Development of intronless expression systems for biomanufacturing

Joshua Zilate

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

The University of Sheffield
Faculty of Science
School of Biosciences

November 2024

Acknowledgments

Firstly, I would like to thank my supervisors: Prof Stuart Wilson, Dr Emma Thomson, Dr Ian Sudbery and Dr Sarah Dunn for their continued guidance and support throughout this project. I'd like to thank all current and previous members of the Wilson lab: Vicky, Ivo, Carmen, Ang, Chiara, Pete, Mikayla, Jack, Charlotte, Justin, and Daniel, who together made working in the Wilson lab a continuously happy place and provided well needed breaks and laughter. Special thanks to Ang for the unending supply of snacks. I am grateful to have met a large group of people across the E floor corridor from the Bose, Thomson and Mitchel labs who have always been lovely to stop and chat with and have made working in Sheffield such a pleasant experience. I'd like to thank my friends Craig and Sam, who were always available for a chat and good walk in the peaks when a break was needed. I'd like to give a heartfelt thanks to my family, especially to my parents, grandparents and uncles who have been a continuous source of support and encouragement throughout this project. Finally, I would like to thank my partner Jess, who has been a constant source of support and motivation. You've never failed to make me smile and made every day of my Ph.D. special.

Abstract

Biotherapeutic molecules, such as monoclonal antibodies (mAbs), represent a sizeable proportion of newly approved therapeutic's available for treatment of disease. Reducing the production costs of these therapeutics, usually produced in massive quantities from Chinese Hamster Ovary (CHO) cells, remains a key focus of the pharmaceutical industry. The presence of introns in the expression cassettes of mAbs are an effective way of achieving high therapeutic protein expression levels. However, frequent mis-splicing events lead to the formation of aberrant protein structures, lowering the yield of the therapeutic protein and requiring filtration from the final product. This leads to increased production costs resulting in increased cost for the consumer and thus lowering the availability of the drug.

To counteract this problem, in this project we set out to develop expression systems with increased therapeutic protein yield from an intronless expression cassette. We demonstrate that improving the GC content of an intronless sequence improves protein expression through improved mRNA export via interactions with the mRNA export factor ALYREF. We go on to show that inclusion of RNA-binding protein (RBPs) recruitment motifs for SRSF1/SRSF7 and SF3b1 into the intronless sequence improves therapeutic protein expression, showing that mRNA processing is a good expression stage to target to in this process. We further demonstrate the potential for both GC content and RBP recruitment motifs to work in tandem to improve therapeutic protein expression and provide evidence that these improvements work in an industrial setting.

Finally, we investigate the role of the HUSH complex in mediated silencing of intronless expression sequences, showing that this HUSH complex may not play a key role in reducing protein expression levels from intronless sequences. Additionally, we show the potential of HNRNPU silencing as a method of improving therapeutic production, but further development of this system is required to alleviate certain drawbacks associated with HNRNPU silencing.

Contents

List of figures

Abbreviations

3′ SS	3′ splice site
5′ SS	5′ splice site
m^7G	7-methylguanosine
ALS	Amyotrophic lateral sclerosis
BP	Branchpoint adenosine
CBC	Binding Complex
CM	Cell Media
CHO	Chinese Hamster Ovary
CDS	Coding Sequence
Cryo-EM	Cryo-electron microscopy
CAR	Cytoplasmic localisation regions
EJC	Exon Junction Complex
ESS	Exonic silencer sequences
ESE	Exonic splicing enhancers
Fc-FP	Fc- fusion protein
hnRNP	heterogeneous nuclear ribonucleoprotein
HGC	High GC
HPLC	High Performance Liquid Chromatography
ISE	Intron splicing enhancers
lncRNAs	Long noncoding RNAs
LGC	Low GC
3′ splice site mRNA	Messenger RNA
MPP8	M-phase phosphoprotein 8
NTC	NineTeen Complex
NTR	NTC related complex
PABPN1	Nuclear poly(A) binding protein
NPC	Nuclear pore complex
PAP	Poly(A)polymerase
PAS	Polyadenylation signal
CPSF	Polyadenylation specificity factor
qRT-PCR	Quantitative reverse transcriptase-polymerase chain reaction
RNPs	Ribonucleoprotein particles
RBD	RNA binding domain
RIP	RNA immunoprecipitation
Pol II	RNA polymerase II
RRM	RNA recognition motif
PAGE	Polyacrylamide gel electrophoresis
snRNAs	Small nuclear RNAs
SMA	Spinal muscular atrophy
SMN1	Survival of motor neurons 1
SMN2	Survival of motor neurons 2
TDP-43	TAR DNA binding protein 43
TREX	TRanscription and EXport
TASOR	Transcription activation suppressor
TEs	Transposable elements
WT	Wild Type

Chapter 1: Introduction

1.1 Co-transcriptional processing of mRNA

The central dogma of molecular biology states that DNA is converted to RNA which becomes protein, but this dogma fails to capture the complexity of the myriad of RNA processing steps. The transcription of protein coding genes in eukaryotes is performed by RNA polymerase II (pol II).

Upon transcription, the generated pre-mRNA immediately undergoes co-transcriptional processing, with 5′ capping of mRNA occurring once the nascent transcript reaches 25nt – 30nt in length (Moteki and Price. 2002). The common cap structure is a 7-methylguanosine (m^7G) that attaches to the first 5′ nucleotide of the transcript via a triphosphate bridge (Furuichi, Y. 2015). The cap structure is generated through the sequential activities of 3 enzymes: an RNA triphosphatase, a guanylyl transferase and a guanine-N7 methyltransferase (Ramanathan et al. 2016). The addition of the m^7G cap acts immediately to protect the pre-mRNA from degradation by 5′ - 3′ exonucleases and acts as a landing pad for a variety of RNA binding proteins through which regulatory roles of the 5′ cap are conducted (Gonatopoulos-Pournatzis and Cowling. 2014).

The m^7G is bound by the Cap Binding Complex (CBC), consisting of CBP20 and CBP80, originally identified based of their affinity for m^7G and shown to co-precipitate from HeLa cells, with both proteins required in complex to bind the 5′ cap (Izaurralde et al. 1994). The CBC effects pre-mRNA processing by binding a variety of pre-mRNA processing factors involved in multiple stages in mRNA biogenesis. The CBC helps recruit splicing factors, specifically the U4/U6.U5 snRNP complex, with knockdown of the CBC in HeLa cells causing inefficient spliceosome assembly on a reporter pre-mRNA (Pabis et al. 2013). Furthermore, the CBC has been identified as a key step in RNA export through protein-protein interactions between ALYREF and UAP56, key members of the TRanscription and EXport (TREX) complex, and CBP80 (Cheng et al. 2006). The recruitment of ALYREF to the CBC is a key fate determination event for mRNA, where competition between ALYREF and MTR4 decides if a mRNA is destined for nuclear export or degradation (Fan et al. 2017).

Whilst capping occurs at the 5′ end of an mRNA immediately, non-coding sections of DNA, called introns, are removed co-transcriptionally by the spliceosome. Removal of

introns by the spliceosome acts as a key regulatory event in the lifetime of an mRNA, with a vast network of proteins interacting with a pre-mRNA because of intron removal and spliceosome associations.

At the 3′ end of eukaryotic mRNA, a poly(A) tail is added co-transcriptionally and acts as an additional regulator of mRNA behaviour and stability, adding another layer of complexity to the transcriptome. Polyadenylation first begins with recognition and selection of a polyadenylation signal (PAS), a hexamer sequence typically consisting of AAUAAA, 25-30nt upstream from the 3′ end of an mRNA. The PAS is bound by the cleavage and polyadenylation specificity factor (CPSF) complex which is responsible for the endonucleolytic cleavage of the nascent RNA and the recruitment of poly(A)polymerase (PAP) to the 3’end of the pre-mRNA to generate the poly(A) tail (Eckmann et al. 2011). Following mRNA cleavage and transcription termination, the growing tail is bound by nuclear poly(A) binding protein (PABPN1) to control the length of the tail and facilitate termination of tail synthesis by PAP. PABPN1 binding along the poly(A) tail holds CPSF in place at the PAS, maintains the interaction between CPSF/PAP and interacts directly with PAP to stabilise the polyadenylation complex. Following addition of around 250 nucleotides, the complex becomes too large for PABPN1 to stabilise the interaction between CPSF and PAP, causing a slowdown and loss of polyadenylation, terminating the process (Kühn et al. 2009, Eckmann et al. 2011).

The poly(A) tail of an mRNA acts as a key hallmark of correct co-transcriptional processing and has a variety of post transcriptional regulatory functions. The poly(A) tail protects the mRNA from degradation at the 3′ end by the exosome. Over an mRNAs lifespan the poly(A) tail is gradually shortened by deadenylases, such as CCR4-NOT, until the tail reaches around 10-12nt at which point the 3′ end of the mRNA is accessible to the exosome (Garneau et al. 2007). Therefore, the length of the poly(A)tail not only protects an mRNA from degradation, but also contributes to it’s half-life (Passmore and Coller. 2022). Furthermore, the poly(A) tail contributes to the initiation of translation through the protein PABPC1, the cytoplasmic version of PABPN1, and through interactions with the 5′ cap. A long-standing observation is the binding between PABPC1 and the translation initiation factor eIF-4G, which is recruited to the mRNP by eIF-4A binding to the 5′ cap (Tarun and Sachs. 1996, Passmore and

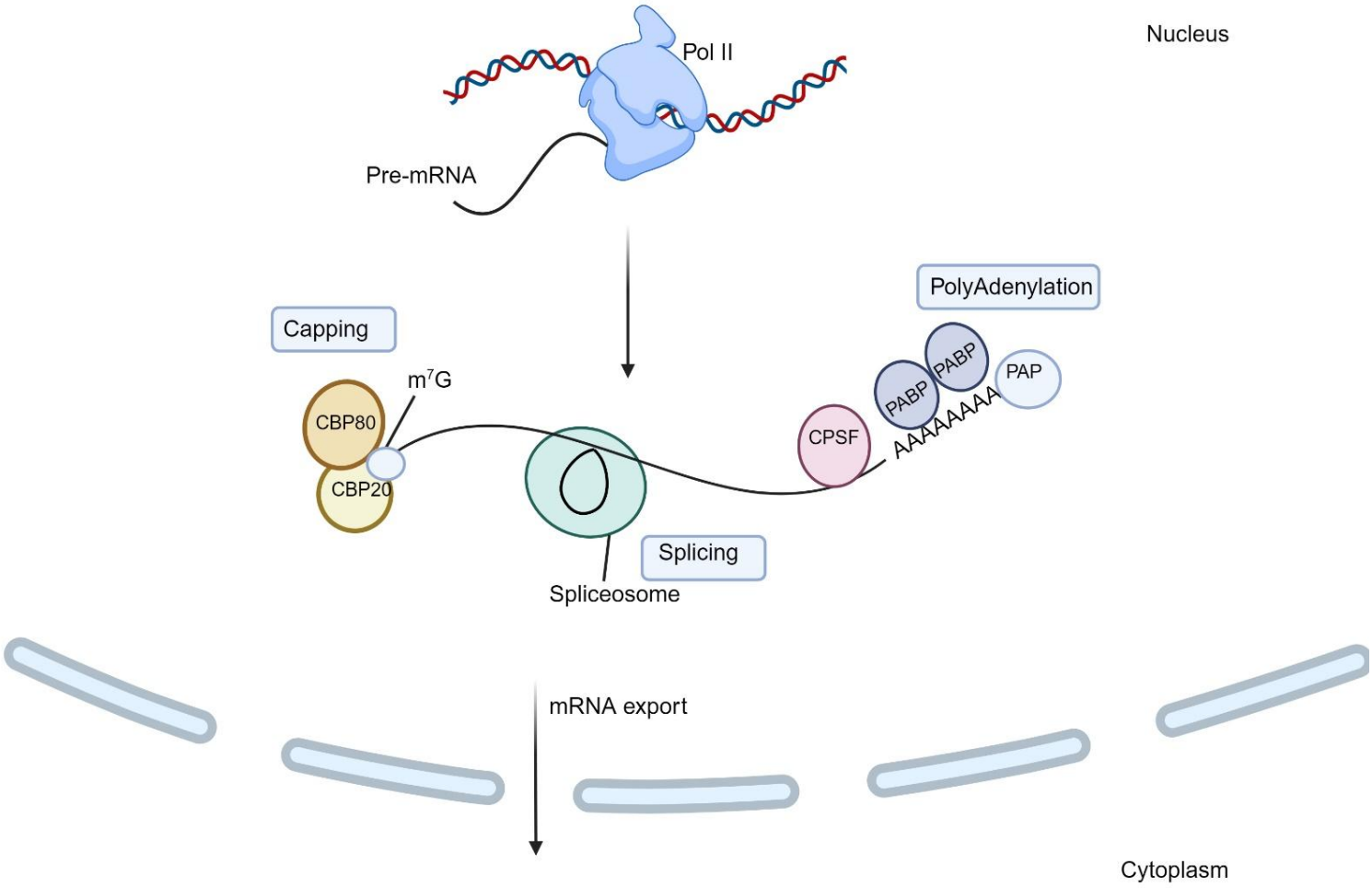


Figure 1.1: Co-transcriptional processing of mRNA.

During transcription of a gene to produce mRNA by RNA polymerase II, the nascent pre-mRNA undergoes co-transcriptional processing. This involves 5′ capping of the pre-mRNA with a m^7G cap, removal of introns via the spliceosome and addition of a poly(A) tail at the 3′ end of the mRNA. Together, these processes signal the correct formation of a mRNA for its export to the cytoplasm.

Coller. 2022). This has led to a model in which the mRNA forms a loop structure during translation initiation, in which the 5′ cap and poly(A)tail fold together to allow binding of the translation initiation complex to the mRNA (Biziaev et al. 2024).

1.2. Introns

The presence of introns has been reported for several decades, with the first observations of exons being spliced together observed in viral DNA (Berget et al. 1977) but the full extent of intron presence in mammals was revealed by the sequencing of the human genome (Lander et al. 2001). How introns evolved is still unclear, but what has been elucidated is the immense impact that they have on gene expression (Irimia and Roy. 2014). Despite the initial view that introns, as non-coding sections of DNA, were simply ‘junk’ a variety of role for introns in gene expression have been discovered (Shaul, O. 2017). This is also supported by the long-standing observation that expression from a spliced gene is better than that of its cDNA counterpart, demonstrating that introns do form a vital part of eukaryotic gene expression. While naturally occurring intronless genes do exist, their expression is in general lower than intron containing genes, with some notable exceptions including histone mRNAs (Shabalina et al. 2010).

1.2.1 Spliceosome components

Removal of intronic sequences from the pre-mRNA is performed by the spliceosome, a large multiunit complex that goes through stages of dynamic remodelling to splice introns. The most common eukaryotic spliceosome consists of the U2 type spliceosome, composed of a core of the small nuclear RNAs (snRNAs) U1, U2, U4, U5 and U6 in Ribonucleoprotein particles (RNPs) with a broad range of proteins associated at distinct stages of splicing. Additionally, most eukaryotes possess the U12 type spliceosome, consisting of the snRNAs: U11, U12, U4atac, U5 and U6atatc. While U5 is a common feature of both spliceosomes, U11, U12, U4atac and U6atac are distinct in their structure but identical in function to their counterparts in the U2 spliceosome (Patel and Steitz. 2003). Which spliceosome is recruited to the splice site is sequence dependent, with the major U2 spliceosome recognising the canonical GU-AG splice sites and the U12 minor spliceosome preferring the AU-AC splice sites (Patel and Steitz. 2003).

Alongside the above-mentioned core of snRNPs, vast numbers of additional proteins shape spliceosomal configurations, structure, and function during the process. Examples of these include the NineTeen Complex (NTC) and NTC related complex (NTR), two large protein complexes associated with the spliceosome during configuration of the active site and involved in shifting the spliceosome structure through the C and C* stages and PRP19 (the core protein for which NTC was named) is important in maintaining U5/U6 association within the spliceosome. Additionally, PRP19 plays role in assembly of U4/U6.U5 tri-snRNP through ubiquitination of PRP3 within the U4 snRNP to stabilise the tri-snRNP interaction (Maréchal et al. 2014, de Moura et al. 2018). These make up key components of the spliceosome alongside a variety of helicases and ATPases who provide the driving force behind structural configurations. In addition to these major components, a significant number of ancillary splicing factors play roles at different stages in the splicing process.

1.2.2 The mechanism of splicing

While splicing of precursor mRNA has been studied for a long time, the development of cryo-electron microscopy (cryo-EM) and its application in resolving spliceosomal components in action has provided clear mechanistic insight into the functions of this complex (Yan et al. 2019). For example, recent cryo-EM structures from Zhan et al (2018) of the spliceosomal C and C* complex show how the ATPase Prp16 triggers the shift from C to catalytically active C* by causing two splicing factors, CCDC49 and CCDC94 to dissociate. This allows the recruitment of distinct factors to the spliceosome, such as the ATPase Prp22 which allows formation of the catalytic C* complex (Zhan et al. 2018).

Splicing begins with the recognition of the 5′ splice site (5′ SS) by the U1 snRNP achieved through RNA: RNA binding at the 5′ end of the U1 snRNA. This leads to recognition of the branchpoint adenosine (BP) by SF1 and binding of U2AF to the 3′ splice site (3′SS) and formation of the E-complex (Berglund et al. 1998, Wilkinson et al. 2019). Next, SF1 and U2AF are displaced from the BP and the 3′ SS through the action of the DEAD box helicase Prp5 with U2 snRNP recruited to the BP, forming the A-complex (Liang and Cheng. 2015). The interaction with the BP leaves the 5′ end of the U2 snRNA free to recruit the U4/U6.U5 snRNP trimer to the pre-B spliceosomal complex through base pairing interactions at the 3′ end of the U6 snRNA. Driven by

the action of the ATPase PRP28, the pre-B spliceosomal complex undergoes a series of structural rearrangements resulting in the U1 snRNA being removed from the 5′ SS site leading to the formation of the B complex. The unwinding of the U4/U6 snRNP duplex leads to the dissociation of the U4 snRNP from the B complex and the recruitment of the NTC/NTR complexes to form the pre-B^{act} spliceosome, allowing U2/U6 snRNPs to fold together to form the active site of the spliceosome leading to the formation of the B^{act} spliceosome (Boesler et al. 2016). This undergoes further remodelling driven by PRP2 so that the BP is in contact with the spliceosomal active site to form the B^{*} complex for the first catalytic reaction. In this first catalytic step, termed branching, the 5′ of the intron is cleaved from the pre-mRNA and leaves an intron lariat attached to the 3′ SS. Further, remodelling of the spliceosome then occurs driven by the action of PRP16, a DEAH-box ATPase, which forms the catalytically active C^{*} spliceosome (Tholen and Galej. 2022). At this point, the now exposed 5′ end of the pre-mRNA attacks the 3′ SS, ligating the two exons together in the second catalytic reaction, known as exon ligation (Fica et al. 2017). Following this stage, the mRNA is released from the spliceosome by the actions of PRP22, after which the spliceosome undergoes rapid disassembly via the helicase PRP43 and allows for subsequent recycling of spliceosomal components and degradation of the resulting intron lariat (Tholen and Galej. 2022).

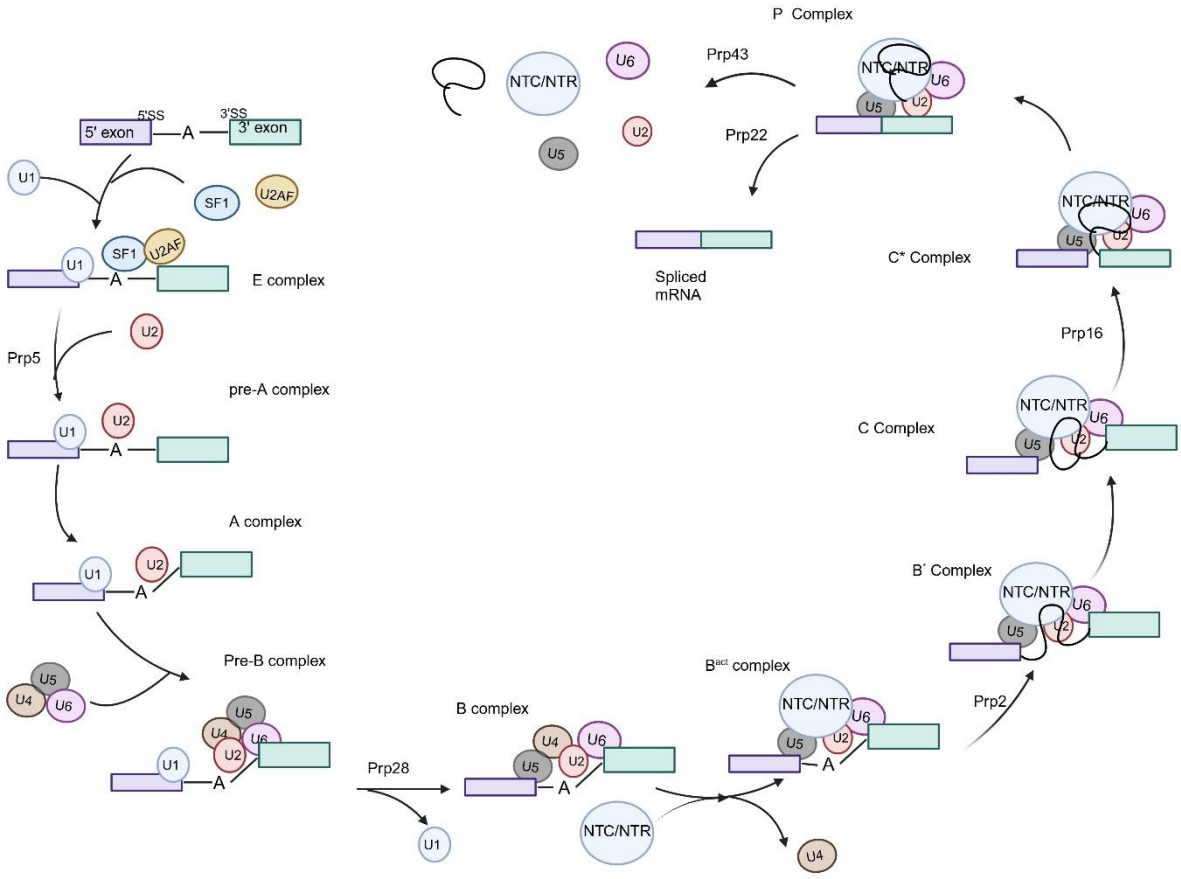


Figure 1.2: an overview of the splicing mechanism.

A schematic representation of splicing and dynamic remodelling of spliceosomal proteins at an intron is shown, based on the illustration by Tholen and Galej (2022). Spliceosome assembly begins with formation of the E complex through recognition of the 5′ SS, 3′ SS and the branchpoint adenosine. Recruitment of U2 snRNP forms the A complex, leading to recruitment of the U4/U6.U5 tri-snRNP which undergoes remodelling via the action of PRP28 and displaces the U1 snRNP to form the B complex. Loss of the U4 snRNP and recruitment of the NTC/NTR complexes leads to formation of the B^{act} spliceosome, which undergoes remodelling by PRP2 to form the B* complex for the first catalytic stage. This results in formation of the C complex, which is shifted into the catalytically active C* complex by PRP16. The actions of PRP22 and PRP43 cause dissociation of the spliced mRNA and the spliceosomal components.

1.3 Alternative splicing

The protein diversity of eukaryotes is further expanded by alternative splicing, causing multiple mRNA sequences and structures to be expressed from a single gene cassette, increasing the diversity of the transcriptome and proteome. Alternative splicing plays a significant role in diversity between tissues with differential expression of proteins that regulate splicing responsible for the differences in protein/RNA isoforms (Chen and Manley. 2009). There are several highly characterised forms of alternate splicing which include: alternate 5′ splice site usage, alternate 3′ splice site usage, exon skipping, intron retention and exclusive exons (Figure 1.3).

Alongside the core spliceosomal components, the Serine/Arginine (SR) protein family play an important part in splicing regulation and initiation. All SR proteins contain two conserved domains - one serine and arginine rich (SR) domain, for which the protein family is named, and one RNA recognition motif (RRM) domain (Shepard and Herterl. 2009).

The recognition of splice sites plays a key role in alternative splicing, with selection of splice sites and pairing of the 5′SS and 3′SS occurring during formation of the Spliceosomal A complex, characterised by U2 snRNP binding to the 3′SS (Kotlajich et al. 2009). Splice point selection is controlled by SR proteins whose activity can be controlled through phosphorylation to recruit the U1 and U2 snRNPs to different splice sites for alternative splicing. For example, the protein SRp38 switches between a splicing activator and repressor depending on its phosphorylation state (Feng et al. 2008) thus allowing phosphorylation states of SR proteins to control spliceosome formation. The main regulation in choosing which splice sites to use comes from splicing factors or regulators that can bind *cis* – acting elements that can either enhance or repress splicing at specific splice sites (Chen and Manley. 2009., Yu et al. 2008).

SR proteins are capable of binding to exonic splicing enhancers (ESE) or intronic splicing enhancers (ISE) through their RRM domain to promote splicing and spliceosome recruitment (Schaal and Maniatis. 1999). ESE are found adjacent to the SS, up to 100nt away, and SR binding to exonic or intronic ESEs can influence intron removal or inclusion. For example, when SRSF7 was tethered to a β-globin pre-mRNA upstream of a 5’ SS, the related intron underwent efficient splicing. However,

downstream tethering inhibited splicing of that intron, leading to its inclusion in the mRNA (Erkelenz et al. 2013).

Alongside SR proteins, the heterogeneous nuclear ribonucleoprotein (hnRNP) protein family also regulate alternative splicing, although playing a more antagonistic role. A varied family of proteins, they have roles at several levels of mRNA processing including splicing, mRNA stability, export, and translation (Geuens et al. 2016). hnRNP A1 was originally shown to act in alternative splicing by binding exonic silencer sequences (ESS), with binding of hnRNP A1 to these sites inhibiting exonic splicing (Del Gatto-Konczak et al. 1999). The interplay between SR proteins and hnRNPs builds an intricate picture of alternative splicing regulation, with a complex network of protein-protein and protein-RNA interactions determining the splicing isoform for each mRNA (Wang et al. 2015).

Additionally, hnRNP A1 has been shown to be involved in 3′ SS recognition and selection by U2AF. U2AF recognises the branchpoint adenosine and the conserved AG sequence at the 3′ end, however U2AF binding at sequences away from splice sites has also been reported. hnRNP A1 is also recruited to the 3′ SS by recognising sequences sharing similar homology to the canonical 3′ SS, forming a complex with U2AF and displacing it from unsuitable or unwanted sequences (Tavanez et al. 2012).

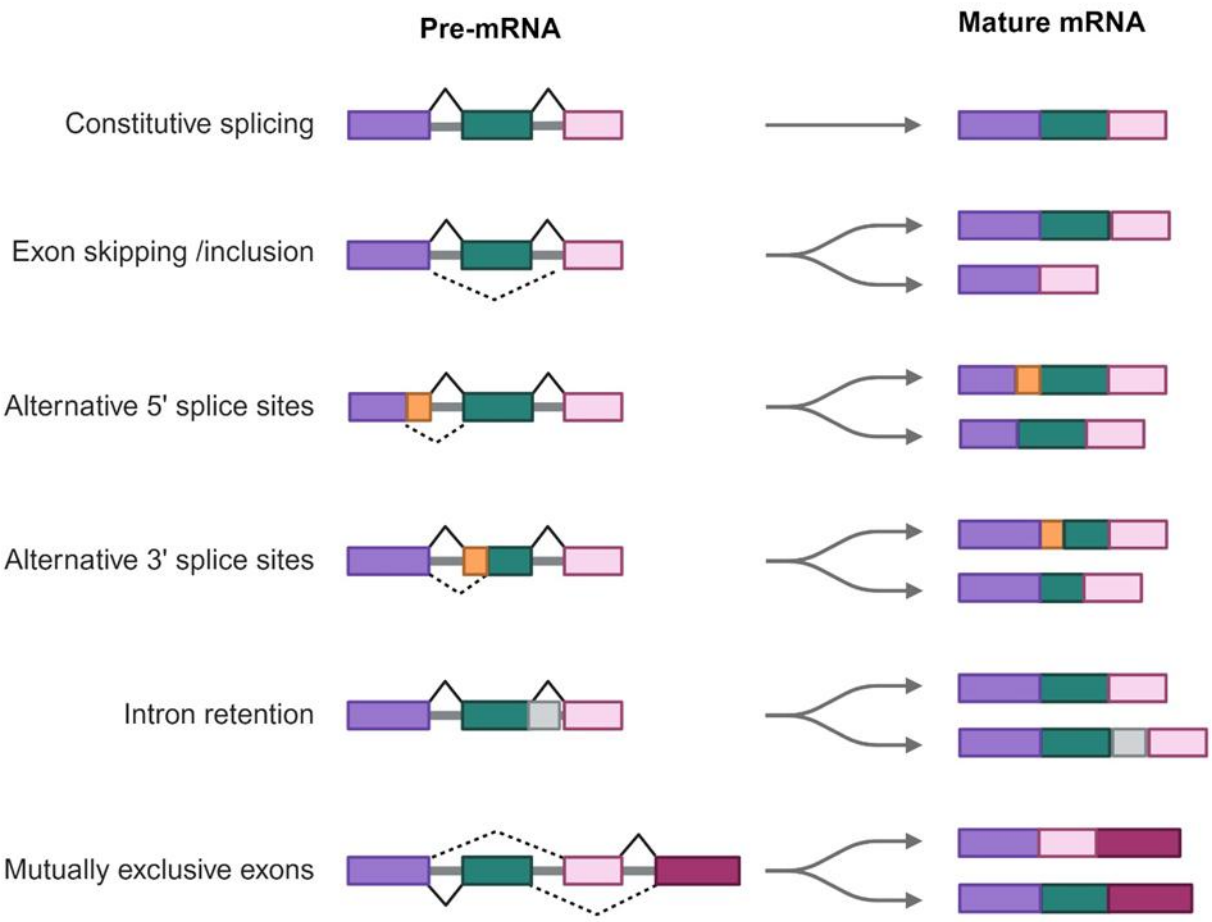


Figure 1.3: A schematic of alternative splicing.

Alternative splicing increases the diversity of the transcriptome. Mechanisms can include removing an exon from the mRNA (exon skipping), increasing or decreasing exon length through alternate splice site usage, leaving parts of an intron in the mRNA (intron inclusion) or using different exons from the same gene depending on end protein function (mutually exclusive exons).

1.3.1 Splicing in disease

While splicing/alternative splicing increases the diversity of the transcriptome, improper regulation or mutations can lead to mis-splicing, creating aberrant protein isoforms that may cause disease.

Several diseases have pathologies characterised by mutations in splice sites or affected by alternatively spliced isoforms. Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disorder characterised by loss of motor neurons leading to muscle atrophy. TAR DNA binding protein 43 (TDP-43) is a member of the HNRNP family and thus is an essential splicing factor, TDP-43 has been connected to ALS through both increases and decreases in protein levels (Fratta et al. 2018). One example of mis-regulation of alternate splicing in this disease comes from HNRNP A1, where TDP-43 is required for binding to HNRNP A1 transcripts to regulate the exclusion of exon 7B. Loss of function of TDP-43 leads to inclusion of exon 7B in the HNRNP A1 protein, which creates a larger protein size leading to the formation of HNRNP A1 aggregates that causes reduced cellular viability (Fratta and Isaacs. 2018).

Another well characterised disease resulting from alternative splicing is Spinal muscular atrophy (SMA). SMA is caused by mutations in the gene survival of motor neurons 1 (SMN1), causing the cells to rely on the second copy of the gene SMN2. However, SMN2 undergoes alternative splicing, frequently skipping exon 7 which results in an unstable and quickly degraded form of SMN, leading to the disease (Cho and Dreyfuss. 2010). This skipping is caused by a single nucleotide mutation, a C to a T, leads to disruption of the intron 7 5′ SS, leading to skipping of exon 7 (Lorson et al. 1999, Singh and Singh. 2011). While traditional treatments of SMA have focused on gene therapies via introduction of another gene copy to improve protein levels, more recent drug advances have focused on improving the splicing efficiency. Spinraza is an antisense oligonucleotide (ASO) that binds to a complimentary downstream regulatory intronic element that recruits splicing factors, such as hnRNPA1, that cause exon 7 skipping. ASO’s binding prevents splicing factor recruitment leading to exon 7 inclusion and generation of full-length protein and alleviating symptoms of the disease (Shorrock et al. 2018).

1.4 Exon junction complex

Efficient recognition of spliced mRNAs is a crucial step in ensuring correct RNA processing. To fulfil this role, the Exon Junction Complex (EJC) is deposited 20-24nt upstream of exon-exon junctions to function as a molecular marker of splicing (Le Hir et al. 2000, Saulière et al. 2012). The EJC has a stable tetramer of core proteins: the dead box helicase eIF4AIII, MLN51, MAGOH and Y14 (Tange et al. 2005). MAGOH and Y14 forms a heterodimer that locks eIF4AIII on RNA in an ATP pre-hydrolysis conformation (Andersen et al. 2006). CLIP studies have showed that the majority of eIF4AIII is bound in the CDS of protein coding genes at the canonical EJC sites, although eIF4AIII binding on other locations along an mRNA has been detected (Saulière et al. 2012).

The EJC is deposited on mRNA in a sequence independent manner following splicing, indicating that it's localisation/binding to mRNA comes from interactions with the splicing machinery. The splicing factor CWC22 directly interacts with eIF4AIII when it's in an open conformation, as opposed to the closed conformation it adopts when bound to ATP, RNA and the MAGOH/Y14 heterodimer (Barbosa et al. 2012). Furthermore, recruitment of eIF4AIII to the spliceosomal C complex and to the spliced RNA is abolished upon knockdown of CWC22, showing that direct interactions with the splicing machinery are required for EJC recruitment (Barbosa et al. 2012). Subsequent cryo-EM studies have confirmed the association of eIF4AIII in the spliceosome during the catalytically active stages, indicating that the sequence independent binding 20 – 24nt upstream of exon-exon junctions of the EJC comes from molecular and steric interactions with the spliceosome during splicing (Galej et al. 2016., Schlautmann and Gehring. 2020).

A well reported observation that spliced genes are better expressed than unspliced genes can be explained, in part, by the variety of functions the EJC participates in during mRNA processing (summarised in figure 1.4). These effects are achieved through peripheral members of the EJC, proteins that are found regularly associated with the complex but are not considered key members (Le Hir et al, 2016).

Knockdown of EJC members in both *Drosophila* and Humans has been linked to alternate splicing events, differential usage of 5' and 3' splice sites and exon skipping (Schlautmann and Gehring. 2020). In *Drosophila,* knockdown of the EJC components

caused a decrease in polymerase pausing at promotors, with increased amounts of elongating Pol II found along the gene. Mago (the *Drosophila* homologue of MAGOH) was found to inhibit Pol II elongation, by blocking the recruitment of CDK9 to PolI II and preventing premature release of pol II into transcriptional elongation along the gene. Knockout of Mago caused increases Ser 2 phosphorylated Pol II, leading to early transcriptional elongation which gives rise to the increase in exon-skipping and abnormal splicing events seen upon knockdown Mago (Akhtar et al. 2019). A similar effect on the rate of Pol II transcription in the absence of MAGOH has also been reported in humans (Wang et al. 2014, Akhtar et al. 2019) demonstrating that the EJC effects alternate splicing and exon definition through control of the transcription rate of Pol II.

Several key export factors, such as the TREX complex members UAP56 and ALYREF, have been identified as EJC peripheral proteins (Le Hir et al. 2016). The EJC, specifically eIF4AIII, has direct interactions with the key export factor ALYREF and is an essential part of ALYREF recruitment to mRNPs and the subsequent recruitment of the key export heterodimer NXF1:NXT1 to facilitate nuclear export (Viphakone et al. 2019). The EJC also has well documented interactions with several SR proteins, SRSF1, SRSF3 and SRSF7, all of which are well known export adaptors and capable of nuclear export of an mRNA (Singh et al. 2012).

