Zn72D seems to act as the key regulator for most RNA editing sites in fruit flies, and these editing sites might require cooperative interaction between Zn72D and ADAR at the editing site.

This work aimed to study the biological role of these proteins. Therefore, recombinant human ZFR was co-expressed with NF45 and used in an RNA Bind-n-Seq experiment (RBNS) performed by the Hogg lab to identify binding motifs. ZFR/NF45 constructs lacking zinc finger domains showed weak RNA binding activity than constructs with zinc fingers. *In vitro* verification of binding of RBNS-obtained motifs indicated that ZFR has a strong preference for dsRNA over ssRNA, suggesting that *in vivo* binding sites that determine alternative splicing sites are likely to contain substantial secondary structure. ZFR/NF45 binds dsRNA and RNA-DNA duplexes without sequence specificity.

To gain a better insight into the mechanistic understanding of the role of ZFR in alternative splicing, structural studies of the various RNA-protein complexes might help us find out more about its precise biological role in regulating alternative splicing and A-to-I RNA editing. Therefore, stable *in vitro* RNA-protein complexes for biochemical and structural studies were produced using a 24-mer dsRNA and several protein components: ADAR, ZFR/NF45, and Zn72D/NF45.

A-to-I RNA editing reactions were reconstituted using *in vitro* transcribed RNA and recombinant *Drosophila* ADAR and Zn72D/NF45 proteins. A poisoned primer extension and an EndoV assay were used to detect adenosine to inosine conversions in such *in vitro* RNA editing reactions. After establishing a protocol for successfully detecting *in vitro* adenosine to inosine conversions, the reason for the existence of Zn72D-dependent and Zn72D-independent RNA editing sites might be determined in the future using this assay.

Lay summary

Human cells store their genetic information on how to make proteins in a molecule called DNA. This information will be converted into an intermediate instruction before being used to make numerous proteins, the building units of cells, with different functions. The intermediate instruction is called mRNA and can be modified to increase the number of different proteins. The human protein ZFR regulates these modification processes and is essential for a cell's survival. The same version of this protein is also found in fruit flies, called Zn72D.

Since it is not fully understood what these two proteins do, this work aimed to study their biological role. To do so, both proteins were produced in bacteria to study their function outside of cells. It is known that they bind to mRNA, thereby influencing the modification process. Therefore, the aim of my work was to understand why they bind mRNA and how they affect modification processes. A central goal of this thesis was to establish an *in vitro* test to determine what these proteins do in detail.

Overall, the research presented here describes certain aspects of regulating these mRNA modification processes and forms a foundation for future studies.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

2'O MTase	Cap-specific mRNA (nucleoside-2'-O-) methyltransferase
ADAR	Adenosine deaminase acting on RNA
ADAT	Adenosine deaminases acting on tRNA
ApoB	Apolipoprotein B
APOBEC	Apolipoprotein B mRNA-editing enzyme, catalytic polypeptide
CARD	Caspase activation and recruitment domain
CBC	Cap-binding complex
CCR4	Carbon catabolite repressor protein 4
CFI	Cleavage factor I
CFII	Cleavage factor II
cGAS	Cyclic GMP-AMP synthase
CPSF	Cleavage and polyadenylation specificity factor
CstF	Cleavage stimulation factor
CTD	C-terminal domain
DCPC	Decapping by the decapping complex
DNA	Deoxyribonucleic acid
dsRBD	dsRNA binding domain
DZF	Domain associated with zinc fingers
eIF4E	Eukaryotic translation initiation factor 4E
eIF4F	Eukaryotic initiation factor 4F
EJC	Exon junction complex
EMSA	Electrophoretic mobility shift assay

gRNA	Guide RNAs
GTase	mRNA guanylyltransferase
GTP	Guanosine-5'-triphosphate
IRF	Interferon regulatory factor
LEAPER	Leveraging endogenous ADAR for programmable editing of RNA
LIC	Ligation independent cloning
lncRNAs	long non-coding RNAs
MAVS	Mitochondrial antiviral-signaling protein
MDA5	Melanoma differentiation-associated protein 5
MRB1	Mitochondrial RNA Binding 1 complex
mRNA	Messenger RNA
N7MTase	mRNA guanine-N7 methyltransferase
NES	Nuclear export signal
NF45	Nuclear factor 45
NF90	Nuclear factor 90
NGD	No-go decay
NMD	Nonsense-mediated decay
NOT	Negative on TATA-less
NSD	Non-stop decay
NTases	Nucleotidyltransferases
OAS	2'-5'-oligoadenylate synthetase
ON	Overnight
ORF	Open reading frame
PABII	Polyadenylate binding protein 2
PABPN1	Poly(A) binding protein nuclear 1
PAP	Polyadenylate polymerase
PKR	Protein kinase R
PRR	Pentatricopeptide repeat
PTBP1	Poly-pyrimidine tract binding protein 1

PTC	Premature termination codon
RBNS	RNA Bind-n-Seq
RECC	RNA editing core complex
RESTORE	Recruiting endogenous ADAR to specific transcripts for oligonucleotide-mediated RNA editing
RIG-I	Retinoic acid-inducible gene I
RLR	RIG-I like receptor
RNA	Ribonucleic acid
RT	Reverse transcriptase
SAM	S-adenosyl-L-methionine
SEC-SAXS	Size exclusion chromatography coupled small-angle X-ray scattering
Sf	Spodoptera frugiperda
SMA	Spinal muscular atrophy
snRNP	Small nuclear ribonucleoproteins
SPNR	Spermatid perinuclear RNA-binding protein
TBK1	Serine/threonine-protein kinase 1
TPase	RNA triphosphatase
tRNA	Transfer RNA
UPF	Up frameshift
YFP	Yellow fluorescent protein
ZFR	Zinc finger RNA-binding protein
Zn72D	Zinc-finger protein at 72D

1 Introduction

1.1 The fundamental basis of eukaryotic gene expression

The flow of genetic information within a biological system is postulated as following: from deoxyribonucleic acid (DNA) to ribonucleic acid (RNA) to protein. DNA is the biological macromolecule essential for all known higher forms of life as it contains the information needed to make proteins. The other vital nucleic acid, RNA, is the messenger that carries this genetic information to the ribosomes that serve as protein production facilities. This process of generating proteins from DNA is called gene expression and has two key stages: transcription and translation.

During transcription, the information from the DNA is rewritten into small, portable messenger RNAs (mRNAs). In eukaryotes these mRNAs are transported from the nucleus to the ribosomes in the cytosol, where they are used as a template to make specific proteins giving this process its name: translation.

1.2 mRNA processing

The specific focus of this thesis is not on transcription or translation but on nuclear mRNA processing in eukaryotes. Between transcription and translation pre-mRNAs undergo a complex cascade of numerous enzymatic reactions, which involve many proteins and RNA components. These reactions are co-transcriptional as they happen during transcription (Bentley, 2014).

The first step in the maturation of pre-mRNAs requires the attachment of a 7-methylguanosine cap to the 5' end (section 1.2.1). Next, excision of introns is followed by exon ligation (section 1.2.2) and the formation of a 3' end by cleavage and addition of a non-templated poly(A) tail (section 1.2.3). Additionally, mRNAs are also edited by selective deamination of adenosines and cytosines

(section 1.3).

1.2.1 Capping

The first processing step in the maturation of a eukaryotic pre-mRNA is the attachment of a 7-methylguanosine cap (five-prime cap or 5' cap or m7G) by a 5'–5' connection. mRNA capping requires three enzymatic activities: RNA triphosphatase, guanylyltransferase, and 7-methyltransferase (Shuman, 2000). The 5' cap is found at the 5' end of an mRNA transcript and consists of a guanine nucleotide linked to pre-mRNA through an unusual 5' to 5' triphosphate link. This guanosine is methylated on position 7 directly after capping by a methyltransferase (Sonenberg et al., 1998) (figure 1.1). Capping takes place co-transcriptionally in the nucleus as soon as the first 20–30 nucleotides are incorporated into the nascent transcript (Moteki et al., 2002) and can be divided into five separate reaction steps.

1. RNA triphosphatase (TPase) removes the terminal γ-phosphate from the 5' triphosphate of the distal guanosine to generate a 5' diphosphate end of the mRNA (Shatkin et al., 2000)

2. mRNA guanylyltransferase (GTase) uses a guanosine-5'-triphosphate (GTP) to form a covalent intermediate containing a 5'-phosphoguanosine

3. The GTase then transfers the 5'-phosphoguanosine to the 5' diphosphate of the mRNA, thereby forming a 5'-5' triphosphate connection between the first nucleotide of the mRNA and the m7G cap

4. mRNA guanine-N7 methyltransferase (N7MTase) adds a methyl group to the N7 amine of the guanine cap donated from an S-adenosyl-L-methionine (SAM) molecule

5. Cap-specific mRNA (nucleoside-2'-O-)-methyltransferase (2'O MTase) methylates the two adjacent ribonucleotides at their C2 position of the ribose to generate the final cap structure using methyl group-donating SAM

The cap structure protects the nascent mRNA transcript from degradation by exonucleases (Burkard et al., 2000). It also regulates the nuclear export of the mRNA and the association of factors that bind specifically to the mRNA cap structure: the cap-binding complex (CBC) in the nucleus and eukaryotic translation initiation factor 4E (eIF4E) in the cytoplasm (Ramanathan et al., 2016).

eIF4E is part of a heterotrimeric protein complex called eukaryotic initiation factor 4F (eIF4F) that promotes eukaryotic translation initiation.

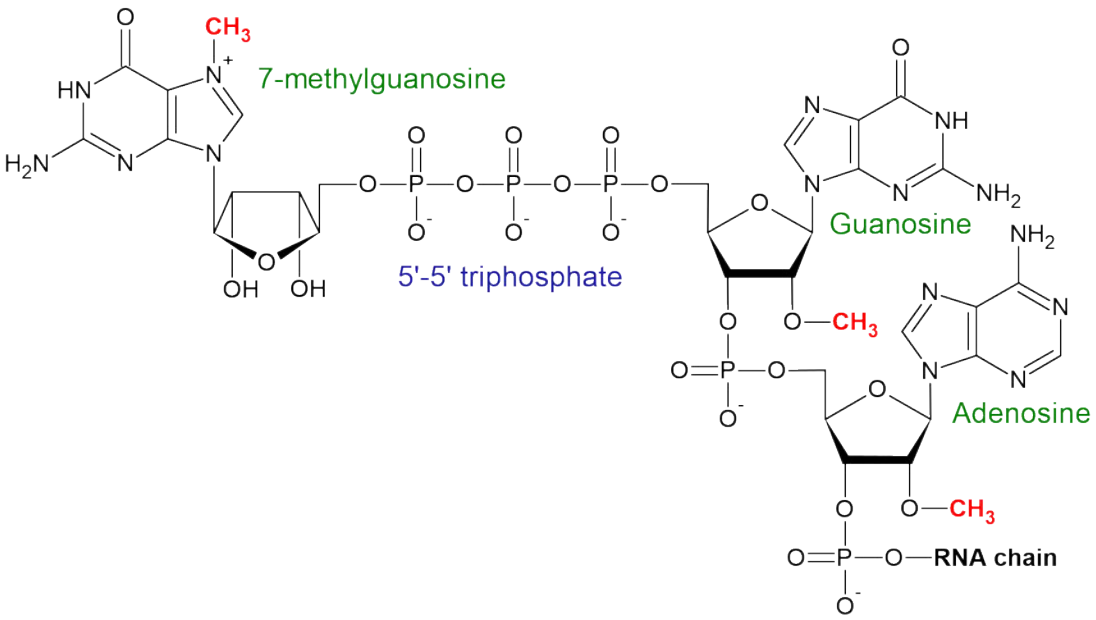


Figure 1.1: Eukaryotic mRNA cap structure. The 5' end of most eukaryotic RNAs possesses a 7-methylguanosine cap at the distal guanine mainly for stability reasons. This cap structure has a unique pattern as the first nucleotide of the RNA chain is connected to the 7-methylguanosine via a 5'-5' triphosphate bridge (blue). Furthermore, the first and second nucleotides are methylated at the free hydroxyl group of C2. All methylation sites are highlighted in red.

1.2.2 Splicing

The first evidence of introns in eukaryotes was revealed by the ground-breaking work of Berget et al., 1977, and Chow et al., 1977. Richard Roberts and Phil Sharp won a joint Nobel Prize in Physiology or Medicine in 1993 for discovering split genes in their labs. Upon hybridisation of a piece of adenovirus DNA with its corresponding mRNA after transcription, intervening DNA segments forming loops were observed under a microscope proving the discrepancy in size between template and transcript (Chow et al., 1977). These gaps in mRNA molecules, called introns (intragenic regions), are non-coding regions in human genes which means they do not carry information from DNA to proteins. Introns represent up to 90% of the entire gene sequence (E. T. Wang et al., 2008). Intronic sequences must be removed entirely before translation which is called splicing and an essential step in the maturation of pre-mRNA. Spliced introns will be degraded while the pre-mRNA remains in the nucleus. The

boundaries between exons and introns are defined by 5' splice sites (exon-intron junction) and 3' splice sites (intron-exon junction). Both splice sites exhibit conserved two nucleotide motifs: GU for 5' splice sites and AG for 3' splice sites (Horowitz, 2011). Any mutation of these splicing-defining nucleotides or additional regulatory splicing sequences might erroneously shift the reading frame of the rejoined exons, thus resulting in a truncated or dysfunctional protein (Hsu et al., 2009). 99% of splice sites use the canonical splice sites GU and AG. This canonical splice site usage is called the lariat pathway, named after the intron's shape upon splicing (Ng et al., 2004).

Splicing in eukaryotes is catalysed by a large RNA-protein complex composed of five small nuclear ribonucleoproteins (snRNPs) U1, U2, U4, U5, and U6 called the spliceosome (Will et al., 2010; Hsu et al., 2009; Anna et al., 2018). The spliceosome must undergo precise stepwise assembly by forming a series of spliceosomal complexes associated with additional protein factors, including U2AF1, U2AF2, and SF1, to catalyse the splicing of eukaryotic pre-mRNA. The spliceosome forms different complexes based on assembly and composition during the splicing process, which have been named A, pre-B, B, B^{act}, B*, C, C*, P, and ILS complexes. Advances in structural and cross-linking mass spectrometry helped to resolve the splicing mechanism. Even though some details of the mechanism were known before the structural work, cryo-EM structures offered a lot of novel information. However, not all parts of the splicing mechanism are yet known. The following enumeration is an overview of the stepwise assembled complexes and their functional mechanism of splicing (Wilkinson, Charenton et al., 2020; Shi, 2017; Zhang et al., 2019; Wilkinson, Fica et al., 2021; Wan et al., 2019; Guth et al., 2000).

1. Complex E
    a) The U1 snRNP binds to the GU sequence at the 5' splice site of an intron;
    b) Splicing factor 1 binds to the intron branch point sequence
    c) U2AF1 binds at the 3' splice site of the intron;
    d) U2AF2 binds to the polypyrimidine tract
2. Complex A (pre-spliceosome)
    a) SF1 is displaced by the U2 snRNP to form the pre-spliceosome (A complex)
3. pre-B complex

4. Complex B (pre-catalytic spliceosome)
    a) The U5/U4/U6 snRNP trimer binds
    b) The U5 snRNP binds exons at the 5' site, with U6 binding to U2
5. Complex B^{act}
6. Complex B* (catalytically activated spliceosome)
    a) The U1 snRNP is released
    b) U5 shifts from exon to intron
    c) U6 binds at the 5' splice site
7. Complex C (catalytic spliceosome)
    a) U4 is released, U6/U2 catalyzes transesterification, making the 5'-end of the intron ligate to the A on intron and form a lariat, U5 binds exon at 3' splice site, and the 5' site is cleaved, resulting in the formation of the lariat
8. Complex C* (post-spliceosomal complex)
    a) U2/U5/U6 remain bound to the lariat
    b) The 3' site is cleaved
    c) Severed exons are ligated using ATP hydrolysis
9. Complex P (post-catalytic spliceosome)
    a) The spliced mRNA is released
10. Intron lariat spliceosome (ILS)
    a) The intron lariat is released and degraded
    b) snRNPs are recycled

Contrary to eukaryotes, prokaryotes splice rarely, lacking sophisticated spliceosomal machinery. However, since there is no spatial separation of transcription and splicing from translation, prokaryotes have developed a mechanism that allows some RNA molecules to splice themselves. The joint Nobel Prize in Chemistry was awarded to Thomas Cech in 1989 for this discovery as he became the first scientist to show that RNA molecules can have a catalytic function (Kruger et al., 1982). Such catalytic RNA molecules are

called ribozymes (ribonucleic acid enzymes), and several hundreds of these ribozymes, such as Ribonuclease P or hammerhead ribozyme, have been identified in the last 40 years (Weinberg et al., 2021).

Bacterial mRNAs exclusively contain group I or group II introns (Belfort, 1990). The self-splicing introns found in *Tetrahymena thermophila* by the Cech lab are now referred to as Group I introns. Group I introns have a complex secondary structure with nine loops and catalyse their excision from mRNA, tRNA and rRNA precursors through transesterification reactions (Hausner et al., 2014). However, group II introns found in mitochondrial genes required assistance from proteins for successful *in vivo* splicing (Lambowitz et al., 2010). Since this excision mechanism shows structural similarity to eukaryotic U6/U2 snRNA and also includes the production of lariats, it has been hypothesised that pre-mRNA introns and splicing mechanisms in eukaryotes evolved from group II introns (Valadkhan, 2010).

Alternative splicing

Eukaryotes use alternative splicing of pre-mRNA as a crucial mechanism to regulate their gene expression and expand the protein isoform repertoire without having to extend their genome length drastically. It allows individual genes to produce multiple protein isoforms, which in many cases encode functionally distinct proteins (Maniatis et al., 2002; Modrek, 2001). The majority of the approximately 20000 genes encoded by the human genome have been shown to produce more than one kind of mRNA transcript through alternative splicing (Johnson et al., 2003). This discrepancy between the number of genes and transcripts is the reason for the vast size of the proteome of eukaryotic organisms and is estimated to exceed the number of genes significantly (Uhlén et al., 2015).

The different types of alternative splicing events can be classified into five main categories (figure 1.2):

1. Exon skipping: an exon can be spliced out of the pre-mRNA transcript altogether with its 5' and 3' flanking introns
2. Alternative 3' splice site: recognition of a different 3' splice site at the downstream exon
3. Alternative 5' splice site: recognition of a different 5' splice site of the upstream exon

4. Intron retention: less common, but an intron can remain in the spliced transcript. However, most intron-retaining transcripts are not exported to the cytoplasm and are frequently turned over in the nucleus

5. Mutually exclusive exons: only one of two or more mutually exclusive exons is spliced out of the pre-mRNA

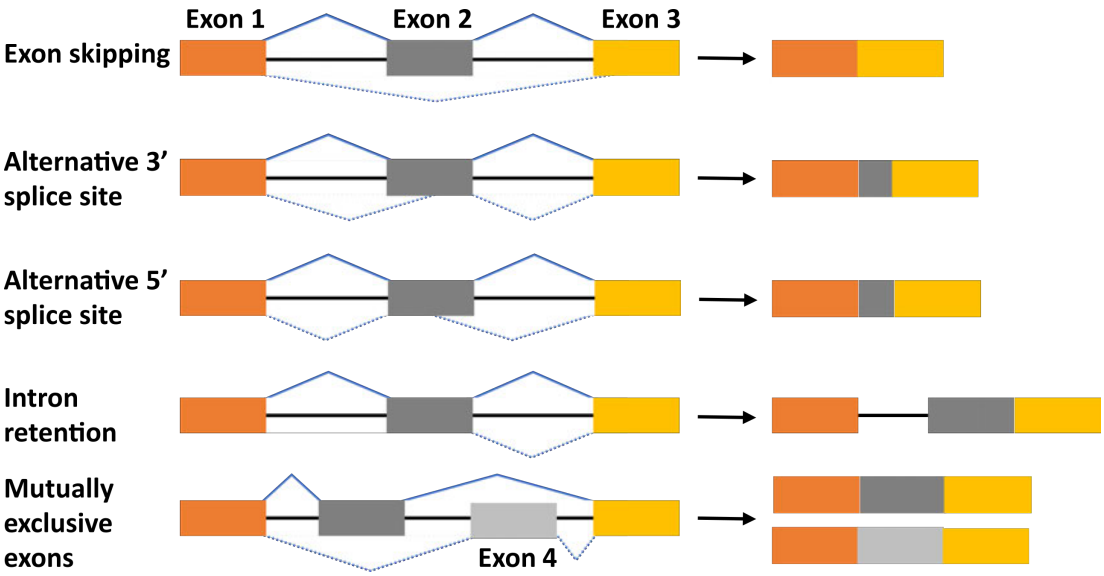


Figure 1.2: Major types of alternative splicing. Coloured boxes define exons 1, 2, 3, and 4 and horizontal lines represent introns. Solid lines above and dashed lines below the pre-mRNA represent five different alternative splicing events of the same transcript.

1.2.3 Polyadenylation

Polyadenylation is a co-transcriptional modification of eukaryotic pre-mRNAs. The 3'-end of a pre-mRNA is cleaved, and a poly(A) tail is added. The length of poly(A) (adenosine) tails vary between higher eukaryotes (about 250 nt) and *Saccharomyces cerevisiae* (*S. cerevisiae*) (around 60 nt) (Casañal et al., 2017). Almost all eukaryotic mRNAs are polyadenylated (Hunt et al., 2008) except for histone mRNAs (López et al., 2007). The poly(A) tail influences stability and protects mRNA transcripts from enzymatic degradation in the cytoplasm. Moreover, it is also required for the export of the mRNA into the cytoplasm and stimulates translation (Guhaniyogi et al., 2001). Several human diseases, including cancer and β-thalassemia, are associated with defects in the 3' end processing of pre-mRNA (Curinha et al., 2014).

The sequence AAUAAA serves as a polyadenylation signal in most human transcripts (Beaudoing et al., 2000). In addition, cleavage and polyadenylation specificity factor (CPSF) responsible for the formation of 3' ends of eukaryotic mRNAs (Stewart, 2019) recruits other proteins to form a larger complex. One of these recruited factors, cleavage stimulation factor (CstF), binds to RNA polymerase II, indicating RNA is cleaved before transcription termination is initiated (Glover-Cutter et al., 2007). Cleavage factors I (CFI) and II (CFII) then cleave pre-mRNAs at a specific site that defines the 3' end of the mature mRNA (J. Zhao et al., 1999). Hence, the activity of the endonuclease is highly regulated (Hill et al., 2019). Upon RNA cleavage, polyadenylate polymerase (PAP) starts building the poly(A) tail by adding adenosine monophosphate units from adenosine triphosphate to the RNA. Polyadenylate binding protein (PABII) binds to short poly(A) tails thereby increasing the affinity of PAP for the RNA. The length of the poly(A) tail is regulated by poly(A) binding protein nuclear 1 (PABPN1), and polyadenylation is stopped after addition of approximately 250 nucleotides to the 3' end of the mRNA. In eukaryotes, each PABPN1 protein is thought to bind 12–15 adenines (Keller et al., 2000). Once a sufficient number of PABPN1 proteins are bound to the poly(A) tail, steric clashes result in the dissociation of PAP from CPSF, thereby ending the polyadenylation process as well as mRNA processing (Eckmann et al., 2010).

1.2.4 Quality control of mRNA processing events

Due to the potentially toxic consequences of dysregulated RNA processing, cells have developed several quality control mechanisms to deal with these errors before they cause harm. For example, in spinal muscular atrophy (SMA), a motor neuron disease that is the leading genetic cause of death among infants (Tisdale et al., 2015), a gene called SMN2 is spliced incorrectly, resulting in an only partially functional protein. Therefore, premature stop codons or frameshift mutations that might have been introduced to the mature mRNA transcript by splicing errors must be recognised by RNA surveillance mechanisms. These cellular quality control mechanisms include nonsense-mediated decay (NMD) (Kurosaki et al., 2019), non-stop decay (NSD) (Doma et al., 2006), no-go decay (NGD) (Powers et al., 2020), or the rapid mammalian deaden-ylation-dependent decay pathway (Tisdale et al., 2015).

NMD is the main RNA surveillance pathway used by cells to target deleterious transcripts encoding aberrant proteins with no, or harmful functions,

thereby preventing the accumulation of such transcripts and their corresponding proteins (Pawlicka et al., 2020). However, the NMD pathway is not exclusively dedicated to mRNAs. A number of long non-coding RNAs (lncRNAs) were shown to be substrates of NMD in *Arabidopsis thaliana* (Kurihara et al., 2009) and *S. cerevisiae* (J. E. Smith et al., 2014). In eukaryotes, the activation of NMD depends on the position of premature termination codons (PTC) in mRNA. Any PTC more than 50 to 54 nucleotides upstream of the last exon-exon junction is recognised as defective and degraded through NMD (Amrani et al., 2006). Multi-protein complexes called exon junction complexes (EJCs) mark the exon-exon boundaries and assemble during splicing at a position of about 20 to 24 nucleotides upstream from the splice junction. The three main factors in NMD are RNA helicase up frameshift 1 (UPF1), UPF2 and UPF3 (Behm-Ansmant et al., 2006). If an EJC is left on the mRNA due to the transcript containing a premature stop codon, the NMD pathway is activated trough UPF1 phosphorylation by the kinase SMG1 (Hug et al., 2016). Phosphorylated UPF1 then recruits SMG5, SMG6 and SMG7, and other general mRNA degradation factors that lead to mRNA degradation (Lykke-Andersen et al., 2015). The entirety of these recruited factors is then able to initiate mRNA degradation. This is achieved through endonucleolytic cleavage through SMG6, deadenylation by the carbon catabolite repressor protein 4 (CCR4)–negative on TATA-less (NOT) deadenylase complex and decapping by the decapping complex (DCPC). Upon removal of the 5' cap of the mRNA transcript, its degradation is executed by XRN1, a 5'-3' exoribonuclease.

1.3 A-to-I RNA editing

Thirty years ago, the enzymatic activity of adenosine deaminase acting on RNA (ADAR) was discovered in frog oocytes (B. L. Bass et al., 1987; Rebagliati et al., 1987). The consequence of A-to-I RNA editing is the introduction of inosine nucleotides in the position of adenosines (Nishikura, 2010). The mechanism of A-to-I RNA editing involves the hydrolytic deamination of adenosine at the C6 site in double-stranded RNA (ADAR is not able to edit single-stranded RNA) substrates by ADAR, resulting in a conversion to inosine (figure 1.3). If the resulting inosine is located within an exon, the translational machinery will read this as guanosine due to inosine's ability to basepair with cytosine, potentially resulting in an amino acid change. For example, if the codon ACC were edited to ICC, this would be translated as a GCC codon. This modification would change the encoded amino acid from threonine to alanine.

Codon changes and consequent alterations of protein-coding sequences of selected genes diversify their protein functions. Over the years, only about 30 human genes were discovered to be subjected to RNA editing within their protein-coding sequence, as most editing events happen within non-coding elements. Non-coding regions such as Alu repeats (Levanon et al., 2005) are present in introns or 3'-UTRs of about 75% of all human genes (Kim et al., 2004). Two Alu repeats in an inverted orientation can adopt secondary structures, thus generating an RNA substrate for ADAR (Lev-Maor et al., 2007). Many of the affected genes play an essential role in neuronal function (Behm et al., 2016), and dysregulation of these editing events can result in various severe diseases (Gallo et al., 2017; Slotkin et al., 2013). A-to-I editing is, for example, essential in humans for glutamate receptor (GluR2) mRNA modification in the central nervous system to preserve glutamatergic neurotransmission (Barbon, 2003). In the case of GluR2, ADAR2 is effectively correcting for a genetic error at the level of the transcript and is the only known essential ADAR2-mediated editing event in humans (Wright et al., 2012). However, disruption of the editing pathways in *Caenorhabditis elegans* (*C. elegans*) and *D. melanogaster* results only in severe behavioural and neural defects as *C. elegans* and *D. melanogaster* that lack ADARs are still viable (Goldstein et al., 2016; H.-Q. Zhao et al., 2014).

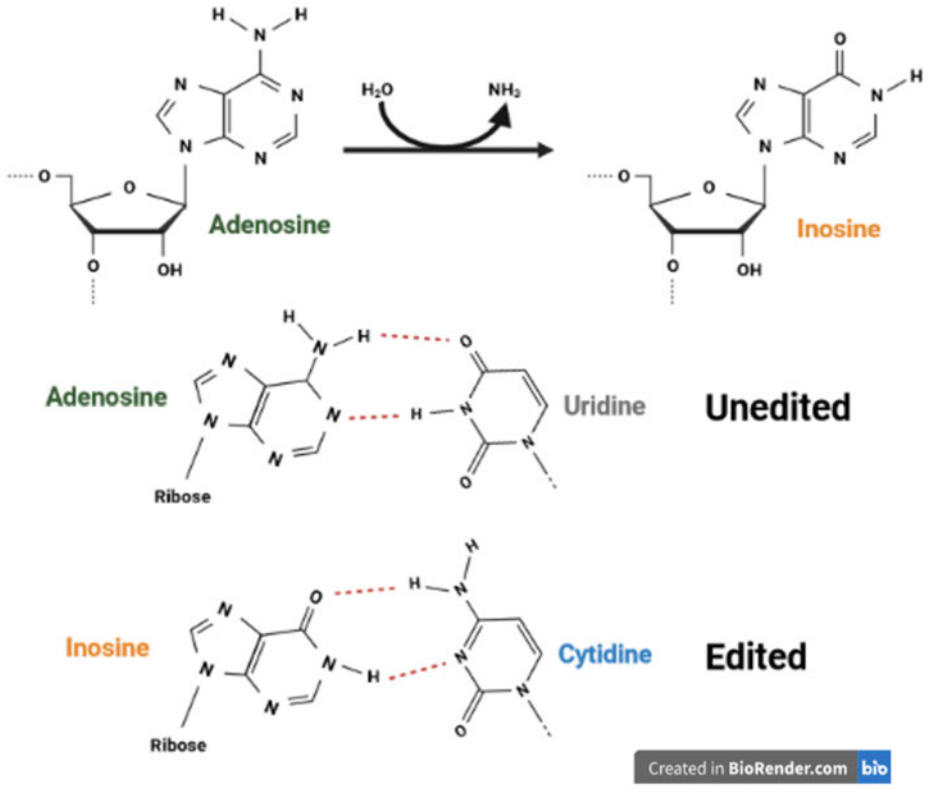


Figure 1.3: Overview adenosine deamination and its effect on base pairing. Hydrolytic deamination reaction catalysed by ADAR converts adenosine to inosine. Adenosine base pairs with uridine, whereas inosine base pairs with cytidine. Figure created with BioRender.com.

1.3.1 Members of the ADAR enzyme family

Enzymes of the ADAR family are responsible for converting adenosine to inosine in pre-mRNA by deamination while binding to double-stranded RNA. ADAR enzymes only exist in metazoans, and the family (figure 1.4) seems to be highly conserved (figure 1.5), which indicates that RNA editing is a crucial regulatory mechanism and essential for metazoan organisms (B. Bass, 1997). A common feature among all ADAR proteins is at least one N-terminal dsRNA binding domain (dsRBD) and a C-terminal deaminase domain (Yablonovitch et al., 2017). There are three ADAR proteins in mammals, ADAR1, 2 and 3. ADAR1 has two isoforms (p110 and p150), and ADAR3 is known not to be enzymatically active. However, only one protein of the ADAR enzyme family is present in *D. melanogaster* making it an ideal organism to use in knock-down studies. ADAR2 is the most likely orthologue of *D. melanogaster* ADAR as they have the closest evolutionary relationship and are likely to have an equivalent function. ADAR1 has an additional N-terminal domain that can bind both DNA

and dsRNA in a Z conformation (Herbert et al., 1997). In *C. elegans* there are two genes coding for ADAR, CeADR1 and CeADR2 (Tonkin et al., 2002), and squids express two ADAR2 orthologous isoforms due to alternative splicing (Palavicini et al., 2009).

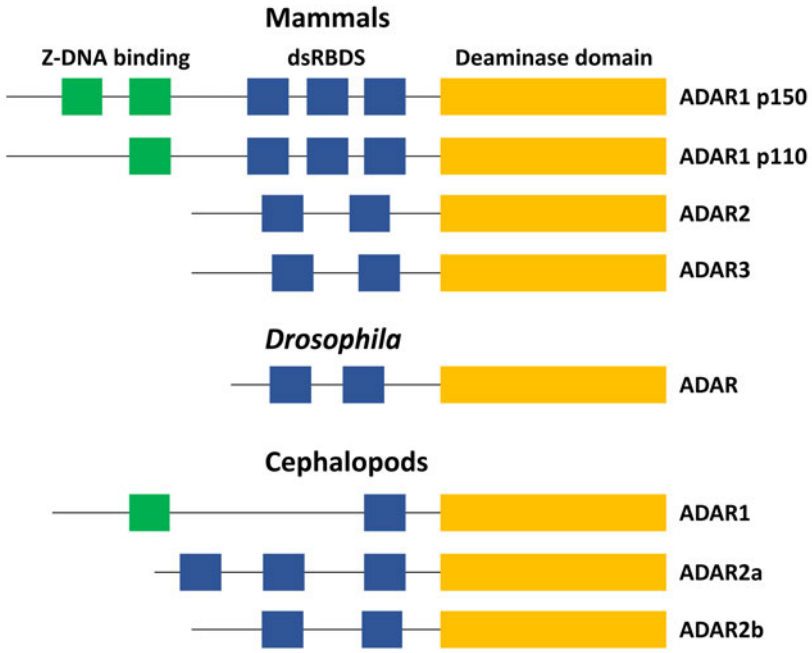


Figure 1.4: Domain architecture of ADAR proteins. Protein domains (represented as boxes) found in members of the ADAR enzyme family are: deaminase domain (yellow), dsRNA binding domain (blue) and Z-DNA binding domain (green).

Human ADAR1 is widely expressed in different tissues and plays a prominent role in the promiscuous editing of long dsRNA, preventing deleterious innate immune responses activated by MDA5 and MAVS (section 1.3.2). The ADAR1 isoforms ADAR1p110 exclusively localizes to the nucleus (Desterro et al., 2003), while ADAR1p150 shuttles between the nucleus and the cytoplasm due to its nuclear export signal (NES) (Poulsen et al., 2001).

Human ADAR2 is localised to the nucleus, lacks the Z-DNA binding domain of ADAR1 and contains only two dsRBDs. It is most highly expressed in the brain (Cenci et al., 2008) and is primarily required for site-specific editing of transcripts of the central nervous system, such as the mammalian glutamate receptor pre-mRNA.

Unlike the other two human editing enzymes, ADAR1 and ADAR2, human ADAR3 lacks catalytic residues, making it a pseudo-enzyme. However, it can suppress RNA editing of the 5-HT2C receptor *in vitro* through an unknown

mechanism (Hong et al., 2015) acting as a negative regulator. In addition, ADAR3 might have other unknown functions unrelated to A-to-I RNA editing mediated by its role as an RNA-binding protein.

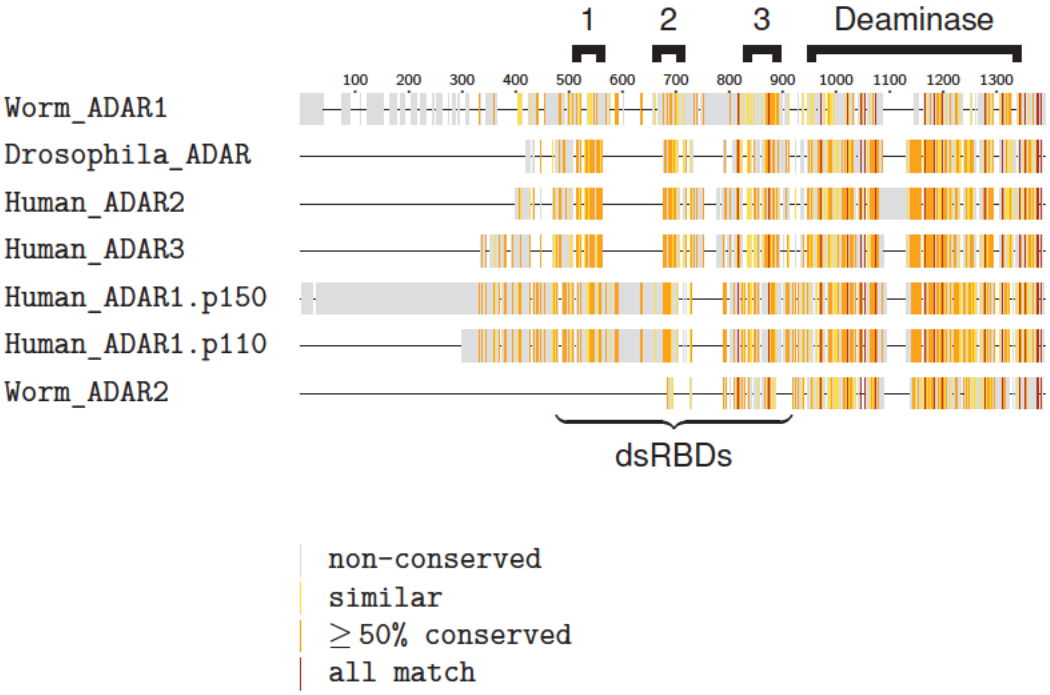


Figure 1.5: Conservation of members of the ADAR enzyme family. Multiple sequence alignment of of human, fruit fly and worm ADAR proteins using Clustal Omega was generated using Clustal Omega (input sequences are collected in table 7.1). The sequence identity is defined by four classes (non-conserved, similar, ≥50% conserved and all match) visualised by colour as indicated in the figure legend.

1.3.2 The role of A-to-I RNA editing in the prevention of a deleterious innate immune response

Perhaps the primary function of RNA editing in mammalian cells is to prevent the activation of an innate immune response triggered by long endogenous dsRNA (Mannion et al., 2014; Pestal et al., 2015). Long stretches of dsRNA can be a signature of viral replication in the cytoplasm, particularly for positive-strand RNA viruses. As an intracellular defence system, eukaryotic cells have evolved a pathway to cope with these threats. ADAR1 is responsible for RNA editing at many specific sites, but none of them can explain the embryonic lethality of ADAR1 knock-out in mice (Q. Wang et al., 2004; Hartner et al., 2004). ADAR1 edits the cellular transcriptome millions of times, and the vast majority of these editing sites are located within primate-specific repetitive elements called Alu repeats (Bazak et al., 2013; Kim et al., 2004). Unedited long endo-

genous dsRNAs resemble structures commonly found in viruses and can activate the melanoma differentiation-associated protein 5 (MDA5)-dependent innate immune response (Weber et al., 2006). As a consequence, edited dsRNA will no longer be recognised by MDA5. Therefore, extensive RNA editing of non-coding regions of mRNA transcripts prevents the deleterious recognition of the cell's dsRNA as non-self by MDA5 and thus inhibits false activation of the cytosolic innate immune response. The only way to rescue the lethal phenotype in ADAR1-deficient mice is to also delete MDA5 to prevent it from recognising long unmodified dsRNA and activating an innate immune response (Heraud-Farlow et al., 2017). Upon additional MDA5 deletion, mice show an average lifespan and normal behaviour, suggesting that prevention of a MDA5-mediated immune response and hence protection against deleterious nucleic acid-sensing is the crucial role of ADAR1-mediated A-to-I RNA editing (Ahmad et al., 2018; Rice et al., 2012).

The viral nucleic acid sensor MDA5 is a member of the retinoic acid-inducible gene I (RIG-I) like receptor (RLR) family and requires hundreds of base pair long dsRNA to become active within the cytosol (Kato et al., 2008; Liddicoat et al., 2015). Other known members of the RLR family detecting various kinds of dsRNA within the cytosol are RIG-I and LGP2 (Satoh et al., 2010). MDA5 belongs to the superfamily II of RNA helicases and consists of two N-terminal caspase activation and recruitment domains (CARD), a central helicase domain, and a C-terminal domain (CTD) (Rawling et al., 2014). Upon activation by long dsRNA, MDA5 interacts with and activates the mitochondrial antiviral signalling protein (MAVS). MAVS is a key mediator of the innate immune response during RNA viral infection, and its activation leads to an interferon response that can severely damage the host cell (Refolo et al., 2020). MAVS, however, is a transmembrane protein localised to the mitochondrial outer membrane and peroxisomes. It consists of a single N-terminal CARD domain followed by a proline-rich region and a transmembrane C terminal domain (Vazquez et al., 2015). Tetramerisation of MDA5's tandem CARD domains is required for further activation of downstream signalling via MAVS (Peisley et al., 2014). If activated by MDA5, MAVS then recruits the inhibitor of nuclear factor kappa-B kinase subunit epsilon (IKKε) and serine/threonine-protein kinase 1 (TBK1) (Fang et al., 2017). This cascade induces phosphorylation and transport of the interferon regulatory factors 3 and 7 (IRF3 and IRF7) into the nucleus, where they induce transcription of type I interferon genes IFN-β and IFN-α (Brisse et al., 2019).

1.3.3 Other known types of RNA editing

Several types of RNA editing other than A-to-I have been discovered in the last four decades. The parasitic kinetoplastid protist *Trypanosoma brucei*, for example, is well-known to modify the nucleotide sequence of about half of its mitochondrial pre-mRNAs through a mechanism by which uracils are being inserted or deleted (Benne et al., 1986; Benne, 1994). This unique mechanism is essential for correcting frameshifts, introducing start and stop codons, and correcting the open reading frames of these mitochondrial transcripts. Trypanosomal RNA editing requires guide RNAs (gRNA) (Blum et al., 1990) for the three stages of mRNA cleavage, U-deletion or insertion and mRNA ligation (Aphasizhev et al., 2014). gRNAs determine the position and the number of uracils being inserted or deleted through hybridisation to pre-mRNA and form a series of mismatches (Cruz-Reyes et al., 2001). Then, two multi-protein and multi-enzyme complexes, the RNA editing core complex (RECC) and the Mitochondrial RNA Binding 1 (MRB1) complex, are recruited by the gRNA-mRNA duplex to execute all of the required enzymatic steps for a successful editing of mitochondrial transcripts (Read et al., 2015; Aphasizheva et al., 2014).

The second type of mRNA editing in mammals is characterised trough cytosine to uracil (C-to-U) deamination. It does not rely on ADAR enzymes for deamination but on the apolipoprotein B mRNA-editing enzyme, catalytic polypeptide (APOBEC) family (Wedekind et al., 2003). This enzyme family received its name from its most prominent editing substrate: the mRNA of apolipoprotein B (ApoB). A CAA is changed to a UAA stop codon, hence generating a truncated isoform called apoB-48 (Blanc et al., 2003). This isoform occurs only in two tissues, the small intestine of all mammals and the liver of some species (Greeve et al., 1993; H. C. Smith et al., 2004). ApoB editing has essential effects on lipoprotein metabolism since it defines distinct pathways for intestinal and hepatic lipid transport in mammals (Salter et al., 2016).

The only types of RNA editing seen in plant organelles (mitochondria and plastids) are either C-to-U in flowering plants or rare cases of U-to-C in ferns and mosses (Hoch et al., 1991). C-to-U editing events of these specific transcripts can vary between 10 to 2000 editing events in different plant species (Takenaka, Zehrmann et al., 2013). One of the functions of RNA editing in plants seems to be the production of modified proteins which are then incorporated into the respiration pathway (Price et al., 2014). An important group

of proteins involved in plant RNA editing is the RNA-binding pentatricopeptide repeat (PRR) protein family (Takenaka, Verbitskiy et al., 2014). Most of the PRR proteins are located in mitochondria and are thought to serve as essential factors for RNA sequence recognition (Andrés-Colás et al., 2017; Barkan et al., 2014).

Adenosine-to-inosine modification of transfer RNAs (tRNAs) is found in tRNAs in all domains of life, prokaryotes as well as eukaryotes, and is essential for codon-anticodon interaction and tRNA thermal stability (Lorenz et al., 2017). A-to-I RNA editing happens mainly at several positions on tRNAs, while position 34 as the first nucleotide of the anticodon is the most important. This modification means that inosines present at the wobble position can base-pair with uracils, cytosines, or adenosine (Agris et al., 2007). The other two RNA editing positions of tRNAs are at 37 following the anticodon and position 57 within the TψC-loop. Since ADAR is specifically and exclusively modifying mRNAs, another enzyme family of deaminases called adenosine deaminases acting on tRNAs (ADAT) is responsible for altering tRNAs. The first ADAT to be identified was the prokaryotic enzyme TadA which forms homodimers and shows sequence homology to subunit Tad2p of yeast tRNA deaminases (Wolf et al., 2002). However, in eukaryotes, the deamination reaction is catalysed by the ADAT2/ADAT3 heterodimer (Torres et al., 2014; Gerber et al., 1999). Both subunits of ADAT in budding yeast are essential as deletion of each subunit is lethal, showing that inosine tRNA editing is paramount for cell viability (Tsutsumi et al., 2007).

1.3.4 RNA-editing related diseases

ADAR1-mediated RNA editing is associated with several specific cancer types. In many tumour types, RNA editing levels are increased compared to normal tissues, suggesting that editing may drive tumorigenesis (Paz-Yaacov et al., 2015). ADAR1 silencing diminishes thyroid cancer cell aggressiveness in both *in vitro* and *in vivo* (Ramirez-Moya et al., 2020). Another cancer to be associated with altered A-to-I RNA editing is acute myeloid leukaemia (Beghini et al., 2000). Tumour cells might exploit ADAR1's role as an immune suppressor for evasion from immune cells (Galore-Haskel et al., 2015). Various cell types in the tumour microenvironment induce interferons to help the innate immune system recognise tumour cells at an early stage and induce the expression of

both MDA5 and protein kinase R (PKR) (Chung et al., 2018). Elevated levels of A-to-I RNA editing of endogenous long dsRNA by increased ADAR1 expression may hide tumour cells from recognition by MDA5 and PKR pathways. ADAR1 activation in cancer is a novel mechanism of tumorigenesis, and silencing ADAR1 could make tumour cells more susceptible to immunotherapy. Consequently, ADAR1 has been identified as a new promising target for anticancer therapy (Ishizuka et al., 2018).

Since ADAR2 is highly expressed in the brain, it is primarily required for site-specific editing of transcripts in the central nervous system. ADAR2-mediated RNA editing is linked to various neurological diseases including schizophrenia (Silberberg et al., 2011), epilepsy and the Prader–Willi syndrome (Slotkin et al., 2013). Further, filamin A, which prevents cardiac remodelling and hypertension, was identified as a target of ADAR2, as RNA editing of lamin A is significantly reduced in cardiovascular patients (Jain et al., 2018).

1.4 The DZF protein family

Eukaryotic cells have developed sophisticated mechanisms to maintain the functional integrity of all proteins made from mRNA transcripts. As a failure to do so may have severe consequences, tight regulation of mRNA processing is critical. The DZF protein family is essential for the regulation of mRNA processing, and in particular alternative splicing and A-to-I RNA editing regulation.

The four essential proteins, nuclear Factor 45 (NF45), nuclear factor 90 (NF90), spermatid perinuclear RNA-binding protein (SPNR), and zinc finger RNA-binding protein (ZFR), share the same dimerisation domain found throughout the metazoans called the DZF domain (domain associated with zinc fingers) and belong to a larger family of proteins that regulate RNA processing (Wolkowicz et al., 2012). All members of the DZF family exhibit a similar domain architecture consisting of the same dimerisation domain as NF45, but with either additional N-terminal zinc fingers (for ZFR and Zn72D) or C-terminal dsRBDs (for SPNR and NF90) (figure 1.6). The DZF domain is highly conserved in metazoans (figure 1.7).

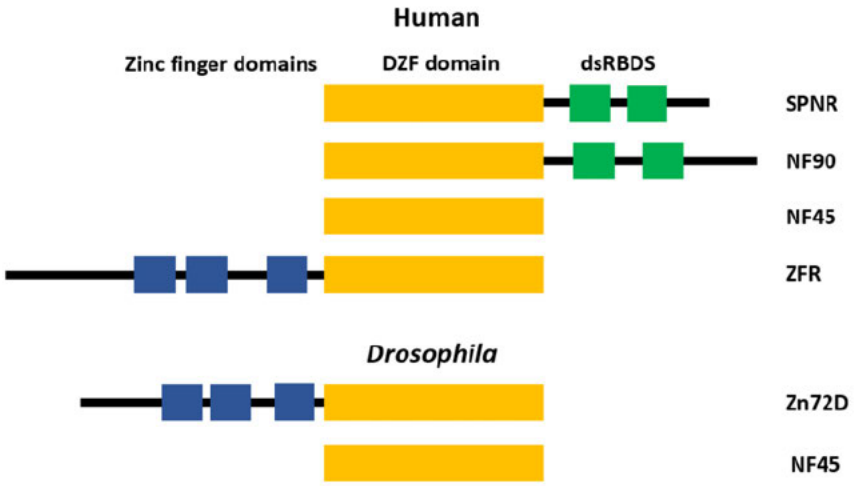


Figure 1.6: Domain architecture of members of the DZF protein family. Representation of Human and *Drosophila* proteins containing a DZF domain. Protein domains are represented as boxes: DZF domain (yellow), zinc finger domains (blue), double-stranded RNA binding domains (dsRBDs) (green).

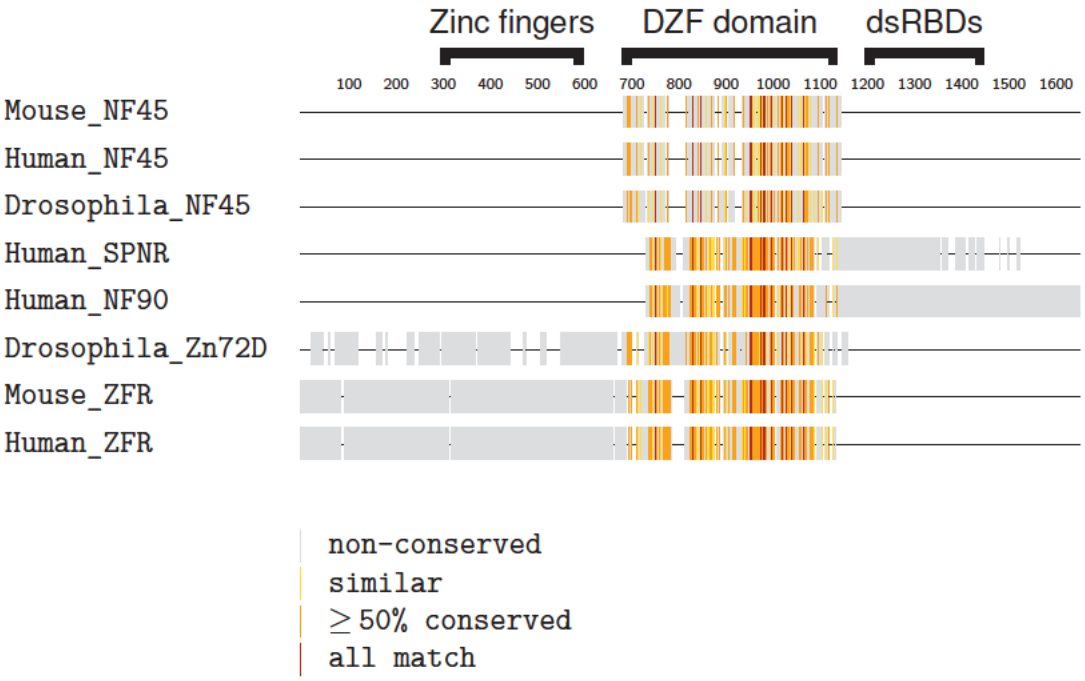


Figure 1.7: Comparison of metazoan proteins containing a DZF domain. Multiple sequence alignment of metazoan proteins containing a DZF domain including NF45 and ZFR from mouse and humans, human NF90 and SPNR and *Drosophila* Zn72D (ZFR orthologue) and NF45 was generated using Clustal Omega (input sequences are collected in table 7.1). The sequence identity is defined by four classes (non-conserved, similar, ≥50% conserved and all match) visualised by colour as indicated in the figure legend.

The DZF domain exhibits structural similarities (Wolkowicz et al., 2012) to template-free nucleotidyltransferases (NTases) such as cyclic GMP-AMP synthase (cGAS) and 2'-5'-oligoadenylate synthetase (OAS), two important nucleic acid sensors in the human innate immune system (Ablasser et al., 2013). However, the lack of conserved catalytic residues suggests that the DZF domain encodes a pseudoenzyme that cannot catalyse the transfer of nucleotides. The crystal structure of an NF45-NF90 dimer (Wolkowicz et al., 2012) was used for a structural alignment of human NF45 and OAS3, highlighting their structural similarities (figure 1.9).

Figure 1.8: Structural alignment of human NF45 and OAS3. Human NF45 (grey, PDBID: 4ATB, chain C, Wolkowicz et al., 2012) and OAS3 (magenta, PDBID: 4S3N, chain, A, Donovan et al., 2015) were used for structural alignment of their nucleotidyltransferase fold domains.

1.4.1 NF90, SPNR and NF45

NF90, SPNR, and ZFR dimerise with NF45 through their DZF domain (Doerks et al., 2002; Parker et al., 2001) to form RNA-binding heterodimers. NF90 and SPNR exhibit a similar domain architecture consisting of an N-terminal DZF domain followed by two double-stranded RNA-binding domains (dsRBDs) and a low-complexity region at the C-terminus (figure 1.6). NF90 and its testes specific paralogue SPNR are only found in vertebrates (Schumacher et al., 1998), while NF45 and ZFR are found throughout the metazoa. In mammals, NF90 and NF45 form a protein complex involved in mRNA processing thereby affecting gene expression both on the transcriptional and translational level

(Nakadai et al., 2015). The DZF domain is highly conserved in metazoans (figure 1.4.1).

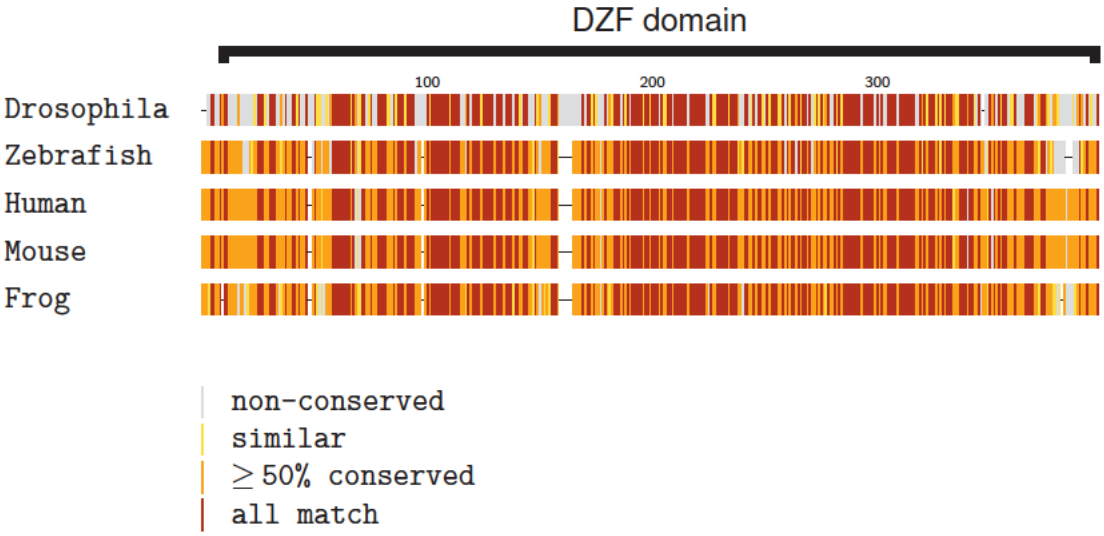


Figure 1.9: Multiple sequence alignment of NF45 proteins. Multiple sequence alignment of metazoan NF45 proteins from human, fruity fly, zebrafish, mouse and frog generated using Clustal Omega (input sequences are collected in table 7.1). The sequence identity is defined by four classes (non-conserved, similar, ≥50% conserved and all match) visualised by colour as indicated in the figure legend.

1.4.2 ZFR

ZFR is found throughout the metazoans, and knockout studies showed loss of this protein is lethal in mice (Meagher et al., 2001). ZFR has three zinc finger RNA binding domains and is known to heterodimerise with NF45 through its DZF domain. In *Drosophila melanogaster*, the function of ZFR is executed by its orthologue zinc-finger protein at 72D (Zn72D). Zn72D also contains a DZF heterodimerisation domain similar to NF45, NF90 and SPNR, but being an orthologue of ZFR, Zn72D also possesses the same highly conserved Cys2-His2-like zinc finger domains (figure 1.10).

Based on recent publications, both human ZFR (Freund et al., 2020) and *Drosophila* Zn72D (Sapiro et al., 2020) were reported to be involved in the regulation of A-to-I RNA editing. Although many RNA editing sites are known, few trans regulators have been identified. Emily Freund performed BioID experiments followed by mass spectrometry to identify regulators of ADAR1 and ADAR2 in HeLa and M17 neuroblastoma cells and identified known and novel ADAR-interacting proteins, among which ZFR was a broadly influential positive regulator of editing.

Anne Sapiro screened about 50 RNA-binding proteins that were thought to be potential candidates that regulate A-to-I RNA editing in *D. melanogaster* and identified Zn72D as a key regulator. While many editing sites were solely dependent on ADAR-mediated RNA editing, more than half of all editing sites also required Zn72D for proper editing. Zn72D was also shown to interact with ADAR in an RNA-dependent fashion, and similar to ADAR, Zn72D is necessary to maintain proper neuromuscular junction architecture and fly mobility.

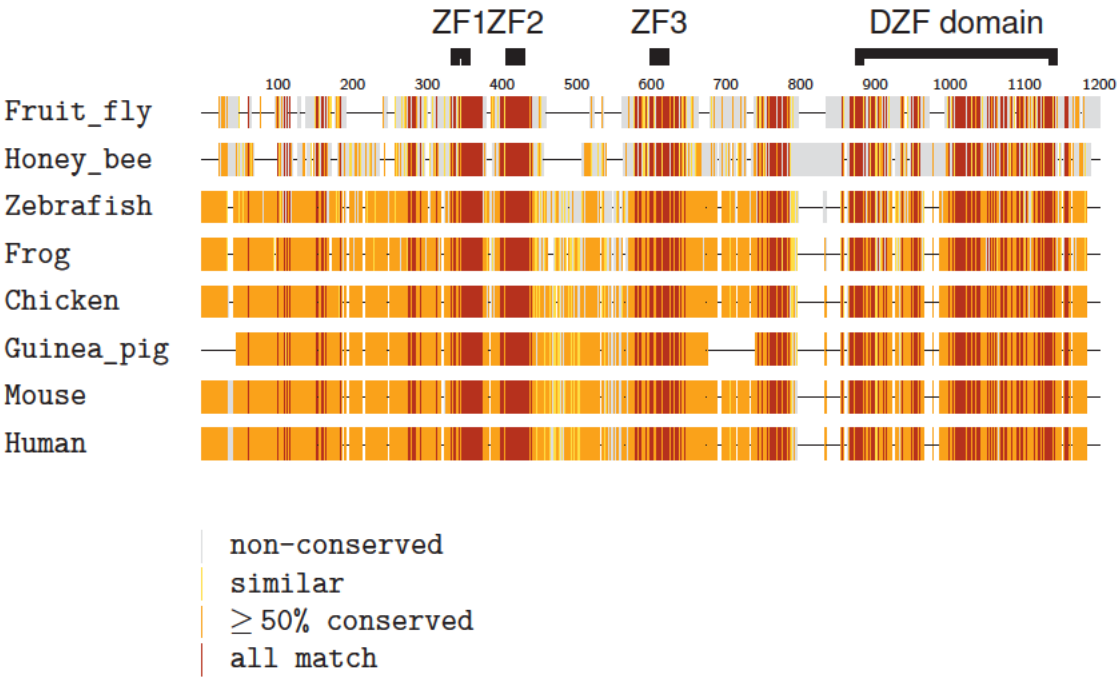


Figure 1.10: Multiple sequence alignment of ZFR and Zn72D proteins. Multiple sequence alignment of mammalian ZFR proteins and Zn72D proteins from invertebrates was generated using Clustal Omega (input sequences can be found in table 7.1). The sequence identity is defined by four classes (non-conserved, similar, ≥50% conserved and all match) visualised by colour as indicated in the figure legend.

1.5 Modulation of the interferon ß response by alternative splicing of mH2A1 mRNA

ZFR has recently also been identified as an inhibitor of the type I interferon response by regulation of alternative pre-mRNA splicing of the specific histone variant macroH2A1 in macrophages. Upon knock-down of ZFR in HEK293 cells, our collaborators from the Hogg lab observed the loss of macroH2A1

(Haque et al., 2018). MacroH2A1 RNA and protein levels were examined by qPCR and WB, and upon ZFR knock-down, showed a significant reduction in RNA levels, and no more protein was detected. This observation can be explained by exon E6 of macroH2A1 being skipped in the absence of ZFR, which leads to the introduction of premature termination codon and degradation of the resulting transcript through NMD (figure 1.11). In the presence of ZFR, macroH2A1 is translated and can inhibit interferon β transcription and subsequent induction of interferon-stimulated genes. However, this is only one of more than 200 hundred alternative splicing events regulated by ZFR, including alternative 5' or 3' splice sites, mutually exclusive exons or intron retention.

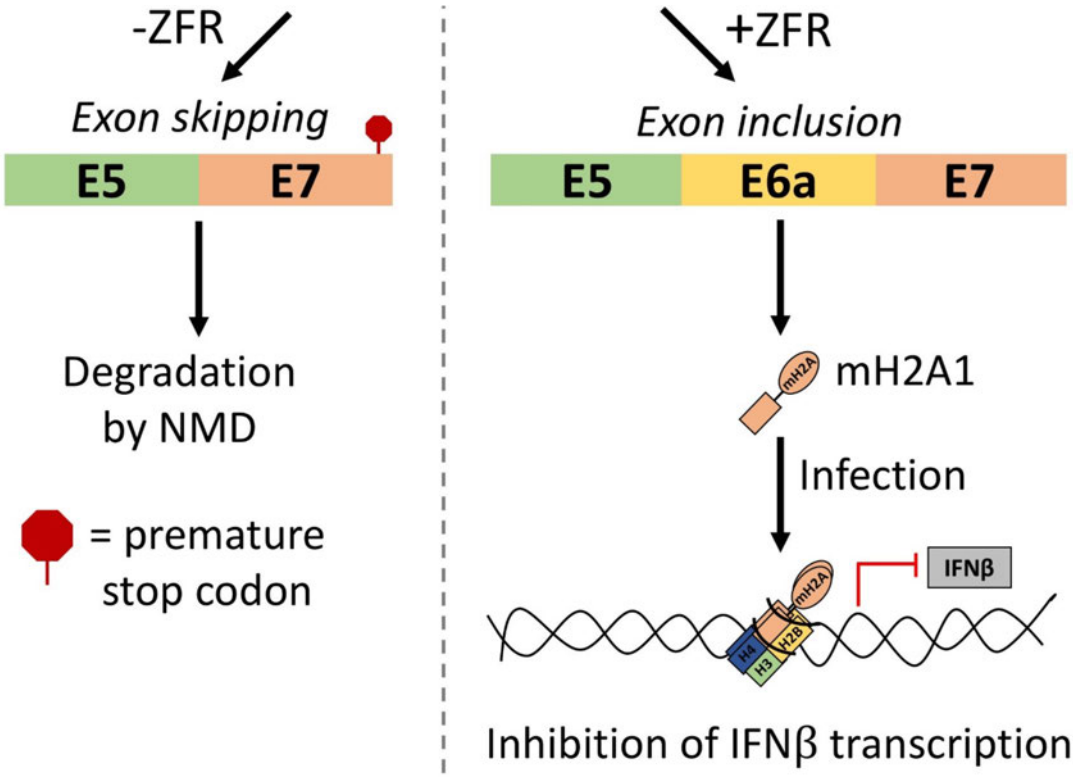


Figure 1.11: Alternative splicing of macroH2A1 mRNA is regulated by ZFR. Alternative splicing of macroH2A1 pre-mRNA in the absence ZFR leads to exon skipping and degradation of the resulting transcript by NMD due to the introduction of a premature stop codon or in the presence of ZFR to exon inclusion, resulting in successful translation.

1.6 Objectives

1.6.1 Gaining an understanding of RNA recognition by ZFR

The identification of RNA motifs that bind ZFR/NF45 might help us to gain a mechanistic understanding of the precise role of ZFR in the regulation of alternative splicing. Therefore, recombinant human ZFR was co-expressed with NF45, and constructs containing or lacking the three zinc-finger domains were purified. These recombinant proteins were sent to the Hogg group for further identification of RNA binding motifs. Based on this data we aimed to work on the following objectives:

1. Identification of RNA binding motifs of ZFR/NF45
2. *In vitro* verification of RNA binding of RBNS-obtained ZFR/NF45 motifs by electromobility shift assays (EMSA)
3. Structural characterisation of human ZFR/NF45 to characterise their molecular mechanism in splicing through RNA binding

1.6.2 Gaining an understanding of the context in which Zn72D is required for full RNA editing

The starting hypothesis for the second part of this work was based on the RNAi screen from Anne Sapiro demonstrating lower editing levels upon Zn72D knock-down while ADAR is still fully active. This observation indicated that more than half of all A-to-I RNA editing sites in *Drosophila* might require co-operative interaction between ADAR and Zn72D at the editing site. To verify this hypothesis that Zn72D/NF45 might be needed for full editing of particular sites in a specific context, and to gain insight into the mechanism of RNA editing on a molecular level, we wanted to reconstitute an A-to-I RNA editing reaction *in vitro* using *in vitro* transcribed RNA templates. A primer-extension-based assay was planned to identify adenosine to inosine changes in transcripts. The objectives of this project were:

1. Establishing and optimisation of an *in vitro* RNA editing assay
2. Expression and purification of full-length fly ADAR for the editing assay
3. Expression and purification of different Zn72D/NF45 versions for the editing assay

4. Structural characterisation of RNA-protein complexes of ADAR and Zn72D/NF45

2 Material and methods

2.1 Materials

2.1.1 Common buffers

10X IVT buffer

Reagent	Final concentration
Tris-HCl pH 8.0	400 mM
Magnesium chloride	500 mM
Triton X-100	0.1% (v/v)
Spermidine	10 mM
DTT	50 mM
ddH_2O	

10X TBE pH 8.3

Reagent	Final concentration
Tris-HCl	0.89 M
Boric acid	0.89 M
EDTA	20 mM
ddH_2O	

50X TAE

Reagent	Final concentration
Acetic acid	1 M
Tris	2 M
EDTA	50 mM
ddH_2O	

10X SDS running buffer

Reagent	Final concentration
SDS	35 mM
Tris-HCl	250 mM
Glycine	2 M
ddH_2O	

2X SDS sample buffer

Reagent	Final concentration
Tris-HCl pH 6.8	100 mM
Glycerol	20% (v/v)
SDS	4% (w/v)
β-mercaptoethanol	10% (v/v)
Bromophenol blue	0.2% (w/v)
ddH_2O	

Lysis buffer for GST-tagged proteins

Reagent	Final concentration
Tris-HCl pH 7.5	20 mM
NaCl	200 mM
DTT	1 mM
ddH_2O	

Lysis buffer for His-tagged proteins

Reagent	Final concentration
Tris-HCl pH 8.0	20 mM
NaCl	150 mM
Imidazole	10 mM
β-mercaptoethanol	0.5 mM
ddH_2O	

RNA elution buffer

Reagent	Final concentration
NaCl	300 mM
NaOAc pH 5.2	60 mM
SDS	0.2% (w/v)
ddH_2O	

2.1.2 Media

Lysogeny broth (LB)

Reagent	Final concentration
Tryptone	1% (w/v)
Yeast extract	0.5% (w/v)
NaCl	0.5% (w/v)
ddH_2O	

2xTY

Reagent	Final concentration
Tryptone	1.6% (w/v)
Yeast extract	1% (w/v)
NaCl	0.5% (w/v)
ddH_2O	

2.1.3 Solutions

6X Native loading dye

Reagent	Final concentration
Glycerol	50% (w/v)
Bromophenol blue	0.25% (w/v)
Xylene cyanol	0.25% (w/v)
ddH_2O	

2X Denaturing RNA loading dye

Reagent	Final concentration
Formamide	95% (v/v)
SDS	0.02% (w/v)
Bromophenol blue	0.02% (w/v)
Xylene cyanol	0.01% (w/v)
EDTA pH 8.0	1 mM
ddH_2O	

Blue silver stain

Reagent	Final concentration
Coomassie brilliant blue G250	0.03% (w/v)
$(NH_4)_2SO_4$	10% (v/v)
Methanol	20% (v/v)
Phosphoric acid	10% (v/v)
ddH_2O	

Coomassie stain

Reagent	Final concentration
Coomassie brilliant blue R250	0.00075% (w/v)
Ethanol	10% (v/v)
Acetic acid	5% (v/v)
ddH_2O	

2.2 Molecular biology

2.2.1 Ligation independent cloning

Ligation independent cloning (LIC) is an alternative to classical restriction enzyme and ligase cloning. T4 DNA polymerases create complementary overhangs for both inserts and vectors. Vector and insert can then be annealed without needing an additional ligation reaction. Joined fragments have nicks that are repaired by *E. coli* during transformation. For designing suitable primers for PCR, the following sequences were added for cloning into LIC vectors:
forward primer: `5'-ccaggggcccgactcg-3'`
reverse primer: `5'-cagaccgccaccgactgc-3'`

ZraI-digestion of LIC vectors

To prepare vectors for subsequent LIC processing, 2 µg of plasmid DNA were digested by ZraI thus excising the GFP gene and linearising the vectors. All reaction components were assembled according to table 2.1, mixed, centrifuged, and then transferred to a heat block preheated to 37 °C. The digestions were incubated for at least 3 hours.