The continued association of the EJC following nuclear export, alongside the observations that spliced mRNA produced more protein molecules than unspliced mRNAs, suggested a role for the complex in translation. Physical tethering of EJC core members to intronless RNAs was shown to improve the translation efficiency and ribosome association (Nott et al. 2004). The exact mechanism of EJC regulation of translation is still not fully understood, however several proteins have been shown to interact with the EJC to effect translation. Translation initiation is controlled by signalling from the mTORC1 pathway. This pathway leads to phosphorylation and activation of S6K1 which contributes to the initiation of scanning by the 40s ribosomal subunit (Hay and Sonenberg. 2004). The link between S6K1 and the EJC comes from the S6K1 interaction protein SKAR, which can bind to eIF4AIII and recruit S6K1 to the mRNPs and cause subsequent phosphorylation activation of proteins in the mRNPs and recruit ribosomes, however the exact mechanism for this is still unclear. In support of this, knockdown of S6K1, SKAR and eIf4AIII all decrease the translational rate of

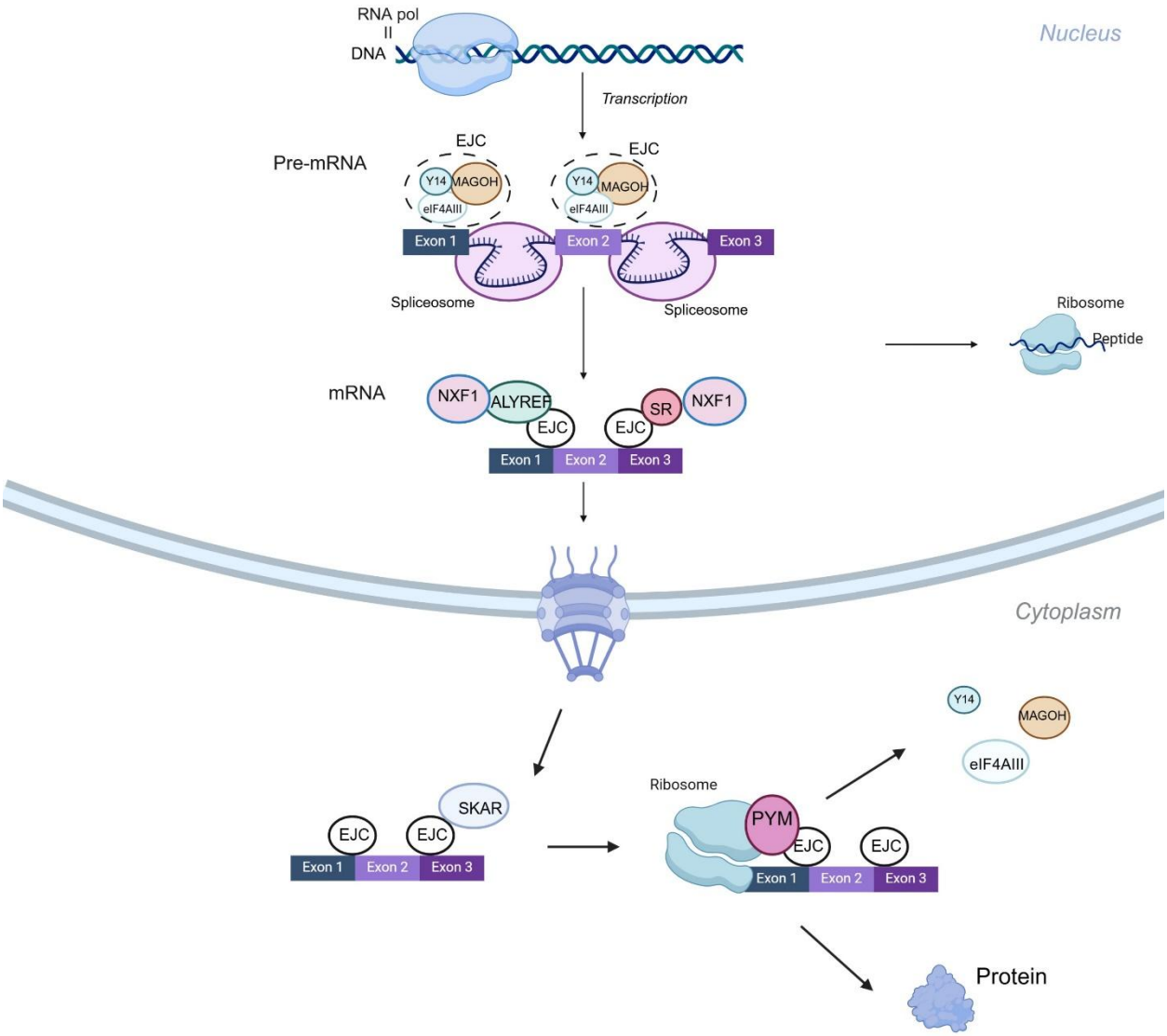


Figure 1.4: The functions of the EJC.

The EJC is a hallmark of correct splicing, deposited 20-24nt upstream of splice sites after intron removal. The EJC can interact with export factors, such as ALYREF or SR protein to facilitate NXF1 recruitment to an mRNA for export to the cytoplasm. Once exported, the EJC can interact with SKAR, leading to translational activation and ribosomal recruitment to an mRNA. PYM also assists in both recruiting the ribosome to an mRNA through interactions with the EJC, which also destabilise the complex and allow its recycling.

spliced mRNAs without effecting the RNA level, supporting the idea of a role of the EJC in improving translation of spliced mRNAs (Max ma et al. 2008).

A further link between the EJC and translations comes from the interactions of the protein PYM and the ribosome. PYM binds to the MAGOH/Y14 heterodimer through it's N-terminal domain (Bono et al. 2004) and interacts to the 40s subunit of the ribosome through it's C-terminal domain. Remarkably, the C-terminal domain of PYM has homology with the C-terminal end of eIF2A, explaining it's interaction with the ribosomal subunit (Diem et al. 2007). Diem et al (2007) showed that knockdown of

PYM using RNAi caused no changes in translational efficiency of an intronless mRNA but caused a significant reduction in the translational efficiency of spliced RNA, suggesting that PYM provides the link between splicing and translation through its interactions with both the EJC and the 40S ribosomal subunit (Diem et al. 2007). This interaction with PYM is also important for the disassembly of the EJC from mRNA and recycling of the EJC components. PYM association with the ribosome allows PYM to interact with an EJC on the mRNA, binding the Y14/MAGOH heterodimer and destabilising its interaction with eIF4AIII. This allows eIF4AIII to enter a open conformation by releasing ADP, thus allowing disassembly of the EJC from the mRNA (Gehring et al. 2009, Woodward et al. 2017).

One final processing function of the EJC lies in nonsense mediated mRNA decay (NMD), a failsafe process for degradation of mRNA that undergo premature translational termination (Kolakada et al. 2024). When translation terminates 50 nucleotides upstream of the last exon-exon junction, and thus the last EJC bound to the mRNA, this acts as a signal for termination of the mRNA to prevent further aberrant protein structures being generated. Y14 present in the EJC acts as a binding point for SMG-1 and formation of the SMG-1-Upf1-eRF1-eRF3 (SURF) complex to mediate the phosphorylation of Upf1 (Kashima et al. 2006). Hyperphosphorylation of Upf1 results in a cascade of downstream effects, such as the inhibition of further ribosomal recruitment to the targeted mRNA and recruitment of de-capping enzymes and deadenylation to allow degradation of the target mRNA by exonucleases (Hwang et al. 2021).

1.5 mRNA export

1.5.1 TREX

The three main stages of pre-mRNA processing occurring co-transcriptionally are all key stages of recruitment for members of the TREX complex. TREX consists of the conserved THO complex, consisting of THO1, 2, 3, 5, 6 and 7 that forms the core scaffold of TREX (Heath et al. 2016). THO associates with the DEAD-box helicase DDX39b, also referred to as UAP56, to interact with a variety of RNA binding proteins that function as 'export adaptors' and load them onto pre-mRNA in an ATP dependent manner via DDX39b (Dufu et al. 2010).

A large compliment of export adaptors has been identified as interacting with TREX however the protein ALYREF is well established to be one of the key export adaptors of the mammalian cell (Masuda et al. 2005). ALYREF contains a central RBM motif flanked by two disordered Arginine/glycine rich domains capable of binding to RNA, with N and C terminal helices necessary to bind DDX39b (Golovanov et al. 2006). The importance of ALYREF to the cell is highlighted by its known redundancy with the export adaptor UIF, an RNA binding protein that is upregulated in the event of ALYREF siRNA knockout. To achieve a significant export block, UIF also had to be knocked out alongside ALYREF, indicating the importance of ALYREF mediated export (Hautbergue et al. 2009).

ALYREF recruitment has been shown to occur at all stages of pre-mRNA processing, with the different mechanisms of recruitment working in tandem (Viphakone et al. 2019). A protein:protein interaction between ALYREF/TREX and CBP20/80 was shown using GST pulldowns with this implying the binding of TREX is predominantly at the 5′ end of pre-mRNA (Cheng et al. 2006). At the time, this contrasted with other reports that suggested that DDX39b and ALYREF were members of the EJC and that it's deposition onto pre-mRNA was a consequence of splicing (Gatfield et al. 2001, Masuda et al. 2005).

More recent work has made use of individual nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP) which allows examination of direct protein-nucleotide interactions (Viphakone et al. 2019, Konig et al. 2011). This allowed for *in vivo* resolution of how ALYREF binds to pre-mRNA. Strikingly, Viphakone et al (2019) found that both stories were true. ALYREF iCLIP showed that the protein was enriched -70

to -20nt upstream of splice sites, matching the iCLIP pattern of eIF4A3, a key component of the EJC. However, they observed prominent ALYREF binding at the 5′ end of a pre-mRNA leading to the idea that ALYREF is originally recruited by the CBC before being distributed along the mRNA to the sites of splicing through interactions with eIF4A3 (Figure 1.5) showing that both proposed methods of ALYREF recruitment were correct (Viphakone et al. 2019).

The ability of ALYREF to bind the CBC as a transient loading system is additionally important as a mRNA degradation maker. Human MTR4 is a key member of the nucleoplasmic NEXT degradation complex which targets a large variety of RNA transcripts for exosome degradation following transcription (Lingaraju et al. 2019). MTR4 can bind a member of the CBC, ARS2, through direct protein:protein interactions and plays a key role in recruiting the exosome to target mRNA (Fan et al. 2017). It was also shown that ALYREF is also capable of protein:protein interactions with ARS2, this interaction assisting in the recruitment of the export factor to the 5′ end of mRNA. As inefficiently exported mRNAs are subject to extensive degradation, it was shown that ALYREF and MTR4 compete for binding of ARS2, with overexpression of ALYREF resulting in a reduction in degradation of exosome target RNAs with the opposite effect observed in overexpression MTR4 (Fan et al. 2017). Hence ALYREF and MTR4 compete for binding at the CBC, with this competition determining the fate of a pre-mRNA. This suggests that the transient ALYREF binding to the CBC is to mark the pre-mRNA for export to the cytoplasm and prevent exosome degradation and allowing the handover of ALYREF along the pre-mRNA to splice sites (Viphakone et al. 2019).

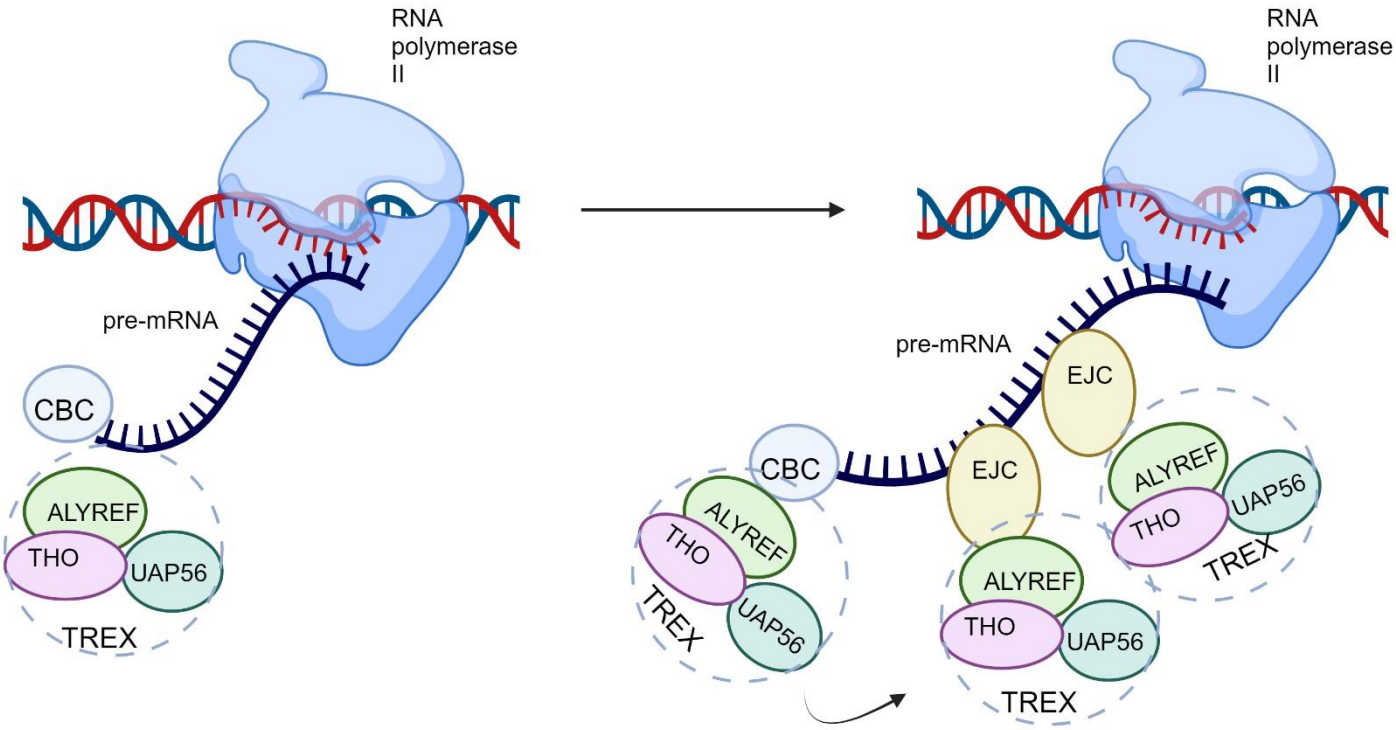


Figure 1.5: **Co-transcriptional ALYREF recruitment to an mRNA.**

ALYREF/TREX is recruited to the 5′ end of mRNA through direct interactions with the CBC. As splicing occurs and the EJC binds to the pre-mRNA, ALYREF shifts along the pre-mRNA, moving from the CBC to bind along the length of the pre-mRNA at site where the EJC is present.

1.5.2 NXF1 facilitated export.

Whilst TREX has a prominent role in mRNA processing by acting as an export adaptor, the role of chaperoning mRNA across the nuclear pore falls to NXF1(also called TAP). Originally discovered from a yeast two hybrid screen as an interactor of tyrosine interacting protein, NXF1 shared a similar homology to the yeast protein Mex67 which had a demonstrated participation in nuclear export of yeast mRNA and was shown to be a conserved export factor (Yoon et al. 1997, Katahira et al. 1999).

NXF1 a has well established modular structure, with structured domains linked by unstructured linker regions (Hautbergue et al. 2008). NXF1 populations within the cell exist in two conformations: an open state in which the N-terminal RNA binding domain (RBD) is freely accessible and a closed state in which the RBD is bound to the NTF2L domain which prevents NXF1 from binding to RNA (Viphakone et al. 2012). ALYREF, alongside THOC5, plays a key role in allowing NXF1 to enter an open conformation and bind mRNA. The binding sites in ALYREF that are responsible for binding to RNA and to NXF1 overlap, highlighting this ability to handover RNA to NXF1 (Hautbergue et al. 2008). ALYREF binds to NXF1s RBD and destabilises the interaction between RBD and NTF2L domain and opens up NXF1 so that the RBD is now capable of RNA binding. THOC5 binds to the NTF2L domain to hold NXF1 in this open conformation and allow the handover of RNA from ALYREF to NXF1 (Viphakone et al. 2012). As TREX recruitment occurs during the co-transcriptionally modification of pre-mRNA, this conformational change acts as a molecular switch to ensure that only correctly processed mRNAs are exported to the cytoplasm.

To export mRNA through the nuclear pore complex (NPC), NXF1 heterodimerises with p15 through direct interactions with its NTF2L domain (Viphakone et al. 2012). The NXF1: p15 heterodimer can interact with the NPC via two FG nucleoporin repeat binding domains, one present in the NTF2l domain and the other in the C-terminal UBA like domain (Fribourg et al. 2001). The domains act synergistically in facilitating mRNA export, with knockout of each FG binding pocket reducing mRNA export by about 20%. However, the FG binding pocket of NFT2L domain is only active when bound to p15, demonstrating that the heterodimer structure is essential in allowing NXF1 bound to mRNA to move through the nuclear pore (Fribourg et al. 2001).

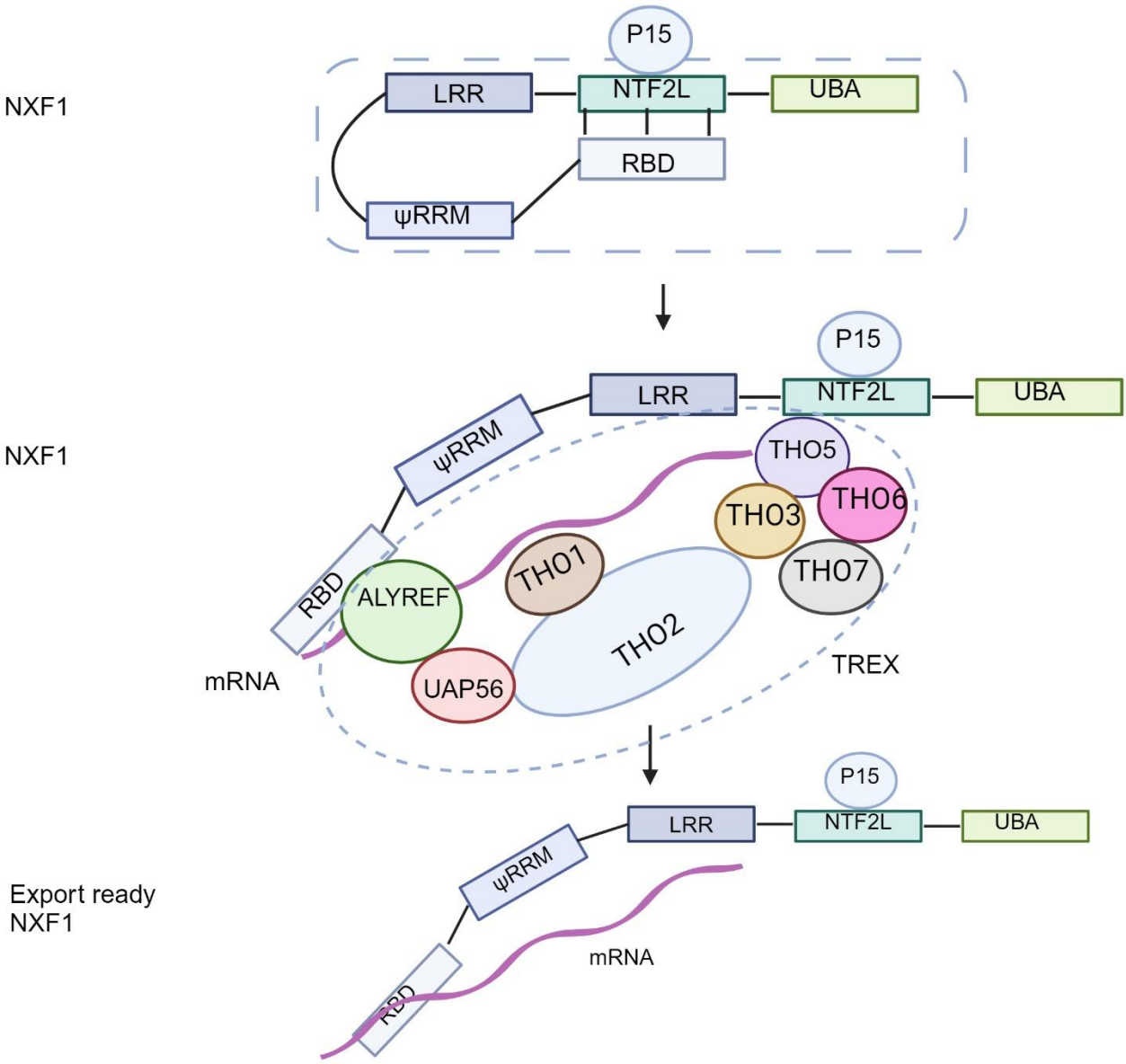


Figure 1.6: The activation of NXF1.

NXF1 is found in the cell in a closed conformation, unable to bind to mRNA as it's RNA binding domain is bound to the NTF2L domain. TREX can bind NXF1, with ALYREF recognising the RBD and binding it, while THO5 stabilises the NXF1- TREX interaction by binding to the NTF2L domain. This places the NXF1 in an open conformation where the mRNA can be handed over from TREX to the RBD of NXF1 where it can export the mRNA to the cytoplasm.

1.5.3 Export of intronless mRNAs

The coupling of pre-mRNA splicing, and mRNA export provides a clear explanation of the long-standing observation that cDNAs are exported to the cytoplasm less efficiently than their intron containing counterparts. A classic example of this is the β-globin gene, in which the spliced version of the gene shows high cytoplasmic localisation while the cDNA remains nuclear localised and subject to rapid degradation (Valencia et al. 2008). This raised the question: how do single-exon genes get efficiently exported to the cytoplasm?

Initial work to understand expression of intronless mRNAs originally used viral mRNAs to model intronless mRNA export. These studies showed that viruses used host export machinery, such as the NXF1 or CRM1 export pathways, to efficiently export their RNAs to the cytoplasm (Chen et al. 2002). However, analysis of naturally intronless genes in eukaryotes has centred on the use of Cytoplasmic Localisation Regions (CARs) to facilitate nuclear export (Khan et al. 2021).

One of the first examples of a CAR came from mouse Histone 2a, where a 100bp sequence was necessary for cytoplasmic localisation of H2a mRNA. Insertion of this sequence was also capable of causing increased export of intronless HIV RNA (Huang and Carmichael. 1997). Several more CARs have been discovered for single exon genes, for example HSPB3 and IFN-α1, with these regions able to cause export of separate cDNAs showing that sequence specificity is required. Additionally, knockdown of TREX components, such as UAP56, is sufficient to abolish export of their mRNAs, suggesting that naturally occurring intronless mRNAs use sequence dependent motifs to recruit export factors from the TREX/NXF1 export pathway (Lei et al. 2011, Lei et al. 2013).

Recent advancements in single exon RNA export have stemmed from the analysis of long noncoding RNAs (lncRNAs). As some lncRNAs are cytoplasmic in *Drosophilla* and humans, export of these RNAs provides further clues to single exon RNA export. A recent example is the lncRNA NKILLA, a cytoplasmic lncRNA involved in regulation of cytokine and inflammation by stabilising the NF-kB negative feedback loop, NKILLA down regulation is an indicator of poor prognosis in breast cancers (Liu et al. 2015). NKILLA cytoplasmic localisation is a result of a 200nt CAR close to the 5′ end of NKILLA that contains clusters of SRSF1 and SRSF7 binding sites. This then allows

recruitment of UAP56 and ALYREF to the lncRNA, allowing export to the cytoplasm via the NXF1 pathway (Khan et al. 2021). This provides further evidence that single exon genes/RNAs use sequence dependent motifs to help recruit export factors, rather than relying on a processing reaction or EJC deposition onto an mRNA.

While SR proteins have key roles in the spliceosome, the role of SR proteins as export adaptors has been reported for some time. Originally discovered as splicing factors, SRp20 (SRSF3) and 9G8 (SRSF7) were shown to bind a 22nt sequence in the Histone 2a (H2a) and promote its export to the cytoplasm (Huang and Steitz. 2001). The two proteins, alongside another SR protein SF2 (SRSF1) were shown to directly interact with NXF1 in the same domain as ALYREF interacts with NXF1 (Huang et al. 2003). SRSF3 and SRSF7 can both interact with NXF1 through a C-terminal region in their RRM domain that is not involved in mRNA binding, allowing them to hand mRNA over to NXF1 in a similar manner to ALYREF (Hargous et al. 2006, Hautbergue et al. 2008).

Another contributing factor to export of single exon genes has emerged in the form of GC content. Recent work by Zuckerman et al (2020) using RNAi followed by RNA sequencing from different cellular compartments showed that single exon genes were especially affected by knockdown of NXF1, which seemed to be recruited by short GC rich sections which had predicted secondary structures (Zuckerman et al. 2020). This is also supported by the observations of Mordstein et al (2020) who reported increases in RNA export to the cytoplasm from single exon transcripts when the GC content was increased (Mordstein et al. 2020). Both studies also reported that increases in mRNA export were also seen when the increased GC content was localised to the 5′ end of the mRNA. However, how GC content can cause increased export of a single exon transcript remains unclear.

1.6 Genome defence against transposable elements

Given the complexity of eukaryotic genomes, the need to distinguish self from foreign elements is critical. A molecular army is dedicated to the detection and destruction of foreign nucleotides that may attempt to infiltrate the genome (De Oliveria Mann and Hornung. 2021). However, throughout evolution parasitic pieces of DNA have become established within eukaryotic genomes. These Transposable Elements (TEs) come in two forms, DNA transposons that replicate and directly insert themselves into the genome and retrotransposons which replicate via RNA intermediates (Cordaux and Batzer. 2009).

1.6.1 HUSH complex

Silencing of transposable elements in the genome is carried out by the HUSH complex. Originally identified in genetic screens looking at epigenetic silencing elements, HUSH was shown to be involved in the silencing of retrotransposons through epigenetic interactions (Tchasovnikarova et al. 2015). The HUSH complex is a trimeric complex consisting of M-phase phosphoprotein 8 (MPP8), Transcription activation suppressor (TASOR) and Periphilin (Tchasovnikarova et al. 2015).

TASOR has been identified as the central scaffold of the complex, with separate binding pockets for MPP8 and Periphilin, with MPP8 and periphilin unable to bind each other without the presence of TASOR (Douse et al. 2020). Demonstrating its importance as a member of HUSH, TASOR mutation is embryonically lethal in mice (Harten et al. 2014.) As HUSH is localised to chromatin to enact its role in epigenetic repression of transposons and other unfavourable DNA elements, recruitment to Histones is vital to its function. This is achieved through the chromodomain of MPP8 which can bind H3K9me3 but not unmethylated H3. However, in the absence of MPP8 and TASOR, periphilin is still able to localise to chromatin, indicating multiple recruitment mechanisms for the complex (Tchasovnikarova et al. 2015).

The final component of HUSH is SETDB1, a peripheral component of the complex. SETDB1 is a histone methyltransferase that deposits the repressive H3K9me3 onto histones, an activity which is inhibited in the event of knockdown of HUSH components (Tchasovnikarova et al. 2015).

1.6.2 HUSH mechanism and roles

HUSH actively represses the transcription of L1 retrotransposons and foreign DNA elements by increasing the deposition of the repressive H3K9me3, but how this complex is recruited and maintained at sites of repression is still not fully understood. HUSH is localised to chromatin by the ability of MPP8 to recognise H3K9me3 where it can recruit SETDB1 to further increase histone methylation (Müller and Helin. 2024). The localisation of HUSH to target loci is additionally helped by the RNA binding ability of Periphilin, which is proposed to bind nascent RNA transcribed from the transgene. This explanation of HUSH recruitment is explaining the observation that Periphilin can still localise to the chromatin in the event of a knockdown of other HUSH core proteins.

HUSH recruitment to target loci is dependent on transcription, resulting from decreasing methylation of the target, occurring during natural cellular events such as mitosis. The transcription of pre-mRNA enables periphilin binding to the RNA which results in the targeting of HUSH to the transcriptionally active loci (Bloor et al. 2025). Recruitment of HUSH to the loci then enables MPP8 to bind H3K9me3 and recruit SETDB1 and MORC2, a chromatin remodeler, to increase H3K9me3 repression of the loci. To degrade the nascent RNA already transcribed, HUSH recruits the nuclear exosome targeting (NEXT) complex via direct protein interactions between MPP8 and ZCCHC8. However, as this interaction is RNA independent, how NEXT is recruited to the RNA from repressed loci is currently unknown (Garland et al. 2022).

As HUSH is responsible for the silencing of mobile elements, especially L1 retrotransposons which are still mobile in the human genome, HUSH must have a manner of sequence identification to facilitate its recruitment. Recent work has shown that introns can protect against HUSH repression. Insertion of an intron into a HUSH repressed reporter abolished repression through reduction of periphilin binding and H3K9me3 deposition on the reporter. Interesting, mutations in intronic splice sites still prevented HUSH repression, indicating it is the presence of the introns rather than their processing that prevents HUSH binding (Seczynska et al. 2022). As foreign DNA and retrotransposons don’t contain introns but simply long stretches of coding DNA due to their limited size and genomic selection pressures, this allows the HUSH complex to distinguish between self and non-self. (Seczynska et al. 2022).

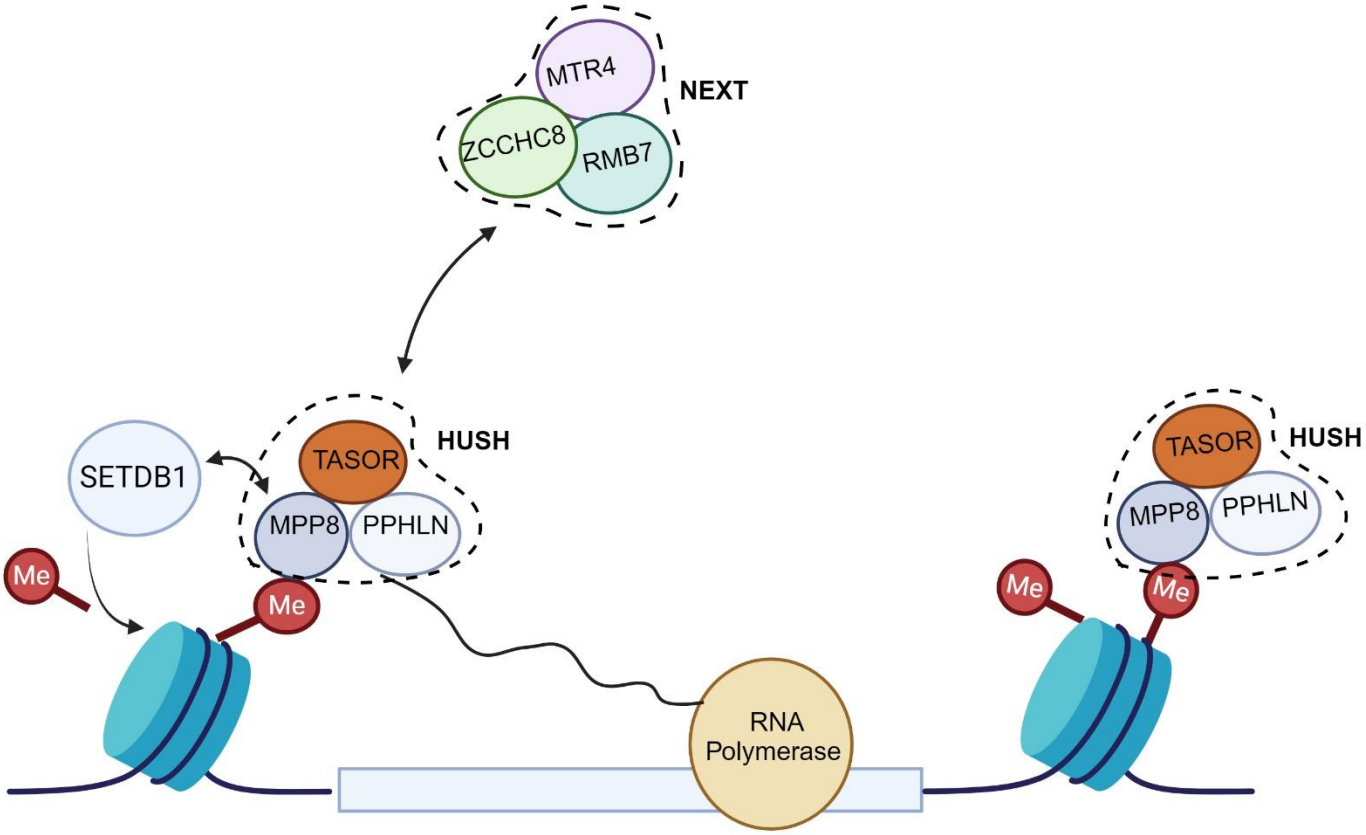


Figure 1.7: Schematic of HUSH repression.

The HUSH complex is recruited target sites by recognition of target mRNA by PPHLN. HUSH binds to H3K9me3 through the chromodomain of MPP8 to stabilise at the transcriptional locus, while recruiting SETDB1 to increase H3K9me3 deposition to repress transcription. Any transcribed mRNA is degraded by the NEXT complex which HUSH recruits to the transcriptional locus. Based on an illustration by Garland et al (2022).

1.7 Monoclonal antibodies as a treatment for disease

Since the approval of Muromonab CD3 as the first monoclonal antibody (mAb) for therapeutic uses in humans, the range of antibodies approved for treatment of disease has continued to grow (Todd and Brogden. 1989, Lu et al. 2020). Industrial production of antibodies for therapeutic use was made possible by the work of Köhler and Milstein (1975) where they fused B cells to immortalised mouse hybridoma cells to create a cell line capable of long-term antibody production (Köhler and Milstein. 1975).

Since then, the uses of mAbs as treatments for disease has grown exponentially. As of June 2024, the number of approved mAbs is approaching 200, with the number of approvals by relevant drug authorities increasing each year (Kelley, B. 2024). One of the most well-known examples of a mAb being used in cancer treatment is bevacizumab (Avastin), an angiogenesis inhibitor. Vascular Endothelial Growth Factors (VEGF) play a key role in tumour angiogenesis, by stimulation the expansion of endothelial cells. VEGF has several isoforms, generated through alternative exon splicing, with the different isoforms having different cellular localisations and secretion properties (Tischer et al. 1991, Ferrara et al. 2003). VEGF-A was identified as a good therapeutic target, causing inhibition of angiogenesis upon it's knockout (Gordon et al. 2001, Garcia et al. 2020). Bevacizumab acts to inhibit tumour angiogenesis by preventing VEGF binding to its tyrosine receptors and prevent the stimulation of endothelial growth and angiogenesis. Since its approval by the FDA in 2006, Bevacizumab has remained a highly used and effective treatment in cancers that have increased VEGF levels, and often poor prognosis, such as metastatic colorectal cancer or metastatic breast cancer (Garcia et al. 2020). Currently, mAbs are used to treat an ever-expanding variety of cancers, immune related disease, and infectious diseases (Lyu et al. 2022).

1.7.1 Types of therapeutic antibody

When discussing the uses of antibodies for disease treatment, most drugs are referred to as mAbs. However, there are a variety of differences in the way these drugs are designed, constructed, and made. A standard IgG consists of 4 polypeptide chains, with 2 light chains (weighing 25 kDa) and 2 heavy chains (weighing 50kDa) coming together to form the iconic Y shape of an antibody (Figure 1.8). The two heavy chains are connected through disulfide bonds, with each heavy chain binding a light chain.

Antibody specificity is determined by the variable domains of the heavy and light chains at the terminal end of the amino acid sequence. Antibodies can be broken down further into the Fragment antigen binding (Fab) and the Fragment crystallisable (Fc) domains. The Fc fragment is the section of the antibody responsible for localisation and interaction with target cells (Schroeder jr and Cavacini. 2010).

The first mAbs produced by Köhler and Milstein (1975) were murine antibodies, reducing their uses in disease treatment due to the elevated levels of immunogenicity (Köhler and Milstein. 1975). The immunogenicity of these early antibodies was improved by generating chimeras, antibodies that contained a variable region generated from mouse/rat hybridoma cell lines and were ligated to constant regions derived from humans (Morrison et al. 1984). This was further improved upon by the generation of 'humanised' antibodies, which contains the ß-sheet backbone from a human antibody and a complementary determining region (CDR) from a mouse (Jones et al. 1986). Eventually, production of fully human mAbs was achieved using transgenic mice, genetically modified so that the endogenous IgG genes were replaced with human ones. This allowed generation of human mAbs from mice exposed to the target antigen (Lonberg et al. 1994).

An additional class of therapeutic antibody exist in the form of Fc-fusion protein (Fc-FP) molecules. These drugs rely on the fusion of an antibody Fc domain to an effector protein that performs the targeted function of the drug. The first reported Fc-FP came from Capon et al (1989) where they generated an antibody containing the Fc fragment from a human IgG 1 heavy chain, fused with the globulin like domain of CD4, the cell surface receptor to gp120 – the HIV-1 envelope protein (Capon et al. 1989, Duivelshof et al. 2021). Their previous work had demonstrated the ability of fragments of CD4 (termed rCD4) to prevent HIV-1 infection, but these fragments suffered from incredibly short plasma half lives in rabbits (around 15 minutes). Fusion of the Fc fragment to the same regions had a dramatic effect on plasma stability, increasing half-life up to 48hours for some molecules (Capon et al. 1989).

The fusion of biological active drugs to Fc domains confers several advantages. As stated above, fusion to an Fc fragment great improves stability of the conjugated drug in serum, improving the pharmacological activity of the drug. Alongside increased pharmacological value, the fusion of a protein to an Fc domain allows for greater yield

of a protein owing to improved stability during production (Duivelshof et al. 2021). The addition of an Fc fragment also allows purification of a drug conjugate in an easier manner, using Protein A chromatography in an analogous manner to mAb purification (Chapon et al. 1989).

The first Fc-FP drug to be granted license for medical use was Etanercept which was approved in 1998, still used today, as a treatment for Rheumatoid Arthritis (RA). Tumour Necrosis Factor-α (TNF-α) is an inflammatory cytokine that forms a key part of the innate immune response that can upregulated/ overexpressed in RA, causing chronic inflammation and affecting patient lifestyles (Maini et al. 1995). Ethanercept contains the human IgG1 Fc fragment fused to the ligand binding domain of the TNF-α cell surface receptor, allowing the drug to act as a competitive inhibitor to TNF-α and reduce the immune response (Mitoma et al. 2018). While other mAbs are on the market for TNF-α, Ethanercept remains a widely used drug due to it's extended serum half life and reduced immunogenicity compared to other mAb treatments (Zhao et al. 2018).

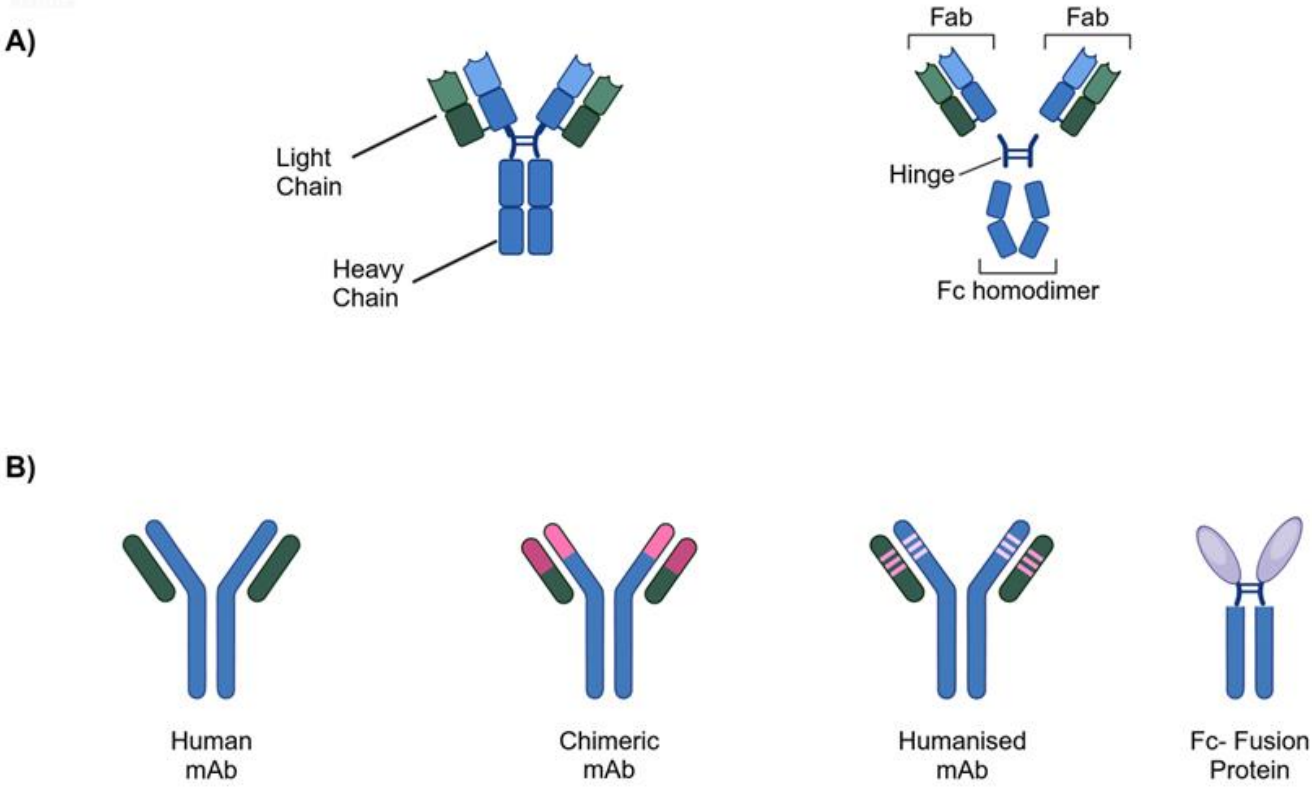


Figure 1.8: Antibody structures used as therapeutic antibodies

A) Structure of a human IgG antibody. The two peptide heavy chains are linked by disulfide bonds are bound by two light chain peptides. This structure can be further broken up into different regions: The Fc fragment, responsible for protein stability and localisation, and the Fab domain which is responsible for antigen recognition. **B)** Examples of therapeutic mAbs. These include fully human mAbs, chimeric mAbs which contain human structure but the variable binding domains from mice, humanised mAbs which only contain the antigen recognition sequences from mice and Fc-fusion proteins which contain an Fc region fused to an effector protein.

1.7.2 CHO cells in mAb production

As antibodies and biologics were approved as a disease treatment in humans, questions were asked of the production process, concerns focussing on the ability to produce sufficient quantities of the drugs in a safe manner. Today, mammalian cells are the most used expression system for drug production, with the Chinese Hamster Ovary (CHO) expression system the most popular. Originally isolated in 1956, the CHO *ori* cell has been adapted to improve protein production (Puck et al. 1958). Many cell lineages have been developed from this original cell line that have been selected for efficient protein production or cell line stability (Tihanyi and Nyitray. 2020). The CHO lineages CHO-K1, CHO-DXB11, CHO-S and CHO-DG44 are all extensively used in biomanufacturing for expression of mAbs. The substantial number of mutations and genetic alterations needed to generate these cell lines combined with the natural plasticity exhibited by the CHO genome result in different behaviours and production rates of these cell lines. For example, the CHO-K1 cell line was shown to produce a higher final mAb titre than the CHO-S line, which favoured production of biomass but had a lower cellular productivity (Reinhart et al. 2019).

The common use of CHO cell lines in mAb production arises for several reasons. Firstly, and arguably the most important reason, CHO expression platforms are capable of high expression productivity, able to produce g/L in culture. The ability to produce high quantities of mAb is important for both drug availability and to reduce the overall production cost of the drug and therefore the final cost on the consumer (Tihanyi and Nyitray. 2020). Production has further been improved by the generation of gene amplification and selection systems, such as the methotrexate (MTX) or glutamine synthetase systems. Both amplification systems require the knockout of an internal CHO enzyme, dihydroflorate refuctase (DHFR) in the MTX system and glutamine synthetase can be inhibited by methionine sulphoximide (MSX). Both systems use expression cassettes which contain the recombinant protein of interest for expression and a gene for the selection marker, in these cases the enzyme being inhibited. This allows selection of cells expressing the target protein and allows further amplification of expression, achieved by increasing the drug concentration which causes cells to increase expression from the cassette in a dose dependent manner to survive drug inhibition, resulting in increased protein production (Schimke, R. 1984, Lai et al. 2013).

Secondly, CHO cell proteins exhibit a similar post-translational protein glycosylation profile as humans. Correct glycosylation of mAbs can affect stability, function, and immunogenicity of the protein. For example, in IgG, the Fc heavy chain regularly undergoes N-linked glycosylation, a common glycosylation modification found in external proteins such as cell surface receptors. This modification is commonly observed in mAb production from CHO cells, lowering the immunogenicity of the produced biotherapeutic in humans (Kim et al. 2012, Zheng et al. 2014). Additionally, CHO cells can be easily scaled up to grow in serum free medium, in suspension and in large cultures allowing greater cellular growth and therapeutic antibody production (Kim et al. 2012).

1.8 Aims of the Project

Alternative splicing or mis-splicing of the pre-mRNA during mAb production from CHO cells poses a significant challenge for biomanufacturing, generating unwanted or aberrant protein structures that require removal from the final product. This increases both the production time and cost of the process for the manufacturer and the consumer. Removal of introns from the expression cassette alleviates these problems at the cost of reduced yield of biotherapeutic protein. This project aims to investigate methods to increase production of a Fc-fusion protein in CHO cells from an intronless sequence.

We examined the effects of increasing/decreasing the GC content of the intronless sequence at both protein and mRNA level, using cellular fractionation and RNA immunoprecipitation (RIP) to characterise changes in mRNA export dynamics. To further improve mRNA processing and protein production from the intronless sequence, we explored the possibility of using known recruitment motifs for SR proteins and SF3b1. Through titre analysis and fed batch cultures of stable CHO cell lines, we compared protein production in an industrial environment and analysed the improvements on expression from an intronless sequence. Finally, we used siRNA silencing to investigate the effects of the HUSH complex on protein production from an intronless sequence.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Bacterial strains

The DH5α *E. coli* strain (Invitrogen) was used for all DNA generation and cloning.

2.1.2 Bacterial growth media

Lysogeny Broth (LB): 10g Tryptone, 5g Yeast Extract, 10g NaCl.

Agar Plates: LB with 2% Agar added.

2.1.3 Primers

Primer name	Sequence	Function
TASOR FWD	5′ TAGCAAATGCTCGGCACTCA 3′	CHO TASOR qRT-PCR fwd primer
TASOR REV	5′ GCTTCACAGGTACCACTGCT 3′	CHO TASOR qRT-PCR rev primer
CH2 FWD	5′ CGAAGACCCTGAGGTCAAGT 3′	Fc-FP mRNA qRT-PCR fwd primer
CH2 REV	5′ CGGTACGTGCTGTTGTACTG 3′	Fc-FP mRNA qRT-PCR rev primer
18S FWD	5′ GTGGAGCGATTTGTCTGGTT 3′	18S rRNA qRT-PCR fwd primer
18S REV	5′ CGGACATCTAAGGGCATCAC 3′	18s rRNA qRT-PCR rev primer
PPHLN1 FWD	5′ TGTGCCAAAGAGACCACCTC 3′	CHO PPHLN1 qRT-PCR fwd primer
PPHLN1 REV	5′ CCTCCTTCATCAGGTCTGGC 3′	CHO PPHLN1 qRT-PCR rev primer
GAPDH FWD	5′ ATTGACCTCAACTACATGGTCTACAT 3′	CHO GAPDH qRT-PCR fwd primer
GAPDH REV	5′ GTTGATGACAAGCTTTCCATTCTCAG 3′	CHO GAPDH qRT-PCR rev primer
5GC FWD	5′ AGCACACAGGACCTCGTACGGCCGCCACCATGGCCATCAT 3′	Used to amplify the 5′ end of the HGC sequence
5GC REV	5′ TCCACGCCGTCCACGTACCAGTTGAACTTGACCTCGGGGT 3′	Used to amplify the 5′ end of the HGC sequence
3GC FWD	5′ CTGTACCTGATCTACGGCCAGGTGGCCCCCAACGCCAACT 3′	Used to amplify the 3′ end of the HGC sequence

3GC REV	5′ TCTAGACCTGCAGGTCATCAGGAGATGAACTG GGGGTTGG 3′	Used to amplify the 3′ end of the HGC sequence
4089 5GC FWD	5′ TGGTACGTGGACGGCGTGGA 3′	Used to amplify the ΔINT-4089 plasmid for 5GC generation
4089 5GC REV	5′ GTACGAGGTCCTGTGTGCTC 3′	Used to amplify the ΔINT-4089 plasmid for 5GC generation
4089 3GC FWD	5′ TGATGACCTGCAGGTCTAGA 3′	Used to amplify the ΔINT-4089 plasmid for 3GC generation
4089 3GC REV	5′ AGTTGGCGTTGGGGGCCACC 3′	Used to amplify the ΔINT-4089 plasmid for 3GC generation
SR FWD	5′ ACAGGACCTCGTACG CACCCGAACCCAAGGAGGACGACCAGGAAA GACGGGAACTCGCGTAGACACGCCCGG CGTACGGCCGCCACC 3′	Used to generate the SR sequence
SR REV	5’ GGTGGCGGCCGTACG CCGGGCGTGTCTACGCGAGTTCCCGTCTTTC CTGGTCGTCCTCCTTGGGTTCGGGTG CGTACGAGGTCCTGT 3’	Used to generate the SR sequence
SF3b (X1) FWD	5′ ACAGGACCTCGTACG GAAGAAG CGTACGGCCGCCACC 3′	Used to generate the SF3b (X1) sequence
SF3b (X1) REV	5′ GGTGGCGGCCGTACG CTTCTTC CGTACGAGGTCCTGT 3′	Used to generate the SF3b (X1) sequence
SF3b (X3) FWD	5′ ACAGGACCTCGTACG GAAGAAG GAAGAAG GAAGAAG CGTACGGCCGCCACC 3′	Used to generate the SF3b (X3) sequence
SF3b (X3) REV	5′ GGTGGCGGCCGTACG CTTCTTC CTTCTTC CTTCTTC CGTACGAGGTCCTGT 3′	Used to generate the SF3b (X3) sequence

2.1.4 Plasmids

Plasmid	Resistance	Source
pCLD-1544	Ampicillin	AstraZeneca
pCLD-4089	Ampicillin	AstraZeneca

2.1.5 Molecular cloning kits

Kit	Supplier	Application
QIAquick Gel Extraction	Qiagen	DNA extraction from gel
QIAprep Spin Miniprep	Qiagen	Small scale plasmid purification
Qiagen Plasmid *plus* Midi Kit	Qiagen	Medium scale plasmid purification

2.1.6 DNA Buffers

5 x TBE: 4.4 M Boric Acid, 4.4 M Tris, 0.01 M EDTA, pH 8.0

2.1.7 Cell lines

CHO: Suspension Chinese Hamster Ovary cell line.

HCT 116: Human colorectal cancer cell line.

HnRNPU - AID: Human colorectal cancer cell expressing hnRNPU – AID tagged protein and constitutively expressed TIR1 protein from *Oryza Sativa*.

2.1.8 Miscellaneous buffers

10 x Phosphate Buffered Saline (PBS): 1.4 M NaCL, 100 mM Na_2HPO_4, 27 mM KCL, 18 mM KH_2HPO_4, adjusted to pH 7.4 with HCL.

2.1.9 Lysis buffer

Cell lysis buffer: 50 mM Hepes – NaOH, pH7.5, 100 mM NaCl, 1 mM EDTA, pH 8, 0.5% Triton – X100, 10 % glycerol, 1 x protease inhibitors, 1 mM DTT.

2.1.10 SDS-PAGE buffers

4 x Acrylamide Resolving Gel Buffer: 1.5 M Tris HCL, pH 8.8, 0.15 % Sodium dodecyl sulphate (SDS).

4 x Acrylamide Stacking Gel Buffer: 0.5 M Tris HCL, pH 6.8, 0.15 % SDS.

4 x SDS loading dye: 200 mM Tris HCL, pH 6.8, 50 % glycerol, 1 % bromophenol blue, 10 % SDS.

SDS page running buffer: 250 mM glycine, 25 mM Tris, 0.1 % SDS.

2.1.11 Western blotting buffers

10 x TG Transfer buffer: 1.9 M Glycine, 250 mM Tris, 35 mM SDS.

1 x TBST: 140 mM NaCl, 20 mM Tris, pH 7.6, 0.2 % Tween- 20.

Blocking solution: 1 x TBST with 5% skimmed milk powder

ECL solution 1: 2.5 mM luminol, 400 µM p-coumaric acid, 100 mM Tris- HCL, pH 8.5.

ECL solution 2: 5.3 mM H_20_2, 100 mM Tris- HCL, pH 8.5.

2.1.12 Antibodies

Primary Antibody	Reactivity	Clonality	Source (catalogue number)
Tubulin	α- Mouse	Monoclonal	Sigma (T5168)
FUS	α- Mouse	Monoclonal	Santa Cruz (sc-47711)
HnRNPU	α- Mouse	Monoclonal	Abcam (3G6)
ALYREF	α- Mouse	Monoclonal	Sigma (11G5)
NXF1	α- Mouse	Monoclonal	Abcam(ab129160)
Periphilin 1	α- Rabbit	Polyclonal	Abcam (ab69569)

Secondary Antibody	Clonality	Source (catalogue number)	HRP conjugate
α-Human	Polyclonal	Sigma (AP113P)	Yes
α-Mouse	Monoclonal	Promega (W4021)	Yes
α-Rabbit	Monoclonal	Promega (W4011)	Yes

2.1.13 Cellular Fractionation buffers

Sucrose lysis buffer: 0.5 M sucrose, 3 mM $CaCl_2$, 0.2 mM $MgCl_2$, 10 mM Tris HCL (pH8), 10 % glycerol.

2.1.14 RNA immunoprecipitation buffers

Fixing solution: 1 % formaldehyde (W/V) - 1 x PBS

Quenching solution: 1 M glycine - 1 x PBS

RIP Lysis buffer: 50 mM Hepes NaOH, 150 mM NaCl, 10 % glycerol, 1 % NP-40, 0.1 % SDS, 0.5 % sodium deoxycholate.

RIP High salt lysis buffer: 50 mM Hepes NaOH, 500 mM NaCl, 10 % glycerol, 1% NP-40, 0.1 % SDS, 0.5 % sodium deoxycholate.

3 x Reverse crosslinking buffer: 3 X PBS, 6 % N-lauroyl sarcosine, 3 mM EDTA, 15 mM DTT.

2.1.15 Polysome gradient buffers

Sucrose gradient buffer: 300 mM NaCl, 15 mM $MgCl_2$, 15 mM Tris pH7.5, 0.5 mM DTT, 0.1 mg/ml Cycloheximide, RNase inhibitors.

Sucrose gradient lysis buffer: 300 mM NaCl, 15 mM $MgCl_2$, 15 mM Tris pH 7.5, 0.5 mM DTT, 0.1 mg/ml Cycloheximide, RNase inhibitors, Protease inhibitors, 0.5 % triton X-100.

10% Sucrose gradient solution: 10 % (W/V) - 1 x Sucrose gradient buffer

50% Sucrose gradient solution: 50 % (W/V) - 1 x Sucrose gradient buffer

2.2 Methods

2.2.1 Plasmid cloning and DNA manipulation techniques

2.2.1.1 Polymerase chain reactions

PCR with Q5 DNA polymerase was performed by mixing the reagents in a suitable PCR tube:

Reagent	Volume
5x Q5 reaction buffer	10 µl
5x Q5 GC enhancer	10 µl
Forward primer	2.5 µl
Reverse primer	2.5 µl
DNTPS	1 µl
Q5 polymerase	1 µl
DNA template	100 ng
Water	Volume up to 50 µl

PCR reactions were carried using the following conditions:

Step	Temperature	Time
Initial denaturation	98 °C	30 seconds
Denaturation	98 °C	10 seconds
Primer annealing	50-72 °C	30 seconds
Extension	72 °C	1 minute per Kb
Final Extension	72 °C	2 minutes

Annealing temperature was dependent on primer pair, calculated using the online NEB tm calculator.

2.2.1.2 Enzyme digestions

All enzymatic digestions were carried out using enzymes from New England Biolabs and used the same standard mix:

Reagent	Volume
10x Cutsmart buffer	2 µl
Enzyme	1 µl
Plasmid	1 µg
Water	Up to 20 µl

All digestions were carried out at 37 °C for 1 hour.

2.2.1.3 Gel electrophoresis

Agarose gel electrophoresis was used to analyse PCR products and enzymatic digestions. Between 0.5 and 2.0 grams of agarose is dissolved in 0.5 TBE buffer to achieve a gel percentage of 0.5 % to 2 %, as determined by the size of the DNA fragment to be resolved. Solution was heated until agarose fully dissolved and allowed to cool to about 50 °C. 1 µl Ethidium bromide was added, and solution poured into a prepared gel cassette. Gels were run at 80V for 45 minutes. DNA fragments were extracted using the QIAquick gel extraction kit following the manufactures instructions.

2.2.1.4 Gibson assembly

Gibson assembly was performed using the 2x NEBuilder Hifi DNA Assembly kit from NEB. All reactions were performed in 10µl total volume with 5µl mix and 5µl of DNA for Gibson in a 7:1 ratio of insert to backbone. The reaction was conducted at 50°C for 60 minutes.

2.2.1.5 Bacterial transformation

For bacterial transformation of cloned plasmids, 100µl DH5α *E.coli* cells were mixed with the entire 10 µl Gibson mix and incubated on ice for 30 minutes. The mix was then heat shocked at 42 °C for 35 seconds and immediately transferred to ice for a further 2 minutes. 900 µl of LB media was added and the cells incubated at 37 °C for 1 hour, shaking at 180rpm. Cells were spun down at 1000 x g for 5 minutes before

removing 900 µl media, the final 100 µl was used for re-suspension of cell pellet and plated onto a prepared LB Agar plate with the correct antibiotic resistance.

2.2.2 Routine cell culture of Chinese Hamster Ovary cells

Chinese Hamster Ovary (CHO) cells were routinely grown at 37 °C, 6 % CO2 and shaken at 120 RPM in 125 ml flasks. Cells were passaged every 3-4 days and seeded at $20x10^4$ cells/ml in 10 ml of CD CHO (Invitrogen) media supplemented with 6mM L-Glutamine.

2.2.3 Transfection methods

2.2.3.1 Chemical methods

For Chemical transfection of CHO cells, Lipofectamine Reagent from Thermofisher was used according to the manufacturer's instructions.

For siRNA transfection of CHO cells, cells were transfected upon reaching 50% confluency using Lipofectamine RNAiMAX from Thermofisher following the manufacturer's instructions. After 48 hours, the transfection was repeated and the cells incubated for a further 24 hours before harvesting.

2.2.3.2 Electroporation

For electroporation, the Neon transfection system from Thermofisher was used according to the manufactures instructions. Neon machine settings were 1700 v, 20 ms, 1 pulse. Settings were determined using the online Neon transfection system cell line data from Thermofisher.

2.2.4 Generation of stable lines

To generate stable CHO pools, the Neon transfection system was used as described above to transfect 10µg of the plasmid containing the expression cassette and a Cre-lox recombinase plasmid in a 2:1 ratio. Cells were allowed to recover for 48 hours before the addition of 250µg of Hygromycin B. Cells were selected for 5 days, before increasing the Hygromycin B concentration to 350µg per ml until the Negative control flask was dead. Cells were transferred and allowed to recover until a cell count of 50 $x10^4$ cells/ml was reached, at which point Hygromycin was reintroduced to the cell media and cell culture would proceed as normal.

2.2.5 Cell lysis

Cells were harvested by centrifugation at 500 x g for 5 minutes, cell media was removed and saved for western analysis. Cell pellet was washed with PBS and spun again. PBS was removed and Immunoprecipitation (IP) lysis buffer (50 mM Hepes ph 7.5, 100 mM NACL, 1 mM EDTA ph 8, 10 % glycerol, 0.5 % Triton - x100 with 1 mM DTT and protease inhibitors added on day of use) was added to the cells and incubated on ice for 5 minutes. Lysate was clarified by centrifugation at 17000 x g for 5 minutes and cellular supernatant moved to a new tube.

2.2.6 SDS page

To generate a 10% polyacrylamide gel for SDS-polyacrylamide gel electrophoresis (PAGE), resolving and stacking gels were made as follows:

Resolving gel:

Reagent	Volume
H_2O	4.2 ml
30% acrylamide	3.3 ml
4 x acrylamide resolving buffer	2.5 ml
10% Ammonium persulfate (APS)	110 µl
Temed	20 µl

Stacking gel:

Reagent	Volume
H_2O	6.3 ml
30% acrylamide	1.2 ml
4 x acrylamide stacking buffer	2.5 ml
10% Ammonium persulfate (APS)	110 µl
Temed	20 µl

SDS-PAGE samples were prepared by loading the required protein concentration, usually between 10 – 20 µg, and made up to 20 µl with water and 4 x loading buffer. Denaturation of the protein was carried out at 95 °C for 5 minutes just before loading. SDS-PAGE was performed in a BIO-RAD electrophoresis chamber using a hand cast 10% acrylamide gel and run at 20mA until resolved. Gels were then prepared for western blotting.

2.2.7 Western blotting

Gel transfers were prepared in the following manner: 6 layers of WypAll paper soaked in transfer buffer were placed on the transfer plate, then a soaked nitrocellulose membrane on top, then the gel is placed flat on top of the nitrocellulose before closing the sandwich with 6 more layers of soaked WypAll paper. Transfers were conducted using a current of 25 V, 1.3 mA for 20 minutes in a Trans-Blot Turbo Transfer System. Transferred western blots were stained using Ponceau S solution for 30 seconds, before rinsing with water to remove the stain and cut accordingly. The blot was then blocked for 1 hour using TBS milk with added tween. Milk was refreshed and relevant antibody was added and incubated shaking for 1 hour at room temperature. The primary antibody was washed using 3 x 30 second washes using TBST and 1 x 10 minute was using TBST. Secondary antibody was added to TBST milk and the blot incubated shaking at room temperature for 1 hour. The previous wash step was repeated before equal parts ECL 1 and 2 were added to react with the Horse radish peroxidase bound to the secondary antibody. Blots were visualised in a BIO-RAD ChemiDoc imaging system.

2.2.8 RNA extraction from whole cell lysate/pellets

Cell pellets or Cell lysates were resuspended in 750 µl TRIzol/TRIzol-LS respectively. Samples were incubated at room temperature for 15 minutes and followed by addition of 200 µl of chloroform, shaking for 30 seconds and then incubated at room temperature for 10 minutes. Samples were spun down at 12000 x g for 15 minutes at 4 °C before the upper phase transferred to a fresh tube. Equal volumes of isopropanol and 5 µg of GlycoBlue (Invitrogen) was added and incubated at -20 for at least 10 minutes. Samples were then spun down at 12000 x g for 30 minutes, the supernatant removed, and the pellet washed with 1ml of ice cold 70\% ethanol and spun for further 5 min at 7500 x g at 4 °C. The pellet was then air dried for 15 minutes.

2.2.9 DNAse treatment of RNA

Air dried pellets were resuspended in 43 µl of ice cold MilliQ water, vortexed for 10 seconds and incubated on ice for 10 minutes. 5 µl of 10 x turbo DNAse buffer (Invitrogen), 40 U of RNASE inhibitor (Meridian) and 1 µl of Turbo DNAse were added to the resuspended RNA and incubated at 37 °C shaking at 300rpm for 1 hour. Following treatment, 50 µl of ice cold MilliQ was added followed by 100 µl of acidic

(pH 5.3) phenol. Mixture was shaken and spun down at 4 °C at 12000 x g for 5 minutes. Upper phase (90 µl) was added to 10 µl 3 M NaAc, 5 µg of GlycoBlue and 250 µl of 100 % EtoH. Samples were incubated overnight at -70 °C.

2.2.10 Generation of cDNA

1 µg of purified RNA was added to a 17 µl mix containing 4 mM DNTPs and 10 x Random Hexamers mix from Thermofisher. Denaturing of the mix occurred at 70 °C for 5 minutes followed by a 5-minute incubation on ice. A master mix containing 2 µl of 10 x Reverse transcriptase buffer, 0.5 µl of RNASE inhibitor (Meridian) and 0.5 µl of Reverse transcriptase (Thermofisher) was added for a final volume of 20 µl. The cDNA reaction was incubated for 10 minutes at 25 °C, then incubated at 37 °C for 120 minutes and finally at 85 °C for 5 minutes.

2.2.11 Quantitative Real Time – PCR (qRT-PCR)

Generated cDNA was diluted 10 x with H_2O before addition to the following reaction mix:

Reagent	Volume
cDNA	4 µl
Primer mix (5µM)	1 µl
2 x Sensimix	5 µl

cDNA for 18s levels was diluted an additional 100 x before addition into the qRT-PCR reaction mix.

The qRT-PCR reaction was carried out in a Corbett Rotor-Gene 2000 instrument for 50 cycles with the below settings:

Temperature (°C)	Time (seconds)
95	10
60	15
72	25

2.2.12 Fractionation

A confluent 6 cm dish of cells were harvested and transferred to a 15ml round-bottom falcon tube and spun down at 500 x g for 3 minutes at room temperature. The pellets

were washed with 5 mls 1X PBS and spun down at 350 x g for 3 minutes at room temperature. PBS was removed and pellet transferred to ice cold 2 ml tube. Cytoplasmic fraction was lysed using 7 x cell volume of Sucrose Lysis Buffer (SLB) with additional RNAse inhibitors, 1 mM DTT and 1 x protease inhibitors. Cell pellets were incubated on ice for 5 minutes before spinning down at 500 x g for 5 minutes at 4 °C. The supernatant was harvested and collected as the cytoplasmic fraction. The remaining cell pellet was washed 2 x in SLB, spinning at 300 x g for 5 minutes, then washed once with 0.5 M sucrose in 1 x PBS spinning at 300 x g for 3 minutes. Resulting nuclei pellet was resuspended in SLB and aliquoted off to harvest nuclei for protein analysis, the remaining pellet was resuspended in TRIzol for RNA extraction.

2.2.13 Cell cross linking

A confluent 6 cm dish was harvested and the pellet collected by centrifugation (500 x g for 3 minutes) and washed with 1x PBS. Pellets were cross-linked by incubation at room temperature in 1x PBS – 1% formaldehyde for 10 minutes. Reaction was quenched with addition of 1 M glycine and incubated for 5 minutes. Pellets were washed 2 times with 5 ml PBS before freezing and storage at -80 °C.

2.2.14 RNA immunoprecipitation

Protein G dynabeads (ThermoFisher) were prepared by washing twice with RIP lysis buffer and then incubated for 2 hours at room temperature in lysis buffer with 5 µg antibody and 1 % BSA. Cell pellets were lysed in 400 µl of RIP lysis buffer with 1 mM DTT, 1 x protease inhibitors, RNAse inhibitors and Turbo DNAse added. Cells were incubated on ice before lysis using a Biorupter Pico using High Amplitude, 30s on - 30s off, 5 cycles. Lysates were collected and spun at 16100 x g to pellet debris. Beads were washed 2x with 1 ml RIP lysis buffer following antibody binding and then 300 µl of cell lysate was added to the beads and incubated at 4 °C for 2 hours rotating. 10 % volume cell lysate was saved as input. Following incubation, beads were washed 2 x with RIP lysis buffer, then 2 x with RIP lysis buffer containing 500 mM salt then finally 2 x with RIP lysis buffer. Reverse cross linking of inputs and eluates was performed by addition of H_2O and 3 x reverse crosslinking buffer: 3X PBS, 6 % N-lauryl sarcosine, 30 mM ETDA, 15 mM DTT and 1.9 mg/ml Proteinase K up to a total sample volume 100 µl. Samples were shaken at 1100 rpm at 42 °C for 1 hour and then 55 °C for 1 hour. Inputs and collected eluates were topped up to 250 µl with ice cold H_2O and then

mixed in TRIzol Ls. RNA was then extracted as stated previously. The entire input and eluate were used in the cDNA synthesis reaction.

2.2.15 Polysome Profiling

2.2.15.1 Sucrose gradient preparation

To prepare a sucrose gradient, 25 mls of 10 % and 50 % Sucrose gradient buffer solutions containing DTT and Cycloheximide were prepared at room temperature. Beckman ultra clear centrifugation tubes were labelled using a SW41 marker block. 10 % sucrose solution was added until the line for short gradient previously marked was reached. Using a 6-inch steel syringe needle, 50 % sucrose solution was then added to the gradient, at the bottom of the tube beneath the 10% sucrose solution. Solution was added up to the marked line. Tubes were capped, ensuring no air bubbles, and were mixed using a Biocomp gradient maker using the SW41 10 – 50 % sucrose setting.

2.2.15.2 Polysome profiling

$20x10^6$ cells per gradient were harvested from culture and treated with 0.1 mg/ml cycloheximide for 5 minutes at 37 °C. Cell pellets were harvested by centrifugation at 180 x g for 5 minutes , the supernatant discarded and the pellet washed in cold 1 x PBS containing for 0.1 mg/ml cycloheximide before centrifugation at 180 x g for 5 minutes. Pellets were resuspended in sucrose gradient lysis buffer and incubated on ice for 5 minutes before centrifugation at 5500 x g for 5 minutes at 4 °C. 50 µl of supernatant was harvested as a total cytoplasm control, the remaining 450 µl of supernatant was gently layered on top of a prepared sucrose gradient. Gradients were balanced as necessary using cold ultrapure H_2O, before centrifugation in a SW41 Beckman rotor for 2 hours at 36000 rpm at 4 °C. Following centrifugation, fractions were harvested using a Brandel gradient fractionator into 125 µl aliquots in prepared Eppendorf tubes.

2.2.16 Titre analysis

The cell culture medium was clarified by centrifugation and recombinant proteins were quantified by using a protein A high-performance liquid chromatography (HPLC) on an Agilent HP1100 or HP1200 (Agilent Technologies, Santa Clara, CA) by comparing

peak size from each sample with a calibration curve (Daramola et al. 2014) in which Protein A binds to the CH3 region of the Fc-domain of the Fc-fusion protein.

2.2.17 Fed-batch overgrow

For fed-batch overgrow analysis, CHO cultures stably expressing recombinant protein were seeded and cultured for 10-day. Cultures were maintained using AstraZeneca proprietary medium and bolus additions of nutrient feed. Glucose and lactate levels were monitored throughout the process (YSI 2900D, YSI Inc).

2.2.18 Statistical analysis

For statistical analysis of all qRT-PCR, titre data and specific productivity data an unpaired t-test was used to analyse differences between data sets. All error bars shown represent standard deviation of the data set. The significance as determined by the p value is as follows:

P value	Significance
P value > 0.05	Not significant (ns)
P value < 0.05	*
P value < 0.01	**

2.2.19 Figure generation

Biorender.com was used for the generation of all illustrations and schematics.

Chapter 3: The effect of GC content on Fc-Fusion Protein production from intronless genes.

The work presented in this chapter focuses on altering the GC content of the Coding Sequence (CDS) of an intronless sequence to analyse its subsequent expression patterns. We looked to improve protein production by increasing the GC content of an intronless gene encoding an Fc – fusion protein to assess the effects on expression. We used western blotting from both transient experiments and stable cell lines to show an increase in intronless production when the GC content is increased. Additionally, we used cell fractionation and RNA immunoprecipitation (RIP) to suggest an increase in mRNA export due to export factors preferentially binding intronless RNA with a higher GC content. Finally, we mimicked industrial production to show increased Fc – fusion protein production over time.

3.1 Experimental design of High GC constructs

To improve production of therapeutic protein expression in mammalian systems, every aspect of the protein expression pathway has been scrutinized. Protein glycosylation profiles, growth rates, media components, transfection methods, protein quality control have all been researched and identified as areas of recent improvement or to be improved (Zhu, 2012). Codon optimisation emerged as a major design factor in current mAb production with its potential to be a bottleneck identified early in the process of optimising expression in a host system (Gustafsson et al. 2004). Despite codon optimisation being the major consideration in codon selection and coding sequence design, we wanted to consider other effects that nucleotide sequence could have on expression to improve therapeutic protein expression from an intronless sequence.