Table 2.1: ZraI digestion of LIC vectors

Reagent	Volume [µL]	Final concentration
LIC vector DNA	variable	2000 ng
10X CutSmart buffer	10	1X
20 U/µL ZraI	3	60 U
ddH_2O	variable	
Total	100	

The linearised plasmids were separated from excised GFP fragments on a 0.8% (w/v) agarose gel containing 1X SYBR Safe. The linearised LIC vector was then extracted from the agarose gel using the QIAquick Gel Extraction Kit from Qiagen according to the manufacturer's instructions.

Processing of LIC vectors

After gel extraction, the linearised vectors were processed for LIC. All reaction components were assembled according to table 2.2, mixed, centrifuged, and incubated for 60 minutes at room temperature. LIC-qualified T4 DNA polymerase was heat-inactivated at 75°C for 20 minutes before LIC processed vectors and inserts were annealed. The final vector concentration is 15 ng/µL.

Table 2.2: LIC vector processing

Reagent	Volume [µL]	Final concentration
Linearised LIC vector DNA	variable	450 ng
10X T4 DNA polymerase buffer	3	1X
25 mM dTTP	3	2.5 µM
100 mM DTT	1.5	5 mM
2.5 U/µL T4 DNA polymerase LIC qualified	0.6	0.15 U
ddH_2O	variable	
Total	30	

Phusion PCR

All reaction components for Phusion PCRs were assembled on ice according to table 2.3, mixed, centrifuged prior to use, and then transferred to a thermocycler preheated to 98°C. Each reaction mix contained the following reagents: template DNA, forward as well as reverse primers, Phusion GC buffer, dNTPs,

DMSO, and Phusion DNA polymerase. DMSO was added to facilitate amplification of difficult targets, such as those with GC-rich sequences or secondary structure. After the PCR reaction, PCR products were analysed on a 1% (w/v) agarose gel containing 1X SYBR Safe to check their size.

Table 2.3: Reaction mix for Phusion PCRs

Reagent	Volume [µL]	Final concentration
Ttemplate DNA	variable	100 ng
10 µM forward primer	2.5	0.5 µM
10 µM reverse primer	2.5	0.5 µM
5X Phusion GC buffer	10	1X
10 mM dNTP mix	1	200 µM
100% DMSO	3	3% (v/v)
2 U/ µL Phusion DNA polymerase	1	0.2 U
ddH_2O	variable	
Total	50	

Table 2.4: Thermocycling conditions of Phusion PCR

Temperature [°C]	Time [min:sec]	Cycles
98	3:00	1
98	0:30	25 to 35
variable	0:30	25 to 35
72	30 sec per kb template DNA	25 to 35
72	10:00	1
4	∞	

The desired PCR products were then extracted from the agarose gel using the QIAquick Gel Extraction Kit from Qiagen according to the manufacturer's instructions.

Processing of LIC inserts

After gel extraction, the PCR products were processed for LIC. All reaction components were assembled according to table 2.5, mixed, centrifuged, and incubated for 60 minutes at room temperature. LIC qualified T4 DNA polymerase was then heat-inactivated at 75°C for 20 minutes before LIC processed vectors and inserts were being annealed.

Table 2.5: Reaction mix for LIC insert processing

Reagent	Volume [µL]	Final concentration
Insert DNA	variable	600 ng
10X T4 DNA polymerase buffer	2	1X
25 mM dATP	2	2.5 µM
100 mM DTT	1	5 mM
2.5 U/µL T4 DNA polymerase LIC qualified	0.4	0.1 U
ddH_2O	variable	
Total	20	

LIC annealing

Both LIC processed vector and insert were annealed according to table 2.6 and incubated for 30 min at room temperature. 1 µL of 25 mM EDTA was added to each reaction and further incubated at RT for 30 minutes before transformation.

Table 2.6: Reaction mix for LIC annealing

Reagent	Volume [µL]	Final concentration
Vector DNA	1	15 ng
Insert DNA	2 to 4	variable
Total	20	

Transformation of electrocompetent XL1 cells

2.5 to 5 µL of annealing mix was added to 50 µL of electrocompetent XL1-Blue cells, which were thawed on ice beforehand. The cells were then transferred to an electroporation cuvette and pulsed. 200 µL of pre-warmed SOC medium was added immediately, and then the cells were shaken for 1 h at 37 ℃ to allow expression of antibiotic resistance proteins. All of the cells were spread on LB agar plates containing appropriate selective antibiotics and grown overnight at 37 ℃.

Plasmid isolation

After successful transformation, single colonies were picked with a single plastic tips to inoculate 5 to 10 mL LB medium containing appropriate selective antibiotics. The cultures were grown overnight at 37 ℃ with constant shaking at 200

rpm and pelleted for 10 min at 4 ℃ at 4000 rpm the next morning. For plasmid isolation, the QIAprep Spin Miniprep Kit from Qiagen was used according to the manufacturer's instructions. All plasmids were eluted in water and stored at -20 ℃.

DNA sequencing

All plasmids were completely internally sequenced after cloning using Sanger sequencing to confirm the insertion of the desired construct in the correct open reading frame (ORF) and to detect potential coding sequence mutations. The sequencing reactions were sent for analysis to either one of the three facilities: Edinburgh Genomics, Source BioScience in Cambridge and Genewiz in Leipzig.

2.2.2 Gibson assembly

LIC plasmids can also be used for Gibson assembly if not being processed for LIC. After digestion by ZraI, the linearised LIC vectors can be assembled with PCR products or gBlocks that have compatible overhangs. All reaction components for Gibson assembly were assembled on ice according to table 2.7, mixed, and then transferred to a heat block preheated to 50 ℃. Each reaction mix contained NEBuilder HiFi master mix and vector and insert DNA with a ratio of 1 to 2. Gibson assembly reactions of two fragments were incubated for 60 minutes which is longer than the recommended amount of time for two fragments being assembled. Following incubation, samples were stored at –20 ℃ for further transformation of chemically competent NEB5-alpha cells.

Table 2.7: Reaction mix of Gibson assembly

Reagent	Volume [µL]	Final concentration
Insert DNA	variable	0.015-0.1 pmol total DNA
Vector DNA	variable	0.015-0.1 pmol total DNA
2X NEBuilder HiFi Master mix	5	1X
ddH_2O	variable	
Total	10	

Transformation of chemically competent NEB5-alpha cells

2 µL of each Gibson assembly mix was added to 25 µL of chemically competent NEB5-alpha cells, which were thawed on ice beforehand for 10 min. The cells were incubated on ice for another 30 min following a 30 s heat-shock at 42 ℃ in a heat block and then chilled for 5 min on ice. 200 µL of pre-warmed SOC medium was added immediately before the cells were shaken for 1 h at 37 ℃ to allow expression of antibiotic resistance proteins. All of the cells were spread on LB agar plates containing appropriate antibiotics and grown overnight at 37 ℃.

2.2.3 Restriction enzyme-based cloning

Restriction digestion

All reaction components for restriction digestions were assembled according to table 2.8, mixed, centrifuged, transferred to a heat block preheated to 37 ℃, and incubated for 1 to 2 hours. Typically, 2000 ng of plasmid or PCR product DNA was used in the reaction. For the restriction enzymes cleaving plasmid DNA with lower efficiency (NEB, 2022), either the total amount of enzyme or the incubation time was increased. Restriction enzymes were heat-inactivated according to the manufacturer's protocol. All digested samples were run on a 1% (w/v) agarose gel containing 1X SYBR Safe to check successful cleavage. According to the manufacturer's instructions, the desired fragments were then extracted from the agarose gel using the QIAquick Gel Extraction Kit from Qiagen.

Table 2.8: Reaction mix for restriction digest

Reagent	Volume [µL]	Final concentration
Plasmid DNA or PCR product	variable	2000 ng
10X CutSmart buffer	5	1X
20 U/µL restriction enzyme	1 to 3	20 to 60 U
ddH_2O	variable	
Total	50	

Ligation

All reaction components for ligation reactions were assembled according to table 2.9, mixed, centrifuged, and then incubated at RT for 60 minutes. The

usual amount of vector DNA in a single reaction was 50 ng and the ratios of insert to vector DNA were 3:1, 5:1, and 7:1. T4 DNA ligase was heat-inactivated at 65℃ for 10 minutes. The ligation samples were briefly spun down and chilled on ice for a few minutes. 2 µL of the ligation mix was used to transform either XL1-Blue or NEB5-alpha cells as described under 2.2.2 and 2.2.1.

Table 2.9: Reaction mix for T4 DNA ligation

Reagent	Volume [µL]	Final concentration
Vector DNA	1	50 ng
Insert DNA	variable	variable
10X T4 DNA ligase buffer	2	1X
400 U/l T4 DNA ligase	1	400 U
Total	20	

2.3 Protein expression in bacteria

2.3.1 Test expressions

After successful cloning, each protein construct was subject to initial small-scale test expression screens using different *E. coli* expression strains (BL21(*DE3*), BL21-CodonPlus (*DE3*)-RIPL, B834(*DE3*), and T7 Express) to determine its expression levels and solubility. 35 ml of 2xTY medium containing the appropriate antibiotics were inoculated with multiple colonies from the same LB agar plate containing transformed cells and then grown at 37℃ with constant shaking at 250 rpm to an OD_{600} of 0.6 to 0.8.

To check basal levels of protein expression (uninduced sample), 1 ml of bacterial cells was harvested (16000 g at 4℃ for 10 min) before inducing expression with IPTG, resuspended in 50 µL of 2X SDS sample buffer per 1 unit of OD_{600} and frozen for later analysis. Upon IPTG induction, proteins were expressed overnight at 18℃ with constant shaking at 250 rpm. Cells were harvested the following morning after 1 ml was taken to check over-expression upon IPTG induction (induced sample), treated the same way as the uninduced sample.

For comparison of protein levels, each test expression sample was normalised to the volume of the medium instead of the cell pellet weight. Therefore, 25 ml of bacterial cells were pelleted for 10 min at 3900 rpm. The supernatant

was discarded, and the pellets were resuspended in 4 ml of cold lysis buffer (for His- or GST-tagged proteins respectively) and then vortexed. Bacterial cells were lysed by sonication for 3.5 min, with 10 s pulses at 40% amplitude followed by pausing for 5 s.

After cell lysis, the soluble proteins were separated from the insoluble ones by centrifugation for 10 min at 4℃ at 13000 rpm. 10 µL of the cleared supernatant was removed and mixed with 20 µL of 2X SDS sample buffer (soluble). 200 µL 2X SDS sample buffer was added to the pellet (pellet). 1.5 ml of the remaining supernatant was mixed with either 30 µL NiNTA or GSH beads prewashed in buffer and incubated for 1 h at 4℃ on a rotating wheel. The beads were washed three times (5000 rpm at 4℃ for 1 min) with 1 ml lysis buffer each, and 30 µL of SDS loading buffer was added before performing SDS-PAGE analysis. All collected protein samples were denatured in SDS sample buffer for 5 min at 95℃. Protein samples (3.5 µL of uninduced and induced, 5 µL of supernatant, 2.5 µL of pellet and 20 µL of beads fractions) were loaded onto a denaturing polyacrylamide gel and separated by SDS-PAGE to show if the protein is expressed and soluble. The gels were then stained with either one of these two Coomassie solutions (table 2.1.3 and 1 2.1.3) to visualise the enriched proteins.

2.3.2 Large-scale expression in bacteria

For large-scale expression of proteins in bacteria, 20 ml LB or 2xTY medium containing appropriate antibiotics were inoculated with multiple colonies from the same LB agar plate and grown at 37℃ with constant shaking at 200 rpm overnight. For a typical protein expression in 3 L, ten 2 L glass flasks containing 300 ml of 2xTY medium were each inoculated with 1 ml of this starter culture and grown at 37℃ to an OD_{600} of 0.6 to 0.8. 1 ml was taken for subsequent SDS-PAGE to check basal levels of expression. The bacterial culture was then cooled for 1 h at 18℃, and expression was induced with IPTG. The protein was expressed overnight at 18℃ with constant shaking at 250 rpm upon IPTG induction. Bacterial cells were harvested the following morning by centrifugation for 20 min at 4℃ at 6000 xg after 1 ml was taken to check expression upon IPTG induction. Also, 25 ml were taken to perform a pull-down to assess expression levels of the desired protein. The supernatant was discarded, and the pellets were flash-frozen in liquid nitrogen and stored at -80℃ for later purification.

2.4 Protein expression in insect cells

2.4.1 Transformation of EMBacY cells

The successful sequenced LIC plasmids for insect cell work were used to generate a large plasmid containing the baculovirus genome called a bacmid. EMBacY cells, which carry the empty bacmid shuttle vector and a helper plasmid, were thawed on ice for 10 min before adding 100 ng of the sequenced LIC plasmids. The cells were incubated on ice for 30 min following a 30 s heat-shock at 42 °C in a heat block and then chilled for 5 min on ice. 200 µL of SOC medium was added before the cells were shaken for 5 h at 37 °C to allow expression of antibiotic resistance proteins. All of the cells were spread on LB agar plates containing IPTG and Xgal for blue-white screening and kanamycin, gentamicin, tetracycline, and carbenicillin and grown for 48 h at 37 °C.

2.4.2 Bacmid isolation

A few white colonies (positive clones) and one blue colony (negative clones) were picked and restreaked on a new plate to ensure proper blue-white screening. After growing them for another 48 h at 37 °C, single colonies were used for inoculating 5 mL LB overnight (ON) cultures containing the appropriate antibiotics. Each ON culture was pelleted for 10 min at 4 °C at 4000 rpm. Buffers from the QIAprep Spin Miniprep Kit from Qiagen were used to lyse the cells. Pelleted bacterial cells were resuspended in 250 µL P1 buffer containing RNase A and LyseBlue and then transferred to a new Eppendorf tube. 250 µL P2 buffer was added per sample, and the tubes were inverted several times. The cells were lysed for a total of 5 min. Next, 350 µL N3 buffer was added, the tubes were again inverted several times and then centrifuged for 10 min at 4 °C at 13000 rpm. The supernatant was transferred to a new Eppendorf tube, and 800 µL isopropanol was added. The bacmid DNA was pelleted by centrifugation for 10 min at 4 °C at 13000 rpm, washed twice with 500 µL ice-cold 70% EtOH and pelleted again for 5 min at 4 °C at 13000 rpm. The pellet was air-dried for a few minutes and resuspended in 100 µL water.

2.4.3 Generation of baculoviruses for protein expression

After successful bacmid isolation, $2*10^7$ *Spodoptera frugiperda* (Sf) 9 cells were prepared in 10 mL insect cell medium and incubated for 30 min at 27 °C

to let the cells attach to the surface of a T75 flask. Next, 40 µg of the bacmid were added to 400 µL of the medium, and then 16 µL of Xtreme GENE HP transfection reagent were added to the mixture. The whole transfection mix was added to the Sf9 cells and antibiotics (the final concentration of antibiotics was 1X). The plate was rocked to distribute the bacmids and incubated at 27℃ for 7 days. Sf9 cells were pelleted at 500 xg for 5 min, and the supernatant, which contains the virus, was filtered through a syringe unit with 0.2 µm.

To generate virus stock V1, $1*10^8$ Sf9 cells were infected with all 10 mL of the virus stock V0, as the virus was usually not very infectious, in a total volume of 50 mL with 1X antibiotics and grown at 27℃ with shaking at 110 rpm for 3 days while monitoring fluorescence. The expression of fluorescent YFP helps to quantify the number of Sf9 cells infected with the virus. Sf9 cells were pelleted at 500 xg for 5 min, and the supernatant was filtered through a syringe unit with 0.2 µm. To generate virus stock V2, $2*10^8$ Sf9 cells were infected with all of the 20 mL virus stock V1 in a total volume of 100 mL with 1X antibiotics and grown at 27℃ with shaking at 110 rpm for 3 days. Sf9 cells were pelleted at 500 xg for 5 min, and the supernatant was filtered through a syringe unit with 0.2 µm.

2.4.4 Test expressions in Sf9 cells

Each protein construct was subject to initial small-scale test expression screenings in Sf9 cells to determine its expression level and solubility. Insect cell medium was inoculated with virus stock in ratios of 1:25, 1:50, and 1:100 (virus:medium) to make a total volume of 50 ml without antibiotics. The cultures were grown at 27℃ with constant shaking at 100 rpm while monitoring their fluorescence and viability. The cells were harvested after 48 h, or earlier if they showed decreasing viability, by centrifugation at 4℃ at 3900 rpm. The insect cell samples taken for analysis of protein expression were treated the same as bacterial test expressions (section 2.3.1).

2.4.5 Large-scale expression in Sf9 cells

2 L of insect cell medium without antibiotics that contained $2*10^6$ Sf9 cells/ml of medium were infected with virus stock V2 in a ratio of either 1:100 or 1:50 and grown at 27℃ with shaking at 110 rpm for 48 hours while monitoring their fluorescence and viability. Generally, the insect cells were harvested by centrifugation for 20 min at 4℃ at 6000 xg when the cells showed about 85%

fluorescence with higher viability than 90%. The cell pellets were flash-frozen in liquid nitrogen and stored at -80 °C until use.

2.5 Protein purification

2.5.1 Bacterial cell lysis

Bacterial cell pellets were resuspended to a final concentration of 10% (w/v) in the corresponding lysis buffer. Lysis buffers typically contained about 400 µg DNase I, one EDTA free protease inhibitor cocktail tablet (Roche), and 500 µg Pefabloc per 2 L of harvested bacterial cells. The resuspended cells were passed once through a constant cell disruptor at 25 kpsi pre-chilled at 6 °C and then sonicated on ice six times (each sonication lasted for 2 min in total with 20 s pulses at a 40% amplitude followed by pausing for 10 s) to fragment nucleic acids. The cell lysates were centrifuged afterwards to separate soluble proteins from cell debris and insoluble components.

2.5.2 Insect cell lysis

Insect cell pellets were prepared as 20% (w/v) in the corresponding lysis buffer. If no significant proteolysis was visible during protein purification, lysis buffers generally contained 400 µg DNase I, one EDTA free protease inhibitor cocktail tablet (Roche), and 500 µg Pefabloc per 1 L of harvested insect cells. Insect cells were lysed on ice by sonicating them six times (each sonication lasted for 2 min in total with 20 s pulses at a 40% amplitude followed by pausing for 10 s). The cell lysates were centrifuged afterwards to separate soluble proteins from cell debris and insoluble components.

2.5.3 Large-scale purification of human ZFR/NF45

For purification of ZFR/NF45, a five-step protocol was established and all steps were carried out at 4 °C.

1. First immobilised metal affinity chromatography (NiNTA)
2. Second immobilised metal affinity chromatography (NiNTA)
3. Heparin sepharose chromatography

4. Ion exchange chromatography

5. Size exclusion chromatography

Cell lysates were cleared after lysis by centrifugation (50000g, 4℃, 45 min) and bound in batch to Ni-NTA resin, pre-equilibrated in lysis buffer, at 4℃ for 2 hours. Beads were packed into a column, connected to an ÄKTAprime plus system, washed with NiNTA buffer A and eluted in a linear gradient of NiNTA buffer B. The fractions containing ZFR and NF45 with a reasonable low 260/280 ratio were pooled.The N-terminal 6xHis-tag of NF45 was cleaved overnight using 2 mg of GST-TEV protease during dialysis into 20 mM Tris-HCl pH 8.0, 200 mM NaCl and 0.5 mM β-ME.

After overnight tag cleavage and dialysis, the sample was filtered the next day using a 0.22 µm filter and loaded onto the same self-packed column containing NiNTA resin as before to remove excess of tag-less NF45 and GST-TEV protease.

After the second affinity step, ZFR's N-terminal His-tag was cleaved overnight using 1 mg of GST-3C protease during dialysis into 20 mM HEPES pH 7.5, 200 mM NaCl and 1 mM DTT (same as Heparin buffer A). The sample was filtered using a 0.22 µm filter and further purified using a 12 ml heparin sepharose column. The samples were eluted using a linear gradient of 150 to 1000 mM NaCl. ZFR/NF45 was dialysed into 20 mM HEPES pH 7.5, 100 mM NaCl and 1 mM DTT overnight (same as IEX buffer A), applied to a 6 ml S column, washed, and eluted with a 50 to 1000 mM NaCl gradient.

Finally, ZFR/NF45 was concentrated in a spin concentrator (PES, 10 kDa MWCO) and applied to a HiLoad 16/600 Superdex 200 pg equilibrated in EMSA buffer (20 mM HEPES pH 7.5, 150 mM KOAc, 4 mM MgOAc and 1 mM DTT). Main peak fractions with a 260/280 ratio of 0.6 were pooled, concentrated, flash-frozen in liquid nitrogen, and stored at -80℃ until use.

NiNTA buffer A

Reagent	Final concentration
Tris-HCl pH 8.0	20 mM
Sodium chloride	600 mM
Imidazole pH 8.0	10 mM
β-mercaptoethanol	0.5 mM
ddH_2O	

NiNTA buffer B

Reagent	Final concentration
Tris-HCl pH 8.0	20 mM
Sodium chloride	200 mM
Imidazole pH 8.0	1000 mM
β-mercaptoethanol	0.5 mM
ddH_2O	

Heparin buffer A

Reagent	Final concentration
HEPES pH 7.5	20 mM
Sodium chloride	200 mM
DTT	1 mM
ddH_2O	

Heparin buffer B

Reagent	Final concentration
HEPES pH 7.5	20 mM
Sodium chloride	1000 mM
DTT	1 mM
ddH_2O	

IEX buffer A

Reagent	Final concentration
HEPES pH 7.5	20 mM
Sodium chloride	100 mM
DTT	1 mM
ddH_2O	

IEX buffer B

Reagent	Final concentration
HEPES pH 7.5	20 mM
Sodium chloride	1000 mM
DTT	1 mM
ddH_2O	

SEC (EMSA) buffer

Reagent	Final concentration
Tris-HCl pH 7.5	20 mM
Potassium acetate	150 mM
Magnesium acetate	4 mM
DTT	1 mM
ddH_2O	

2.5.4 Large-scale purification of *D. melanogaster* Zn72D/NF45

For purification of fruit fly Zn72D/NF45, a five-step protocol was established and all steps were carried out at 4 ℃.

1. Affinity chromatography (GSH beads)
2. First immobilised metal affinity chromatography (NiNTA)
3. Second immobilised metal affinity chromatography (NiNTA)
4. Heparin sepharose chromatography
5. Size exclusion chromatography

Cell lysates were cleared after lysis by centrifugation (45000 xg, 4 ℃, 45 min) and bound in batch to GSH resin, pre-equilibrated in lysis buffer, at 4 ℃ for 3 hours. Beads were packed into a column, connected to an ÄKTAprime plus system, washed with buffer A, and eluted in GSH buffer B until no more protein was eluted off the beads. The fractions containing Zn72D/NF45 were pooled and diluted before ON dialysis to prevent precipitation.

NF45's N-terminal GST-tag was cleaved overnight at 4 ℃ in a 6-8 kDa MWCO dialysis tubing using 1 mg of GST-TEV protease during dialysis into 20 mM Tris-HCl pH 8.0, 250 mM NaCl, 10 mM imidazole, and 1 mM β-mercaptoethanol (same as NiNTA buffer A). The sample was filtered the next day using a 0.22 µm filter and loaded onto a HisTrap 5 ml FF column. The column was washed with NiNTA buffer A and eluted with a linear gradient of NiNTA buffer B. Zn72D's His-tag was cleaved overnight at 4 ℃ in a 6-8 kDa MWCO dialysis tubing using 1 mg of GST-3C protease during dialysis into 20 mM Tris-HCl pH 8.0, 250 mM NaCl, 10 mM imidazole, and 1 mM β-mercaptoethanol (same as NiNTA buffer A). The sample was filtered the next day using a 0.22 µm filter and loaded onto

the same HisTrap 5 ml FF column from the day before. The FT was collected and dialysed ON at 4℃ in a 6-8 kDa MWCO dialysis tubing into 20 mM Tris-HCl pH 7.5, 100 mM NaCl, and 1 mM DTT (same as Heparin buffer A).

The sample was filtered the next day through a 0.22 µm filter and loaded onto a HiTrap Heparin HP 5 ml column. The column was washed with Heparin buffer A and eluted with a linear gradient of Heparin buffer B.

Finally, Zn72D/NF45 was concentrated in a spin concentrator (PES, 10 kDa MWCO) and applied to a HiLoad 16/600 Superdex 200 pg equilibrated in SEC (EMSA) buffer. Main peak fractions with a 260/280 ratio of 0.6 were pooled, concentrated, flash-frozen in liquid nitrogen, and stored at -80℃ until use.

GSH buffer A

Reagent	Final concentration
Tris-HCl pH 8.0	40 mM
Sodium chloride	750 mM
Urea	50 mM
β-mercaptoethanol	2 mM
ddH_2O	

GSH buffer B

Reagent	Final concentration
Tris-HCl pH 8.0	40 mM
Sodium chloride	750 mM
Glutathione (reduced)	1000 mM
β-mercaptoethanol	1 mM
ddH_2O	

NiNTA buffer A

Reagent	Final concentration
Tris-HCl pH 8.0	20 mM
Sodium chloride	250 mM
Imidazole	10 mM
β-mercaptoethanol	1 mM
ddH_2O	

NiNTA buffer B

Reagent	Final concentration
Tris-HCl pH 8.0	20 mM
Sodium chloride	250 mM
Imidazole	1000 mM
β-mercaptoethanol	1 mM
ddH_2O	

Heparin buffer A

Reagent	Final concentration
HEPES pH 7.5	20 mM
Sodium chloride	100 mM
DTT	1 mM
ddH_2O	

Heparin buffer B

Reagent	Final concentration
HEPES pH 7.5	20 mM
Sodium chloride	1000 mM
DTT	1 mM
ddH_2O	

SEC (EMSA) buffer

Reagent	Final concentration
Tris-HCl pH 7.5	20 mM
Potassium acetate	150 mM
Magnesium acetate	4 mM
DTT	1 mM
ddH_2O	

2.5.5 Large-scale purification of *D. melanogaster* ADAR

For purification of *D. melanogaster* ADAR, a three-step protocol was established and all steps were carried out at 4℃.

1. Affinity chromatography (GSH beads)

2. Heparin sepharose chromatography

3. Size exclusion chromatography

Cell lysates were cleared after lysis by centrifugation (45000 xg, 4 °C, 60 min) and bound in batch to GSH resin, pre-equilibrated in lysis buffer, at 4 °C for 3 hours. Beads were washed with wash buffers with a decreasing KCl concentration (750, 500, and 250 mM) and ADAR was eluted off the beads with GSH buffer B in several fractions. The fractions containing ADAR with a reasonable low 260/280 ratio were pooled.

ADAR's N-terminal His-GST-tag was cleaved overnight in a 6-8 kDa MWCO dialysis tubing using 1 mg of GST-3C protease during dialysis into 20 mM HEPES pH 7.5, 150 mM KCl, glycerol 10% (v/v), and 1 mM DTT (same as Heparin buffer A). The sample was filtered the next day using a 0.22 µm filter and loaded onto a 12 ml heparin sepharose column. The column was washed with Heparin buffer A and eluted with a linear gradient of Heparin buffer B.

Finally, ADAR was concentrated in a spin concentrator (PES, 10 kDa MWCO) and applied to a HiLoad 16/600 Superdex 75 pg equilibrated in EMSA buffer (20 mM HEPES pH 7.5, 150 mM KOAc, 4 mM MgOAc and 1 mM DTT). Main peak fractions with a 260/280 ratio of 0.6 were pooled, concentrated, flash-frozen in liquid nitrogen, and stored at -80 °C until use.

GSH buffer A

Reagent	Final concentration
Tris-HCl pH 7.5	20 mM
Potassium chloride	1000 mM
Glycerol	20% (v/v)
Urea	100 mM
β-mercaptoethanol	1 mM
ddH_2O	

GSH buffer B

Reagent	Final concentration
Tris-HCl pH 7.5	20 mM
Potassium chloride	1000 mM
Glycerol	20% (v/v)
Urea	100 mM
Glutathione (reduced)	20 mM
β-mercaptoethanol	1 mM
ddH_2O	

Heparin buffer A

Reagent	Final concentration
HEPES pH 7.5	20 mM
Potassium chloride	150 mM
Glycerol	10% (v/v)
DTT	1 mM
ddH_2O	

Heparin buffer B

Reagent	Final concentration
HEPES pH 7.5	20 mM
Potassium chloride	1000 mM
Glycerol	10% (v/v)
DTT	1 mM
ddH_2O	

SEC (EMSA) buffer

Reagent	Final concentration
HEPES pH 7.5	20 mM
Potassium acetate	150 mM
Magnesium acetate	4 mM
DTT	1 mM
ddH_2O	

2.5.6 Large-scale purification of human Endonuclease V

For purification of human EndoV, a four-step protocol was established and all steps were carried out at 4℃.

1. Immobilised metal affinity chromatography (NiNTA)
2. Affinity chromatography (GSH beads)
3. Heparin sepharose chromatography
4. Size exclusion chromatography

Cell lysates were cleared after lysis by centrifugation (45000xg, 4℃, 45 min) and filtered through 0.22 µm. The cleared lysate was loaded onto a HisTrap 5 ml FF column using the sample pump of an ÄKTA pure system, washed with NiNTA buffer A and eluted in a linear gradient of NiNTA buffer B. The pooled fractions were diluted 1 to 4 to prevent precipitation ON and bound in batch to GSH resin the next day, pre-equilibrated in GSH buffer A, at 4℃ for 3 hours. Beads were packed into a column, connected to an ÄKTAprime plus system, washed with buffer A and eluted in GSH buffer B until no more protein was eluted off the beads.

EndoV's His-GST-tag was cleaved overnight using 1 mg of GST-3C protease during dialysis into 20 mM HEPES pH 7.5, 100 mM NaCl and 1 mM DTT (same as Heparin buffer A). The sample was filtered the next day using a 0.22 µm filter and loaded onto a HiTrap Heparin HP 5 ml column. The column was washed with Heparin buffer A and eluted with a linear gradient of Heparin buffer B.

Finally, EndoV was concentrated in a spin concentrator (PES, 10 kDa MWCO) and applied to a HiLoad 16/600 Superdex 75 pg equilibrated in SEC buffer. Main peak fractions with a 260/280 ratio of 0.6 were pooled, concentrated, flash-frozen in liquid nitrogen, and stored at -80℃ until use.

NiNTA buffer A

Reagent	Final concentration
Tris-HCl pH 8.0	40 mM
Sodium chloride	500 mM
Immidazole	10 mM
β-mercaptoethanol	2 mM
ddH_2O	

NiNTA buffer B

Reagent	Final concentration
Tris-HCl pH 8.0	40 mM
Sodium chloride	500 mM
Immidazole	1000 mM
β-mercaptoethanol	2 mM
ddH_2O	

GSH buffer A

Reagent	Final concentration
Tris-HCl pH 8.0	20 mM
Sodium chloride	250 mM
DTT	1 mM
ddH_2O	

GSH buffer B

Reagent	Final concentration
Tris-HCl pH 8.0	20 mM
Sodium chloride	250 mM
Glutathione (reduced)	30 mM
DTT	1 mM
ddH_2O	

Heparin buffer A

Reagent	Final concentration
HEPES pH 7.5	20 mM
Sodium chloride	100 mM
DTT	1 mM
ddH_2O	

Heparin buffer B

Reagent	Final concentration
HEPES pH 7.5	20 mM
Sodium chloride	1000 mM
DTT	1 mM
ddH_2O	

SEC buffer

Reagent	Final concentration
Tris-HCl pH 7.5	20 mM
Potassium acetate	50 mM
Manganese acetate	4 mM
DTT	1 mM
ddH_2O	

2.6 Electrophoretic mobility shift assay (EMSA)

All RNA oligonucleotides were purchased from Biomers.net (Ulm, Germany) with either a DY681- or Cy5-fluorophore attached to the 5' end. Oligonucleotides were reconstituted in water at 100 µM. Complementary strands were mixed and heated to 95℃ with cooling overnight to RT to form duplexes. Binding reactions contained protein and RNA or RNA:DNA oligos in 20 mM HEPES pH 7.5, 150 mM potassium acetate, 4 mM magnesium acetate and 1 mM DTT. EMSAs with different proteins used identical conditions. Each reaction was assembled in 10 µL, containing a final concentration of 0.5 µM RNA or duplexes and an increasing concentration of protein. Samples were incubated on ice for 1 hour. A 20 cm x 20 cm 8% native polyacrylamide gel in 0.5X TBE buffer was pre-run at 2 W at 4℃ for 1 hour. To each sample, 2 µL 6X native gel loading buffer was added. 4 µL of each reaction was loaded onto the gel and run at 2 W at 4℃ for 1 hour. Migration of fluorescently labelled RNA oligos was visualized using the Odyssey® CLx (LI-COR) imaging system at 700 nm. Images were converted to grayscale using Image Studio.

2.7 RNA editing assay

2.7.1 Large-scale plasmid isolation

XL1-Blue cells were transformed with the plasmids destined for large-scale isolation as described in section 2.2.1. 150 ml of LB medium were inoculated with a single colony each from the transformed LB agar plates and grown ON at 37℃ with constant shaking at 250 rpm. The cells were harvested for 10 min at 4℃ at 3900 rpm the next day. The cell pellets were lysed, and the plasmids purified using the ZymoPURE II Plasmid Maxiprep Kit according to the manufacturer's protocol .

2.7.2 Linearisation of plasmids

For run-off *in vitro* transcriptions, the DNA templates were linearised first. 100 µg of pUC19 plasmid containing the sequence to be transcribed were digested using 200 U SmaI and incubated at 25℃ for 4 hours according to table 2.10. After the incubation, the linearised plasmids were separated on a 1% (w/v) agarose gel containing 1X SYBR Safe. The linearised DNA template was then extracted from the gel using the QIAquick Gel Extraction Kit from Qiagen according to the manufacturer's instructions and further purified by phenol:chloroform extraction and ethanol precipitation following the same protocol as described for extracting RNA in section 2.7.4. However, for DNA extractions, the pH of the phenol:chloroform:isoamyl alcohol solution was 7.8. The air-dried DNA pellet was dissolved in 100 µL nuclease-free water for use in run-off *in vitro* transcriptions.

Table 2.10: SmaI digestion of pUC19 plasmids

Reagent	Volume [µL]	Final concentration
Plasmid DNA	variable	2000 ng
10X CutSmart buffer	20	1X
10 U/µL SmaI	20	200 U
Nuclease-free water	variable	
Total	200	

2.7.3 *In vitro* transcription of short RNAs

All reaction components for run-off *in vitro* transcription were assembled on ice according to table 2.11, mixed, centrifuged, transferred to a heat block preheated to 37℃, and incubated for 4 hours. Each reaction contained the following components: template DNA, transcription buffer, rNTPs, RNase inhibitor, inorganic pyrophosphatase (IPP) from *E. coli*, and T7 RNA polymerase. In general, our own laboratory stock of T7 RNA polymerase was used for *in vitro* transcription reactions. After 4 hours of incubation, 4 U of Turbo DNase was added per reaction and incubated for another hour at 37℃. IVT reactions were upscaled accordingly as needed. The final RNA products were purified by one of the three methods: phenol:chloroform extraction followed by ethanol precipitation; lithium chloride precipitation; or gel extraction with subsequent phenol:chloroform extraction and ethanol precipitation.

Table 2.11: Run-off in vitro transcription reaction

Reagent	Volume [µL]	Final concentration
Linearised template DNA	variable	100 nM
25 mM rNTPs	6	8.3 mM
10X IVT buffer	2	1X
40 U/µL RiboLock RNase inhibitor	1	40 U
0.1 U/µL Inorganic pyrophosphatase	1	0.1 U
5 µg/µL T7 RNA polyermase	1	10 µg
Nuclease-free water	variable	
Total	20	

2.7.4 Phenol:chloroform extraction and ethanol precipitation

For RNA purification after *in vitro* transcription, an equal volume of 300 µL phenol:chloroform:isoamyl alcohol (ratios of 25:24:1) pH 4.5 was made up to a volume of 300 µL with nuclease-free water. The tubes were vortexed vigorously and centrifuged for 5 min at 21℃ at 13000 xg. The RNA-containing upper aqueous phase was gently transferred to a new Eppendorf tube without disturbing the other phases. RNA extraction was performed a second time as described. Residual phenol was removed from the sample using pure chloroform. An equivalent volume of 300 µL chloroform to phenol was added to the RNA-containing aqueous phase. The solution was vortexed and then centrifuged for 5 min at 21℃ at 13000 xg. This step was executed again. To precipitate RNA, 0.1X of the sample volume 3 M NaOAc pH 5.2 and 2.5X of the sample volume ice-cold 100% ethanol were added to the RNA containing solution and precipitated for at least 1 hour at -80℃. The RNA was pelleted for 30 min at 4℃ at 13000 xg. The RNA pellet was washed with 1 mL ice-cold 70% ethanol and pelleted again for 10 min at 4℃ at 13000 xg. The RNA pellet was air-dried for a few minutes and dissolved in a small volume of 20 µL nuclease-free water.

2.7.5 Lithium chloride precipitation

An equal volume of 4 M lithium chloride (LiCl) was added to IVT samples, to adjust the LiCl concentration to 2 M, and incubated at -20℃ for 30 min. The RNA was pelleted at 4℃ for 20 min at 13000 xg. The supernatant was

discarded, the RNA pellet was washed with 500 µL ice-cold 70% EtOH, and spun at 4 ℃ for 10 min at 13000 xg. The RNA pellet was air-dried for a few minutes and dissolved in 20 µL nuclease-free water.

2.7.6 Gel extraction of RNA

IVT samples were separated on a 20% denaturing urea acrylamide gel (1X TBE, 7 M Urea) in 1X TBE buffer at 20 W for up to 3 hours, wrapped in cling film, and placed on a Fluor-coated TLC plate (Ambion). The RNA bands were visualised using UV shadowing under short wave (254 nm) light and marked on the cling film. The bands were excised from the gel and incubated in 500 µL of RNA elution buffer overnight. The following day, the RNA was phenol:chloroform extracted and ethanol precipitated (section 2.7.4). The RNA pellet was air-dried for a few minutes and dissolved in a small volume of 20 µL nuclease-free water.

2.7.7 RNA editing reaction

RNAs were incubated with *D. melanogaster* ADAR in one of two buffer systems (EMSA or Paper buffer (see Vik et al., 2013) to reconstitute adenosine deamination reactions *in vitro*. For subsequent analyses by poisoned primer extension assays, the RNAs do not require any modification, but a fluorescent label is needed for visualisation for endonuclease V assays.

All reaction components for RNA editing reactions were assembled according to table 2.12 at RT, mixed, centrifuged, and incubated for 3 hours in total. The reactions were either incubated at 21 ℃, 30 ℃, 37 ℃, or alternating between 30 ℃ and37 ℃ for 30 min each.

Table 2.12: RNA editing reaction

Reagent	Volume [µL]	Final concentration
1 µM RNA oligo	1	1 pmol
40 U/µL RiboLock RNase inhibitor	0.5	20 U
Fruit fly ADAR	variable	10 pmol
1X buffer	variable	1X
Total	5	

2.7.8 Poisoned primer extension assay

Fluorescently-labelled DNA primers that hybridise a short distance downstream from the editing site of interest were extended by reverse transcription. The reactions contain three nucleotides for extension (dATP, dGTP, and dCTP) and a chain-terminating dideoxynucleotide, ddTTP. The primer (1 pmol) was mixed with the corresponding *in vitro* transcribed RNA in a 1:1 ratio. This mix was heated to 85℃ for 3 minutes to allow annealing and forming of a DNA:RNA duplex that is needed for primer extension by reverse transcriptase. Reaction components for the poisoned primer extension were assembled according to table 2.13, mixed, centrifuged, transferred to a heat block preheated to 37℃, and incubated for 1 hour. The extended DNA primers were separated on a 20% denaturing urea acrylamide gel (1X TBE, 7 M Urea). They were first heated to 95℃ for 2 min and then chilled on ice. The gel was run at 20 W for 2 to 3 hours.

Table 2.13: Poisoned primer extension assay

Reagent	Volume [µL]	Final concentration
RNA	variable	2 ng
DNA primer	variable	1X
Nuclease-free water	variable	
Total	5	

Table 2.14: Poisoned primer extension assay

Reagent	Volume [µL]	Final concentration
Annealing reaction	5	2000 ng
10X Reverse transcription buffer	20	1X
40 U/µL RiboLock RNase inhibitor	0.5	20 U
1 mM dNTP mix	20	0.1 mM
1 mM dTTP	20	0.1 mM
Reverse transcriptase	20	200 U
Nuclease-free water	variable	
Total	200	

2.7.9 Endonuclease V assay

Human endonuclease V is a highly active ribonuclease specific for inosine in RNA. To visualise A-to-I RNA editing, fluorescently-labelled RNAs were incubated with this enzyme to cleave RNAs at converted inosines. For full enzymatic activity of EndoV in EMSA buffer at pH 7.5, divalent manganese ions are required rather than magnesium ions according to Vik et al., 2013. To address this, magnesium and mangenese ions were added to EMSA buffer and the Paper buffer (10 mM Tris-HCl pH 7.5, 50 mM potassium chloride, 0.5 mM manganese chloride, 1 mM DTT and 5% glycerol) to see if this affects cleavage of inosine-containing RNAs.

All reaction components for endonuclease V assays were assembled according to table 2.15, mixed, centrifuged, transferred to a heat block preheated to 37℃, and incubated for 2 hours. 10 µL of denaturing 2X RNA loading dye were added to each sample and heated to 50℃ for 5 min. Cleaved (products) and uncleaved (substrates) RNA oligos were separated on a 20% denaturing urea acrylamide gel (7 M urea, 1X TBE buffer). Migration of fluorescently labelled RNA oligos was visualized using the Odyssey® CLx (LI-COR) imaging system at 700 nm. Images were converted to grayscale using Image Studio.

Table 2.15: Endonuclease V assay

Reagent	Volume [µL]	Final concentration
1 µM RNA oligo	1	0.1 µM; 1 pmol
40 mM divalent ions	1	4 mM
40 U/µL RiboLock RNase inhibitor	1	40 U
Human EndoV	variable	5 µM; 50 pmol
1X buffer	variable	1X
Total	10	

2.8 Western blotting

After SDS-PAGE, the separated proteins were transferred from the gel onto a nitrocellulose membrane for western blotting using a tank blotting system. The transfer was run at a constant current of 200 mA for 75 min with a block of ice in 1X transfer buffer. The membrane was blocked with 10 mL 3% (w/v) BSA in PBS-T at 4℃ with rocking and washed with PBS-T to remove the remaining BSA. The membrane was incubated with primary antibody or NiNTA-HRP

conjugate for 1 h at RT. All further steps were carried out at room temperature. To remove the primary antibody or NiNTA-HRP conjugate, the membrane was washed three times with PBS-T for 5 min whilst shaking. The secondary antibody was diluted in 5 mL 3% (w/v) BSA in PBS-T. Next, the membrane was incubated for 30 min whilst rocking. After incubation with the secondary antibody, the membrane was washed three times with PBS-T for 5 min to remove the secondary antibody. The membrane was further drained to eliminate the excess PBS-T and was then covered with freshly prepared ECL solution for NiNTA-HRP western blots. After a short incubation of 1 to 5 min, western blot images were obtained.

2.9 SEC-SAXS

All samples destined for SEC-SAXS were set up at a protein concentration of 5 mg/ml. For samples with RNA, RNA was added in a 1:2 molar ratio of RNA to protein and incubated on ice for 1 h to form *in vitro* RNA-protein complexes. All samples were loaded onto an S200 2.4 mL column at the B21 beamline at Diamond Light Source (Harwell). The resulting datasets were analysed using BioXTAS RAW. The radius of gyration (Rg) was estimated across the peaks after buffer subtraction. Selected frames from regions with a stable radius of gyration were averaged for further analyses of Guinier plots, P(r) functions, molecular weights from Porod volumes and Krakty plots.

3 Establishing an *in vitro* A-to-I RNA editing assay

This chapter aimed to reconstitute an editing reaction using *in vitro* transcribed RNA templates to test the hypothesis that *D. melanogaster* Zn72D might be required for editing of particular RNA editing sites and to solve the mechanism of RNA editing on a molecular level. To visualise adenosine to inosine conversions *in vitro* and to show if a cooperative interaction between Zn72D and ADAR at the site of editing is required, a fluorescence-based *in vitro* A-to-I RNA editing was established and optimised. One essential aspect of this chapter focuses on producing the recombinant protein components needed for this assay: ADAR and Zn72D/NF45. ADAR from *D. melanogaster*, the crucial enzymatic component for an editing assay must be expressed in insect cells to incorporate an inositol hexakisphosphate molecule which is buried within the core and required for maintaining the catalytic activity of ADAR. On the contrary, Zn72D and NF45 were expressed in bacteria as they are not known to require important post-translational modifications.

3.1 Optimisation of conditions for expression of recombinant *D. melanogaster* Zn72D and NF45

Before this work, there were no data on protein expression and purification conditions of Zn72D and NF45 from *D. melanogaster*. One of the first tasks of this project was to test different conditions for protein production and then establish a robust protocol to make soluble protein for subsequent *in vitro* assays and structural studies.

Since the N-terminus of Zn72D is unstructured, the expression of full-length constructs was not expected to give high yields of soluble protein. Therefore, several truncated constructs of Zn72D were designed for individual pro-

tein expression in bacteria (figure 3.1). Most constructs lacked the unstructured N-terminal part of the protein, which was removed due to the typical difficult nature of expressing long and flexible regions. Other Zn72D constructs lacked their DZF domain, zinc fingers, or a small part of the C-terminus. Another reason for designing multiple constructs of the same protein is their possible use in binding studies or biochemical assays to get more information on which domain contributes to specific functions like the RNA binding. However, most Zn72D constructs still possess the three zinc finger domains required for nucleic acid-binding. NF45 was the full-length version since the DZF domain makes up almost the whole protein. All Zn72D and NF45 constructs carried either a 6xHis- or a 6xHis-GST-tag at their N-terminus.

Pull-down assays of these test expressions were used to monitor and compare their protein levels. All of these constructs were expressed in several bacterial expression strains (BL21(*DE3*), B834(*DE3*), BL21-CodonPlus (*DE3*)-RIPL, and T7 Express) that possess individual features advantageous for recombinant protein expression. All four strains are derivates from the historic *E. coli* B strain. BL21(*DE3*) strains are deficient in Lon (cytoplasm) as well as OmpT proteases (outer membrane) and contain the λDE3 prophage that encodes the gene for T7 RNA polymerase under the control of the lacUV5 promoter. IPTG is required to induce expression of the T7 RNA polymerase with the aim of high-level expression of recombinant genes cloned downstream of a T7 promoter. B834 is the parental strain for BL21 and is a methionine auxotroph. This allows specific labelling of target proteins with 35S-methionine and selenomethionine for crystallography. BL21-CodonPlus (*DE3*)-RIPL is a derivative strain from BL21 that is beneficial if codon bias is a problem. RIPL cells carry an additional plasmid with extra copies of tRNA genes for four rare codons (AGA = arginine (R), AUA = isoleucine (I), CCC = proline (P), CUA = leucine (L)). T7 Express is also a BL21 derivative that lacks the λ*DE3* prophage that can be activated under certain circumstances and thus cause unwanted cell lysis. Its T7 RNA polymerase is under the control of a lac operon, ensuring protein expression will only be initiated upon IPTG addition by inactivating the lac repressor.

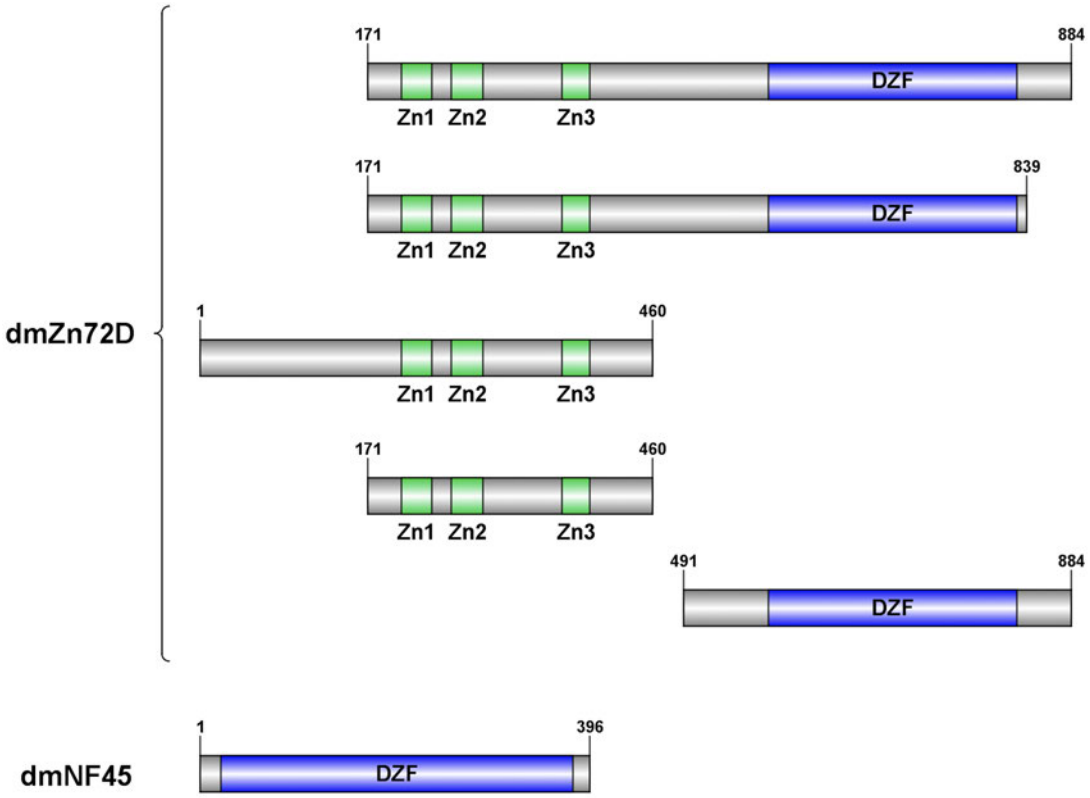
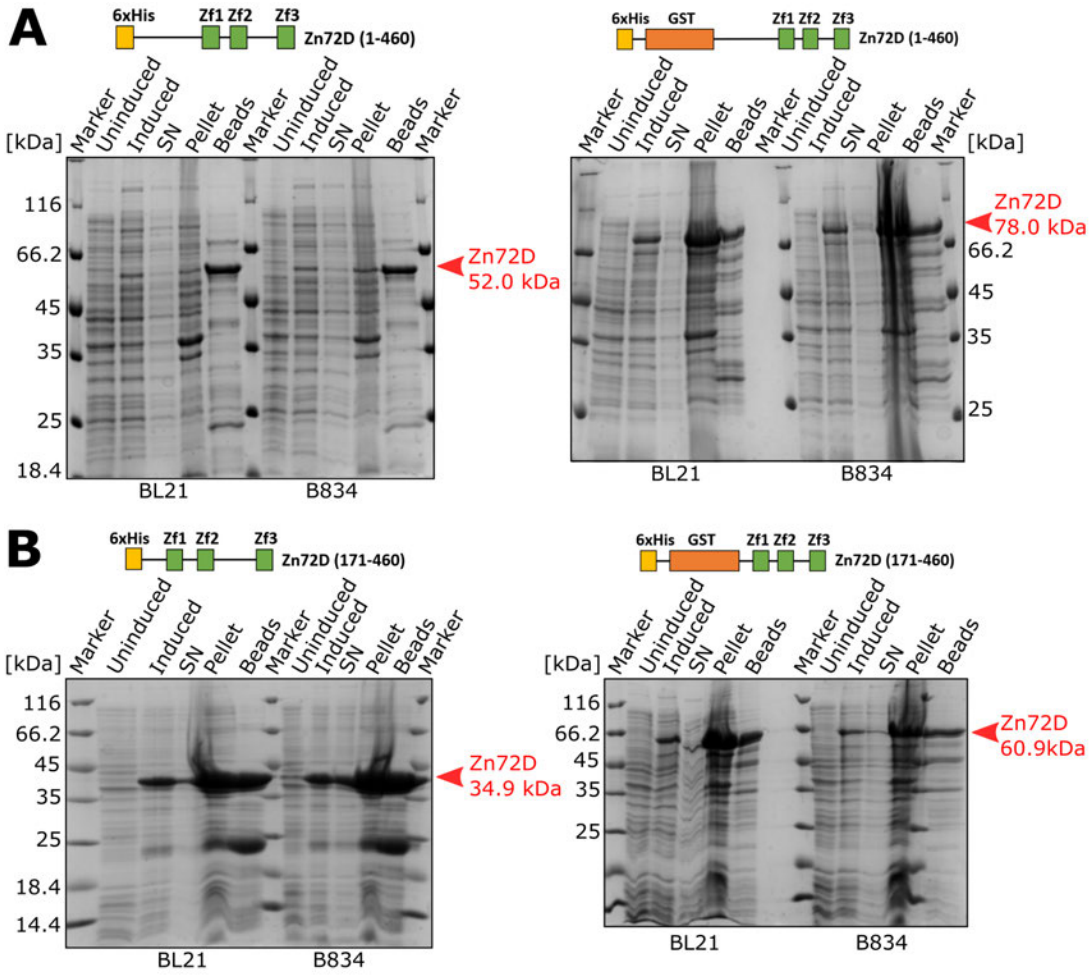


Figure 3.1: Overview of recombinant Zn72D and NF45 constructs used for expression in bacteria. Several constructs of truncated Zn72D and full-length NF45 were tested for protein production in bacteria. All constructs carry either an N-terminal 6xHis- or 6xHis-GST-tag for affinity purification that could be cleaved off by 3C proteases. The different versions of Zn72D and NF45 are shown with their domains, such as zinc fingers and the DZF domain required for heterodimerisation of Zn72D and NF45.

3.1.1 Test expressions of individual Zn72D constructs

Several truncated versions of Zn72D carrying an N-terminal 6xHis-tag or 6xHis-GST-tag were expressed in a panel of *E. coli* strains including BL21, B834, T7 and RIPL. Constructs, for example, that correspond to Zn72D from amino acids 171 to 884 are denoted here "Zn72D (117-884)" and accordingly to other truncated constructs. The constructs lacking a DZF domain, Zn72D (171-460) (figure 3.2A) and Zn72D (1-460) (figure 3.2B), showed good levels of soluble protein. In contrast, other Zn72D constructs showed poor levels of expression and limited solubility such as Zn72D (491-884) (figure 3.2C). These observations indicate that the Zn72D DZF domain might have a destabilising effect on the construct. Of the long constructs, the His-tagged Zn72D (171-884) (figure 3.2D) and GST-tagged Zn72D (171-839) showed some low expression levels

of soluble protein. Two further observations from this collection of test expressions were that His-tagged Zn72D (171-460) was expressed better than the GST-tagged constructs and using either BL21 or B834 as expression strains did not influence the amount of protein expressed. As large proteins were not well expressed in *E. coli*, and some of the GST-tagged constructs had a molecular weight of over 100 kDa, this might explain the differences in protein expression levels. However, for some constructs the GST-tag might enhance their translation rate and solubility. The results of all test expressions of Zn72D constructs were summarised in table 3.1.



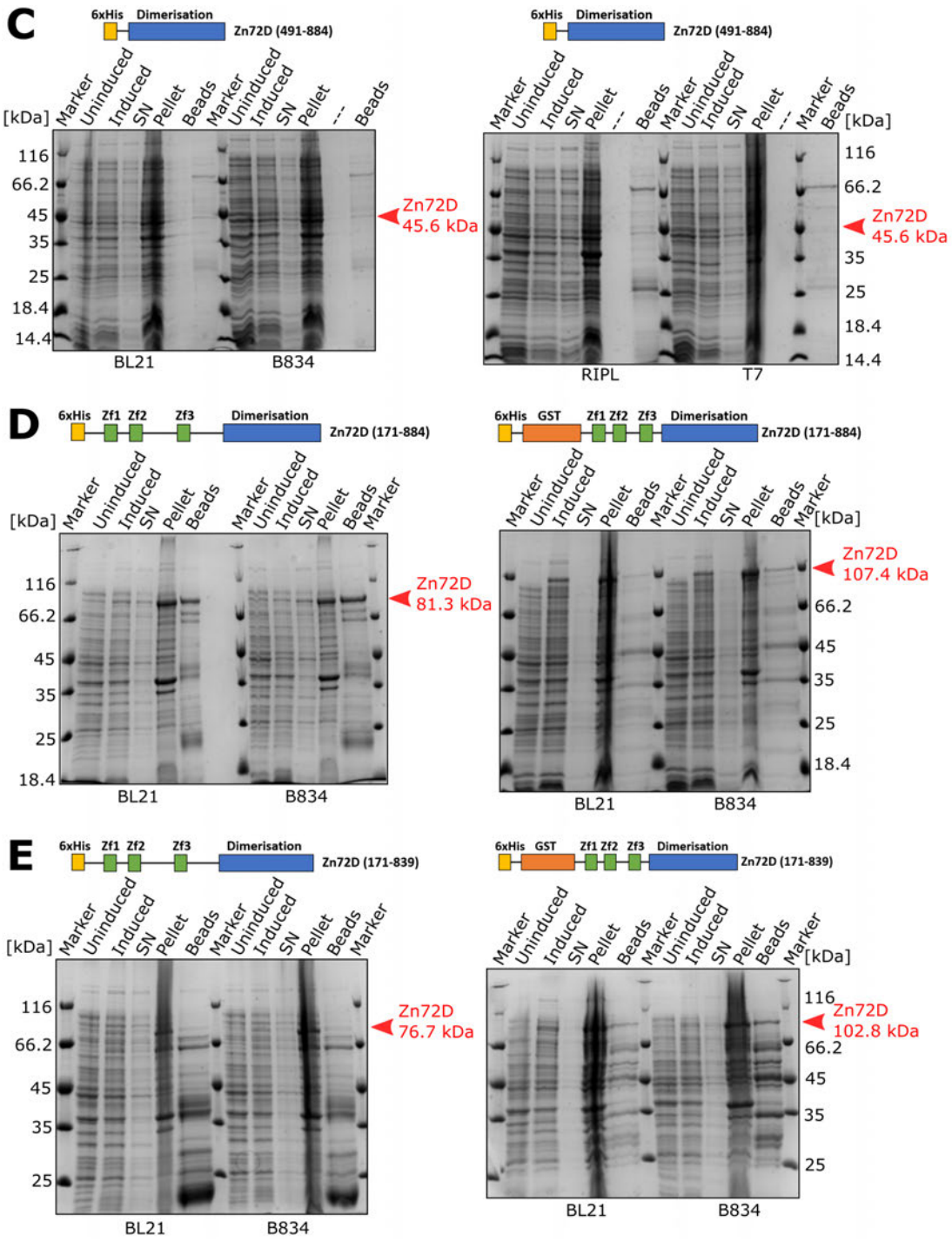


Figure 3.2: Test expressions of Zn72D constructs with DZF domain in bacteria. Several truncated versions of Zn72D were expressed in various bacterial strains such as BL21, B834, T7 and RIPL carrying either a 6xHis- or a 6xHis-GST-tag at their N-terminus to compare protein levels after cell lysis and protein enrichment using NiNTA or GSH beads. SDS-PAGE gels of A) His-tagged Zn72D (1-460), B) His- and GST-tagged Zn72D (171-460), C) His- and GST-tagged Zn72D (491-884), D) His- and GST-tagged Zn72D (171-884), E) His- and GST-tagged Zn72D (171-839). The specific bacterial strain used for protein expression is indicated below the gel. SN - supernatant.

3.1.2 Individual expression of full-length NF45

Full-length NF45 from *D. melanogaster* was expressed in BL21, B834, T7 and RIPL strains. The N-terminally 6xHis- and 6xHis-GST-tagged were highly expressed. However, contrary to the human orthologue, both constructs are insoluble and found in the pellet fraction (figure 3.3A). An additional construct of NF45 without a tag for potential co-lysis was tested (figure 3.3B). Since the untagged NF45 cannot be pulled-down with NiNTA or GSH beads, the expression levels were determined by the amount of protein in the soluble fraction after cell lysis and centrifugation (SN - supernatant). Expression of untagged NF45 did not show a significant improvement of protein solubility. The results of all individual test expressions of NF45 were summarised in table 3.1.

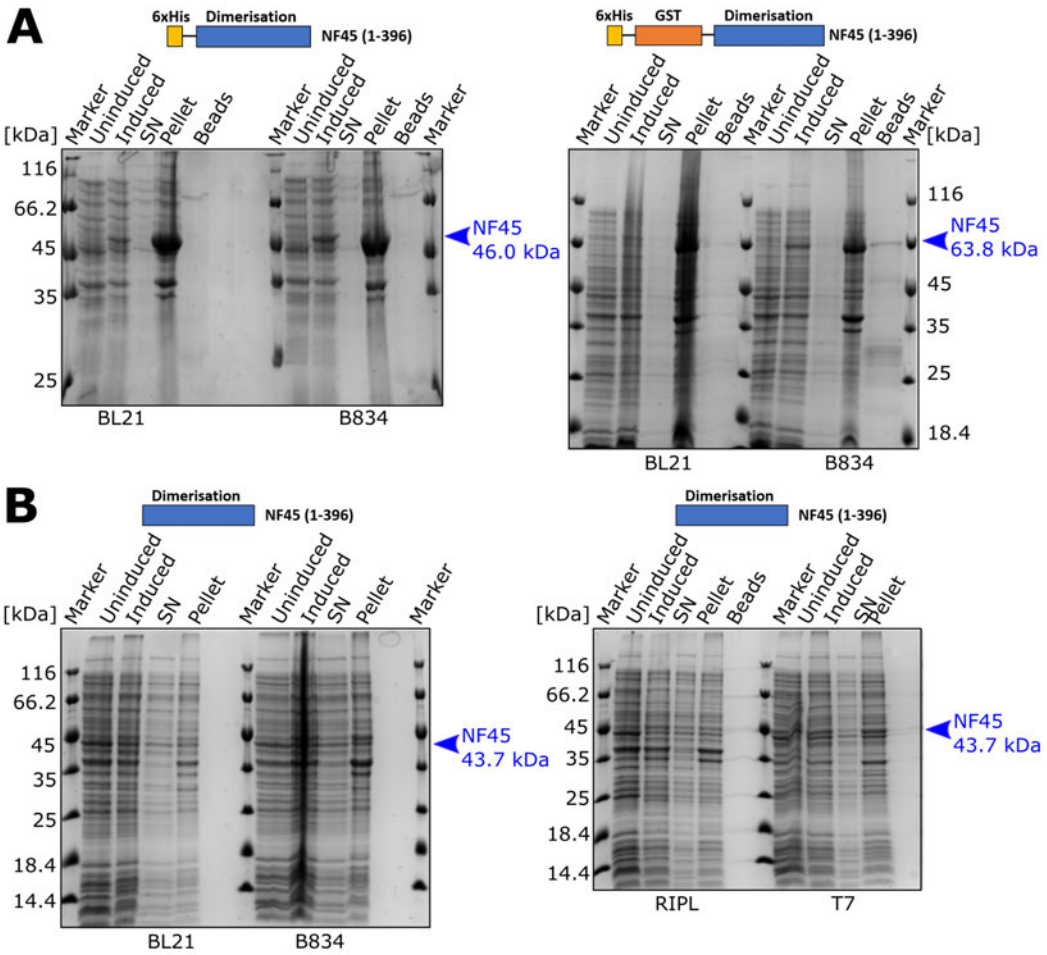


Figure 3.3: Test expressions of full-length NF45 in bacteria. Full-length NF45 was expressed in different bacterial strains (BL21, B834, T7 and RIPL). SDS-PAGE gels of A) 6xHis- or a 6xHis-GST-NF45 (1-396) and B) untagged NF45 (1-396). The specific bacterial strains used for protein expression are indicated below the gel. SN - supernatant.

Table 3.1: Overview of individual test expressions of different Zn72D and NF45 constructs

Protein	Tag	Start	End	Size [kDa]	BL21	B834	RIPL	T7	Expression	Solubility	Figure
Zn72D	6xHis	171	884	81.3	yes	yes	no	no	okayish	partially	3.2D
Zn72D	6xHisHis-GST	171	884	107.4	yes	yes	no	no	very low	very low	3.2D
Zn72D	6xHis	171	839	76.7	yes	yes	no	no	low	very low	3.2E
Zn72D	6xHis-GST	171	839	102.8	yes	yes	no	no	very low	partially	3.2E
Zn72D	6xHis	171	460	34.9	yes	yes	no	no	very high	yes	3.2B
Zn72D	6xHis-GST	171	460	60.9	yes	yes	no	no	high	yes	3.2B
Zn72D	6xHis	1	460	52.0	yes	yes	no	no	good	yes	3.2A
Zn72D	6xHis-GST	1	460	78.0	yes	yes	no	no	good	yes	3.2A
Zn72D	6xHis	491	884	45.6	yes	yes	yes	yes	very low	no	3.2C
NF45	6xHis	1	396	46.0	yes	yes	no	no	very low	no	3.3A
NF45	6xHis-GST	1	396	63.8	yes	yes	no	no	-	-	3.3A
NF45	-	1	396	43.7	yes	yes	yes	yes	yes	no	3.3B

3.1.3 Co-expression of Zn72D and NF45

Previous observations indicate that neither Zn72D nor NF45 constructs containing DZF domains are soluble since it might have a destabilising effect (figure 3.2 and 3.3). Since they are known to dimerise, it was hypothesised that they might need to co-fold for stability. Consequently, Zn72D constructs containing a DZF domain were now co-expressed with NF45 (without affinity tag) in different strains (figure 3.4) to assess formation of heterodimerisation. Co-expression of two proteins using individual plasmids required different antibiotic resistances to select for cells carrying both plasmids simultaneously. In addition, the plasmids need to have two compatible replication origins since origins from the same group are not compatible as they compete for the same segregation machinery.