It is well established that intronless cDNAs tend to be expressed at lower levels than their spliced counterparts and this has been attributed to relatively poor nuclear export of intronless mRNAs (Valencia et al. 2008). A major consequence of splicing comes from the EJC, which is deposited on the nascent RNA anywhere from 50 to 5 nucleotides upstream of the splice site and is locked in position by the core EJC component EIF4A3 (Le Hir et al. 2016). iCLIP studies have revealed that ALYREF and EIF4A3 bind in close proximity to each other upstream of exon junctions. Additionally, knockdown of EIF4A3 reduces ALYREF binding to both intron containing and, contrary

to expectations, naturally intronless mRNAs, demonstrating the ability of EIF4A3 to assist in the recruitment of ALYREF to RNA co-transcriptionally (Viphakone et al. 2019). Natural intronless mRNAs appear able to recruit both EIF4A3 and ALYREF sufficiently for their export which suggests they have evolved strategies for such recruitment in the absence of a splicing dependent mechanism. Such a mechanism may be inefficiently reproduced when intronless cDNAs are used for expression.

Knockdown of TREX components has been shown to have a significant effect on the export of both intron containing and single exon genes, especially those that have a higher GC content (Zuckerman et al. 2020). Recent work has shown that in a library of intronless GFP variants, the sequences with increased GC content over the full length of the gene or the 5' end have increased expression when compared to the AT rich variants (Mordstein et al. 2020). They also reported that expression of the AT variants could be improved with the inclusion of an intron in the 5'UTR or in the coding sequence but this had less effect on the sequences with high GC content (Mordstein et al. 2020).

All this published data leads us to the hypothesis that increasing the GC content of an intronless nucleotide sequence will improve its expression by having positive effects on mRNA export, potentially through interactions with ALYREF and the TREX complex. To increase the GC content of a therapeutic protein without affecting the coding sequence, we adopted the same method that was used by Mordstein et al to generate their intronless GFP library (Mordstein et al. 2020). This method focuses on altering the final base of each codon, allowing us to maintain the integrity of the CDS but increase the GC content (Figure 3.1). Industrially produced proteins are usually highly codon optimised so we tried to produce 2 constructs, one construct containing as many GC alterations as we could make to the sequence, the other only containing changes that had minimal disruption in codon usage.

However, the constructs that we designed proved to be very difficult to synthesise, likely due to the existing high GC content of the starting intronless Fc-FP (60%). This meant that we could only get the construct with as many changes as possible. synthesised as the codon optimised failed during synthesis. However, the codon optimality of the CHO genome is GC favoured so effects on expression from sub –

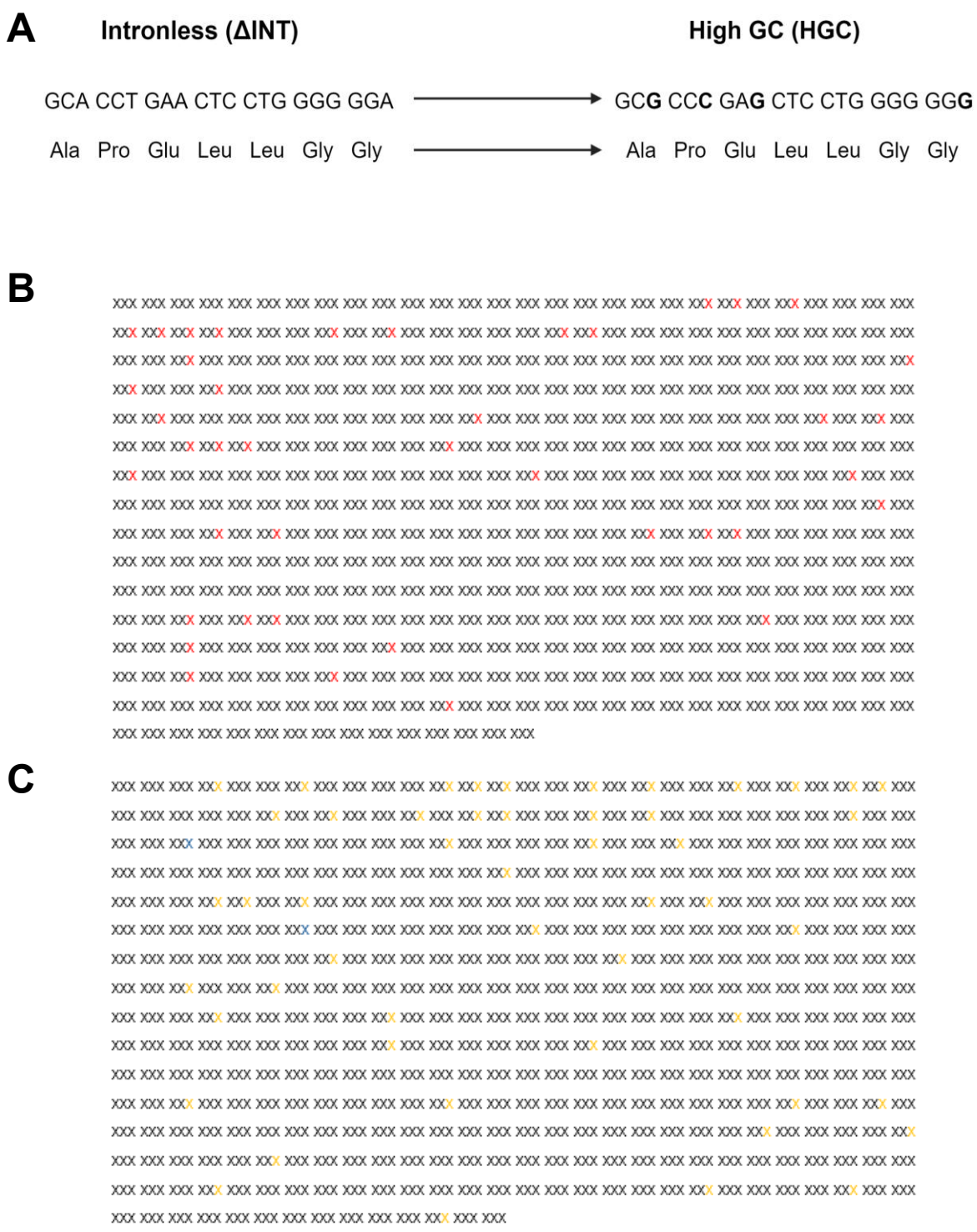


Figure 3.1. Design Schematic of High GC design.

A) Schematic of how the High GC construct was designed, with alterations in GC content achieved through changing the 3rd base in each codon. This allows changes in the GC content without altering the protein sequence. **B)** Schematic of the ΔINT sequence, bases highlighted in red have been changed to a G or C in the HGC. **C)** Schematic of the ΔINT sequence, bases highlighted in yellow have been changed to a A or T in the low GC (LGC) sequence.

optimal codon usage should be minimal. Using this method, we increased the overall GC of the coding sequence from 60% to 64%.

Alongside this, we generated another sequence where we lower the GC content by changing the nucleotides to A/T, which we hypothesise will reduce expression further to use as a control for expression.

3.2 Increasing GC content shows potential to improve intronless protein and RNA expression in a transient system.

Firstly, we used transient transfections to make sure that the Fc-FP sequences we had cloned, as well as the newly designed High GC sequence, expressed correctly. We found that chemical transfection using PEI and Lipofectamine resulted in poor transfection efficiency in CHO cells, so we used electroporation with the NEON transfection system. We transiently transfected the 'Wild Type' (WT) which contained two introns and the intronless (ΔINT), which was the same sequence but with the introns removed, into CHO cells (Figure 3.2.B). Our Fc-FP constructs contained an N-terminal signal peptide to allow secretion from the cell, to make our results more applicable to industrial manufacturing. This meant that we could use western blotting to look for protein in the cell media (CM) and cellular lysate. Alongside our constructs we co-transfected a plasmid containing a FLAG-tagged GFP to act as a transfection marker for two reasons: firstly, to allow quick identification of successful transfections using fluorescent microscopy and secondly to check how similar the transfection efficiencies were between plasmids via western blot analysis. As we expected, the intronless sequences had lower levels of protein expression than the WT containing introns in the cell media and lysate, confirming the long-observed result that the removal of introns does reduce protein expression (Palmiter at al. 1991).

To test the effect GC content had on intronless Fc-FP expression, we transfected our Higher GC content (HGC) sequence into CHO cells alongside the previous two constructs. We observed more protein expression in the HGC than the intronless control indicating that increasing the GC content does have a positive effect on protein production (Figure 3.2.B). However, the differing levels of GFP between transfections make it difficult to say with real certainty how much better the expression is. A better method would have been to insert the GFP construct into the same plasmid that our

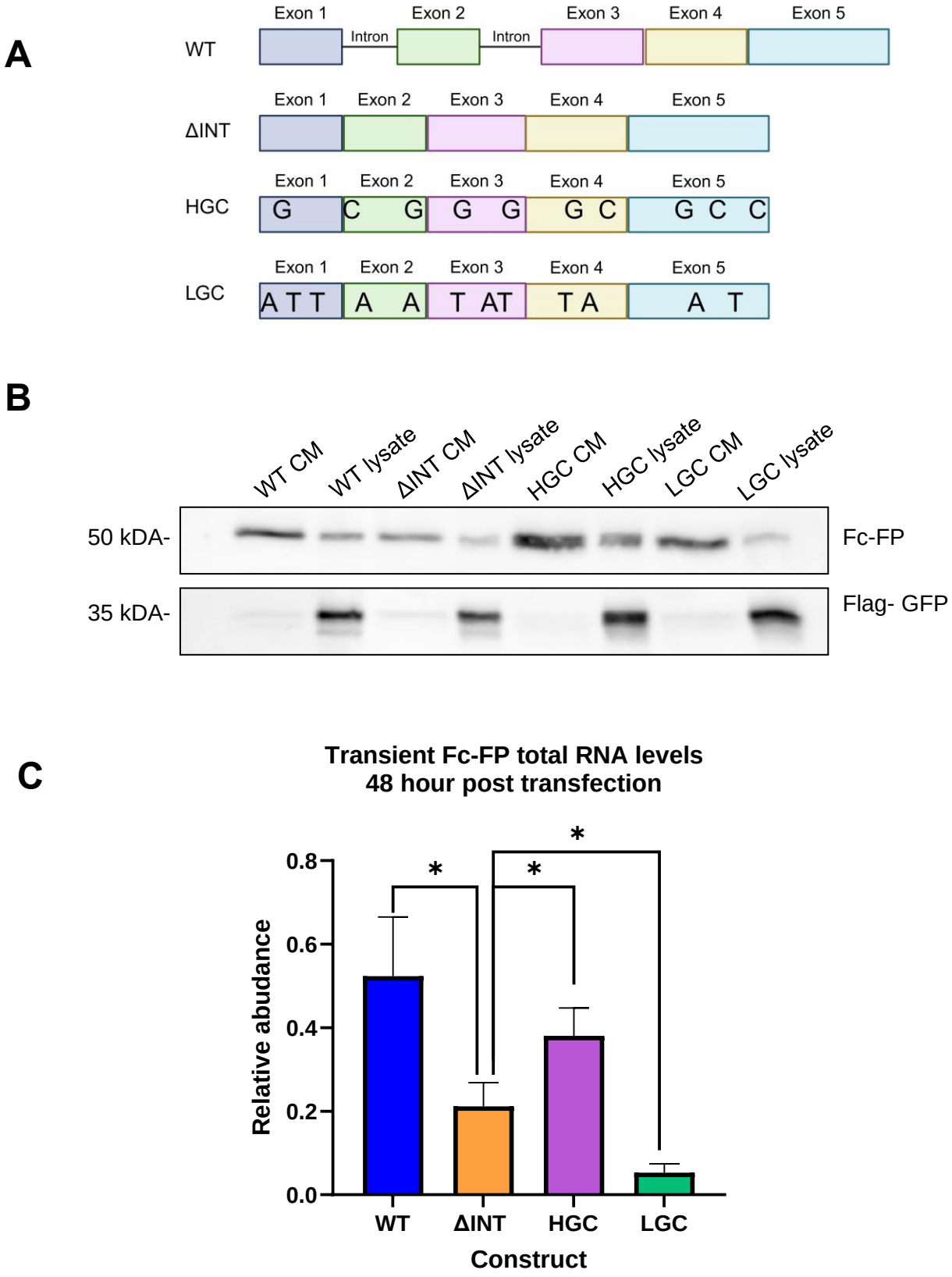


Figure 3.2: Transient expression of Fc- FP in CHO cells.

A) Schematics of Fc-FP constructs transfected. B) Western blot showing protein expression 48 hours post transfection in the cell media (CM) and cell lysate. Flag- GFP is shown as transfection control. C) qRT-PCR panel of transiently expressed Fc-FP RNA levels 48 hours after transfection (n = 3). All samples were normalised to the level of 18S rRNA.

sequence was being expressed from to allow an accurate comparison of plasmid transfection. Surprisingly, we observed increased protein expression from our lowered GC content construct in the CM but lower amounts in the cell lysate which may indicate that this modification had somehow altered the ability to secrete the protein, though how a change in the mRNA sequence would alter protein secretion is unclear. Alternatively, the effects observed with the lowered GC content may have resulted from the modestly increased transfection efficiency as indicated by the GFP levels.

To establish whether differences in protein expression could be accounted for by changes in mRNA levels, we examined total RNA samples from transfected cells using quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) (Figure 3.2.C). We found reduced mRNA levels in the ΔINT when compared to the total mRNA from the WT, but that increasing the GC content improves the expression from an intronless sequence. This is additionally supported by the reduction on mRNA levels seen in the cells transfected with the Low GC (LGC) sequence. Thus, increasing the GC content influences the RNA expression pathway resulting in improved expression of therapeutic protein.

3.3 High GC may affect export of RNA to the cytoplasm by recruiting export factors.

One possible reason for reduced total mRNA levels with lower GC content may be lower stability of the mRNA. This may arise due to inefficient export of the low GC rich RNA and nucleoplasmic surveillance and destruction by the exosome. Consistent with this prediction, Mordstein et al (2020) reported an increase in cytoplasmic localisation of GC mRNA when compared to AT rich mRNA sequences. To test this idea, we isolated nuclear and cytoplasmic RNA fractions. Originally, we fractionated the whole nuclei into nucleoplasm/chromatin fractions, however as the expression constructs were transiently expressed and not integrated into the chromatin, we debated if analysing the chromatin fraction was necessary, resulting in our decision to use whole nuclei. This would also be sufficient to examine the movement of RNA to the cytoplasm. By analysing the ratio of RNA in the cytoplasm to that of the nucleus, we saw an increase in the export of intronless Fc-FP RNA when the GC content is increased (Figure 3.3.A). The main mRNA export pathway utilises co-transcriptional loading of the export adaptor ALYREF across all stages of RNA processing, including

capping, splicing and Poly(A) tailing (Viphakone et al. 2019). ALYREF acts as an export adaptor in tandem with members of the THO complex and DDX39 (forming the core of the TREX complex) to recruit NXF1, although other proteins can act as export adaptors such as CHTOP (Heath et al. 2016). We hypothesised that an increase in ALYREF recruitment to HGC mRNA could account for the increase in export to the cytoplasm. To investigate this, we performed an ALYREF RIP to evaluate ALYREF binding to our transfected constructs. In support of our hypothesis and our previously generated fractionation data, we see improved binding of ALYREF to HGC RNA when compared to the ΔINT. The ΔINT RNA still showed some ALYREF binding, however this is to be expected since we see protein expression and RNA export. Additionally, Viphakone et al (2019) reported that ALYREF was still recruited to intronless RNA in humans. We detected significantly more ALYREF loading onto the HGC RNA, suggesting that ALYREF prefers bind to RNA with increased GC content.

While enhanced ALYREF binding may lead to improved mRNA export, the NXF1/NXT1 (also known as TAP: p15) heterodimer is responsible for the chaperoning of RNA across the nuclear pore. ALYREF, alongside THOC5, expose the RNA binding domain of NXF1 and facilitate its handover onto mRNA (Viphakone et al. 2012). Hence, we performed an NXF1 RIP to examine the amount of NXF1 bound to the different Fc – FP sequences. Surprisingly, we found that significantly less spliced WT RNA bound to NXF1 than the ΔINT, showing an inverse result to the ALYREF RIP (Figure 3.3.C). The HGC RNA was bound slightly less to NXF1 than the ΔINT but not to a significant degree. This suggests that RNAs that are inefficiently exported remain bound to NXF1 for a longer time than those RNAs that are efficiently shuttled to the cytoplasm, where NXF1 dissociates from the mRNA. Thus, we propose that increasing the GC content improves intronless biotherapeutic production which may be due to enhanced recruitment of the mRNA export factor ALYREF.

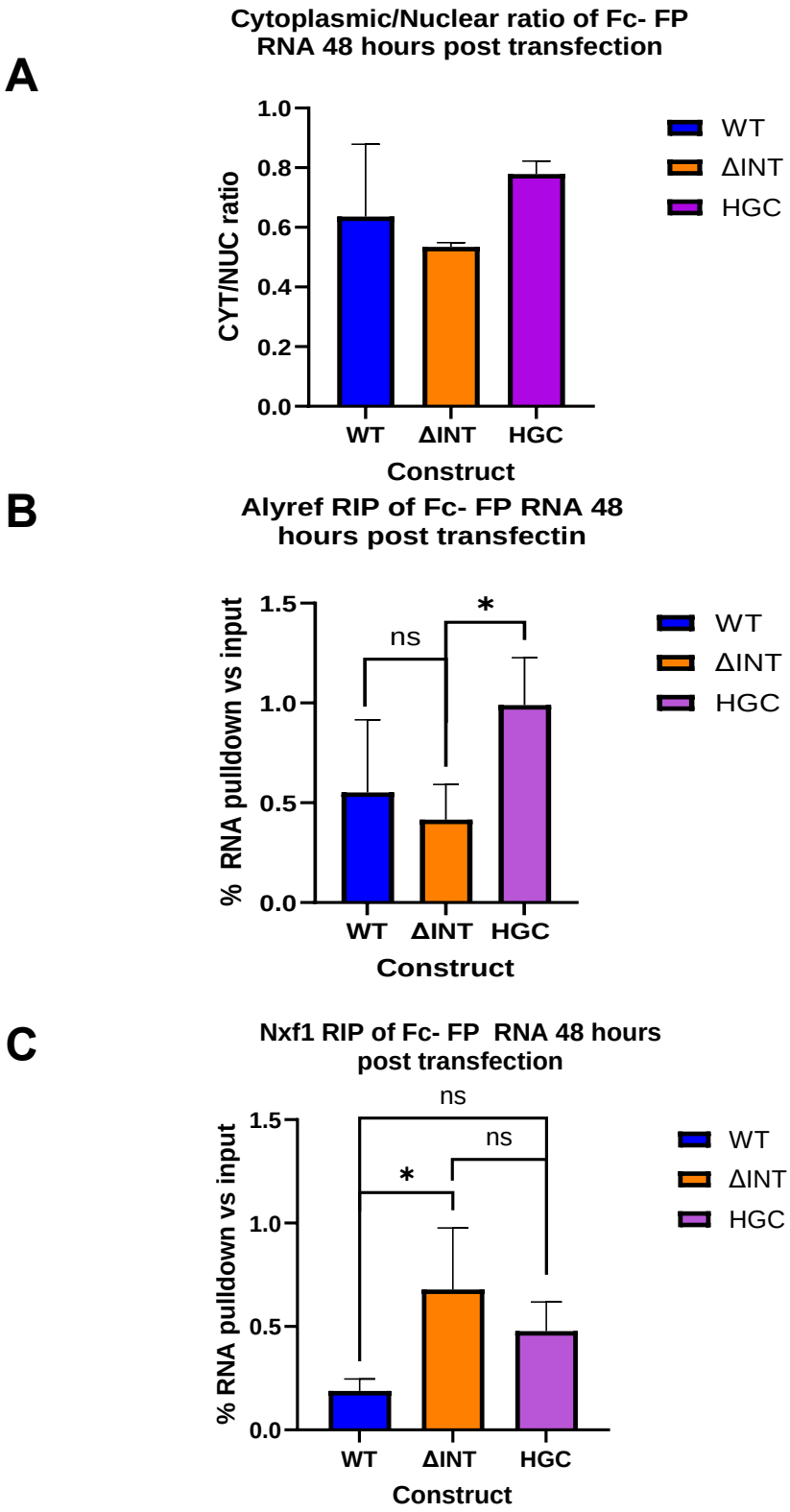


Figure 3.3: The effect of increased GC content on RNA export.

A) qRT-PCR panel of CYT/NUC ratio 48 hours after transfection (n=2). Factions were normalised to GAPDH mRNA. **B)** Pulldown efficiency of Fc-FP RNA in a ALYREF RIP (n=3). RNA was normalised to 18S rRNA levels from the input sample. **C)** Pulldown efficiency of Fc – FP RNA in a NXF1 RIP (n=3). RNA was normalised to 18S rRNA levels from the input sample.

3.4 Expression of intronless biotherapeutic protein in a stable system

Monoclonal antibody production in CHO cells is done using fed-batch systems, usually taking place in bioreactors. Nutrient rich feed is added to cell cultures after cells have been grown for a few days in normal growth media to prolong cell survivability and increase the yield of therapeutic protein produced (Xu et al. 2023, Schellenberg et al. 2022). To make our work more representative of industrially used growth system, we developed stable pools of CHO cells expressing the same Fc-FP sequences that we tested transiently.

A cre-LoxP system was used to integrate our sequences into cells to ensure that the Fc- FP sequences were integrated into the same genomic locus and successful clones were selected with hygromycin. Initially, this was done with the WT and ΔINT Fc-FP to establish control lines we would use for future experiments. This generated cell pools expressing our Fc-FP sequences which we then used to select cells and build clonal cell lines. Upon testing protein expression from the single clones via western blot analysis, we discovered that an unexpected protein was present in the WT cell line (Figure 3.4.B). We theorised that this came from a mis-splicing event in which the first 5' splice site of intron 1 was used alongside the 3' splice site of intron 2, leading to the excision of exon 2. (Figure 3.4.A). This would result in the aberrant protein of reduced molecular weight we see in our western blots. This mis-splicing was present in every single clone we tested, indicating that it was a common splicing issue with the WT construct. However, this mis- splicing issue provides biological context for the project and demonstrates the problem facing biomanufacturing. We saw no evidence that this aberrant protein was capable of being secreted from the cell, nevertheless we decided to stop using this line as a control as we could not be sure how much the protein levels are affected by this mis-splicing event making comparison to other cell lines difficult. The development of the ΔINT line would be sufficient to act as a control line to compare expression of any other stable lines we generated. Additionally, we took the decision to use stable pools of cells rather than single clones as the targeted integration should ensure that all cells have similar expression levels (Zeh et al. 2024). This decision was also taken due to time constraints as single cell recovery took a long period of time.

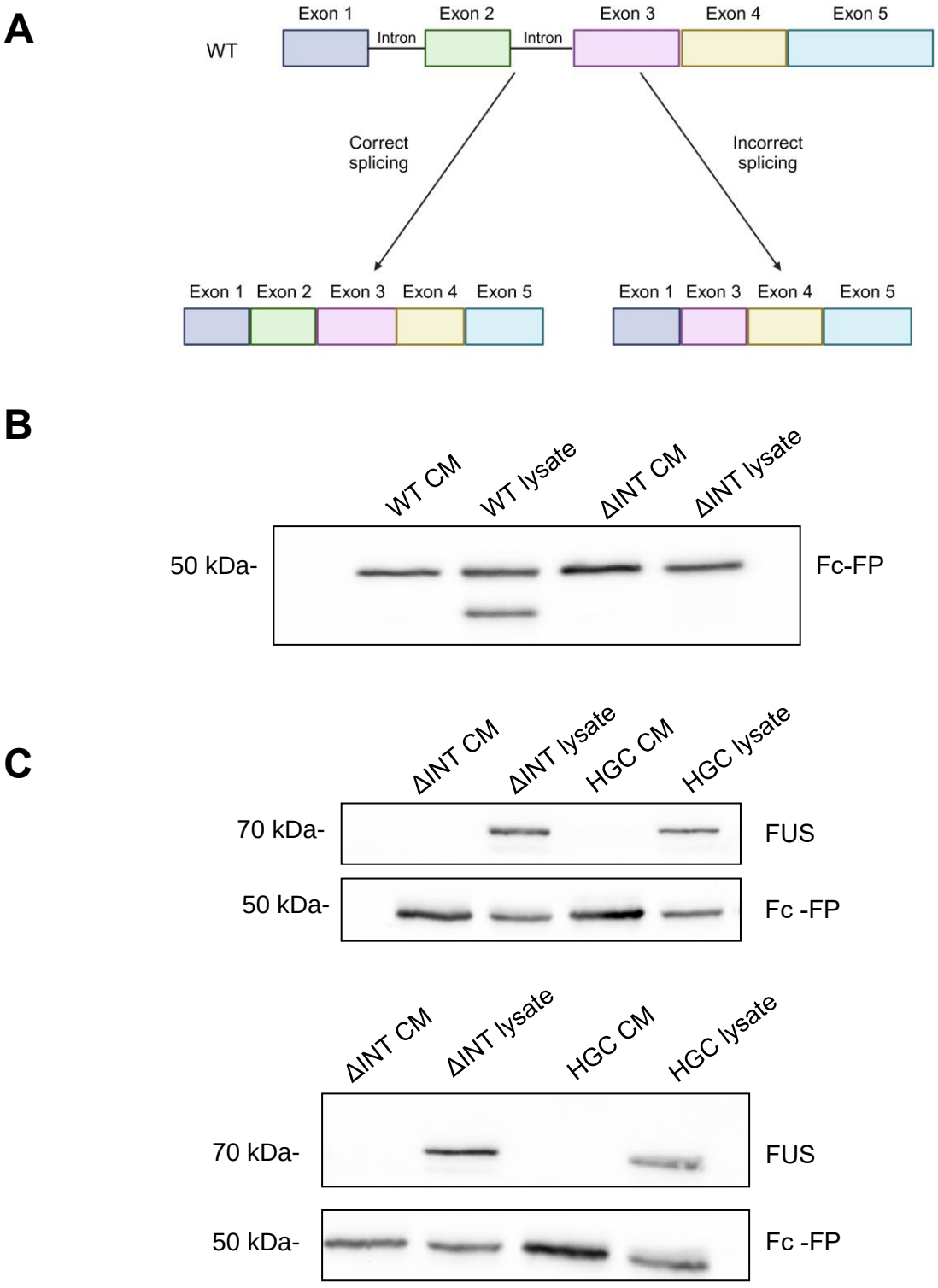


Figure 3.4: Expression of Fc-FP protein in a stable system.

A) Schematic showing the splicing defect present in our WT Fc-FP cell line when expressed from a stable cell. **B)** Western blot showing the splicing defect present in the WT lysate. **C)** Western blots comparing Fc-FP production from the ΔINT and HGC stable cell line. FUS is shown as a loading control for cellular lysates. 2 separate replicates shown. CM samples were loaded equally by volume, cellular lysate samples were loaded according to protein concentration.

We then generated a clonal CHO pool expressing the HGC Fc-FP sequence to compare against the ΔINT cell line. We seeded cells in dishes and let them grow for 48 hours in a similar manner to our transient experiments. Western blot analysis (Figure 3.4.C) showed that the HGC line either expresses similarly or better than the ΔINT cell line. Differences in the cell media, where equal volumes of media were loaded onto the blot rather than equal protein, show the largest change between the two cell lines.

3.5 Increasing GC content effects RNA expression in a stable context.

As we see improved protein expression in our stable pools, mirroring what we saw from the transient data, we hypothesised that the effects we saw transiently at the RNA level would also be seen in the stable pools. We seeded cells in dishes for 48 hours and isolated total RNA from the cell pellets. Interestingly, and in contrast to the transient data, we could detect no significant differences in the levels of mRNA being produced between the ΔINT line and the HGC Fc-FP cell line (Figure 3.5.A).

The lack of differences in RNA levels could be for a variety of reasons: The level of Fc-FP RNA in these cell lines is significantly higher than the levels we could detect during our transient experiments, especially for the intronless construct, therefore any slight increases in mRNA levels in the HGC line may be masked by this increase in ΔINT expression. We also hypothesised that increased therapeutic expression could be due to the differences in the distribution of the mRNA, a theory supported by our transient work showing increase export and export factor binding.

To test this, we performed cytoplasmic/ nuclear fractionation to look at the distribution of the RNA. As we expected, and in support of our transient data, we see improved export of the high GC Fc-FP RNA to the cytoplasm when compared to the ΔINT control line (Figure 3.5.B). Thus, we see improved expression of an intronless sequence when the GC content of that sequence is increased because of an increase of mRNA export to the cytoplasm. Due to time restrictions, we were not able to analyse the binding of export factor proteins such as ALYREF in a stable line, however increased export factor binding would provide a molecular basis for this improved expression. As we continue to see an increase in protein production and mRNA export to the nucleus as

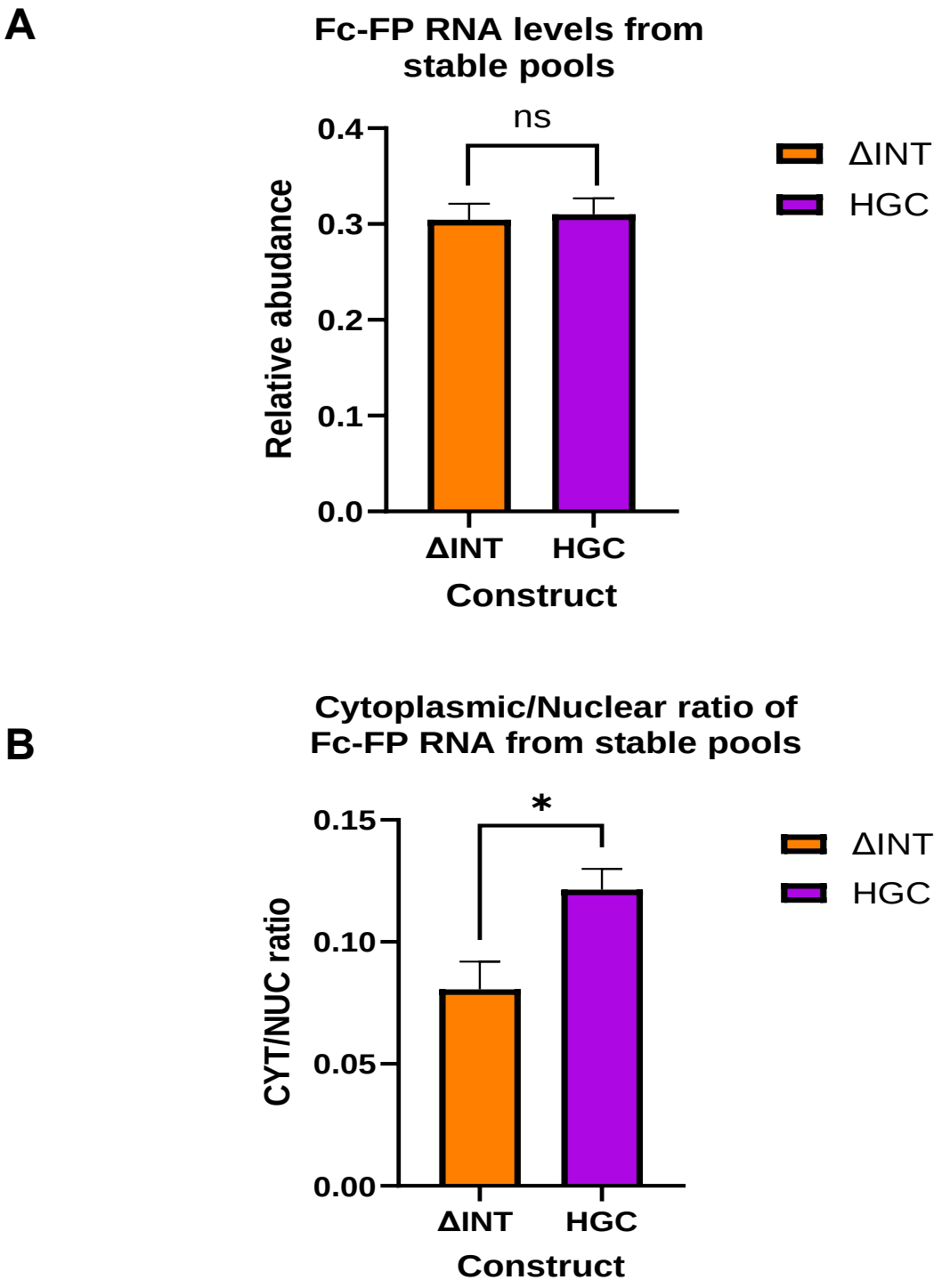


Figure 3.5. Fc-FP mRNA levels in stable CHO pools.

A) qRT-PCR of Fc-FP RNA levels after 48 hours of growth in stable pools of CHO cells (n= 4). All samples were normalised to the levels of 18S rRNA. **B)** Cytoplasmic/Nuclear ratios of Fc-FP RNA after 48 hours of growth in CHO cell stable pools (n= 5). All samples were normalised to the levels of GAPDH RNA.

we did transiently, we expect that we would see an increase in ALYREF binding to RNA with increased GC content.

3.6 Increased GC content shows potential to improve Fc-FP production over longer time periods

As this work hopes to improve production from intronless constructs in an industrial environment, we simulated the initial stages of biotherapeutic production by growing up our stable pools over longer time periods. Cells were seeded at a known density and allowed to grow under regular growth conditions for 6 days. While this time point is still shorter than would usually be analysed in industry, this was the longest time point we could do before cell viability dropped off without using nutrient rich feed media. Cell counts, cell media and cell pellets were taken every 2 days to analyse the cell growth and protein production over time.

Western blotting was used to analyse the production of Fusion protein by analysing levels in the cell media. Samples for both the ΔINT and HGC cell lines were taken at 2, 4 and 6 days (every two days to allow changes to be detectable via western blot anaylsis) and run alongside each other (Figure 3.6). Equal volumes of cell media were loaded onto the western rather than equal protein amounts so that we can analyse total production of protein. While there is variability between replicates, which is to be expected given that these are pools of cells, at the 6-day mark the HGC line had produced more protein than the ΔINT cell line we compared it to. An increase was also visible in some replicates at the 4-day mark, but this increase is less obvious. Interestingly, the HGC line appeared to grow slower than the INT line and cell growth was stagnant by the 6th day. This could be an indicator that the increased production of biotherapeutic protein puts an additional strain on the cells that could be an interesting point of further investigation.

The same cell media samples from the 6-day batch cultures were then sent to our industrial partners at AstraZeneca. Using protein A High-performance liquid chromatography (HPLC), they measured the protein titre from these cultures (Figure 3.7). We see a slight increase in protein titre in the HGC compared to the ΔINT at the 6-day mark in each replicate, however the differences between replicates makes it difficult to confirm this increased production with any degree of significance. We also see no differences in protein titre at the 2- or 4-day mark, matching our western

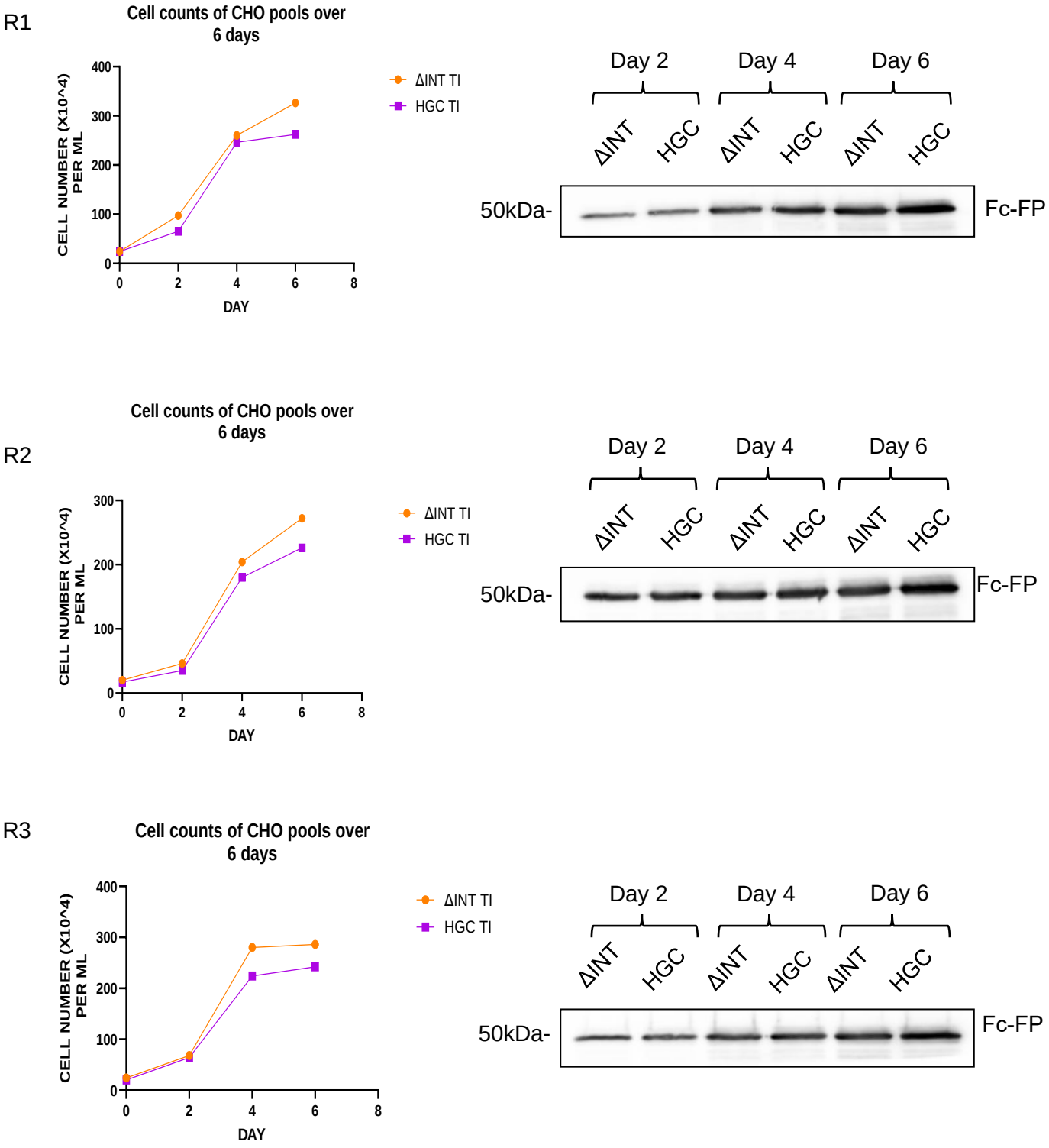


Figure 3.6: Production of Fc-FP protein over 6 days from stable lines.

Figure 3.6 shows 3 separate replicates of protein production from ΔINT and HGC cell pools. On the left is the cell count, taken every two days, which corresponds to the western blot to its right, which shows protein expression in the cell media over 6 days.

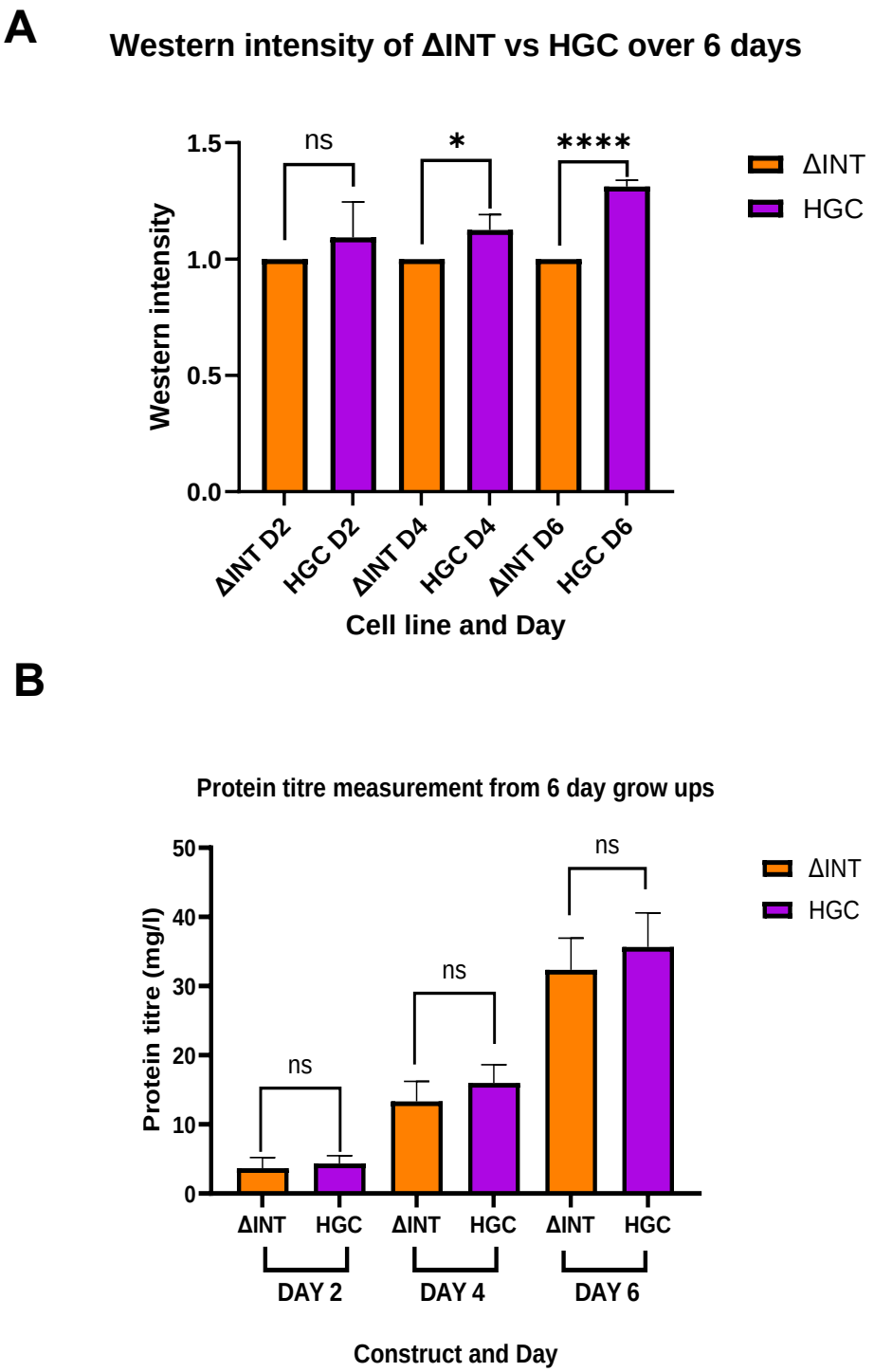


Figure 3.7: Protein titre measurements.

A) Western intensities measured from the western blots presented in figure 3.6. Each intensity was normalised to the ΔINT control from the same day on each blot. **B)** Protein titre from the cell media samples generated from our 6-day batch cultures (n=3). Titre data were collected using HPLC by Dr Sarah Dunn at AstraZeneca.

observations. Hence, we conclude that increasing the GC content of an intronless sequence has positive effects on protein production over longer time periods, but to what extent requires further investigation.

3.7 Partial increases in GC production show slight changes in Fc-FP protein expression

Designing a therapeutic protein for production and use in a clinical setting must consider a variety of properties to ensure correct functioning of the antibody in the clinic. Examples of these properties include prolonged stability in the patient to ensure the drug can reach the intended target and low immunogenicity to decrease off-target binding and reduce the number of adverse effects (Lu et al.2020). To assist in the design process, we wanted to test if increasing the GC content of the 5' end or the 3' end of the intronless sequence was enough to facilitate the increases in expression we saw with the full length HGC construct. This idea was supported by evidence in the literature that having a GC rich 5' end is enough to facilitate increased expression from an intronless construct (Mordstein et al. 2020). We took the full-length HGC Fc-FP sequence and PCR amplified the 5' and 3' end, 220bps at each end and inserted it into the ΔINT sequence to generate sequences that had GC rich 5′ or 3′ ends respectively and used them to generate stable expression pools.

We initially tested the expression of the Fc-FP protein at the 48-hour mark (Figure 3.8.1) and found, as with the HGC, better expression of the Fc protein in the cell media was detected. We were unable to detect any real differences between the rich 5' and rich 3' sequences, suggesting that the effects are simply due to the increase in GC content rather than positioning within the sequence.

As before with the HGC cell lines, we then compared the general RNA levels in these cell lines against the ΔINT and once more found that we could detect no real differences in the total levels of RNA in the cell. This matched up with the effects we see with the full-length GC construct. Thus, we assumed that the increase in protein would come from a difference in RNA export to the cytoplasm as we saw before. In contrast to our expectation, we found that there was no significant difference between the export ratios from cytoplasmic/ nuclear fractionation. This suggests a model in which the GC content of an intronless sequence affects the RNA expression in a cumulative manner, the higher the GC content the better RNA export and protein

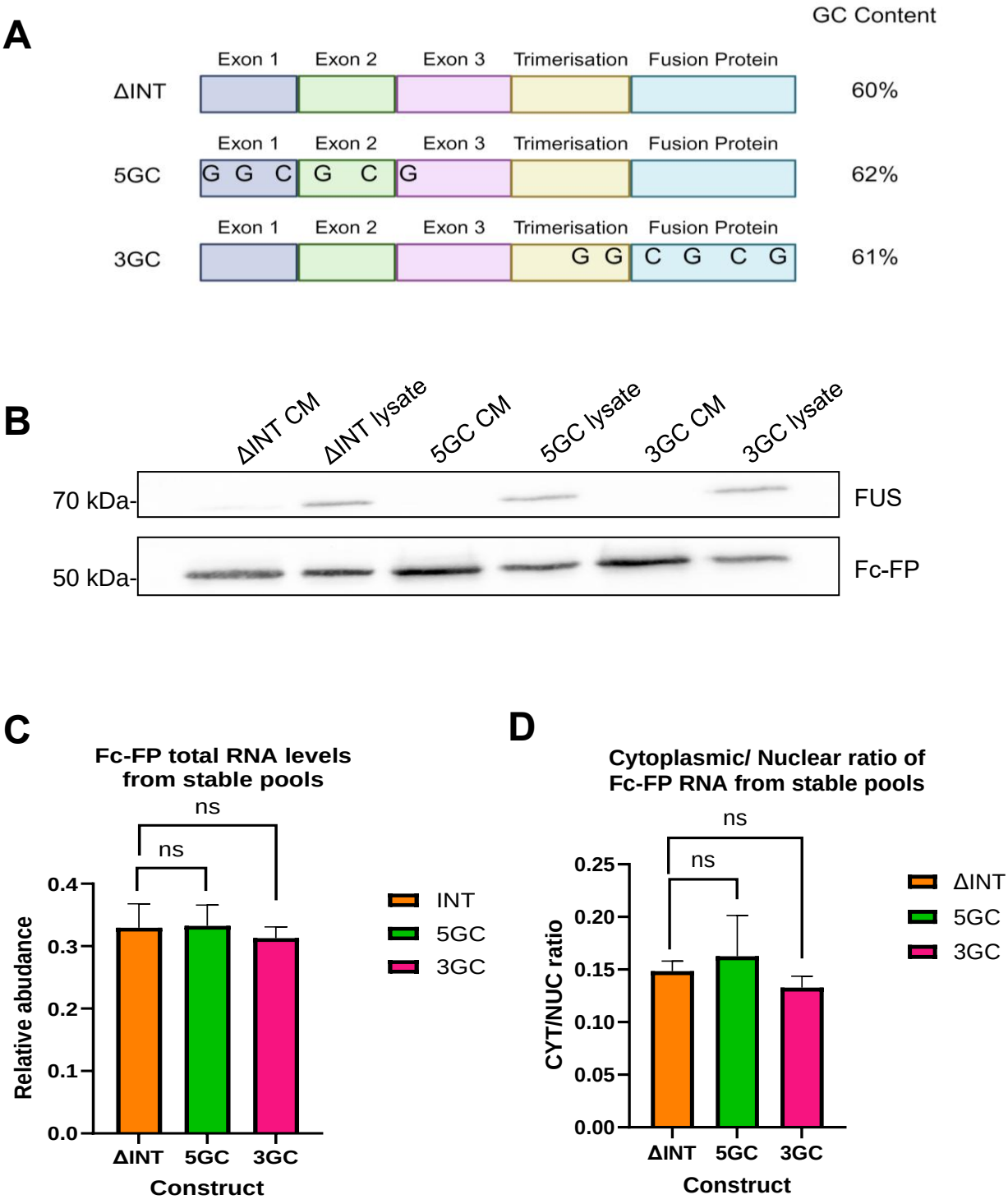


Figure 3.8: Expression of 5GC and 3GC constructs in a stable system.

A) Schematic of ΔINT, 5GC and 3GC sequences. **B)** Western blot showing Fc-FP expression from Stable pools of CHO cells. FUS is shown as a loading control. **C)** qRT-PCR panel of total RNA levels from stable pools (n=6). All RNA is normalised to the levels of 18S rRNA. **D)** Cytoplasmic/nuclear ratio of stable CHO pools (n=4). All RNA is normalised to the level of GAPDH mRNA.

production rather than any specific positional effects that may occur. We can also not rule out that other cellular effects may be contributing to these effects, such as effects on the stability or translation of the mRNA which in turn may lead to the increased protein levels.

3.8 Increasing the 5′ GC content effects protein titre production

While we observed no changes in mRNA behaviour from the 5GC and 3GC cell line, we did see an increase in protein production compared to the ΔINT control line. To test this increased production over a longer time and its benefit to industrial production, we performed 6- day batch cultures. As before, cells were seeded at a known density and cell counts and cell media samples were taken at day 2, 4 and 6. Equal volumes of cell media were then used for western blotting to analyse protein production over the time course (Figure 3.9.A).

From the western blots we observe increased protein titre at the day 6 timepoint in the 5GC and 3GC when compared to the ΔINT. As with full length GC, there appears to be no differences at the day 2 and 4 timepoint between the ΔINT and 5GC/3GC.

We then sent the cell media samples from these grow ups for titre analysis at AstraZeneca (Figure 3.9.B). We observe increased titre production from the 5GC and slightly increased production from the 3GC cell line at the 6-day mark compared to the ΔINT, matching our observations from the time course westerns. While we cannot confirm a significant increase in protein titre due to only having 2 replicates, the increase in protein production from an intronless sequence by increasing GC content at the 5′ end matches the observations of Mordstein et al (2020) who also reported increased protein expression from intronless constructs with a GC rich 5′ end. Thus, we conclude that increasing the GC content of the 5′ end has a positive effect on the protein expression of an intronless Fc-FP, although the reason for this increase requires further investigation.

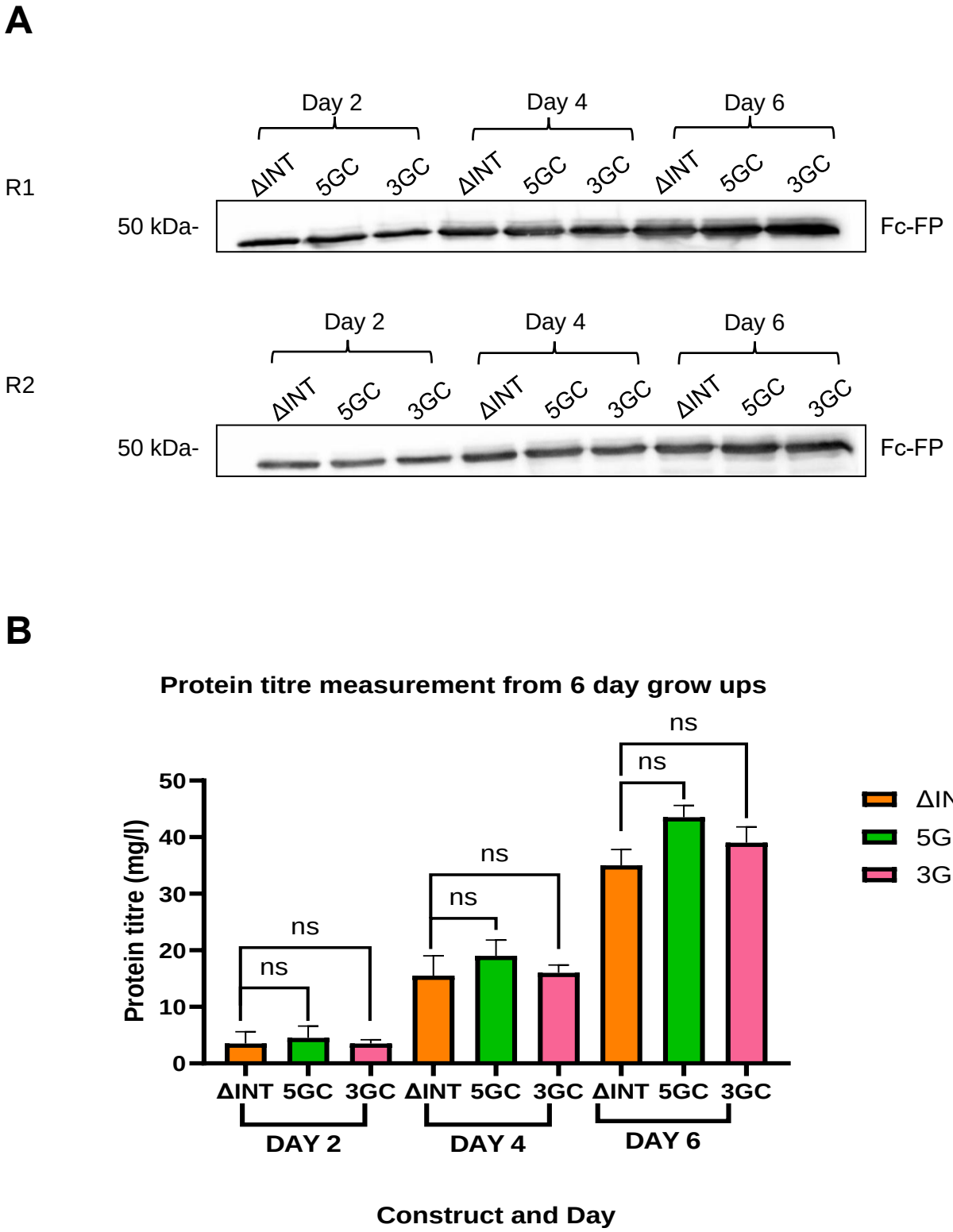


Figure 3.9: Production of Fc-FP protein from ΔINT and 5GC/3GC over 6 days from stable cell lines.

A) western blots showing ΔINT, 5GC and 3GC Fc-FP protein production from stable CHO pools over 6 days. **B)** Protein titre from the cell media samples generated from the 6-day batch cultures (n=2). Titre data were collected using HPLC by colleagues at AstraZeneca.

3.9 Summary

In this chapter we analysed the effects of GC content on intronless monoclonal antibody production from CHO cells. We showed that increasing the GC content of an intronless sequence improves its expression by improving the export of mRNA to the cytoplasm and this is accompanied by increased recruitment of the mRNA export factor ALYREF to the mRNA

We designed an intronless Fc-FP sequence that contained an identical amino acid sequence but altered codon sequence using 3rd position wobble to increase the GC content of an intronless sequence. Transient transfection in CHO cells and subsequent analysis of protein levels via western blot analysis showed that removal of introns reduced protein expression but increasing the GC content helped rescue expression, especially in the amount of protein secreted into the cell media. Analysis of total mRNA levels using qRT-PCR showed that increasing the GC content of an intronless sequence improved total RNA expression. We used cytoplasmic/ nuclear fractionation to study the distribution of RNA within the cell and found that there was an increase in mRNA export to the cytoplasm when the GC content was increased, consistent with the observations in the literature (Mordstein et al. 2020).

We took this a stage further and used RNA-IPs to show that the improved export efficiency was due to an increase in the binding of the key export adaptor ALYREF. Increasing the GC content of an intronless Fc- FP improves ALYREF recruitment to an mRNA which most likely contributes to the export effect that we see. This result is supported by the observations of Mordstein et al (2020) and Zuckerman et al (2020) that were discussed earlier in the chapter

While this evidence supports our result, we can only hypothesise about the mechanism that results in this increased binding in the presence of increased GC content. ALYREF contains an RNA Recognition Motif (RRM) flanked on both sides by unstructured Arginine - Glycine rich (RGG) domains which can bind RNA. The RRM has relatively little RNA binding activity and instead the principal RNA binding activity resides within the unstructured RGG domains (Golovanov et al. 2006). This series of structures is also found in the RNA binding protein FUS, a protein which has been

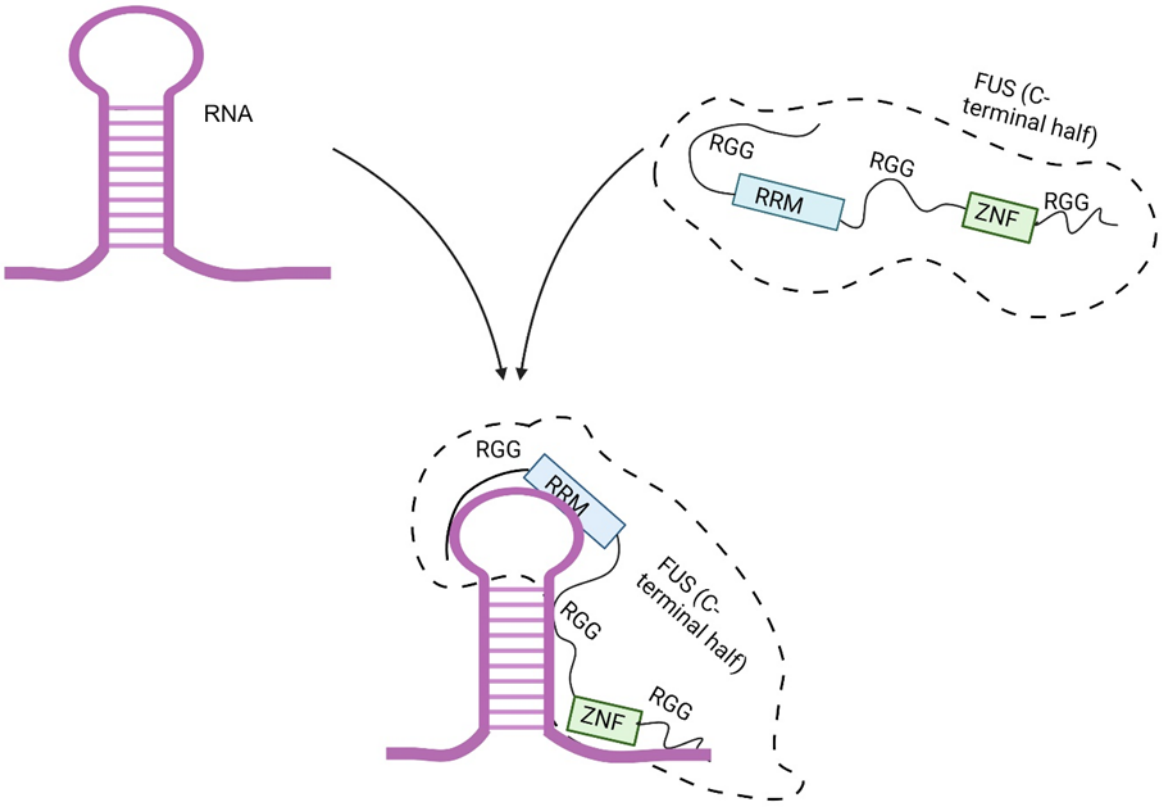


Figure 3.10: A model of the protein FUS binding an RNA stem loop.

Figure 3.10 shows a schematic of how the protein FUS can recognise RNA secondary structure, such as a stem loop, rather than specific sequences. The ZNF domain can recognise target sequences in an RNA allowing the RRM of FUS to interact with non-specific sequences in the stem loop, allowing the disordered RGG domains to wrap around the stem loop.

linked to a variety of cellular and RNA related functions such as: transcriptional regulation, splicing and RNA phase separation and granule formation (Schwartz et al. 2012, Burke et al. 2015). FUS, through interactions with its RBM and flanking RGG sequences can bind to highly structured regions of RNA such as RNA stem loops (Figure 3.10). The RGG domains in FUS dramatically increases the affinity for structured RNAs and associate with the structured regions within the RNA. GC rich mRNAs have a greater propensity to form secondary structures such as hairpin and G-quadraplexes, which FUS binds preferentially (Loughlin et al. 2019, Takahama et al. 2013). Thus, we predict that ALYREF, in a comparable manner to FUS, may bind to structured regions of RNA rather than sequence specific elements which would provide a molecular basis for the preferential binding of ALYREF to GC-rich intronless sequences.

This increased association between ALYREF and GC rich mRNA may also increase the stability of the mRNA, further improving its expression. ALYREF competes for binding to ARS2, a member of the CBC, against hMTR4 which is a key component of the NEXT and PAXT degradation complexes. Efficient recruitment of ALYREF to ARS2 is a key marker of efficient export, while binding of hMTR4 leads to degradation of inefficiently exported mRNAs (Fan et al. 2017). Hence, the increased ability of GC rich intronless mRNA may also protect the mRNA from degradation and increase the stability of mRNA transcripts by outcompeting hMTR4 and marking the mRNA for export.

We used further RNA-IPs to show that removal of introns increases the amount of NXF1 binding to the mRNA. This result was contrary to our predictions, as we expected reduced NXF1 binding to the mRNA upon removal of introns from the sequence (Reed and Hurt, 2002). Therefore, our data indicates that efficiently exported mRNAs are bound to NXF1 for a lower amount of time than inefficiently exported mRNAs. This may be because the increased nuclear residence time for inefficiently exported RNAs allows greater NXF1 binding, whereas for efficiently exported RNAs, the NXF1 interacts with the RNA transiently and is stripped off during transit through the nuclear pore. However, ALYREF does not appear to accumulate on inefficiently exported RNAs, suggesting that other factors contribute to the binding events we observed.

To make our work more appropriate for biomanufacturing in industry, we built pools of stable CHO cells expressing the intronless Fc-FP and the high GC counterpart to compare the expression of therapeutic protein. At this point we discovered a mis-splicing event in our WT control that meant a substantial amount of protein produced remained in the cell. This means that while we can compare the effect of GC content on an intronless sequence, we cannot make any statements about expression levels when compared to a spliced RNA. We compared the expression of the ΔINT and the HGC Fc-FP in a stable context and showed that the protein increases we saw transiently are present in a stable environment. There were no differences in the total RNA levels in stable pool, but cytoplasmic/nuclear fractionation showed a similar effect on RNA export as demonstrated transiently, indicating that the increase in protein expression comes from increased mRNA export efficiently. Mimicking industrial growth conditions, we saw the increases in protein production were present over longer time periods showing that the work has applications in industrial production.

Finally, we developed sequences with only the 5’ and 3’ end of the intronless CDS having altered GC content. We built stable cell lines expressing these mAbs and saw that while there was an increase in Fc-FP production levels, there was no significant effect on RNA levels or on the export ratio. This data indicates that the improvements of intronless Fc-FP expression seen by increasing the GC content are because of the overall GC content as opposed to positional effects in the sequence.

Thus, from our data, we conclude that increasing the GC content of an intronless sequence can improve the therapeutic protein expression by improving the nuclear export of the mRNA with a concomitant increase in the level of the ALYREF nuclear export factor associated with the intronless mRNA.

Chapter 4: Improving therapeutic protein expression from an intronless sequence using RNA binding protein recruitment motifs.

Following our experiments analysing the effects of high GC content and its effects on mRNA export, we looked at other ways that we could improve mRNA export of an intronless mRNA. The work in this chapter focuses on the use of two protein binding motifs for SRSF1/SRSF7 and Sf3b1, and their interactions in the RNA processing pathway to improve protein production. We further combine these protein binding motifs with increased GC content to evaluate their cumulative effects on expression of therapeutic protein from an intronless sequence.

4.1 Insertion of an SR sequence shows minor changes in transient ΔINT protein expression.

The loss of splicing, and the subsequent loss of exon junction complex deposition on the mRNA and its interaction with ALYREF/TREX, accounts for the reduction in export seen upon removal of introns (Valencia et al. 2008). Our work on GC content identified mRNA export as a targetable molecular process that can successfully improve expression of an intronless sequence.

While the TREX/NXF1 pathway is regarded as the main export pathway for processed mRNAs, ALYREF is not the only factors capable of facilitating mRNA export. Serine/Arginine (SR) rich proteins have been shown to interact with a variety of RNA binding proteins, such as Nxf1, to export RNA to the cytoplasm. SRSF3 (SRp20) and SRSF7 (9G8) have well characterised interactions with TAP and RNA using their RRMs (which are 80% similar between both proteins) which allows them to serve as export adaptors (Huang et al. 2003, Hargous et al. 2006).

As the work presented in the previous chapter focused on alterations in the CDS to improve expression, we wanted to identify binding motifs for known export adaptors that could be inserted into the 5’ untranslated region (UTR). By using a sequence in the UTR, this left open the possibility of combining these sequences with our high GC constructs to assess cumulative effects on expression. Specifically, we looked for sequences or motifs that originated from single exon genes or as we expected these to have the highest chance of success.

Khan et al (2021) discovered a cytoplasmic accumulation region (CAR) present in the long non-coding RNA (lncRNA) NKILLA, a mono-exonic and cytoplasmic localised

lncRNA. They identified a cluster of binding sites for the proteins SRSF1 and SRSF7 and used siRNA knockdowns to show that loss of these proteins reduces NKILLAs export to the cytoplasm. Removal of this binding site cluster also causes nuclear retention, showing that the recruitment of SRSF1/SRSF7 to the lncRNA is sequence specific. Interestingly, they found that export of NKILLA was reduced when TREX subunits (ALYREF and UAP56) or NXF1 were knocked down using RNA interference, suggesting that SRSF1 and SRSF7 recruit or interact with TREX to facilitate the export of NKILLA to the cytoplasm (Khan et al. 2021).

We inserted the 55nt region from NKILLA identified as containing clusters of SRSF1/SRSF7 binding sites into the 5' UTR of the intronless Fc-FP sequence to analyse its effects on expression. We transiently transfected CHO cells with either the WT, ΔINT or SR construct (Figure 4.1.A) alongside flag-tagged GFP as a transfection control and analysed protein expression after 48 hours using western blot (figure 4.1.B). As we showed previously, protein levels were lowered in the ΔINT when compared to the WT, especially in the cell media. Additionally, inclusion of the SR sequence in the 5' UTR had a positive effect on protein expression when compared to the ΔINT in both the CM and cellular lysate.

Following on, we wanted to analyse the effects of the SR sequence on total RNA levels. We transfected the same constructs and analysed the Fc-FP mRNA levels 48 hours post transfection using qRT-PCR (Figure 4.1.B). As demonstrated previously, the deletion of introns has a negative effect on RNA levels when compared to the WT. However, we could detect no significant differences between the RNA levels in the ΔINT and SR constructs. This observation suggests that the increases we see are a result of either increased export of the RNA or from alternate factors such as increased RNA stability or translation.

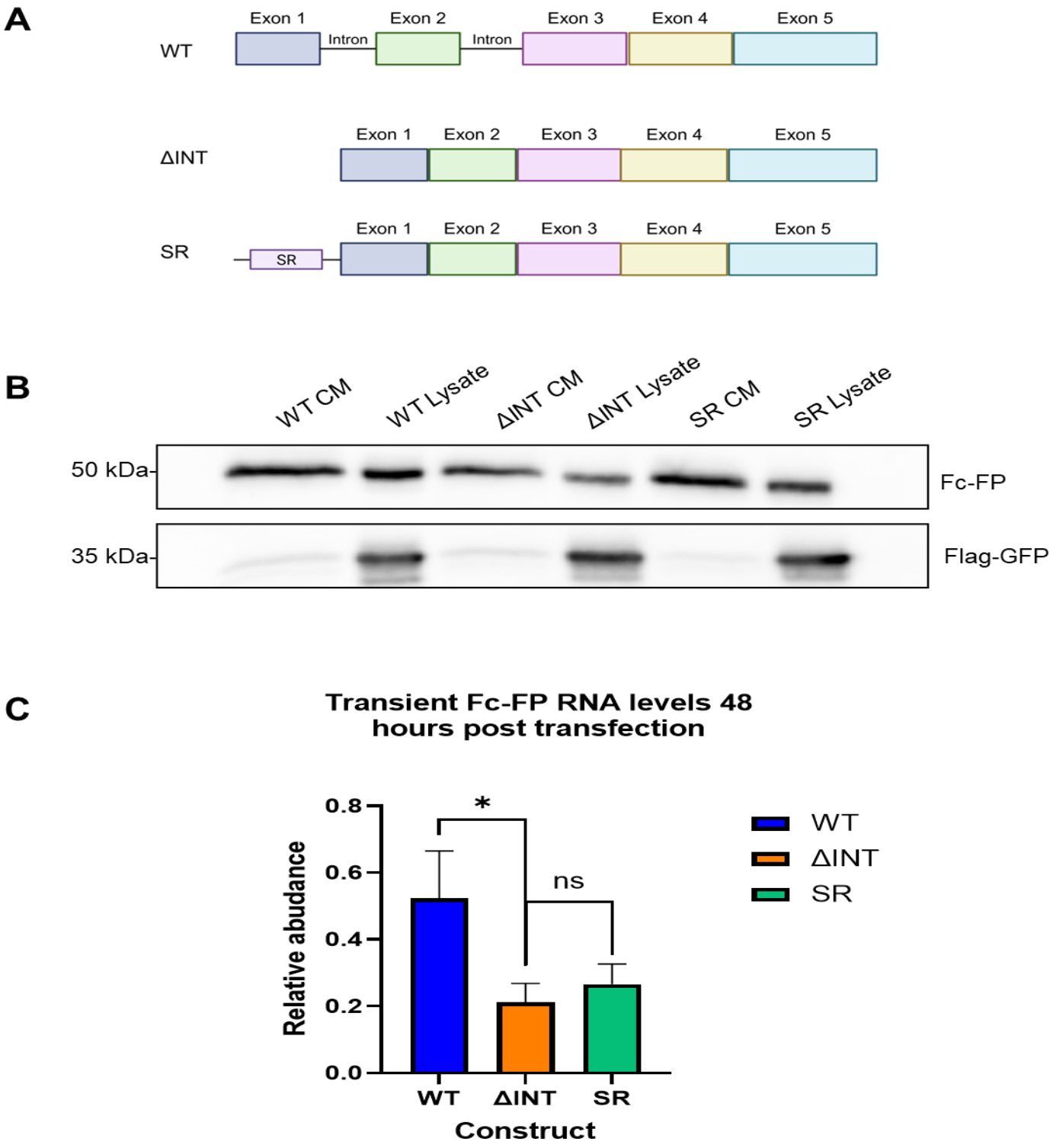


Figure 4.1: **Transient expression of Fc-FP protein and RNA.**

A) Schematics of the Fc-FP sequences transfected into CHO cells. **B)** Western blot showing transient Fc-FP expression from the 3 constructs. Cell Media (CM) and cell lysate for each sample are shown. Flag-GFP is shown as a transfection control. **C)** qRT-PCR of panel of Fc-FP RNA 48 hours after transfection (n = 3). All samples were normalised to the level of 18s rRNA.

4.2 Insertion of the SR motif affects stable protein expression of an intronless sequence

We want to make our work as applicable to industrial biomanufacturing as possible and we saw some differences between the transient and stable results in the high GC experiments. Therefore, we decided to generate stable cell pools in CHO cells expressing the intronless Fc-FP containing the SR binding sequence to monitor their performance.

We seeded cells and allowed to grow for 48 hours and then harvested the cell media and cellular lysate and analysed the protein levels after using western blotting (Figure 4.2.A). As observed with transient transfections, we saw an increase in protein expression from the SR cell line compared to ΔINT, especially in the cell media. We then analysed the total RNA levels from the cell pools using qRT-PCR (Figure 4.2.B). As we detected from our transient experiments, there were no differences in the total RNA levels between the ΔINT and SR cell lines.

4.3 The inclusion of SR motif may improve loading of ribosomes onto intronless Fc-FP mRNA

SR proteins have been implicated in a variety of functions. Alongside their roles in pre mRNA splicing and in RNA shuttling, several SR proteins have been implicated in translation. SRSF2 has been shown to help initiate translation through interactions with the mTOR pathway, which leads to phosphorylation of 4E-BP and causes the activation of eIF-4E and activates translation (Michlewski et al. 2008). SRSF2 and SRSF7 have also been shown to associate with ribosomal machinery and enhance translation of a reporter gene (Sanford et al. 2003). As we saw no improvements in total RNA levels transiently or stably after inclusion of the SR motif in the 5′ UTR, we hypothesised that the recruitment of SR proteins may not improve the expression of RNA but may improve ribosomal assembly onto the RNA. An increase in translational efficiency would account for an increase in protein production but no changes in RNA levels.