Upon co-expression of Zn72D (491-884) and NF45, both proteins could be expressed at low levels, indicating heterodimer formation. However, heterodimer formation was only observed in B834 (figure 3.4A) and T7 cells (figure 3.4B) and neither in BL21 nor RIPL.

The same co-expression strategy was also applied to Zn72D (171-884) (figure 3.4C) and Zn72D (171-839) (figure 3.4D) constructs, containing all three zinc-fingers and the DZF domain. The amount of contaminating proteins for both pull-downs made an interpretation of the SDS-PAGE gel difficult. It seemed that both Zn72D (171-884) and Zn72D (171-839) could have co-expressed with NF45 in B834 cells (figure 3.4D), but if so showed only very low levels of soluble protein. An alternative expression strategy was applied because all four test expressions did not yield sufficient soluble protein levels for further large-scale expression: the open reading frames of Zn72D and NF45 were fused.

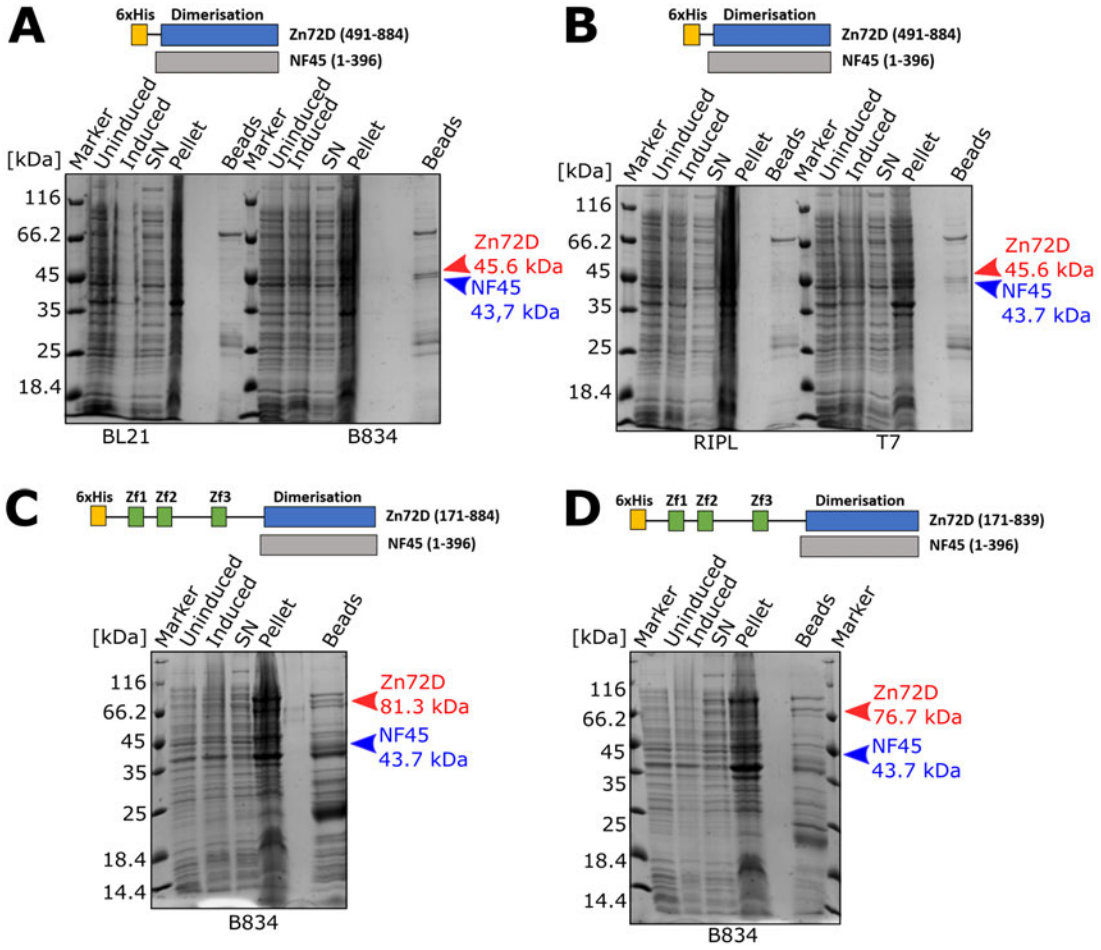


Figure 3.4: Test co-expressions of Zn72D and NF45 constructs in bacteria. SDS-PAGE gels of co-expression of 6xHis-Zn72D (491-884) and NF45 in A) BL21 and B834, B) RIPL and T7, C) 6xHis-Zn72D (171-884) with NF45 in B834 and D) 6xHis-Zn72D (171-839) with NF45 in B834. The specific bacterial strains used for protein expression are indicated below the gel. SN - supernatant.

3.1.4 Expression of a Zn72D and NF45 fusion protein

Neither individual expression of Zn72D and NF45 nor their co-expression succeeded in producing usable amounts of soluble protein (figures 3.2, 3.3, and 3.4). A possible way to increase their low expression levels and solubility could be the expression of both proteins as a single open reading frame with an N-terminal GST-tag, thereby, enhancing the stability of the corresponding fusion protein. A homology model of human ZFR superposed onto the structure of a complex of the dimerisation domains of NF90/NF45 showed spatial proximity of the C-terminus of ZFR and the N-terminus of NF45 (figure 3.5A). Since both termini were spatially close and could be connected by a flexible short linker,

the two open reading frames of both proteins (full-length Zn72D and full-length NF45) were combined into a single one. The C-terminus of Zn72D and the N-terminus of NF45 are connected via two TEV-cleavable linker sequences (figure 3.5B) for subsequent separation if required.

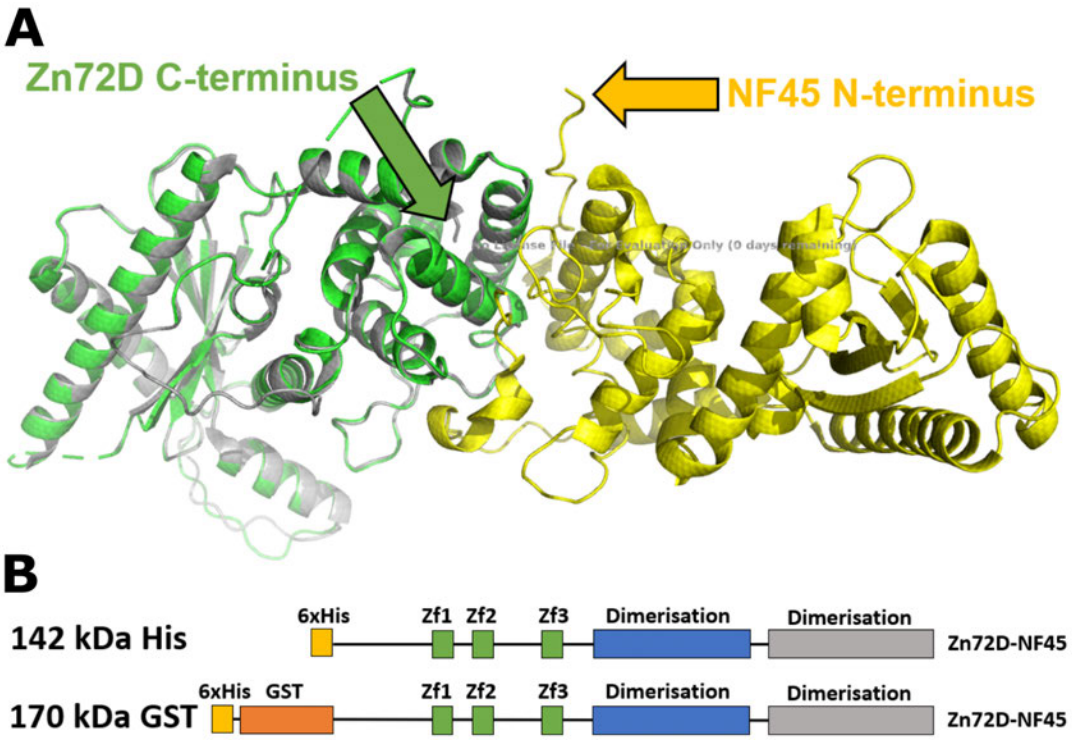


Figure 3.5: Design principle of Zn72D-NF45 fusion protein. A) A superposed homology model of ZFR (in green) using Phyre2 onto the structure of a complex of the dimerisation domains of NF90/NF45 (NF90 in gray and NF45 in yellow) from mouse (PDB entry: 4AT7) showed spatial proximity of the C-terminus of ZFR and the N-terminus of NF45. B) The final Zn72D-NF45 fusion constructs for protein production in bacteria. Both full-length Zn72D and NF45 are separated via a short heptapeptide linker with the sequence GSSGSSG followed by a TEV cleavage site and another decapeptide linker with the sequence GGGGSGGGS.

Expression of 6xHis- and 6xHis-GST-tagged constructs of the fusion proteins was tested in a panel of *E. coli* strains (figure 3.6). Expression levels of the His-tagged fusion protein (His) were significantly lower than the His-GST-tagged fusion protein (GST), which showed soluble protein production for expression in BL21, B834, and RIPL, but not in T7. Therefore, B834 cells were chosen as the bacterial strain for further protein production on a large scale. The Zn72D-NF45 fusion might be a case where GST enhances solubility of the construct in contrast to large proteins usually being not well expressed in *E. coli*.

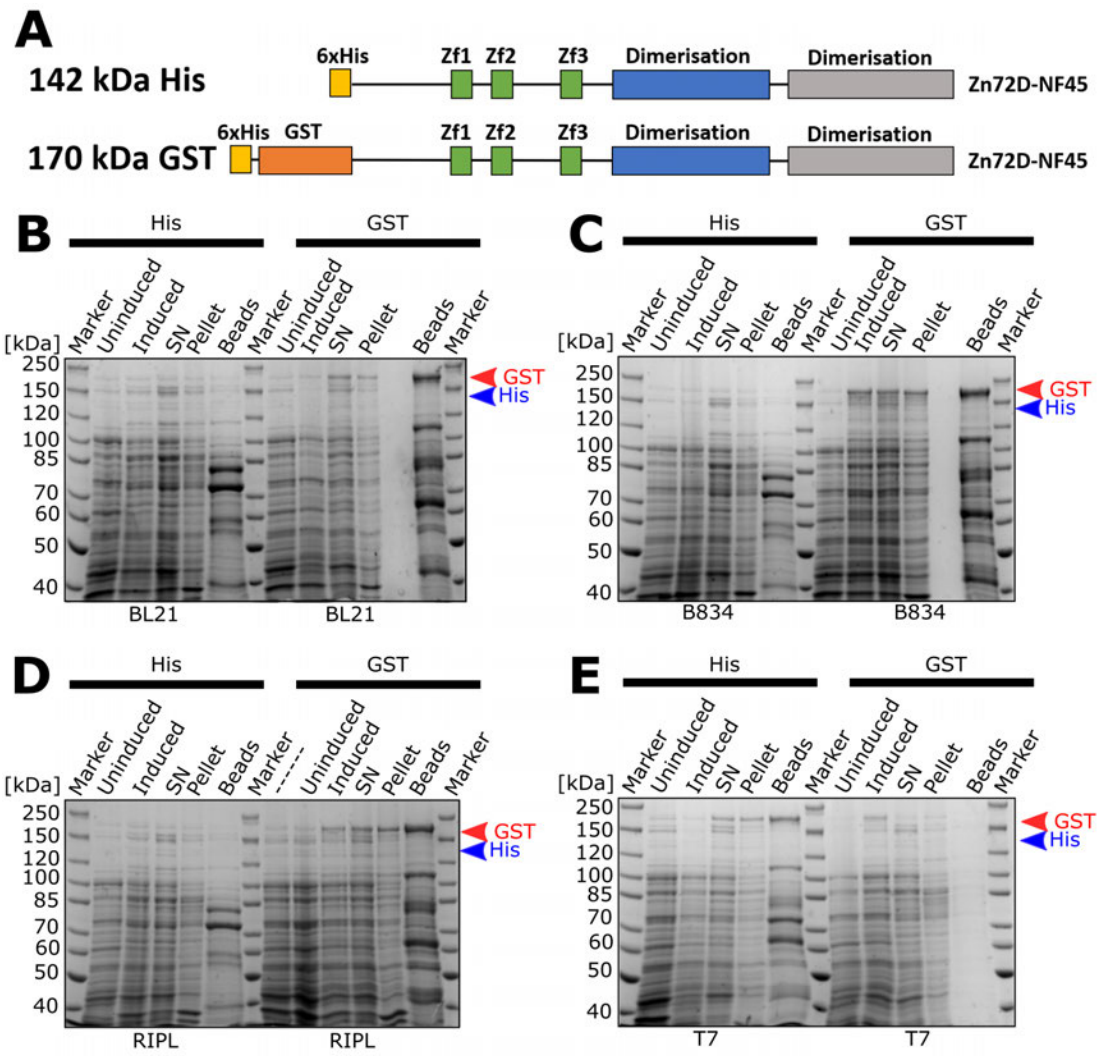


Figure 3.6: Test expressions of Zn72D-NF45 fusion protein. SDS-PAGE gels of (A) His- and GST-tagged Zn72D-NF45 fusion protein test expressions in several bacterial strains: B) BL21, C) B834, D) RIPL, and E) T7. The specific bacterial strain used for protein expression is indicated below the gel. SN - supernatant. His- and GST-tagged fusion proteins are indicated by arrow heads.

Size exclusion chromatography during large-scale purification of the His-GST-tagged fusion protein in B834 cells showed that the yields of protein were meagre (figure 3.7) for using 3 L of bacterial cell culture. The low yield of recombinant protein in peak 1 might be a consequence of low expression levels in bacteria due to its size of 170 kDa. Also, the TEV cleavage site seemed to be partially inaccessible as even a significant excess of GST-TEV protease could not cleave all fusion proteins overnight in solution into a Zn72D and an NF45 fragment (figure 3.7). Another issue encountered during the purification process was that the fusion protein remained bound to GSH beads (figure 3.7)

used for affinity chromatography and made it difficult to elute the majority of the bound fusion protein. Hence, the next step in producing soluble Zn72D/NF45 was to rely again on a co-expression strategy as this might mitigate the issues arising from expressing and purifying a huge fusion protein.

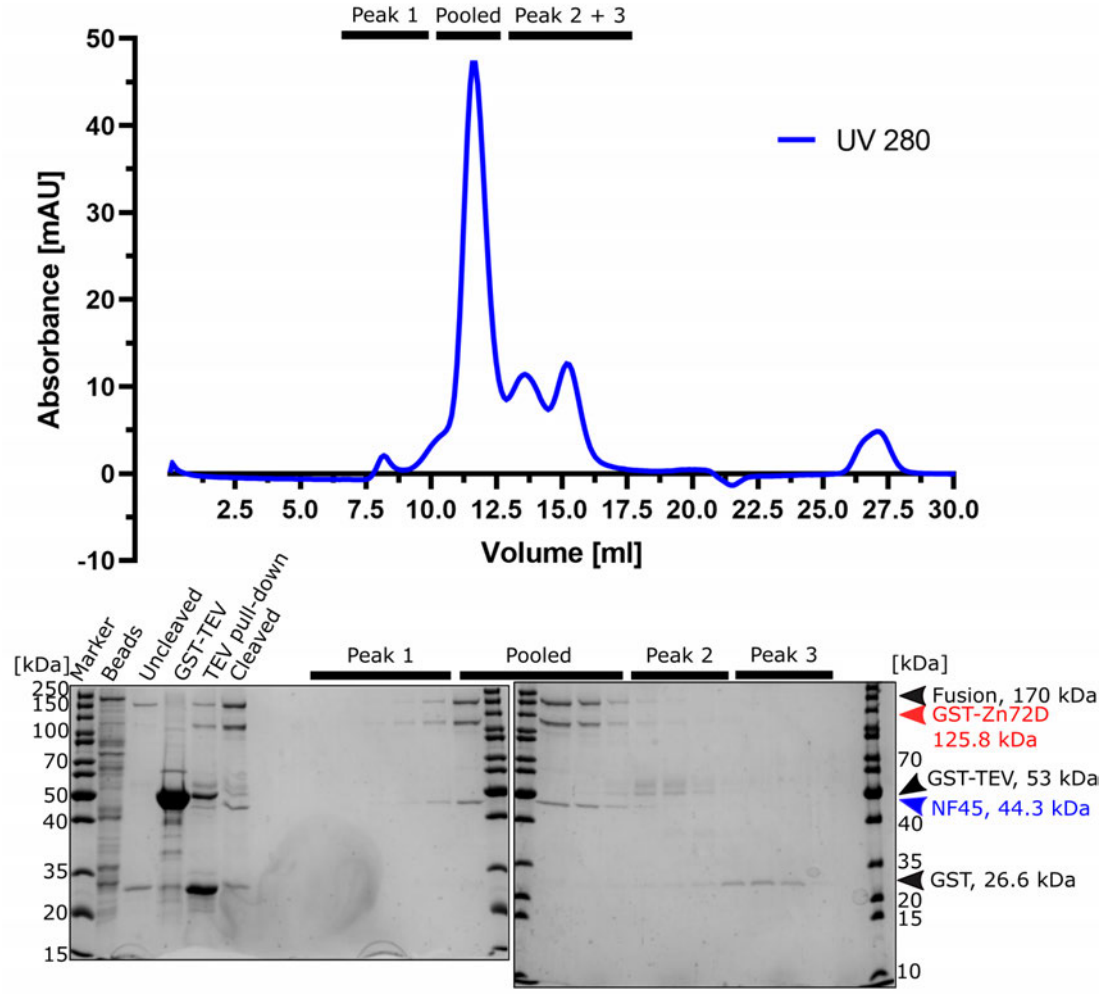


Figure 3.7: Size exclusion chromatography of Zn72D-NF45 fusion protein purification. Size exclusion chromatogram with absorbance at 280 nm in blue and corresponding SDS-PAGE gels. All relevant proteins and their sizes are indicated by arrow heads.

3.2 Optimisation of co-expression conditions

3.2.1 Effect of IPTG concentration on protein production

Multiple screening trials of different conditions of either individual or co-expression of Zn72D and NF45 did not produce enough soluble protein for structural studies. Since the fusion protein purification is unyielding, the co-expression strategy was tested again. However, this time it became evident that the OD_{600} after overnight co-expression of Zn72D/NF45 was relatively low, resulting in very little biomass for further protein extraction. Therefore, deleterious effects on bacterial cell density might be due to the toxic expression of Zn72D or its RNA-binding activity. To test the toxic effect of protein expression, B834 cells were co-transformed with and without plasmids for protein co-expression of His-tagged Zn72D and GST-tagged NF45. LB agar plates containing a titration of IPTG concentration were used for plating the transformed cells. Since there was a decreasing number of colonies on agar plates with an inversely proportional concentration of IPTG, the effect of IPTG concentration on protein expression levels was tested in pull-downs (figure 3.8). Since this effect was not observed to such an extent with the full-length Zn72D-NF45 fusion protein and the protein expression levels are high under low IPTG conditions, it can only be speculated about the exact reason. However taking these observations into account, the deleterious effect on bacterial biomass seems to be caused by the rate of protein expression rather than the presence or amount of the individual or fusion proteins.

Testing four different concentrations of IPTG for inducing protein expression showed that there is indeed a relationship between IPTG concentration and the density and biomass of bacterial cells, which might ultimately determine the amount of soluble protein. For example, cell density and the amount of soluble protein were very low for the higher IPTG concentrations (figure 3.8B) but increased significantly with lower IPTG concentrations (figure 3.8C). Therefore, subsequent co-expression of Zn72D/NF45 was induced with 0.025 mM IPTG.

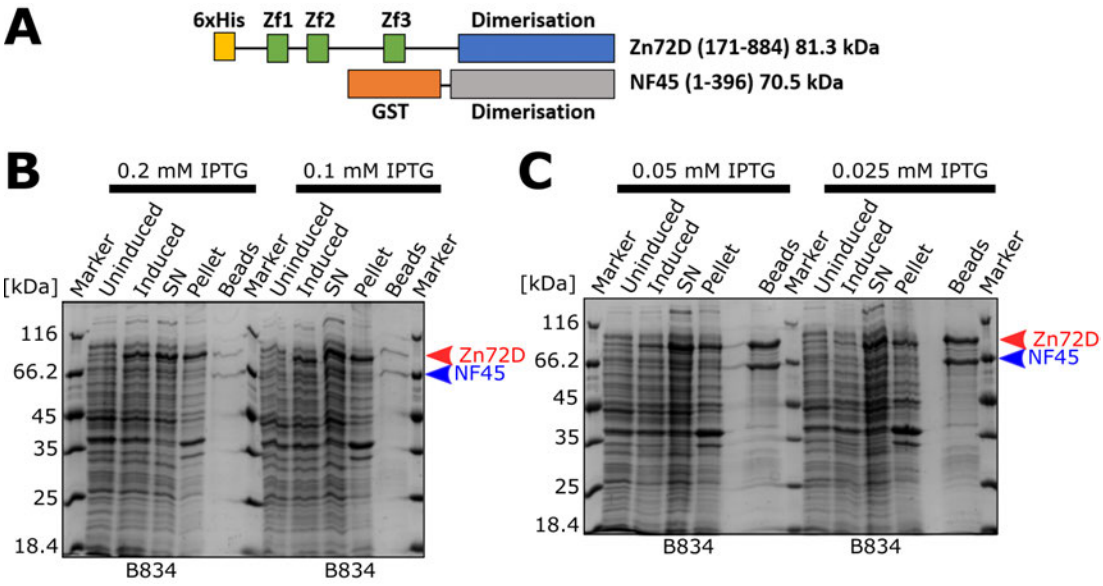


Figure 3.8: Co-expression of Zn72D/NF45 with low IPTG concentrations increases bacterial biomass. A) His-Zn72D and GST-NF45 constructs were co-expressed in B834 for testing impact of IPTG on biomass and protein expression levels. B) and C) SDS-PAGE gels of co-expression trials of His-Zn72D and GST-NF45. The specific bacterial strain used for protein expression is indicated below the gel. SN - supernatant. Zn72D and NF45 are indicated by arrow heads.

3.2.2 Lysis buffer optimisation

The expression levels of recombinant Zn72D/NF45 were much improved after lowering the IPTG concentration (section 3.2.1). Lysis buffer optimisation was performed to enhance the protein yield further, and the 260/280 ratio of the eluted fractions was measured using a NanoDrop to quantify nucleic acid contamination. Figure 3.9A and B show that there is a significant increase in the amount of soluble His-tagged Zn72D with increasing salt concentrations in the lysis buffer (250, 500, 750, and 1000 mM). High NaCl concentrations of 750 and 1000 mM seem to give the highest protein levels.

Since there is no further increase of protein levels detectable at 1000 mM NaCl, 750 mM was chosen for lysis buffer to be tested with increasing concentrations of urea (figure 3.9B and table 3.2). Following advice from Dr Sutapa Chakrabarti, urea was added to the lysis buffer for RNA removal. In her experience, concentrations of 50 to 200 mM were not harmful to the protein she used to work with and proved effective in RNA removal. There is a decrease in the amount of RNA contamination if the NaCl concentration is 750 mM while the amount of urea in the lysis buffer is 50 mM (figure 3.9B, C and table 3.2). In-

creasing urea concentrations seems to adversely affect RNA contamination as the 260/280 ratio increases corresponding to a higher amount of nucleic acids. To summarise these protein production and lysis optimisations, Zn72D/NF45 lysis buffers have been set up with a high salt concentration (750 mM) and 50 mM urea, and dialysis buffers contained more salt than before as the concentration was increased to 100 mM.

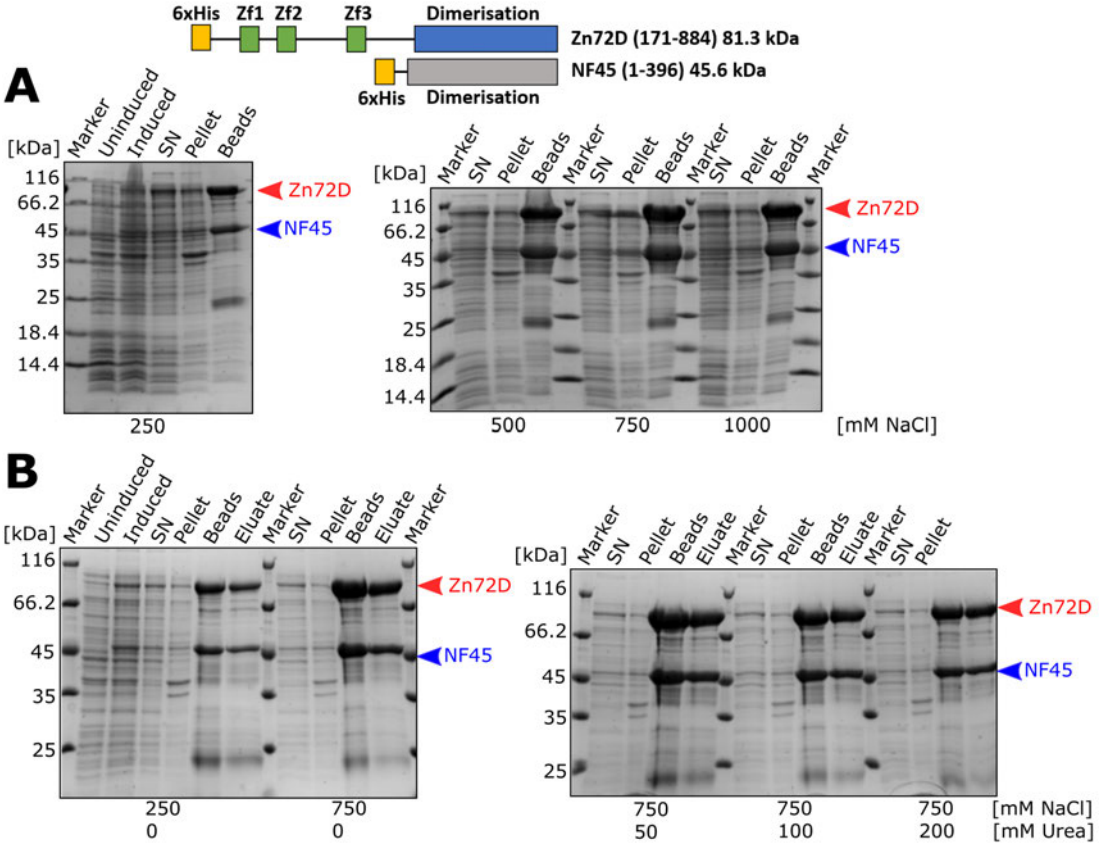


Figure 3.9: Co-expression of Zn72D and NF45 for lysis buffer optimisation. A) and B) Co-expressed His-tagged Zn72D (indicated by red arrow head) and His-tagged NF45 (blue) were lysed with different concentrations of NaCl (250, 500, 750, and 1000 mM) in the buffer. B and C) Zn72D/NF45 was lysed with different concentrations of urea in buffer containing 750 mM NaCl.

Table 3.2: Decrease of RNA contamination of Zn72D/NF45 purification after lysis buffer optimisation measured by Nanodrop

NaCl [mM]	750	750	750	750
Urea [mM]	0	50	100	200
260/280	1.36	1.03	1.50	1.50

3.3 Protein purification of Zn72D/NF45 heterodimers

The findings from the previous lysis buffer optimisation (section 3.2.2) were used as a starting point for purifying the *D. melanogaster* complex comprised of His-tagged Zn72D and GST-tagged NF45. Following cell lysis and separation of soluble and insoluble components of the lysate by centrifugation, the supernatant was incubated with GSH beads before being packed into a column (section 2.5.4). The column was then washed with GSH buffer A to remove any non-specific bound proteins until it reached a steady UV baseline. Removal of unbound (Unbound) and non-specific bound proteins (Wash) are shown in the chromatogram and the corresponding SDS-PAGE gel (figure 3.10). Bound proteins were eluted with 100% GSH buffer B containing 30 mM reduced glutathione, until UV signal reached baseline again. A single peak containing Zn72D/NF45 was observed upon elution. The gel shows a relatively clean eluate with little free GST (26.6 kDa). The indicated pooled fractions of the eluate had the lowest 260/280 ratio. The GST-tag of NF45 was cleaved using GST-TEV protease prior to immobilised metal affinity chromatography for further purifcation.

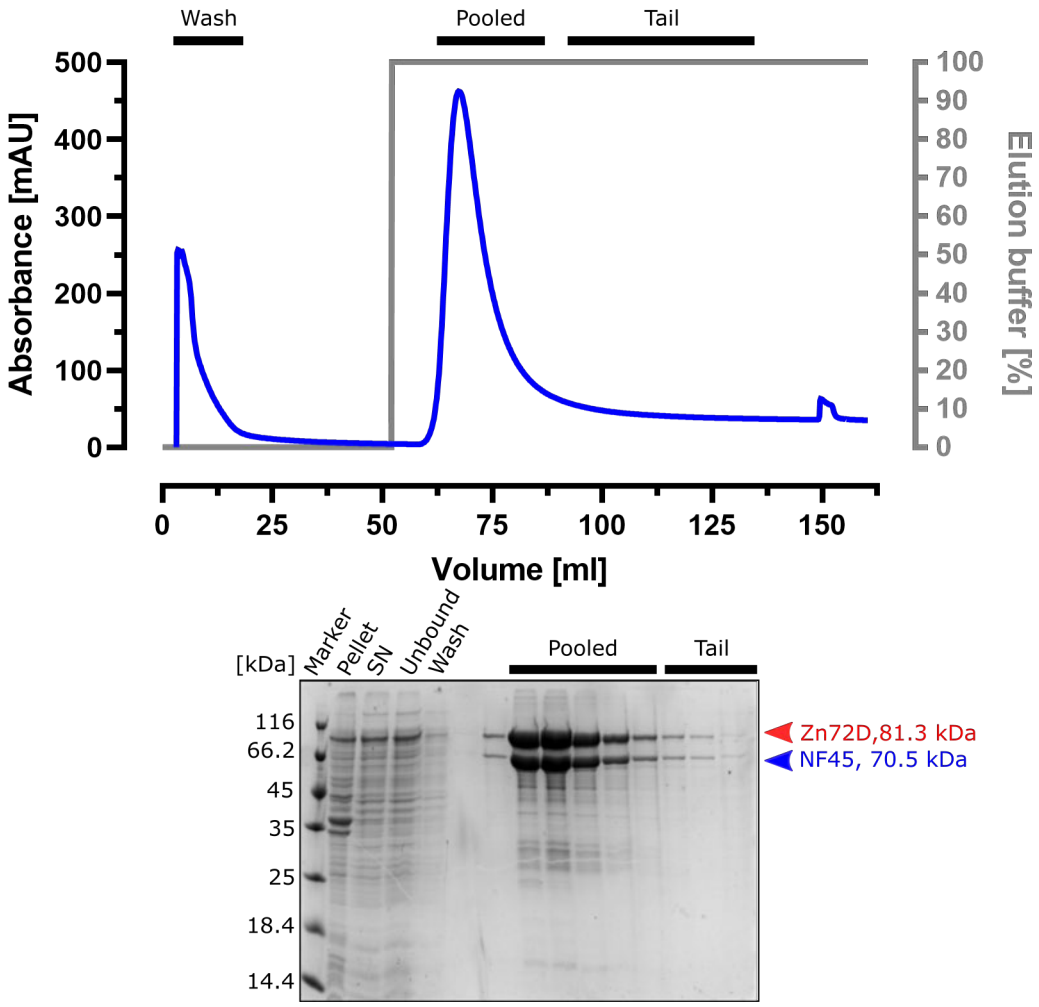


Figure 3.10: Affinity chromatography of Zn72D/NF45 using GSH beads. The chromatogram shows the absorbance at 280 nm (in blue) and the elution buffer ([in %] in grey). The corresponding SDS-PAGE gel shows proteins and their sizes indicated by arrow heads. SN - supernatant.

After GST-tag cleavage, Zn72D/NF45 was loaded onto a HisTrap column. Samples before and after TEV cleavage were analysed by SDS-PAGE (figure 3.11) and showed partial removal of the GST-tag of NF45. The cleaved NF45 band migrates at a lower molecular weight, close to the 45 kDa band of the marker. There is also a new band present after TEV treatment, representing cleaved GST (26.6 kDa). The column was washed with NiNTA buffer A until it reached a steady UV baseline. Bound Zn72D/NF45 was eluted with a linear gradient of NiNTA buffer B to 90%. A single peak containing Zn72D/NF45 was observed upon elution. Upon further increase of the concentration of elution buffer, no more protein could be eluted from the column. The SDS-PAGE gel shows a relatively pure product with free GST and GST-TEV (53 kDa). The His-tag of Zn72D was cleaved using GST-3C, and more GST-TEV was added

to cleave the remaining GST-tagged NF45.

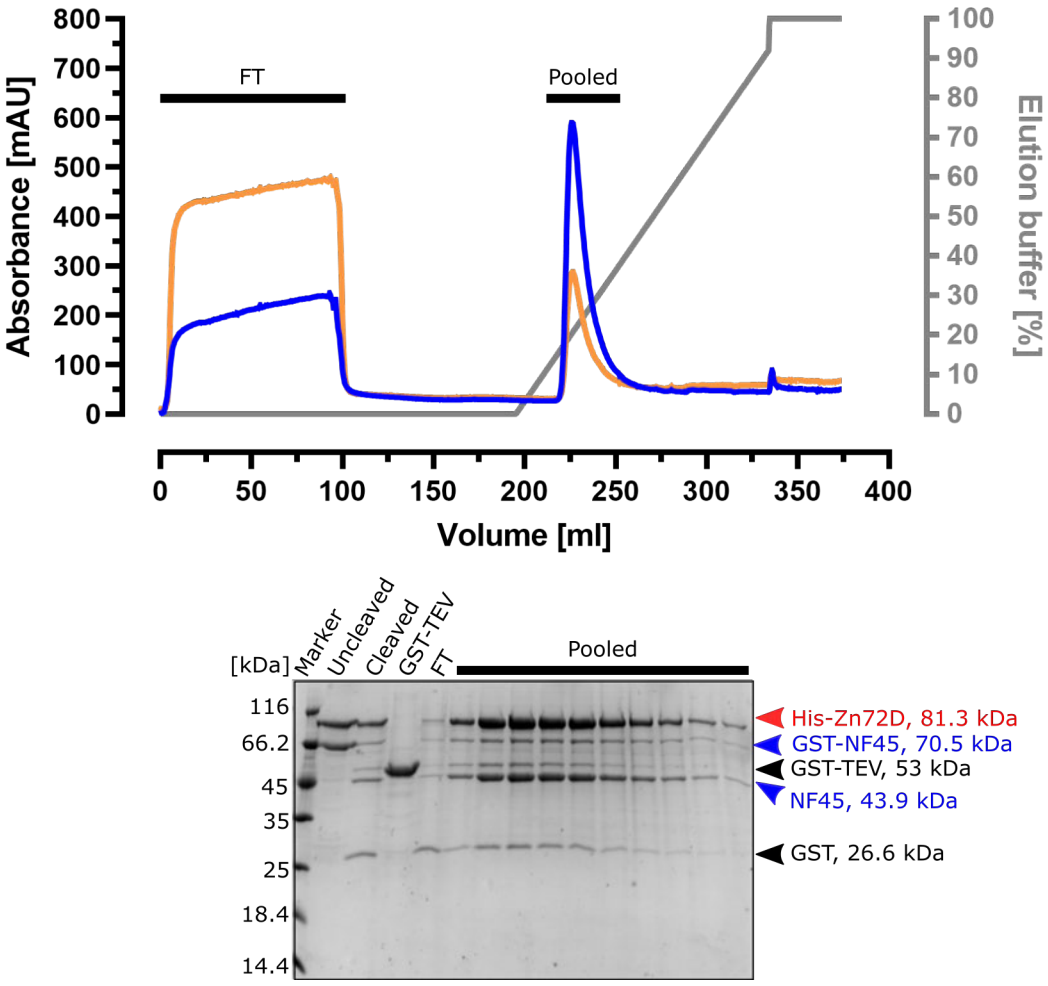


Figure 3.11: First immobilised affinity chromatography of Zn72D/NF45. Chromatogram of the first IMAC shows absorbances at 280 nm (in blue), at 254 nm (in orange) and the elution buffer ([%] in grey). The corresponding SDS-PAGE gel shows proteins and their sizes indicated by arrow heads. FT - flow-through.

Zn72D/NF45 was re-loaded onto a HisTrap column to separate Zn72D with and without His-tag, and samples before and after 3C cleavage were analysed by SDS-PAGE. Cleavage of GST-tagged NF45 had reached completion as there was no GST-tagged NF45 visible (figure 3.12). The column was washed with NiNTA buffer A until it reached a steady UV baseline, resulting in a long peak containing NF45 and Zn72D in the flow-through fractions. Bound His-tagged Zn72D/NF45 was eluted with a 100% NiNTA buffer B and discarded upon elution. The SDS-PAGE gel shows a quite clean product with minor contamination of free GST (26.6 kDa), GST-3C (42 kDa), and GST-TEV (53 kDa)

contaminations. The indicated pooled fractions were used for subsequent heparin chromatography.

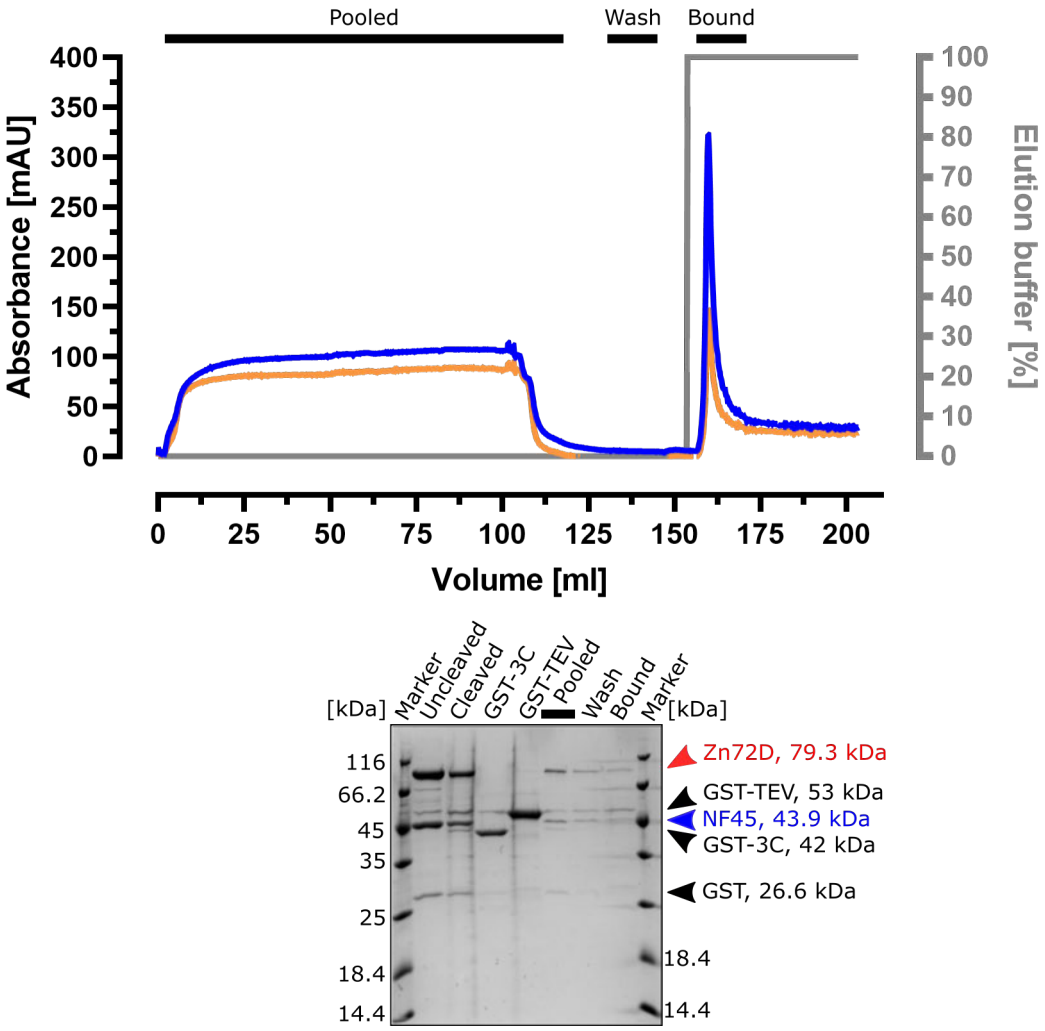


Figure 3.12: Second immobilised affinity chromatography of Zn72D/NF45. Chromatogram of the second IMAC shows absorbances at 280 nm (in blue), at 254 nm (in orange) and the elution buffer ([%] in grey). The corresponding SDS-PAGE gel shows proteins and their sizes indicated by arrow heads.

Zn72D/NF45 was loaded onto a heparin sepharose column to remove the contaminating nucleic acids. The corresponding chromatogram and SDS-PAGE gel (figure 3.13) indicate that contaminating nucleic acids could be separated from Zn72D/NF45 as the 260/280 ratio of the flow-through (FT) is above 1, while no protein is detectable by SDS-PAGE. First, the column was washed with heparin buffer A until it reached a steady UV baseline. Next, bound Zn72D/NF45 was eluted with a linear gradient of heparin buffer B to 100% (1000 mM KCl). A single peak containing Zn72D and NF45 was observed

upon elution, and the SDS-PAGE gel shows that the two proteins co-elute in fractions with a low 260/280 ratio (0.6). Therefore, the indicated pooled fractions were used for size exclusion chromatography.

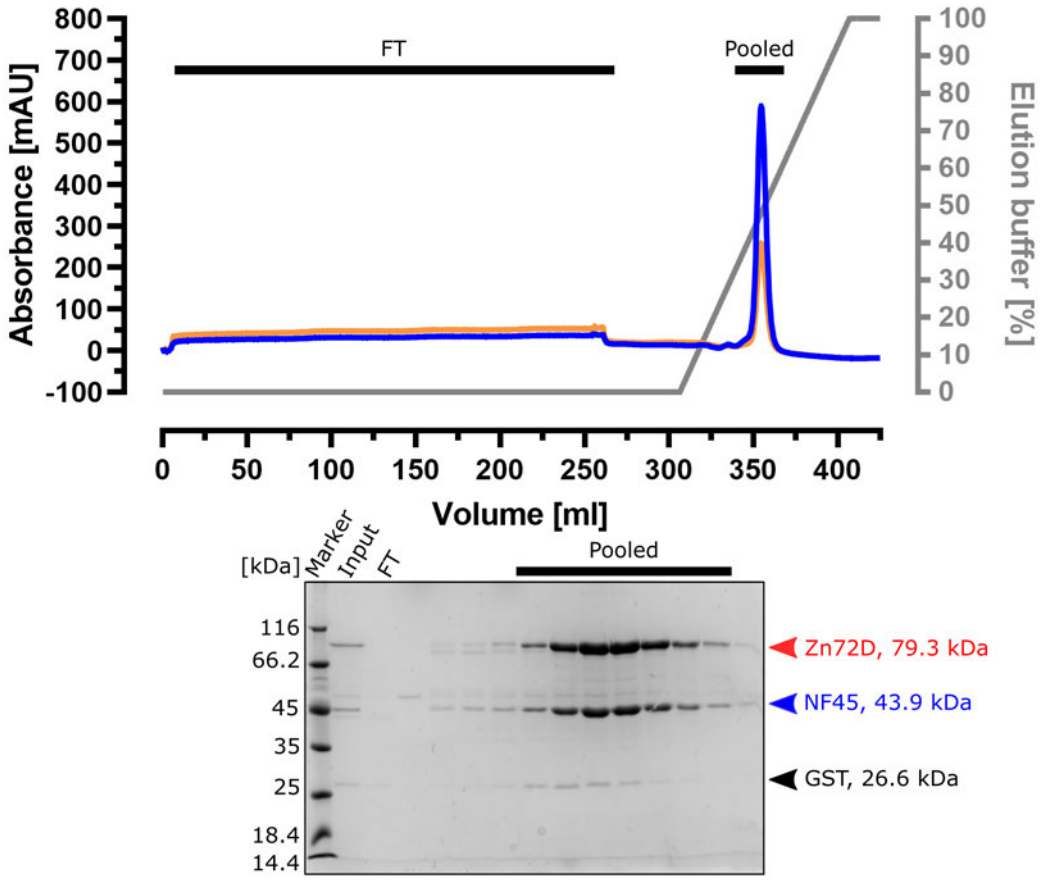


Figure 3.13: Heparin sepharose chromatography of Zn72D/NF45. Chromatogram of the heparin step shows absorbances at 280 nm (in blue), at 254 nm (in orange) and the elution buffer ([%] in grey). The corresponding SDS-PAGE gel shows proteins and their sizes indicated by arrow heads. FT - flow-through.

After heparin sepharose chromatography, Zn72D/NF45 was loaded onto a 24 ml S200 column. A single peak containing Zn72D/NF45 was observed in the chromatogram and the corresponding gel (figure 3.14) upon elution. The SDS-PAGE gel shows that the two proteins co-elute in fractions with a low 260/280 ratio (0.6). The final yield was typically around 2 mg of pure protein per 6 L of bacterial cells. The protein sample was flash-frozen in liquid nitrogen for further experiments.

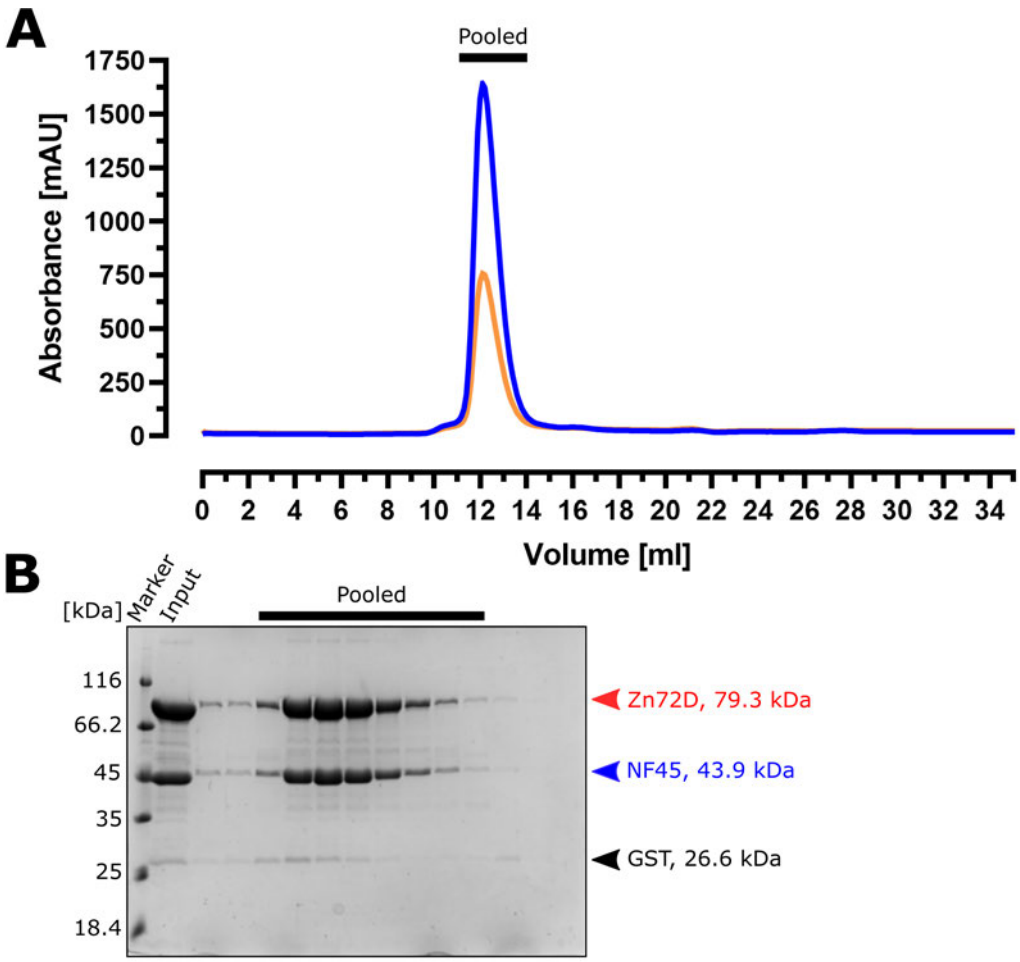


Figure 3.14: Size exclusion chromatography of Zn72D/NF45. Chromatogram of the size exclusion step shows absorbances at 280 nm (in blue) and at 254 nm (in orange). The corresponding SDS-PAGE gel shows proteins and their sizes indicated by arrow heads.

3.4 Optimisation of conditions for expression of recombinant *D. melanogaster* ADAR

To express ADAR from *D. melanogaster* in insect cells, it was first cloned into two different helper plasmids, pFastBac (figure 7.4) and pFL-GST (figure 7.4), to incorporate either an N-terminal hexahistidine (6xHis) (pFastBac) or a GST (pFL-GST) tag to the protein. These donor plasmids were then used to generate a bacmid containing the integrated ADAR gene for the infection of insect cells and for producing the virus stock. As the bacmids also encode for yellow fluorescent protein (YFP), the number of infected insect cells can be estimated using a fluorescent microscope. YFP fluorescence helps determine a suitable

virus concentration that produces high levels of proteins without a high rate of infection that prevents cell growth. Possible considerations for test expressions in Sf9 cells are the specific virus stock, the virus concentration, the length of incubation after infection, or the tag attached to the protein. Harvesting cells after two to three days with viability over 90% gave the best results for protein expression and was usually independent of the virus stock used for infection.

The expression of ADAR was determined by a pull-down assay (figure 3.16). SDS-PAGE confirmed that both His- and GST-ADAR could be expressed as soluble proteins that can be enriched on NiNTA or GSH beads. However, His-ADAR seemed to be expressed at higher levels than GST-ADAR, but it also shows a much higher background of endogenous proteins with at least two more major bands of unknown origin at around 60 and 100 kDa (figure 3.16A). Enrichment of His-tagged ADAR using NiNTA beads seems to pull down these additional insect cell proteins as they might bind to NiNTA. GST-ADAR, on the other hand, seemed to be expressed in relatively low levels with a strong band of free GST at around 25 kDa (figure 3.16B), indicating partial translation or degradation of the protein construct.

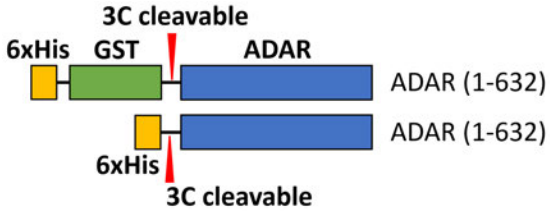


Figure 3.15: ADAR constructs used for protein purification. His- and His-GST-ADAR constructs used for protein expression and purification. Protease cleavage sites (3C) are indicated by a red arrowhead.

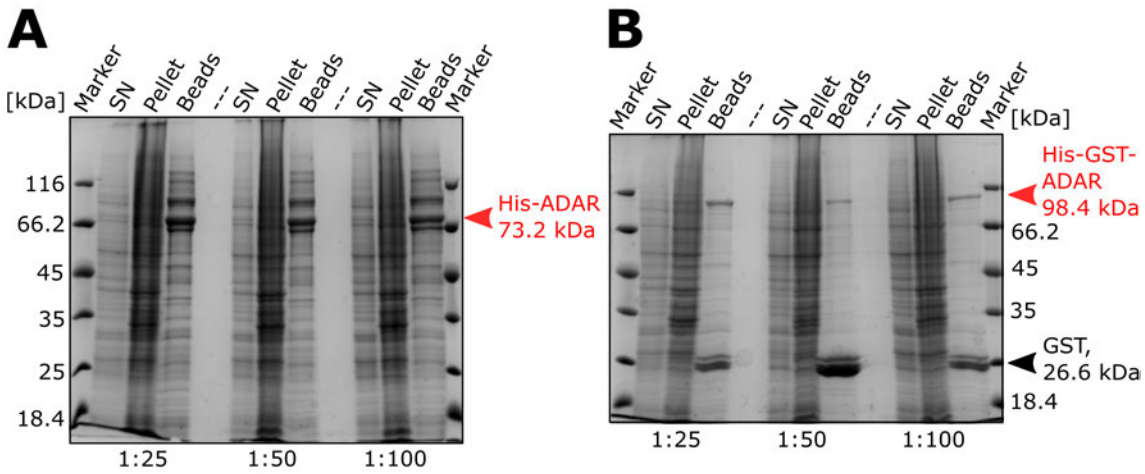


Figure 3.16: Test expressions of His- and His-GST-tagged ADAR in Sf9 cells A) Test expression of virus stock V2 of His-ADAR using different ratios of viral stock to cell culture volume as indicated below. His-ADAR has a size of about 73.2 kDa. B) Test expression of virus stock V2 of His-GST-ADAR with different ratios. His-GST-ADAR has a size of about 98.4 kDa.

Based on these test expressions (figure 3.16A), virus stock V2 of His-ADAR in a ratio of 1:50 was used to set up a large-scale expression in 1.6 L insect cell medium with 2x 10^6 cells per ml. Contrary to the test expressions, the fluorescence of the large-scale expression was low, indicating low infection and the Sf9 cells were harvested two days after infection due to their decreasing viability. In addition, SDS-PAGE of the batch purification using NiNTA beads indicated that expression levels of ADAR were poor and that the level of ADAR eluted off the NiNTA beads did not match that expected from the test expression (figure 3.17A).

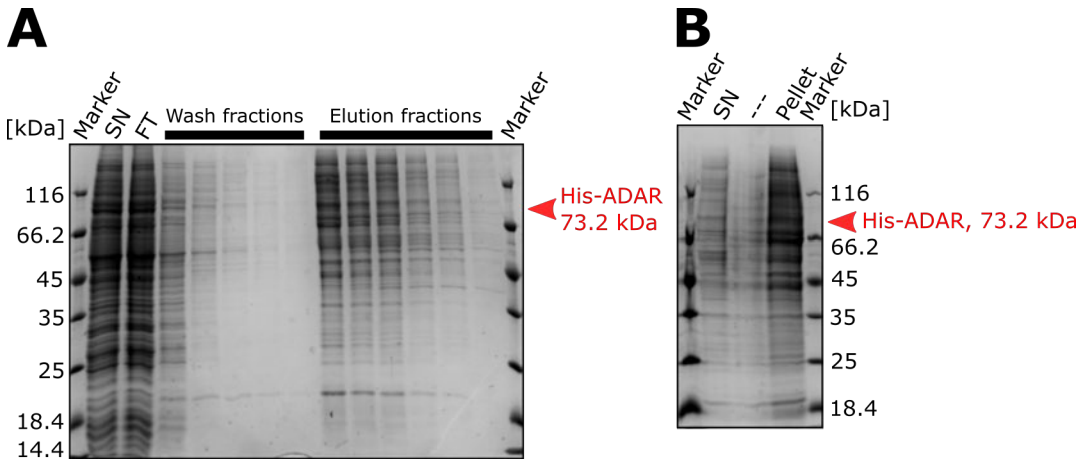


Figure 3.17: Batch purification of ADAR A) NiNTA batch purification of His-tagged ADAR (indicated by red arrow head) expressed in 1.6 L insect cells. B) ADAR sample after dialysis into low salt (50 mM NaCl).

However, since biochemical assays can be conducted with relatively little protein compared to that required for crystallisation, the purification continued, and ADAR was dialysed in 2 L of low-salt buffer (containing 50 mM NaCl) to remove excess imidazole and to lower the salt concentration in the buffer before using a heparin sepharose column. Nonetheless, some protein in the eluate precipitated, and it was not possible to re-solubilise ADAR upon adding more salt or other additives such as urea and zinc acetate. The most likely explanation is that the low salt conditions of the dialysis caused the precipitation of ADAR. Therefore, a small sample of the precipitation was taken for centrifugation and analysed on SDS-PAGE. Most proteins of the eluate (pooled elution fractions) precipitated in low salt buffer, including ADAR (figure 3.17B).

To further assess whether the bulk of ADAR was to be found in the soluble (SN) or precipitated fraction (Pellet), a western blot (figure 3.18) using NiNTA-HRP for the detection of His-tagged proteins was performed. Three different bands between 60 and 75 kDa (figure 3.18B) were found in the insoluble fraction (Pellet). These bands also seemed to reflect the same pattern of bands observed in the test expressions (figure 3.16A), which indicated that a significant amount of ADAR is either being degraded during cell lysis and purification or it is not being expressed homogeneously as the full-length protein.

Taking all these findings into account, the GST-tagged version of ADAR was expressed, known to have a lower background in insect cells (figure 3.16B).

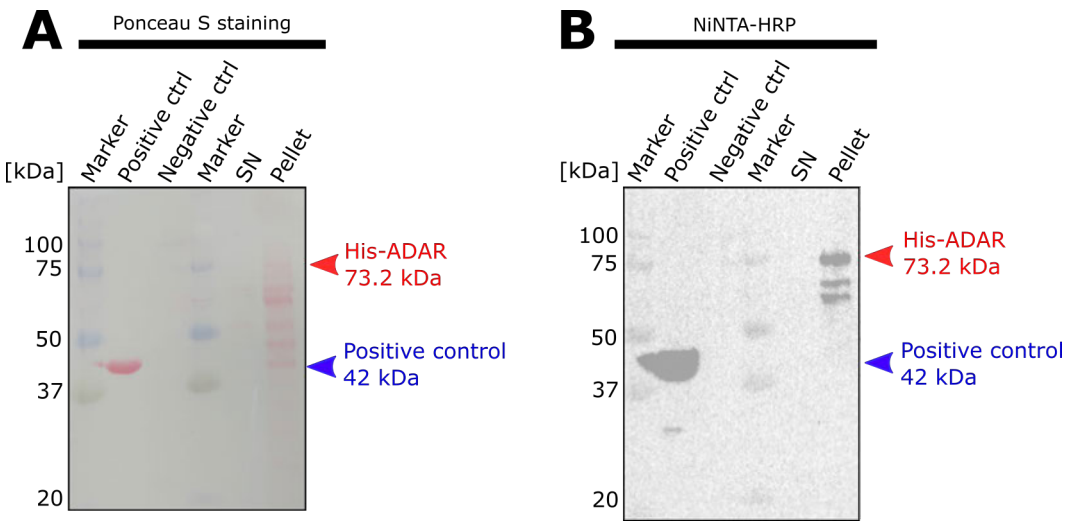


Figure 3.18: Western blot of the supernatant and pellet of the precipitated ADAR sample. Western blot of the supernatant and pellet of the precipitated ADAR sample. A) The nitrocellulose membrane was stained with Ponceau S. B) NiNTA-HRP western blot for detection of His-tagged proteins.

3.4.1 Lysis buffer optimisation

In summary, the main issues of ADAR expression and purification seemed to be: 1) a heterogeneous expression, 2) degradation of the full-length His-tagged version, 3) a high background of insect proteins, 4) a tendency to precipitate during dialysis under low-salt conditions, and 5) a high amount of nucleic acids after cell lysis.

Expressing His-GST-tagged ADAR produced a more homogenous protein species with a lower background contamination of insect proteins. Removing most of the contaminating nucleic acids at an early stage, such as cell lysis or the first affinity chromatography step, might be helpful for further purification of ADAR. Previous work from Macbeth et al., 2007 and Park et al., 2020 on human ADAR2 purification showed that high amounts of salt for cell lysis (750 mM NaCl) and subsequent overnight dialysis (350 mM NaCl) did not lead to protein precipitation. Since Ring et al., 2004 used 200 mM KCl when purifying ADAR from *D. melanogaster* during both cell lysis and purification, 50 mM NaCl in the dialysis buffer might be too low for maintaining the stability of *D. melanogaster* ADAR. Based on these protocols, the lysis conditions for ADAR from Sf9 cells were optimised using higher concentrations of KCl in the lysis buffer (figure 3.19A).

Test expressions using His-GST-ADAR (figure 3.19A) indicated higher KCl concentrations did not adversely affect the protein amounts of soluble His-GST-ADAR that were pulled down. On the contrary, soluble protein levels of His-GST-ADAR are highest with a 1000 mM KCl concentration in the lysis buffer. The outstanding expression levels of His-GST-ADAR might be due to the excellent quality of the cells used in the test expressions with a passage number lower than 20. In addition, cell viability and fluorescence were high (>90%), allowing for a longer protein expression time before harvesting the cells. As for Zn72D/NF45 before, concentrations of 50, 100, and 200 mM urea in the lysis buffer were tested (figure 3.19B). Indeed, there is a decrease in the amount of RNA contamination if the KCl (750 mM) and urea (100 mM) concentration in the lysis buffer are high (figure 3.19B and table 3.3). Therefore, for ADAR lysis buffers, all further protocols used a high concentration of KCl (1000 mM) with 100 mM urea, and dialysis buffers contained 150 mM KCl.

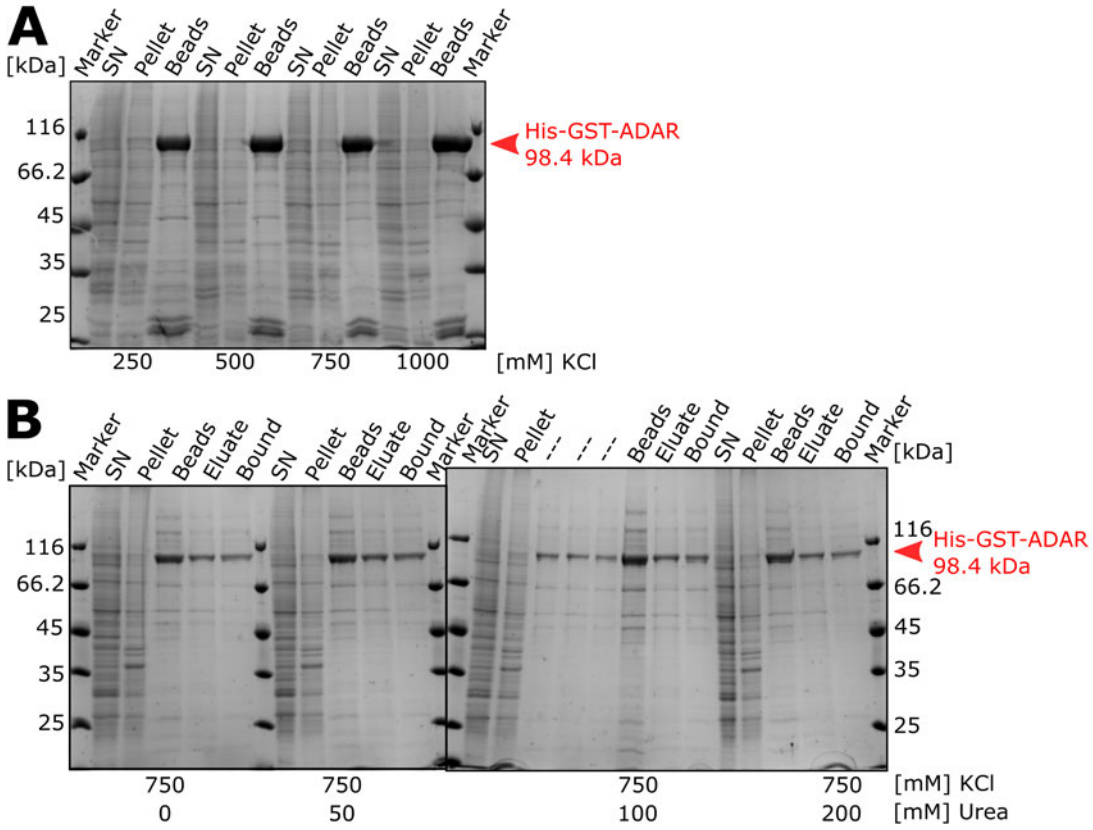


Figure 3.19: Lysis buffer optimisation for purification of His-GST-ADAR
A) Insect cells expressing His-GST-tagged ADAR (indicated by red arrow head) were lysed with different concentrations of KCl in the buffer. B) and C) Insect ells expressing His-GST-tagged ADAR was lysed with different concentrations of urea in buffer containing 750 mM KCl.

Table 3.3: Effect of lysis buffer on RNA contamination of ADAR purifcation as measured by Nanodrop

KCl [mM]	750	750	750	750
Urea [mM]	0	50	100	200
260/280	1.36	1.50	1.03	1.50

3.5 Purification of ADAR

The findings from the previous lysis buffer optimisation (section 3.4.1) were applied for purifying His-GST-tagged ADAR. Following cell lysis and separation of soluble and insoluble components of the lysate by centrifugation, the supernatant was bound to GSH beads in batch before being packed into a column. The column was then washed with GSH buffer A to remove non-specific bound

proteins until it reached a steady 280 nm UV signal baseline. The removal of non-specific bound proteins was observed in the first few washing fractions on the SDS-PAGE (figure 3.20). Next, specifically bound proteins were eluted with 100% GSH buffer B containing 20 mM reduced glutathione run until the UV signal reached baseline again. A single peak containing ADAR was observed upon elution. The gel shows a much cleaner product than obtained previously with the His-tagged construct (figure 3.17), as it was possible to separate ADAR from most contaminants. The eluate, however, still contained a decent amount of free GST (26 kDa). Since the ÄKTAprime plus system measures only at a wavelength of 280 nm, RNA contamination was later determined by Nanodrop. The His-GST-tag of the fractions indicated in figure 3.20B was cleaved using GST-3C.

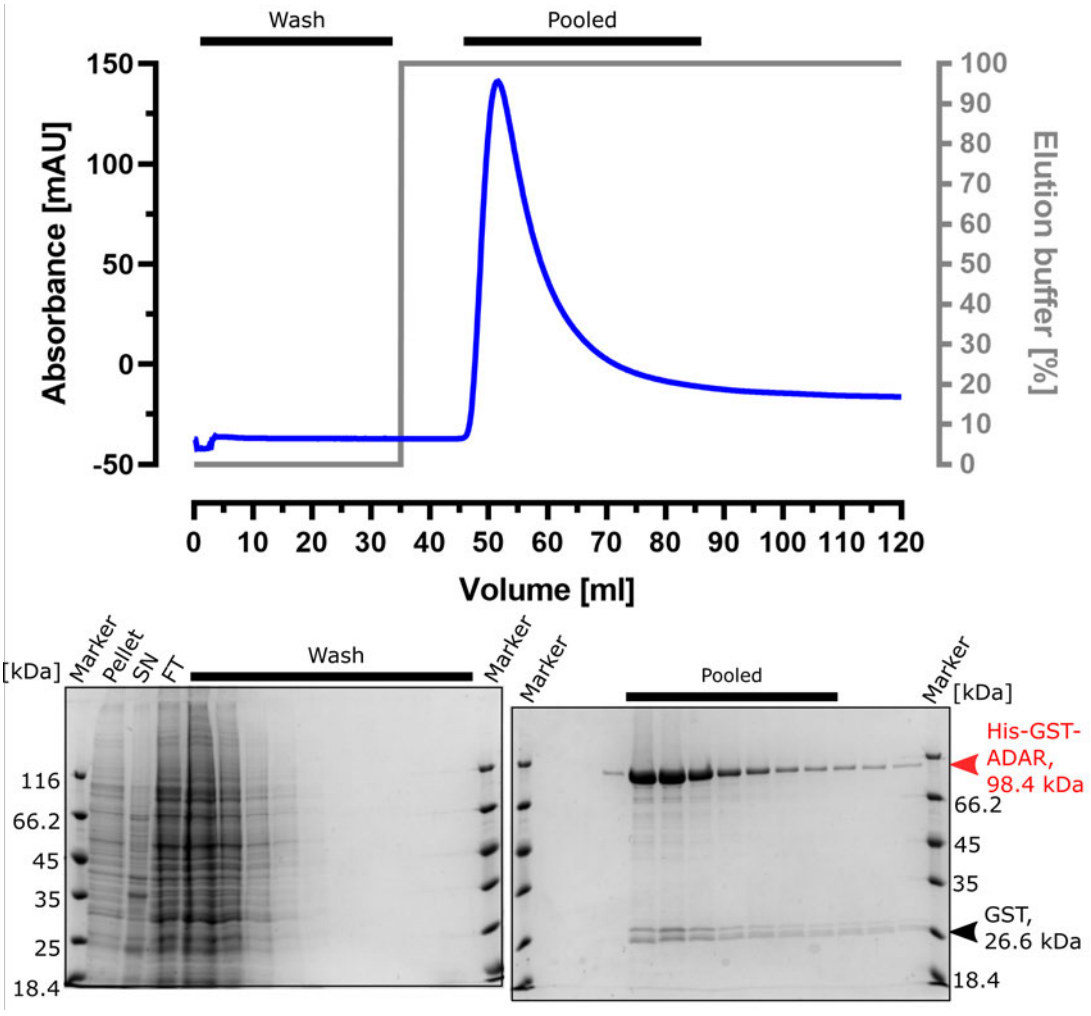


Figure 3.20: Large-scale batch purification of GST-ADAR. The chromatogram shows the UV absorbance at 280 nm in blue and the elution buffer in grey. GST-tagged ADAR is indicated by red arrow head in the SDS-PAGE gel.