To test this, we performed polysome profiling to analyse differences in translational efficiencies of the Fc-FP RNA between the ΔINT and SR cell lines. We used cycloheximide to stop translation and to block the ribosomal release from the mRNA.

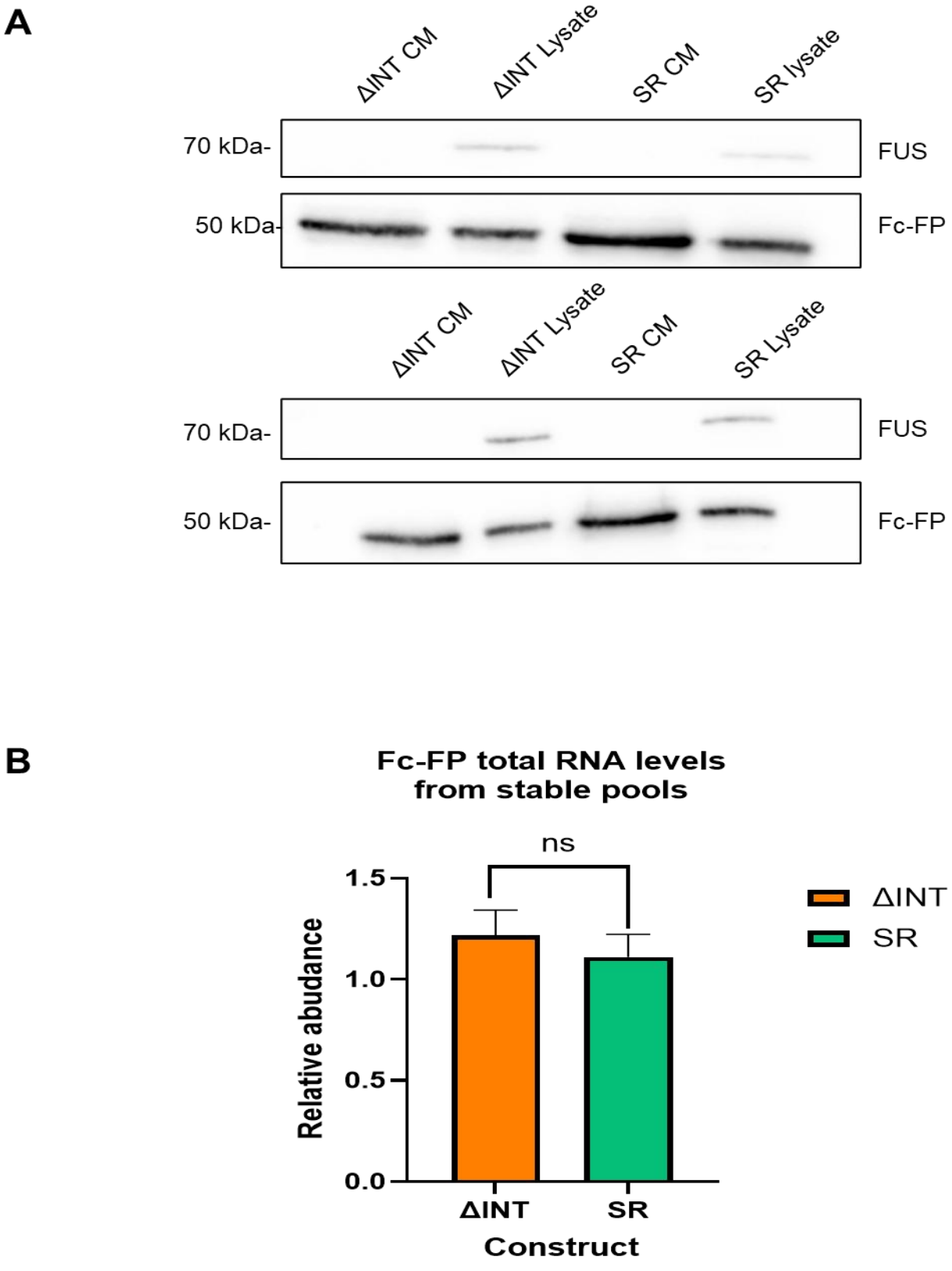


Figure 4.2. Stable expression of ΔINT and SR from stable cell lines.

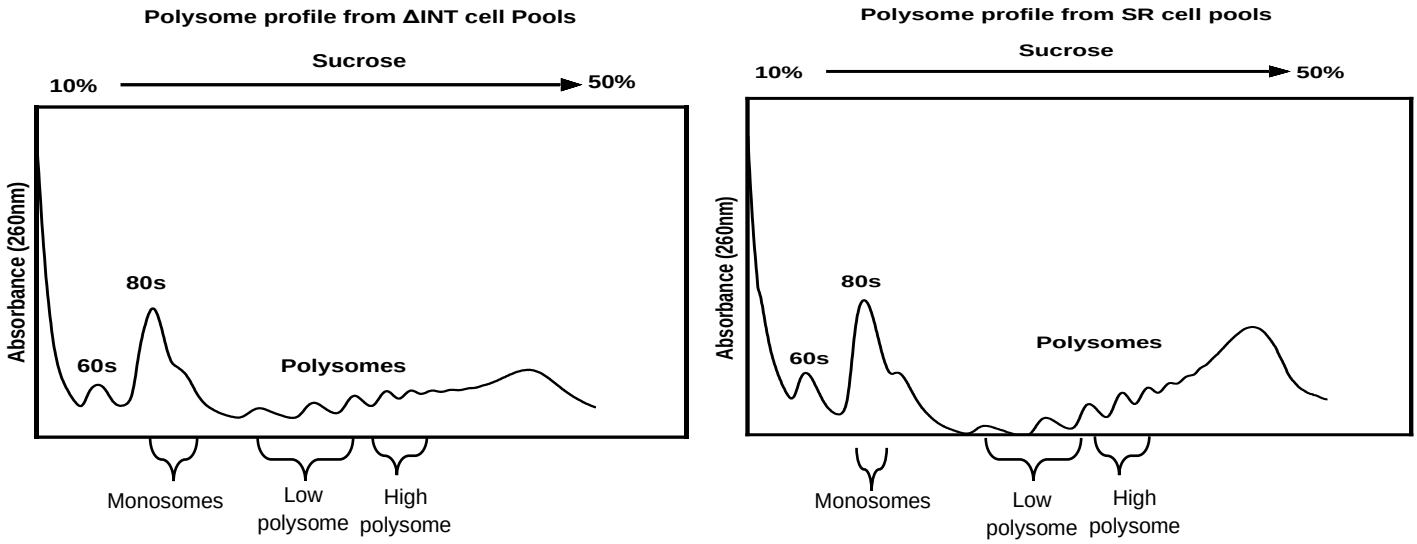
A) Western blots of cell media and cell lysates from ΔINT and SR cell lines. FUS is shown as a loading control for cell lysate samples. **B)** qRT-PCR panel of Fc-FP total RNA from stable pools (n=6). All samples were normalised to the level of 18S rRNA.

The cytoplasmic faction was isolated from the nuclei and loaded onto a sucrose gradient (10% - 50%) and factions were collected. As multiple ribosomes can assemble on a single mRNA, highly translated mRNAs will have multiple ribosomes bound resulting in a heavier molecular weight, allowing separation of well translated and poorly translated mRNAs. Absorbance profiles (Figure 4.3.A) was used to identify the 80s peak and the subsequent polysome peaks. We isolated the RNA from the prominent 80s peak and then pooled the RNA from peaks 2–4 and peaks 5-8. RNA from the 80s pool represents monosomes, mRNAs with a single ribosome attached to it. The pooled RNA from peaks 2-4 represent less efficiently translated RNAs as they have fewer ribosome attached to them while RNA pooled from peaks 5-8 represent better translated mRNAs as they have multiple ribosomes attached. This allows us to compare the translational efficiencies of the Fc-FP mRNA from each set of cell pools.

We then analysed the RNA from each pool using q-PCR. We normalised all RNA to the levels of 18s rRNA from a whole cytoplasm sample taken before loading onto the sucrose gradient. We then calculated the levels of Fc-FP mRNA in each fraction as a percentage of the total Fc-FP mRNA levels in the cytoplasm so that we could fairly compare the levels between cell lines and replicates.

Our initial comparison between the cell lines (n=2) suggests that the inclusion of SR motifs in the intronless 5'UTR has an enhancing effect on translation. We compared the levels of Fc-FP mRNA in the polysome fractions of the ΔINT and SR cell pools. We observe a clear increase in SR mRNA in the low polysome fraction when compared to the ΔINT low polysome fraction. Due to data spread in the SR cell line samples we cannot draw any conclusions about the mRNA levels in the high polysome fractions and we would need a further set of replicates. However, an increase in low polysome fractions would be a good indicator that the SR mRNA is translated better as some of those polysome present may have become bound by more ribosomes if translation had not been halted. Thus, we can initially conclude that the SR sequences in the 5' UTR appear to have a positive effect on the translation rate of intronless mRNAs, but further investigation is needed to clearly see how much of an effect it has.

A



B

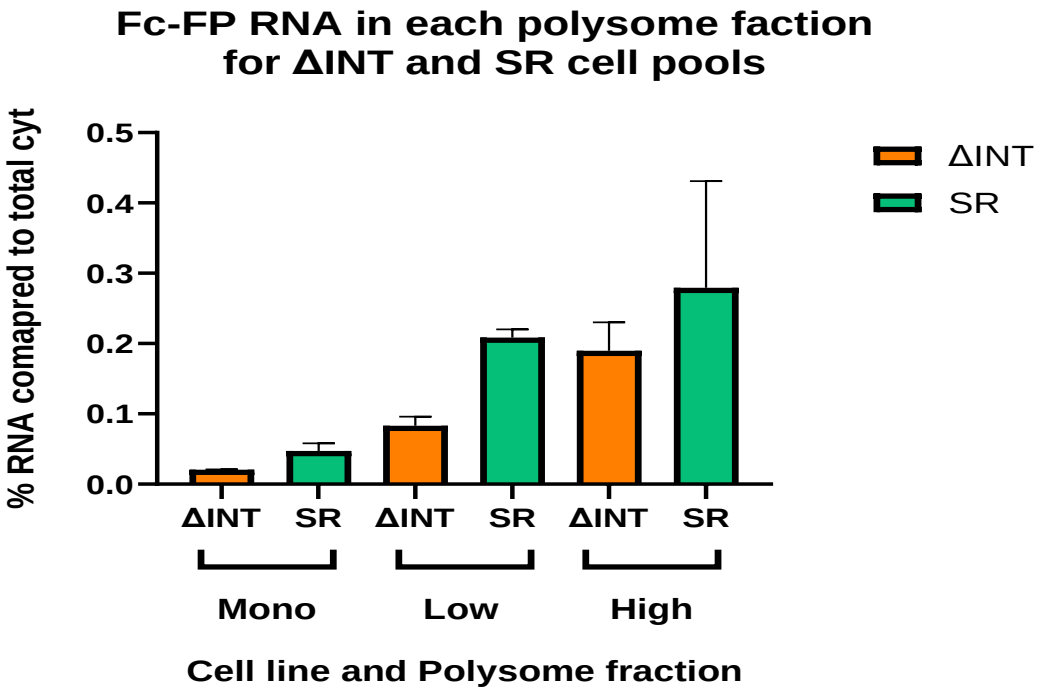


Figure 4.3. The effect of SR on translation in stable CHO pools.

A) Polysome profile from the ΔINT and SR stable CHO cell line. Peaks taken for each fraction are labelled. **B)** qRT-PCR panel of Fc-FP RNA bond to ribosomes from ΔINT and SR stable CHO pools (n=2). Each sample is normalised to the level of 18s rRNA in the total cytoplasm from each polysome.

4.4 Inclusion of SF3b motif into the 5′ UTR of ΔINT sequence positivly effects transient expression in CHO cells.

Binding of mRNA export factors occurs co-transcriptionally, alongside other essential processing actions such as splicing or 5’ capping. As our work with the SR binding motifs in the 5’ UTR showed positive effects on protein expression transiently, we examined the literature to look for more proteins with known binding motifs or known interactions with RNA processing in a similar manner to SR proteins.

The U2 snRNP complex is a key component of the spliceosome. U2 snRNP is involved in branchpoint recognition and initial generation of the pre-spliceosomal complex, one of the first stages of intron removal (Black et al. 1985). The U2 snRNP complex is made up of a few elements, such as SF3a and the SF3b complexes, with the U2 snRNA. The largest member of the SF3b complex, SF3B1, has been shown to bind to some histone pre-RNAs and bring about histone RNA 3’ processing by the U7 snRNP. The interaction between SF3B1 and histone RNA was shown to be sequence specific, with the 7nt sequence C/GAAGAAG shown to be a direct recruitment motif (Friend et al. 2007). More recent work has shown that SF3B1 has direct interactions with members of the THO complex, a core component of TREX, hinting at a role for SF3b in RNA export. It was shown that inclusion of the SF3B1 binding motif found in histones could increase the export of mRNAs to the cytoplasm and that deletion of SF3B1 causes bulk RNA retention in the nucleus (Wang et al. 2019).

This work interested us for several reasons. Firstly, histone RNA is a famous example of an intronless RNA and thus any proteins involved in their processing may be good targets for increasing expression from an intronless sequence (Friend et al. 2007). Secondly, Wang et al (2019) showed that inclusion of the SF3b target sequence could be used to improve the RNA export of intronless RNAs that are typically poorly exported. They also analysed ENSEMBLE human gene sets to show that the target sequence was present in many intronless genes (34%). Finally, they showed that export of intronless RNA was improved in a dose dependent manner, with multiple copies of the SF3b sequence having a cumulative effect on RNA export efficiency (Wang et al. 2019). The idea of a dosage-dependent sequence was interesting to us, as it would allow for greater regulatory potential in construct design when designing molecules for expression at the manufacturing level.

We designed ΔINT constructs that contained either one copy or three copies of the SF3b binding motif to test if they enhance expression of an intronless sequence. When designing these constructs, we noticed that the base ΔINT Fc-FP sequence contained one copy of the SF3b binding sequence (GAAGAAG) in the 3rd exon. The presence of this sequence could explain why the ΔINT is still able to express good levels of therapeutic protein despite the removal of introns. CHO cells were transiently transfected with the ΔINT, SF3b (X1) or SF3b (X3) constructs with a flag-GFP to monitor transfection efficiency between the different constructs (Figure 4.4.A). Western blotting was used to examine expression levels in the cellular lysate 48 hours after transfection. We observed better expression of Fc-FP protein from the SF3b (X1) and SF3b (X3) when compared to the expression of ΔINT in the lysate (Figure 4.4.B). However, taking account of the higher transfection of the SF3b (X3) construct, as revealed by examining the expression of the GFP transfection control, the efficacy of the SF3B (x3) motif seems much lower than that for the single SF3B motif for promoting Fc-FP protein expression when compared to the ΔINT control.

We then analysed total mRNA levels 48 hours after transfection using qRT-PCR (Figure 4.4.C). We observed significantly more RNA when the SF3b (X1) sequence was added to the 5' UTR compared to the levels in ΔINT. Interestingly, we saw no significant differences between the ΔINT and the SF3b (X3) transfections indicating that increasing the number of motifs had the opposite effect. This is contrary to what we expected and what was reported in the literature and requires further investigation in how best to optimise expression of an intronless sequence with the SF binding site.

4.5 Stable expression of SF3b into the ΔINT slightly improves Fc-FP expression but has no effect on RNA levels

As with the other constructs previously generated, we wanted to test the effect of the SF motif in the ΔINT 5'UTR in a stable condition. We generated stable pools of CHO cells stably expressing the Sf3b (X1) (now referred to as SF3b) construct to compare expression levels against our previously generated stably expressing ΔINT CHO cell pool. As the effect of multiple copies of the SF3b motif was unclear in our transient experiments, we decided not to generate a stable pool expressing this construct. We harvested cell pellets 48 hours after seeding in wells and analysed their expression levels in the cell media and cell lysate. Western blotting showed slight increases in Fc-FP protein levels when the SF3b binding sequence was present when compared to

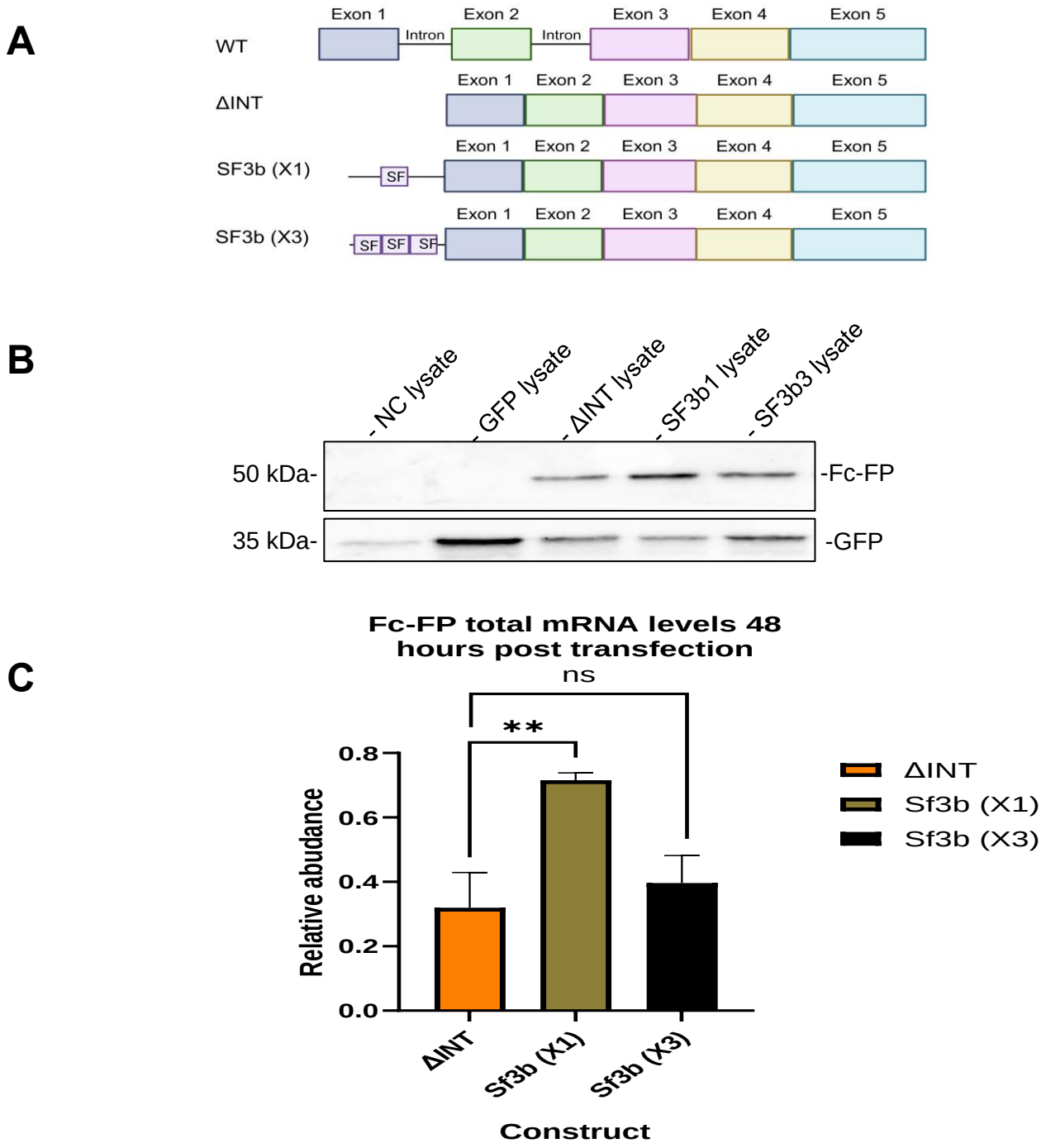


Figure 4.4: Transient expression of SF3b constructs in CHO cells.

A) Schematics of Fc-FP sequences transfected into CHO cells. **B)** western blot showing transient Fc-FP protein expression from CHO cell lysate. GFP is shown as a transfection control. **C)** qRT-PCR panel showing transient Fc-FP mRNA expression in CHO cells (n=3). All RNA levels were normalised to the levels of 18S rRNA.

the ΔINT control (Figure 4.5.A). However, the differences we observe were much less than those we had observed from the HGC, or the SR cell pools shown previously.

We then extracted total RNA levels from cell pellets grown and interestingly we could detect no significant differences between the ΔINT and SF3b pools, showing that the inclusion of the SF3b motif was not having any effect on RNA levels (Figure 4.5.B). While this was contrary to our transient result, we observed the same effect in the stable HGC pools where there was no difference in total mRNA but increased mRNA export efficiency. The interaction of SF3B1 with members of the THO complex was reported and so we concluded that the positive effect on protein levels would be due to increased export of Fc-FP RNA in the presence of the SF motif. We performed cytoplasmic/nuclear fractionation between the two pools two compare mRNA export to the cytoplasm (Figure 4.5.C). To our surprise, we could detect no differences in the export ratios between the ΔINT and the SF construct, indicating that the inclusion of the SF motif was having no effect on RNA export.

4.6 The SR and SF motif show potential to improve longer term protein production from an intronless sequence

Inclusion of the SR and SF3b sequences resulted in increased Fc-FP expression compared to the ΔINT control sequence. As both sequences came to our top attention due to their recruitment of already know mRNA processing factors and their interactions with members of TREX, we wanted to test the cell pools side by side so that we could compare them against each other.

As we did with the High GC line, we analysed the therapeutic protein expression from these two cell pools compared to the ΔINT over a longer period in batch culture. Cells were seeded at a known density and cell counts, media samples and cell pellets were taken every 2 days over a period of 6 days. Cell media samples were analysed using western blot (Figure 4.6.A). Across all three replicates, we observed little difference in the levels of protein being produced at day 2 and day 4 across the 3 cell lines. At day 6, we saw greater differences between the ΔINT and SR and SF cell lines, although this was replicate dependent. In the first 2 replicates, both the SR and SF cell line had produced more protein when compared to the ΔINT cells, while in the final replicate the SR and SF3b appeared to make more protein but the difference between the two cell pools and the ΔINT is more difficult to see.

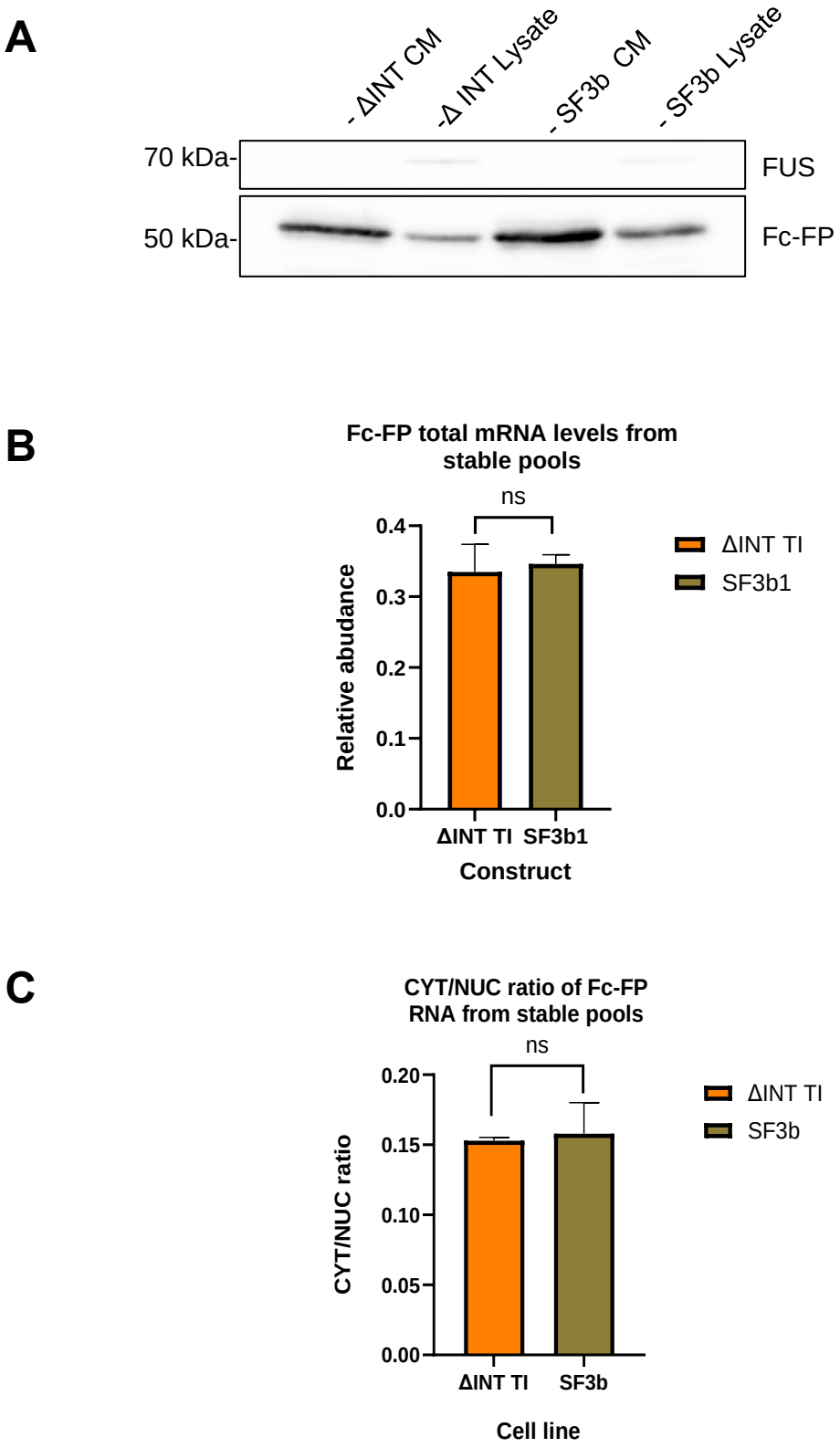


Figure 4.5: Expression of the SF3b vs ΔINT in stable CHO pools.

A) Western blots of cell media and cell lysates from ΔINT and SF3b stable pools to look at Fc-FP expression. **B)** qRT-PCR panel of total RNA from ΔINT and SF3b cell pools (n=5). All samples were normalised to the levels of 18S rRNA**.** CYT/NUC ratio from Fc-FP after 48 hours of growth from stable CHO pools (n=4). All RNA was normalised to the levels of GAPDH mRNA.

We did notice that there were less cells per ml across the pools in the 3rd replicate at day 6 when compared to the first two, so this result may have been influenced by poor cell growth or harsh splitting conditions.

Comparing between the SR and SF cell lines we had generated, it's very difficult from the western data to draw any conclusions about which motif has the greater effect on intronless protein production. The western blots do provide good evidence however that the inclusion of RNA binding protein motifs in the 5'UTR is an effective way of producing more therapeutic protein from an intronless sequence.

As western blots don't provide any reliable quantitative data, we wanted to measure titre data for these cell lines to try and get an accurate measurement of the production differences between the ΔINT and the SR/SF cell lines. We sent the same cell media samples that we analysed using western blot to our industrial partners at AstraZeneca. They used Protein A High-Performance Liquid Chromatography (HPLC) to measure the Fc-FP protein levels from each cell line.

As we saw from the western blots previously, there was little to no change in protein titres at the 2- and 4-day mark (Figure 4.6.B) but an observable difference in titre was detectable at the 6-day time point between both the SR and SF and the ΔINT cell lines. Interestingly, the data sets matched the observations from our westerns, with the differences in bands matching the differences in titre. Replicate 1 and 2 had a large increase in protein titre in SR/SF cell lines when compared to the ΔINT pools, while replicate 3 had only slight improvements in protein expression. Interestingly the SR appears to slightly better than the SF cell line but not by any significant margin.

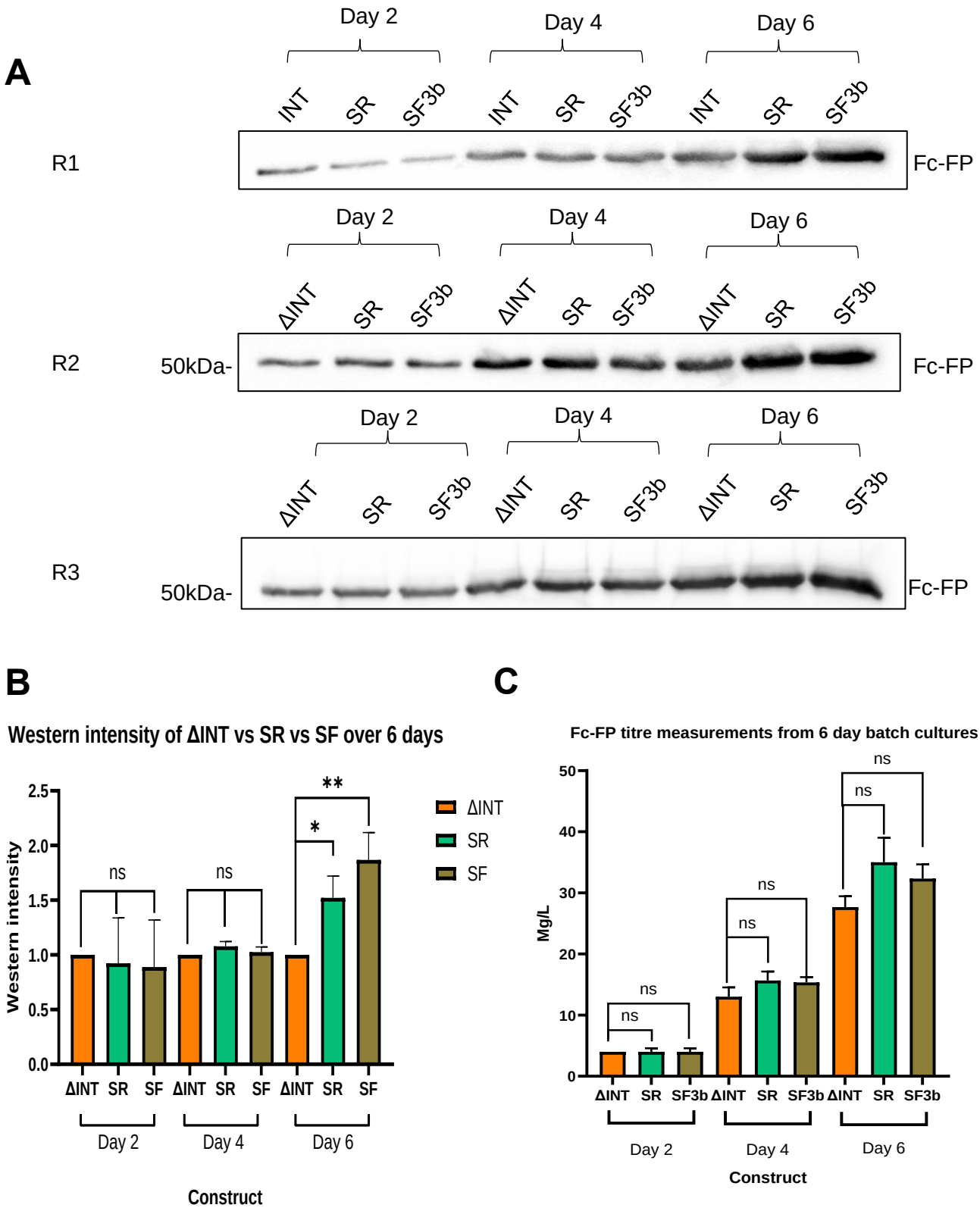


Figure 4.6: **Expression from ΔINT and SR/SF stable CHO pools**

A) western blots showing Fc-FP protein levels in the cell media over 6 days from ΔINT, SR and SF3b stable CHO pools. **B)** Western intensities calculated from the Western blots shown in **A**, for each sample the band intensity was normalised to ΔINT band for each day. **C)** Titre measurements of cell media samples collected from 6-day batch cultures (n=3). Titre data collected by Dr Sarah Dunn at AstraZeneca.

4.7 Combining 5′ UTR motifs with increased GC content shows altered Fc-FP protein expression

In the previous chapter, we have suggested that increased CDS GC content had a positive effect on expression of therapeutic protein from an intronless sequence. As the work presented in this chapter focuses on binding sequences present in the 5'UTR, we wanted to test if combining the two separate areas of investigation would have cumulative effect on expression of an intronless sequence. Theoretically, as both sequences have been shown to have interactions with members of TREX to assist in the export of mRNA to cytoplasm, in the same way that we showed increased GC content caused increased recruitment of ALYREF, we thought the possibility of them working together to recruit members of TREX had merit.

To test this hypothesis, we inserted the binding motifs of SRSF1/SRSF7 and SF3b into the 5' UTR of the high GC Fc-FP sequence before generating stable pools of CHO cells. Due to time restrictions and the limited information, we would be able to gain from western blotting, we decided to proceed immediately into growth assays. We compared therapeutic protein production of the combined cell lines (so termed SR: GC and SF: GC respective of the motif placed in the 5' UTR) against the ΔINT cell line. Limitations on growth space prevented us from growing any other cell lines alongside for comparison.

As we had done previously, we harvested cell media samples at Days 2, 4 and 6 to monitor expression of protein secreted from the cell via western blot anaylsis (Figure 4.8.A). Interestingly, we struggled to observe any clear differences between ΔINT protein expression and the protein levels of the SR: GC or SF: GC pools at the 6-day mark. However, we suspect that the high level of proteins produced at this time point is causing over-exposure of the western and making it difficult to spot any differences between the cell line, highlighting the flaw of using western blotting as a technique for looking at protein production, especially as we loaded equal volumes of media as opposed to know protein concentrations. We did see increases in the SR: GC protein production when compared to the ΔINT protein levels at the day 2 and 4 mark. As we did not see this effect in either of the individual lines (SR or HGC) we saw this as a good sign of improved expression and further evidence that the effect may be masked by the exposure of the western blots.

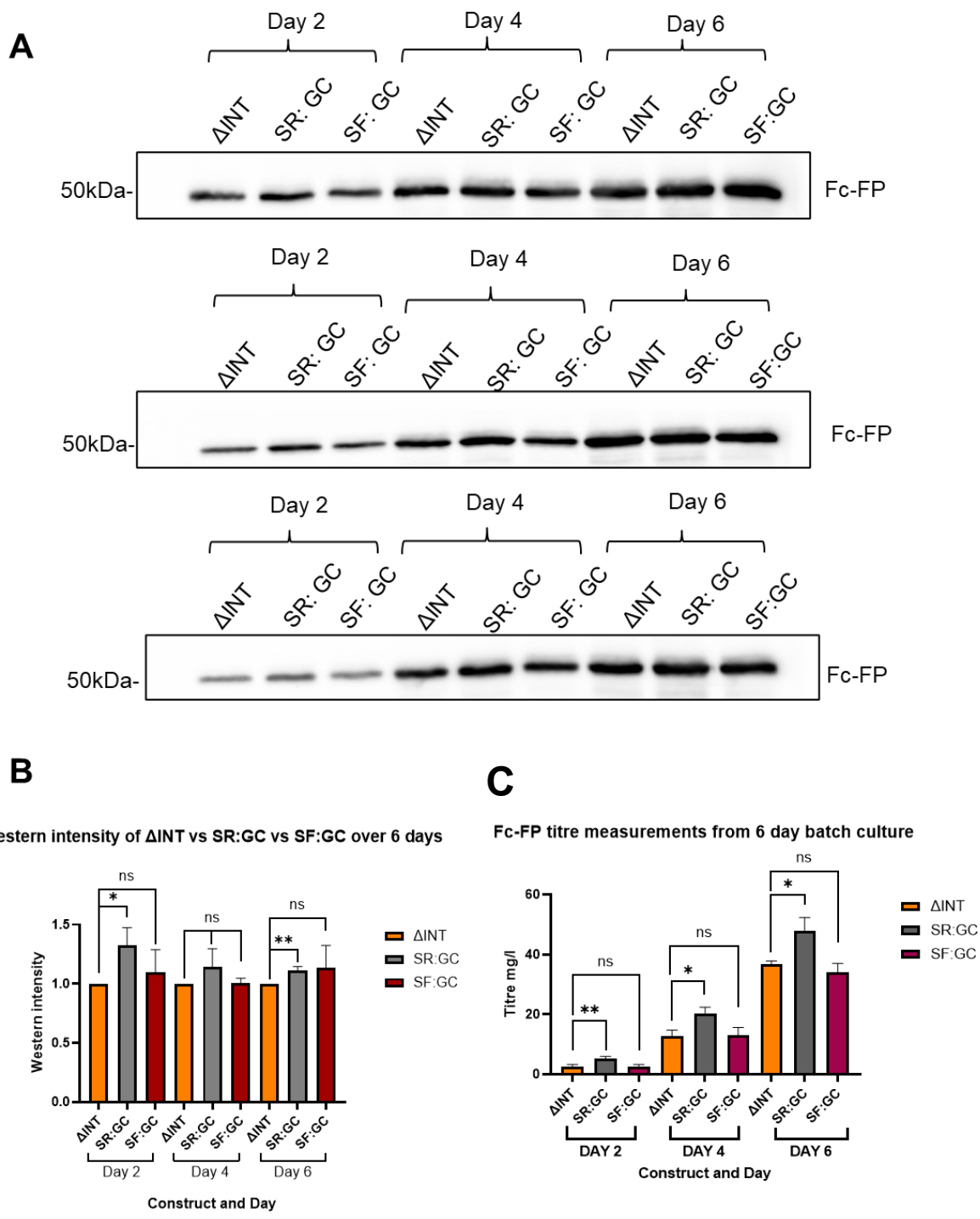


Figure 4.7: Protein production from ΔINT and SR:GC and SF:GC stable lines.