ADAR was loaded onto a heparin sepharose column to remove the contaminating nucleic acids, and samples before and after 3C cleavage were analysed on an SDS-PAGE gel (figure 3.21). The His-GST-tag was successfully cleaved as cleaved ADAR has the expected molecular weight of about 70 kDa. In the chromatogram and the corresponding gel, free GST (28 kDa) and GST-3C (40 kDa) did not bind to the heparin column and were found in the flowthrough (FT). The column was washed with heparin buffer A until it reached a steady UV baseline. Then, bound ADAR was eluted with a linear gradient of heparin buffer B (1000 mM KCl) until 60%. A single peak containing ADAR was observed upon elution with a small shoulder on the left. Upon further increase of elution buffer, no more protein could be eluted from the column. The SDS-PAGE gel shows a clean product with a 260/280 nm ratio of about 0.6, indicating no contaminating nucleic acid. The fractions indicated in figure (figure 3.21) were used for size exclusion chromatography.

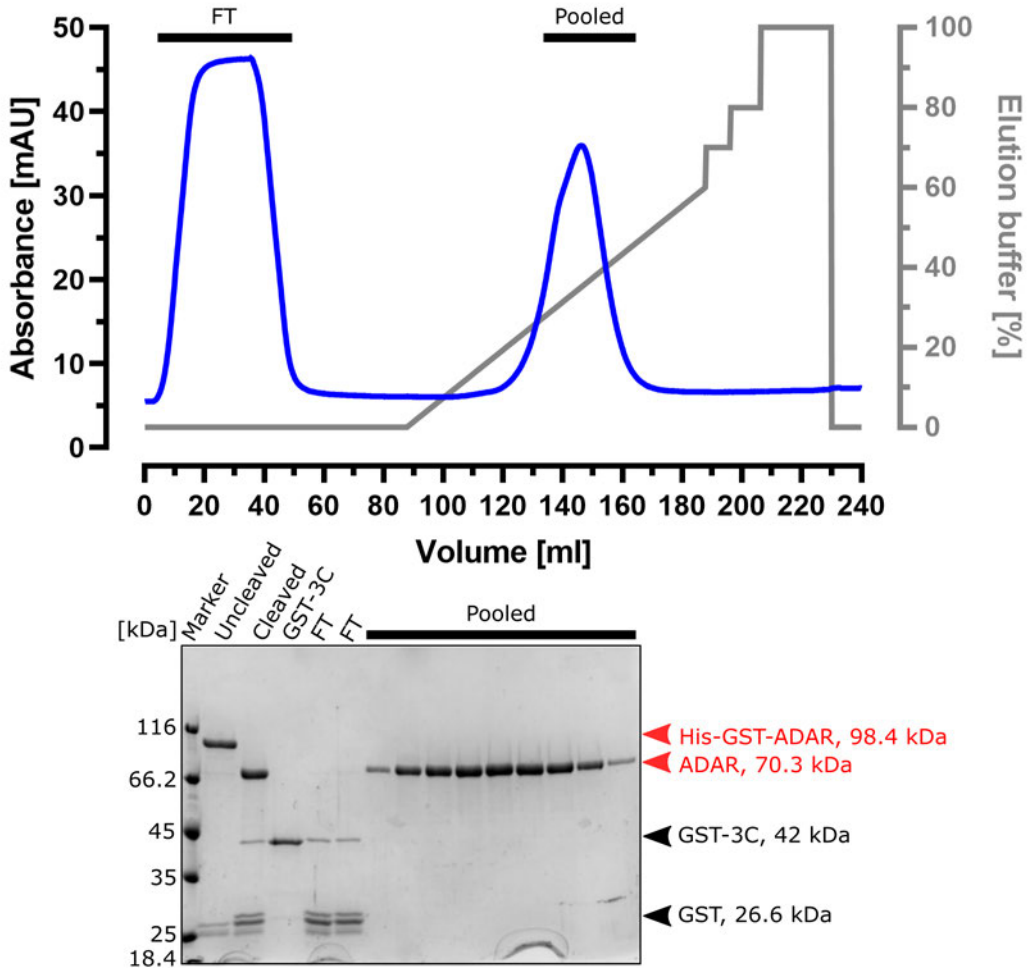


Figure 3.21: Heparin affinity chromatography of ADAR. The chromatogram shows the UV absorbance at 280 nm in blue and the elution buffer in grey. Uncleaved and cleaved ADAR are indicated by red arrow heads and GST and GST-3C by black arrow heads next to the SDS-PAGE gel.

As a final polishing step, ADAR was loaded onto a 24 ml S200 10/300 GL column. Upon elution, a single peak containing ADAR, tailing to the right, was observed in the chromatogram and the corresponding SDS-PAGE (figure 3.22). The SDS-PAGE showed a pure single product with a 260/280 ratio of about 0.6, indicated by the 260/280 nm ratio in the chromatogram. The final yields were typically around 0.5 mg of protein per 1.5 L of insect cell medium.

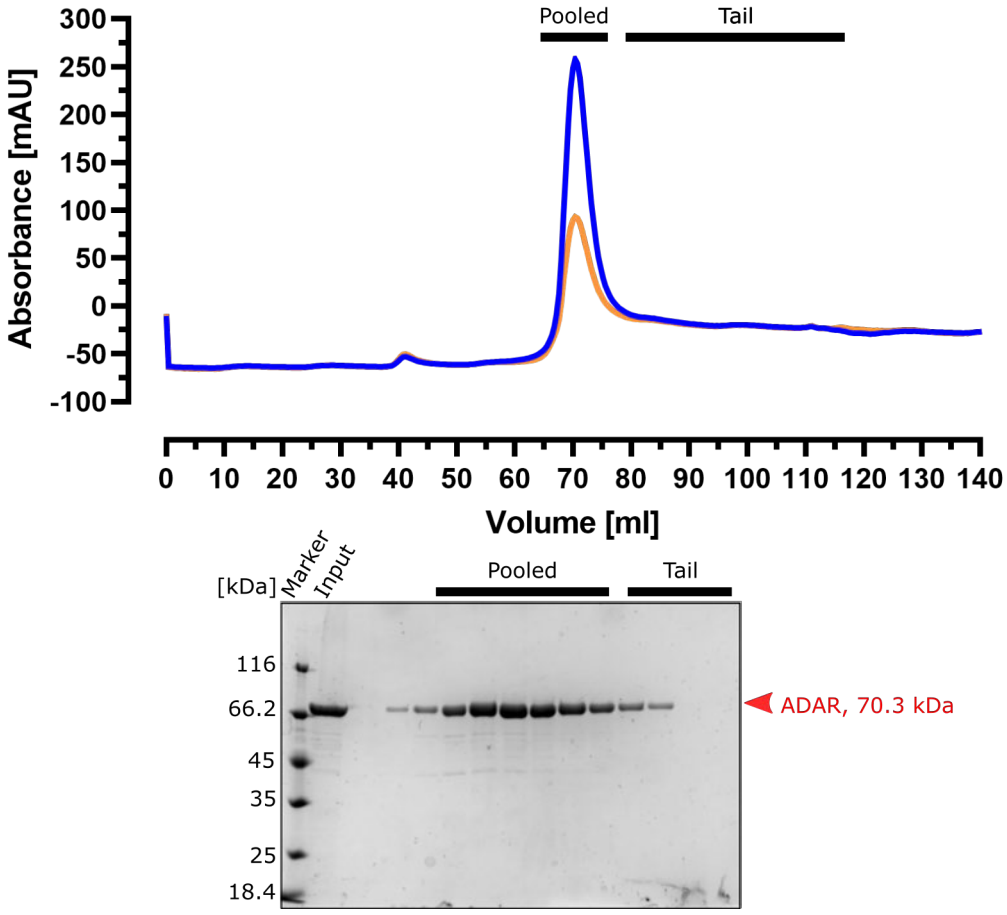


Figure 3.22: Size exclusion chromatography of ADAR. The chromatogram shows the UV absorbance at 280 nm in blue and at 254 nm in orange. ADAR is indicated by a red arrow head next to the SDS-PAGE gel.

3.6 Identification of RNA candidates for an *in vitro* A-to-I editing assay

To gain mechanistic insight into the role of Zn72D in regulating hundreds of A-to-I RNA editing events in *D. melanogaster*, it is crucial to understand why Zn72D is required for proper A-to-I RNA editing for a subset of editing sites. Therefore, this process should be recapitulated in a simplified *in vitro* system with a limited number of components (ADAR and Zn72D/NF45) and chosen RNA editing substrates. For the identification and selection of the RNA editing substrate candidates for this *in vitro* RNA editing assay, we made use of an unpublished preliminary *in vivo* RNA editing dataset at the time, which Anne Sapiro from the Li lab at Stanford kindly provided (Sapiro et al., 2020). For this experiment, she applied RNAi to knockdowns of either ADAR or Zn72D in *D. melanogaster* brains and performed RNA sequencing to identify changes in RNA editing levels. Edited sites appear as guanosines and unedited as adenosines in the sequencing data. In addition, changes in G to A nucleotides indicate editing sites lost upon protein knockdown. The dataset contains a collection of 776 native editing sites, and their quantified and averaged editing levels were ascertained using a GFP control siRNA.

The dataset was screened for the best candidates and then separated into two groups using an online tool prepared by Alistair Kerr from the bioinformatics core facility at the WTCCB. One group is Zn72D-independent (control group), and the other one is Zn72D-dependent for A-to-I RNA editing in *Drosophila* brains. The control group lists editing sites that are ADAR-dependent but Zn72D-independent (called Zn72D-independent). The criteria applied for the Zn72D-independent group were defined as follows:

1. Lower editing levels upon ADAR knockdown compared with GFP control ($\triangle$ADAR)
2. Similar editing levels upon Zn72D knockdown compared with GFP control ($\triangle$Zn72D)
3. Candidates must not show significant reduction in editing in the Zn72D-depleted flies when compared to those lacking ADAR ($\triangle$ADAR-$\triangle$Zn72D)

Table 3.4: Critera for selecting candidates of the Zn72D-independent group

Criterion	Option	Minimum value	Maximum value
1	ΔADAR	-0.8	-0.3
2	ΔZn72D	-0.1	0.05
3	ΔADAR-ΔZn72D	-0.8	0.06

The other group represents ADAR- and Zn72D-dependent editing sites. The criteria applied for the Zn72D-dependent group were defined as

1. Lower editing levels upon ADAR knockdown compared with GFP control (ΔADAR)
2. Lower editing levels upon Zn72D knockdown compared with GFP control (ΔZn72D)
3. Candidates did not show significant reduction in editing in the ZFR-depleted flies when compared to those lacking ADAR (ΔADAR-ΔZn72D)

Table 3.5: Critera for selecting candidates of the Zn72D-dependent group

Criterion	Option	Minimum value	Maximum value
1	ΔADAR	-0.8	-0.3
2	ΔZn72D	-0.9	0.2
3	ΔADAR-ΔZn72D	-0.8	0.06

By applying the above criteria, 17 hits for the Zn72D-independent group (table 3.4) and 21 hits for the Zn72D-dependent group (table 3.5) out of the 776 total entries were extracted. For each of those candidates more information (tables 7.3, 7.4, 7.5, 7.6, 7.7, and 7.8) was collected to help selecting the most promising candidates among these hits. Table 7.2 lists the nature of the additional collated data. Among other information, we were looking at the editing levels of the whole body and head extracted from RADAR, a comprehensive online database for RNA editing in humans, mice and fruit flies (Ramaswami et al., 2013) that is no longer available, or how many more editing sites are within 100 bp up and downstream of the editing site.

RNA editing is known to occur in regions of dsRNA. Therefore, the RNA sequence of each candidate was extracted and used for predicting the minimum free energy (MFE) in kcal/mol as well as its secondary structure using the

RNAfold Server from ViennaRNA Web Services (ViennaRNA, 2022). This was done to allow us to identify shorter regions of structured RNA that are likely to fold *in vitro* to reconstitute short RNAs that can be *in vitro* transcribed for biochemical assays. Sequences of the templates were then truncated to give a double-stranded RNA of about 50 bp length with their editing site in the middle (figure 3.23). When it was also necessary to truncate the RNA sequence of a candidate at one end, the hairpin loop motif UUCG was introduced to cap the RNA molecule and enhance its stability (Baumruk et al., 2001). The length and structure of our RNA templates seem to be suitable for an *in vitro* editing assay compared to the structure of human ADAR2 E488Q bound to an RNA duplex (Matthews et al., 2016a). Furthermore, the farthermost interactions of human ADAR2 (ortholog to *Drosophila* ADAR) with the dsRNA backbone are 11 bp away from the editing site. Thus, our templates should be suitable for mimicking important protein-RNA interactions.

An overview of the final candidates (tables 7.3, 7.4, 7.5, 7.6, 7.7, and 7.8, rows highlighted in orange) for *in vitro* RNA editing assays shows their predicted secondary structures, and their respective editing site is highlighted with a red box (figure 3.23). The last two RNA candidates (G) and (H) represent fully edited versions of CG49442 (A) and Zn72D-dependent Sxl (F).

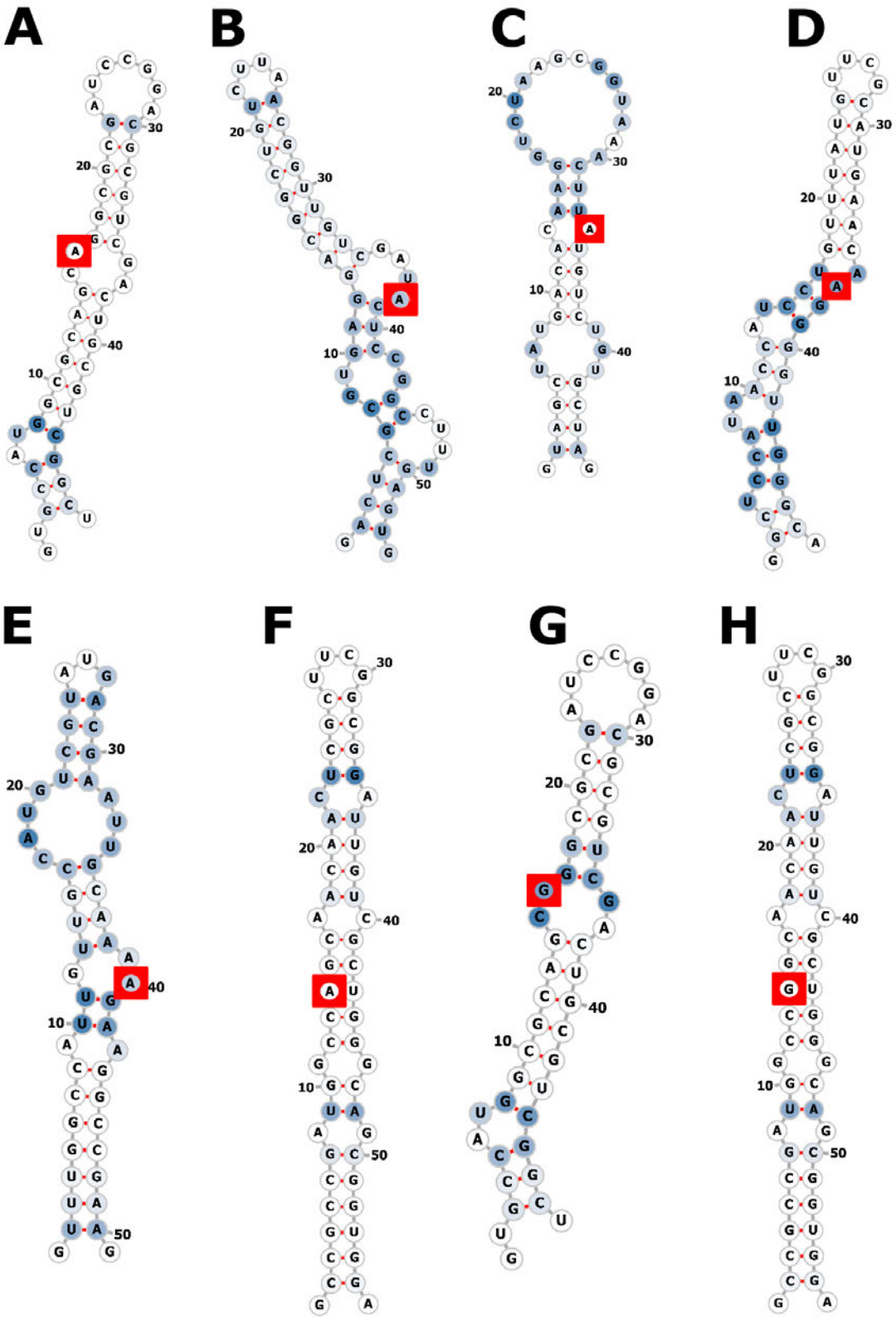


Figure 3.23: Overview of secondary structure predictions of RNA editing assay candidates. RNA candidates were subject to secondary structure predictions using the RNAfold server from ViennaRNA Web services. (A) CG42492, (B) NaCP60E, (C) Caps, (D) Nckx30C, (E) Sxl, (F) para, (G) CG42492 A-to-G, (H) Sxl A-to-G. Base number is given from the in 5' end of a predicted in vitro transcribed product. Red lines connecting circles represent base pairing. Positions of edited sites are shown with red boxes.

After identifying the most promising templates, truncating their sequences, and predicting their secondary structures, *in vitro* transcription templates were designed as described in figure 3.24. Each template starts with a HindIII and ends with an EcoRI restriction site for cloning into a pUC19 vector. In addition, because a linear DNA template is required for run-off *in vitro* transcription, a SmaI restriction site was placed between the RNA template and the EcoRI site used for cloning.

Figure 3.24: Overview of the RNA template for run-off transcription. Overview of the RNA template blueprint for in vitro T7 transcription using SmaI for linearisation and HindIII as well as EcoRI for cloning into pUC19.

For *in vitro* transcription of all eight RNA templates (figure 3.23), the respective cloned plasmids were cut with SmaI, giving blunt end linearised DNA templates.

The control reaction (pTRO-RNA 18s) showed an excellent transcription efficiency 3.25), indicating that our T7 polymerase worked well. Most RNAs were successfully transcribed, and the resulting products migrated at the expected sizes. However, NaCP60E showed extensive smearing, suggesting that a variety of heterogeneous transcripts were made. The Zn72D-dependent candidates (para, Sxl and Nckx30E) show single RNA products. To summarise these test transcriptions, it is apparent that the RNA templates can be *in vitro* transcribed for use in further biochemical assays.

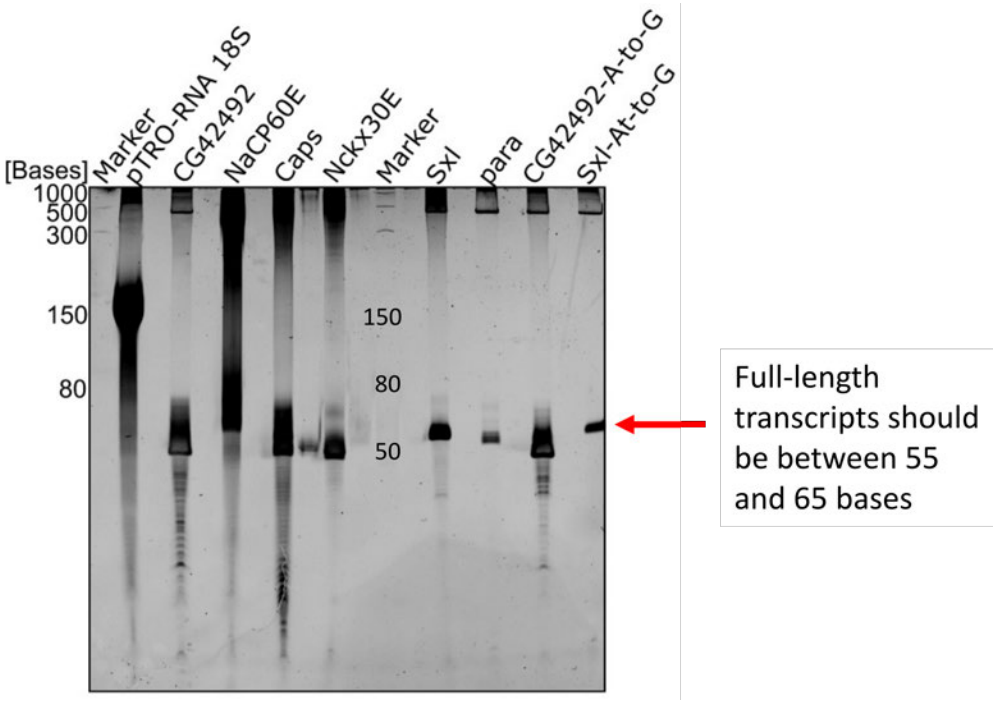


Figure 3.25: Run-off transcription of RNA templates. Denaturing gel of run-off RNA transcriptions of Zn72D-independent and Zn72D-dependent RNA templates using T7 polymerase.

3.7 Poisoned primer extension assay

As it is hard to distinguish between RNAs that are nearly identical in sequence and size (e.g., edited vs non-edited transcripts), a poisoned primer extension assay was adopted for use as our *in vitro* A-to-I RNA editing assay (figure 3.26). The principle of poisoned primer extension is that a reverse transcriptase extends a labelled DNA primer that hybridises 3' to the editing site. The reaction contains three of the four nucleotides for extension (dATP, dGTP, and dCTP) and a chain-terminating dideoxynucleotide (ddTTP). The primer is designed so that the edited site is the first complementary A base after the 3' end of the primer. The extension reaction stops when the reverse transcriptase adds a chain-terminating dideoxynucleotide to the template. Edited RNAs will yield different-sized extension products that can be resolved on a denaturing gel and visualised by its fluorescent label.

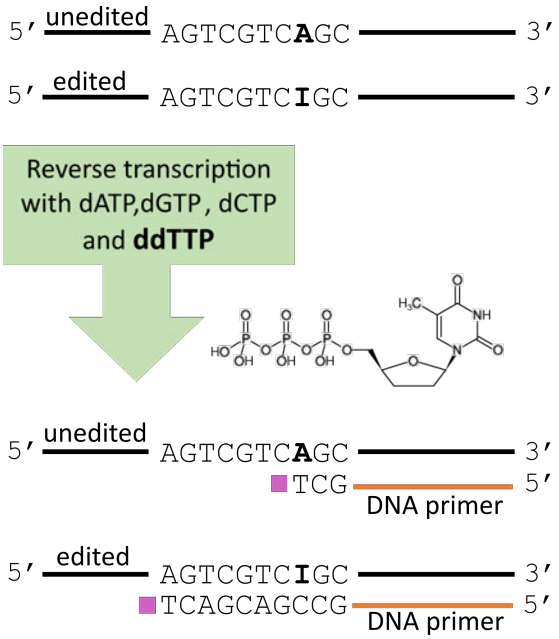


Figure 3.26: Principle of poisoned primer extension assay. A reverse transcriptase extends fluorescently-labelled DNA primers that complement a sequence 3' with respect to their editing site. Edited and unedited RNAs can be differentiated due to products of different length.

As a quality control of the proteins required for the poisoned primer extension assay, the RNA-binding activity of recombinant purified ADAR and Zn72D/NF45 complex was assessed by EMSA (figure 3.27). A fluorescently-labelled C-rich 24-mer RNA probe ("UGACACCCUGACACCCUGACACCC") was used for this. Both ADAR and Zn72D/NF45 showed RNA binding activity *in vitro* and therefore could be used for RNA editing assay. However, ADAR's enzymatic activity is not yet tested.

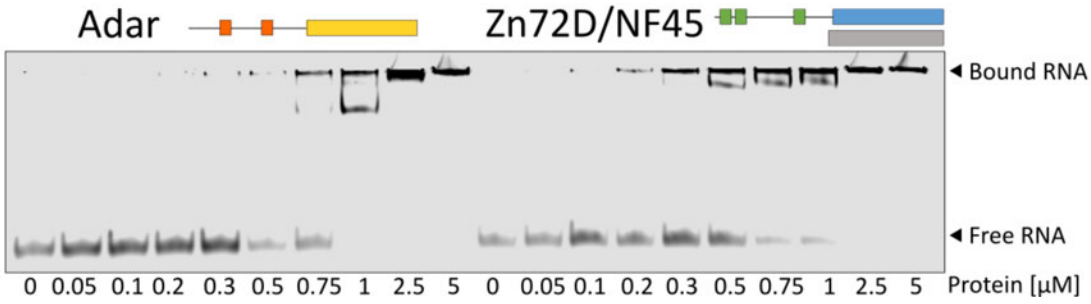


Figure 3.27: Testing RNA-binding capability of recombinant ADAR and Zn72D/NF45 from *D. melanogaster*.

The performance of five reverse transcriptases (RTs) (table 3.6) in poisoned primer extension assays were tested following a common protocol with the manufacturer's recommended reaction temperatures best-suited for each RT.

RTs have to extend the fluorescently-labelled DNA primer with more nucleotides for the edited RNA because instead of ddTTP, a dCTP is used for incorporation at the mutated editing site (figure 3.23). The bypassing of the mutated editing sites until the RT encounters the next adenosine in the RNA leads to different product lengths of 25 nt. Therefore, the Zn72D-independent RNA template CG422492 (unedited) and its counterpart CG422492 A-to-G (fully edited) were first *in vitro* transcribed in a large scale. Then, the poisoned primer extension assay was performed as described in (section 2.7.8). A DNA primer with a length of 20 nt was used for both RNAs for primer extension. After primer extension for unedited samples, the expected size of products was 22 nt, and for fully edited samples, 25 nt. Therefore, two additional fluorescent oligos were used as size controls of 25 and 22 nt and loaded next to the primer extension samples (figure 3.28). As a result, all five RTs could extend fully edited and unedited RNAs by the expected number of nucleotides. However, the variation in the bands' intensity of the final products suggested that the transcription efficiency was highly variable across the different RTs, especially since the amount of template RNAs and DNA primers used for the various conditions were the same. RT1, RT3, RT4, and RT5 showed some intermediate-sized products (21-24 nt) and the expected 25 nt product, indicating that they produce highly variable endpoints. Moreover, these four RTs were extending the DNA primer by one additional nucleotide. This makes the discrimination between edited RNAs that have not been fully extended and unedited RNAs that have been erroneously extended by an extra nucleotide problematic, therefore, obscuring the ability of the assay to differentiate between these two species. RT2, on the other hand, shows a generally weaker performance for both edited and unedited substrates than the four other RTs, making it also a suboptimal choice. But at least there are only two products.

Table 3.6: Reverse transcriptases used for primer extension

Label	Reverse transcriptase
RT1	MulV
RT2	Avian
RT3	Revert
RT4	SuperScript IV
RT5	Maxima

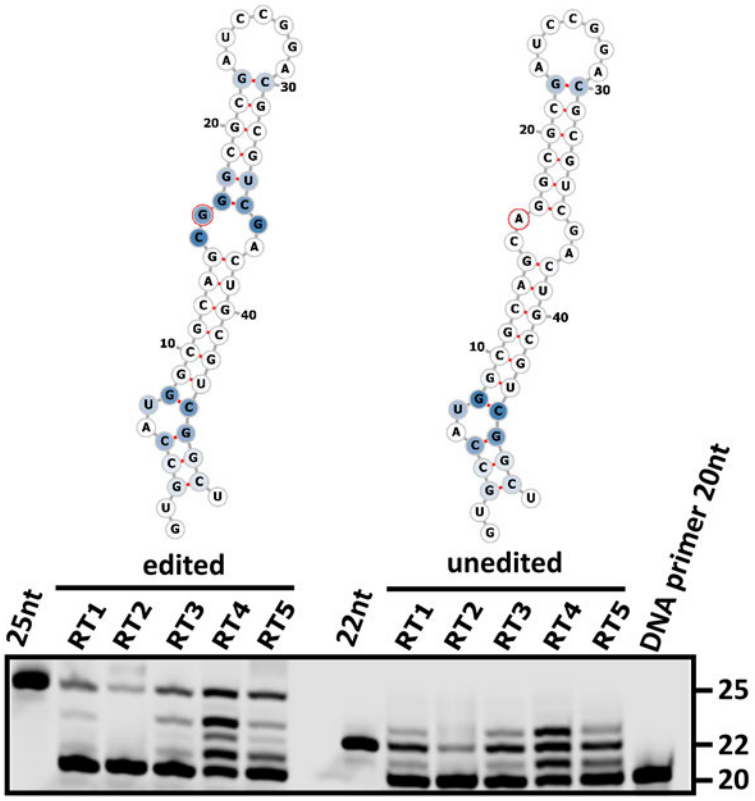


Figure 3.28: Testing reverse transcriptases for primer extension performance. Five different reverse transcriptases were used for testing their primer extension performance.

Despite the apparent issues of poisoned primer extension assays in quantifying A-to-I RNA editing, four out of five RTs incorporated not more than a single additional nucleotide for unedited RNA substrates. For CG42492, the difference between edited and unedited should be three nucleotides which is why it should be possible to make a qualitative statement about *in vitro* RNA editing capability of *D. melanogaster* ADAR. Since all required components for testing ADAR's enzymatic activity were ready, a poisoned primer extension assay was performed. Therefore, CG4292 was chosen as substrate (as it does not require Zn72D for proper editing) was incubated with ADAR (protein molar excess to RNA 20 to 1) for three hours. Four different conditions for incubating RNA substrates with ADAR were tested due to different protocols (Hanswillemenke et al., 2015, Heep et al., 2017, Filippini et al., 2018, Sasaki et al., 2006) for *in vitro* assays using ADAR.

1. 3 hours at 21 ℃
2. 3 hours at 30 ℃

3. 3 hours at 37℃
4. 30 minutes at 30℃ and 30 minutes at 21℃ alternating for a total of 3 hours

The denaturing RNA gel shows a slight distortion of the samples, but size controls were run on both sides of the gel, making it possible to conclude (figure 3.29). The assay control samples (Ctrl) did not contain ADAR but were otherwise treated the same as other samples. Hence, they should be extended by RT1 to the expected size of 22 nt (22nt). However, as seen in the previous assay, RT1 (was the only RT with enough enzyme for 6 reactions at the recommended concentration of RT per reaction) extended the DNA primer by a single extra nucleotide, making it a 23 nt long product. All four samples containing ADAR show an additional band, compared with the control, that migrates at the same size as the fully edited size control (25nt), indicating a successful deamination reaction of the adenosine at the editing site of the RNA substrate. This band is quite faint compared with the shorter products and might suggest that, despite the excess ADAR, the conditions were not optimal, or ADAR is not fully active.

Moreover, another band (32 nt) is present in three samples containing ADAR (30℃, 37℃, and 30℃/37℃). This additional band seems to represent the product after RT1 bypassed the next adenosine after the editing site (ADAR edited that) and extended the DNA primer until it encountered the next adenosine after the editing site. The second adenosine after the editing site is 10 nucleotides away and corresponds with the band present on the gel. This bypassing behaviour might indicate that the second site is also edited, in which case, the assay is even less clear. However, it also demonstrates that there might be more evidence for ADAR activity.

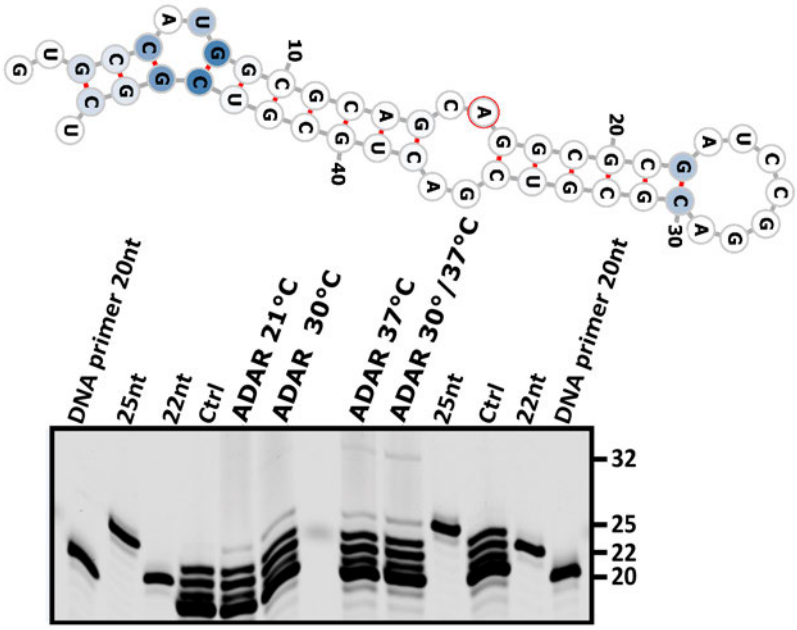


Figure 3.29: Poisoned primer extension assay of RNAs incubated with recombinant ADAR. Zn72D-independent RNA CG42492 was incubated with recombinant ADAR and then used for poisoned primer extension assay to assess ADAR's enzymatic activity *in vitro*.

3.8 Endonuclease V assay

It is difficult to distinguish between RNA molecules nearly identical in sequence and size (e.g., edited vs non-edited products). Contrary to our initial idea, poisoned primer extension assays were not able to accurately differentiate between edited, only partially extended, or non-edited RNAs extended beyond the editing site, making it impossible to quantify A-to-I RNA editing executed by *D. melanogaster* ADAR. Therefore, an alternative assay using endonuclease V (EndoV) was tested. EndoV is a highly active ribonuclease specific for inosine-containing nucleic acids and catalyses the cleavage of the second phosphodiester bond 3' to inosine, leaving 3'-OH and 5'-phosphate fragments. In *E. coli*, EndoV acts as a DNA repair enzyme involved in removing deaminated adenosines from the genome. However, the function of eukaryotic EndoV has not been identified yet.

The cleavage activity of bacterial and human EndoV was compared for single-stranded (ss) and double-stranded (ds) DNA and RNA containing inosine by Vik et al., 2013. The nucleic acid substrates were incubated with wild-type EndoV which cleaved single-stranded RNA containing inosine at position 13 of a 25-mer oligo at even low protein concentrations (1 nM). Human EndoV only cuts RNA with inosine and is active on both ss and dsRNA but not DNA. Therefore, human EndoV's cleavage specificity for inosine-containing RNAs can be used as a tool to show whether *D. melanogaster* ADAR is functional as RNA substrates should only be cleaved if inosines are incorporated by active ADAR. An additional advantage is that no experimental artefacts are expected due to the absence of inosines in unedited RNA substrates.

3.8.1 Optimisation of expression conditions of recombinant human EndoV

For protein expression and purification of human EndoV, two different protocols from Vik et al., 2013 and Morita et al., 2013 were used as starting points for further optimisation. One suggested protein production of a fusion protein with an N-terminal His-MBP tag in BL21-Codon Plus (*DE3*)-RIPL cells, and the other recommended expressing a GST-tagged version in Rosetta 2 cells. Ultimately, adopting the two approaches to our requirements, a full-length His-GST-tagged version of human EndoV was cloned for further protein expression and purification.

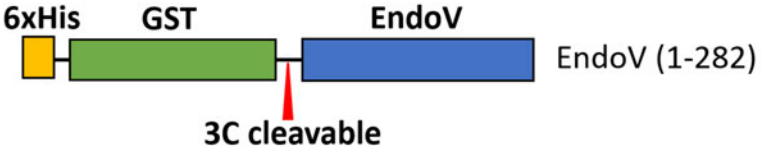


Figure 3.30: His-GST-EndoV construct used for protein purification. Illustration of full-length human EndoV (282 aa) with an N-terminal 6xHis-GST-tag used for protein expression and purification shows protease cleavage site (3C), indicated by a red arrowhead.

The His-GST-tagged EndoV construct (figure 3.30) was expressed in either BL21 (*DE3*) or B834 cells . Small-scale expression trials using GSH beads for pull-down revealed that the EndoV construct expressed well and was soluble with the expected molecular weight in both bacterial strains (figure 3.31). B834 cells were used for further screening trials and large-scale protein expression. The contaminants are more likely degraded His-GST-EndoV than non-specific interactors.

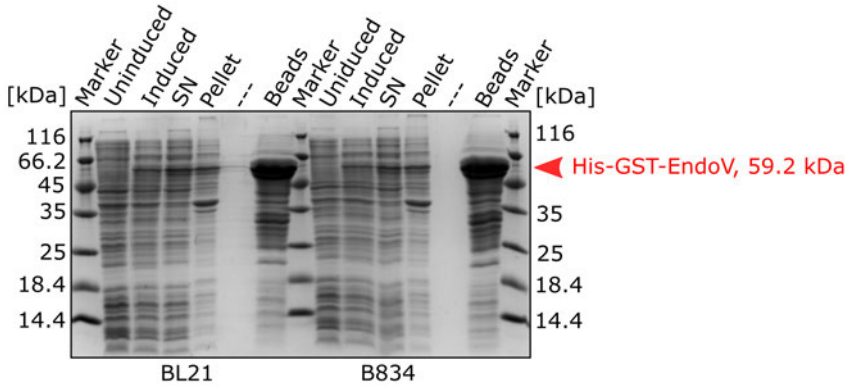


Figure 3.31: Test expressions of human EndoV in bacteria. Protein production of full-length His-GST-tagged human EndoV (59.2 kDa) in two bacterial expression strains. EndoV is indicated by a red arrow head next to the SDS-PAGE gel.

The published protocols used either 300 mM NaCl or 300 mM KCl during cell lysis. Since optimising the amount of salt in the lysis buffer had positive effects for ZFR/NF45, Zn72D/NF45, and ADAR, lysis buffer optimisation was also performed for EndoV. However, figure 3.32 shows little to no increase in the amount of soluble EndoV with increasing salt concentrations in the lysis buffer (250, 500, 750, and 1000 mM). Therefore, for large-scale purification of EndoV, 500 mM NaCl was chosen as it gave slightly higher soluble protein levels than 250 mM.

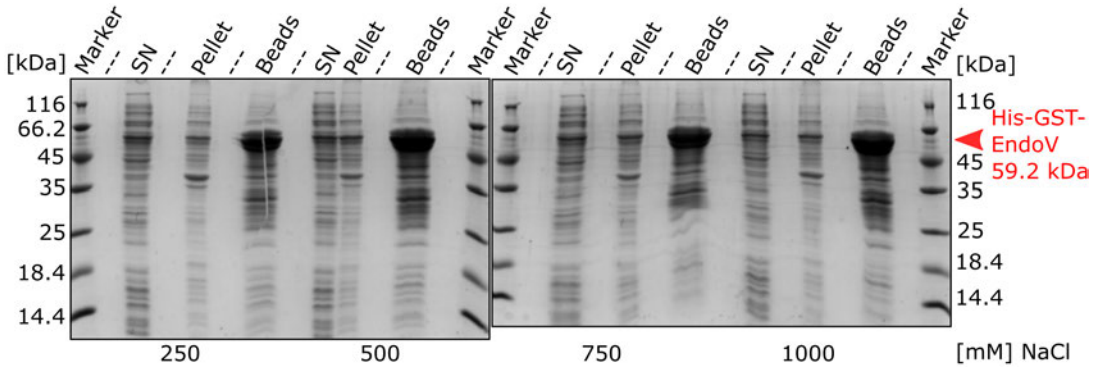


Figure 3.32: Lysis buffer optimisation for purification of His-GST-tagged human EndoV. Protein levels of soluble His-GST-tagged human EndoV (indicated by red arrow head) have been tested for different concentrations of NaCl in the lysis buffer.

3.8.2 Protein purification of recombinant EndoV

The protocol for purification of human EndoV (Vik et al., 2013) as well as the findings from the previous lysis buffer optimisation (section 3.8.1) were used as a starting point for purifying full-length His-GST-tagged human EndoV. Following cell lysis and separation of soluble and insoluble components of the lysate by centrifugation, the supernatant was loaded onto a 5ml HisTrap FF column and washed with NiNTA buffer A to remove any non-specific bound proteins until it reached a steady UV baseline. Removal of unbound (FT) and non-specific bound proteins (Wash) can be seen in the chromatogram and the corresponding SDS-PAGE gel (figure 3.33). Bound proteins were eluted with a linear gradient of NiNTA buffer B containing 1000 mM imidazole. A single peak containing EndoV and other contaminating proteins was observed upon elution. The fractions indicated in figure 3.33 were pooled and used for the subsequent affinity chromatography step (GSH) to remove these protein contaminations.

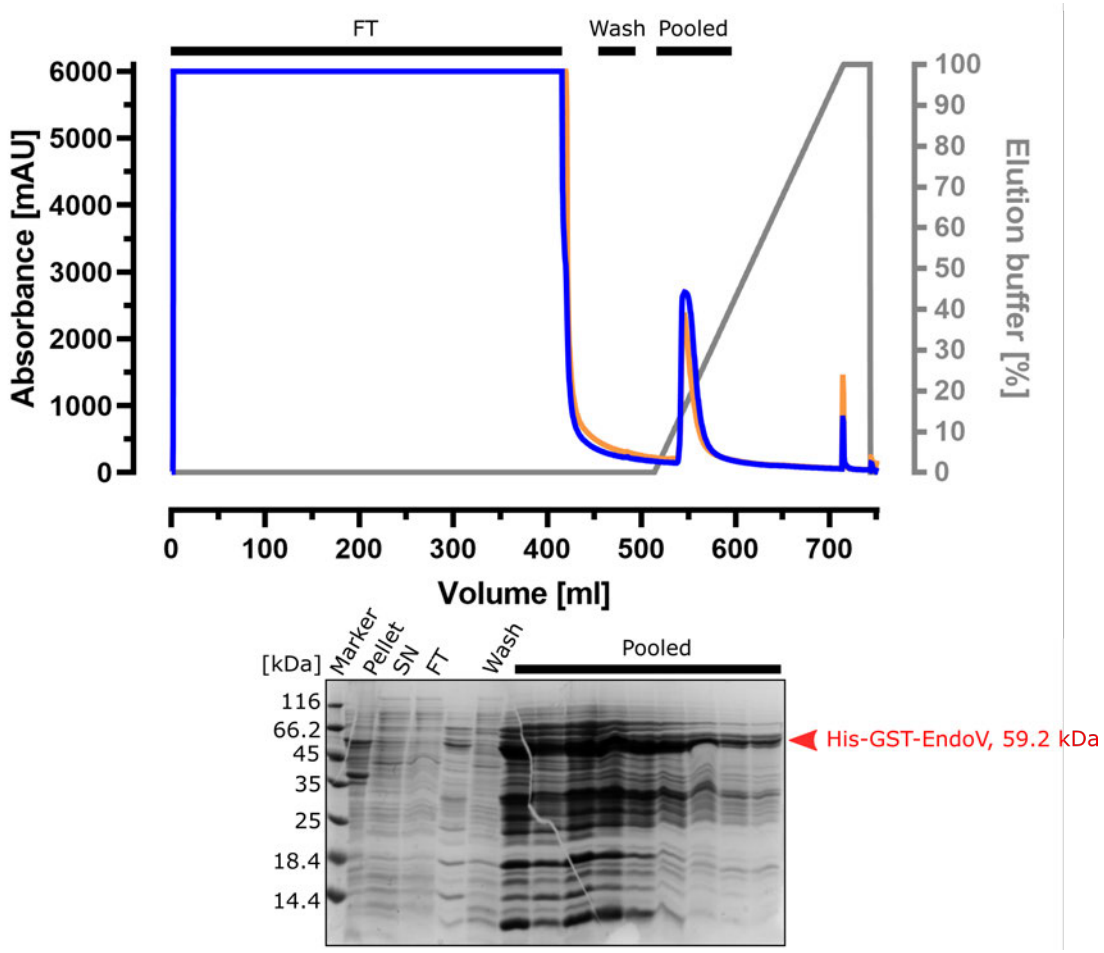


Figure 3.33: Immobilised affinity chromatography of EndoV. The chromatogram shows the absorbance at 280 nm in blue and at 254 nm in orange. EndoV is indicated by a red arrow head next to the SDS-PAGE gel.

After IMAC, EndoV was incubated with GSH beads before being packed into a column and was then washed with GSH buffer A to remove any non-specific proteins until a steady UV baseline was reached. The corresponding chromatogram and SDS-PAGE (figure 3.34) showed that most contaminants did not bind to the beads (Unbound) while some proteins could be washed off (Wash) before elution. Next, bound EndoV was eluted with 100% GSH buffer B. A single peak containing EndoV was observed upon elution. The SDS-PAGE gel shows a cleaner product with some remaining protein contaminations. The His-GST-tag of the EndoV protein in the indicated pooled fractions was cleaved using GST-3C.

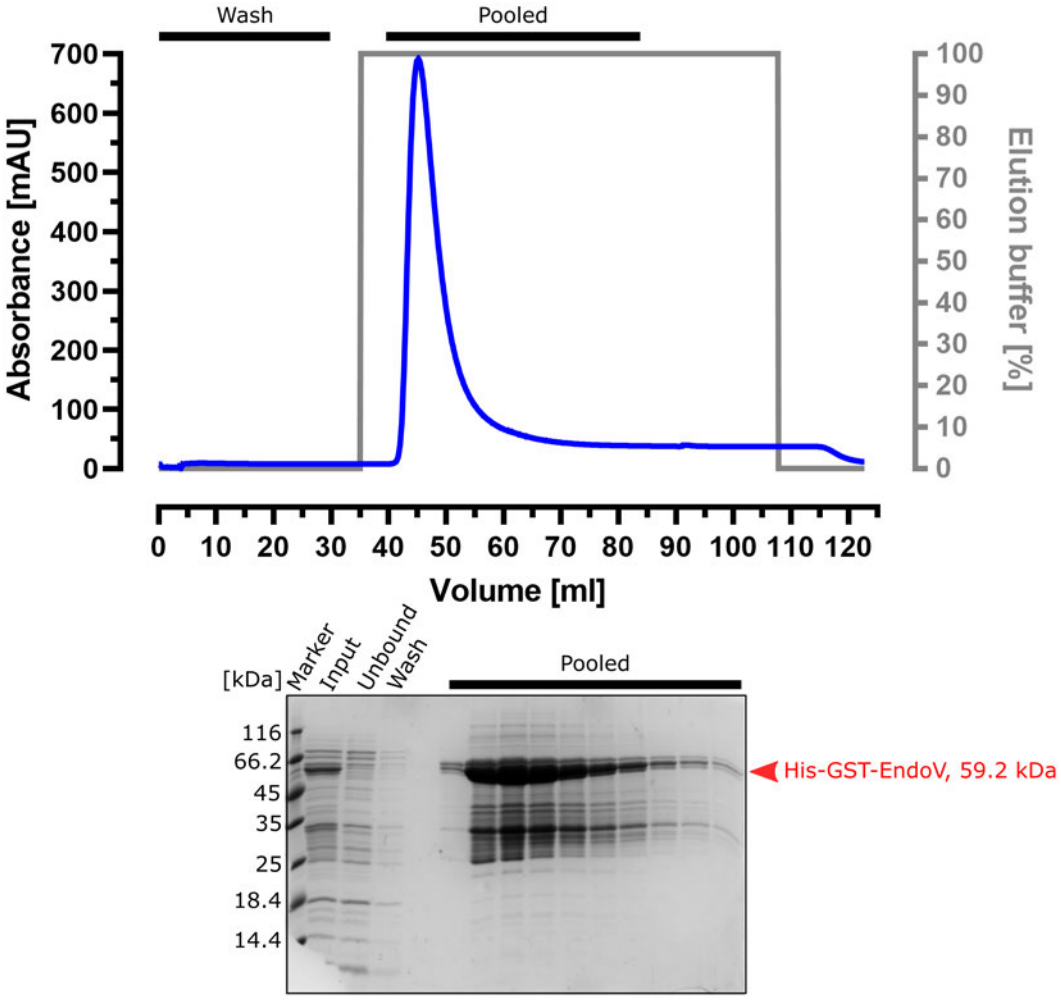


Figure 3.34: Affinity chromatography of EndoV using GSH beads. The chromatogram shows the absorbance at 280 nm in blue and at 254 nm in orange. EndoV is indicated by a red arrow head next to the SDS-PAGE gel.

Following the two affinity purification steps, many nucleic acids were still associated with EndoV as assessed by the 260/80 nm ratio. EndoV was loaded onto a Heparin column and specifically bound EndoV was eluted with a linear gradient of heparin buffer B until 100% (1000 mM NaCl). Two elution peaks contained either free GST and some tagged-EndoV or cleaved EndoV contaminated with some free GST. The corresponding chromatogram and SDS-PAGE (figure 3.35) indicated that contaminating nucleic acids could be separated from EndoV as there was not much protein detectable in the flow-through (FT) by SDS-PAGE and the 260/280 nm ratio is also relatively high in these fractions. The samples before and after 3C cleavage analysed by SDS-PAGE showed that almost all His-GST-tagged EndoV has been successfully cleaved. The fractions indicated in figure 3.35 had a low 260/280 nm ratio and were

pooled for subsequent size exclusion chromatography.

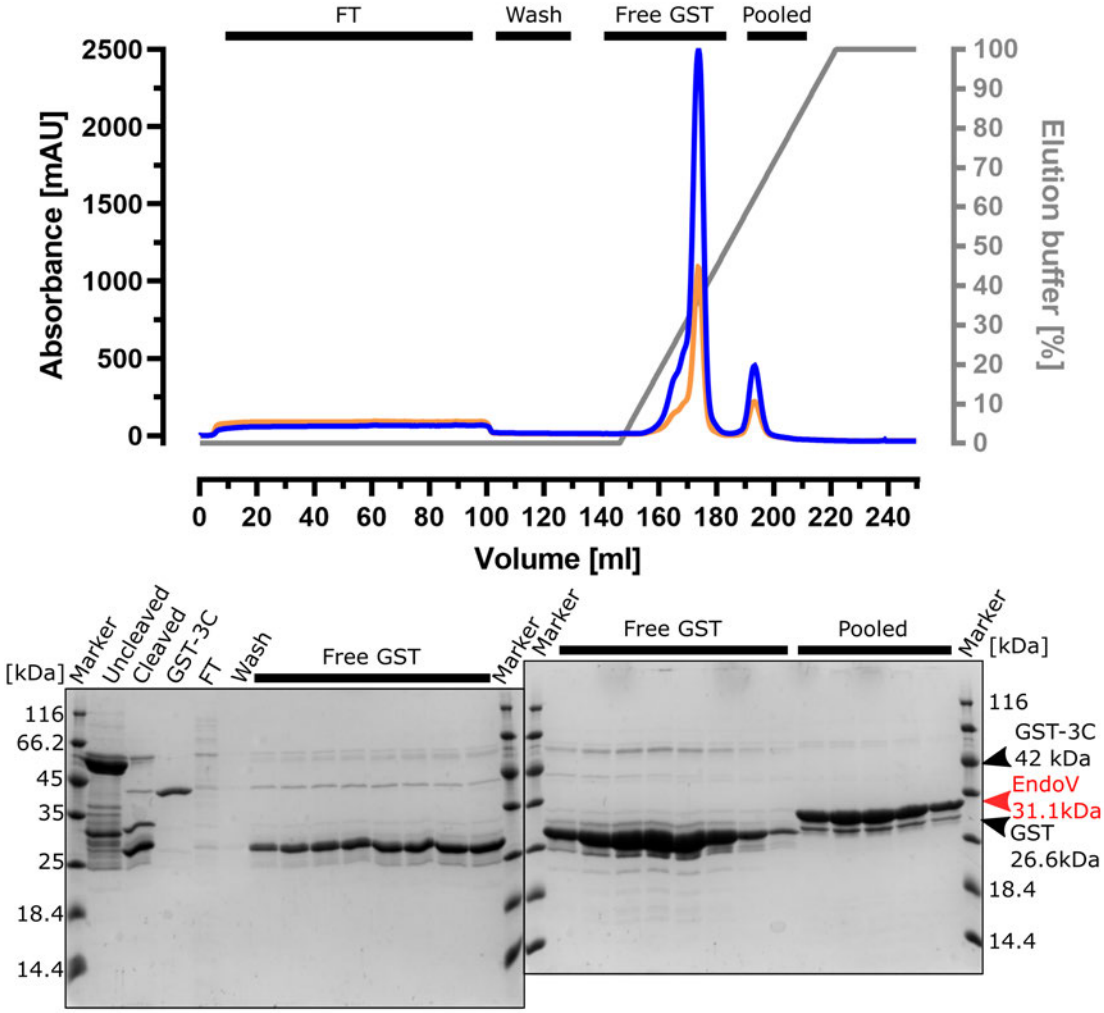


Figure 3.35: Heparin affinity chromatography of EndoV. The chromatogram shows the absorbance at 280 nm in blue and at 254 nm in orange. EndoV is indicated by a red and GST as well as GST-3C by black arrow heads next to the SDS-PAGE gel.

As a final polishing step, human EndoV was loaded onto a 120 ml S200 10/60 column. Upon elution, a single peak containing EndoV, tailing to the right, was observed in the chromatogram and the corresponding SDS-PAGE (figure 3.36). The SDS-PAGE showed a quite pure product with a 260/280 ratio of about 0.6, indicated by the 260/280 nm ratio in the chromatogram. The additional band migrating at around 25 kDa represents the residual cleaved GST-tag that has not been separated from EndoV. The final yields were typically around 2 mg of pure protein per 2 L of bacterial cells. The protein sample was flash-frozen in liquid nitrogen for further experiments.

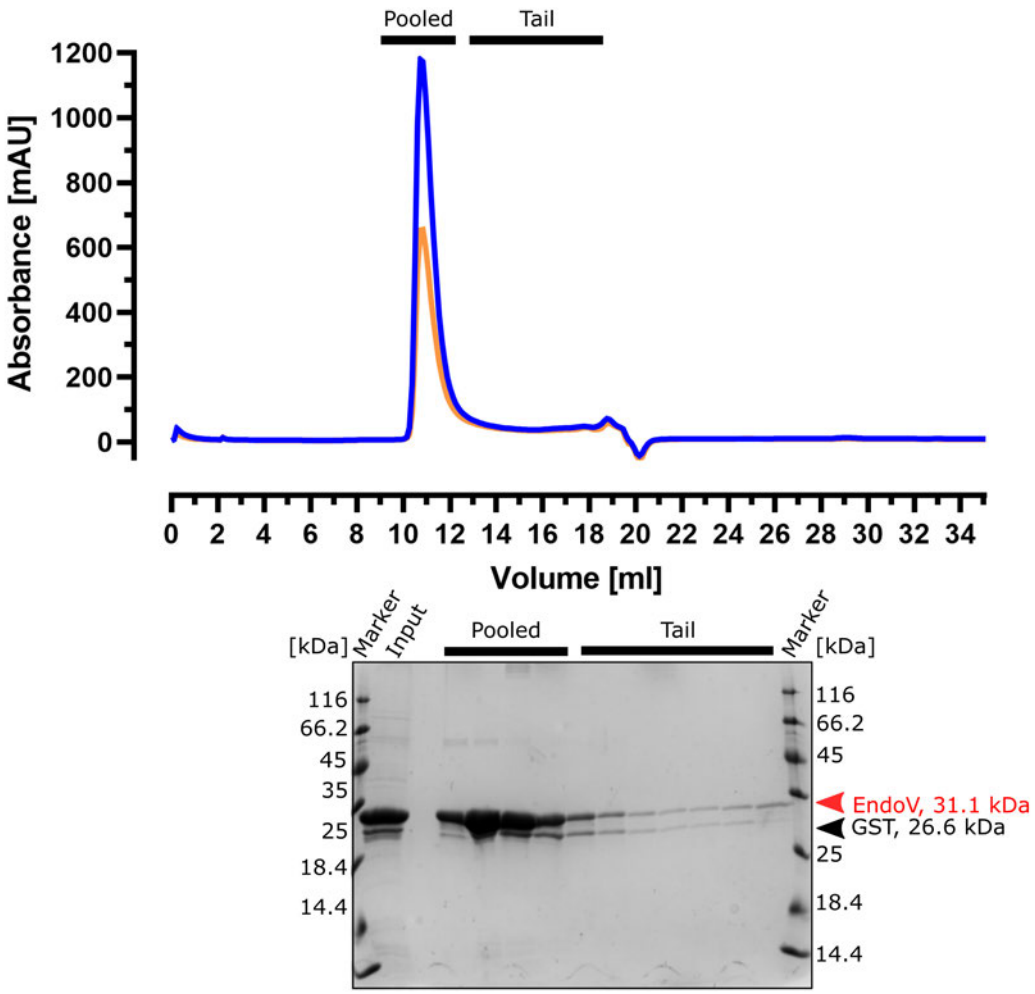


Figure 3.36: Size exclusion chromatography of EndoV. The chromatogram shows the absorbance at 280 nm in blue and at 254 nm in orange. EndoV is indicated by a red and GST by black arrow head next to the SDS-PAGE gel.

3.8.3 Detection of inosines in RNAs using EndoV

After successful purification of full-length human EndoV, its cleavage activity was tested on several RNA substrates. A 5′ fluorescently-labelled (IRDye 800CW) 20mer RNA oligo with either adenosine or inosine at position 15 was used as ssRNA or dsRNA (figure 3.37A) for this. Vik et al., 2013 suggested a specific reaction buffer (10 mM Tris-HCl pH 7.5, 50 mM potassium chloride, 0.5 mM manganese chloride, 1 mM DTT and 5% glycerol). However, it was tested if EndoV would also exhibit full activity in EMSA buffer as all previous experiments, such as RNA binding assays, were carried out in EMSA buffer. Therefore, EndoV cleavage activity in EMSA buffer (20 mM Tris-HCl

pH 7.5, 150 mM potassium acetate 150 mM, 4 mM magnesium acetate, and 1 mM DTT) was assessed. EndoV seems to be active in both buffers (figure 3.37B+C) as well as being inosine-specific as there is no cleavage detectable for all the adenosine-only containing substrates. However, EndoV cleaves ssRNA after inosines in the presence of either manganese, magnesium or both, but it does not cleave dsRNA without manganese. The results of this experiment recapitulate the findings from Vik et al., 2013 as EndoV prefers ssRNA over dsRNA. Also, EndoV seems to be active in the sole presence of magnesium, but only if ssRNA is the substrate and it is not able to cleave dsRNA if there is no manganese present (figure 3.37B+C).

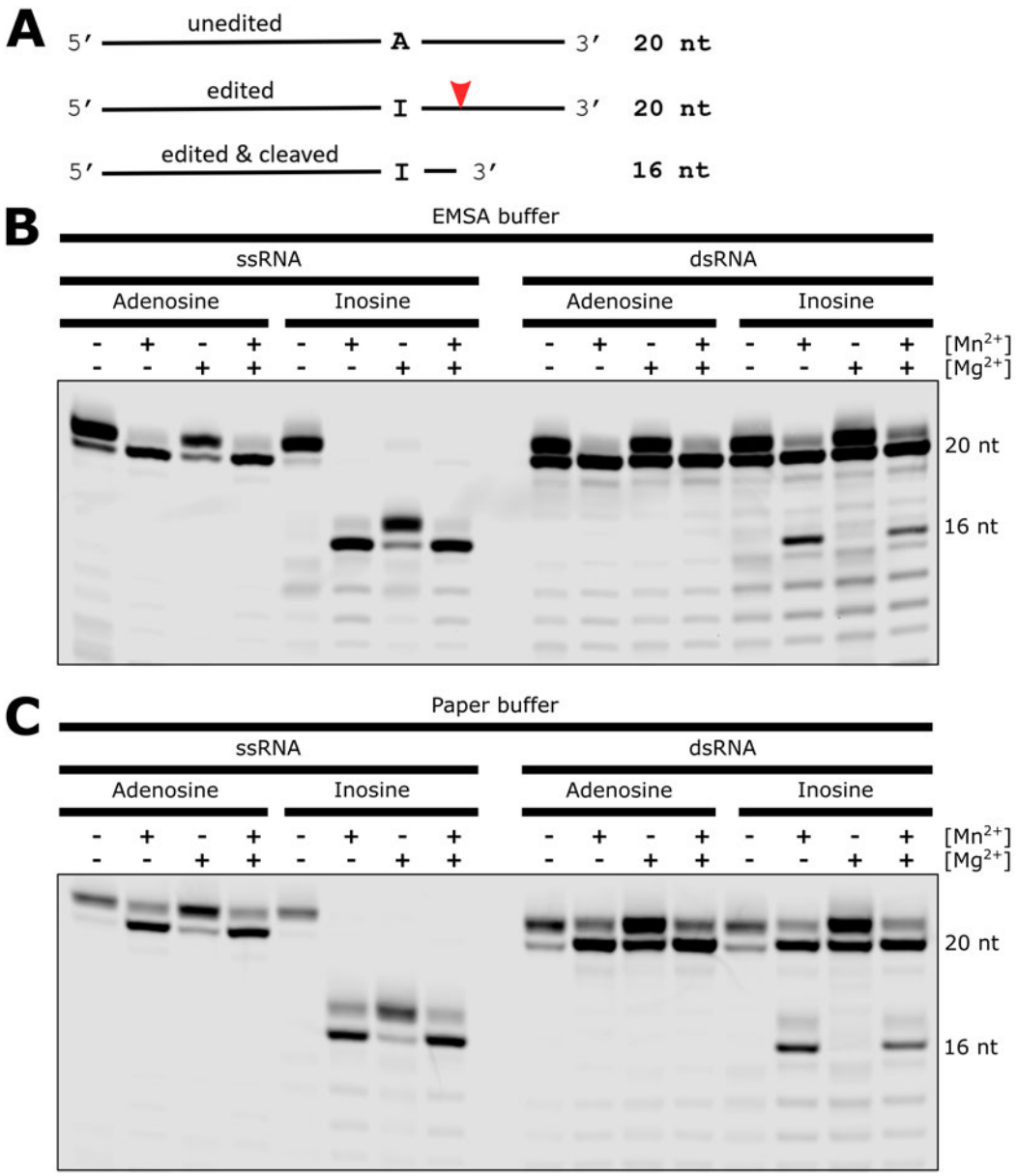


Figure 3.37: Measuring human EndoV cleavage activity *in vitro*. A) Two different 20-mer RNA substrates were used to determine EndoV cleavage, with either an adenosine (non-edited) or an inosine (edited) at position 15. The respective EndoV cleavage site for the edited RNA substrate is indicated by a red arrowhead. Fluorescent scans of denaturing RNA gels of EndoV assays (containing extra added divalent Mg^{2+} and Mn^{2+} ions) cleaving ssRNA and dsRNA in EMSA buffer (B) and Paper buffer (C) containing divalent ions.

To find out the cause of the double band of RNA oligos seen in previous denaturing RNA gels, the impact of EMSA buffer, manganese, and magnesium (figure 3.38A) on the migration behaviour of the RNA probe was tested. ssRNA and dsRNA containing either adenosine or inosine in water or EMSA

buffer containing or lacking divalent ions were separated on a 20% denaturing urea acrylamide gel in the absence of EndoV. It became evident from figure 3.38A that manganese ions influence the migration behaviour of the RNA oligos. Furthermore, manganese ions also alter fluorescence intensity as all RNA oligos were loaded onto the gel in equivalent amounts, but samples with manganese exhibited a higher total intensity. However, a double band also appears in samples containing only water. Since this oligo was only used for assessing the cleavage activity of EndoV and this migration behaviour could not be observed for any other RNA probe, the RNA oligos were used for testing purposes.

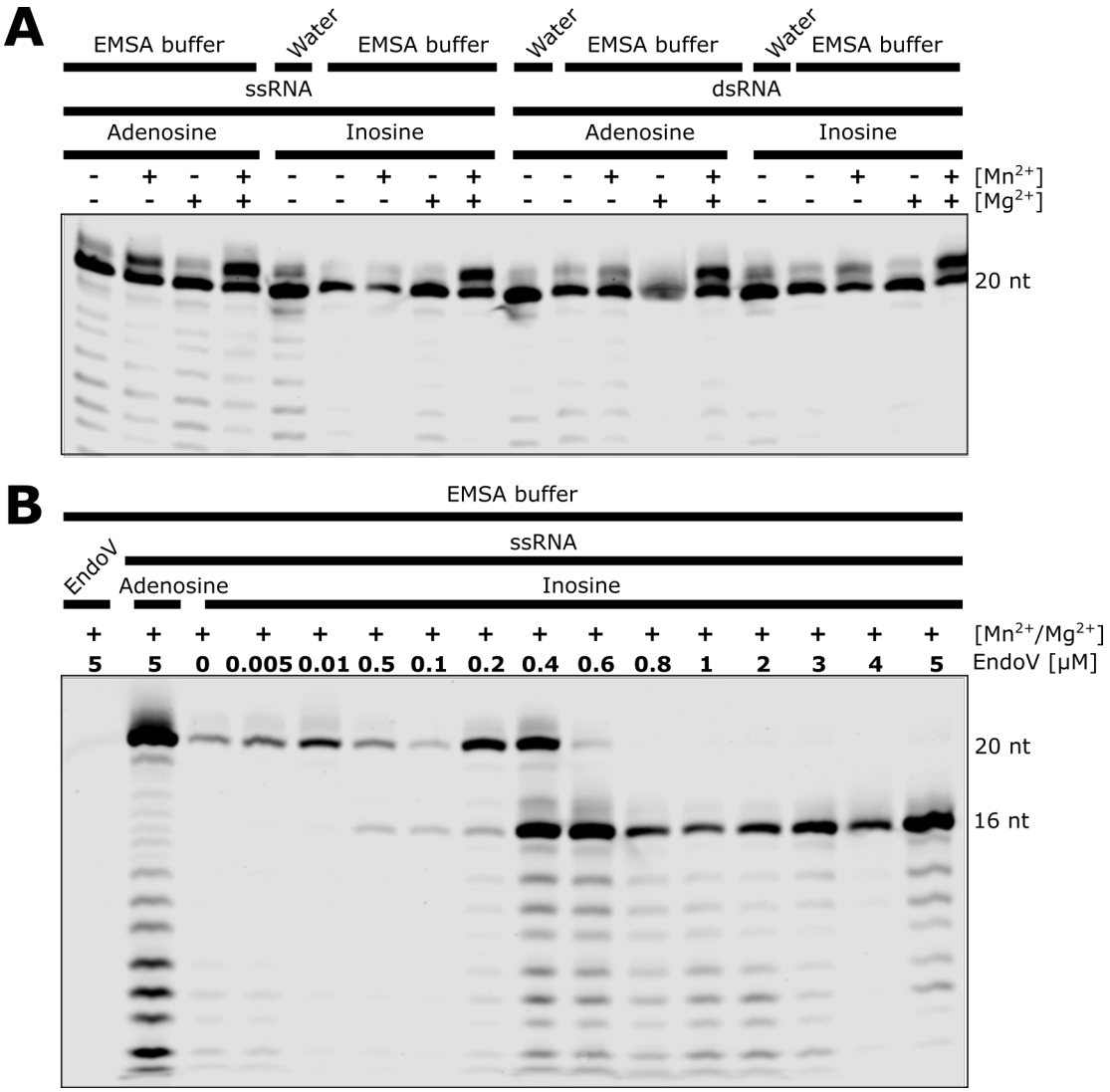


Figure 3.38: Measuring human EndoV activity *in vitro*. Fluorescent scans of denaturing RNA gels. A) Determining the influence of divalent Mg^{2+} and Mn^{2+} ions on migration behaviour of 20-mer RNA oligos. B) The minimum concentration of EndoV to cleave 0.1 µM ssRNA in 2 h was defined.

High amounts of EndoV seemed to retard RNA oligos during electrophoresis of the denaturing RNA gel. This behaviour could interfere with and even obscure precise quantification of RNA editing levels *in vitro*. Hence, the minimum concentration of EndoV required to entirely cleave 0.1 µM ssRNA in 2 hours was determined. Figure 3.38B indicated that equimolar concentrations (0.1 µM) of EndoV to substrate RNA are sufficient to achieve around 50 % substrate cleavage. Differing band intensities could likely be a consequence of pipetting errors as the number of fluorescent oligos was the same. Full cleavage of the RNA substrate is achieved at a 10-fold molar excess of EndoV over its substrate. Other strategies to mitigate the issue would be to use proteinase K to break down all EndoV after the reaction or to extract all RNA substrates by phenol:chloroform.

3.8.4 5' and 3' end fluorescent labelling of *in vitro* transcribed RNA

The previous set of experiments (section 3.8.3) confirmed that recombinant full-length human EndoV could cleave ss and dsRNAs (in the presence of manganese) containing inosines. Therefore, EndoV can be used to detect A-to-I RNA editing executed by *D. melanogaster* ADAR *in vitro*. To use the RNA templates (figure 3.23) as substrates for *in vitro* detection of adenosine to inosine deamination using EndoV, substrates need to be fluorescently labelled first. Therefore, RNA was labelled with fluorescent dyes on either end (5' and 3'), and the labelling efficiency was compared. Different chemical properties of the 5' and 3' end make labelling of *in vitro* transcribed RNAs with fluorescent dyes possible. There are two free hydroxyl groups at the 3' end, and the 5' end can be specifically modified (addition of a thiol group) during either *in vitro* transcription or afterwards using phosphatase and kinase enzymatic activity. The basic principle of 5' end labelling is the introduction of a reactive thiol group at the α-phosphate group of the first nucleotide (due to the transcription start site of the T7 promoter, this is always guanosine) (figure 3.39). Then a maleimide molecule with a fluorescent dye can be coupled with the thiol group to form a thioether bond, covalently attaching the fluorescent label to the RNA.

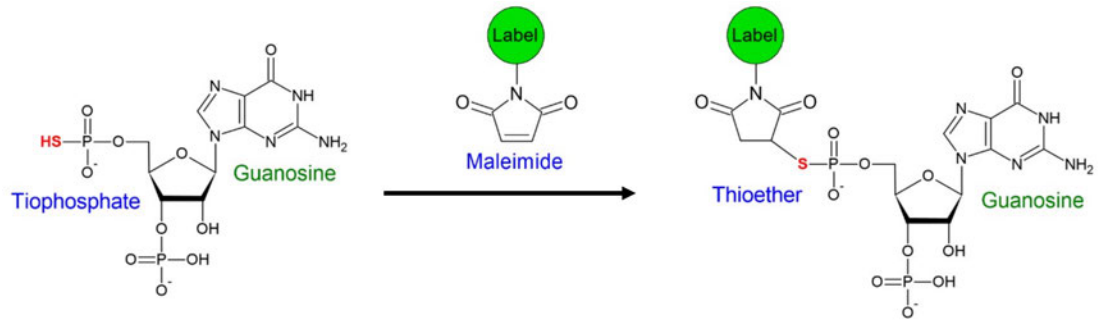


Figure 3.39: 5' end labelling strategy. Principle of 5' end RNA labelling using maleimide-coupled fluorophores to form a covalent bond with the thiolated α-phosphate group of RNA.

The 3' end labelling strategy makes use of a chemical approach, where the most distant ribose is oxidised in a Malaprade reaction using $NaIO_4$. The formed aldehydes can then react with a hydrazide before the reaction product can be reduced using $NaBH_4$ creating the end product (figure 3.40).

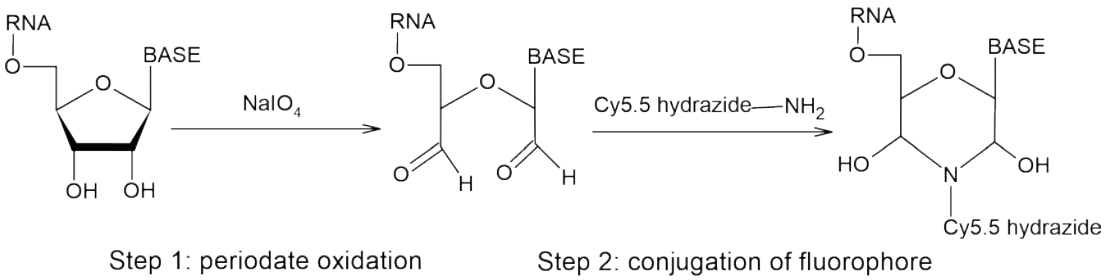


Figure 3.40: 3' end labelling strategy. Principle of 3' end RNA labelling using Cyanine5.5, as the fluorophore coupled to hydrazide, to form a covalent bond with the nucleotide carrying two hydroxyl groups at the 3' end of the RNA.

There are two ways to create an α-phosphate group at the first nucleotide of *in vitro* transcribed RNAs. One approach is to use a GMPαS nucleotide in excess of GTP in transcription reactions to increase the chance for GMPαS incorporation. This approach would make immediate fluorescent labelling possible and additional steps redundant after transcription. Therefore, a 20 to 1 excess of GMPαS to GTP was tested with all 8 RNA templates (figure 3.23) to determine the effect on transcription efficiency and total RNA levels. For all reactions, the same amount of input DNA was used. All samples containing GMPαS could be transcribed (figure 3.41). However, the amount of RNA after transcription was lower for these samples than for the controls (GTP-containing). Particularly low RNA levels were detected for Caps, NaCP60E, and Sxl when

transcribed with GMP. For para there was no transcribed product present when GMP was included.

Due to low or no RNA levels after transcription for most templates, GMPαS was not further used and instead, a second approach was tested using two additional enzymatic reactions for incorporating the necessary thiol group.

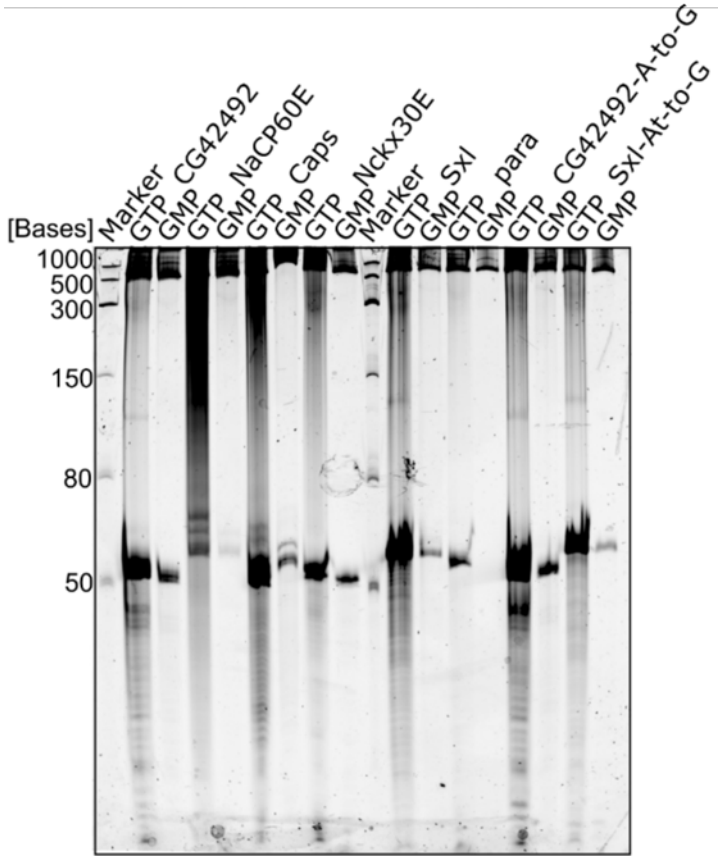


Figure 3.41: Comparison of RNA levels after *in vitro* transcription using either an excess of GMPαS over GTP or solely GTP. All eight RNA templates were *in vitro* transcribed using either an excess of GMPαS over GTP or solely GTP and then separated on a SYBR Safe stained denaturing polyacrylamide:urea gel.

5' end labelling with this method has several advantages: 1) uniform labelling of the 5' end RNA due to enzymatic activities, 2) preservation of the unmodified 3' end of the oligonucleotide, and 3) it is compatible with several labels, including fluorophores and biotin. To prepare the *in vitro* transcribed RNAs for 5' end labelling, their triphosphate 5' end needs to be thiolated for coupling with maleimide-fluorophores using two enzymatic steps:

1. 5' phosphorylated ends are converted to 5' -OH groups by alkaline phosphatase

2. T4 polynucleotide kinase transfers a thiophosphate from ATPγS to the 5' -OH group of the *in vitro* transcribed RNA

The two RNA templates CG42492 (Zn72D-independent) and Sxl (Zn72D-dependent) were *in vitro* transcribed and then thiolated at their 5' end using the Vector Laboratories kit according to their protocol. For 5' end labelling of these two RNAs, four different fluorophores (fluorescein, TAMRA, Cy5, and Atto700, see figure 7.6) coupled to maleimide were used. Then, the resulting fluorescence signal and the amount of RNA used for labelling (the same denaturing gel stained with SYBR Safe) were compared. The intensity of the fluorescent signal, rather than the number of RNAs labelled was the basis for deciding which fluorophore is most suited for 5' end labelling of RNA.

SYBR Safe staining of the labelling reactions using CG4292 RNA confirmed that the amount of RNA used is nearly equivalent (figure 3.42) and migrated at the expected size of around 50 nt. The fluorescence scan showed two signals for Atto700-labelled RNA migrating slower than the 40-mer ssRNA. One fluorescent signal corresponds with the migrating front of the Atto700 fluorophore on its own, and the other one represents labelled RNA. Even though the 5' labelled RNAs were phenol:chloroform extracted and ethanol precipitated, there was still some free Atto700 left in the reaction, which might be an undesired side effect for further biochemical or binding assays. Cy5 also indicated successful labelling, but the fluorescent signal was much weaker compared with Atto700. Unfortunately, fluorescein and TAMRA fluorescence was not detected due to the use of a LI-COR Odyssey CLx system and detection at 700 nm. Both their emission spectra are in the shorter wavelength range and do not show emission at 700 nm (figure 7.6).

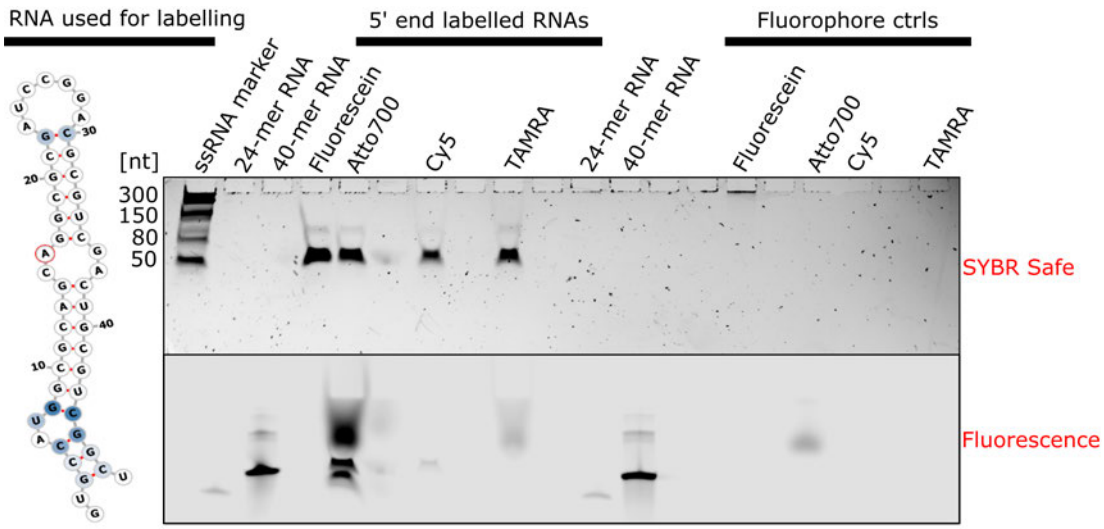


Figure 3.42: 5' end labelling using CG42492 RNA. Transcribed CG42492 RNA (48 nt) was 5' end labelled with one of the four fluorophores: fluorescein, Atto700, Cy5, or TAMRA. The denaturing RNA gel was scanned first for detecting fluorescence at 700 nm and then stained with SYBR Safe.

To corroborate the previous findings, *in vitro* transcribed Sxl RNA was used for 5' end labelling using the same reaction conditions and four fluorophores (figure 3.43). Also, for Sxl, the amount of RNA used for labelling reactions was equivalent after SYBR Safe staining. The fluorescence scan recapitulated the results from the previous labelling approach as Atto700 was successfully attached at the 5' end of Sxl, but none of the other fluorophores showed signs of labelled RNA. As previously mentioned, fluorescein and TAMRA are not suited for detecting fluorescence at 700nm (figure 7.6). However, it is unclear why there is only very little fluorescence detectable for cyanine 5, even though its emission should be detectable at 700 nm.

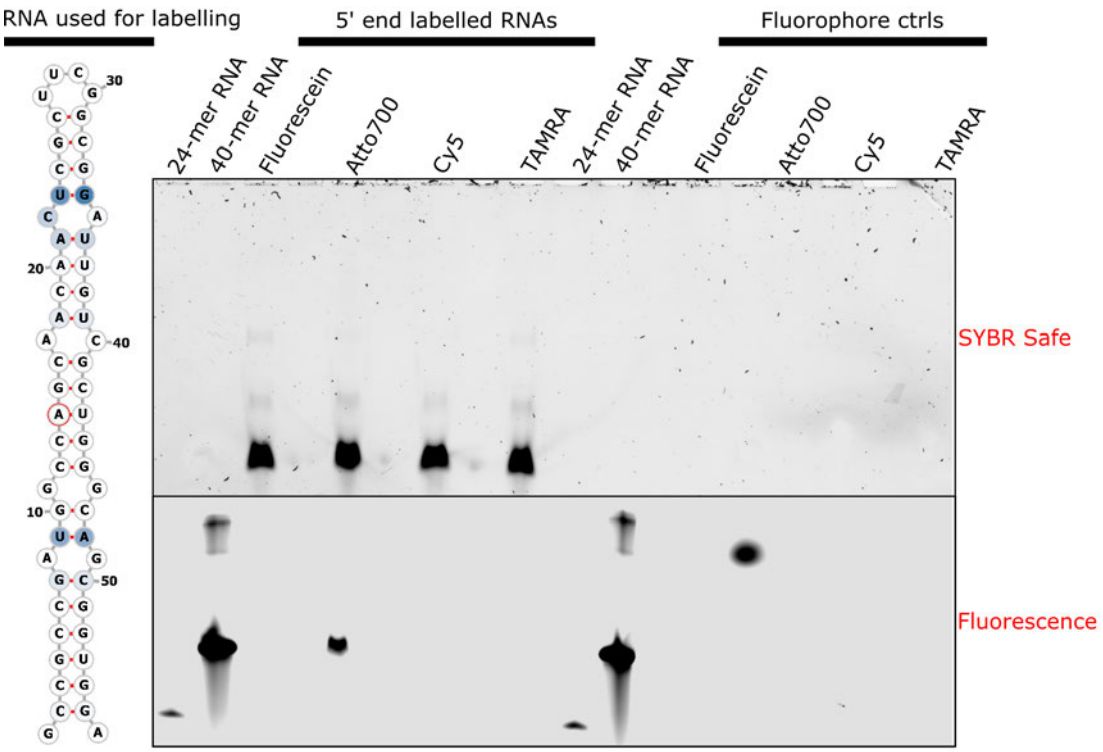


Figure 3.43: 5' end labelling using SxI RNA. Transcribed SxI RNA (56 nt) was 5' end labelled with one of the four fluorophores: fluorescein, Atto700, Cy5, or TAMRA. The denaturing RNA gel was scanned first for detecting fluorescence at 700 nm and then stained with SYBR Safe.

In the previous experiments, 5' end labelling of RNA Atto700 was identified as the fluorophore best suited out of four candidates. Therefore, the amount of Atto700 used for labelling was reduced to facilitate subsequent removal of the fluorophore by phenol:chloroform extraction. SYBR Safe staining of the labelling reactions using CG4292 and CG42492 A-to-G RNAs confirmed that the amount of RNA used is nearly equivalent (figure 3.44) and migrated at the expected size of around 50 nt. Furthermore, the fluorescent signal for both RNAs with the lower excess of Atto700 (50x) showed equivalent fluorescent signals to the higher excess (200x), indicating that there is no disadvantage in using less fluorophore for 5' end labelling.

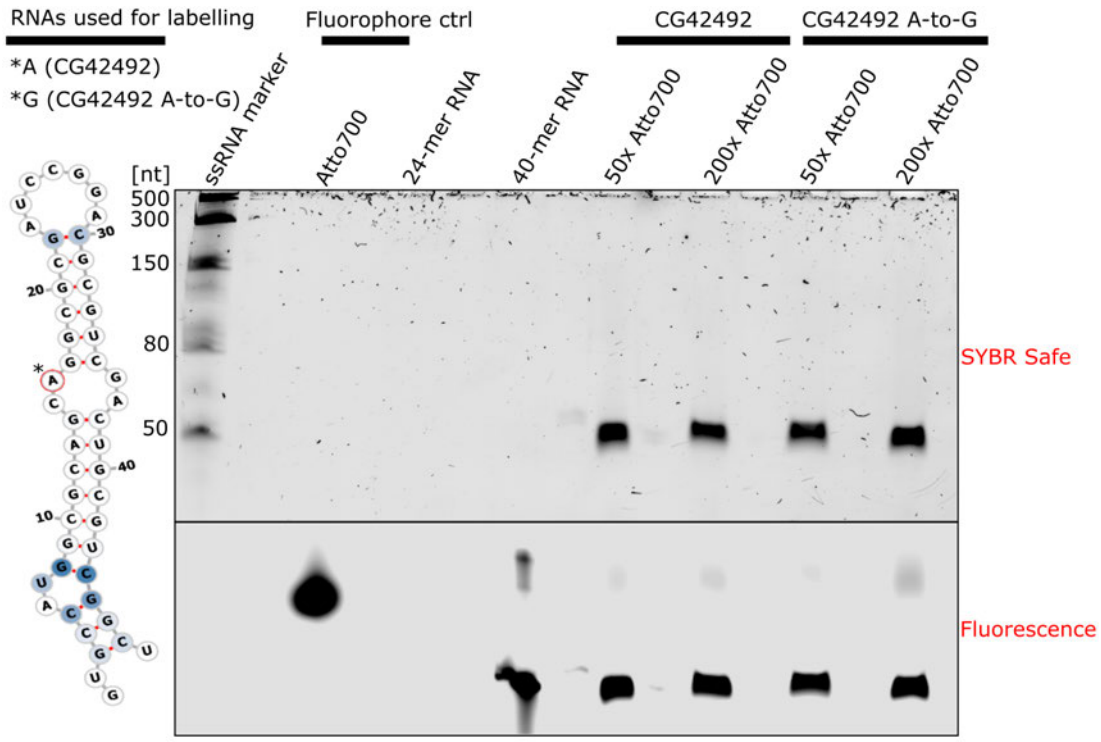


Figure 3.44: 5' end labelling using Atto700 excesses. Transcribed CG42492 RNA (48 nt) was 5' end labelled with Atto700 in 50x or 200x excess. The denaturing RNA gel was scanned first for detecting fluorescence at 700 nm and then stained with SYBR Safe.

After testing a method for 5' end labelling of RNA relying on initial enzymatic activity, a chemical approach was tested by oxidising the 3' ribose and attaching a hydrazide-coupled cyanine 5.5 (figure 7.7) on it. *In vitro* transcribed CG42492 was cross-linked using the strong oxidising agent $NaIO_4$ at two different temperatures and two different concentrations.

1. 4 ℃, 10 mM $NaIO_4$
2. 4 ℃, 1 mM $NaIO_4$
3. RT, 10 mM $NaIO_4$
4. RT, 1 mM $NaIO_4$

SYBR Safe staining of the labelling reactions using CG4292 RNA confirmed that the amount of RNA used is nearly equivalent (figure 3.45) and migrated at the expected size of around 50 nt. Comparing the intensity of the fluorescent signals for the four conditions, they are almost equivalent. Therefore, RNA can be labelled at the 3' end using all four reaction conditions, adding another possibility for RNA labelling.

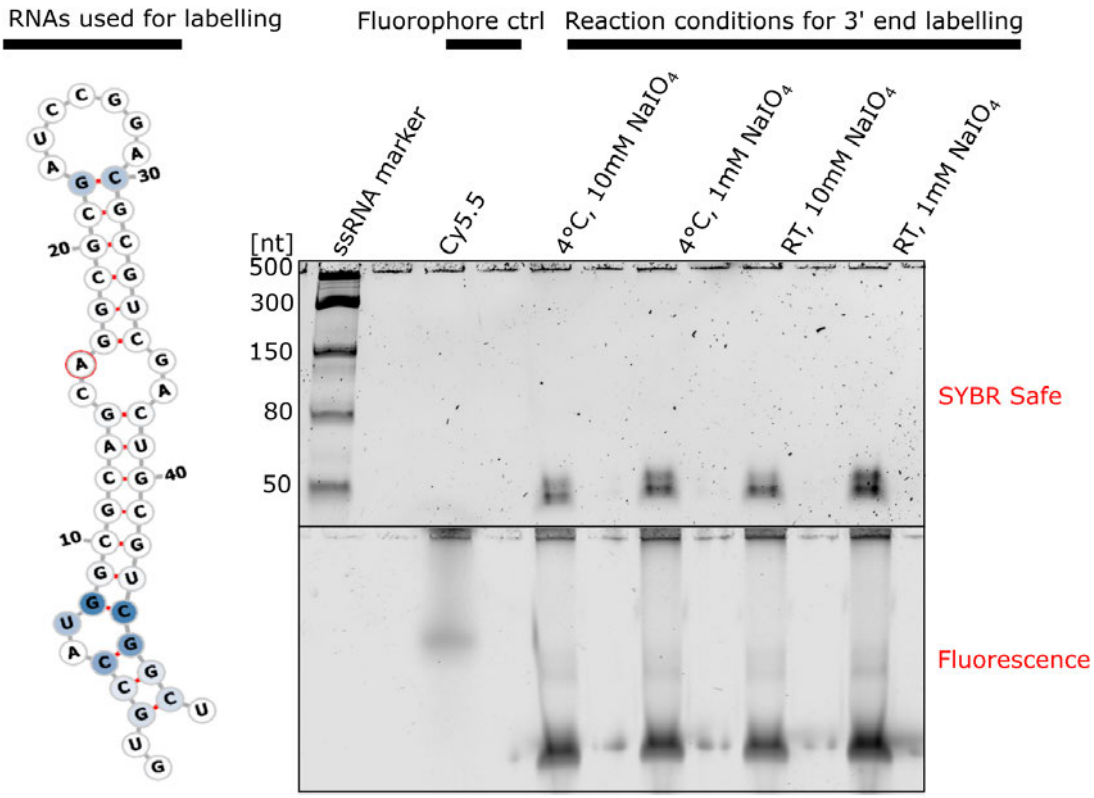


Figure 3.45: Figure. Transcribed CG42492 RNA (48 nt) was 3' end labelled with Cy5.5. The denaturing RNA gel was scanned first for detecting fluorescence at 700 nm and then stained with SYBR Safe.

4 Investigating the mechanism of alternative splicing regulation by ZFR

In order to gain a mechanistic understanding of the role of ZFR in alternative splicing regulation, it is important to characterise its RNA binding specificity. Identifying RNA motifs that are bound by ZFR might help us finding out more about its precise biological role in splicing. Therefore, it is essential to develop a robust method for the expression and purification of recombinant ZFR for being able to carry out *in vitro* assays for biochemical, biophysical and structural characterisation. The ultimate goal is to find a construct that can be used to produce a stable, soluble, and active protein. Human ZFR expression constructs carrying an N-terminal tag (either 6xHis or 6xHis-GST) were tested for their protein production in bacteria. Once this has been successfully optimised, ZFR constructs were planned to be used in an RNA Bind-n-Seq experiment (RBNS) performed by the Hogg group for the identification of RNA binding motifs *in vitro*.

4.1 Optimisation of expression conditions of human ZFR and NF45 in bacteria

Various small-scale test expression screenings were carried out to establish a robust method for protein production of human ZFR and NF45 (figure 4.1). The first 316 amino acids of ZFR's unstructured N-terminal sequence were excluded to enhance its expression and solubility. Two human NF45 constructs were tested. One lacks the first 28 N-terminal amino acids, and the other one was the full-length version (figure 4.1).

Both NF45 constructs carried an N-terminal 6xHis-GST-tag and were expressed in BL21 cells (figure 4.1A). Expression levels of the N-terminally trun-

cated NF45 were higher than for full-length NF45. ZFR (317-1074) carried either an N-terminal 6xHis-tag or 6xHis-GST-tag and was expressed in BL21 cells (figure 4.1A). Both GST- and His-tagged ZFR indicated that most proteins were insoluble, but there seemed to be some soluble His-tagged ZFR (figure 4.1B). Therefore, His-tagged ZFR (317-1074) was expressed in a panel of *E. coli* strains including BL21, B834, T7 and RIPL (figure 4.1C+D). All pull-downs showed that His-tagged ZFR is mostly insoluble, but expression levels are quite high. Therefore, ZFR and NF45 were co-expressed to enhance solubility through dimer formation.

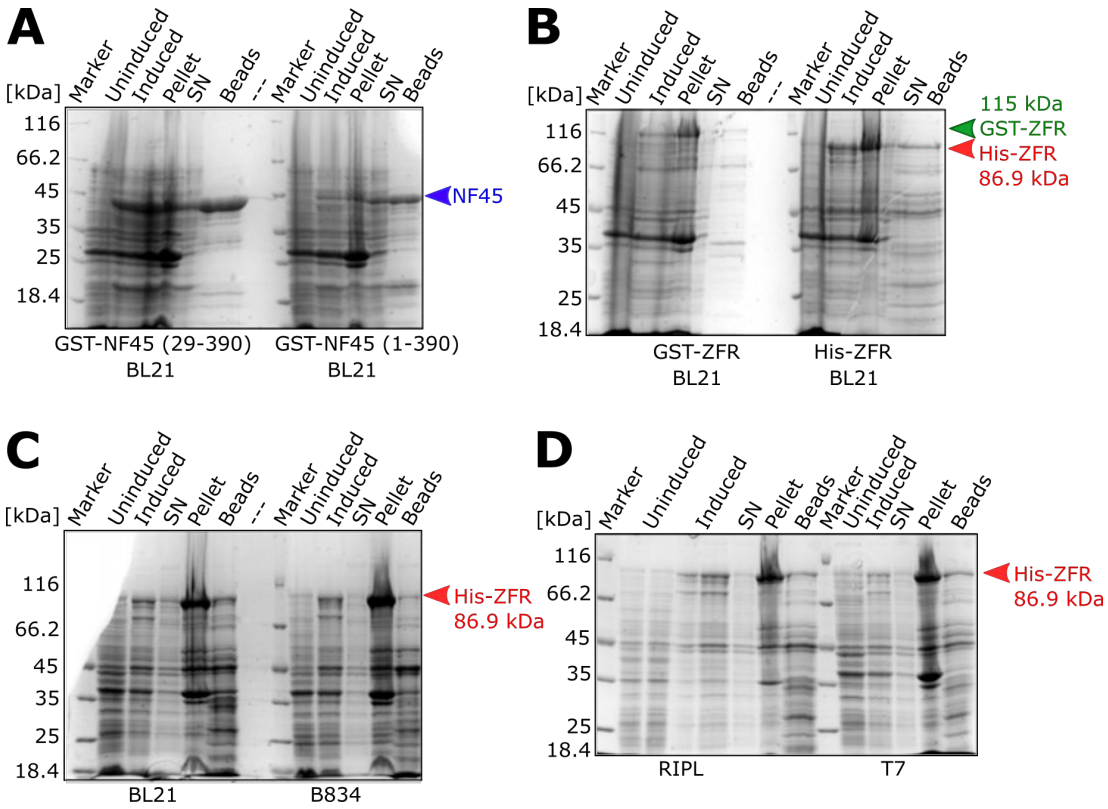


Figure 4.1: Individual test expressions of human ZFR and NF45. SDS-PAGEs of pull-downs of test expressions of human ZFR and NF45 in bacteria. A) GST-NF45 (29-390) in BL21 (left) and GST-NF45 (1-390) in BL21 (right). B) GST-ZFR (317-1074) (left) and His-ZFR (317-1074) in BL21 (right). C) and D) His-ZFR (317-1074) in BL21 (right) and B834 (left) C) and RIPL (right) and T7 (left) D). SN - Supernatant.

For co-expression, both ZFR and NF45 carrying an N-terminal 6xHis-tag were expressed in a panel of *E. coli* strains including BL21, B834, and T7 (figure 4.2A+B). Protein levels for ZFR and NF45 were higher upon co-expression than before for individual expression (figure 4.1A). Co-expression in B834 (figure 4.2A) led to the highest levels of soluble protein of all three strains and was

therefore chosen as the default strain for protein expression.

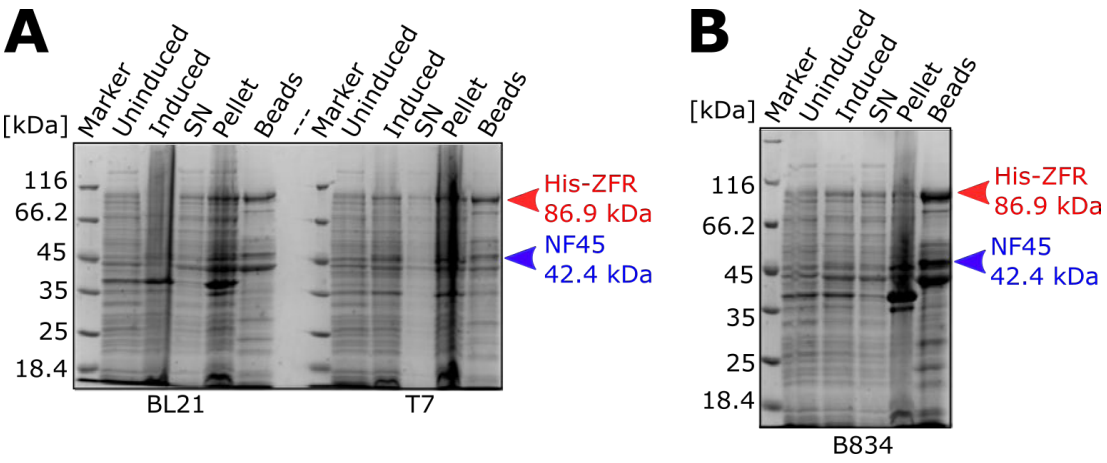


Figure 4.2: Test co-expressions of human ZFR and NF45. SDS-PAGEs of pull-downs of test co-expressions of human His-ZFR (317-1074) and His-NF45 (29-390) in bacteria. A) In BL21 (right) and T7 cells (left). B) In B834 cells. SN - supernatant.

To further validate enrichment of co-expressed His-ZFR and His-NF45 after pull-down, three western blots (figure 4.3) were carried out using two different antibodies, anti-ZFR and anti-NF45, and also a NiNTA-HRP conjugate, to detect recombinant proteins carrying a His-tag.

Both bands corresponding to His-ZFR and His-NF45 were detected by NiNTA-HRP in the enriched fraction (Beads), indicated by the two major bands at around 95 kDa and 40 kDa (figure 4.3A). The additional bands are likely to be degradation products of both ZFR and NF45. The negative control (Uninduced samples) did not show any sign of protein expression. The protein amount (His-tagged 40 kDa protein of unknown origin purified and provided by Valdeko Kruusvee) of the positive control loaded for the NiNTA-HRP western blot led to over-saturation of the signal.

The anti-NF45 western blot (figure 4.3B) could also detect His-NF45 (around 40 kDa) in the enriched fraction (Beads) as in the NiNTA-HRP western blot before. Both western blots confirmed that the 40 kDa band correspond with His-tagged NF45. A GST-tagged version of NF45 (purified and provided by Uma Jayachandran) served as the positive control, seen at its higher molecular weight of around 70 kDa.

Unfortunately, there was no positive control available at the time for the anti-ZFR western blot (figure 4.3C), but His-ZFR was detected again due to a specific signal at around 95 kDa as in the NiNTA-HRP western blot before. This indicates, as for NF45, successful expression of His-tagged ZFR even though

it migrated slower as expected.

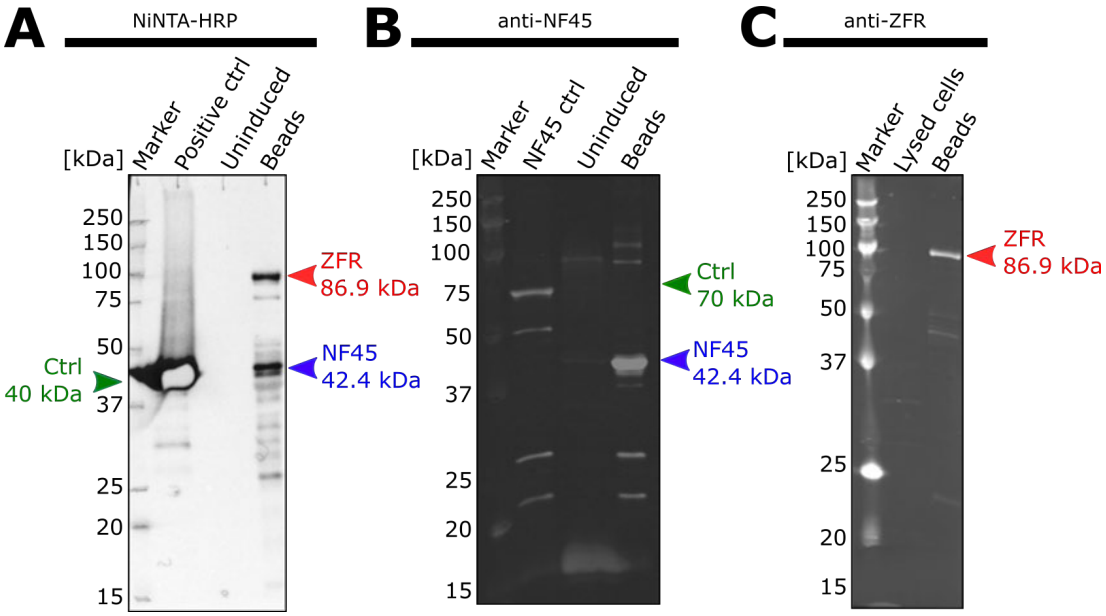


Figure 4.3: Testing expression levels of ZFR and NF45. Western blots using A) a NiNTA-HRP conjugate to detect His-tagged ZFR and NF45, B) an anti-NF45 antibody, and C) an anti-ZFR antibody. Green arrowheads indicate positive controls; red arrowheads are His-ZFR; blue arrowheads are His-NF45.

4.1.1 Lysis buffer optimisation

The recovery of soluble ZFR/NF45 was tried to be improved to enhance the yield of purified protein for further assays. Low solubility represented the bottleneck in co-expression with NF45. The impact of salt concentration during cell lysis on protein recovery was determined to enhance protein solubility and reduce nucleic acid contamination during protein purification. Several concentrations (200, 400, 600, 800, and 1000 mM) of NaCl in the lysis buffer were tested to see if this would increase the protein levels of ZFR/NF45 that could be pulled down (figure 4.4) and whether this improved the 260/280 ratio in the recovered sample (table 4.1). The proteins remaining bound to the beads after elution (Bound) and the eluted proteins (Eluate) were also prepared for SDS-PAGE analysis. There is an evident increase in soluble ZFR/NF45 from lower salt (200 and 400 mM) to higher salt concentration (600, 800, and 1000 mM), indicating that higher salt promoted solubilisation and extraction of ZFR from lysates.

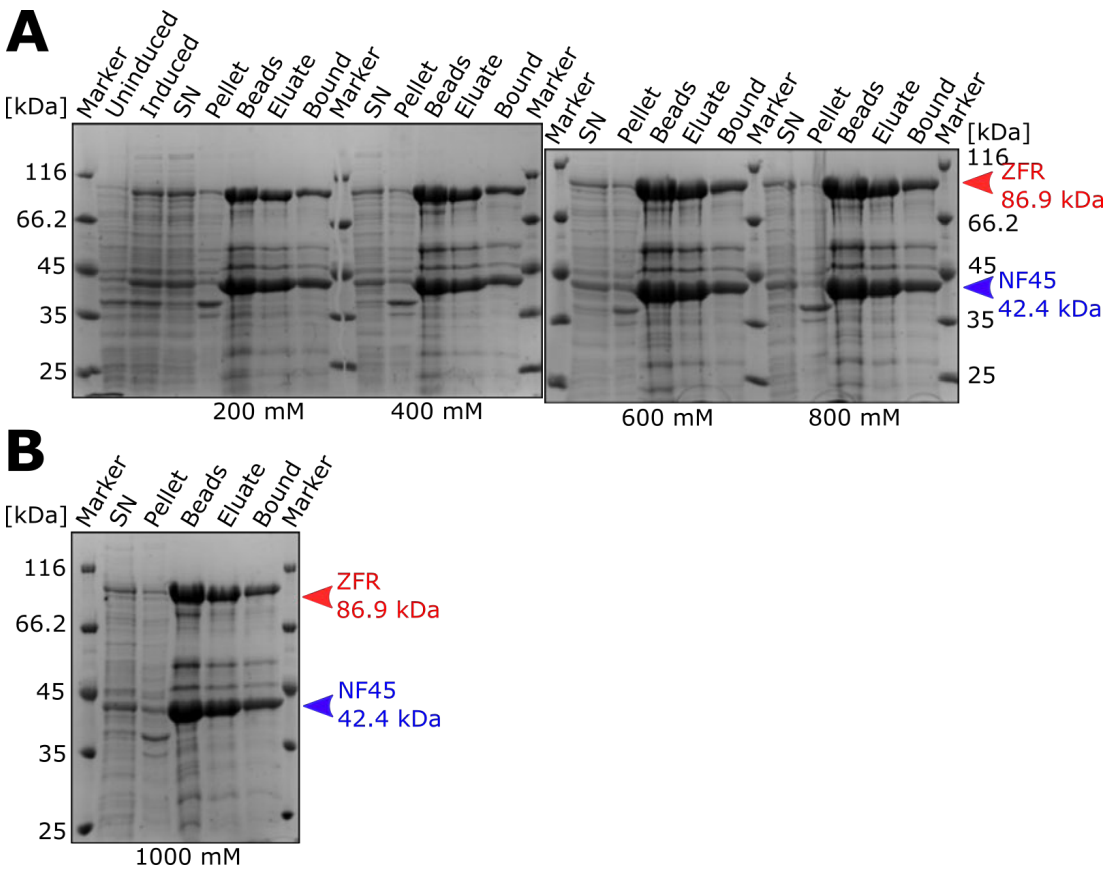


Figure 4.4: Effect of salt concentrations during cell lysis on ZFR/NF45 protein recovery. SDS-PAGEs of pull-downs of co-expressed His-ZFR and His-NF45 lysed with different salt concentrations. A) Salt concentrations of 200, 400, 600 and 800 mM NaCl. B) Salt concentration of 1000 mM NaCl. SN - supernatant.

When nucleic acid contamination was assessed for ZFR/NF45 co-expression (table 4.1), a slight decrease was observed with increasing salt concentrations up to 600 mM. However, the contamination does not decrease further with higher salt concentrations. Consequently, 600 mM NaCl was used as the default concentration in the lysis buffer for large-scale purifications of human ZFR and NF45.

Table 4.1: Effect of salt in lysis buffer on nucleic acid contamination determined by NanoDrop

NaCl [mM]	200	400	600	800	1000
260/280	1.69	1.37	1.18	1.20	1.14

4.2 Protein purification of human ZFR/NF45 heterodimers

The lysis buffer optimisation (4.1.1) was used as a starting point for purifying the human complex comprised of His-tagged ZFR and His-tagged NF45 (figure 4.5). The current protocol consists of five purification steps (1. IMAC, 2. IMAC, heparin, IEX, and SEC) and yields soluble, active heterodimers. The 6xHis-tags were consecutively removed, starting with the His-tag from NF45. The 6xHis-tag on ZFR is retained for the second IMAC step; therefore, passing the sample over a Ni-NTA column again should separate off GST-TEV and excess NF45 but allow binding of the ZFR/NF45 complex.

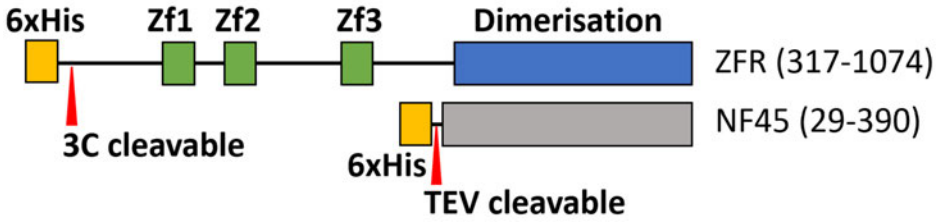


Figure 4.5: His-ZFR and His-NF45 constructs used for protein purification. Human His-ZFR (3C-cleavable) and His-NF45 (TEV-cleavable). Protease cleavage sites are indicated by a red arrowhead.

Following cell lysis and separation of soluble and insoluble components of the lysate by centrifugation, the supernatant was incubated with NiNTA beads for batch purification. The beads were washed several times with NiNTA buffer A, and the removal of unbound (Unbound) and non-specific bound proteins (Wash) can be seen in the corresponding SDS-PAGE gel (figure 4.6). Bound proteins were eluted with NiNTA buffer B containing 1000 mM imidazole. The SDS-PAGE gel shows that the complex was eluted with some additional contaminations. A small proportion of the ZFR/NF45 complex is retained on the NiNTA beads after elution. The indicated pooled fractions of the eluate were used for further purification. The His-tag of NF45 was cleaved using GST-TEV before moving to a second immobilised metal affinity chromatography step to re-capture ZFR/NF45 complexes containing His-tagged ZFR.

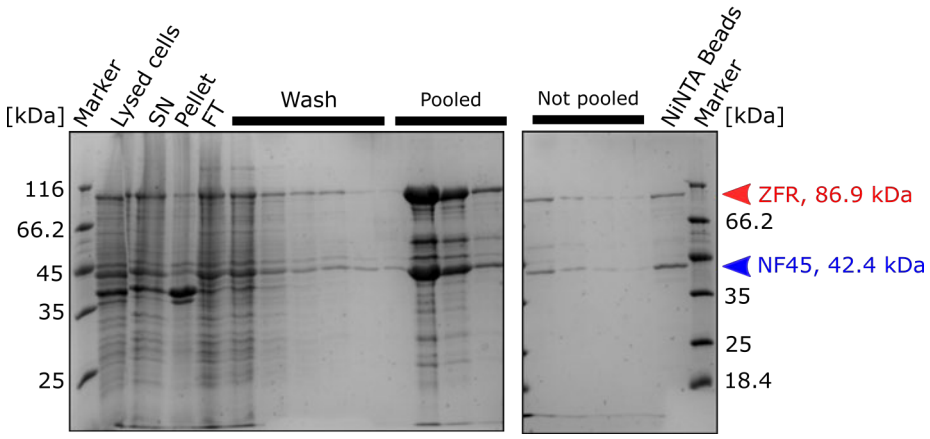


Figure 4.6: First immobilised affinity chromatography step of ZFR/NF45 purification. SDS-PAGE of batch purification of human ZFR/ NF45. ZFR (red) and NF45 (blue) are indicated by arrow heads. SN - supernatant. FT - flow-through.

After cleaving the 6xHis tag from NF45, ZFR/NF45 was incubated again with NiNTA beads and packed into a column to remove GST-TEV and excess NF45 without His-tag but retained the His-ZFR. Cleavage of His-tagged NF45 was difficult to see on an SDS-PAGE gel as the corresponding NF45 (Cleaved) differs only in 1.8 kDa from His-NF45 (Uncleaved) (figure 4.7). Bound ZFR/NF45 was eluted with a linear gradient of NiNTA buffer B until 80%. A single peak containing ZFR and NF45 was observed after elution. The SDS-PAGE showed a quite clean product with a protein present at around 55 kDa that co-elutes and might be a degradation product of ZFR and some GST-TEV (53 kDa). Also, a significant amount of the sample did not bind a second time to the beads (Unbound). The indicated pooled fractions were used for subsequent heparin affinity chromatography.

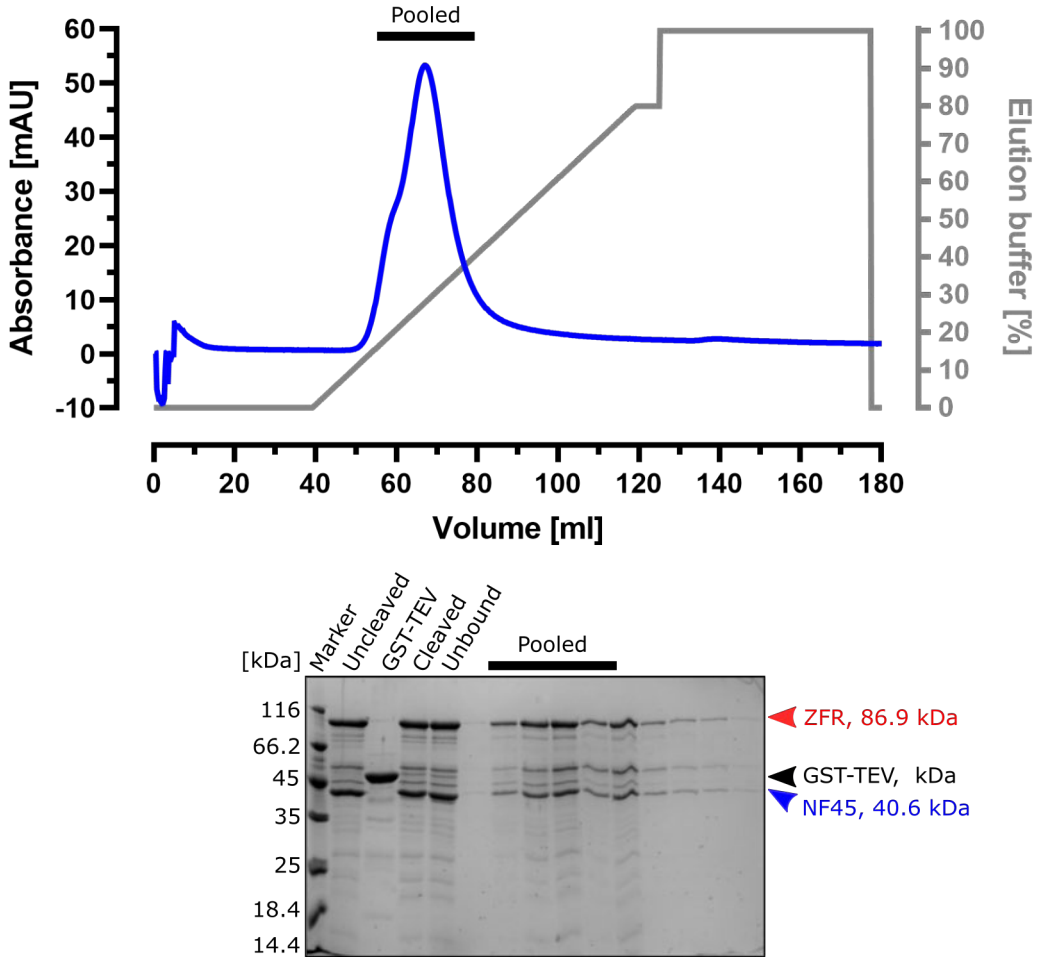


Figure 4.7: Second immobilised affinity chromatography step of ZFR/NF45 purification. Chromatogram representing the UV absorbance at 280 nm (blue) and the increasing percentage of elution buffer (grey) with the corresponding SDS-PAGE. ZFR (red), NF45 (blue), and GST-TEV (black) are indicated by arrow heads.

ZFR/NF45 was loaded onto a heparin sepharose column to remove the contaminating nucleic acids. The corresponding chromatogram and SDS-PAGE gel (figure 4.8) indicated that contaminating nucleic acids could be separated from ZFR/NF45 as there is very little protein detectable in the flow-through fractions (FT) by SDS-PAGE. Bound ZFR/NF45 was eluted with a linear gradient of heparin buffer B to 80% (1000 mM NaCl) and then 100%. A single peak containing ZFR and NF45 was observed upon elution, and the SDS-PAGE gel shows two clean protein products with a low 260/280 ratio of about 0.6. However, there is still some contaminating 55 kDa protein present. Therefore, the indicated pooled fractions were used for subsequent ion exchange chromatography.

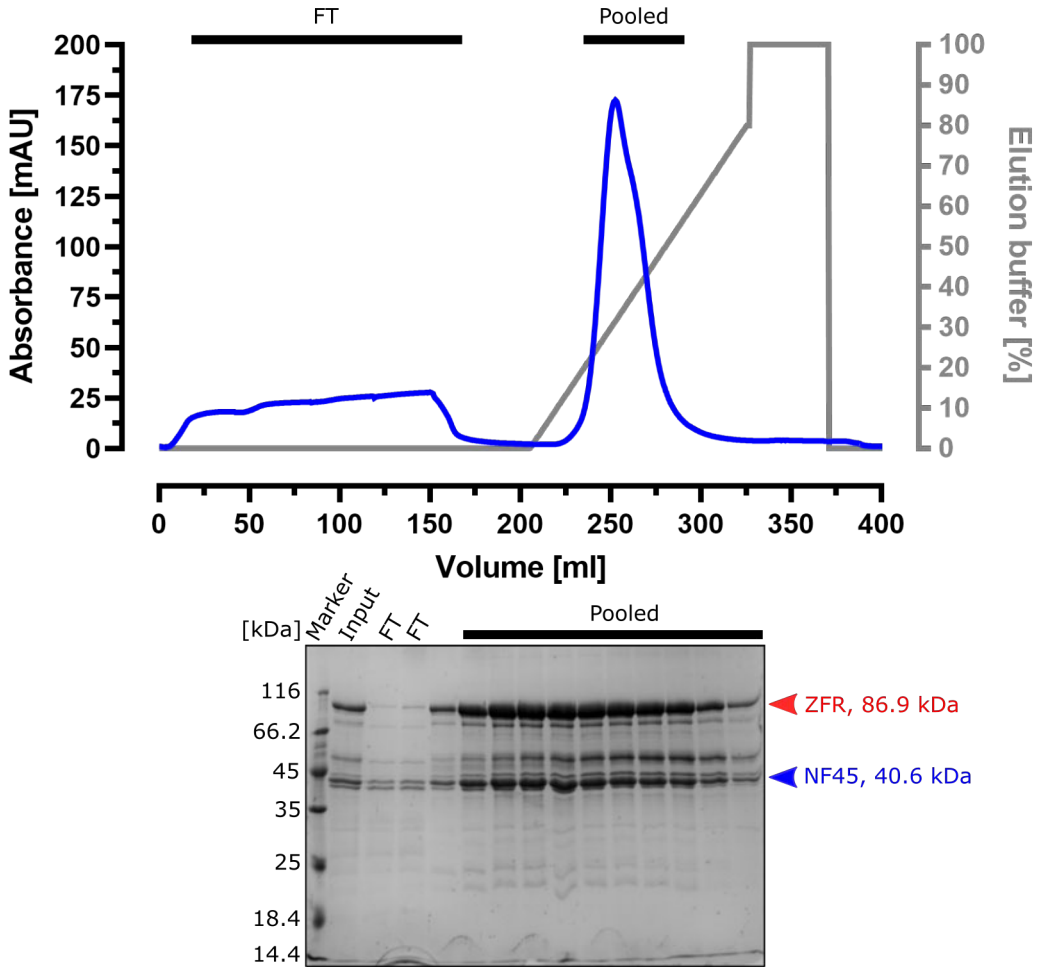


Figure 4.8: Heparin affinity chromatography step of ZFR/NF45 purification. Chromatogram representing the UV absorbance at 280 nm (blue) and the increasing percentage of elution buffer (grey) with the corresponding SDS-PAGE. ZFR (red) and NF45 (blue) are indicated by arrow heads.

ZFR/NF45 was loaded onto a 6 ml S column to remove the remaining protein contaminants. The corresponding chromatogram and SDS-PAGE gel (figure 4.9) indicate that the contaminating 55 kDa protein could be separated from ZFR/NF45. The column was washed with IEX buffer A until it reached a steady UV baseline. Bound ZFR/NF45 was eluted with a shallow linear gradient to 40% IEX buffer B (1000 mM NaCl). Upon further step-wise increase of the amount of elution buffer, no further protein was eluted from the column. A single peak containing ZFR and NF45 was observed upon elution, and the SDS-PAGE gel shows two very clean protein products. The indicated pooled fractions were used for subsequent size exclusion chromatography.

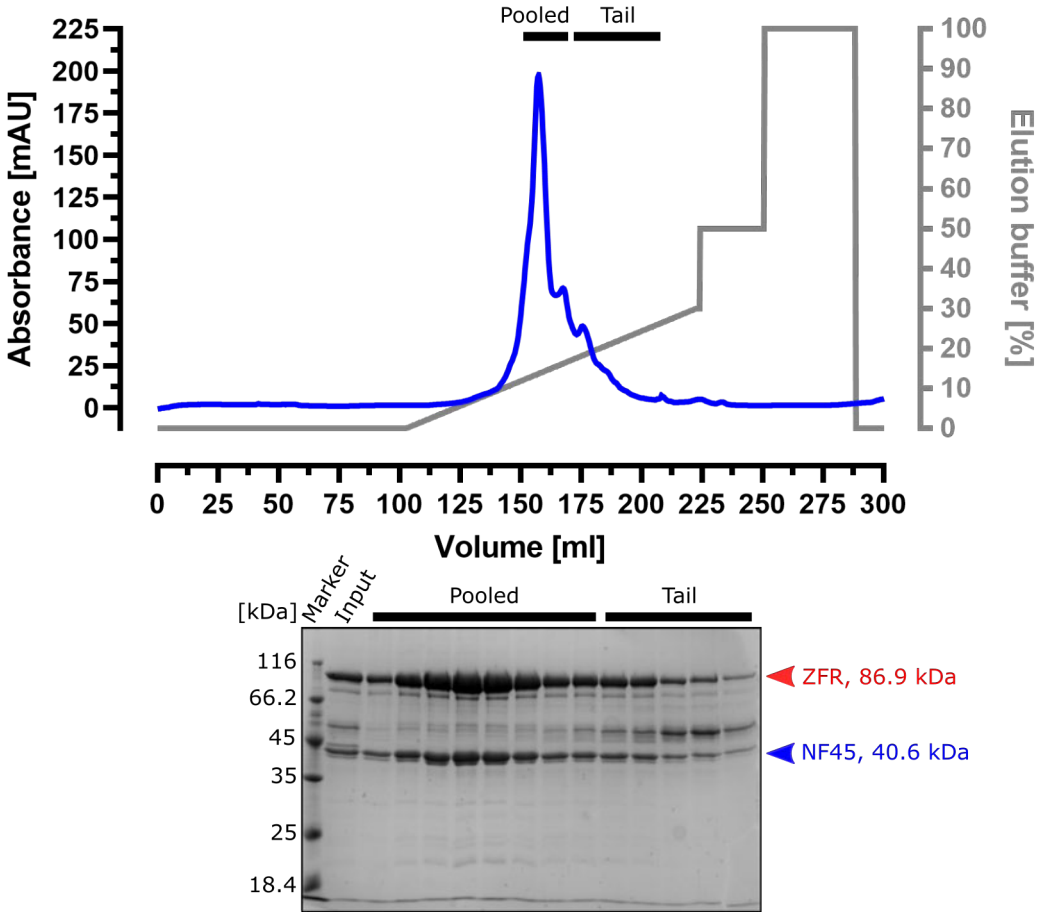


Figure 4.9: Ion exchange affinity chromatography step of ZFR/NF45 purification. Chromatogram representing the UV absorbance at 280 nm (blue) and the increasing percentage of elution buffer (grey) with the corresponding SDS-PAGE. ZFR (red) and NF45 (blue) are indicated by arrow heads.

As a final polishing step, ZFR/NF45 was loaded onto a 120 ml S200 10/60. A small peak around the void volume was observed, with the majority of ZFR/NF45 eluting in a single peak around 56 mL observed in the chromatogram and the corresponding gel (figure 4.10). The first peak likely is a higher molecular weight protein aggregate eluting around the void. The SDS-PAGE gel shows two very clean protein products that co-elute, and the chromatogram shows a 260/280 ratio of about 0.6. The final yields were typically around 2 mg of pure protein per 6 L of bacterial cells.

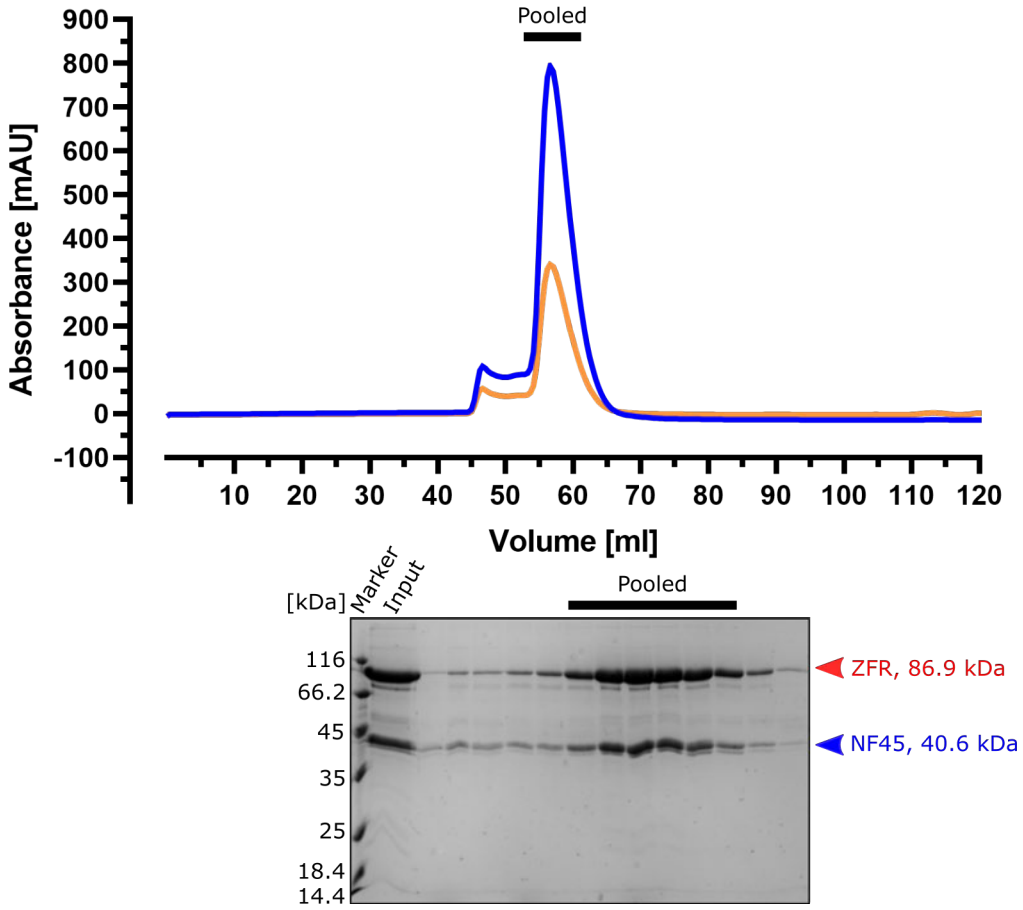


Figure 4.10: Size exclusion chromatography step of ZFR/NF45 purification. Chromatogram representing the UV absorbance at 280 nm (blue) and at 260 nm (orange) with the corresponding SDS-PAGE. ZFR (red) and NF45 (blue) are indicated by arrow heads.

4.3 Identification of RNA-binding motifs of ZFR/NF45

To gain a mechanistic understanding of ZFR's role in regulating hundreds of alternative splicing events (Haque et al., 2018), it is crucial to characterise its RNA-binding specificity. Cross-linking immunoprecipitation-high-throughput sequencing (CLIP-Seq) would be the first-choice method for the identification of RNA-binding motifs of a protein of interest *in vivo*. However, the CLIP-Seq data from the Hogg group indicated that ZFR cross-links through an extended region across skipped exons, making it impossible to extract valid binding motifs. Hence, an alternative approach called RNA Bind-n-Seq (RBNS) (N. Lambert et al., 2014; N. J. Lambert et al., 2015) was chosen, which is an adopted

method to study and quantify protein-RNA interactions *in vitro* based on HT-SELEX (Jolma et al., 2010) and Bind-n-Seq (Zykovich et al., 2009) methods. In this method, a recombinant RBP with a specific tag is purified and added to a randomised pool of nucleic acids at various concentrations. RNA-protein complexes are pulled down using the specific tag, and then all protein-bound RNA is eluted, reverse-transcribed, and sequenced (figure 4.11). Using multiple protein concentrations to optimise analysis at different ranges of affinity helps estimate relative dissociation constants and assess the effects of RNA secondary structure on binding. RBNS should also be able to eliminate systematic biases, and sources of false positive and false negative hits of CLIP-Seq as previous studies have shown that uridines form RNA-protein cross-links easier than other bases, thereby favouring detection of U-rich sequences by CLIP-Seq (Sugimoto et al., 2012).

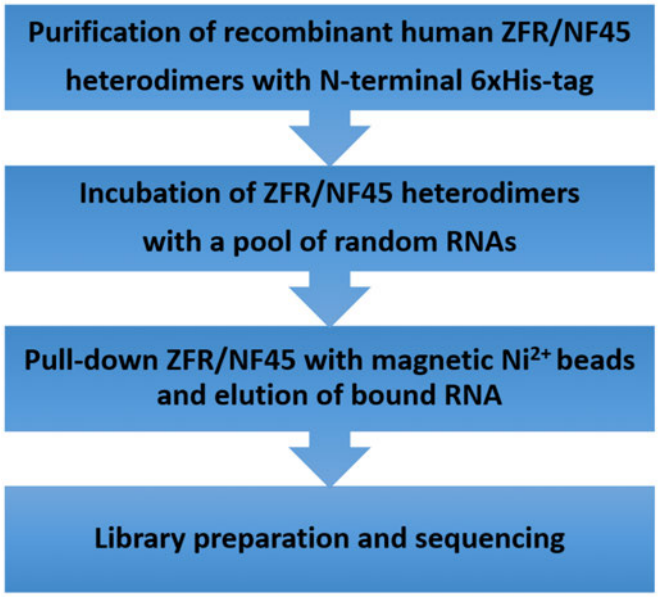


Figure 4.11: Overview of the experimental workflow of RBNS. 6xHis-tagged ZFR/NF45 was incubated with a pool of RNA oligonucleotides. Then, ZFR/NF45 was pulled down, and the associated RNA was eluted, prepared as sequencing library and then sequenced.

Two different recombinant preparations of the human ZFR/NF45 complex, either containing (long-ZFR/NF45) or lacking the three zinc-finger domains (DZF-ZFR/NF45), were produced by co-expression of ZFR with NF45. Both protein constructs were purified using affinity (NiNTA beads and heparin), ion exchange, and size exclusion chromatography steps (figure 4.12) while at the same time keeping their His-tag for RBNS. DZF-ZFR/NF45 indicated the presence of several oligomeric states based on a diverse elution profile (figure

4.12A+B). Long-ZFR/NF45 (figure 4.12C) also showed a diverse elution profile, but several low molecular-weight protein contaminants could be removed (figure 4.12D). Both heterodimers have low background contamination of RNA as indicated by the 260/280 ratio is around 0.5.

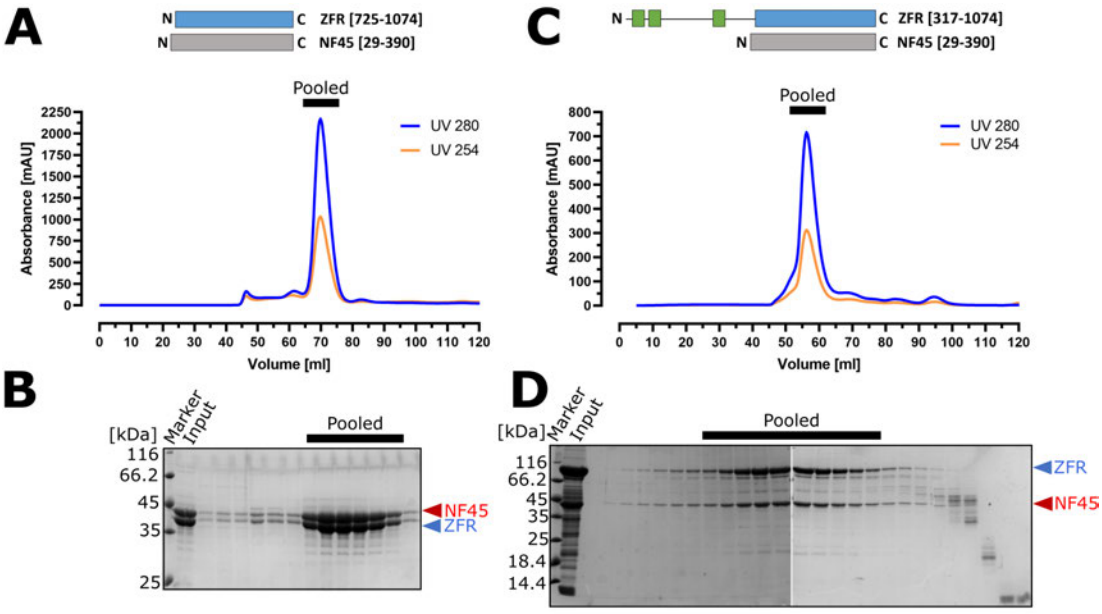


Figure 4.12: Size exclusion chromatography of ZFR/NF45 heterodimers destined for RBNS. Size exclusion chromatograms representing the UV absorbance at 280 nm (blue) and at 260 nm (orange) with corresponding SDS-PAGEs. A) + B) DZF-ZFR/NF45 lacking its zinc fingers. C) + D) long-ZFR/NF45 with the zinc fingers present. ZFR and NF45 are indicated by red and blue arrow heads next to the SDS-PAGE gels. ZFR (blue) and NF45 (red) are indicated by arrow heads.