A) western blot showing Fc-FP protein levels in the cell media over 6 days from stable CHO lines. **B)** Western intensities calculated from blots show in **A**, each band was normalised to the ΔINT control sample for each day **C)** Protein titre measurements from the cell media samples collected over 6 days from stable lines. Titre data collected by Dr Sarah Dunn at AstraZeneca.

To test if protein expression was improved by combining the SR and High GC in a quantifiable way, we sent the same cell media samples shown above to our industrial partners at AstraZeneca. They performed HPLC to measure the protein titre produced across the time course (Figure 4.8.C). Combining the GC content and the SR sequence caused a significant increase in protein production compared to the ΔINT across the entire time course, with the highest increase in protein titre detected at the 6-day time point.

4.8 Fed-batch cultures reveal two cell lines with increased cellular protein production

While we have attempted to replicate industrial biotherapeutic protein production within our lab, there were several limitations to our batch cultures. Our 6-day batch cultures were limited to 6 days due to decreasing cell viability at the final timepoint, and we were limited to single cellular counts which cannot fully reveal cell population fluctuations over the full time course. To overcome these limitations, all the cell lines discussed in this chapter and the GC cell lines discussed previously, were sent to AstraZeneca where expression was analysed in fed-batch cultures over 10 days. This would provide a more accurate measure of how the cells behaved over a longer time and secondly a better measurement of productivity.

The stable CHO cell lines expressing our intronless Fc-FP constructs were expanded for ten days, with nutrient rich feed added to sustain cell growth and viability over the ten day period. Titre measurements were taken on day 7 (Figure 4.8.A) and day 10 (Figure 4.8.B). At both time points, there are clear differences in titre levels between the ΔINT cell line and several of our improved cell lines. The SR, SF, SR:GC and SF:GC cell lines all had increased protein titre at both day 7 and day 10. Interestingly, combining the SF3b motif with increased GC content in the CDS further improved the amount of protein titre while combing the SR motif with GC content decreased the amount of protein produced when compared to just the SR motif.

During the fed-batch cultures, cell counts are measured (Figure 4.8.C). The cell lines with large improvements in protein expression also had significantly more cells present than the ΔINT control. To account for this, the specific productivity of each cell line was calculated (Figure 4.8.D). This gives a measure of how much protein is produced per cell per day, as opposed to total protein titre in the culture which does not take into

account differences in cell number. This revealed that inclusion of the SR motif into the 5'UTR of the ΔINT control caused an increase in therapeutic protein per cell produced compared to the ΔINT control. Interestingly, combining the SR motif with increased GC content in the CDS (SR: GC) had a negative effect, reducing the protein production per cell. This indicates that the SR motif and High GC effect expression of intronless therapeutic in different and non-compatible ways.

Furthermore, both increased GC content and inclusion of the SF3b in the UTR had slight but not significant increase in specific productivity compared to the ΔINT control, but combining the two features (SF: GC) caused a significant increases in specific productivity when compared to the ΔINT control. Therefore, the SF3b and High GC in the CDS must act in a co-operative manner to improve expression of an intronless sequence.

From this data, we can conclude that expression of an intronless sequence can be improved by insertion of binding motifs of known RNA binding proteins into the 5′ UTR and by increasing the GC content of an intronless CDS. Additionally, there is the potential for these two methods to work together to improve expression of an intronless therapeutic protein.

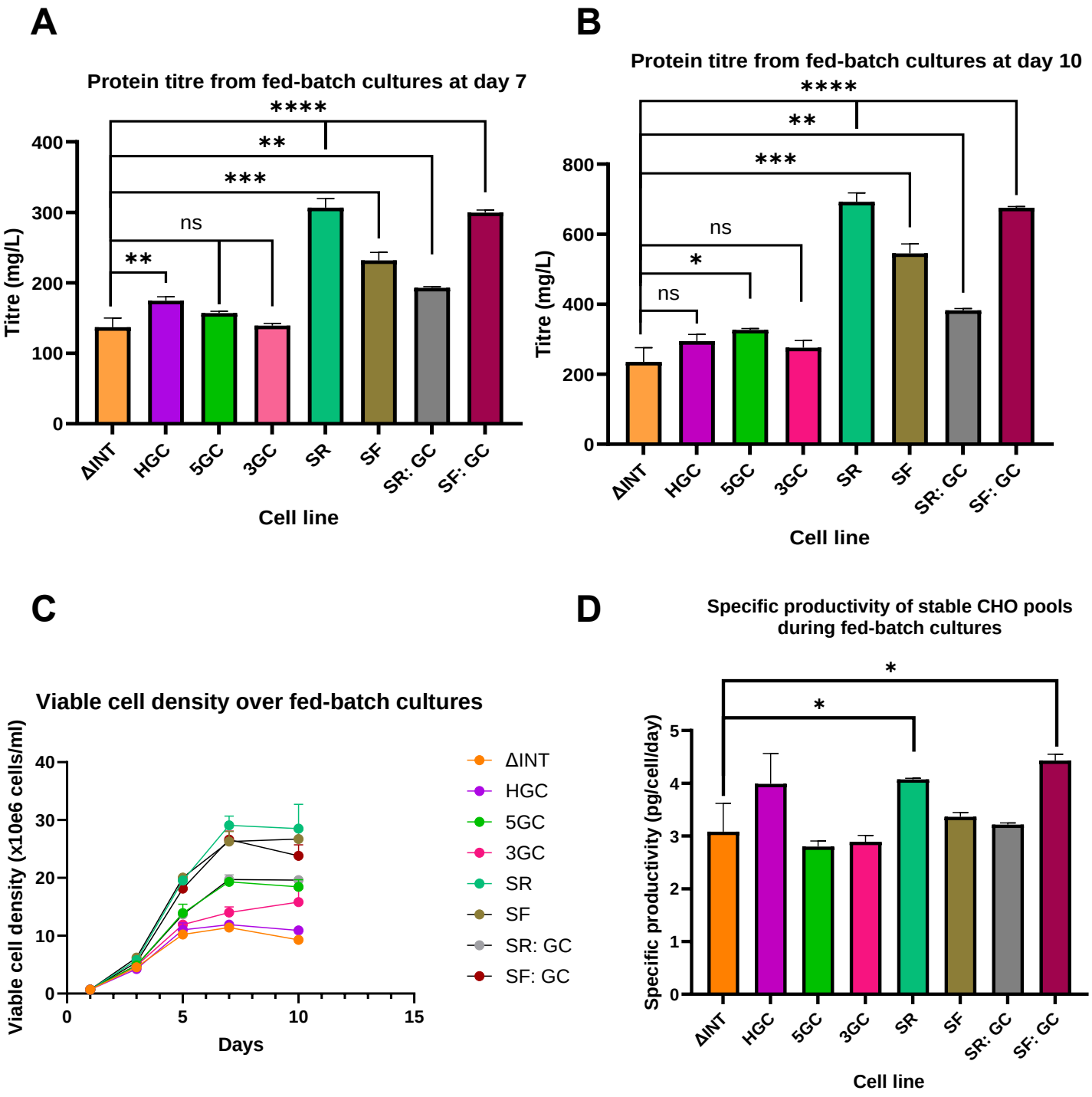


Figure 4.8: Titre, growth and specific productivity from fed-batch cultures of stable CHO pools expressing Fc-FP.

A) Fc-FP titre data at day 7 of a fed-batch culture from stable CHO cell lines (n=3).**B)** Fc-FP titre data at day 10 of a fed-batch culture from stable CHO cell lines (n=3). **C)** Viable cell density of CHO stable pools over fed-batch culture **D)** Specific productivity of the stable CHO cell lines during fed-batch culture (n=3). Data generated by Dr Sarah Dunn at AstraZeneca.

4. 9 Summary

In this chapter, we investigated the effect of placing known RNA binding protein recruitment motifs into the 5'UTR of an intronless Fc-FP sequence to test the effect it has on protein expression for biomanufacturing.

Firstly, we inserted a 55nt sequence previously identified as containing multiple binding sites for SRSF1 and SRSF7 (Khan et al. 2021) into the 5'UTR of an intronless sequence. We saw that addition of this motif series improved protein expression in a transient transfection in CHO cells but had no effect on the total mRNA levels. We then generated a CHO cell line stably expressing the SR sequence to compare to our previously generated ΔINT CHO cell line. We saw slight increases in protein expression, especially in the cell media, but did not see any changes in the total RNA levels from q-PCR analysis.

As SR proteins have been linked to translation, we used polysome profiling to conduct an initial examination of the translation differences between the SR and ΔINT sequences. Our preliminary results suggest that the inclusion of the SR sequence may help increase the translational rate of an intronless mRNA, but further investigation is needed to say this for certain as we lack the statistical power to draw any firm conclusions from our 2 replicates. Additionally, from our observations polysome profiling can produce quite variable date (as seen from our SR data sets) and so the protocol may need further optimisation to improve consistency between replicates.

Finally, we saw that inclusion of the SR sequence increased the Fc-fusion protein titre yield when compared to the yield from the ΔINT cell line. All this data suggests that inclusion of the SR sequence in the 5'UTR improves expression of therapeutic protein from an intronless sequence. However, from our work we cannot say for certain that the effects we see are due to interactions between the mRNA and SRSF1 or SRSF7. In the previous chapter, we investigated the effect of GC content on intronless expression and found that increasing the GC content improved mRNA export through increased binding by Alyref. The sequence containing the SR binding motifs is 55nt long but has a GC content of 64% - the same GC content as our High GC intronless sequence. Thus, the increase in expression may be caused by the increased GC content of the 5' UTR rather than SR binding through better Alyref recruitment. Furthermore, increased GC content in the 5'UTR can have effects, both positive and

negative, on translation rate, secondary structure and thermostability of an mRNA (Babendure et al. 2006). Moving forward, we would generate a random sequence of 55nt with a GC content of 64% to insert into the ΔINT 5'UTR to act as a negative control. If this still caused an increase in expression, we would know that SR binding is not the only part of this sequence effecting the expression of the intronless sequence and further work would be needed to elucidate which part of the sequence is having a positive effect. Additionally, we could use RIP against one of SRSF1 or SRSF7 to check for increased binding to mRNA to prove that the sequence recruits the proteins effectively.

Finally, we showed that combining the SR motif and the HGC sequence used in the previous chapter can significantly improve protein titre from an intronless sequence when compared to the ΔINT control. Thus, the SR motif and high GC can work together to improve protein expression. Khan et al (2021) showed that SRSF1 and ALYREF interact using immunoprecipitation, with the knockout of SRSF1 resulting in a reduction of ALYREF binding the NKILLA lncRNA (Khan et al. 2021). This result means that while we don't know exactly which part of the SR motif is responsible for recruiting ALYREF (the recruitment of SR proteins and subsequent interactions with TREX or the high GC content of the sequence), we can hypothesise that the sequence can improve ALYREF recruitment in a cumulative manner with the HGC content of the intronless CDS.

Alongside the SR sequence, we also investigated the effect of a 7-nucleotide motif that can recruit SF3b155 to mRNA which has reported interactions with members of the THO complex, a key component of TREX (Wang et al. 2019). We saw that inclusion of this motif improved protein expression and total mRNA levels when transiently expressed in CHO cells. We generated a stable pool of CHO cells expressing the SF3b sequence, however we struggled to detect clear differences in the expression levels when compared to the ΔINT control cell line with protein and mRNA levels equal. Additionally, we could detect no differences in the export of mRNA to the cytoplasm indicating that the sequence may not be having much of an effect. We also confirmed that inclusion of the sequence did not cause any cryptic splicing to occur. We measured therapeutic protein expression over a longer time and compared it to the ΔINT control line. Western analysis showed increases in protein production

when compared to the control line and while this improved production was confirmed by titre analysis of the same samples, the increase was not to a significant degree,

We were surprised by the differences we saw between the transient results and the stable cell lines for the SF3b construct. Transiently we saw clear differences at both the protein and mRNA level when we compared the Sf3b cell line to the ΔINT control line, but those differences disappear when placed into a stable context. Firstly, this demonstrates the importance of using stable systems to mimic biomanufacturing rather than transient experiments, as an explanation for our observations lies in better transfection efficiency and plasmid expression of the SF3b Fc-FP sequence compared to the control ΔINT sequence. There may be another observation for these differences. A copy of the SF3b motif naturally lies in the ΔINT control with the sequence GAAGAAG present in the Fc-fusion protein. We have observed that the ΔINT control expresses better from a stable pool than from transient experiments and this sequence may go some way as to explaining that as it may facilitate binding of SF3b1 and recruit TREX members to the ΔINT control, helping to improve expression. A further experiment would be to mutate this sequence in the control ΔINT line and see if that has any effect on the expression of therapeutic protein. Additionally, Wang et al (2019) reported that multiple copies of the SF3b motif had a cumulative increase on nuclear export, Hence, our inclusion may be influencing expression but is masked somewhat by the presence of this motif already.

As both the SR and SF3b sequence were present in the UTR and were shown to bind members of the TREX export pathway (Khan et al. 2021, Wang et al. 2019). The work presented previously showed that increasing the GC content of an intronless sequence improved mRNA export through better ALYREF recruitment. Therefore, we hypothesised that these elements may work cumulatively to improve expression of therapeutic protein from an intronless sequence. We inserted these sequences into the 5′ UTR of the HGC sequence to generate the respective SR: GC and SF: GC cell line. We analysed protein expression over 6 days and observed a significant increase in titre from the SR:GC cell line over the time course compared to the ΔINT but saw no increase in titre from the SF: GC cell line.

To assess the actual therapeutic potential of our cell lines, we sent the stable CHO cell lines to AstraZeneca. This allowed side by side comparison of all the cell lines we

generated to test the production of Fc-FP in an industrial system. From this data, we observe significant increases in protein titre from the SR/SF/ SR: GC and SF: GC cell lines compared to the ΔINT control. However, there were significant differences in the growth rates between the high producing cell lines and the ΔINT which would explain the significant increases in titre production. Therefore, the specific productivity was calculated to work out protein production per cell. From this, we see that the SR cell line and the SF: GC cell line both produce more protein than the ΔINT control line. Thus, we can conclude several points. Firstly, we can conclude that the SR sequence can improve protein production from an intronless sequence. The exact mechanism for this improvement still requires further investigation, however our work suggests a potential increase in translation of the mRNA. Secondly, while both the insertion of the SF3b sequence into the 5′ UTR and increasing the GC content of the CDS cause increases in titre and specific productivity, combing the two features causes greater increases in protein production from an intronless sequence. Thus, these two features can be combined to successfully increase therapeutic protein production from an intronless sequence.

Our work in this chapter demonstrates that the inclusion of RNA binding recruitment motifs in the 5’ UTR of an intronless therapeutic protein can improve its expression, although exact molecular interactions and mechanisms remain unknown. We also conclude that combining the previously presented work on GC content with these recruitment motifs remains a viable strategy for improving therapeutic expression from an intronless Fc-FP.

Chapter 5: Investigating the effects of Periphilin and HnRNPU silencing on intronless Fc-FP production

In this chapter, we analysed the effects of silencing two cellular proteins/complexes: the HUSH complex and HNRNPU. We show that our preliminary work into the HUSH complex, contrary to published work had no effect on the levels of intronless fusion protein when silenced. Additionally, we show that the knockdown of the protein HNRNPU can be used to increase the protein production from an intronless construct, but it's applications in biomanufacturing may be limited due to adverse effects on cells.

5.1 Silencing the HUSH complex

Cells employ a variety of mechanisms and defences in the eternal struggle to defend their genomic integrity from invasive DNA elements. Modification of histones, such as the trimethylation of Histone 3 at the 9th Lysine residue (H3K9me3), and methylation of DNA at CpG islands are well characterised marks of transcription repression (Greenberg & Bourc'his. 2019, Spencley et al. 2023). The Human Silencing Hub (HUSH) is a complex consisting of three proteins: TASOR, MPP8 and PPHLN1 specialising in epigenetic repression by recruitment of the methyltransferase SETDB1 to targeted sites (Figure 5.1). MPP8 contains a chromodomain which, alongside PPHLN1's ability to bind to nascent RNA, coordinates the recruitment of HUSH to loci already enriched in the repressive H3K9me3 modification (Tchasovnikarova et al. 2015).

Retrotransposons are mobile elements that are silenced by the HUSH complex and due to their mobility, HUSH must have mechanisms for detecting and silencing them no matter their position within the genome. Sezcynska et al investigated how the HUSH complex goes about recognising mobile elements. They showed that HUSH can recognise a variety of both mobile elements and foreign DNA, indicating that HUSH repression is not based on epigenetic memory. Additionally, expression from a cDNA is lower than that of its genomic counterpart, indicating HUSH can repress the intronless cDNA sequence. They then concluded that HUSH targeting can be repressed by intronic sequences in the DNA, regardless of whether splicing is active on the target intron (Sezcynska et al. 2022).

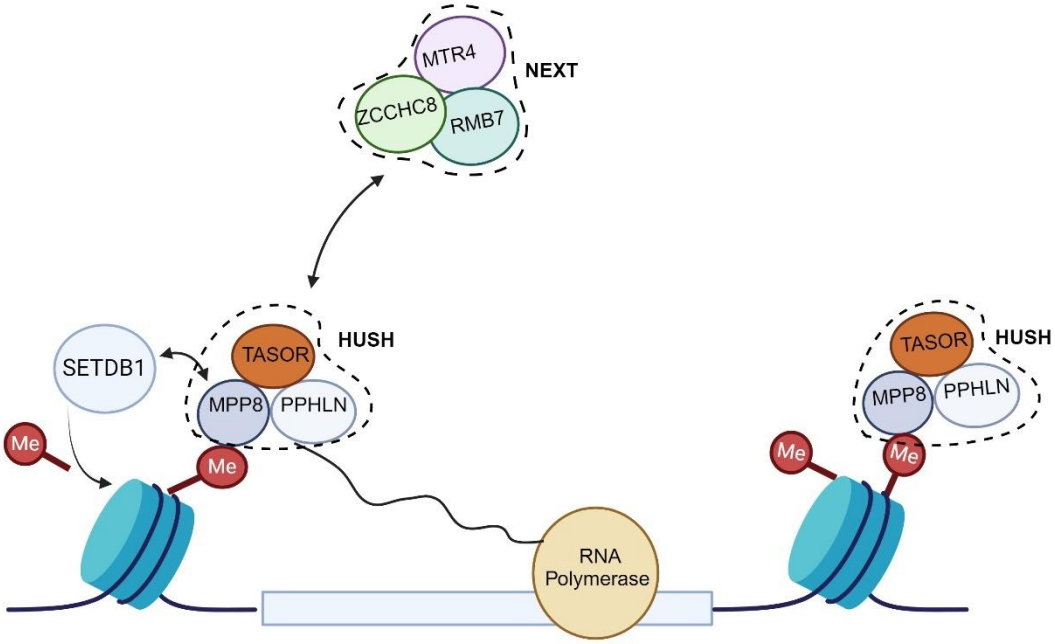


Figure 5.1: **Schematic of the HUSH complex**.

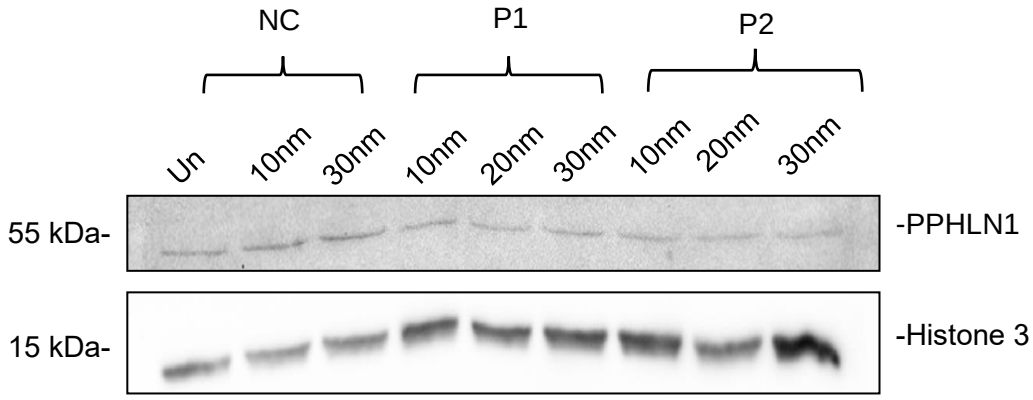
A schematic of the HUSH complex. HUSH is targeted to chromatin sites via the chromodomain of MPP8 and the ability of PPHLN1 to bind nascent RNA. HUSH recruits the methyltransferase SETDB1 to histones to repress transcription while nascent RNA is targeted by the NEXT degradation complex. Based on an illustration by Garland et al (2022).

As this project focuses on improving the expression of intronless biotherapeutics, we identified the HUSH complex as a valid target for silencing that has potential to increase expression from an intronless sequence, especially in a stable cell lines which are used often in industrial production.

To examine the effects of HUSH complex knockdown we selected periphilin as our silencing target. Periphilin binds to nascent RNA from transcribed loci and has a preference in binding RNA from young L1 retrotransposons (Sezcynska et al. 2022). This means that as periphilin engages in selecting the targets of HUSH for silencing, its knockdown may prevent any HUSH binding to the expression locus of our intronless fusion protein.

We deigned two siRNAs against the CHO periphilin sequence (termed P1 and P2) and tested their effectiveness using western blotting and qRT-PCR. We struggled to find an effective antibody that gave a strong signal against CHO periphilin on western blot (Figure 5.2.A) and despite trying different lysis conditions and western loading we found that western blotting to prove the knockdown of the protein was too inefficient and too variable to be useful for testing the knockdown of periphilin. We then decided to use qRT-PCR to measure the PPHLN1's RNA levels as a measure of knockdown effectiveness (Figure 5.2.B). While this could not provide us with actual protein knockdown conformation, an analysis of the RNA levels alongside what information we could get from our attempts to western blot gave us confidence that the siRNAs were influencing periphilin expression.

A



B

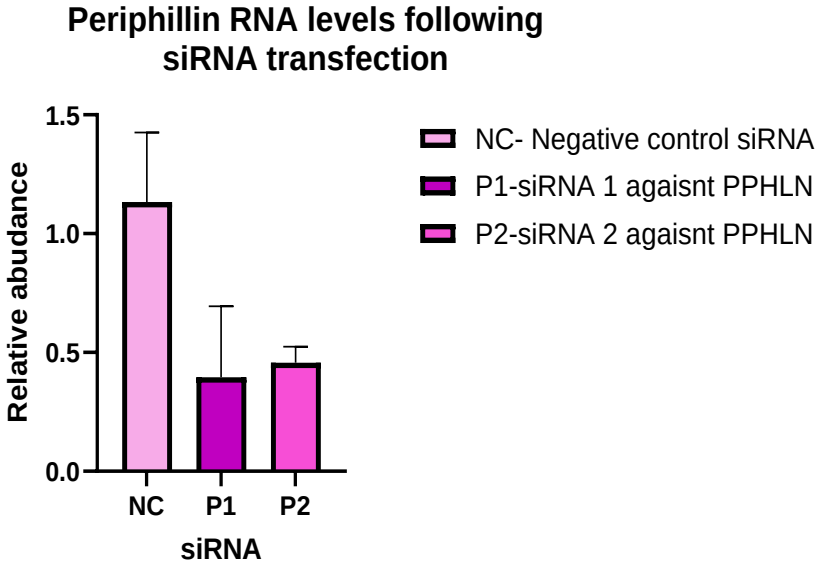


Figure 5.2: Knockout of CHO Periphilin using siRNA.

A) Western blot showing CHO periphilin levels 72 hours after siRNA transfections. Un = untransfected samples, NC = Negative control (scrambled siRNA), P1 = 1st siRNA against periphilin, P2 = 2nd siRNA against periphilin. Histone 3 levels are shown as a loading control. **B)** qRT-PCR panel showing Periphilin 1 mRNA levels (n=2). All samples normalised to the level of GAPDH mRNA.

5.2 The effect of HUSH knockdown on intronless RNA

After confirming that our siRNAs were influencing the levels of periphilin mRNA, we moved forward to analyse the effects of periphilin knockout on intronless therapeutic expression. We initially looked to see what effect the periphilin knockout was having on expression of intronless RNA (Figure 5.3.A). If the HUSH complex participated in repression of the expression, we would expect an increase in the amount of target RNA as knockdown would help relieve the repression. To our surprise, we detected the opposite, with knockdown of periphilin causing a slight reduction in the amount of intronless RNA present in the cells. While this is contrary to our expectations, we theorised this was an experimental effect from the cells being stressed following transfection of the siRNA.

As we could not verify protein levels, we considered the possibility that we were not effectively inhibiting HUSH complex formation to release inhibition. To this end we designed siRNAs against TASOR, the main scaffold of the HUSH complex to further knockdown the complex. We transfected our stable ΔINT CHO cell line with siRNAs against TASOR and analysed the TASOR and PPHLN1 RNA levels via qRT-PCR (Figure 5.3.B). As an additional control we also co-transfected siRNAs against TASOR and PPHLN1 to ensure complete knockdown of the complex. We then looked at ΔINT Fc-FP RNA levels via qRT-PCR. As with the PPHLN1 knockouts, we could observe no increase and in several samples a decrease in the amount of intronless mRNA present, indicating that knocking out a different member of the HUSH complex made no difference. In addition, the double knockout of TASOR and PPHLN1 also caused a decrease in intronless RNA expression compared to both the NC and the ‘Untransfected’. While our work is only 2 replicates and we cannot fully confirm knockdown of the HUSH complex at the protein level, we see no evidence that knocking out the HUSH complex alleviates silencing of an intronless Fc-FP sequence. There may be benefits to silencing of HUSH for biomanufacturing however further work would be needed to examine this.

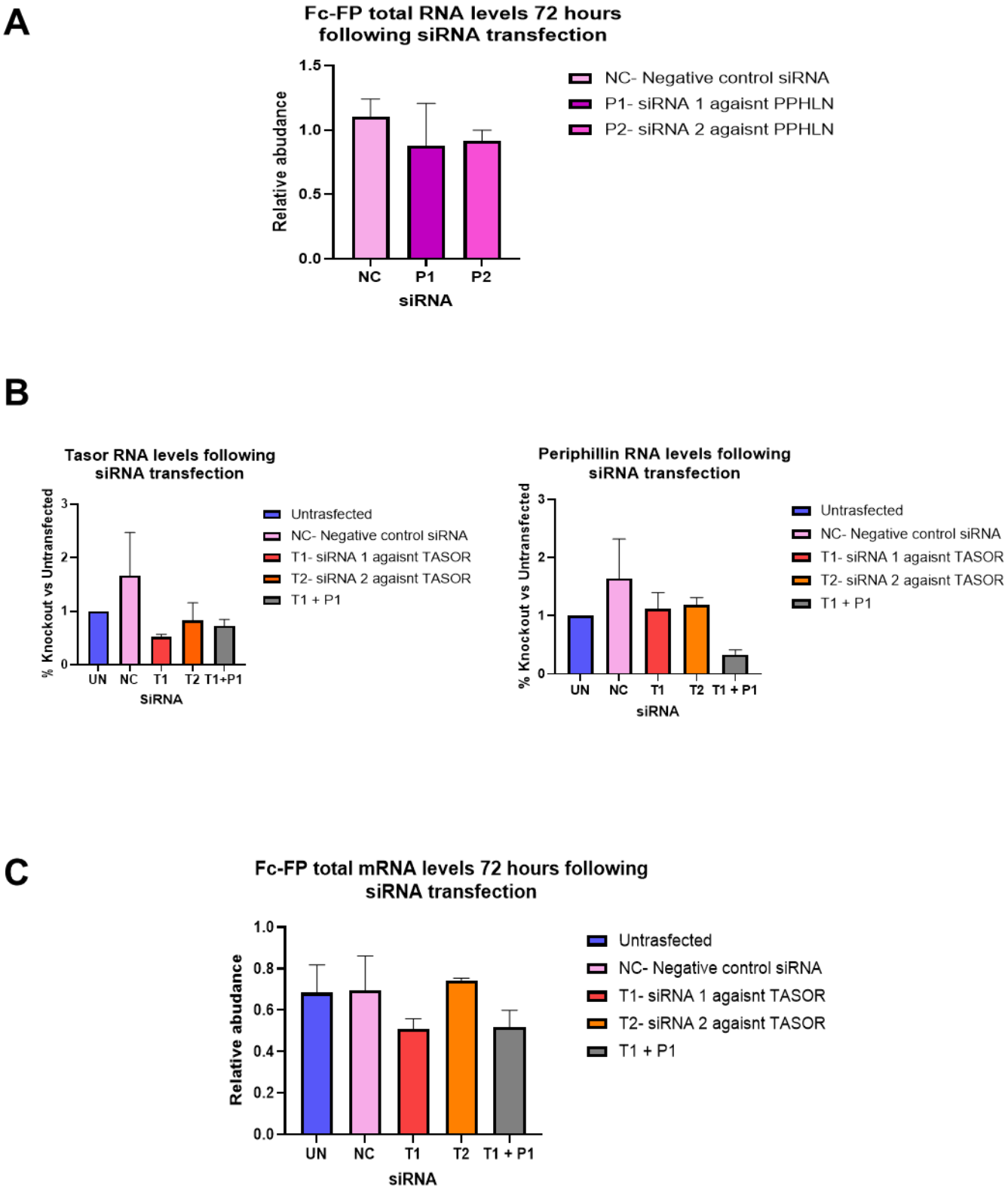


Figure 5.3. The effect of HUSH knockout on intronless RNA expression

A) qRT-PCR panel of ΔINT Fc-FP RNA in stable pools in PPHLN1 knockout (n=2). Samples are normalised to the level of GAPDH mRNA. **B)** qRT-PCR panels of TASOR, PPHLN1 and ΔINT Fc-FP RNA levels following siRNA transfection (n=2). All samples are normalised to the level of GAPDH mRNA.

5.3 Using Knockdown of HNRNPU to improve expression of an intronless therapeutic protein

Whilst we were investigating the effects of HUSH complex silencing on intronless therapeutic expression, we wanted to look for other proteins/ complexes that may impact protein production. HNRNPU was originally identified as 120kDa protein involved in chromatin scaffolding (Romig et al, 1992) but has since been implicated in a variety of cellular processes including regulation of topologically associated domains (TAD) boundaries, chromatin loop regulation and roles in transcription initiation and elongation (Zhang et al. 2019). Knockdown of the protein hnRNPU has also been shown to cause an increase in transcription and splicing rates of mRNA and cause an increase in expression, with a noticeable effect on intronless transcripts (Wilson lab, unpublished). We decided to test if knockdown of HNRNPU had any effect on the expression of an intronless Fc-FP construct.

To initially test if knockdown of HNRNPU was a viable strategy for increasing the production of therapeutic protein from an intronless sequence, we utilised an existing HCT116 cell line generated by Catherine Heath (Wilson lab) that contained an AID tag to allow degradation of HNRNPU within 24 hours upon addition of Auxin. We transiently transfected this cell line with the WT Fc-FP containing two introns and the ΔINT cell line which had the introns removed (Figure 5.4.A). The cells were transfected and then 24 hours later auxin was added to cause the degradation of HNRNPU in the transfected cell for a further 24 hours to allow complete degradation of the protein and then analysed the cell lysates using western blot (Figure 5.4.B). We observed slight increases in Fc-FP protein production in the WT and greater increases in the ΔINT in the HNRNPU-AID cell line vs the HCT116 control samples. There is a certain degree of constitutive degradation of HNRNPU-AID in the absence of auxin which may explain the differences in protein expression between HCT116 and HNRNPU-AID before auxin addition. We also wanted to analyse the effect of HNRNPU knockdown on protein levels in the cell media, so we transfected the ΔINT construct into HCT116 and the HNRNPU-AID line (Figure 5.3.C). We saw improved protein expression in both the CM and cell lysates when compared to the HCT116 control, giving us an preliminary result suggesting that depletion of HNRNPU improves expression of an intronless therapeutic protein.

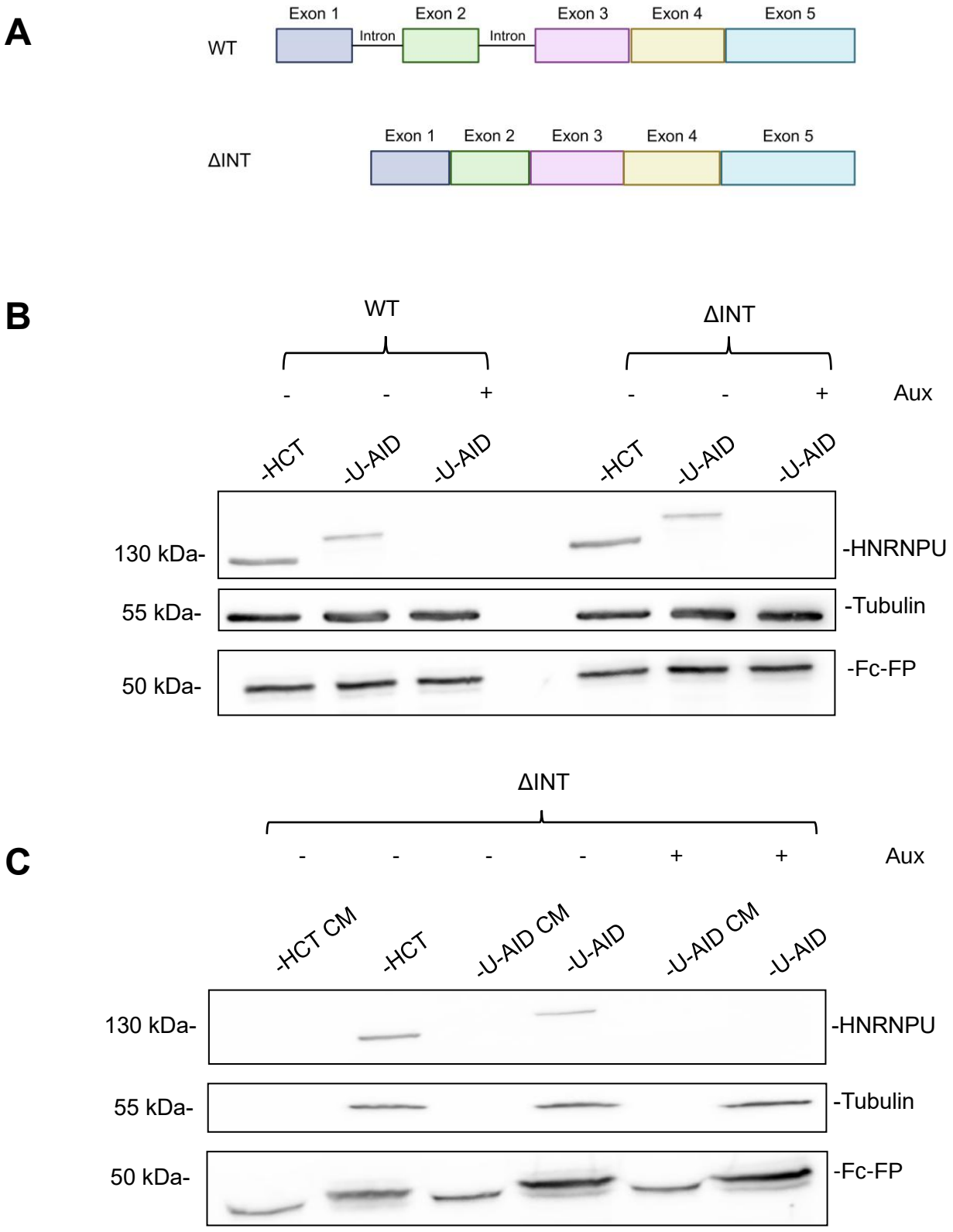


Figure 5.4: Knockdown of HNRNPU to improve intronless protein expression.