To ensure that RNA binding is mediated via the the zinc finger domains, single-stranded oligo-U and double-stranded U/A RNA probes were used for EMSAs with long-ZFR/NF45 and DZF-ZFR/NF45 (figure 4.13). Both RNAs are fluorescently-labelled and should display a shift on the gel upon RNA binding, representing an RNA-protein complex with higher molecular weight. Long-ZFR/NF45 exhibited a strong binding preference for dsRNA over ssRNA. In contrast, DZF-ZFR/NF45 lacking zinc fingers showed very weak RNA binding with both probes. There is almost no binding to the ssRNA and very little binding to the dsRNA which is quite difficult to observe. Since the long form of ZFR showed tight RNA-binding ability, it passed this initial quality control step. Furthermore, the short construct of ZFR/NF45 showed weak RNA binding activity, indicating that it was a suitable negative control for the experiment. Both preparations were sent to the Hogg lab for subsequent RBNS experiments.

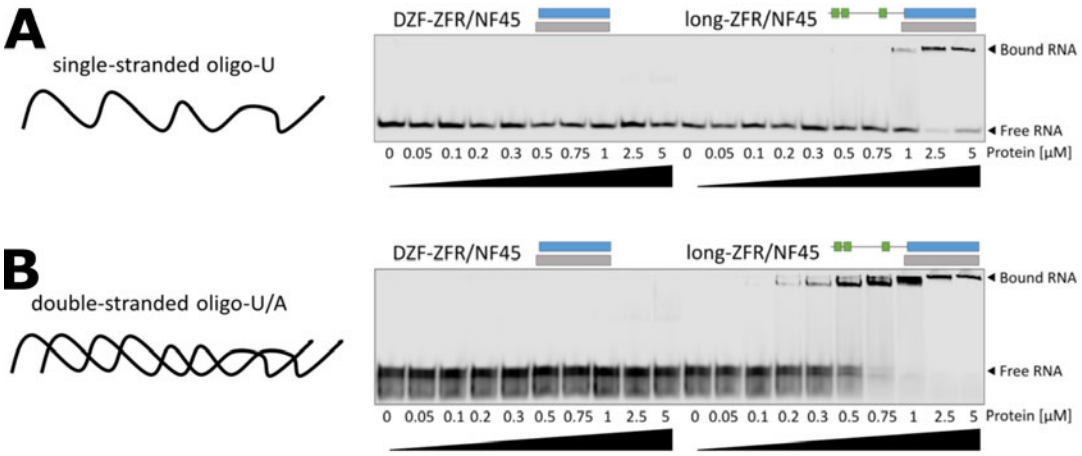


Figure 4.13: Assessing RNA-binding of ZFR/NF45 heterodimers as quality control for RBNS experiments. Fluorescent scan of EMSAs using recombinant DZF-ZFR/NF45 (left) and long-ZFR/NF45 (right) heterodimers and fluorescently-labelled RNA probes. A) EMSA with with single-stranded oligo-U 20-mer probes. B) EMSA with double-stranded oligo-U/A 20-mer probes.

To perform RBNS, the Hogg lab incubated a library of *in vitro* transcribed RNAs containing 40 nt long random sequences flanked by primer binding sites to prepare a sequencing library with purified long-ZFR/NF45 while DZF-ZFR/NF45 and recombinant poly-pyrimidine tract binding protein 1 (PTBP1) served as controls. DZF-ZFR/NF45 served as control as it is soluble, stable, well-behaved and does not bind RNA. The recombinant proteins were affinity purified using magnetic Ni^{2+} beads, and sequences enriched or depleted from the starting pool were identified by high-throughput sequencing (figure 4.11). In addition, all motifs derived from DZF-ZFR/NF45 were removed from the long-ZFR/NF45 data to account for the background RNA-binding of the DZF domains.

The initial analysis of the RBNS data performed by the Hogg lab formed the basis for further *in vitro* RNA-binding and structural studies using motifs from RBNS. Consequently, presented ZFR RBNS data refers to long-ZFR/NF45.

Results have shown that PTBP1 preferred binding to relatively unstructured RNAs, with a propensity for ssRNA in the motif region (figure 4.14C), matching previous findings using RBNS (N. Lambert et al., 2014). Since the known sequence preference of PTBP1 was recovered in this experiment, it was concluded that the RBNS protocol was successful. Long-ZFR/NF45 motifs were examined in the form of 6-mers or 8-mers, indicating a preference for C-rich sequences. Those 6-mer and 8-mer ZFR motifs that were highly enriched at various concentrations of the protein (figure 4.14A+B) seemed much more likely

to be structured than the surrounding sequence, explaining ZFR's preference for dsRNA over ssRNA. However, some ZFR-enriched 8-mer motifs are more structured (UGACACCC) than others (CCCCCCCC). These findings were not reproducible with the 6-mer RNA oligo. The sequence logo of ZFR RBNS indicates an RNA motif mainly consisting of cytosines (figure 4.14D). Among the top hits of recovered 8-mers were sequences such as UGACACCC, CCCCCCCC, or CUCCCUCC. To further investigate potential sequence-dependent RNA-binding of ZFR/NF45, EMSAs were performed using selected C-rich versus U-rich sequences.

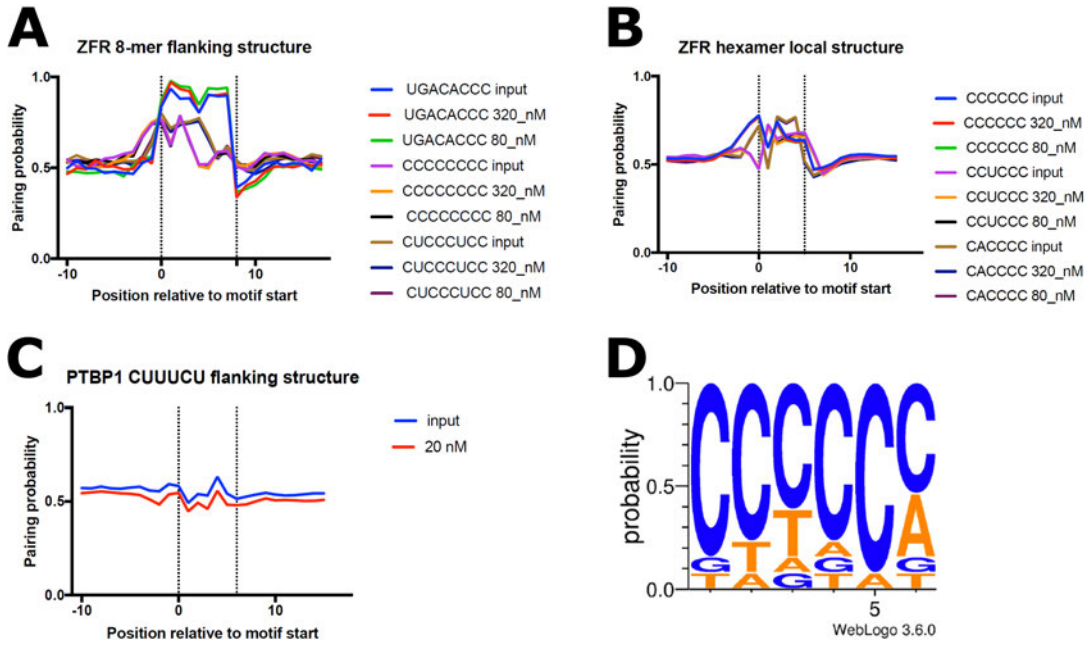


Figure 4.14: Predicted pairing probabilities for the most highly enriched 6- and 8-mers for long-ZFR/NF45. Potential for pairing to sequences found in all RNAs for ZFR-enriched sequences. A) 8-mer base-pairing probability of long-ZFR/NF45. B) 6-mer base-pairing probability of long-ZFR/NF45. C) 6-mer base-pairing probability of PTBP1 for CUUUCU. D) Sequence logo of 6-mer RBNS motifs for long-ZFR/NF45.

The RNA binding motif enrichment analysis suggested a propensity for ZFR to bind structured RNAs with a C-rich context because of the corresponding sequence logo (figure 4.14D). Based on these data, one C-rich (UGACACCC) motif from the top and one U-rich (UGUUGUAA) motif from the bottom of the list of recovered sequences were constructed as 12-mer and 24-mer fluorescently-labelled RNA oligo to verify this sequence-specificity by EMSA (figure 4.15). Two additional adenosines were added on both 5' and 3' ends (C-rich: "AAUGACACCCAA" and U-rich: "AAUGUUGUAAAA") for the 12-mer as shorter RNA

oligos could not be synthesised. For the 24-mer the binding motif was triplicated (C-rich: “UGACACCCUGACACCCUGACACCC” and U-rich: “UGUUGUAAUGUUGUAAUGUUGUAA”). During transcription, RNA can anneal to its DNA template to generate an RNA:DNA duplex and a displaced DNA strand, termed R-loop. R-loop binding could have an important implication for ZFR’s role in alternative splicing regulation. In addition to dsRNA, RNA:DNA duplexes were used for the assessment of RNA binding activity.

In vitro verification of RBNS-obtained ZFR motifs indicated no sequence-specific RNA binding. This means that ZFR shows no clear preference for either C- or U-rich dsRNA 12-mers or 24-mers (figure 4.15A+B). The same is true for RNA:DNA duplexes (figure 4.15C). However, ZFR strongly prefers dsRNA over ssRNA based on RBNS and EMSAs, suggesting that *in vitro* binding sites that determine alternative splicing sites likely contain substantial secondary structure.

Closer examination of the data showed that the most highly enriched sequences for ZFR were complementary to a sequence in the 3’ PCR primer binding site next to the 40 nt random sequence. C-rich sequences are enriched in the RBNS results because they allow secondary structure formation with the linker sequences. The preference of long-ZFR/NF45 for dsRNA caused it to pull-down sequences with base-pairing resulting from PCR primer binding sites.

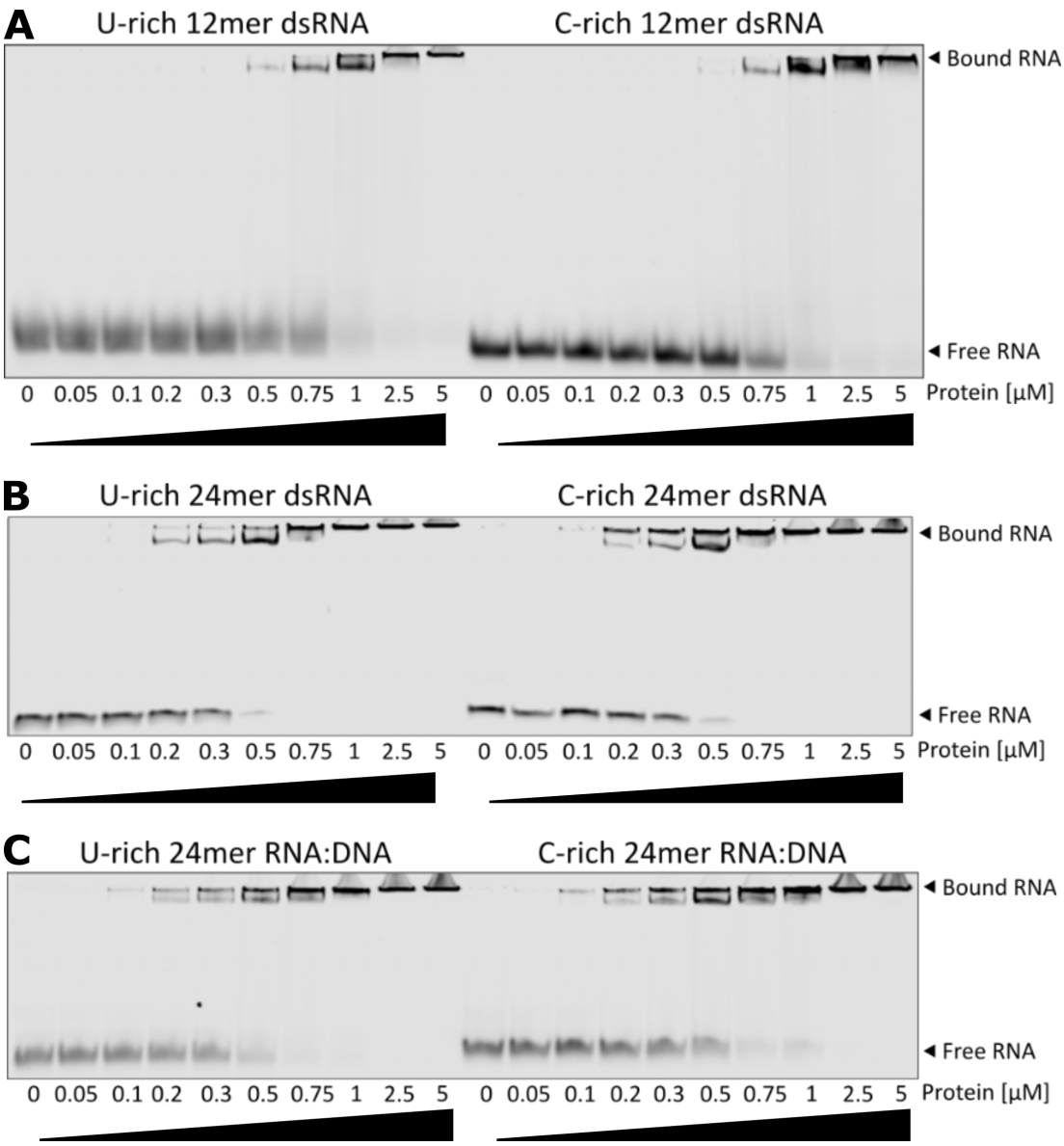


Figure 4.15: Assessing sequence specific RNA binding of ZFR/NF45. Fluorescent scans of EMSA experiments to compare RNA binding activity of ZFR/NF45 to various RNA probes differing in length and sequence context. A) EMSA with U-rich (left) and C-rich (right) dsRNA 12-mer probes. B) U-rich (left) and C-rich (right) dsRNA 24-mer probes. C) U-rich (left) and C-rich (right) 24-mer RNA:DNA duplexes probes.

After testing sequence-specific RNA-binding between C- and U-rich dsRNA as well as RNA:DNA hybrids, RNA-binding of these motifs was compared independent of their sequence context or length (figure 4.16) to see if ZFR prefers dsRNA over RNA:DNA duplexes. Both EMSAs demonstrated that ZFR/NF45 showed no preference for dsRNA over RNA:DNA duplexes and bound both without sequence specificity (figure 4.16A+B). This suggests that ZFR/NF45

prefers A-form nucleic acid duplexes independent of their sequence context.

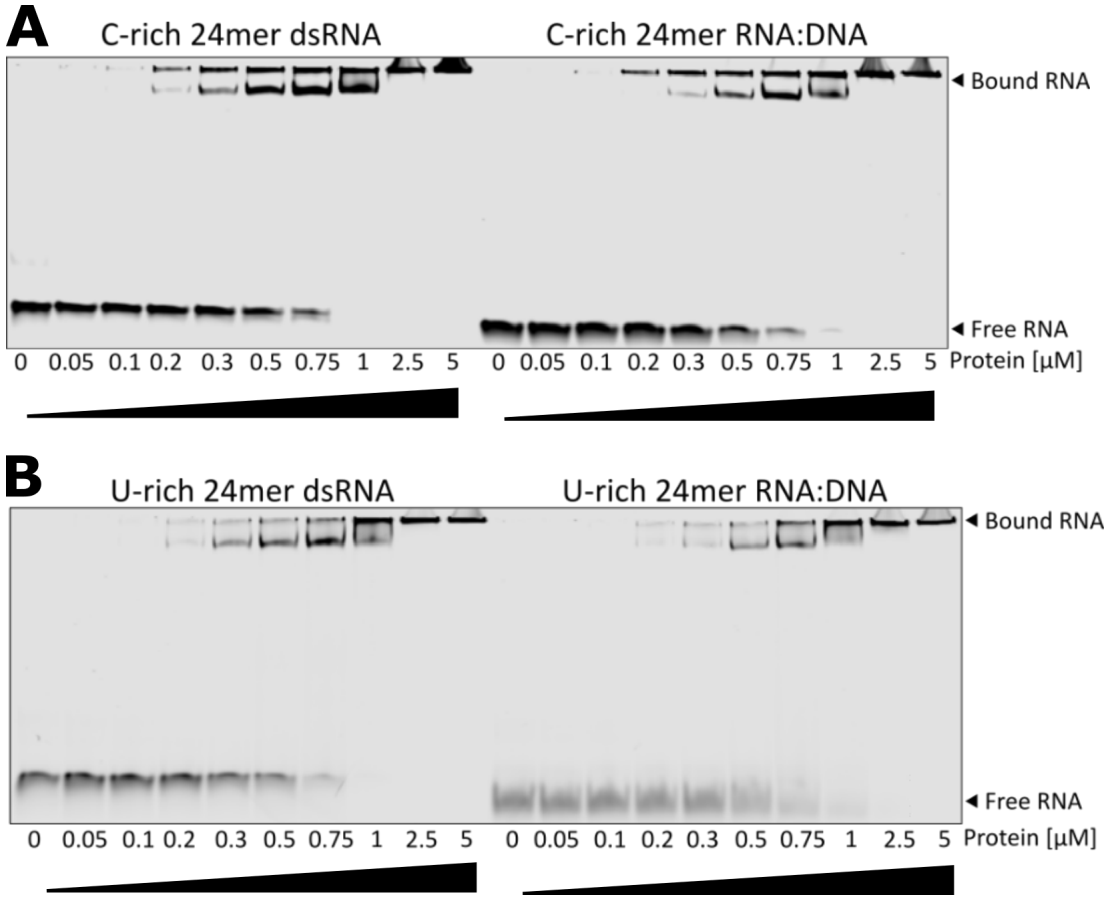


Figure 4.16: Assessing RNA binding of ZFR/NF45 for dsRNA and RNA:DNA duplexes. Fluorescent scans of EMSA experiments to compare RNA binding activity of ZFR/NF45 to various RNA probes differing in length and sequence context. A) C-rich 24-mer dsRNA (left) and C-rich 24-mer RNA:DNA duplex probes (right). B) U-rich 24-mer dsRNA (left) and U-rich 24-mer RNA:DNA duplex probes (right).

To gain reliable quantitation of binding constants, a fluorescence anisotropy assay was used to determine ZFR's binding rate to different types of nucleic acids (ssRNA, dsRNA, and RNA:DNA duplexes). The binding of a small ligand (RNA oligo) to a much bigger molecule (protein) that moves much slower than the free ligand in a solution can cause changes in anisotropy due to mobility changes of the fluorophore. This assay measures the binding of ligands to a protein of interest when a fluorophore is attached to the ligand and is a more quantitative approach compared to EMSAs. It allows the comparison of binding affinities for ssRNA, dsRNA and RNA:DNA duplexes. ZFR/NF45 was incubated at concentrations between 0 and 40 µM with three different Cy3-labelled RNA probes at a constant concentration of 20 nM (figure 4.17). A

significant change in anisotropy values was observed for all three RNA probes. The steep increase in anisotropy at low protein concentrations suggests a high binding affinity to the ligand.

Contrary to a fluorescence anisotropy experiment with NF90/NF45 and dsRNA (Jayachandran et al., 2015), the anisotropy levels for dsRNA and RNA:DNA ligands do not reach a maximum plateau at high concentrations of proteins (figure 4.17A). This behaviour might be associated with non-specific interactions between protein and RNA or even potential protein aggregation. Furthermore, the anisotropy curve of ssRNA at higher protein concentrations looked the same as for dsRNA and RNA:DNA duplexes. Again, this observation contrasts with previous EMSAs and might be attributed to non-specific interactions between protein and RNA.

However, some differences in the anisotropy levels between dsRNA and ssRNA binding affinity were observed for very low concentrations of ZFR/NF45 (under 0.02 µM) (figure 4.17B). Since the concentrations are very low and the changes not significant for ssRNA and RNA:DNA duplexes, it is difficult to reach a quantitative conclusion about ZFR's RNA binding affinities using a fluorescence anisotropy assay.

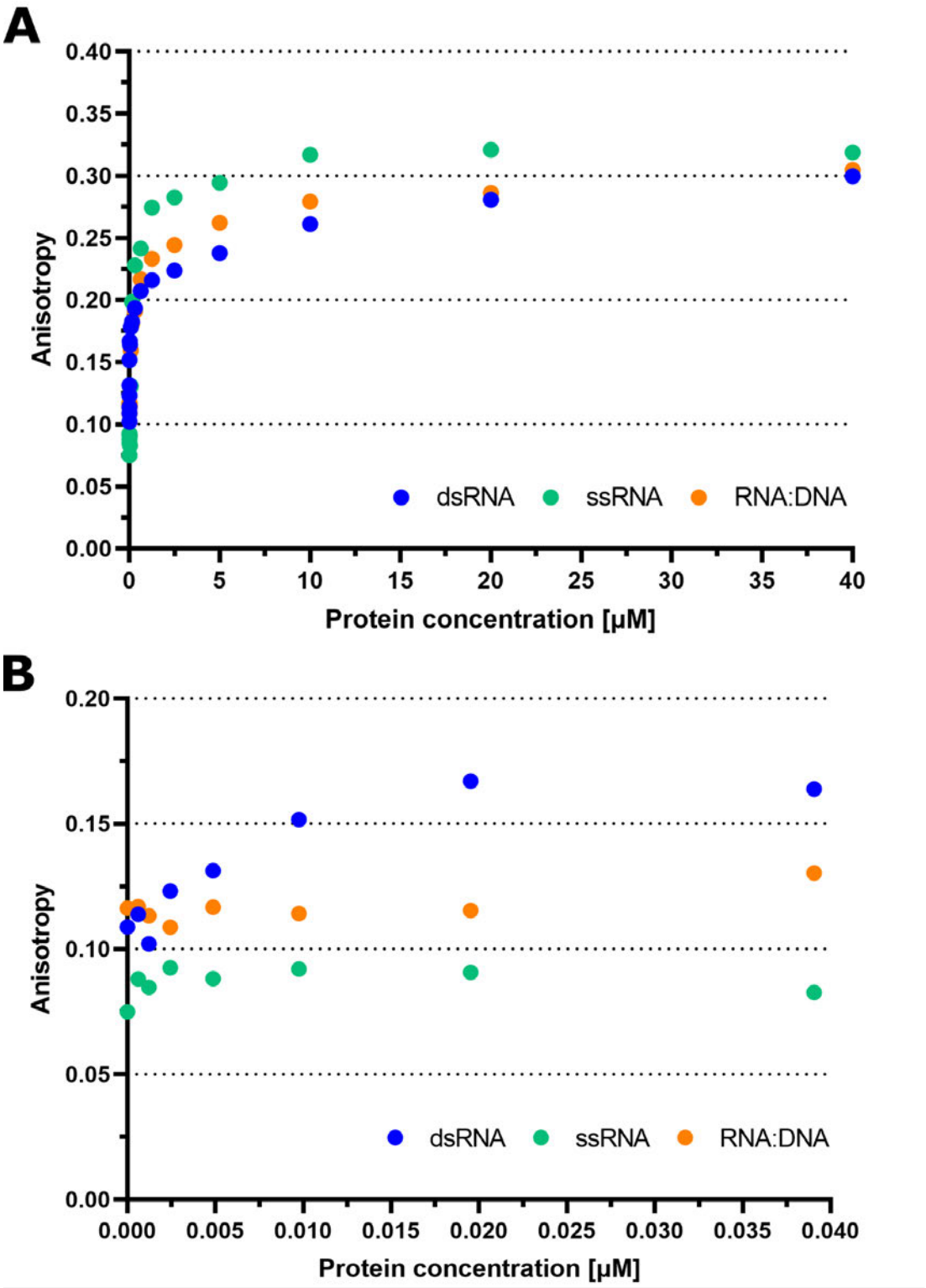


Figure 4.17: Fluorescence anisotropy assays of ZFR/NF45. A) Binding of ZFR(317-1074)/NF45(29-390) to various fluorescently-labelled nucleic acid probes: ssRNA, dsRNA, and RNA:DNA was measured. B) As in A) with a cropped concentration range of ZFR/NF45 to 40 nM.

5 Structural characterisation of RNA-protein complexes

5.1 Protein and RNA interaction studies *in vitro*

After successful optimisation of protein expression and purification conditions of *Drosophila* ADAR and Zn72D/NF45 as well as human ZFR/NF45, the structural characterisation of these proteins was initiated to understand the molecular mechanisms of RNA binding and potential protein-protein interactions in splicing and RNA editing regulation. Therefore, RNA-protein complexes were intended to be made *in vitro* using analytical gel filtration for analysis with low-resolution structural tools such as single-particle negative stain electron microscopy or size exclusion chromatography coupled small-angle X-ray scattering (SEC-SAXS). If these initial characterisations were successful, they could prove vital for obtaining a high-resolution structure by giving valuable information about stoichiometry, molecular weight and shape. Therefore, the purified protein complexes were incubated with a 24-mer C-rich dsRNA (same as used for EMSA experiments), and RNA-protein complex formation was assessed by analytical gel filtration. Finding conditions for *in vitro* formation of stable RNA-protein complexes is a prerequisite for structural analysis.

For ZFR/NF45, four samples were prepared and analysed by analytical gel filtration using a 24 ml S200 column (10 to 600 kDa molecular weight range) (figure 5.1) and SDS-PAGE (figure 5.2) to assess *in vitro* RNA-protein complexes. ZFR/NF45 on its own (Protein only) showed an elution peak (Peak 1 area) at around 11.5 ml and RNA on its own (RNA only) at 18.0 ml (Peak 2 area). If ZFR/NF45 was incubated in an equimolar ratio with RNA (1:1 protein:RNA), the first elution peak (Peak 1 area) shifted to 11.0 ml indicating a higher molecular weight. However, there was still free RNA, indicated by the additional peak (Peak 2 area) at around 17.7 ml. Since the free RNA was not entirely saturated with protein, the molar ratio of protein to RNA was increased to 2:1. The corresponding sample (2:1 protein:RNA) showed an even

earlier elution at around 10.5 ml (Peak 1 area) whereas the free RNA peak was absent. For the Coomassie-stained SDS-PAGEs (figure 5.2), the same fractions of each sample were run at the same position for comparison. The SDS-PAGE gels supported the idea that the proteins shift into higher molecular weight complexes upon RNA binding. To conclude, these interaction studies of ZFR/NF45 with dsRNA show that ZFR/NF45 forms a stable RNA-protein complex *in vitro* with a 2:1 stoichiometry of protein to RNA. This complex was used for further structural characterisation by single-particle negative stain electron microscopy (figure 5.5) and small-angle X-ray scattering (figure 5.6).

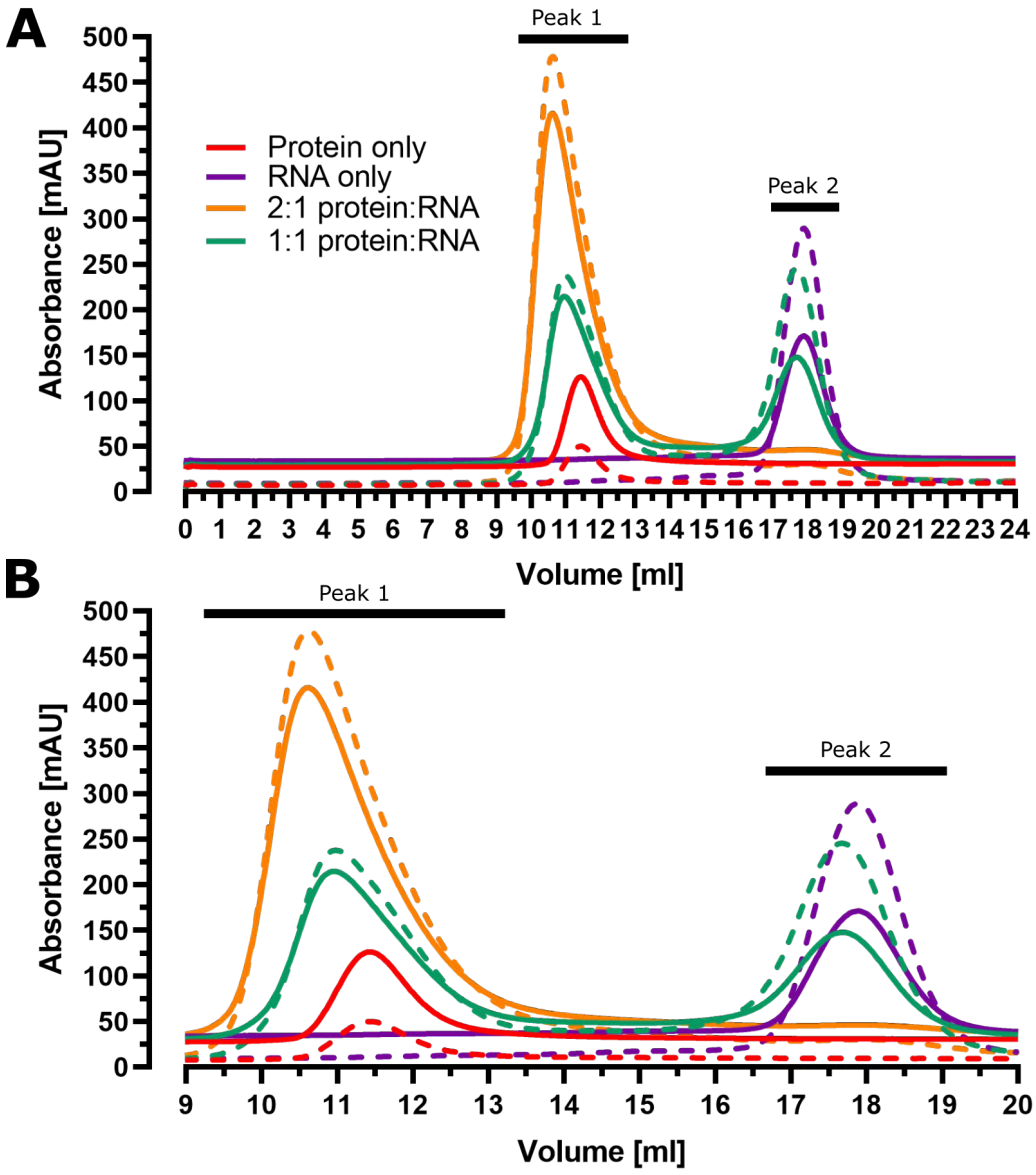


Figure 5.1: Analytical SEC experiments with ZFR/NF45 and RNA. Analytical SEC experiments were performed to demonstrate the interactions between ZFR/NF45 and a 24-mer dsRNA (C-rich) using a 24 ml S200 10/300 GL column. A) Elution volume from 0 to 24 ml. B) Elution volume from 9 to 20 ml. Solid lines indicate the 280 nm and dashed lines the 260 nm. Samples used for SEC are coloured as: ZFR/NF45 (protein only) in red, dsRNA (RNA only) in violet, ZFR/NF45 in a 1:1 molar ratio with dsRNA (1:1 protein:RNA) in green, and ZFR/NF45 in a 2:1 molar ratio with dsRNA (2:1 protein:RNA) in orange. The peaks indicated on the chromatograms (Peak 1 and Peak 2) were analysed by SDS-PAGE (figure 5.2).

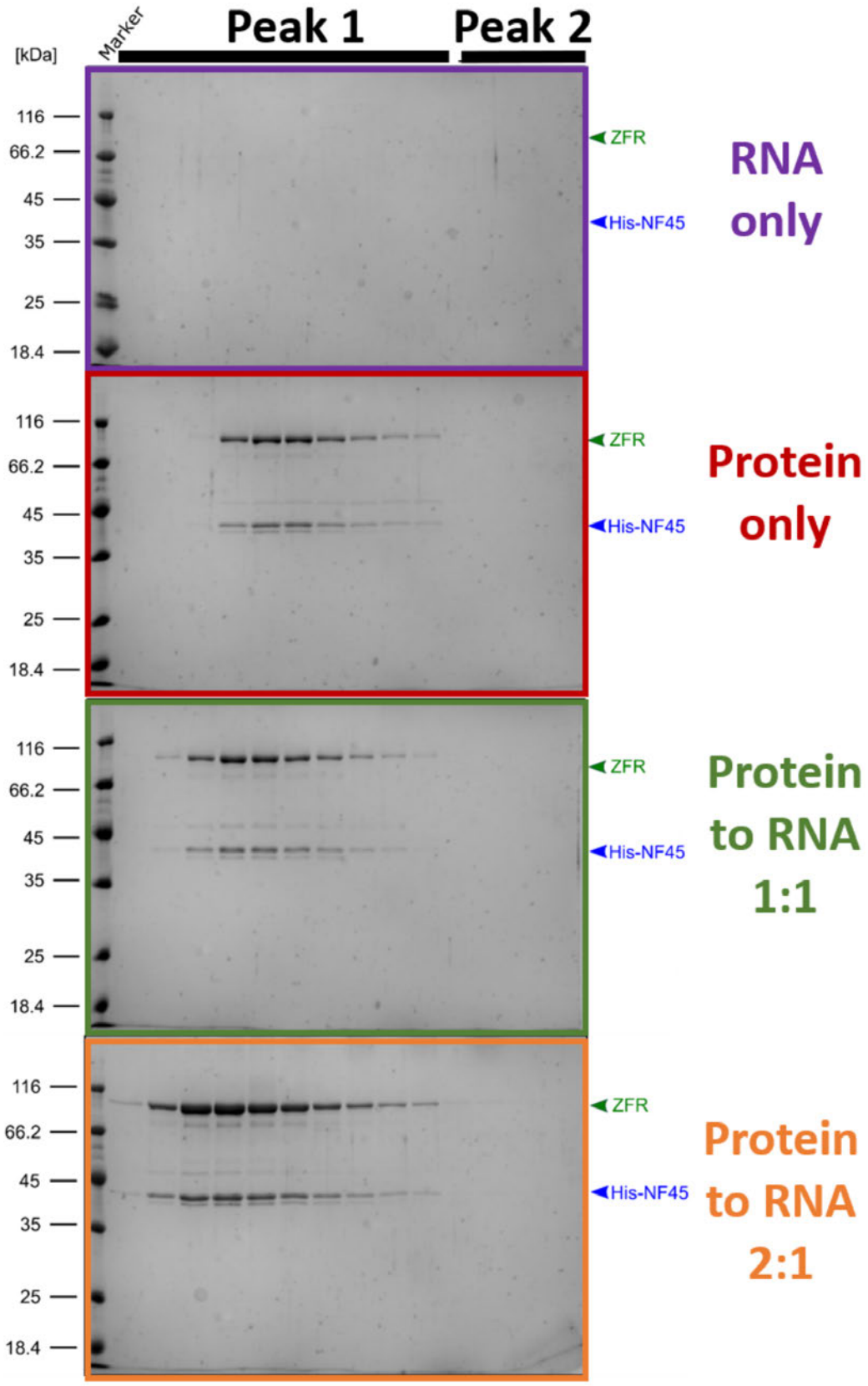


Figure 5.2: SDS-PAGE analysis of analytical SEC experiments with ZFR/NF45 and RNA. SDS-PAGE analysis of the indicated fractions in the analytical SEC chromatograms (figure 5.1).

Equivalent RNA-protein interaction studies were also executed with *Drosphila* ADAR and Zn72D/NF45. Four samples were prepared and analysed by analytical gel filtration using a 2.4 ml Superose 6 column (5 to 5000 kDa molecular weight range) (figure 5.3) and SDS-PAGE (figure 5.4). Zn72D/NF45 on its own (Zn72D only) showed an elution peak (Peak 1 area) at around 1.70 ml and RNA on its own (RNA only) at 1.85 ml (Peak 2 area). If Zn72D/NF45 was incubated in a 2:1 molar ratio with dsRNA, it showed an earlier elution at around 1.62 ml (Peak 1 area), whereas no more RNA was detected in peak area 2. The last sample was Zn72D/NF45 mixed with ADAR in equimolar ratios (2:2) and dsRNA in a 2:2:1 molar ratio. The corresponding sample eluted even earlier than the Zn72D/NF45 plus RNA sample at around 1.56 ml which can be interpreted as a complex with an even bigger molecular weight. This observation might be the first indication that ADAR, Zn72D/NF45, and RNA can form a complex *in vitro*. Due to sample limitations, whether the complex formation is RNA-dependent *in vitro* was unfortunately not tested. However, RNA-dependent interaction between ADAR and Zn72D/NF45 was shown in a pull-down experiment (Sapiro et al., 2020). Looking at the Coomassie-stained SDS-PAGE gels (figure 5.4), it was evident that the elution profile of Zn72D/NF45 changed upon incubation with RNA. If adding ADAR to Zn72D/NF45 and RNA, the three proteins seemed not only to co-elute on the SDS-PAGE but also to elute earlier than the previous RNA-protein complex (Zn72D/NF45 + RNA).

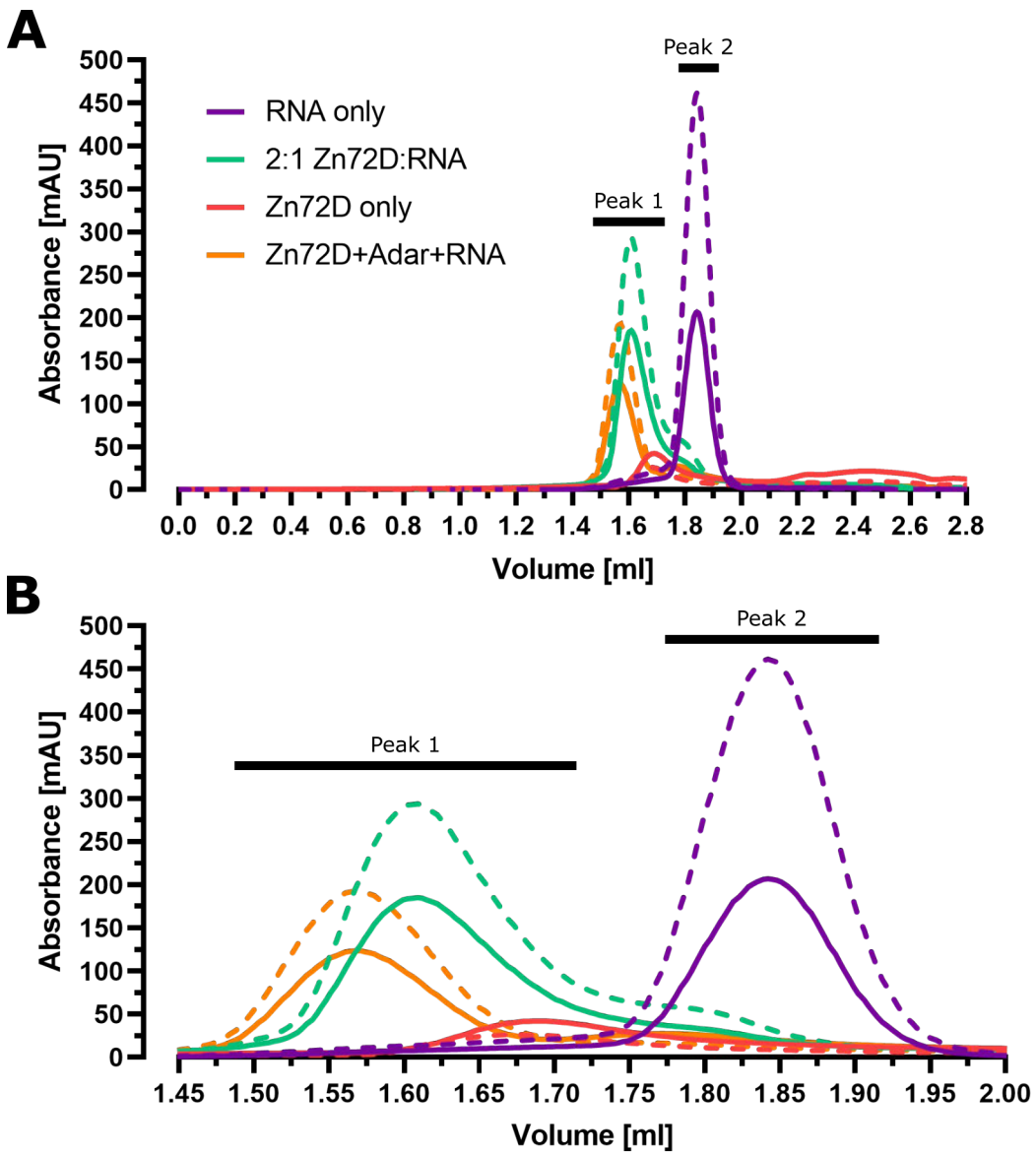


Figure 5.3: Analytical SEC experiments with Zn72D/NF45, ADAR, and RNA. Analytical SEC experiments were performed to demonstrate the interactions between Zn72D/NF45, ADAR and a 24-mer dsRNA (C-rich) using a 2.4 ml Superose 6 column. A) Elution volume from 0 to 2.8 ml. B) Elution volume from 1.45 to 2.00 ml. Solid lines indicate the 280 nm and dashed lines the 260 nm. Samples used for SEC are coloured as: Zn72D/NF45 (Zn72D only) in red, dsRNA (RNA only) in violet, Zn72D/NF45 in a 2:1 molar ratio with dsRNA (2:1 Zn72D:RNA) in green, and Zn72D/NF45 in a 2:2:1 molar ratio with ADAR and dsRNA (Zn72D+Adar+RNA) in orange. The peaks indicated on the chromatograms (Peak 1 and Peak 2) were analysed by SDS-PAGE (figure 5.4).

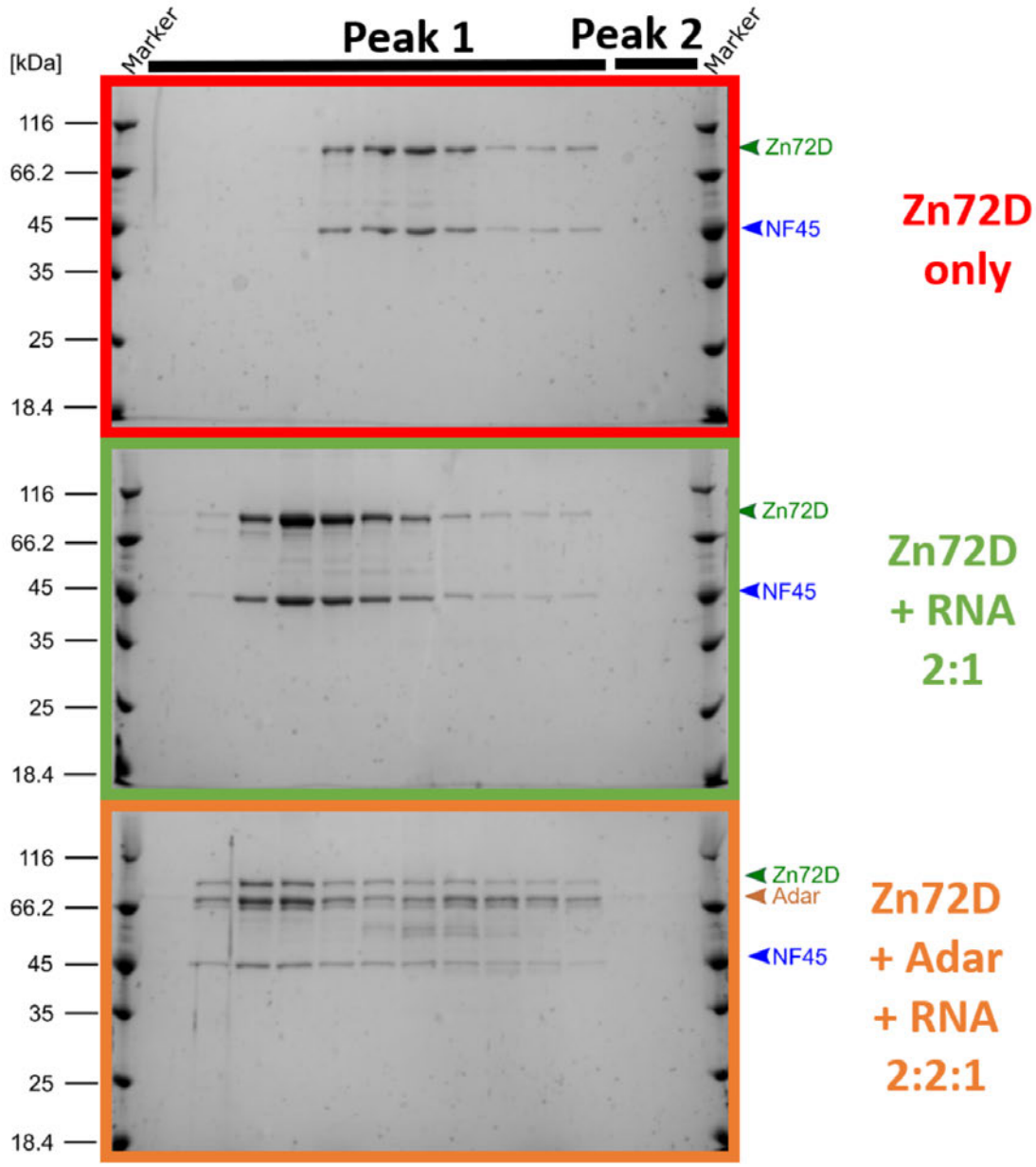


Figure 5.4: SDS-PAGE analysis of analytical SEC experiments with Zn72D/NF45, ADAR, and RNA. SDS-PAGE analysis of the indicated fractions in the analytical SEC chromatograms (figure 5.1).

5.2 Single particle analysis of RNA-protein complex of ZFR/NF45

After successful RNA-protein complex formation using ZFR/NF45 and a 24-mer dsRNA (figure 5.1), the middle fraction of the peak was used negative

stained for single particle analysis using electron microscopy. Two representative images are shown from a dataset (protein concentration 0.01 mg/ml) that was collected at 75000x magnification at the JEOL JEM 1400 Plus TEM. These images showed very little to no aggregation, and particles are separated (figure 5.5). These images indicate that the ZFR/NF45 RNA-protein complex formed *in vitro* is a good candidate for further structural characterisation by single-particle analysis. Ultimately, this might help solve a high-resolution structure of the complex using cryo-EM.

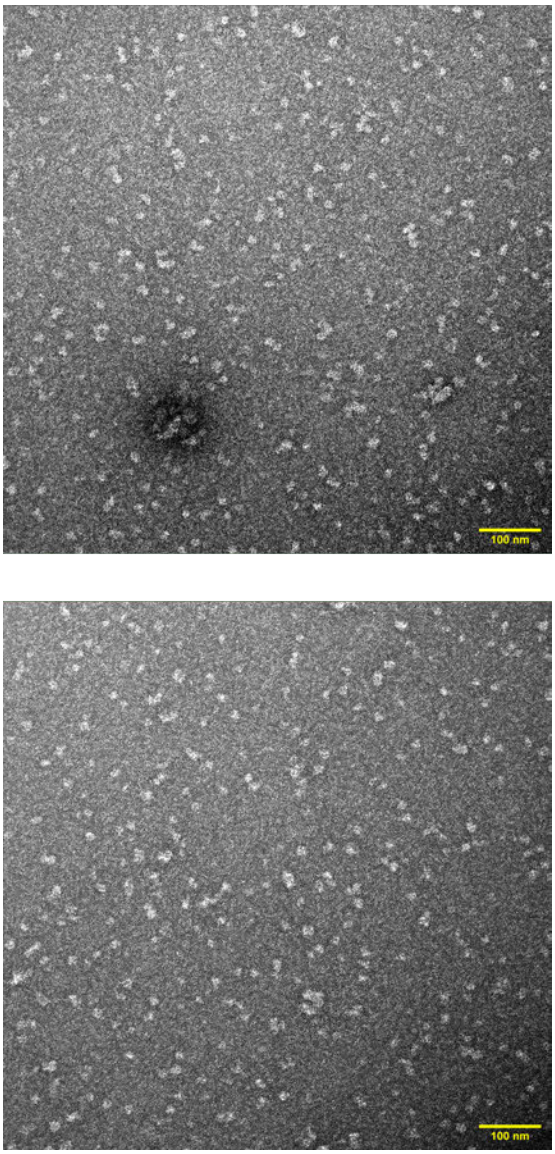


Figure 5.5: Negative stain electron microscopy images of a 2 ZFR/NF45:1 dsRNA complex. Two representative images of negative staining data of ZFR/NF45 RNA-protein complex at 75 000x magnification.

5.3 SEC-SAXS experiments and analysis of RNA-protein complexes

SEC-SAXS experiments were carried out to characterise the mass and shape of the proteins and corresponding RNA-protein complexes in solution. Therefore, two samples for each protein (ADAR, Zn72D/NF45 and ZFR/NF45) were prepared: one with and one without dsRNA (24-mer C-rich). The SAXS samples were all applied to a S200 2.4 mL column with a protein concentration of 5 mg/ml concentration at the B21 beamline at Diamond Light Source (Harwell). The samples containing RNA were set up in 2:1 molar ratios of protein to RNA.

5.3.1 ZFR/NF45 RNA-protein complex

The SEC profile of ZFR/NF45 (figure 5.6) indicated that the two samples did not elute as symmetrical peaks as both had a trailing edge. However, the sample of ZFR/NF45 without RNA did not differ from ZFR/NF45 with RNA in its molecular weight across the elution profile, indicating potential protein-protein interactions in the absence of RNA. The curves showed the presence of higher-molecular weight aggregates and therefore were found unsuitable for further analysis. The two SEC-SAXS samples might have been switched, the RNA might have degraded, or the binding would have been affected by the experiment at room temperature.

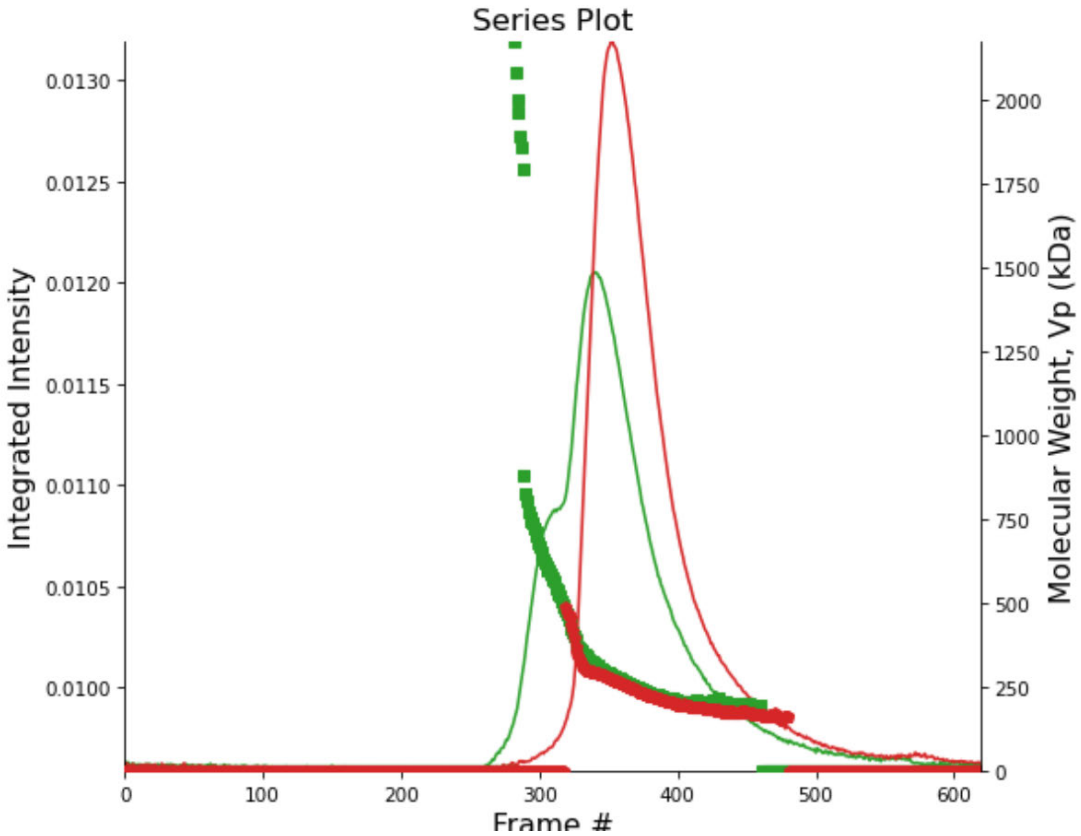


Figure 5.6: Size exclusion profile of ZFR/NF45 from SEC-SAXS experiments. SEC trace for ZFR/NF45 with RNA (red) and without (green) eluted from a S200 Increase 3.2 (2.4 mL) column. The buffer was subtracted for both curves. The left Y-axis shows the molecular weight in kDa.

5.3.2 Zn72D/NF45 RNA-protein complex

The SEC profiles of Zn72D/NF45 (figure 5.7) indicated that the two samples did not elute as symmetrical peaks, with both having a trailing edge. In contrast to the ZFR/NF45 samples, Zn72D/NF45 showed different molecular weights across the peaks for samples with and without RNA (figure 5.7). The sample with RNA showed the presence of higher-molecular weight aggregates, which made selecting a suitable area of the peak difficult. The radius of gyration (Rg) was estimated across the peaks and only frames from regions with a stable radius of gyration selected to average for further analysis (table 5.1).

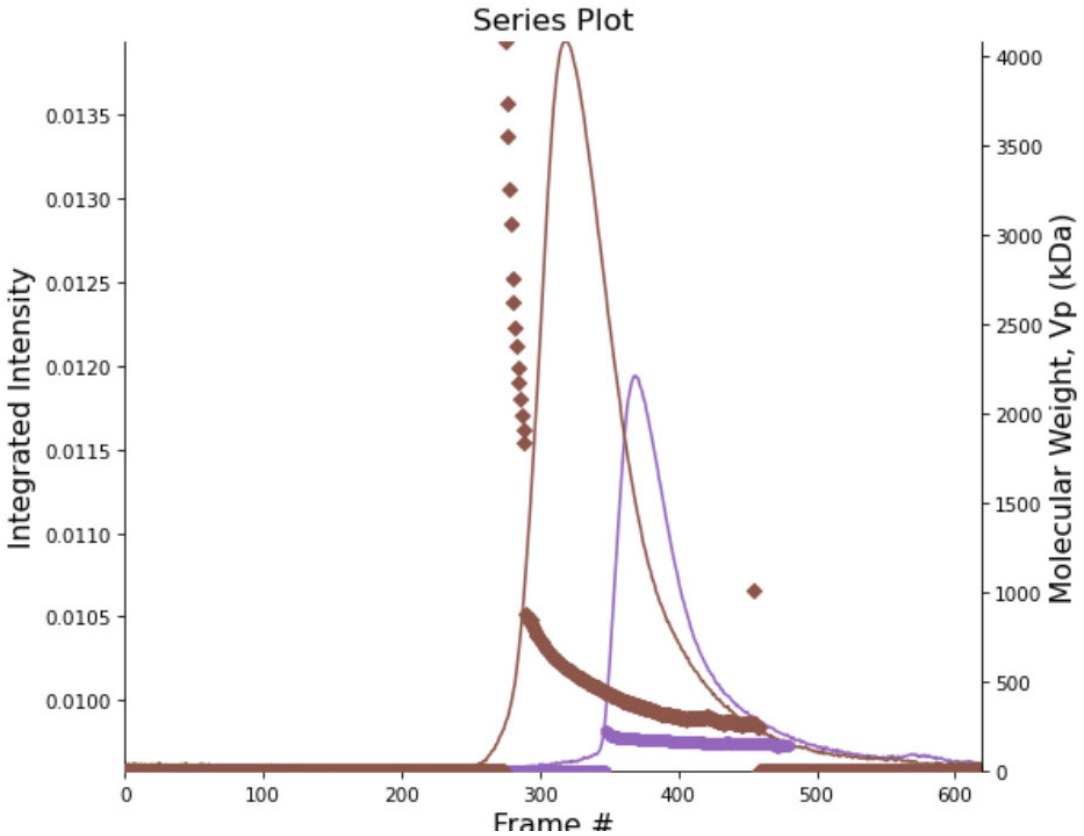


Figure 5.7: Size exclusion profile of Zn72D/NF45 from SEC-SAXS experiments. SEC trace for Zn72D/NF45 with RNA (brown) and without (violet) eluted from a S200 Increase 3.2 (2.4 mL) column. The buffer was subtracted for both curves. The left Y-axis shows the molecular weight in kDa.

A Guinier analysis was carried out on the averaged scattering curves to determine the quality of the data. The Guinier plots of both Zn72D/NF45 samples were linear (figure 5.8 and 5.9), showing the absence of aggregation and the estimated Rg for Zn72D/NF45 without RNA was 46.4Å and 59.6Å with RNA.

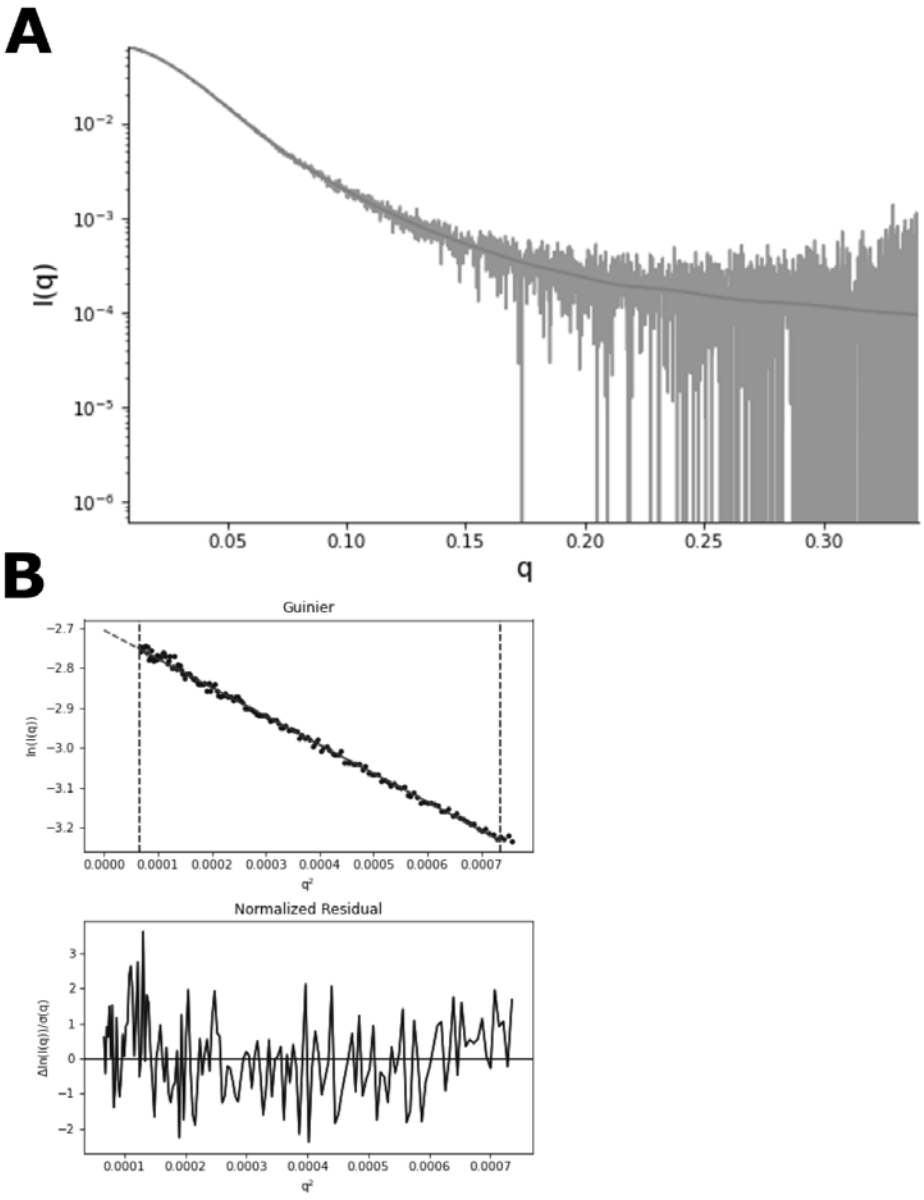


Figure 5.8: Scattering plot and Guinier analysis of Zn72D/NF45 without RNA from SEC-SAXS experiments. A) Scattering plot and B) Guinier analysis of Zn72/NF45 without RNA.

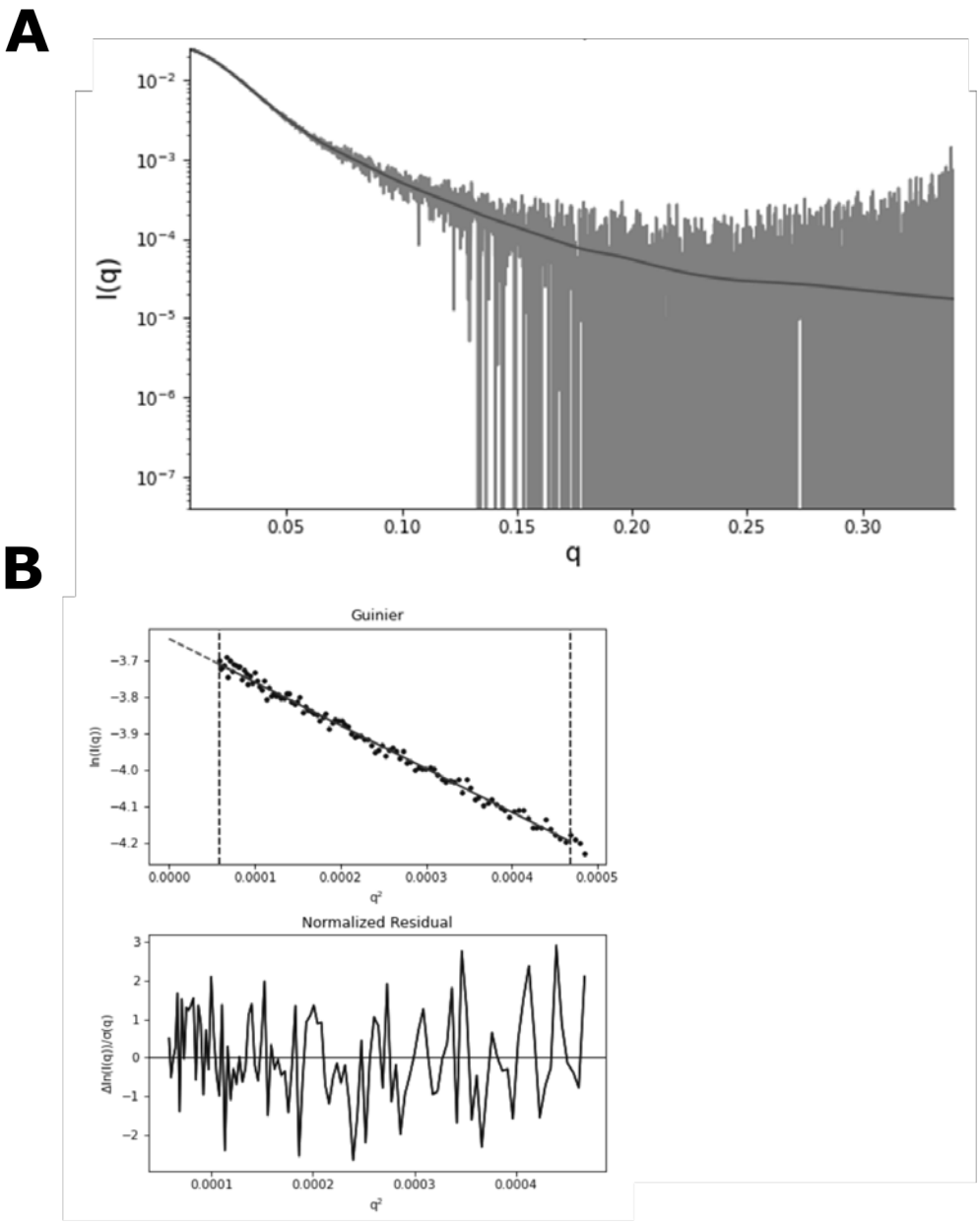


Figure 5.9: Scattering plot and Guinier analysis of Zn72D/NF45 with RNA from SEC-SAXS experiments. A) Scattering plot and B) Guinier analysis of Zn72/NF45 with RNA.

The pairwise distribution functions (P(r)) were calculated from the scattering curves were used to estimate the maximum diameter (Dmax) of the two samples and to gain an understanding of the overall shape of the proteins in solution (5.10). The Dmax for Zn72D/NF45 without RNA was 186Å and 290Å with RNA (table 5.1). The two P(r) functions have a single peak with a trailing tail. Then a dimensionless Kratky plot was generated to estimate the degree

of flexibility shown by the proteins in the solution.

Finally, the molecular weights of the samples were estimated from the Porod volume: Zn72D/NF45 without RNA (179.2 kDa) and with RNA (298.6 kDa). This indicated an almost two-fold molecular weight difference for Zn72D/NF45 with RNA that seemed to replicate the previous gel filtration results (figure 5.3). Also, the calculated molecular mass of a monomer (1 Zn72D/NF45) (123.2 kDa) and dimer with RNA (2 Zn72D/NF45:1 dsRNA) (264.4 kDa) are close to the expected sizes but slightly higher than expected. This might be due to the dynamic range of Rg across the peaks from SEC profiles.

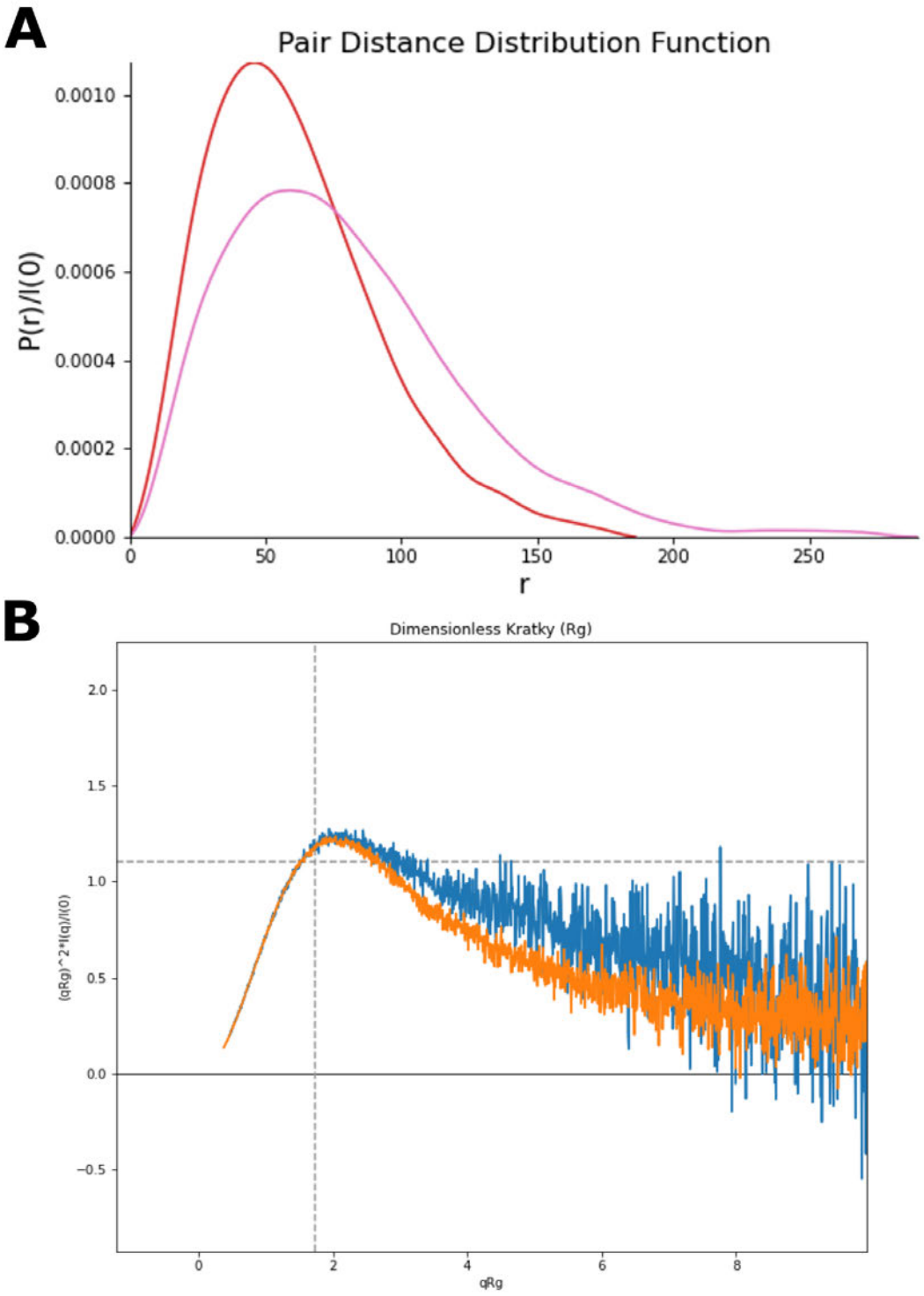


Figure 5.10: P(r) function and dimensionless Krakty plot of Zn72D/NF45 SEC-SAXS experiments. A) P(r) function of Zn72D/NF45 with (pink) and without RNA (red). B) Dimensionless Krakty plot of Zn72D/NF45 with (blue) and without RNA (orange).

5.3.3 ADAR RNA-protein complex

The SEC profiles of ADAR (figure 5.11) indicated that the two samples did not migrate as symmetrical peaks, with both having a trailing edge. In contrast to the ZFR/NF45 samples, ADAR had an extreme gradient of Rg values across the peaks for samples with and without RNA. The sample with RNA showed presence of higher-molecular weight aggregates, which made the selection of a suitable area of the peak difficult.

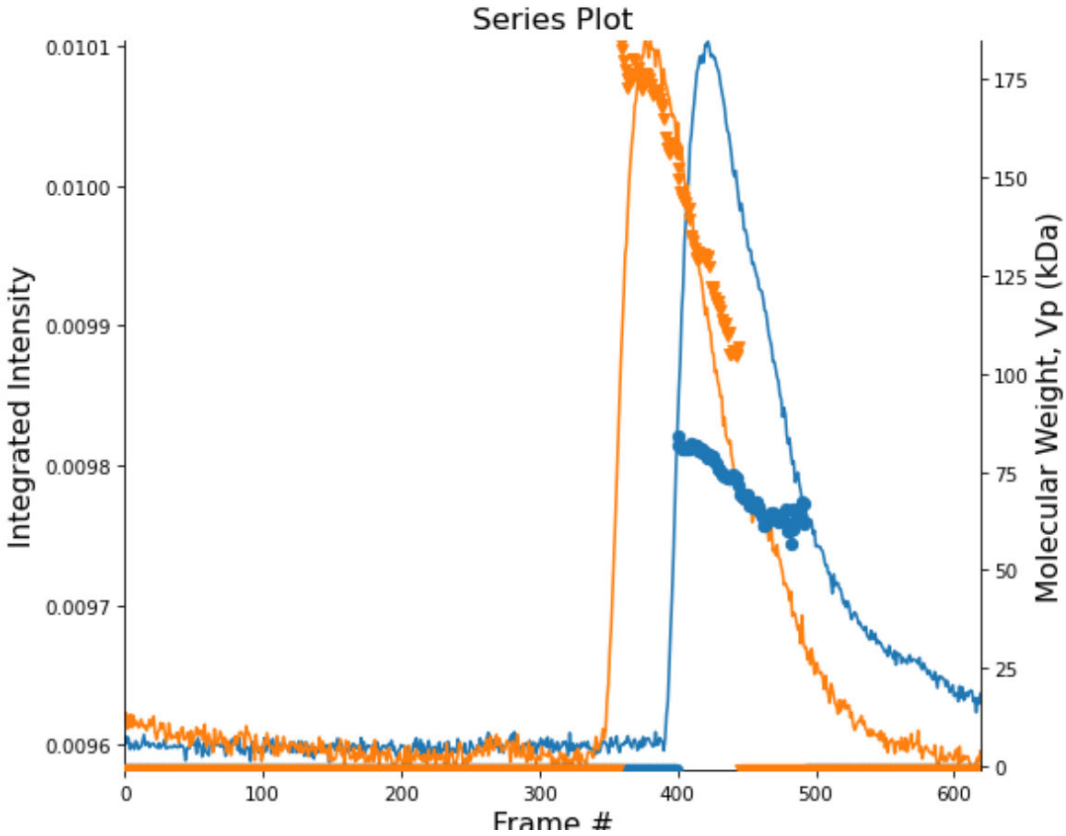


Figure 5.11: Size exclusion profile of ADAR from SEC-SAXS experiments. SEC trace for ADAR with RNA (orange) and without (blue) eluted from a S200 Increase 3.2 (2.4 mL) column. The buffer was subtracted for both curves. The left Y-axis shows the molecular weight in kDa.

A guinier analysis was carried out on the averaged scattering curves to determine the quality of the data. The Guinier plots of both ADAR samples were linear (figure 5.12 and 5.13), showing the absence of aggregation and the estimated Rg for ADAR without RNA was 37.63Å and 48.75Å with RNA (table 5.1), indicating an increase of Rg upon RNA binding.

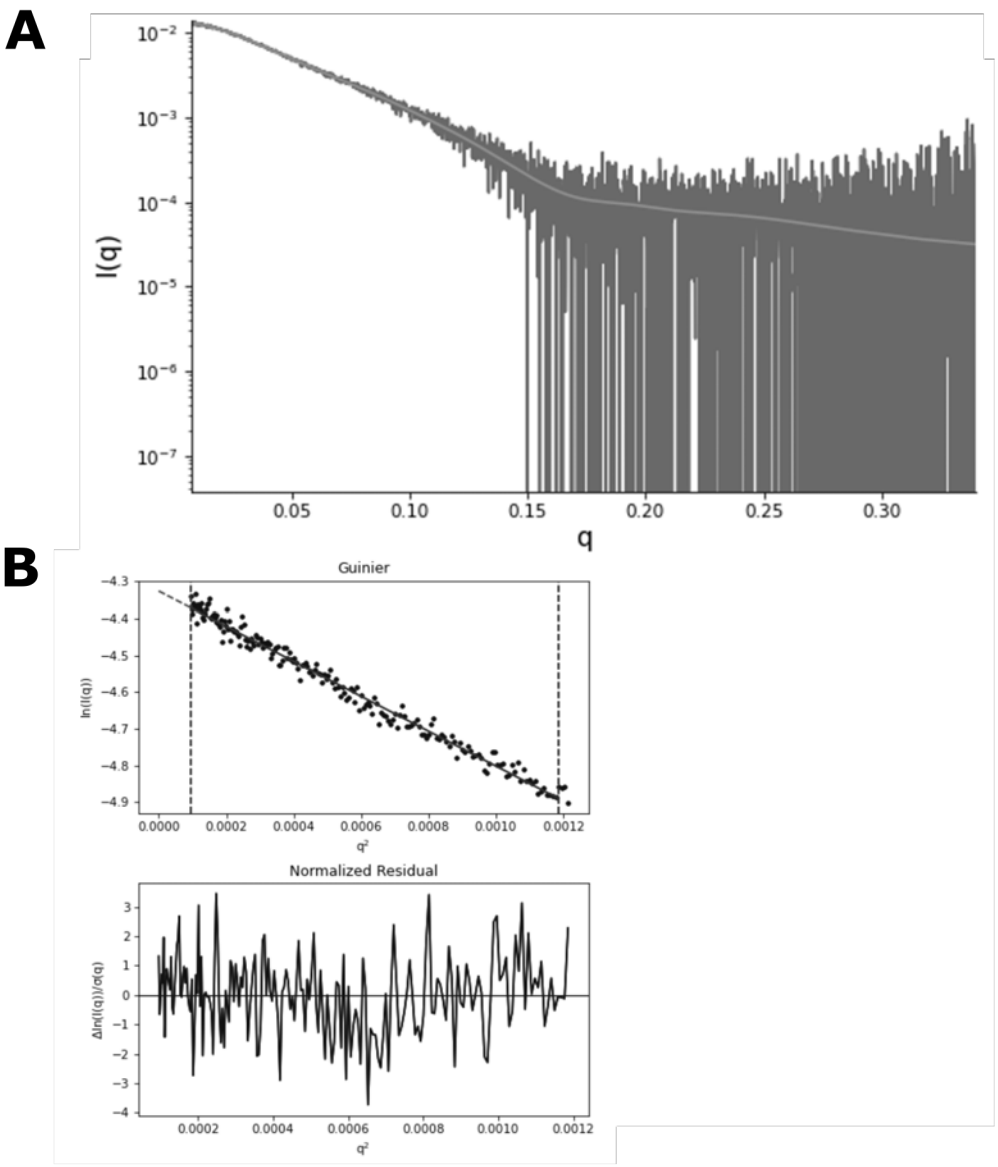


Figure 5.12: Scattering plot and Guinier analysis of ADAR without RNA from SEC-SAXS experiments. A) Scattering plot and B) Guinier analysis of ADAR without RNA.

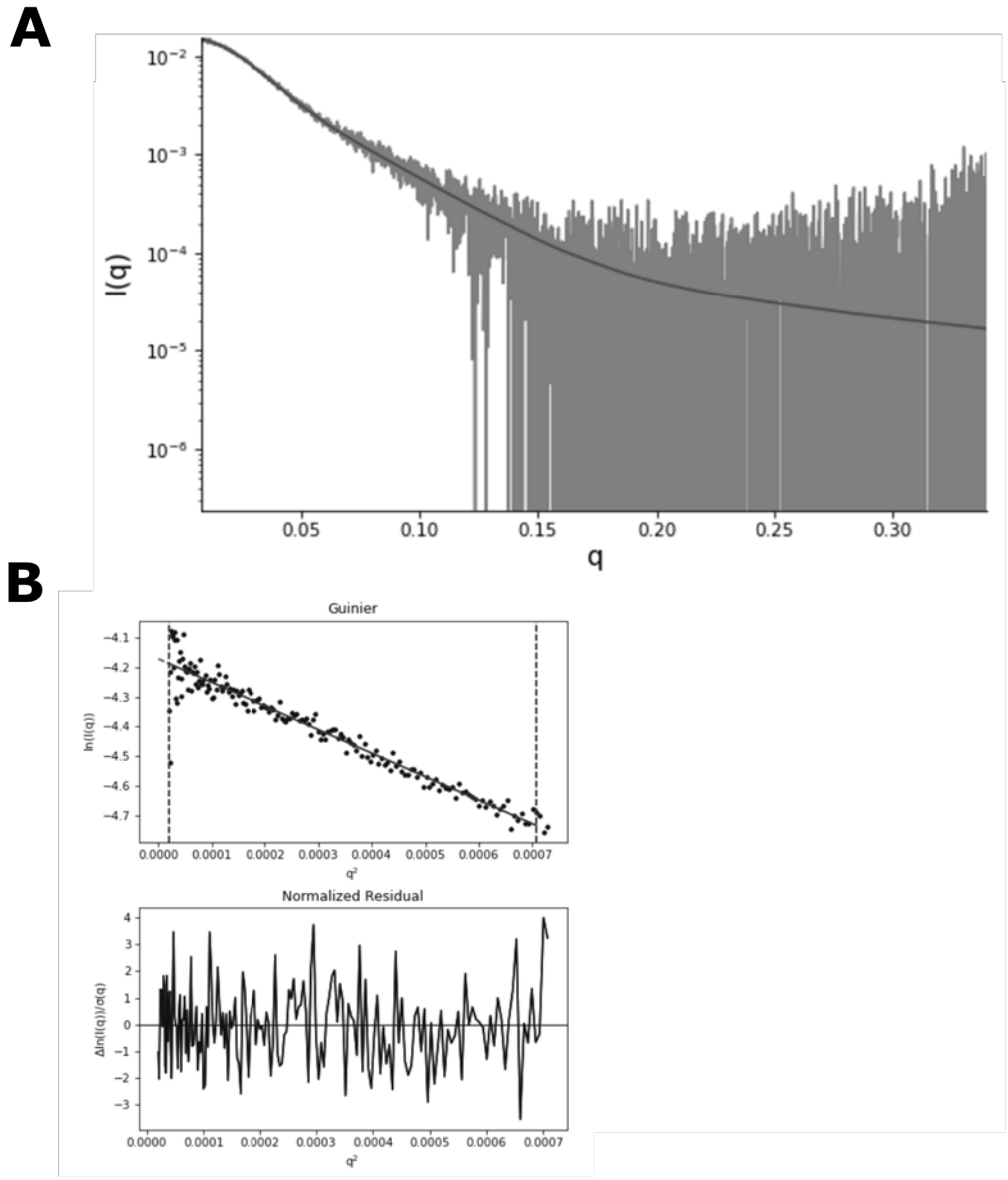


Figure 5.13: Scattering plot and Guinier analysis of ADAR with RNA from SEC-SAXS experiments. A) Scattering plot and B) Guinier analysis of ADAR with RNA.

The pairwise distribution functions (P(r)) was also calculated for ADAR (figure 5.14). The Dmax for ADAR without RNA was 185Å and 210Å with RNA (table 5.1). The two P(r) functions have a single peak with a trailing tail. Then a dimensionless Kratky plot was generated to estimate the degree of flexibility shown by the proteins in the solution.

The molecular weight of the samples from the Porod volume was calculated for ADAR with RNA (79.1 kDa) and without RNA (172.3 kDa). This indicated a

two-fold molecular weight difference for ADAR with RNA that seemed to replicate the previous gel filtration results (figure 5.3). Also, the calculated molecular mass of a monomer (70.3 kDa) and dimer with RNA (148.6 kDa) were close to the expected sizes but slightly higher than expected. This might be due to the dynamic range of Rg across the peaks from SEC profiles.

The SEX-SAXS confirmed the previous gel filtration experiments and showed that the calculated molecular wight more than doubled upon RNA binding, which has been shown previously for solving the structure of human ADAR2 with RNA (Matthews et al., 2016a). Further, this is indicates that these RNA-protein complexes could be used for further SAXS experiments.

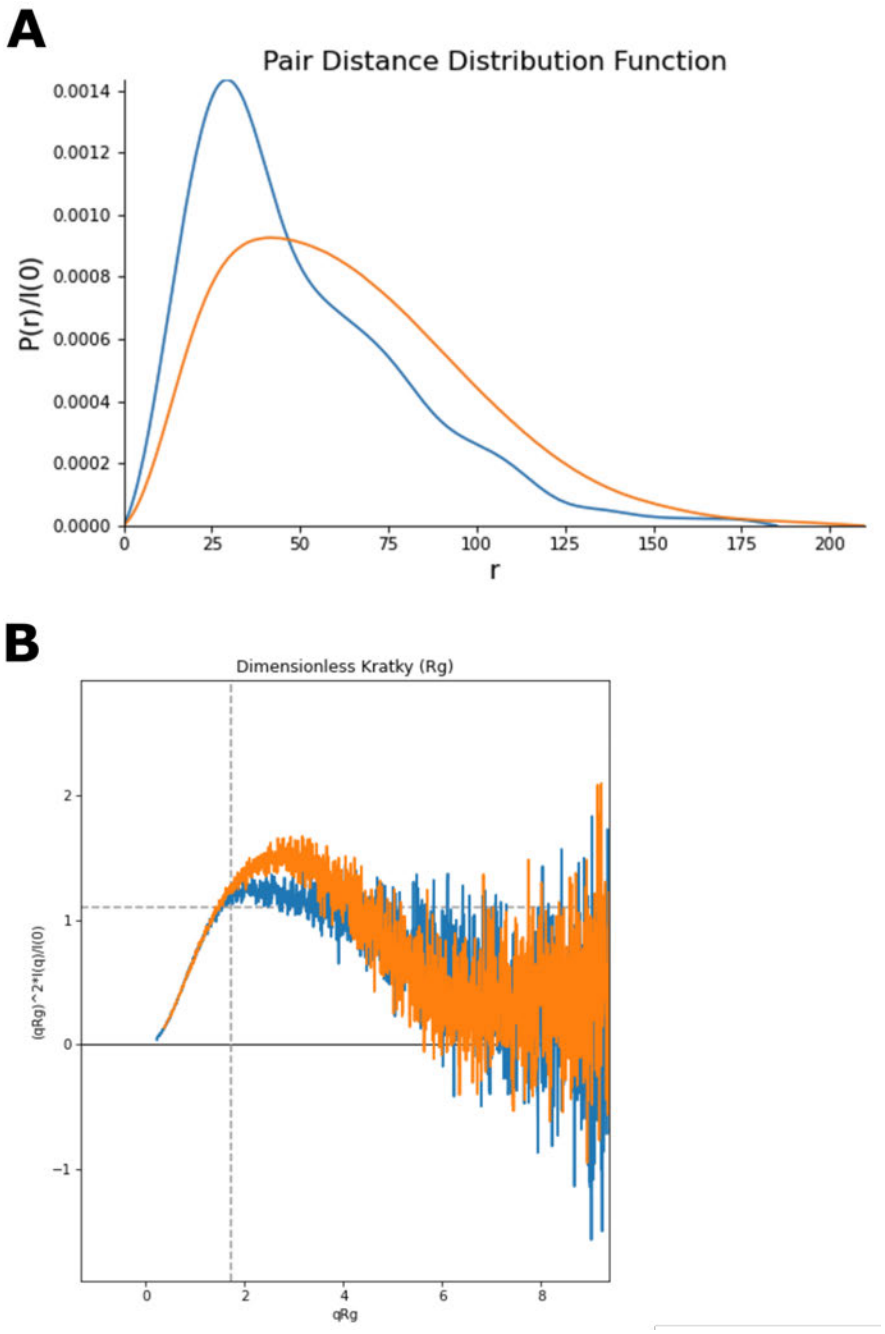


Figure 5.14: P(r) function and dimensionless Krakty plot of ADAR SEC-SAXS experiments. A) P(r) function of ADAR with (orange) and without RNA (blue). B) Dimensionless Krakty plot of ADAR with (blue) and without RNA (orange).

Table 5.1: Overview of analyses of SEC-SAXS data of ADAR and Zn72D/NF45

	ADAR no RNA	ADAR + RNA	Zn72D/NF45 no RNA	Zn72D/NF45 + RNA
Files	02750-02766	03316-03327	03947-03955	05831-05849
Frames	414-430	371-389	372-380	396-414
Guinier analysis				
n_min	20	19	28	24
n_max	189	169	173	131
q_min	0.01	0.01	0.01	0.01
q_max	0.03	0.03	0.03	0.02
q_min*Rg	0.47	0.34	0.38	0.46
q_max*Rg	1.30	1.30	1.26	1.29
I(0)	0.01 +/- 3.3e-05	0.02 +/- 4.3e-05	0.07 +/-8.15e-05	0.03 +/- 6.77e-05
Rg	37.63 +/- 0.15	48.7513 +/- 0.2091	46.414 +/- 0.0912	59.63 +/- 0.23
r^2 (fit)	0.98	0.98	0.99	0.99
P(r) analysis				
Dmax (GNOM)	185	210	186	290
Rg P(r) GNOM	41.6 +/- 0.49	51.8 +/- 0.33	48.26 +/- 0.17	59.63 +/- 0.23
I(0) GNOM	0.01 +/- 5.47E-05	0.02 +/- 5.18E-05	0.07 +/- 1.09E-04	0.03 +/- 6.7E-05
Chi^2	1.46	2.52	1.22	1.59
MW				
Vc MW	58.3	143.0	145.5	244.4
Vc	521.5	926.5	911.8	1.34E+03
Vp MW	79.1	172.3	179.2	298.6
Vp	1.64E+05	3.69E+05	3.77E+05	6.66E+05
Corrected Vp	9.53E+04	2.08E+05	2.16E+05	3.60E+05

6 Discussion

The first aim of this thesis was the investigation of A-to-I RNA editing regulation using proteins from *D. melanogaster*. The *Drosophila* RNA-binding protein, Zn72D, acts as a key regulator for many RNA editing events *in vivo* (Sapiro et al., 2020). Dissecting the mechanism of RNA editing on a molecular level will ultimately help us to understand why those particular sites require Zn72D for efficient editing. Our original hypothesis was that a specific subset of editing sites might require cooperative interaction between Zn72D and ADAR, the enzyme responsible for the adenosine-to-inosine conversion, for proper RNA binding and/or substrate recognition. One could imagine that Zn72D recruits ADAR to substrates that it would not otherwise easily recognise. An ambitious goal was to reconstitute an A-to-I RNA editing reaction *in vitro* using recombinant proteins and transcribed RNA. For this goal, both expression and purification conditions of ADAR and Zn72D/NF45 were optimised to produce soluble proteins. Since ADAR needs to be expressed in insect cells to maintain its full enzymatic function and due to the lack of previous data for expression and purification conditions of Zn72D and NF45, these optimisations did become a main focus of this thesis.

Reported *Drosophila* RNA editing sites were divided into Zn72D-dependent and Zn72D-independent sites using the data from an *in vivo* RNAi screen looking for RNA-binding proteins that regulate A-to-I RNA editing in *Drosophila* (Sapiro et al., 2020). Good candidates for each group were identified using several criteria such as the absence of surrounding editing sites, editing levels in the head, and secondary structure prediction. The best candidates were transcribed *in vitro* to determine if these RNAs could serve as substrates in editing assays. Once protein production was optimised, and RNA templates selected, all necessary components for an *in vitro* editing assay were available. As edited vs non-edited RNA products are nearly identical in sequence and size, it was a challenge to find and establish an *in vitro* method that enables the distinction between both forms needed to report on editing.

Most common RNA editing assays in the literature are cell-based and typic-

ally make use of a GFP reporter gene carrying a stop codon that needs to be edited for successful translation of the fluorescent protein (Heep et al., 2017; Hanswillemenke et al., 2015) or a GFP-RFP reporter construct separated by a stop codon that expresses one fluorescent protein when unedited and produces both when edited (Garncarz et al., 2013). Other RNA editing assays are based on reverse transcription of edited RNA followed by DNA sequencing (Palavicini et al., 2009) or poisoned primer extension using radio-labelled DNA primers (O'Connell et al., 1997). A fluorescence-based method incorporated fluorescent adenosine analog thieno[3,4-d]-6-aminopyrimidines into ADAR substrates and measured changes in fluorescence due to the deamination reaction (Mizrahi et al., 2015). Our initial idea of detecting A-to-I RNA editing focussed on poisoned primer extension. We initially intended to use a poisoned primer extension assay for the detection of adenosine to inosine conversion *in vitro* which seemed simple, did not rely on specific nucleotides, could be adapted for use with a fluorescent probe and was easily applicable to most RNA templates. However, this poisoned primer extension approach proved not ideal: we found that most reverse transcriptases do not faithfully extend their templates, and many include additional non-template nucleotides. These issues made identification and precise quantification of inosine incorporation unfeasible. We developed an alternative method for the detection of inosine incorporation using human endonuclease V, a highly active ribonuclease specific for ribonucleic acids that contain inosine, based on the findings of Vik et al., 2013. We were able to demonstrate that the adapted concept could be successfully used for our purposes. This method showed beneficial properties compared with poisoned primer extension. It is a direct method to detect inosines in RNA, which should not be present without A-to-I RNA editing, averting the possibility of false-positive or false-negative detection. Unfortunately, due to time restrictions we were not able to continue assembling the *in vitro* RNA editing reactions or to compare editing levels for Zn72D-dependent and Zn72D-independent RNA editing sites to draw mechanistic insights into Zn72D/NF45's role in regulating RNA editing.

One of the main achievements of the project was the reconstitution of protein-protein and RNA-protein complexes. Both ADAR and Zn72D/NF45 formed RNA-protein complexes with a 2:1 (protein to RNA in molar ratio) stoichiometry based on analytical gel filtration experiments. The dimerisation of ADAR upon RNA binding has been shown in the crystal structure of human ADAR2 bound

to dsRNA (Matthews et al., 2016b) and might also be an underlying mechanism for Zn72D-dependent editing regulation. ADAR and Zn72D/NF45 complexes with dsRNA are ideal candidates for further structural studies. The next critical steps for this project are the use of these RNA-protein complexes for further SEC-SAXS and single particle electron microscopy (EM) experiments to ultimately apply cryo-EM to get a high resolution structure. These data will give substantial insights into dsRNA substrate recognition by ADAR and Zn72D/NF45 and may reveal changes that occur when these complexes form. Interestingly, the previous analytical gel filtration experiments also indicated that ADAR and Zn72D/NF45 (in equimolar ratios) incubated with RNA (2 ADAR:2 Zn72D/NF45:1 RNA in molar ratios) could form a complex with an even higher molecular weight than the Zn72D/NF45-RNA complex alone. This observation could lead to a novel structure and help solve how ADAR and Zn72D/NF45 interact upon RNA binding.

Another important consideration for the investigation of Zn72D-regulated RNA editing is the impact of tertiary structure on RNA editing sites. Secondary structure predictions give valuable information, but a determining factor in ADAR and Zn72D cooperation might be related to the three-dimensional structure of the RNA. Therefore, as another following project, Zn72D-dependent and Zn72D-independent RNA editing sites could be structurally characterised by RNAthor (Gumna et al., 2020) and QuShape (Karabiber et al., 2012).

In the last few years, A-to-I and C-to-U RNA editing have become promising therapeutic targets. Loss of RNA editing causes the activation of multiple dsRNA-sensing pathways mediated by MDA5, OAS, and PKR (see section 1.3.2). APOBEC3 proteins are capable of catalysing hypermutations in cancers that drive tumorigenesis and therapy resistance (Roberts et al., 2013). Knock-out of ADAR1 caused cancer cells to be more vulnerable to immunotherapy and overcome resistance to checkpoint blockade (M. Wang et al., 2019). This makes ADAR1 an attractive target to improve the therapeutic outcome in immuno checkpoint-resistant patients. The activation of these pathways leads to type I interferon (IFN) production and translational arrest, making cancer cells sensitive to cancer immunotherapy involving, for example, anti–programmed cell death 1 antibodies (Nakahama et al., 2020).

ADAR and APOBEC enzymes might also be exploited as tools to restore disease-causing point mutations by introducing single nucleotide changes, as

more than half of all known genetic disorders are caused by point mutations (Chen et al., 2019). In contrast to ADAR proteins, APOBEC3 can edit DNA, increasing the number of potential therapeutic approaches even more. This strategy is called site-directed RNA editing and provides the benefit of precise nucleotide modifications. There are several strategies on how RNA-editing enzymes could be used as a tool to manipulate the transcriptome in a programmable way. Two of these strategies envision use of endogenous ADAR enzymes for A-to-I RNA editing by introducing a single guide RNA into the cell. One approach is called recruiting endogenous ADAR to specific transcripts for oligonucleotide-mediated RNA editing (RESTORE). Introduced by Merkle et al., 2019 who were able to demonstrate the repair of the clinically relevant PiZZ mutation in primary human cells through endogenous human ADAR enzymes. These enzymes were recruited via guide RNAs that partly mimic an endogenous target (GluR2 mRNA) and a specificity domain that allows target mRNA binding next to the mutation (figure 6.1). The second approach, called leveraging endogenous ADAR for programmable editing of RNA (LEAPER), utilises long antisense RNAs (up to 200 nt) to generate double-stranded regions that resemble typical ADAR1 substrates to stimulate binding and editing by endogenous ADAR1 (Qu et al., 2019) (figure 6.1). In addition, off-target RNA editing, an undesired side-effect of these therapies, can be further reduced by the introduction of chemical modifications to the guide RNAs. RESTORE and LEAPER are both excellent examples of the potential medical applications of RNA editing. Factors that regulate RNA editing might also be harnessed for further development of similar therapeutic applications and might help in the selection of specific editing sites *in cellulo*. For example, knowledge on the mechanism by which ZFR/NF45 enhances editing could be harnessed to generate guide RNAs that work best in human tissues with high ZFR expression.

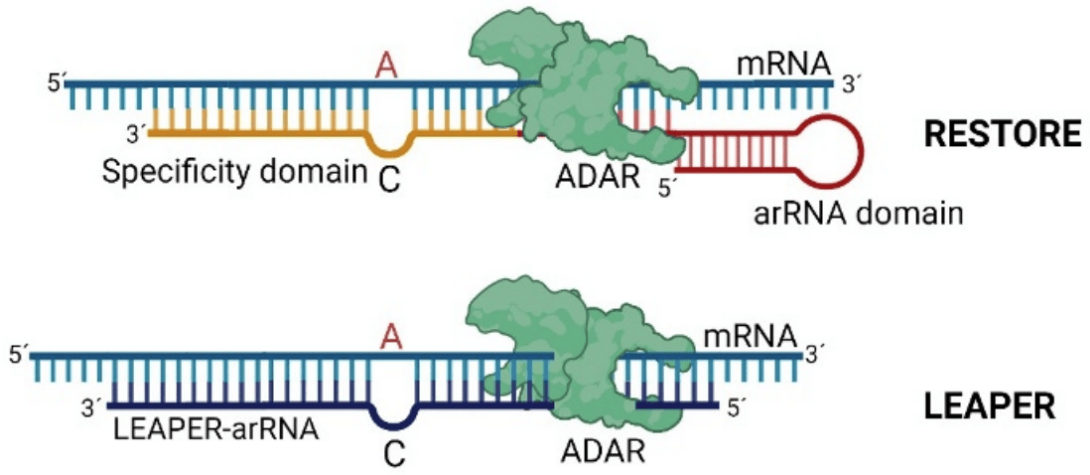


Figure 6.1: Novel therapeutic approaches based on RNA editing. Endogenous ADARs can be recruited via guide RNAs that either partly mimic an endogenous target (GluR2-arRNA, RESTORE) or that generate double-stranded RNAs of different lengths that recruit ADARs (LEAPER). Figure adapted from (Khosravi et al., 2021).

The second major focus of this thesis was to gain a mechanistic understanding of the role of ZFR in the regulation of alternative splicing. Human ZFR has been identified as an inhibitor of the type I interferon response upon viral infection, acting by regulation of alternative pre-mRNA splicing of the histone variant macroH2A1 (Haque et al., 2018). Activation of IFNß is regulated predominantly on the level of transcription in response to infection. macroH2A1 not only suppresses the initiation of type I interferon signalling by inhibition of IFNB1 transcription, but is also able of binding and repressing promoters of interferon stimulated genes (Lavigne et al., 2015). Moreover, ZFR regulates several hundred other alternative editing events, including intron retention, alternative 5' and 3' splice sites, and exon skipping. We hypothesised that the RNA-binding ability of ZFR is essential for its biological function. Consequently, the characterisation of its RNA binding specificity was a main goal of this thesis. Identification of RNA motifs bound by ZFR might help us learn more about its precise biological role in splicing.

To this end, recombinant human ZFR was produced by co-expression with NF45, and constructs that contained or lacked the three zinc-finger domains were purified. We were able to show with EMSAs that constructs lacking zinc finger domains have weak RNA binding activity compared to constructs with zinc fingers. These two constructs were used in an RNA Bind-n-Seq experiment (RBNS) performed by the Hogg group (NIH, Bethesda, USA). The RBNS data suggested a propensity for ZFR to bind structured RNAs with a C-rich

context. *In vitro* verification of binding of RBNS-obtained motifs indicated that ZFR has a strong preference for dsRNA over ssRNA. This suggests that *in vivo* binding sites that determine alternative splicing are likely to contain substantial secondary structure. EMSA-based observations indicated that ZFR/NF45 complexes bind dsRNA and RNA-DNA duplexes without sequence specificity. ZFR's ability to bind RNA-DNA duplexes is most likely due to duplexed nucleic acids generally adopting A-form helices. Given that ZFR/NF45 showed no apparent sequence specificity *in vitro*, why did the RBNS data give an apparent preference for C-rich motifs over other sequences? A closer examination of the data revealed that the sequences most strongly enriched for ZFR/NF45 were complementary to a sequence in the 3' PCR primer linker sequence required for the RBNS amplification step. C-rich sequences are enriched in the RBNS results as they allow secondary structure formation with the linker sequences. The preference of long-ZFR/NF45 for dsRNA caused it to pull-down sequences with base-pairing potential to the PCR primer binding sites. This observation explains why we were not able to detect any obvious differences in RNA binding despite using dsRNAs of different sequence compositions.

During the structural characterisation, human ZFR/NF45 formed RNA-protein complexes with a 2:1 (protein to RNA in molar ratios) stoichiometry based on anlytical gel filtration experiments, but aggregated at room temperature during SEC-MALS experiments. This 2:1 stoichiometry suggests that the formation of ZFR/NF45 heterodimers on RNA is a conserved feature that has also been observed with related NF90/NF45 complexes (Jayachandran et al., 2015). This behaviour suggests that dimerisation of members of the DZF protein family on dsRNA is of functional importance. Therefore, the proposed experiments for structural studies of ADAR and Zn72D/NF45 should also be applied to the ZFR/NF45 RNA-protein complexes and would be a great addition to the biochemical data obtained from RBNS and RNA binding studies.

New cross-linking mass spectrometry data for RNA-protein complexes suggests a novel dimer-dimer interface between the DZF domains of human NF90 and NF45, whose distances are not compatible with a 24Å linker distance. This data supports a model that DZF proteins oligomerise with critical implications for their biological function. Further, unpublished data from Uma Jayachandran based on negative stain electron microscopy of an RNA-protein complex of human NF90/NF45 with long dsRNA (over 300 nt) showed oligomerisation resembling beads on a string, dependent on the length of substrate RNA. A sim-

ilar cooperative binding model has recently been proposed for human ADAR. Multiple human ADAR proteins cooperate to bind dsRNA and perform A-to-I RNA editing (Song et al., 2020). These observations of oligomerisation of DZF proteins are also consistent with the eCLIP data from the Hogg group showing ZFR cross-links all across skipped exons. This oligomerisation behaviour should also be tested for other members of the DZF family such as ZFR/Zn72D. The incubation of purified recombinant Zn72D/NF45 and ZFR/NF45 as well as ADAR with long dsRNA for negative stain electron microscopy represents one of the subsequent key experiments to investigate potential oligomerisation on double-stranded RNA in editing and splicing regulation.

Another important experiment to unravel more of ZFR's biological role would be the identification of protein-protein interactions of ZFR as well as Zn72D in cells. This would give insights into how ZFR is regulated upstream and could perhaps explain under which conditions its splicing activity will be induced or inhibited, thereby modulating interferon β response. Also, since ZFR and Zn72D do not have an enzymatic function, they might instead act as a mediator or adaptor protein for ADAR. Emily Freund from the Li group at Stanford used BioID in a human cell line to identify proteins interacting with ADAR1 and ADAR2. Among the regulators that she identified were all four members of the DZF protein family: NF90, NF45, SPNR, and ZFR, which indicated that the DZF proteins interact with human ADAR enzymes and modulate RNA editing levels.

Interestingly, NF90 was reported to be a negative regulator of editing, whereas ZFR, in contrast, appeared to be a positive regulator (Freund et al., 2020). Results obtained from our cooperation with the Hogg group (unpublished data) showed that ZFR and NF90 have opposing roles in a subset of alternative splicing events. Furthermore, NF90 and NF45 have recently also been reported to play an essential role in mitosis by competing with the SMD machinery to regulate the expression of cell cycle mRNAs (Nourreddine et al., 2020). In conclusion, BioID experiments to identify protein-protein interactions could help test the hypothesis that DZF proteins such as ZFR, Zn72D and NF90 interact with editing enzymes and splicing regulators to modulate their functions.

7 Appendix

7.1 Plasmid maps

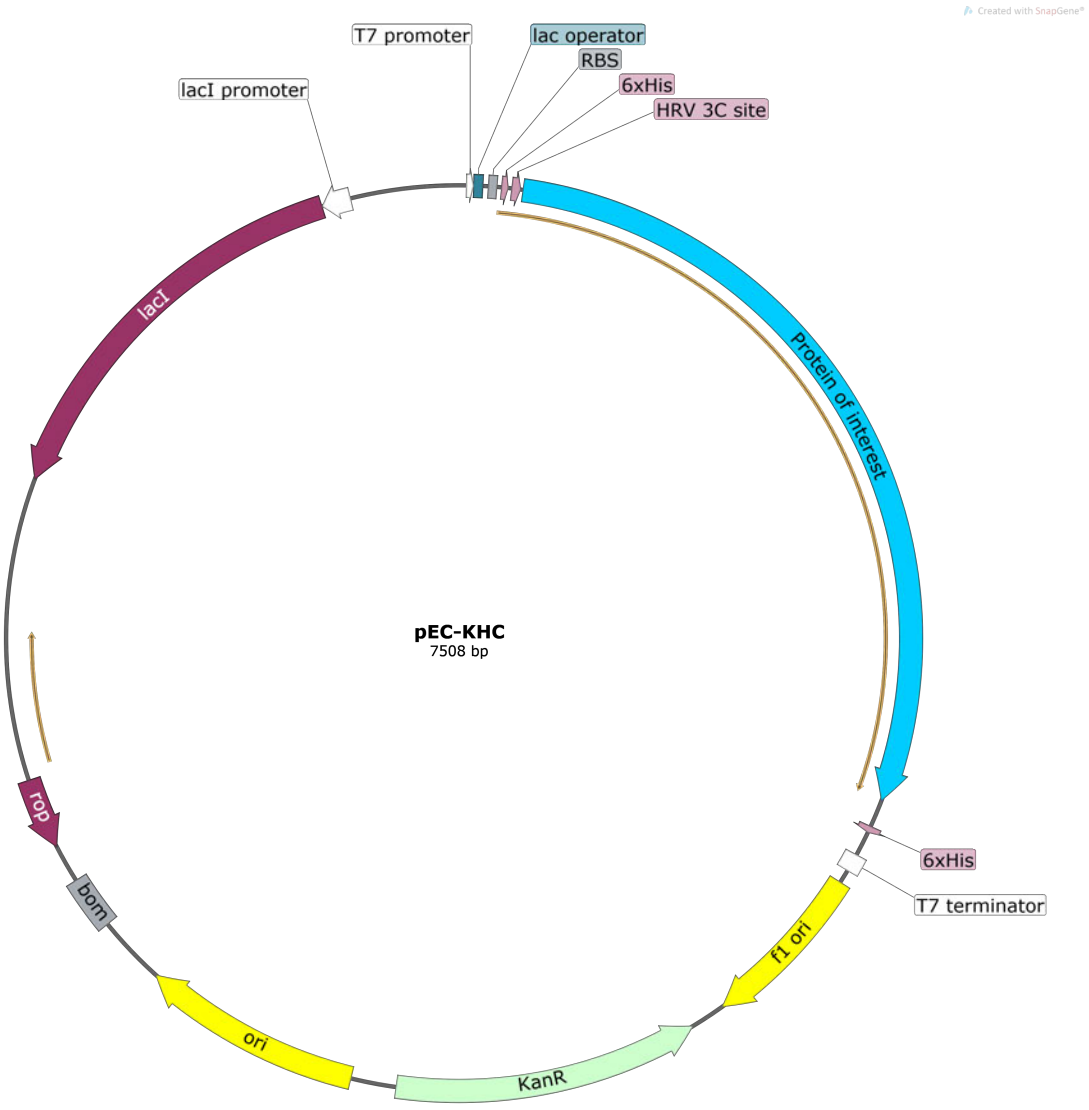


Figure 7.1: Plasmid map of pEC-KHC. pEC-KHC plasmid encodes for a kanamycin resistance gene (K) and a 3C protease-cleavable (C) N-terminal 6xHis-tag (H).

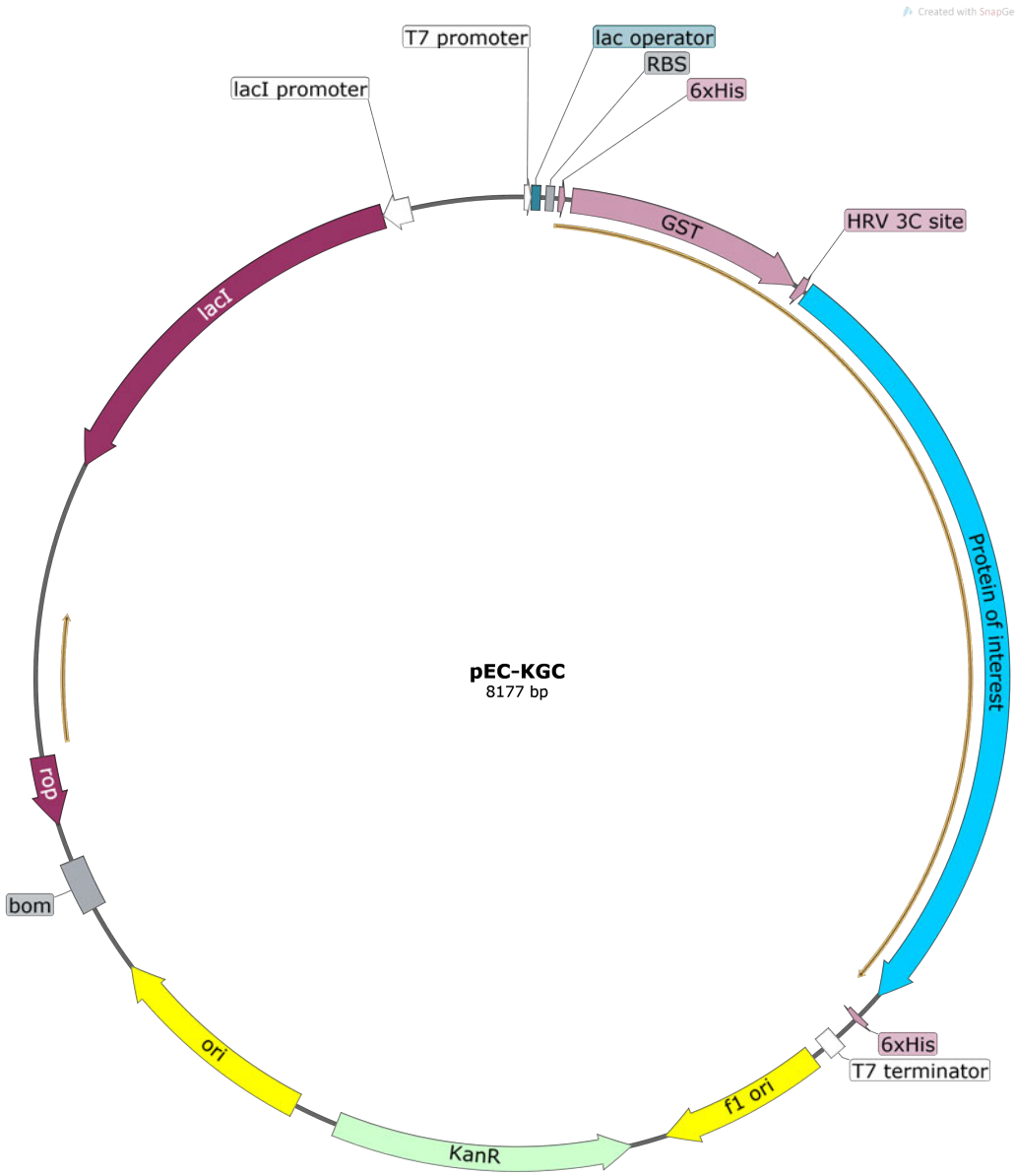


Figure 7.2: Plasmid map of pEC-KGC. pEC-KHC plasmid encodes for a kanamycin resistance gene (K) and a 3C protease-cleavable (C) N-terminal 6xHis-GST-tag (G).

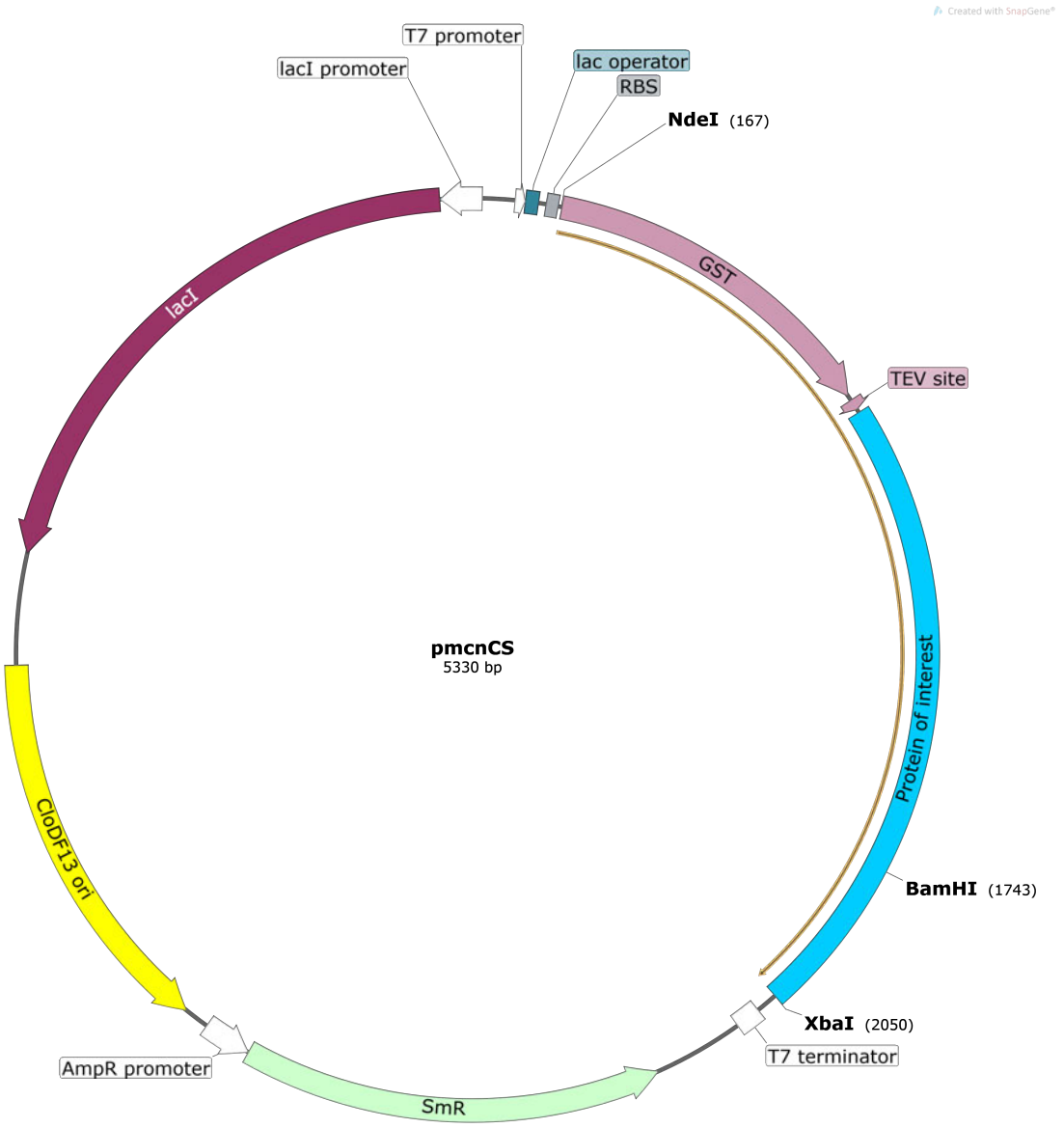


Figure 7.3: Plasmid map of pmcncCS. pmcnCS plasmid encodes for a streptomycin resistance gene and a TEV protease-cleavable N-terminal GST-tag.

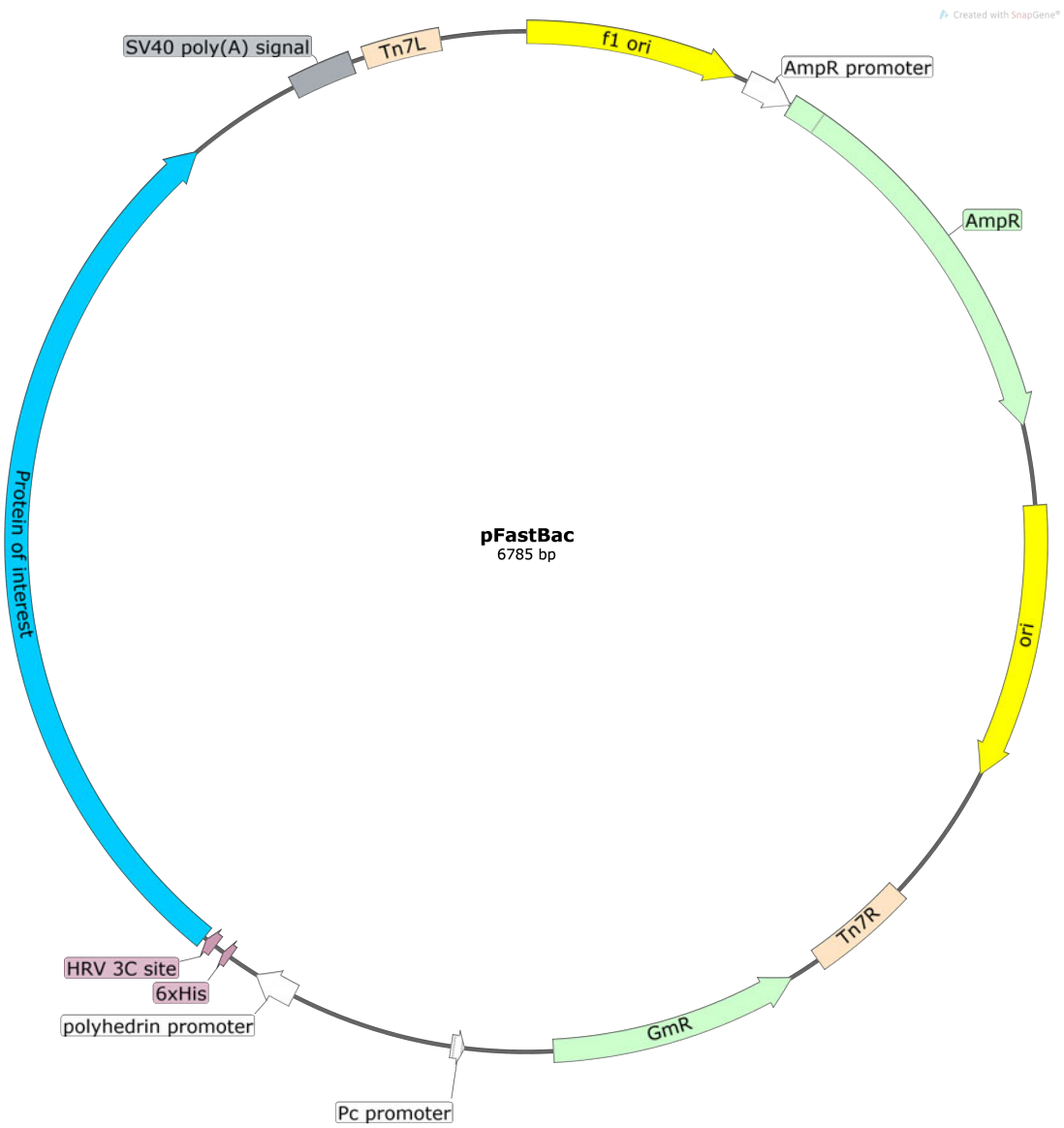


Figure 7.4: Plasmid map of pFastBac. pFastBac plasmid encodes for an ampicillin and a gentamicin resistance gene and 3C protease-cleavable N-terminal 6xHis-tag within Tn7L and Tn7R cassettes.

Figure 7.5: Plasmid map of pFL-GST. pFL-GST plasmid encodes for an ampicillin and a gentamicin resistance gene and 3C protease-cleavable N-terminal 6xHis-GST-tag within Tn7L and Tn7R cassettes.

7.2 Amino acid sequences of (wild-type) proteins used for recombinant protein production

Human ZFR (1-1074 aa)

```
   1 MIPICPVVSF TYVPSRLGED AKMATGNYFG FTHSGAAAAA AAAQYSQQPA
  51 SGVAYSHPTT VASYTVHQAP VAAHTVTAAY APAAATVAVA RPAPVAVAAA
 101 ATAAAYGGYP TAHTATDYGY TQRQQEAPPP PPPATTQNYQ DSYSYVRSTA
 151 PAVAYDSKQY YQQPTATAAA VAAAAQPQPS VAETYYQTAP KAGYSQGATQ
 201 YTQAQQTRQV TAIKPATPSP ATTTFSIYPV SSTVQPVAAA ATVVPSYTQS
 251 ATYSTTAVTY SGTSYSGYEA AVYSAASSYY QQQQQQQKQA AAAAAAAAAT
 301 AAWTGTTFTK KAPFQNKQLK PKQPPKPPQI HYCDVCKISC AGPQTYKEHL
 351 EGQKHKKKEA ALKASQNTSS SNSSTRGTQN QLRCELCDVS CTGADAYAAH
 401 IRGAKHQKVV KLHTKLGKPI PSTEPNVVSQ ATSSTAVSAS KPTASPSSIA
 451 ANNCTVNTSS IATSSMKGLT TTGNSSLNST SNTKVSAVPT NMAAKKTSTP
 501 KINFVGGNKL QSTGNKAEDI KGTECVKSTP VTSAVQIPEV KQDTVSEPVT
 551 PASLAALQSD VQPVGHDYVE EVRNDEGKVI RFHCKLCECS FNDPNAKEMH
 601 LKGRRHRLQY KKKVNPDLQV EVKPSIRARK IQEEKMRKQM QKEEYWRRRE
 651 EEERWRMEMR RYEEDMYWRR MEEEQHHWDD RRRMPDGGYP HGPPGPLGLL
 701 GVRPGMPPQP QGPAPLRRPD SSDDRYVMTK HATIYPTEEE LQAVQKIVSI
 751 TERALKLVSD SLSEHEKNKN KEGDDKKEGG KDRALKGVLR VGVLAKGLLL
 801 RGDRNVNLVL LCSEKPSKTL LSRIAENLPK QLAVISPEKY DIKCAVSEAA
 851 IILNSCVEPK MQVTITLTSP IIREENMREG DVTSGMVKDP PDVLDRQKCL
 901 DALAALRHAK WFQARANGLQ SCVIIIRILR DLCQRVPTWS DFPSWAMELL
 951 VEKAISSASS PQSPGDALRR VFECISSGII LKGSPGLLDP CEKDPFDTLA
1001 TMTDQQREDI TSSAQFALRL LAFRQIHKVL GMDPLPQMSQ RFNIHNNRKR
1051 RRDSDGVDGF EAEGKKDKKD YDNF
```

Human NF45 (1-390 aa)

```
   1 MRGDRGRGRG GRFGSRGGPG GGFRPFVPHI PFDFYLCEMA FPRVKPAPDE
  51 TSFSEALLKR NQDLAPNSAE QASILSLVTK INNVIDNLIV APGTFEVQIE
 101 EVRQVGSYKK GTMTTGHNVA DLVVILKILP TLEAVAALGN KVVESLRAQD
 151 PSEVLTMLTN ETGFEISSSD ATVKILITTV PPNLRKLDPE LHLDIKVLQS
 201 ALAAIRHARW FEENASQSTV KVLIRLLKDL RIRFPGFEPL TPWILDLLGH
 251 YAVMNNPTRQ PLALNVAYRR CLQILAAGLF LPGSVGITDP CESGNFRVHT
 301 VMTLEQQDMV CYTAQTLVRI LSHGGFRKIL GQEGDASYLA SEISTWDGVI
 351 VTPSEKAYEK PPEKKEGEEE EENTEEPPQG EEEESMETQE
```

Human EndoV (1-282 aa)

```
  1 MALEAAGGPP EETLSLWKRE QARLKAHVVD RDTEAWQRDP AFSGLQRVGG
 51 VDVSFVKGDS VRACASLVVL SFPELEVVYE ESRMVSLTAP YVSGFLAFRE
101 VPFLLELVQQ LREKEPGLMP QVLLVDGNGV LHHRGFGVAC HLGVLTDLPC
151 VGVAKKLLQV DGLENNALHK EKIRLLQTRG DSFPLLGDSG TVLGMALRSH
201 DRSTRPLYIS VGHRMSLEAA VRLTCCCCRF RIPEPVRQAD ICSREHIRKS
251 LGLPGPPTPR SPKAQRPVAC PKGDSGESSA LC
```

Drosophila melanogaster Zn72D (1-886 aa)

```
  1 MANNNYAGFN YGGTQYNTGQ VSYPAVTNAT YANAAAYQNA AVAAGQGGYG
 51 GAAGAGAGAT SGPGAGGYGG YGDYRSAMQY DATKTFYQQS PASYNASGST
101 ASVSKTHYSA PPVKNQVKGK MDKSNGGPKP PSSAPPAGNN YSGYDTALYN
151 AASMYVAQQH QGNPNQKPNG GANNWYQRKM GATIPGATAI RGMRPKAPPR
201 PQQLHYCEVC KISCAGPQTY REHLEGQKHK KREASLKMSA SANSATQNRG
251 NNYHCELCDV TCTGTDAYAA HVRGAKHQKV VKLHQKLGKP IPSEEPKKMG
301 KINFVPAAAG GAGVAKTEGG ANESDAAGDL DDNLDDSLGE NTDNIKPVGG
351 EYIEEVKDEE GKILSFNCKL CDCKFNDPNA KEMHMKGRRH RLQYKRKVQP
401 DLVVDFKPTP RQRRLAEARA NRAMMSSHRG GDDHDGGYWE EQRNRQYNEE
451 YDYNNWMSRS FGGAQRFGRM GNGPPPHFGM MPGGNVRRPE STDDRHAIAR
501 HAEIYPKEEE LQTIQRIVSH TERALKLVSD ALAEQPSDAG AANKKDKAEK
551 PSEKDGRDNQ IFSFQKDADN GGNVVRILKG VMRVGYLAKG LLLHGDNAVE
601 LVVLCAEKPT SGLLQRVANV LPDKLKEVAG DTQVNYRVEV NAEEAALIVL
651 DESVSVKITL TSPLLRDANP GEAATTDEGE GDSEFLPREP CLRALADLRH
701 AKWFQARATG LQSCVMVIRI LRDLCQRVSS WQSLPQWSLE LLVEKVISSA
751 GFPISPGDCM RRIMEALSSG FLINGPGLLD PCEKDPTDAL LELTKQERED
801 LTVSAQLFLR YIAFRQIYKV LGMEPLPAMK FPMRPWRVNR KRRRSSGKAG
851 GAPGTETESN DIDETGSDEK VAKKEGVGGT PAST
```

Drosophila melanogaster NF45 (1-396 aa)

```
  1 MVRGALRGGR PMRGGIRPPF KKTFVPRHPF DLTLAEVFFP KVPSAGAVDD
 51 SALTAALLKR NQDLSPTPSE QTAIGNLVTK VQAVLDNLVV APGDLTTCQL
101 EEVRQVGSFK KGTILTGNNV ADVVVILKTL PTKEAVDALA KKVEADLKAS
151 MKTEVLTKGD QHTVQIHERG FDIANVHAKV RILIATLPQN LRKLEPEIHL
201 DHKLMQSHLA AIRHTRWFEE NAHHSSIKVL IRILKDLTRR FDAFSPLSAW
251 MLDLIAHLAI MNNPSRQALP INLAFRRVFQ LLSAGLFLPG SAGITDPTEP
```

```
301 GHIRVHTAMT LEQQDVCCYT SQTLLRVLAH GGYKHILGLE GNTSVVREMS
351 VWNGVCISPL TAVYEKPTDK KEGDLEEDID MIENENEEEG SDDGAE
```

***Drosophila melanogaster* ADAR (1-632 aa)**

```
  1 MLNSANNNSP QHPVSAPSDI NMNGYNRKLP QKRGYEMPKY SDPKKKMCKE
 51 RIPQPKNTVA MLNELRHGLI YKLESQTGPV HAPLFTISVE VDGQKYLGQG
101 RSKKVARIEA AATALRSFIQ FKDGAVLSPL KPAGNLDFTS DEHLENDVSK
151 SAITVDGQKK VPDKGPVMLL YELFNDVNFE CINIDGAQNN CRFKMTVTIN
201 EKKFDGTGPS KKTAKNAAAK AALASLCNIS YSPMVVPQKN VPLPIDDKSS
251 SMELPQIHAD TIGRLVLEKF MEVIKGQEAY SRRKVLAGIV MTENMNFCEA
301 KVISVSTGTK CVSGEHMSVN GAVLNDSHAE IVSRRCLLKY LYAQLDLQCN
351 QATAYQSIFV RNTDGQYPYK LKSGVHFHLY INTAPCGDAR IFSPHENDTG
401 VDKHPNRKAR GQLRTKIESG EGTIPVKSSD GIQTWDGVLQ GQRLLTMSCS
451 DKIARWNIVG IQGSLLSSII EPVYLHSIVL GSLLHPEHMY RAVCGRIEKS
501 IQGLPPPYHL NKPRLALVTS AEPRNQAKAP NFGINWTIGD TELEVVNSLT
551 GRTIGGQVSR ITKQAFFVKY GFLMANLPGI LVRKVTTDYG QTKANVKDYQ
601 IAKLELFSAF KREDLGSWLK KPIEQDEFGL AE
```

7.2.1 Input sequences for multiple sequence alignments

Table 7.1: UniProtKB entries of input sequences used for generating multiple sequence alignments

Protein	Organism	Label	UniProtKB	Number of amino acids
ADAR	*Drosophila melanogaster*	Drosophila_ADAR	M9NFQ6	632
ADAR	*Caenorhabditis elegans*	Worm_ADAR1	Q9U3D6	964
ADAR	*Caenorhabditis elegans*	Worm_ADAR2	Q22618	495
ADAR1 p110	*Homo sapiens*	Human_ADAR1.p110	P55265-5	931
ADAR1 p150	*Homo sapiens*	Human_ADAR1.p150	P55265-1	1226
ADAR2	*Homo sapiens*	Human_ADAR2	P78563	741
ADAR3	*Homo sapiens*	Human_ADAR3	Q9NS39	739
NF45	*Danio rerio*	Zebrafish	Q6NZ06	387
NF45	*Drosophila melanogaster*	Fruit_fly	Q9VG73	396
NF45	*Homo sapiens*	Human	Q12905	390
NF45	*Mus musculus*	Mouse	Q9CXY6	390
NF45	*Xenopus laevis*	Frog	Q7ZY44	388
NF90	*Homo sapiens*	Human	Q12906	894
SPNR	*Homo sapiens*	Human	Q96SI9	672
ZFR	*Cavia porcellus*	Guinea_pig	H0UUW7	972
ZFR	*Danio rerio*	Zebrafish	Q6PCR6	1074
ZFR	*Gallus gallus*	Chicken	F1NLI6	1067
ZFR	*Homo sapiens*	Human	Q96KR1	1074
ZFR	*Mus musculus*	Mouse	O88532	1074
ZFR	*Xenopus laevis*	Frog	Q6GPM1	1054
Zn72D	*Apis mellifera*	Honey_bee	A0A088AKW6	997
Zn72D	*Drosophila melanogaster*	Fruit_fly	Q86BI3	884

7.3 Additional data for RNA template identification

Table 7.2: Description for the rows of the RNAi screen raw data and additional information

Row name	Description
chr	Chromosomal location of the editing site
pos	Genomic coordinate of the editing site (dm3)
gene	Gene name
annot1	Location of editing site in regard to mRNA (exonic, intronic, 5' UTR, and 3' UTR)
annot2	Effect of editing site (nonsynonymous changes amino acid)
ecs	Editing site complementary sequence
gfp_avg	Editing levels after GFP siRNA treatment
gfp_sem	Standard error of mean of gfp_avg
adar_avg	Editing levels after ADAR siRNA treatment
adar_sem	Standard error of mean of adar_avg
zn72d_avg	Editing levels after Zn72D siRNA treatment
zn72d_sem	Standard error of mean of zn72d_avg
$\triangle$Zn72D	Editing levels upon ADAR knockdown compared with GFP control
$\triangle$ ADAR	Editing levels upon Zn72D knockdown compared with GFP control
$\triangle$Zn72D-$\triangle$ADAR	Difference of editing levels between ADAR and Zn72D knockdown compared with GFP control
Zn72D KG [%]	Change of editing levels upon Zn72D knockdown in %
ADAR KG [%]	Change of editing levels upon ADAR knockdown %
Whole body [%]	Editing levels in *Drosophila* for the whole body
Head [%]	Editing levels in *Drosophila* for the head
Editing sites	Number of additional editing sites within 100 bp up and downstream of the editing site
Effect	Amino acid change of editing site
Strand	Location of editing site (positive or negative strand)
MFE [kcal/mol]	Minimum free energy for secondary structure formation

Table 7.3: Overview of *in vivo* RNAi screen data of selected Zn72D-independent RNA editing sites (part 1)

chr	pos	gene	annot1	annot2	ecs	gfp_avg	gfp_sem	adar_avg	adar_sem	zn72d_avg
2L	11830365	Pde1c	exonic	nonsyn	proximal	0.487	0.021	0.123	0.025	0.42
X	**5524267**	**CG42492**	**exonic**	**nonsyn**	**unknown**	**0.657**	**0.017**	**0.296**	**0.016**	**0.651**
2R	**20798584**	**NaCP60E**	**exonic**	**nonsyn**	**proximal**	**0.526**	**0.019**	**0.205**	**0.016**	**0.508**
4	**1268833**	**Caps**	**exonic**	**nonsyn**	**proximal**	**0.425**	**0.008**	**0.112**	**0.009**	**0.419**
3L	2925255	Shab	exonic	nonsyn	proximal	0.836	0.015	0.364	0.022	0.738
3R	14797034	CG42613	exonic	nonsyn	distal	0.719	0.01	0.33	0.015	0.702
2R	20798298	NaCP60E	exonic	nonsyn	proximal	0.743	0.017	0.359	0.02	0.771
2R	11406217	Khc-73	exonic	unknown	proximal	0.517	0.021	0.139	0.027	0.431
X	18064159	Frq1	exonic	nonsyn	unknown	0.769	0.013	0.445	0.02	0.765
3L	4586715	CG42540	exonic	nonsyn	unknown	0.773	0.006	0.458	0.01	0.729
X	11863002	cac	exonic	nonsyn	distal	0.399	0.019	0.087	0.023	0.402
X	5525081	CG42492	exonic	nonsyn	unknown	0.718	0.015	0.412	0.022	0.689
3L	4429963	nAcRbeta-64B	UTR3	UTR3	unknown	0.92	0.011	0.445	0.033	0.898
3R	16819563	Calx	exonic	nonsyn	unknown	0.611	0.013	0.145	0.02	0.582
X	16412866	para	exonic	nonsyn	distal	0.696	0.017	0.231	0.033	0.734
3L	5689230	sif	exonic	nonsyn	unknown	0.495	0.016	0.153	0.012	0.402
2R	7335122	qvr	exonic	nonsyn	distal	0.808	0.008	0.478	0.028	0.732

Table 7.4: Overview of *in vivo* RNAi screen data of selected Zn72D-independent RNA editing sites (part 2)

chr	pos	gene	zn72d_sem	△Zn72D	△ADAR	△Zn72D-△ADAR	Zn72D KD [%]	ADAR KD [%]
2L	11830365	Pde1c	0.015	-0.067	-0.364	0.297	13.8	74.7
X	**5524267**	**CG42492**	**0.03**	**-0.006**	**-0.361**	**0.355**	**0.9**	**54.9**
2R	**20798584**	**NaCP60E**	**0.03**	**-0.018**	**-0.321**	**0.303**	**3.4**	**61**
4	**1268833**	**Caps**	**0.019**	**-0.006**	**-0.313**	**0.307**	**1.4**	**73.6**
3L	2925255	Shab	0.021	-0.098	-0.472	0.374	11.7	56.5
3R	14797034	CG42613	0.007	-0.017	-0.389	0.372	2.4	54.1
2R	20798298	NaCP60E	0.018	0.028	-0.384	0.412	-3.8	51.7
2R	11406217	Khc-73	0.034	-0.086	-0.378	0.292	16.6	73.1
X	18064159	Frq1	0.039	-0.004	-0.324	0.32	0.5	42.1
3L	4586715	CG42540	0.036	-0.044	-0.315	0.271	5.7	40.8
X	11863002	cac	0.019	0.003	-0.312	0.315	-0.8	78.2
X	5525081	CG42492	0.011	-0.029	-0.306	0.277	4.0	42.6
3L	4429963	nAcRbeta-64B	0.011	-0.022	-0.475	0.453	2.4	51.6
3R	16819563	Calx	0.028	-0.029	-0.466	0.437	4.7	76.3
X	16412866	para	0.034	0.038	-0.465	0.503	-5.5	66.8
3L	5689230	sif	0.023	-0.093	-0.342	0.249	18.8	69.1
2R	7335122	qvr	0.016	-0.076	-0.33	0.254	9.4	40.8

Table 7.5: Overview of *in vivo* RNAi screen data of selected Zn72D-independent RNA editing sites (part 3)

chr	pos	gene	Whole body [%]	Head [%]	Editing sites	Effect	Strand	MFE [kcal/mol]
2L	11830365	Pde1c	14.3	62	3	Ser to Gly	-	-60.3
X	**5524267**	**CG42492**	**16.7**	**52**	**3**	**Gln to Arg**	**-**	**-85.7**
2R	**20798584**	**NaCP60E**	**20.9**	**50**	**1**	**Thr to Ala**	**+**	**-58.5**
4	**1268833**	**Caps**	**8.7**	**41**	**1**	**Met to Val**	**+**	**-44.2**
3L	2925255	Shab	42.7	77	2	Ser to Gly	+	-62.8
3R	14797034	CG42613	30.5	53	1	Ile to Met	+	-28.7
2R	20798298	NaCP60E	13.6	78	4	Tyr to Cys	+	-69.3
2R	11406217	Khc-73	2.9	44	3		-	-80.3
X	18064159	Frq1	67.9	54	4	Ile to Met	+	-34.9
3L	4586715	CG42540	48	76	3	Asn to Asp	+	-48
X	11863002	cac	N/A	27	2	Asn to Ser	-	-51
X	5525081	CG42492	20.5	65	4	Thr to Ala	-	-54.1
3L	4429963	nAcRbeta-64B	45.8	81	3		+	-31.8
3R	16819563	Calx	9	56	5	Glu to Gly	+	-62.9
X	16412866	para	75	60	2	Gln to Arg	-	-12
3L	5689230	sif	17.6	31	2	Thr to Ala	+	-82
2R	7335122	qvr	27.6	51	5	Asn to Ser	+	-49

Table 7.6: Overview of *in vivo* RNAi screen data of selected Zn72D-dependent RNA editing sites (part 1)

chr	pos	gene	annot1	annot2	ecs	gfp_avg	gfp_sem	adar_avg	adar_sem	zn72d_avg
X	15286718	cngl	exonic	nonsyn	unknown	0.829	0.016	0.32	0.01	0.622
2L	9717614	Nckx30C	exonic	nonsyn	proximal	0.832	0.01	0.337	0.017	0.308
X	6979849	Sxl	exonic	nonsyn	proximal	0.781	0.022	0.334	0.051	0.469
X	16399583	para	exonic	nonsyn	unknown	0.671	0.017	0.244	0.014	0.36
2L	14090945	nAcRalpha-34E	UTR3	UTR3	unknown	0.683	0.024	0.261	0.014	0.407
3R	11238976	Atx2	exonic	nonsyn	proximal	0.613	0.03	0.222	0.03	0.183
X	8134123	sdt	UTR3	UTR3	unknown	0.649	0.032	0.259	0.051	0.221
2R	20802225	NaCP60E	exonic	nonsyn	proximal	0.614	0.016	0.258	0.016	0.368
X	14992625	rab3-GEF	exonic	nonsyn	unknown	0.409	0.022	0.081	0.016	0.025
3R	23500458	CG18437	exonic	nonsyn	proximal	0.825	0.013	0.377	0.008	0.375
2R	20802387	NaCP60E	exonic	nonsyn	proximal	0.712	0.015	0.271	0.007	0.419
X	15852620	mmd	exonic	nonsyn	distal	0.789	0.011	0.365	0.031	0.346
3R	6071147	CG42795	exonic	nonsyn	proximal	0.686	0.023	0.28	0.031	0.269
2R	20802397	NaCP60E	exonic	nonsyn	proximal	0.553	0.021	0.153	0.007	0.235
X	11871571	cac	exonic	nonsyn	unknown	0.794	0.013	0.396	0.046	0.546
3L	20766250	Pitslre	UTR5	UTR5	unknown	0.662	0.019	0.305	0.019	0.338
X	2522899	PI4KIIIalpha	exonic	nonsyn	proximal	0.547	0.014	0.197	0.032	0.149
2R	20802388	NaCP60E	exonic	nonsyn	proximal	0.585	0.021	0.244	0.009	0.345
3R	11239027	Atx2	exonic	nonsyn	unknown	0.559	0.025	0.225	0.017	0.172
3L	4369184	DopEcR	exonic	nonsyn	unknown	0.842	0.024	0.517	0.024	0.491
3R	13707142	Rim	exonic	syn	unknown	0.654	0.015	0.337	0.034	0.332

Table 7.7: Overview of *in vivo* RNAi screen data of selected Zn72D-dependent RNA editing sites (part 2)

chr	pos	gene	zn72d_sem	△Zn72D	△ADAR	△ADAR-△Zn72D	Zn72D KD [%]	ADAR KD [%]
X	15286718	cngl	0.035	-0.207	-0.509	-0.302	25	61.4
2L	9717614	Nckx30C	0.014	-0.524	-0.495	0.029	63	59.5
X	6979849	Sxl	0.048	-0.312	-0.447	-0.135	39.9	57.2
X	16399583	para	0.021	-0.311	-0.427	-0.116	46.3	63.6
2L	14090945	nAcRalpha-34E	0.024	-0.276	-0.422	-0.146	40.4	61.8
3R	11238976	Atx2	0.057	-0.43	-0.391	0.039	70.1	63.8
X	8134123	sdt	0.056	-0.428	-0.39	0.038	65.9	60.1
2R	20802225	NaCP60E	0.021	-0.246	-0.356	-0.11	40.1	58
X	14992625	rab3-GEF	0.006	-0.384	-0.328	0.056	93.9	80.2
3R	23500458	CG18437	0.022	-0.45	-0.448	0.002	54.5	54.3
2R	20802387	NaCP60E	0.023	-0.293	-0.441	-0.148	41.2	61.9
X	15852620	mmd	0.009	-0.443	-0.424	0.019	56.1	53.7
3R	6071147	CG42795	0.031	-0.417	-0.406	0.011	60.8	59.2
2R	20802397	NaCP60E	0.022	-0.318	-0.4	-0.082	57.5	72.3
X	11871571	cac	0.027	-0.248	-0.398	-0.15	31.2	50.1
3L	20766250	Pitslre	0.029	-0.324	-0.357	-0.033	48.9	53.9
X	2522899	PI4KIIIalpha	0.019	-0.398	-0.35	0.048	72.8	64
2R	20802388	NaCP60E	0.016	-0.24	-0.341	-0.101	41	58.3
3R	11239027	Atx2	0.037	-0.387	-0.334	0.053	69.2	59.7
3L	4369184	DopEcR	0.015	-0.351	-0.325	0.026	41.7	38.6
3R	13707142	Rim	0.019	-0.322	-0.317	0.005	49.2	48.5

Table 7.8: Overview of *in vivo* RNAi screen data of selected Zn72D-dependent RNA editing sites (part 3)

chr	pos	gene	Whole body [%]	Head [%]	Editing sites	Effect	Strand	MFE [kcal/mol]
X	15286718	cngl	27.5	80	1	Thr to Ala	-	-80.2
2L	**9717614**	**Nckx30C**	**49.3**	**82**	**2**	**Lys to Arg**	**-**	**-73.3**
X	**6979849**	**Sxl**	**25.2**	**73**	**1**	**Ser to Gly**	**-**	**-82.5**
X	**16399583**	**para**	**29.4**	**66**	**1**	**Lys to Arg**	**-**	**-55**
2L	14090945	nAcRalpha-34E	N/A	N/A	3		+	-30.1
3R	11238976	Atx2	5.6	N/A	2	Lys to Arg	+	-80.3
X	8134123	sdt	N/A	N/A	2		+	-65.1
2R	20802225	NaCP60E	19.9	61	4	Ile to Val	+	-31.2
X	14992625	rab3-GEF	10.1	42	2	Ser to Gly	+	-69.2
3R	23500458	CG18437	38.7	80	3	Gln to Arg	+	-71.5
2R	20802387	NaCP60E	29.6	36	8	Lys to Glu	+	-58.1
X	15852620	mmd	86	79	2	Asn to Ser	-	-39.4
3R	6071147	CG42795	29	60	3	Ser to Gly	+	-69.5
2R	20802397	NaCP60E	20.8	24	8	Gln to Arg	+	-56.5
X	11871571	cac	81.8	71	1	Ile to Met	-	-38.6
3L	20766250	Pitslre	N/A	65	1		+	-44.7
X	2522899	PI4KIIIalpha	5.6	38	1	Tyr to Cys	+	-94.1
2R	20802388	NaCP60E	19.2	18	8	Lys to Arg	+	-58.5
3R	11239027	Atx2	5.9	41	2	Lys to Arg	+	-72.6
3L	4369184	DopEcR	40.9	65	6	Ile to Val	-	-26
3R	13707142	Rim	N/A	N/A	1	Lys to Lys	-	-45.9

7.4 Additional information for RNA end labelling

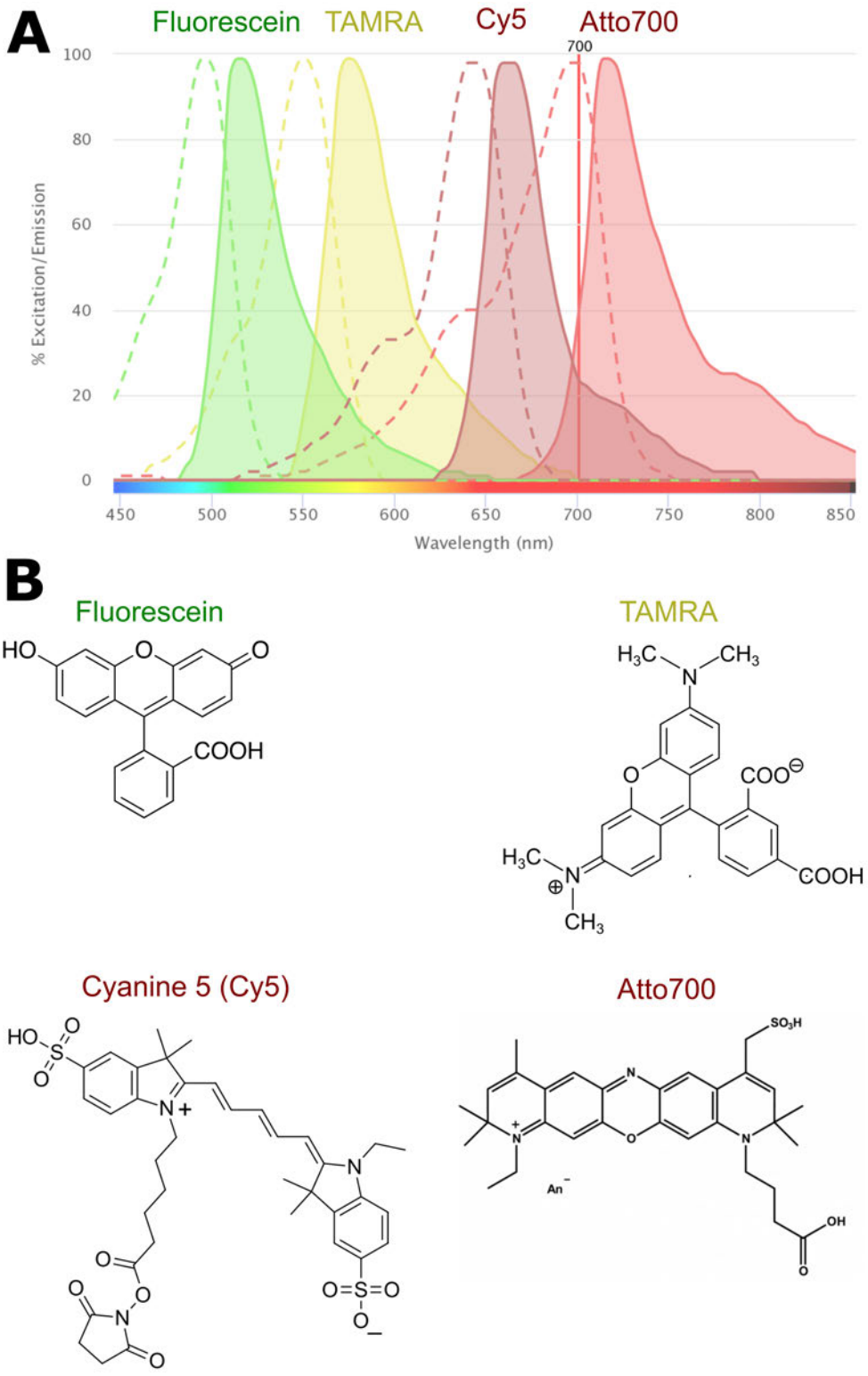


Figure 7.6: Fluorophores used for 5' end labelling. Absorption and emission spectra (A) (created with FluoroFinder) and structure (B) for fluorescein (green), TAMRA (yellow), cyanine 5 (dark red) and Atto700 (red) used for 5' end labelling of *in vitro* transcribed RNA. Detection wavelenght at 700 nm is indicated with a vertical red line.

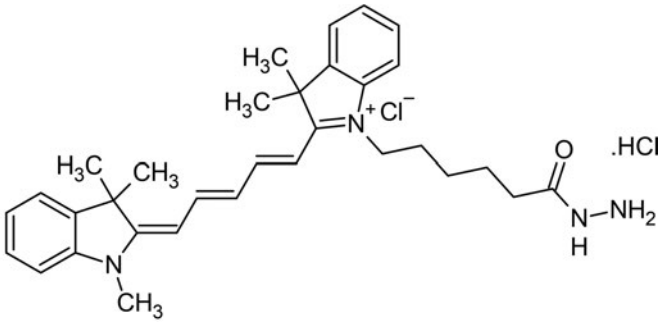


Figure 7.7: Fluorophores used for 3' end labelling. Structure of cyanine 5.5 hydrazide used for 3' end labelling of *in vitro* transcribed RNA.

Bibliography

Ablasser, Andrea, Marion Goldeck, Taner Cavlar, Tobias Deimling, Gregor Witte, Ingo Röhl, Karl-Peter Hopfner, Janos Ludwig and Veit Hornung (May 2013). 'cGAS produces a 2'-5'-linked cyclic dinucleotide second messenger that activates STING'. In: *Nature* 498.7454, pp. 380–384. DOI: 10.1038/nature12306.

Agris, Paul F., Franck A.P. Vendeix and William D. Graham (Feb. 2007). 'tRNA's Wobble Decoding of the Genome: 40 Years of Modification'. In: *Journal of Molecular Biology* 366.1, pp. 1–13. DOI: 10.1016/j.jmb.2006.11.046.

Ahmad, Sadeem, Xin Mu, Fei Yang, Emily Greenwald, Ji Woo Park, Etai Jacob, Cheng-Zhong Zhang and Sun Hur (Feb. 2018). 'Breaching Self-Tolerance to Alu Duplex RNA Underlies MDA5-Mediated Inflammation'. In: *Cell* 172.4, 797–810.e13. DOI: 10.1016/j.cell.2017.12.016.

Amrani, Nadia, Matthew S. Sachs and Allan Jacobson (June 2006). 'Early nonsense: mRNA decay solves a translational problem'. In: *Nature Reviews Molecular Cell Biology* 7.6, pp. 415–425. DOI: 10.1038/nrm1942.

Andrés-Colás, Nuria, Qiang Zhu, Mizuki Takenaka, Bert De Rybel, Dolf Weijers and Dominique Van Der Straeten (July 2017). 'Multiple PPR protein interactions are involved in the RNA editing system in mitochondria and plastids'. In: *Proceedings of the National Academy of Sciences* 114.33, pp. 8883–8888. DOI: 10.1073/pnas.1705815114.

Anna, Abramowicz and Gos Monika (Apr. 2018). 'Splicing mutations in human genetic disorders: examples, detection, and confirmation'. In: *Journal of Applied Genetics* 59.3, pp. 253–268. DOI: 10.1007/s13353-018-0444-7.

Aphasizhev, Ruslan and Inna Aphasizheva (May 2014). 'Mitochondrial RNA editing in trypanosomes: Small RNAs in control'. In: *Biochimie* 100, pp. 125–131. DOI: 10.1016/j.biochi.2014.01.003.

Aphasizheva, Inna, Liye Zhang, Xiaorong Wang, Robyn M. Kaake, Lan Huang, Stefano Monti and Ruslan Aphasizhev (Dec. 2014). 'RNA Binding and Core Complexes Constitute the U-Insertion/Deletion Editosome'. In: *Molecular and Cellular Biology* 34.23, pp. 4329–4342. DOI: 10.1128/mcb.01075-14.

Barbon, A (Oct. 2003). 'Glutamate receptor RNA editing: a molecular analysis of GluR2, GluR5 and GluR6 in human brain tissues and in NT2 cells following in vitro neural differentiation'. In: *Molecular Brain Research* 117.2, pp. 168–178. DOI: 10.1016/s0169-328x(03)00317-6.

Barkan, Alice and Ian Small (Apr. 2014). 'Pentatricopeptide Repeat Proteins in Plants'. In: *Annual Review of Plant Biology* 65.1, pp. 415–442. DOI: 10.1146/annurev-arplant-050213-040159.

Bass, B L and Weintraub H (Feb. 1987). 'A developmentally regulated activity that unwinds RNA duplexes'. In: *Cell* 48.4, pp. 607–613. DOI: 10.1016/0092-8674(87)90239-x.

Bass, Brenda (May 1997). 'RNA editing and hypermutation by adenosine deamination'. In: *Trends in Biochemical Sciences* 22.5, pp. 157–162. DOI: 10.1016/s0968-0004(97)01035-9.

Baumruk, Vladimir, Catherine Gouyette, Tam Huynh-Dinh, Jian-Sheng Sun and Mahmoud Ghomi (Oct. 2001). 'Comparison between CUUG and UUCG tetraloops: thermodynamic stability and structural features analyzed by UV absorption and vibrational spectroscopy'. In: *Nucleic Acids Research* 29.19, pp. 4089–4096. DOI: 10.1093/nar/29.19.4089.

Bazak, Lily et al. (Dec. 2013). 'A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes'. In: *Genome Research* 24.3, pp. 365–376. DOI: 10.1101/gr.164749.113.

Beaudoing, Emmanuel, Susan Freier, Jacqueline R. Wyatt, Jean-Michel Claverie and Daniel Gautheret (July 2000). 'Patterns of Variant Polyadenylation Signal Usage in Human Genes'. In: *Genome Research* 10.7, pp. 1001–1010. DOI: 10.1101/gr.10.7.1001.

Beghini, A., C. B. Ripamonti, P. Peterlongo, G. Roversi, R. Cairoli, E. Morra and L. Larizza (Sept. 2000). 'RNA hyperediting and alternative splicing of hematopoietic cell phosphatase (PTPN6) gene in acute myeloid leukemia'. In: *Human Molecular Genetics* 9.15, pp. 2297–2304. DOI: 10.1093/oxfordjournals.hmg.a018921.

Behm, Mikaela and Marie Oehman (Mar. 2016). 'RNA Editing: A Contributor to Neuronal Dynamics in the Mammalian Brain'. In: *Trends in Genetics* 32.3, pp. 165–175. DOI: 10.1016/j.tig.2015.12.005.

Behm-Ansmant, Isabelle and Elisa Izaurralde (Feb. 2006). 'Quality control of gene expression: a stepwise assembly pathway for the surveillance complex that triggers nonsense-mediated mRNA decay'. In: *Genes Development* 20.4, pp. 391–398. DOI: 10.1101/gad.1407606.

Belfort, Marlene (Dec. 1990). 'PHAGE T4 INTRONS: SELF-SPLICING AND MOBILITY'. In: *Annual Review of Genetics* 24.1, pp. 363–385. DOI: 10.1146/annurev.ge.24.120190.002051.

Benne, Rob (Apr. 1994). 'RNA editing in trypanosomes'. In: *European Journal of Biochemistry* 221.1, pp. 9–23. DOI: 10.1111/j.1432-1033.1994.tb18710.x.

Benne, Rob, Janny Van Den Burg, Just P.J. Brakenhoff, Paul Sloof, Jacques H. Van Boom and Marijke C. Tromp (Sept. 1986). 'Major transcript of the frameshifted coxll gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA'. In: *Cell* 46.6, pp. 819–826. DOI: 10.1016/0092-8674(86)90063-2.

Bentley, David L. (Feb. 2014). 'Coupling mRNA processing with transcription in time and space'. In: *Nature Reviews Genetics* 15.3, pp. 163–175. DOI: 10.1038/nrg3662.

Berget, Susan M., Claire Moore and Phillip A. Sharp (Aug. 1977). 'Spliced segments at the 5' terminus of adenovirus 2 late mRNA'. In: *Proceedings of the National Academy of Sciences* 74.8, pp. 3171–3175. DOI: 10.1073/pnas.74.8.3171.

Blanc, Valerie and Nicholas O. Davidson (Jan. 2003). 'C-to-U RNA Editing: Mechanisms Leading to Genetic Diversity'. In: *Journal of Biological Chemistry* 278.3, pp. 1395–1398. DOI: 10.1074/jbc.r200024200.

Blum, Beat, Norbert Bakalara and Larry Simpson (Jan. 1990). 'A model for RNA editing in kinetoplastid mitochondria: RNA molecules transcribed from maxicircle DNA provide the edited information'. In: *Cell* 60.2, pp. 189–198. DOI: 10.1016/0092-8674(90)90735-w.

Brisse, Morgan and Hinh Ly (July 2019). 'Comparative Structure and Function Analysis of the RIG-I-Like Receptors: RIG-I and MDA5'. In: *Frontiers in Immunology* 10. DOI: 10.3389/fimmu.2019.01586.

Burkard, Karina T. D. and J. Scott Butler (Jan. 2000). 'A Nuclear 3'-5' Exonuclease Involved in mRNA Degradation Interacts with Poly(A) Polymerase and the hnRNA Protein Npl3p'. In: *Molecular and Cellular Biology* 20.2, pp. 604–616. DOI: 10.1128/mcb.20.2.604-616.2000.

Casañal, Ana et al. (Oct. 2017). 'Architecture of eukaryotic mRNA 3' end processing machinery'. In: *Science* 358.6366, pp. 1056–1059. DOI: 10.1126/science.aao6535.

Cenci, Caterina, Rita Barzotti, Federica Galeano, Sandro Corbelli, Rossella Rota, Luca Massimi, Concezio Di Rocco, Mary A. O'Connell and Angela

Gallo (Mar. 2008). 'Down-regulation of RNA Editing in Pediatric Astrocytomas'. In: *Journal of Biological Chemistry* 283.11, pp. 7251–7260. DOI: 10.1074/jbc.m708316200.

Chen, Genghao, Dhruva Katrekar and Prashant Mali (Apr. 2019). 'RNA-Guided Adenosine Deaminases: Advances and Challenges for Therapeutic RNA Editing'. In: *Biochemistry* 58.15, pp. 1947–1957. DOI: 10.1021/acs.biochem.9b00046.

Chow, Louise T., James M. Roberts, James B. Lewis and Thomas R. Broker (Aug. 1977). 'A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA:DNA hybrids'. In: *Cell* 11.4, pp. 819–836. DOI: 10.1016/0092-8674(77)90294-x.

Chung, Hachung et al. (Feb. 2018). 'Human ADAR1 Prevents Endogenous RNA from Triggering Translational Shutdown'. In: *Cell* 172.4, 811–824.e14. DOI: 10.1016/j.cell.2017.12.038.

Cruz-Reyes, Jorge, Alevtina Zhelonkina, Laura Rusche and Barbara Sollner-Webb (Feb. 2001). 'Trypanosome RNA Editing: Simple Guide RNA Features Enhance U Deletion 100-Fold'. In: *Molecular and Cellular Biology* 21.3, pp. 884–892. DOI: 10.1128/mcb.21.3.884-892.2001.

Curinha, Ana, Sandra Oliveira Braz, Isabel Pereira-Castro, Andrea Cruz and Alexandra Moreira (Oct. 2014). 'Implications of polyadenylation in health and disease'. In: *Nucleus* 5.6, pp. 508–519. DOI: 10.4161/nucl.36360.

Desterro, Joana M. P., Liam P. Keegan, Miguel Lafarga, Maria Teresa Berciano, Mary O'Connell and Maria Carmo-Fonseca (May 2003). 'Dynamic association of RNA-editing enzymes with the nucleolus'. In: *Journal of Cell Science* 116.9, pp. 1805–1818. DOI: 10.1242/jcs.00371.

Doerks, Tobias, Richard R. Copley, Jörg Schultz, Chris P. Ponting and Peer Bork (Jan. 2002). 'Systematic Identification of Novel Protein Domain Families Associated with Nuclear Functions'. In: *Genome Research* 12.1, pp. 47–56. DOI: 10.1101/gr.203201.

Doma, Meenakshi K. and Roy Parker (Mar. 2006). 'Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation'. In: *Nature* 440.7083, pp. 561–564. DOI: 10.1038/nature04530.

Donovan, J., G. Whitney, S. Rath and A. Korennykh (Mar. 2015). *Crystal structure of human OAS3 domain I in complex with dsRNA*. DOI: 10.2210/pdb4s3n/pdb.

Eckmann, Christian R., Christiane Rammelt and Elmar Wahle (Nov. 2010). 'Control of poly(A) tail length'. In: *Wiley Interdisciplinary Reviews: RNA* 2.3, pp. 348–361. DOI: 10.1002/wrna.56.

Fang, Run, Qifei Jiang, Xiang Zhou, Chenguang Wang, Yukun Guan, Jianli Tao, Jianzhong Xi, Ji-Ming Feng and Zhengfan Jiang (Nov. 2017). 'MAVS activates TBK1 and IKKepsilon through TRAFs in NEMO dependent and independent manner'. In: *PLOS Pathogens* 13.11. Ed. by Pinghui Feng, e1006720. DOI: 10.1371/journal.ppat.1006720.

Filippini, Alice, Daniela Bonini, Edoardo Giacopuzzi, Luca La Via, Fabrizio Gangemi, Marina Colombi and Alessandro Barbon (Feb. 2018). 'Differential Enzymatic Activity of Rat ADAR2 Splicing Variants Is Due to Altered Capability to Interact with RNA in the Deaminase Domain'. In: *Genes* 9.2, p. 79. DOI: 10.3390/genes9020079.

Freund, Emily C., Anne L. Sapiro, Qin Li, Sandra Linder, James J. Moresco, John R. Yates and Jin Billy Li (May 2020). 'Unbiased Identification of trans Regulators of ADAR and A-to-I RNA Editing'. In: *Cell Reports* 31.7, p. 107656. DOI: 10.1016/j.celrep.2020.107656.

Gallo, Angela, Dragana Vukic, David Michalik, Mary A. O'Connell and Liam P. Keegan (Sept. 2017). 'ADAR RNA editing in human disease: more to it than meets the I'. In: *Human Genetics* 136.9, pp. 1265–1278. DOI: 10.1007/s00439-017-1837-0.

Galore-Haskel, Gilli et al. (Aug. 2015). 'A novel immune resistance mechanism of melanoma cells controlled by the ADAR1 enzyme'. In: *Oncotarget* 6.30, pp. 28999–29015. DOI: 10.18632/oncotarget.4905.

Garncarz, Wojciech, Aamira Tariq, Cornelia Handl, Oliver Pusch and Michael F. Jantsch (Feb. 2013). 'A high-throughput screen to identify enhancers of ADAR-mediated RNA-editing'. In: *RNA Biology* 10.2, pp. 192–204. DOI: 10.4161/rna.23208.

Gerber, Andre P. and Walter Keller (Nov. 1999). 'An Adenosine Deaminase that Generates Inosine at the Wobble Position of tRNAs'. In: *Science* 286.5442, pp. 1146–1149. DOI: 10.1126/science.286.5442.1146.

Glover-Cutter, Kira, Soojin Kim, Joaquin Espinosa and David L Bentley (Dec. 2007). 'RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes'. In: *Nature Structural & Molecular Biology* 15.1, pp. 71–78. DOI: 10.1038/nsmb1352.

Goldstein, Boaz, Lily Agranat-Tamir, Dean Light, Orna Ben-Naim Zgayer, Alla Fishman and Ayelet T. Lamm (Dec. 2016). 'A-to-I RNA editing promotes de-

velopmental stage-specific gene and lncRNA expression'. In: *Genome Research* 27.3, pp. 462–470. DOI: 10.1101/gr.211169.116.

Greeve, J, I Altkemper, JH Dieterich, H Greten and E Windler (Aug. 1993). 'Apolipoprotein B mRNA editing in 12 different mammalian species: hepatic expression is reflected in low concentrations of apoB-containing plasma lipoproteins'. In: *Journal of Lipid Research* 34.8, pp. 1367–1383. DOI: 10.1016/s0022-2275(20)36966-2.

Guhaniyogi, Jayita and Gary Brewer (Mar. 2001). 'Regulation of mRNA stability in mammalian cells'. In: *Gene* 265.1-2, pp. 11–23. DOI: 10.1016/s0378-1119(01)00350-x.

Gumna, Julita, Tomasz Zok, Kacper Figurski, Katarzyna Pachulska-Wieczorek and Marta Szachniuk (Oct. 2020). 'RNAthor - fast, accurate normalization, visualization and statistical analysis of RNA probing data resolved by capillary electrophoresis'. In: *PLOS ONE* 15.10. Ed. by Danny Barash, e0239287. DOI: 10.1371/journal.pone.0239287.

Guth, Sabine and Juan Valcárcel (Dec. 2000). 'Kinetic Role for Mammalian SF1/BBP in Spliceosome Assembly and Function after Polypyrimidine Tract Recognition by U2AF'. In: *Journal of Biological Chemistry* 275.48, pp. 38059–38066. DOI: 10.1074/jbc.m001483200.

Hanswillemenke, Alfred, Tahsin Kuzdere, Paul Vogel, Gáspár Jékely and Thorsten Stafforst (Dec. 2015). 'Site-Directed RNA Editing in Vivo Can Be Triggered by the Light-Driven Assembly of an Artificial Riboprotein'. In: *Journal of the American Chemical Society* 137.50, pp. 15875–15881. DOI: 10.1021/jacs.5b10216.

Haque, Nazmul, Ryota Ouda, Chao Chen, Keiko Ozato and J. Robert Hogg (Mar. 2018). 'ZFR coordinates crosstalk between RNA decay and transcription in innate immunity'. In: *Nature Communications* 9.1. DOI: 10.1038/s41467-018-03326-5.

Hartner, Jochen C., Carolin Schmittwolf, Andreas Kispert, Albrecht M. Müller, Miyoko Higuchi and Peter H. Seeburg (Feb. 2004). 'Liver Disintegration in the Mouse Embryo Caused by Deficiency in the RNA-editing Enzyme ADAR1'. In: *Journal of Biological Chemistry* 279.6, pp. 4894–4902. DOI: 10.1074/jbc.m311347200.

Hausner, Georg, Mohamed Hafez and David R Edgell (Mar. 2014). 'Bacterial group I introns: mobile RNA catalysts'. In: *Mobile DNA* 5.1. DOI: 10.1186/1759-8753-5-8.

Heep, Madeleine, Pia Mach, Philipp Reautschnig, Jacqueline Wettengel and Thorsten Stafforst (Jan. 2017). 'Applying Human ADAR1p110 and ADAR1p150 for Site-Directed RNA Editing-G/C Substitution Stabilizes GuideRNAs against Editing'. In: *Genes* 8.1, p. 34. DOI: 10.3390/genes8010034.

Heraud-Farlow, Jacki E. et al. (Sept. 2017). 'Protein recoding by ADAR1-mediated RNA editing is not essential for normal development and homeostasis'. In: *Genome Biology* 18.1. DOI: 10.1186/s13059-017-1301-4.

Herbert, Alan, Jens Alfken, Yang-Gyun Kim, I. Saira Mian, Kazuko Nishikura and Alexander Rich (Aug. 1997). 'A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase'. In: *Proceedings of the National Academy of Sciences* 94.16, pp. 8421–8426. DOI: 10.1073/pnas.94.16.8421.

Hill, Chris H. et al. (Mar. 2019). 'Activation of the Endonuclease that Defines mRNA 3' Ends Requires Incorporation into an 8-Subunit Core Cleavage and Polyadenylation Factor Complex'. In: *Molecular Cell* 73.6, 1217–1231.e11. DOI: 10.1016/j.molcel.2018.12.023.

Hoch, Brigrtte, Rainer M. Maier, Kurt Appel, Gabor L. Igloi and Hans Kössel (Sept. 1991). 'Editing of a chloroplast mRNA by creation of an initiation codon'. In: *Nature* 353.6340, pp. 178–180. DOI: 10.1038/353178a0.

Hong, HuiQi, Jaymie Siqi Lin and Leilei Chen (Mar. 2015). 'Regulatory factors governing adenosine-to-inosine (A-to-I) RNA editing'. In: *Bioscience Reports* 35.2. DOI: 10.1042/bsr20140190.

Horowitz, David S. (Oct. 2011). 'The mechanism of the second step of pre-mRNA splicing'. In: *Wiley Interdisciplinary Reviews: RNA* 3.3, pp. 331–350. DOI: 10.1002/wrna.112.

Hsu, Shu-Ning and Klemens J. Hertel (Nov. 2009). 'Spliceosomes walk the line: Splicing errors and their impact on cellular function'. In: *RNA Biology* 6.5, pp. 526–530. DOI: 10.4161/rna.6.5.9860.

Hug, Nele, Dasa Longman and Javier F. Cáceres (Jan. 2016). 'Mechanism and regulation of the nonsense-mediated decay pathway'. In: *Nucleic Acids Research* 44.4, pp. 1483–1495. DOI: 10.1093/nar/gkw010.

Hunt, Arthur G et al. (2008). 'Arabidopsis mRNA polyadenylation machinery: comprehensive analysis of protein-protein interactions and gene expression profiling'. In: *BMC Genomics* 9.1, p. 220. DOI: 10.1186/1471-2164-9-220.

Ishizuka, Jeffrey J. et al. (Dec. 2018). 'Loss of ADAR1 in tumours overcomes resistance to immune checkpoint blockade'. In: *Nature* 565.7737, pp. 43–48. DOI: 10.1038/s41586-018-0768-9.

Jain, Mamta et al. (Aug. 2018). 'RNA editing of Filamin A pre-mRNA regulates vascular contraction and diastolic blood pressure'. In: *The EMBO Journal* 37.19. DOI: 10.15252/embj.201694813.

Jayachandran, Uma, Heather Grey and Atlanta G. Cook (Dec. 2015). 'Nuclear factor 90 uses an ADAR2-like binding mode to recognize specific bases in dsRNA'. In: *Nucleic Acids Research* 44.4, pp. 1924–1936. DOI: 10.1093/nar/gkv1508.

Johnson, Jason M., John Castle, Philip Garrett-Engele, Zhengyan Kan, Patrick M. Loerch, Christopher D. Armour, Ralph Santos, Eric E. Schadt, Roland Stoughton and Daniel D. Shoemaker (Dec. 2003). 'Genome-Wide Survey of Human Alternative Pre-mRNA Splicing with Exon Junction Microarrays'. In: *Science* 302.5653, pp. 2141–2144. DOI: 10.1126/science.1090100.

Jolma, Arttu et al. (Apr. 2010). 'Multiplexed massively parallel SELEXnfor characterization of human transcription factor binding specificities'. In: *Genome Research* 20.6, pp. 861–873. DOI: 10.1101/gr.100552.109.

Karabiber, Fethullah, Jennifer L. McGinnis, Oleg V. Favorov and Kevin M. Weeks (Nov. 2012). 'QuShape: Rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis'. In: *RNA* 19.1, pp. 63–73. DOI: 10.1261/rna.036327.112.

Kato, Hiroki, Osamu Takeuchi, Eriko Mikamo-Satoh, Reiko Hirai, Tomoji Kawai, Kazufumi Matsushita, Akane Hiiragi, Terence S. Dermody, Takashi Fujita and Shizuo Akira (June 2008). 'Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid-inducible gene-I and melanoma differentiation-associated gene 5'. In: *Journal of Experimental Medicine* 205.7, pp. 1601–1610. DOI: 10.1084/jem.20080091.

Keller, Rebecca W., Uwe Kühn, Mateo Aragón, Larissa Bornikova, Elmar Wahle and David G. Bear (Mar. 2000). 'The nuclear poly(A) binding protein, PABP2, forms an oligomeric particle covering the length of the poly(A) tail'. In: *Journal of Molecular Biology* 297.3, pp. 569–583. DOI: 10.1006/jmbi.2000.3572.

Khosravi, Hamid Mansouri and Michael F. Jantsch (Sept. 2021). 'Site-directed RNA editing: recent advances and open challenges'. In: *RNA Biology* 18.sup1, pp. 41–50. DOI: 10.1080/15476286.2021.1983288.

Kim, Dennis D.Y., Thomas T.Y. Kim, Thomas Walsh, Yoshifumi Kobayashi, Tara C. Matise, Steven Buyske and Abram Gabriel (Sept. 2004). 'Widespread RNA Editing of Embedded Alu Elements in the Human Transcriptome'. In: *Genome Research* 14.9, pp. 1719–1725. DOI: 10.1101/gr.2855504.

Kruger, Kelly, Paula J. Grabowski, Arthur J. Zaug, Julie Sands, Daniel E. Gottschling and Thomas R. Cech (Nov. 1982). 'Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena'. In: *Cell* 31.1, pp. 147–157. DOI: 10.1016/0092-8674(82)90414-7.

Kurihara, Yukio et al. (Feb. 2009). 'Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in Arabidopsis'. In: *Proceedings of the National Academy of Sciences* 106.7, pp. 2453–2458. DOI: 10.1073/pnas.0808902106.

Kurosaki, Tatsuaki, Maximilian W. Popp and Lynne E. Maquat (Apr. 2019). 'Quality and quantity control of gene expression by nonsense-mediated mRNA decay'. In: *Nature Reviews Molecular Cell Biology* 20.7, pp. 406–420. DOI: 10.1038/s41580-019-0126-2.

Lambert, Nicole, Alex Robertson, Mohini Jangi, Sean McGeary, Phillip A. Sharp and Christopher B. Burge (June 2014). 'RNA Bind-n-Seq: Quantitative Assessment of the Sequence and Structural Binding Specificity of RNA Binding Proteins'. In: *Molecular Cell* 54.5, pp. 887–900. DOI: 10.1016/j.molcel.2014.04.016.

Lambert, Nicole J., Alex D. Robertson and Christopher B. Burge (2015). 'RNA Bind-n-Seq'. In: pp. 465–493. DOI: 10.1016/bs.mie.2015.02.007.

Lambowitz, A. M. and S. Zimmerly (May 2010). 'Group II Introns: Mobile Ribozymes that Invade DNA'. In: *Cold Spring Harbor Perspectives in Biology* 3.8, a003616–a003616. DOI: 10.1101/cshperspect.a003616.

Lavigne, Matthieu D., Giannis Vatsellas, Alexander Polyzos, Evangelia Mantouvalou, George Sianidis, Ioannis Maraziotis, Marios Agelopoulos and Dimitris Thanos (May 2015). 'Composite macroH2A/NRF-1 Nucleosomes Suppress Noise and Generate Robustness in Gene Expression'. In: *Cell Reports* 11.7, pp. 1090–1101. DOI: 10.1016/j.celrep.2015.04.022.

Lev-Maor, Galit, Rotem Sorek, Erez Y Levanon, Nurit Paz, Eli Eisenberg and Gil Ast (2007). 'RNA-editing-mediated exon evolution'. In: *Genome Biology* 8.2, R29. DOI: 10.1186/gb-2007-8-2-r29.

Levanon, Keren, Eli Eisenberg, Gideon Rechavi and Erez Y. Levanon (Sept. 2005). 'Letter from the editor: adenosine-to-inosine RNA editing in Alu repeats in the human genome'. In: *EMBO reports* 6.9, pp. 831–835. DOI: 10.1038/sj.embor.7400507.

Liddicoat, Brian J., Robert Piskol, Alistair M. Chalk, Gokul Ramaswami, Miyoko Higuchi, Jochen C. Hartner, Jin Billy Li, Peter H. Seeburg and Carl R. Walkley (Sept. 2015). 'RNA editing by ADAR1 prevents MDA5 sensing of

endogenous dsRNA as nonself'. In: *Science* 349.6252, pp. 1115–1120. DOI: 10.1126/science.aac7049.

López, Marcela Dávila and Tore Samuelsson (Nov. 2007). 'Early evolution of histone mRNA 3' end processing'. In: *RNA* 14.1, pp. 1–10. DOI: 10.1261/rna.782308.

Lorenz, Christian, Christina Lünse and Mario Mörl (Apr. 2017). 'tRNA Modifications: Impact on Structure and Thermal Adaptation'. In: *Biomolecules* 7.4, p. 35. DOI: 10.3390/biom7020035.

Lykke-Andersen, Søren and Torben Heick Jensen (Sept. 2015). 'Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes'. In: *Nature Reviews Molecular Cell Biology* 16.11, pp. 665–677. DOI: 10.1038/nrm4063.

Macbeth, Mark R. and Brenda L. Bass (2007). 'Large-Scale Overexpression and Purification of ADARs from Saccharomyces cerevisiae for Biophysical and Biochemical Studies'. In: pp. 319–331. DOI: 10.1016/s0076-6879(07)24015-7.

Maniatis, Tom and Bosiljka Tasic (July 2002). 'Alternative pre-mRNA splicing and proteome expansion in metazoans'. In: *Nature* 418.6894, pp. 236–243. DOI: 10.1038/418236a.

Mannion, Niamh M. et al. (Nov. 2014). 'The RNA-Editing Enzyme ADAR1 Controls Innate Immune Responses to RNA'. In: *Cell Reports* 9.4, pp. 1482–1494. DOI: 10.1016/j.celrep.2014.10.041.

Matthews, Melissa M, Justin M Thomas, Yuxuan Zheng, Kiet Tran, Kelly J Phelps, Anna I Scott, Jocelyn Havel, Andrew J Fisher and Peter A Beal (Apr. 2016a). 'Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity'. In: *Nature Structural & Molecular Biology* 23.5, pp. 426–433. DOI: 10.1038/nsmb.3203.

– (Apr. 2016b). 'Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity'. In: *Nature Structural Molecular Biology* 23.5, pp. 426–433. DOI: 10.1038/nsmb.3203.

Meagher, Madeleine J. and Robert E. Braun (Apr. 2001). 'Requirement for the Murine Zinc Finger Protein ZFR in Perigastrulation Growth and Survival'. In: *Molecular and Cellular Biology* 21.8, pp. 2880–2890. DOI: 10.1128/mcb.21.8.2880-2890.2001.

Merkle, Tobias, Sarah Merz, Philipp Reautschnig, Andreas Blaha, Qin Li, Paul Vogel, Jacqueline Wettengel, Jin Billy Li and Thorsten Stafforst (Jan. 2019). 'Precise RNA editing by recruiting endogenous ADARs with antisense oli-

gonucleotides'. In: *Nature Biotechnology* 37.2, pp. 133–138. DOI: 10.1038/s41587-019-0013-6.

Mizrahi, Rena A., Dongwon Shin, Renatus W. Sinkeldam, Kelly J. Phelps, Andrea Fin, Dean J. Tantillo, Yitzhak Tor and Peter A. Beal (June 2015). 'A Fluorescent Adenosine Analogue as a Substrate for an A-to-I RNA Editing Enzyme'. In: *Angewandte Chemie International Edition* 54.30, pp. 8713–8716. DOI: 10.1002/anie.201502070.

Modrek, B. (July 2001). 'Genome-wide detection of alternative splicing in expressed sequences of human genes'. In: *Nucleic Acids Research* 29.13, pp. 2850–2859. DOI: 10.1093/nar/29.13.2850.

Morita, Yoko, Toshihiro Shibutani, Nozomi Nakanishi, Kazuko Nishikura, Shigenori Iwai and Isao Kuraoka (Aug. 2013). 'Human endonuclease V is a ribonuclease specific for inosine-containing RNA'. In: *Nature Communications* 4.1. DOI: 10.1038/ncomms3273.

Moteki, Shin and David Price (Sept. 2002). 'Functional Coupling of Capping and Transcription of mRNA'. In: *Molecular Cell* 10.3, pp. 599–609. DOI: 10.1016/s1097-2765(02)00660-3.

Nakadai, Tomoyoshi, Aya Fukuda, Miho Shimada, Ken Nishimura and Koji Hisatake (Oct. 2015). 'The RNA Binding Complexes NF45-NF90 and NF45-NF110 Associate Dynamically with the c-fos Gene and Function as Transcriptional Coactivators'. In: *Journal of Biological Chemistry* 290.44, pp. 26832–26845. DOI: 10.1074/jbc.m115.688317.

Nakahama, Taisuke and Yukio Kawahara (Jan. 2020). 'Adenosine-to-inosine RNA editing in the immune system: friend or foe?' In: *Cellular and Molecular Life Sciences* 77.15, pp. 2931–2948. DOI: 10.1007/s00018-020-03466-2.

NEB (2022). URL: https://international.neb.com/tools-and-resources/selection-charts/cleavage-of-supercoiled-dna.

Ng, Bernard, Fan Yang, David P. Huston, Yan Yan, Yu Yang, Zeyu Xiong, Leif E. Peterson, Hong Wang and Xiao-Feng Yang (Dec. 2004). 'Increased noncanonical splicing of autoantigen transcripts provides the structural basis for expression of untolerized epitopes'. In: *Journal of Allergy and Clinical Immunology* 114.6, pp. 1463–1470. DOI: 10.1016/j.jaci.2004.09.006.

Nishikura, Kazuko (June 2010). 'Functions and Regulation of RNA Editing by ADAR Deaminases'. In: *Annual Review of Biochemistry* 79.1, pp. 321–349. DOI: 10.1146/annurev-biochem-060208-105251.

Nourreddine, Sami et al. (May 2020). 'NF45 and NF90 Regulate Mitotic Gene Expression by Competing with Staufen-Mediated mRNA Decay'. In: *Cell Reports* 31.7, p. 107660. DOI: 10.1016/j.celrep.2020.107660.

O'Connell, Mary A., André Gerber and Walter Keller (Jan. 1997). 'Purification of Human Double-stranded RNA-specific Editase 1 (hRED1) Involved in Editing of Brain Glutamate Receptor B Pre-mRNA'. In: *Journal of Biological Chemistry* 272.1, pp. 473–478. DOI: 10.1074/jbc.272.1.473.

Palavicini, Juan Pablo, Mary A. O'connell and Joshua J.C. Rosenthal (Apr. 2009). 'An extra double-stranded RNA binding domain confers high activity to a squid RNA editing enzyme'. In: *RNA* 15.6, pp. 1208–1218. DOI: 10.1261/rna.1471209.

Park, SeHee, Erin E. Doherty, Yixuan Xie, Anil K. Padyana, Fang Fang, Yue Zhang, Agya Karki, Carlito B. Lebrilla, Justin B. Siegel and Peter A. Beal (Oct. 2020). 'High-throughput mutagenesis reveals unique structural features of human ADAR1'. In: *Nature Communications* 11.1. DOI: 10.1038/s41467-020-18862-2.

Parker, Lisa M., Ivo Fierro-Monti and Michael B. Mathews (Aug. 2001). 'Nuclear Factor 90 Is a Substrate and Regulator of the Eukaryotic Initiation Factor 2 Kinase Double-stranded RNA-activated Protein Kinase'. In: *Journal of Biological Chemistry* 276.35, pp. 32522–32530. DOI: 10.1074/jbc.m104408200.

Pawlicka, Kamila, Umesh Kalathiya and Javier Alfaro (Mar. 2020). 'Nonsense-Mediated mRNA Decay: Pathologies and the Potential for Novel Therapeutics'. In: *Cancers* 12.3, p. 765. DOI: 10.3390/cancers12030765.

Paz-Yaacov, Nurit, Lily Bazak, Ilana Buchumenski, Hagit T. Porath, Miri Danan-Gotthold, Binyamin A. Knisbacher, Eli Eisenberg and Erez Y. Levanon (Oct. 2015). 'Elevated RNA Editing Activity Is a Major Contributor to Transcriptomic Diversity in Tumors'. In: *Cell Reports* 13.2, pp. 267–276. DOI: 10.1016/j.celrep.2015.08.080.

Peisley, Alys, Bin Wu, Hui Xu, Zhijian J. Chen and Sun Hur (Mar. 2014). 'Structural basis for ubiquitin-mediated antiviral signal activation by RIG-I'. In: *Nature* 509.7498, pp. 110–114. DOI: 10.1038/nature13140.

Pestal, Kathleen, Cory C. Funk, Jessica M. Snyder, Nathan D. Price, Piper M. Treuting and Daniel B. Stetson (Nov. 2015). 'Isoforms of RNA-Editing Enzyme ADAR1 Independently Control Nucleic Acid Sensor MDA5-Driven Autoimmunity and Multi-organ Development'. In: *Immunity* 43.5, pp. 933–944. DOI: 10.1016/j.immuni.2015.11.001.

Poulsen, Hanne, Jakob Nilsson, Christian K. Damgaard, Jan Egebjerg and Jørgen Kjems (Nov. 2001). 'CRM1 Mediates the Export of ADAR1 through a Nuclear Export Signal within the Z-DNA Binding Domain'. In: *Molecular and Cellular Biology* 21.22, pp. 7862–7871. DOI: 10.1128/mcb.21.22.7862-7871.2001.

Powers, Kyle T, Jenn-Yeu Alvin Szeto and Christiane Schaffitzel (Dec. 2020). 'New insights into no-go, non-stop and nonsense-mediated mRNA decay complexes'. In: *Current Opinion in Structural Biology* 65, pp. 110–118. DOI: 10.1016/j.sbi.2020.06.011.

Price, David H. and Michael W. Gray (Apr. 2014). 'Editing of tRNA'. In: *Modification and Editing of RNA*. ASM Press, pp. 289–305. DOI: 10.1128/9781555818296.ch16.

Qu, Liang et al. (July 2019). 'Programmable RNA editing by recruiting endogenous ADAR using engineered RNAs'. In: *Nature Biotechnology* 37.9, pp. 1059–1069. DOI: 10.1038/s41587-019-0178-z.

Ramanathan, Anand, G. Brett Robb and Siu-Hong Chan (June 2016). 'mRNA capping: biological functions and applications'. In: *Nucleic Acids Research* 44.16, pp. 7511–7526. DOI: 10.1093/nar/gkw551.

Ramaswami, Gokul and Jin Billy Li (Oct. 2013). 'RADAR: a rigorously annotated database of A-to-I RNA editing'. In: *Nucleic Acids Research* 42.D1, pp. D109–D113. DOI: 10.1093/nar/gkt996.

Ramirez-Moya, Julia, Allison R. Baker, Frank J. Slack and Pilar Santisteban (Mar. 2020). 'ADAR1-mediated RNA editing is a novel oncogenic process in thyroid cancer and regulates miR-200 activity'. In: *Oncogene* 39.18, pp. 3738–3753. DOI: 10.1038/s41388-020-1248-x.

Rawling, David C and Anna Marie Pyle (Apr. 2014). 'Parts, assembly and operation of the RIG-I family of motors'. In: *Current Opinion in Structural Biology* 25, pp. 25–33. DOI: 10.1016/j.sbi.2013.11.011.

Read, Laurie K., Julius Lukeš and Hassan Hashimi (Nov. 2015). 'Trypanosome RNA editing: the complexity of getting U in and taking U out'. In: *WIREs RNA* 7.1, pp. 33–51. DOI: 10.1002/wrna.1313.

Rebagliati, M.R. and D.A. Melton (Feb. 1987). 'Antisense RNA injections in fertilized frog eggs reveal an RNA duplex unwinding activity'. In: *Cell* 48.4, pp. 599–605. DOI: 10.1016/0092-8674(87)90238-8.

Refolo, Giulia, Tiziana Vescovo, Mauro Piacentini, Gian Maria Fimia and Fabiola Ciccosanti (Feb. 2020). 'Mitochondrial Interactome: A Focus on Anti-

viral Signaling Pathways'. In: *Frontiers in Cell and Developmental Biology* 8. DOI: 10.3389/fcell.2020.00008.

Rice, Gillian I et al. (Sept. 2012). 'Mutations in ADAR1 cause Aicardi-Goutieres syndrome associated with a type I interferon signature'. In: *Nature Genetics* 44.11, pp. 1243–1248. DOI: 10.1038/ng.2414.

Ring, Gillian M., Mary A. O'Connell and Liam P. Keegan (2004). 'Purification and Assay of Recombinant ADAR Proteins Expressed in the Yeast Pichia pastoris or in Escherichia coli'. In: pp. 219–238. DOI: 10.1385/1-59259-775-0:219.

Roberts, Steven A et al. (July 2013). 'An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers'. In: *Nature Genetics* 45.9, pp. 970–976. DOI: 10.1038/ng.2702.

Salter, Jason D., Ryan P. Bennett and Harold C. Smith (July 2016). 'The APOBEC Protein Family: United by Structure, Divergent in Function'. In: *Trends in Biochemical Sciences* 41.7, pp. 578–594. DOI: 10.1016/j.tibs.2016.05.001.

Sapiro, Anne L., Emily C. Freund, Lucas Restrepo, Huan-Huan Qiao, Amruta Bhate, Qin Li, Jian-Quan Ni, Timothy J. Mosca and Jin Billy Li (May 2020). 'Zinc Finger RNA-Binding Protein Zn72D Regulates ADAR-Mediated RNA Editing in Neurons'. In: *Cell Reports* 31.7, p. 107654. DOI: 10.1016/j.celrep.2020.107654.

Sasaki, Tadamasa, Yasushi Yukawa, Tatsuya Wakasugi, Kyoji Yamada and Masahiro Sugiura (July 2006). 'A simple in vitro RNA editing assay for chloroplast transcripts using fluorescent dideoxynucleotides: distinct types of sequence elements required for editing of ndh transcripts'. In: *The Plant Journal* 47.5, pp. 802–810. DOI: 10.1111/j.1365-313x.2006.02825.x.

Satoh, Takashi, Hiroki Kato, Yutaro Kumagai, Mitsutoshi Yoneyama, Shintaro Sato, Kazufumi Matsushita, Tohru Tsujimura, Takashi Fujita, Shizuo Akira and Osamu Takeuchi (Jan. 2010). 'LGP2 is a positive regulator of RIG-I - and MDA5-mediated antiviral responses'. In: *Proceedings of the National Academy of Sciences* 107.4, pp. 1512–1517. DOI: 10.1073/pnas.0912986107.

Schumacher, Jill M., Karen Artzt and Robert E. Braun (July 1998). 'Spermatid Perinuclear Ribonucleic Acid-Binding Protein Binds Microtubules In Vitro and Associates with Abnormal Manchettes In Vivo in Mice1'. In: *Biology of Reproduction* 59.1, pp. 69–76. DOI: 10.1095/biolreprod59.1.69.

Shatkin, Aaron J. and James L. Manley (Oct. 2000). 'The ends of the affair: capping and polyadenylation'. In: *Nature Structural Biology* 7.10, pp. 838–842. DOI: 10.1038/79583.

Shi, Yigong (Sept. 2017). 'Mechanistic insights into precursor messenger RNA splicing by the spliceosome'. In: *Nature Reviews Molecular Cell Biology* 18.11, pp. 655–670. DOI: 10.1038/nrm.2017.86.

Shuman, Stewart (2000). 'Structure, mechanism, and evolution of the mRNA capping apparatus'. In: *Progress in Nucleic Acid Research and Molecular Biology*. Elsevier, pp. 1–40. DOI: 10.1016/s0079-6603(00)66025-7.

Silberberg, Gilad, Daniel Lundin, Ruth Navon and Marie Öhman (Oct. 2011). 'Deregulation of the A-to-I RNA editing mechanism in psychiatric disorders'. In: *Human Molecular Genetics* 21.2, pp. 311–321. DOI: 10.1093/hmg/ddr461.

Slotkin, William and Kazuko Nishikura (2013). 'Adenosine-to-inosine RNA editing and human disease'. In: *Genome Medicine* 5.11, p. 105. DOI: 10.1186/gm508.

Smith, Harold C., Joseph E. Wedekind, Kefang Xie and Mark P. Sowden (Dec. 2004). 'Mammalian C to U editing'. In: pp. 365–400. DOI: 10.1007/b105432.

Smith, Jenna E., Juan R. Alvarez-Dominguez, Nicholas Kline, Nathan J. Huynh, Sarah Geisler, Wenqian Hu, Jeff Coller and Kristian E. Baker (June 2014). 'Translation of Small Open Reading Frames within Unannotated RNA Transcripts in Saccharomyces cerevisiae'. In: *Cell Reports* 7.6, pp. 1858–1866. DOI: 10.1016/j.celrep.2014.05.023.

Sonenberg, Nahum and Anne-Claude Gingras (Apr. 1998). 'The mRNA 5' cap-binding protein eIF4E and control of cell growth'. In: *Current Opinion in Cell Biology* 10.2, pp. 268–275. DOI: 10.1016/s0955-0674(98)80150-6.

Song, Yulong, Wenbing Yang, Qiang Fu, Liang Wu, Xueni Zhao, Yusen Zhang and Rui Zhang (Mar. 2020). 'irCLASH reveals RNA substrates recognized by human ADARs'. In: *Nature Structural - Molecular Biology* 27.4, pp. 351–362. DOI: 10.1038/s41594-020-0398-4.

Stewart, Murray (Mar. 2019). 'Polyadenylation and nuclear export of mRNAs'. In: *Journal of Biological Chemistry* 294.9, pp. 2977–2987. DOI: 10.1074/jbc.rev118.005594.

Sugimoto, Yoichiro, Julian König, Shobbir Hussain, Blaž Zupan, Tomaž Curk, Michaela Frye and Jernej Ule (2012). 'Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions'. In: *Genome Biology* 13.8, R67. DOI: 10.1186/gb-2012-13-8-r67.

Takenaka, Mizuki, Daniil Verbitskiy, Anja Zehrmann, Barbara Härtel, Eszter Bayer-Császár, Franziska Glass and Axel Brennicke (Nov. 2014). 'RNA editing in plant mitochondria - Connecting RNA target sequences and acting

proteins'. In: *Mitochondrion* 19, pp. 191–197. DOI: 10.1016/j.mito.2014.04.005.

Takenaka, Mizuki, Anja Zehrmann, Daniil Verbitskiy, Barbara Härtel and Axel Brennicke (Nov. 2013). 'RNA Editing in Plants and Its Evolution'. In: *Annual Review of Genetics* 47.1, pp. 335–352. DOI: 10.1146/annurev-genet-111212-133519.

Tisdale, S. and L. Pellizzoni (June 2015). 'Disease Mechanisms and Therapeutic Approaches in Spinal Muscular Atrophy'. In: *Journal of Neuroscience* 35.23, pp. 8691–8700. DOI: 10.1523/jneurosci.0417-15.2015.

Tonkin, Leath A., Lisa Saccomanno, Daniel P. Morse, Thomas Brodigan, Michael Krause and Brenda L. Bass (Nov. 2002). 'RNA editing by ADARs is important for normal behavior in Caenorhabditis elegans'. In: *The EMBO Journal* 21.22, pp. 6025–6035. DOI: 10.1093/emboj/cdf607.

Torres, Adrian Gabriel, David Piñeyro, Liudmila Filonava, Travis H. Stracker, Eduard Batlle and Lluis Ribas de Pouplana (Sept. 2014). 'A-to-I editing on tRNAs: Biochemical, biological and evolutionary implications'. In: *FEBS Letters* 588.23, pp. 4279–4286. DOI: 10.1016/j.febslet.2014.09.025.

Tsutsumi, Satoshi, Reiko Sugiura, Yan Ma, Hideki Tokuoka, Kazuki Ohta, Rieko Ohte, Akiko Noma, Tsutomu Suzuki and Takayoshi Kuno (Nov. 2007). 'Wobble Inosine tRNA Modification Is Essential to Cell Cycle Progression in G1/S and G2/M Transitions in Fission Yeast'. In: *Journal of Biological Chemistry* 282.46, pp. 33459–33465. DOI: 10.1074/jbc.m706869200.

Uhlén, Mathias et al. (Jan. 2015). 'Tissue-based map of the human proteome'. In: *Science* 347.6220. DOI: 10.1126/science.1260419.

Valadkhan, Saba (May 2010). 'Role of the snRNAs in spliceosomal active site'. In: *RNA Biology* 7.3, pp. 345–353. DOI: 10.4161/rna.7.3.12089.

Vazquez, Christine and Stacy M. Horner (July 2015). 'MAVS Coordination of Antiviral Innate Immunity'. In: *Journal of Virology* 89.14. Ed. by C. S. Sullivan, pp. 6974–6977. DOI: 10.1128/jvi.01918-14.

ViennaRNA (2022). URL: http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi.

Vik, Erik Sebastian, Meh Sameen Nawaz, Pernille Strøm Andersen, Cathrine Fladeby, Magnar Bjørås, Bjørn Dalhus and Ingrun Alseth (Aug. 2013). 'Endonuclease V cleaves at inosines in RNA'. In: *Nature Communications* 4.1. DOI: 10.1038/ncomms3271.

Wan, Ruixue, Rui Bai, Chuangye Yan, Jianlin Lei and Yigong Shi (Apr. 2019). 'Structures of the Catalytically Activated Yeast Spliceosome Reveal the Mech-

anism of Branching'. In: *Cell* 177.2, 339–351.e13. DOI: 10.1016/j.cell.2019.02.006.

Wang, Eric T., Rickard Sandberg, Shujun Luo, Irina Khrebtukova, Lu Zhang, Christine Mayr, Stephen F. Kingsmore, Gary P. Schroth and Christopher B. Burge (Nov. 2008). 'Alternative isoform regulation in human tissue transcriptomes'. In: *Nature* 456.7221, pp. 470–476. DOI: 10.1038/nature07509.

Wang, Manni and Xiawei Wei (Mar. 2019). 'Deletion of the RNA-editing enzyme ADAR1A: new strategy to potentiate responses to PD-1 immune checkpoint blockade'. In: *Signal Transduction and Targeted Therapy* 4.1. DOI: 10.1038/s41392-019-0039-8.

Wang, Qingde, Mana Miyakoda, Weidong Yang, Jaspal Khillan, David L. Stachura, Mitchell J. Weiss and Kazuko Nishikura (Feb. 2004). 'Stress-induced Apoptosis Associated with Null Mutation of ADAR1 RNA Editing Deaminase Gene'. In: *Journal of Biological Chemistry* 279.6, pp. 4952–4961. DOI: 10.1074/jbc.m310162200.

Weber, Friedemann, Valentina Wagner, Simon B. Rasmussen, Rune Hartmann and Soren R. Paludan (May 2006). 'Double-Stranded RNA Is Produced by Positive-Strand RNA Viruses and DNA Viruses but Not in Detectable Amounts by Negative-Strand RNA Viruses'. In: *Journal of Virology* 80.10, pp. 5059–5064. DOI: 10.1128/jvi.80.10.5059-5064.2006.

Wedekind, Joseph E., Geoffrey S.C. Dance, Mark.P. Sowden and Harold C. Smith (Apr. 2003). 'Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business'. In: *Trends in Genetics* 19.4, pp. 207–216. DOI: 10.1016/s0168-9525(03)00054-4.

Weinberg, Christina E, V Janett Olzog, Iris Eckert and Zasha Weinberg (June 2021). 'Identification of over 200-fold more hairpin ribozymes than previously known in diverse circular RNAs'. In: *Nucleic Acids Research* 49.11, pp. 6375–6388. DOI: 10.1093/nar/gkab454.

Wilkinson, Max E., Clément Charenton and Kiyoshi Nagai (June 2020). 'RNA Splicing by the Spliceosome'. In: *Annual Review of Biochemistry* 89.1, pp. 359–388. DOI: 10.1146/annurev-biochem-091719-064225.

Wilkinson, Max E., Sebastian M. Fica, Wojciech P. Galej and Kiyoshi Nagai (Apr. 2021). 'Structural basis for conformational equilibrium of the catalytic spliceosome'. In: *Molecular Cell* 81.7, 1439–1452.e9. DOI: 10.1016/j.molcel.2021.02.021.

Will, C. L. and R. Luhrmann (Dec. 2010). 'Spliceosome Structure and Function'. In: *Cold Spring Harbor Perspectives in Biology* 3.7, a003707–a003707. DOI: 10.1101/cshperspect.a003707.

Wolf, Jeannette, Andre P. Gerber and Walter Keller (July 2002). 'tadA, an essential tRNA-specific adenosine deaminase from Escherichia coli'. In: *The EMBO Journal* 21.14, pp. 3841–3851. DOI: 10.1093/emboj/cdf362.

Wolkowicz, Urszula M. and Atlanta G. Cook (July 2012). 'NF45 dimerizes with NF90, Zfr and SPNR via a conserved domain that has a nucleotidyltransferase fold'. In: *Nucleic Acids Research* 40.18, pp. 9356–9368. DOI: 10.1093/nar/gks696.

Wright, Amanda and Bryce Vissel (2012). 'The essential role of AMPA receptor GluR2 subunit RNA editing in the normal and diseased brain'. In: *Frontiers in Molecular Neuroscience* 5. DOI: 10.3389/fnmol.2012.00034.

Yablonovitch, Arielle L., Patricia Deng, Dionna Jacobson and Jin Billy Li (Nov. 2017). 'The evolution and adaptation of A-to-I RNA editing'. In: *PLOS Genetics* 13.11. Ed. by Jianzhi Zhang, e1007064. DOI: 10.1371/journal.pgen.1007064.

Zhang, Xiaofeng, Xiechao Zhan, Chuangye Yan, Wenyu Zhang, Dongliang Liu, Jianlin Lei and Yigong Shi (Feb. 2019). 'Structures of the human spliceosomes before and after release of the ligated exon'. In: *Cell Research* 29.4, pp. 274–285. DOI: 10.1038/s41422-019-0143-x.

Zhao, Han-Qing, Pan Zhang, Hua Gao, Xiandong He, Yanmei Dou, August Y. Huang, Xi-Ming Liu, Adam Y. Ye, Meng-Qiu Dong and Liping Wei (Nov. 2014). 'Profiling the RNA editomes of wild-type C. elegans and ADAR mutants'. In: *Genome Research* 25.1, pp. 66–75. DOI: 10.1101/gr.176107.114.

Zhao, Jing, Linda Hyman and Claire Moore (June 1999). 'Formation of mRNA 3' Ends in Eukaryotes: Mechanism, Regulation, and Interrelationships with Other Steps in mRNA Synthesis'. In: *Microbiology and Molecular Biology Reviews* 63.2, pp. 405–445. DOI: 10.1128/mmbr.63.2.405-445.1999.

Zykovich, Artem, Ian Korf and David J. Segal (Oct. 2009). 'Bind-n-Seq: high-throughput analysis of in vitro protein-DNA interactions using massively parallel sequencing'. In: *Nucleic Acids Research* 37.22, e151–e151. DOI: 10.1093/nar/gkp802.