A) Schematics of WT Fc-FP and ΔINT Fc-FP transfected. **B)** Western blot showing Fc-FP protein expression upon depletion of HNRNPU. Tubulin is shown as a loading control. **C)** Western blot showing Fc-FP protein expression upon depletion of HNRNPU. Tubulin is shown as a loading control. Cell Media (CM) is loaded in equal volumes while lysate are loaded according to protein concentration.

5.4 Depletion of HNRNPU in CHO cells can affect intronless therapeutic expression

As our initial tests of HNRNPU in HCT116 cells yielded positive results, we transferred this work into CHO cell cells. While the HNRNPU-AID line was effective at degrading the protein, due to time and experimental challenges in generating this degron line in CHO cells, we decided to use siRNAs against HNRNPU to reduce protein levels. We designed 2 siRNAs against the CHO HNRNPU to knockdown the protein and test the effect of the knockdown. We initially tested the effectiveness of the siRNAs by transfecting them into our stable CHO cell pool stably expressing the ΔINT Fc-FP (Figure 5.5.A) and analysing protein levels after 72 hours via western blot analysis (Figure 5.5.B). We observed good knockdowns of HNRNPU from both transfected siRNAs but observed no real changes in the level of ΔINT protein in the cell media or cell lysate. However, we were sceptical about the protein production observed from this western blot as we observed a significant reduction in growth amongst the CHO cells transfected with the siRNAs.

To more accurately measure the protein production differences when HNRNPU is knocked down, we counted how many cells were alive in the wells after 72-hour post transfection and took equal amounts of cells to extract lysates from and analysed protein levels via western blotting anaylsis (Figure 5.5.C). From this, we saw one replicate with slight increases in protein production compared to the NC siRNA and one replicate with much better protein production compared to transfection with the NC siRNA. We chose not to analyse the amount of protein present in the cell media, as the differences in cell counts between the knockouts and the NC siRNA would mean any changes in protein production would most likely be masked, as we saw in initially (Figure 5.5.C) This suggests that knockout of HNRNPU in CHO cells may be a viable strategy for improving expression of an intronless therapeutic protein.

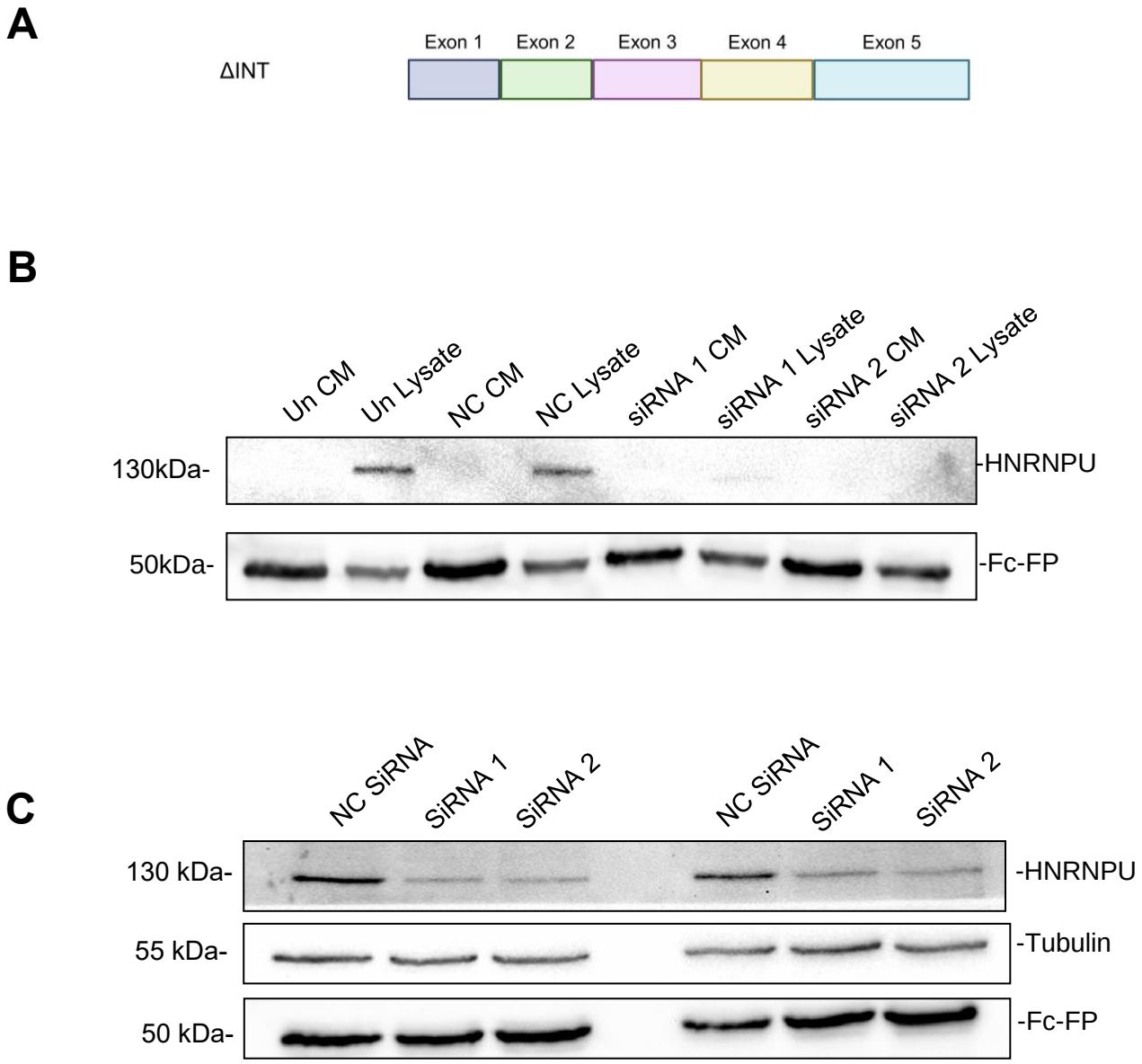


Figure 5.5. Depletion of HNRNPU affect expression of an intronless protein in CHO cells.

A) Schematic of the ΔINT Fc-FP expressed from the CHO cell pools transfected. **B)** Western blot showing HNRNPU knockout in a ΔINT stable CHO pool. Untransfected cells (Un) and Negative control scrambled siRNA (NC) are show as controls. **C)** Western blot showing ΔINT Fc-FP expression from siRNA knockout with equal amounts of cells lysed. Tubulin is shown as a protein loading control.

5.5 Summary

In this chapter, we analysed the effects of siRNA knockdowns of 2 different proteins/complexes on expression of a stably expressed Intronless fusion protein.

Firstly, we designed and tested siRNAs against PPHLN1, a member of the HUSH complex, which is involved in the silencing of intronless retrotransposons and foreign DNA elements (Sezcynska et al. 2022). Thus, we hypothesised that the HUSH complex may have a role in silencing expression from an intronless locus and could be a good target to knockout to improve expression of an intronless therapeutic protein. However, knockdown in PPHLN1 resulted in reduced Fc-FP mRNA levels, having the opposite effect to our hypothesis. We then tried to knockout the scaffolding protein of the HUSH complex, TASOR, theorising that this may have more of a destabilising effect on the complex, however this caused a similar effect on the Fc-FP mRNA as knocking down PPHLN1. Based on this, we conclude that knocking down the HUSH complex has no effect on expression of an intronless therapeutic protein. However, this result could be explained by experimental/technical difficulties and may not be as clear cut. Firstly, we were unable to find good antibodies for members of the HUSH complex that reacted well with the CHO forms of this complex. Thus, while we can conclude that our siRNAs were able to reduce the RNA levels of these proteins, we cannot draw any conclusions on how well these knockdowns were mirrored at the protein level and therefore how depleted the complex was.

Another explanation lies in the actual cellular role of HUSH. HUSH defends the genome from invasion from foreign DNA by targeting long intronless section of DNA. Knockout of TASOR is lethal in the preliminary stages of development in mice (Harten et al. 2014) and MPP8 knockdown has been implicated in several cancers with roles in regulating interferon response and increases in DNA damage (Tunbak et al. 2020). Hence, the reason for the reduction of Fc-FP RNA seen may be due to stress effects in the cells upon knockdown of HUSH due to loss of repression of other HUSH targets. As the aim of this work is to improve production of therapeutic protein in biomanufacturing, we reason that the knockdown of the HUSH complex may not be viable for industrial production as loss of HUSH can alleviate transposon repression which could have adverse effects on the therapeutic protein produced.

Alongside the effects of HUSH knockdown, we investigated the effects of knocking down HNRNPU on expression of intronless therapeutic protein. As knockdown of HNRNPU causes a general, global improvement of transcription and splicing rates, we identified it as a potential target for increasing protein yield from an intronless sequence (Ang li, Wilson lab. unpublished). To evaluate the viability of this strategy, we initially used a HCT116 HNRNPU-AID cell line previously generated in the lab by Catherine Heath. This cell line causes complete degradation of cellular HNRNPU compared to HCT116 cells within 24 hours of Auxin addition to the media. Using transient transfection, we saw that knockdown of HNRNPU caused an increase in transient ΔINT Fc-FP expression via western blot analysis from both cellular lysate and in the cell media. This improvement was visible in the HNRNPU-AID cell line before the addition of auxin; this is explained by the decreased base HNRNPU levels in the cell line resulting from the tagging process.

We then moved this work into CHO cells, using a stably expressing ΔINT Fc-FP pool. We used siRNAs to knockout CHO HNRNPU to examine the effect this had on Fc-FP expression via western blot analysis. While we observed good knockdown of CHO HNRNPU upon transfection of our siRNAs, we struggled to see any changes in protein production when compared to untransfected or NC siRNAs, especially in the cell media samples. One observation from this experiment was a clear decrease in cell growth rates and viability upon knockdown of HNRNPU in the CHO cells. This effect was more severe than the slightly decreased growth rate displayed by the HNRNPU-AID line. To counteract this effect, we repeated the experiment but collected equal amounts of cells from the NC and the HNRNPU knockouts to give a better representation of protein production per cell. This showed an increase in ΔINT Fc-FP levels from cellular lysates upon knockdown of HNRNPU compared to the NC. However, to achieve equal cell counts for the western, the entirety of the 24 well was harvested for the HNRNPU knockout samples, while only half the cell content of the well was harvested for the NC.

Experimentally, while this provides the information of how HNRNPU knockout affects intronless Fc-FP production per CHO cell, from a biomanufacturing point of view it is difficult to justify the knockout of HNRNPU to improve production due to the observed growth defect. Even with the indication that knockout of HNRNPU improves cellular production, the better growth observed from the control would cancel out this effect,

as seen in our initial CHO tests. Thus, we concluded that while knockdown of HNRNPU does show potential for improving therapeutic protein expression from an intronless sequence, further work is needed to understand and optimise this system for uses in biomanufacturing.

Chapter 6: Discussion

In this project we set out to improve protein production of an Fc-fusion protein from an intronless sequence in CHO cells. In the biopharmaceutical industry, high levels of protein expression are important for improving yield of therapeutic protein, which in turn reduces the production costs of biotherapeutic drugs. This then allows a reduced final cost for the consumer, making the drugs more available for patients. High expression levels are usually achieved by the inclusion of introns in the CDS due to the positive effects they have on gene expression. However, as we demonstrated in our Fc-FP stable line, mis-splicing events can cause the generation of aberrant protein structures. This both decreases the yield of therapeutic protein and the aberrant protein has to be removed to avoid contamination of the product. Therefore, generation of intronless sequences with increased protein expression is of great interest to the biopharmaceutical industry.

We initially investigated the effect of increasing the GC content of the CDS and used cellular fractionation and RIPs to characterise the improved mRNA export effects we observed. Through this work, we identified RNA export as a targetable process, leading us to use recruitment motifs for known RNA binding proteins (SRSF1/SRSF7 and SF3b1) that have been implicated in mRNA export. From these we saw improved expression of an intronless Fc-FP, however the exact molecular basis for this improvement requires further elucidation. We then combined the sections of work, showing the potential of combining the increased GC content and the RBP motifs to work in tandem to increase Fc-FP production from an intronless sequence, further supported by work completed by our industrial partners measuring titre and growth. Finally, we investigated the effects of siRNA silencing on the HUSH complex and HNRNPU to improve expression from an intronless sequence, showing the potential of silencing these proteins but also the limitations. Here, we discuss these results and their implications in more detail.

6.1 Increased GC content improves expression of an intronless sequence

The association between highly transcribed genomic loci and high GC content has been well reported, alongside reports that increased GC content improves protein expression from an intronless reporter (Mordstein et al. 2020). We modified the CDS by utilising 3rd codon base wobble to increase the GC content of an intronless

sequence while maintaining the sequence integrity. While a suitable approach, this method was limited by the already high GC content of the ΔINT control we used, meaning an increase of 4% in GC content was the max we could generate while enabling synthesis of the sequence. On this basis, we also generated a lowered GC control, allowing us to confirm the opposite effect and giving us confirmation that GC content was affecting sequence expression.

Transient transfections showed an increase in protein production, total mRNA levels and mRNA export to the cytoplasm upon increasing the GC content when compared to the levels of the ΔINT control. Additionally, protein expression and total mRNA levels from the low GC control were drastically reduced compared to the high GC and ΔINT sequences, confirming that the GC content of an intronless sequence effects it's expression. We considered that the increased levels of mRNA export resulted from an increase in RNA export factor binding. We used RNA immunoprecipitation to observe an increase in binding of the key export factor ALYREF to GC rich mRNA over the ΔINT control. While no canonical binding site for ALYREF exist, due to its involvement in the export of most mammalian mRNAs, we propose this binding is due to increased secondary structures present within the mRNA. This theory is supported by several pieces of evidence: ALYREF contains a similar RNA binding domain structure to that of the protein FUS, a nuclear RNA binding protein shown to bind hairpin and G-quadraplex RNA secondary structures. Thus, it is likely that ALYREF can bind to mRNA in a similar fashion, recognising RNA secondary structure to bind to mRNA. Additionally, knockdown of ALYREF or other TREX members using siRNAs has shown a preferential export reduction for mRNAs high in GC content, further implicating GC content as a factor in ALYREF recruitment (Zuckerman et al, 2020). To prove this theory, it would be prudent to design an intronless sequence that is capable of greater variability in GC content than the Fc-FP sequence used here, with the potential to design specific secondary structures, such as hairpins or G-quadraplexes, into the mRNA and evaluate their ability to recruit ALYREF.

As therapeutic protein is normally produced from CHO cell lines with expression cassettes incorporated into their genome, we generated stably integrated cell lines expressing the ΔINT and HGC sequence. We continued to observe increased protein production and mRNA export from the HGC sequence, but in contrast to our transient observations we could not detect significant differences between the total mRNA

levels. These results support our transient observations that an increase in RNA export improves expression of a therapeutic protein. This suggests that the differences in total mRNA level seen transiently could be in part caused by the cellular processing of a transient, intronless cDNA, potentially through repression of the intronless transcript through cellular systems such as the HUSH silencing complex (Sezcynska et al. 2022). If this is true, this also suggests that GC content enables better expression in a transient system by avoiding cellular repression systems, however further investigation would be required to confirm this effect. Importantly, we continued to see improved protein production when the cells were grown for a longer period of 6 days and while titre analysis of these samples failed to show an increase in cell media protein titre, we were encouraged by the fact each titre sample from the HGC line had improved protein production compared to the ΔINT control grown alongside it, though variance between replicates prevented this from being statistically significant.

To allow further optimisation for uses in biomanufacturing, we evaluated the effect on expression of increasing the GC content of just the 5′ or 3′ end of the sequence. Interestingly we observed increased protein expression at both 48 hours and over 6-day time courses, but this increased expression was not mimicked at the RNA level, with no significant differences between the 5GC/3GC compared to the ΔINT control in either total mRNA levels or export ratios. These observations suggest that overall GC content may be a more defining factor in expression from an intronless sequence rather than its localisation. These observations contrast with that of Mordstein et al (2020) who reported that increased GC content at the 5′ end of an intronless GFP sequence was enough to improve protein and mRNA expression (Mordstein et al. 2020). The low GC increase present between the ΔINT control and 5GC sequence used in our experiments could explain why we fail to see a significant increase in expression as they reported.

Furthermore, these results suggest that the changes in expression seen are not entirely down to mRNA export, suggesting a further role for GC content in improving expression from an intronless sequence. The answer could lie in RNA stability, something we did not assess in this project, which could be investigated using a 4-Thiouridune (4SU) time course to measure the stability and degradation rate of the mRNA from low and high GC sequences. The idea of increased mRNA stability when GC is increased is further supported by the competition between ALYREF and MTR4

for determining the export or degradation of an mRNA. The improved ALYREF binding detected from the full-length GC mRNA could provide improved mRNA stability by protecting it from degradation by the NEXT complex (Fan et al. 2017). Whilst our data provide evidence that increased GC content can improve therapeutic protein expression from an intronless sequence through improved export factor binding, the interplay between RNA export and stability in this context requires further investigation.

6.2 Recruiting SRSF1/7 and SF3b1 to intronless mRNA improves protein production

Our work on GC content identified RNA processing as a suitable target for improving therapeutic protein production from an intronless sequence. Recent work exploring the cytoplasmic localisation of an intronless lncRNA, NKILLA, revealed a 55nt sequence containing known binding motifs for the proteins SRSF1/SRSF7, SR proteins with roles in mRNA processing, export, and translation (Khan et al. 2021).

Insertion of this motif into the 5′ UTR of the ΔINT control showed increases in protein production compared to ΔINT control but showed no difference in the RNA levels, observations that were mirrored in the stable integrated CHO cell lines we generated. This led us to the hypothesis that the SR sequence was affecting mRNA processing and translational efficiency in the cytoplasm rather than nuclear processing. To investigate, we used polysome profiling to assess the translational efficiency of the mRNA and the ribosomal association. Due to time and experimental limitations, such as the large numbers of cells required for this technique, we could only collect 2 replicates from the ΔINT and SR stable cell lines.

Thus, we present preliminary data showing an increase in polysome association in the SR motif RNA compared to the ΔINT control. The two replicates were sufficient to see an increase in loading of SR mRNA into low efficiency polysomes when compared to the ΔINT mRNA. While we cannot draw any conclusions from the high efficiency polysome data, an increase in the amount of low polysome associated mRNA in the SR cell line could be a good clue as to the behaviour in the high polysome associated fractions. Further polysome replicates would allow clarification of this data and confirmation of this result. However, an improvement in translational efficiency of the SR containing mRNA over the ΔINT control fits our observations and matches reports

from the literature suggesting SR proteins help initiate translation (Michlewski et al. 2008).

Subsequently, encouraged by the effects we saw using the SR motif, we inserted a 7nt sequence in the 5′UTR previously identified as recruiting the SF3b1 complex to mRNA. We were drawn to this motif due to the small size of the inserted sequence, making it of more valuable use and flexibility in expression construct design for biomanufacturing. Additionally, insertion of this sequence was shown to be capable of improving export in a dosage-dependent manner, with 3 copies of the sequence causing greater cytoplasmic localisation of mRNA from a cDNA sequence than 1 copy (Wang et al. 2019). Finally, the complex has direct protein-protein interactions with members of the TREX complex, providing a clear method for improved export. These direct interactions come as no surprise considering that splicing, where SF3b forms part of the U2 snRNP, participates in recruitment of TREX to mRNA (Masuda et al. 2005). Upon transient transfection of CHO cells, we observed improved protein production and increased total mRNA levels from the SF3b construct when compared to ΔINT control. Interestingly these effects were not mirrored when 3 copies of the sequence were inserted into the UTR of the ΔINT, instead we observed a reduction in total mRNA levels. While contrary to our expectations, we hypothesise this is due to a design flaw, in which the motifs were placed back-to-back and might have caused steric clashes between the recruited SF3b1 complexes at the 5′ end of the mRNA, reducing the overall expression from the sequence.

We then generated a stable CHO line expressing the SF3b binding sites and compared its expression to the ΔINT control. Surprisingly, we saw slight increases in protein production, but no significant changes in total mRNA levels or in the RNA export ratios. This observation could be explained by the presence of one copy of the SF motif naturally present in the ΔINT sequence. As Wang et al (2019) showed the motif was able to facilitate RNA export in a cumulative manner, it’s possible the lower than expected increased in protein production was a consequence of this, with two copies of the sequence slightly increasing expression over the one copy present in the ΔINT control.

As the RBP recruitment motifs were located in the 5′ UTR, this makes them compatible from a design point of view with the increased GC content of the CDS we showed in

our first experiments. Additionally, we showed the GC content acts to improve mRNA processing to increase therapeutic protein expression, theoretically functioning in an analogous manner to the SR and SF sequences and allowing compatibility between the design features. Generation of these cell lines and analysis of protein levels over a 6-day time yielded satisfactory results. In fact, the SR: GC line was the only cell line from the 6-day batch culture performed in our lab that showed a significant increase in protein titre compared to the ΔINT, and at every stage of the time course as opposed to just the final time point at 6 days. From the data we have presented we envisage two complementary effects occur with the SR:GC cell line. Firstly, the improved GC content allows better mRNA export to the cytoplasm, where the mRNA is more efficiently translated by the ribosomes via the interactions from the recruited SR proteins. It's possible that the SR motif also contributes to improved mRNA export, a possibility that we could investigate with further time through cellular fractionation, which is also given plausibility by the reported roles of SR proteins as export adaptors (Hargous et al. 2006). In contrast to the SR: GC line the SF: GC combination appeared, from the titre data, to have the opposite effect, reducing Fc-FP expression down to comparable levels of the ΔINT control. As the SF3b motif is predicted to recruit SF3b1 which in turn recruits members of TREX, like our predicted function of the HGC sequence, we speculate that the two elements fail to work together due to an oversaturation of TREX components to the point that further recruitment has little effect.

6.3 Fed-batch cultures show increases in therapeutic protein production

As this project aimed to improve protein production for biomanufacturing, it remained unclear how applicable our results where to protein production at an industrial scale. To investigate this, the 8 cell lines we generated were sent to our industrial partners at AstraZeneca. This provided two advantages over the 6-day growth experiments we had performed. Firstly, it allowed cells to be grown over a longer period, 10 days to be precise, to allow the experiments to more closely align with those used in biomanufacturing. This also allowed all cell lines to be evaluated side by side, allowing direct comparison between all the constructed cell lines and the ΔINT control.

Analysing the titre data at the 10-day time point, we saw 4 cell lines with clear titre increases, with some samples several hundred mg/l higher, compared to the ΔINT

control from which all expression constructs are derived. The SR, SF, SR: GC and SF: GC all showed large increases in titre, while increases from the HGC, 5GC, 3GC were observable compared to the ΔINT, they were much less/not significant. While these substantial increases in protein titre look very promising, they may be misleading. The specific productivity calculations show that only two cell lines had actual increases in protein production per cell when compared to the ΔINT: the SR and SF: GC. However, the differences between the specific productivity are lower than could be expected for such significant differences in titre levels. This means the increase in titre can be attributed to the differences in cell growth between the SR, SF, SR: GC and SF: GC cell lines and the ΔINT control.

The increases in specific productivity highlight the SR only construct and the SF: GC cell line as the two key cell lines for improving protein production from an intronless sequence. The SR sequence having a significant effect on protein production is not surprising from our previously generated data, showing improved protein production and a potential improvement in mRNA translation. What was surprising is the increased specific productivity of the SF: GC cell line compared to the ΔINT control, rather than the SR: GC cell line we expected from our 6-day batch cultures.

One interesting experiment would be to repeat these fed-batch cultures, looking to analyse the growth rates between the cell lines. If the 4 cell lines that showed major increases in growth and titre did so once again, it would suggest that the changes in ΔINT sequences in these lines may have improved cell growth. While the sequence changes or inserted motifs we have used should have no effect on cell division rates, a potential explanation lies in the cellular production cost. As we see from the titre data, the ΔINT cell line is still able to express good amounts of protein. However, this may come at a large resource cost to the cell, lowering its growth rate. The changes we have implemented could improve expression from the same loci, reducing the protein production cost for the cell and leaving it more readily able to divide. Additionally, it would be informative to repeat this experiment alongside the WT cell line (containing two introns) that we built. While this cell line has a splicing defect reducing its protein production, it would be interesting to see how it compared to the protein levels from the intronless sequences we had built.

Furthermore, these observations are limited to just one biotherapeutic protein. As both the GC content modifications and recruitment motifs can be applied to another sequence, it would be important to evaluate if these methods improved expression of other Fc-fusion proteins from intronless sequences. Given the variety of mAbs, Fc-fusion proteins and therapeutics available on the market, it would also be prudent to test these methods on other forms of biologics, such as mAbs, to see if these are universal methods of improving expression from intronless sequences for therapeutic protein production.

One further point of discussion revolves around the differences between the smaller scale protein expression studies lasting 6 days and the fed-batch cultures. We expected the SR:GC to be the highest producing cell line in the fed-batch cultures based off the data from the 6 day grow ups, however we observed higher expression from the SR, SF and SF:GC cell lines than the SR:GC when compared to the ΔINT. These differences are intriguing; however, they may be a product of the experimental process involved. The fed-batch cultures were grown in larger media volumes, with greater amounts of cells and additional nutrient feed added, making the cultures hardier and less subject to variation than the 6-day cultures. Additionally, the fed-batch cultures were not analysed for protein titre until day 7, a day after the end point of the 6-day grow ups, meaning how fair it is to compare the data from the two experiments is unclear.

6.4 HUSH and HNRNPU silencing may have therapeutic production potential but further work is needed

In the final results chapter, we presented the use of siRNAs in silencing two internal cellular targets, the HUSH complex and HNRNPU. We identified the HUSH complex as a potential inhibitor of protein expression from an intronless sequence due to its role in silencing mobile elements such as L1 retrotransposon. Recent work has identified introns as a protective element from HUSH silencing, where the complex uses the presence of introns to differentiate self from non-self DNA. From an evolutionary perspective, this feature may arise from the lack of introns present in viruses, as they have compact genomes and naturally select against introns(Sezcynska et al. 2022).

The work presented focused on using siRNA knockdown of members of the HUSH complex in the stable ΔINT line to test if the loci containing the intronless sequence expressing a therapeutic Fc-FP was repressed by HUSH. Contrary to our predictions, we could detect no changes in ΔINT Fc-FP mRNA expression when members of HUSH (periphilin or TASOR) were targeted with siRNAs. In fact, knockdown of HUSH proteins had a negative effect, reducing the levels of target mRNA compared to the control samples. This observed negative effect on our mRNA levels could be for a variety of reasons. Firstly, it's possible that the reduction of mRNA is a side effect of the siRNA transfection causing stress on the cells. However, this hypothesis lacks evidence, based on the fact there were no observable changes in cell growth during the knockdown time course and no decrease in Fc-FP levels in the NC siRNA samples. A more likely explanation is that knockdown of HUSH proteins is having a negative effect on the cell due to cellular effects. As stated previously, HUSH is implicated in silencing mobile elements to protect the genome integrity. Recent work on the HUSH complex has been presented in human or mouse cell lines, which are subject to much less genomic flexibility and change than the CHO genome (Dahodwala and Lee. 2019, Seczynska et al. 2022)). The genomic flexibility of CHO leads to a further point of discussion. The aims of this project were to improve intronless protein production in biomanufacturing. Even if silencing the CHO HUSH complex did improve expression from an intronless sequence, what side effects could there be on product quality or protein production over several passages if mobile elements are allowed to roam unchecked through the genome. The potential applications of HUSH to improve intronless expression could benefit from further investigation but should be weighed up against the negatives of silencing such a complex in therapeutic protein production.

Finally, we investigated the effect of HNRNPU silencing and its potential uses to improve intronless Fc-FP expression. We used siRNAs against CHO HNRNPU to knockdown the protein in our stable cell line expressing ΔINT Fc-FP mRNA. We observed good knockdown at the protein level and saw an improvement in ΔINT Fc-FP production in the cellular lysate compared to the controls. However, we also detected a cell growth defect in the HNRNPU knockdown cells when compared to the NC siRNA, and the improvement in protein production was only visible when samples were prepared taking this growth effect into account.

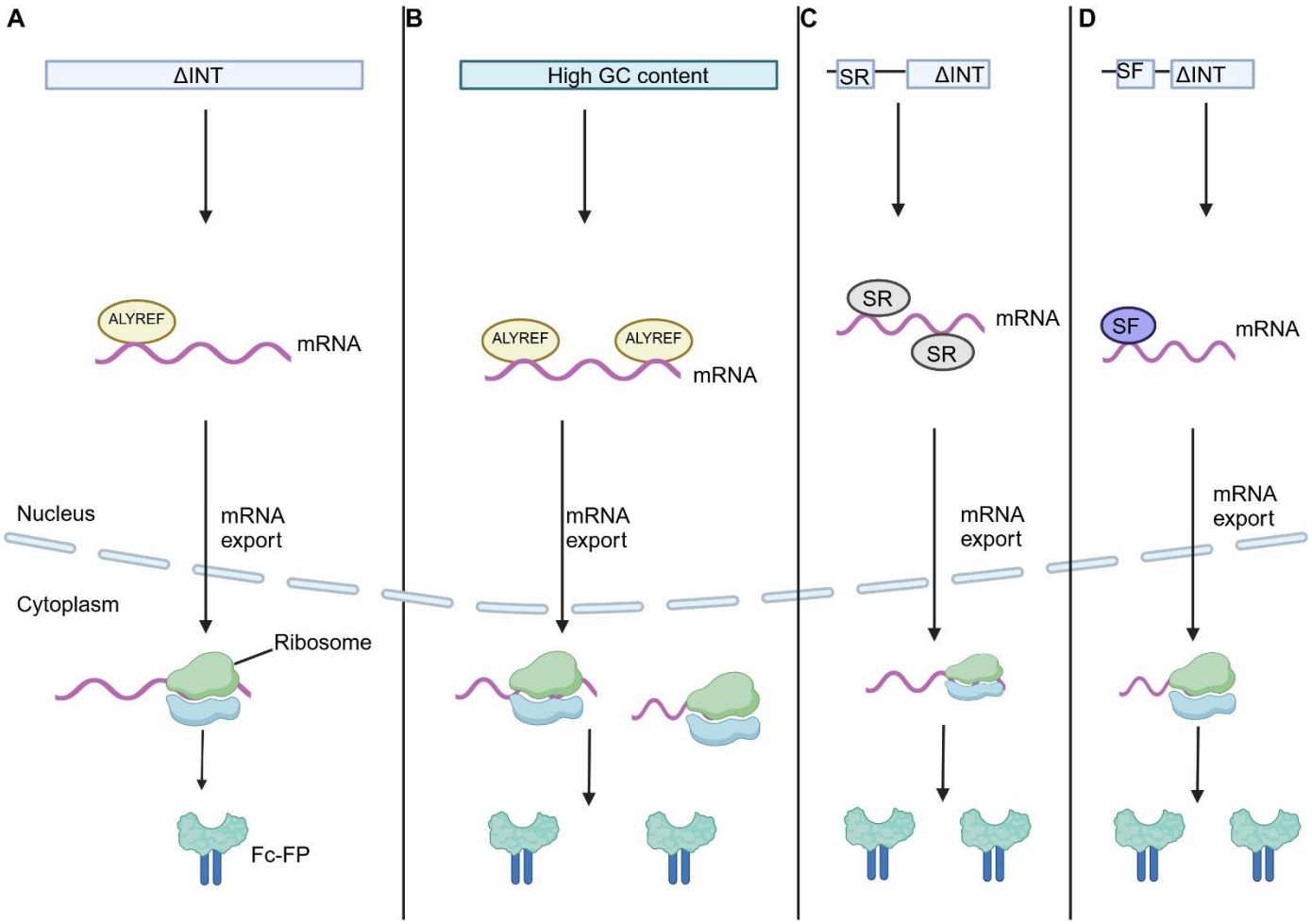


Figure 6.1: A summary of methods to improve therapeutic protein production.

A) Fc-FP protein expression from an intronless sequence. **B)** Improving the GC content of an intronless sequence improves ALYREF recruitment to the mRNA, leading to improved mRNA export and better Fc-FP protein expression. **C)** Inclusion of SR recruitment motifs for SRSF1/ SRSF7 in the 5′ UTR improves therapeutic protein production. **D)** Inclusion of SF3B1 recruitment motif in the 5′ UTR improves protein production from an intronless sequence.

Initial conclusions suggest that this makes HNRNPU an unsuccessful target for improving protein expression, as untransfected cell will make more protein due to the greater cell number, as demonstrated by our own CHO fed-batch cultures. However, it could be possible to make use of the HNRNPU knockdown towards the end of a fed-batch cultures when cell growth is reducing naturally, by using an inducible siRNA knockout system to knockout HNRNPU. An inducible system could be used to give one final push to protein production towards the end of a culture, where the growth defect would be less consequential. Therefore, from our work we suggest that the knockout of HNRNPU may be of use in the production of therapeutic protein from an intronless sequence, but further work is required to investigate the viability of this system.

In this project, we have developed methods of improving protein expression in CHO cells from an intronless sequence. By increasing the GC content of the CDS or by including the SR/SF3b1 recruitment motifs into the 5′ UTR we can improve production of a therapeutic protein. The non-specific nature of these improvements and ease of incorporation into an expression cassette make them easily adaptable to aid production of biologics from intronless sequences in both transient and stable CHO expression systems.

References

Akhtar, J., Kreim, N., Marini, F., Mohana, G., Brüne, D., Binder, H., and Roignant, J. Y. (2019). 'Promoter-proximal pausing mediated by the exon junction complex regulates splicing'. *Nature Communications*, **10**(521).

Andersen, C. B. F., Ballut, L., Johansen, J. S., Chamieh, H., Nielsen, K. H., Oliveira, C. L. P., Pedersen, J. S., Séraphin, B., le Hir, H., and Andersen, R. G. (2006). 'Structure of the Exon Junction Core Complex with a Trapped DEAD-Box ATPase Bound to RNA'. *Science*, **313**(5795), 1968-1972.

Babendure, J. R., Babendure, J. L., Ding, J. H., and Tsien, R. Y. (2006). 'Control of mammalian translation by mRNA structure near caps'. *RNA*, **12**(5), 851-861.

Barbosa, I., Haque, N., Fiorini, F., Barrandon, C., Tomasetto, C., Blanchette, M., and le Hir, H. (2012). 'Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon junction complex assembly'. *Nature Structural and Molecular Biology*, **19**(10), 983–991.

Bensaude, O., Barbosa, I., Morillo, L., Dikstein, R., & le Hir, H. (2024). 'Exon-junction complex association with stalled ribosomes and slow translation-independent disassembly'. *Nature Communications*, **15**(1).

Berget, S. M., Moore, C., and Sharp, P. A. (1977). 'Spliced segments at the 5' terminus of adenovirus 2 late mRNA* (adenovirus 2 mRNA processing/5' tails on mRNAs/electron microscopy of mRNA-DNA hybrids)'. *Biochemistry*, **74(**8), 3171-3175.

Berglund, J. A., Abovich, N., and Rosbash, M. (1998). 'A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition'. *Genes and development*. **12**(6), 858-867.

Biziaev, N., Shuvalov, A., Salman, A., Egorova, T., Shuvalova, E., and Alkalaeva, E. (2024). 'The impact of mRNA poly(A) tail length on eukaryotic translation stages'. *Nucleic Acids Research*, **52**(13), 7792–7808.

Black, D. L., Chabot, B., and Steitz, J. A. (1985). 'U2 as well as Ul Small Nuclear Ribonucleoproteins Are Involved in Premessenger RNA Splicing'. *Cell*, **42**(3), 737-750.

Bloor, S., Wit, N., and Lehner, P.J. (2025). ‘RNA binding by Periphilin plays an essential role in initiating silencing by the HUSH complex’. *Nucleic Acids Research,* **53**(2).

Boesler, C., Rigo, N., Anokhina, M. M., Tauchert, M. J., Agafonov, D. E., Kastner, B., Urlaub, H., Ficner, R., Will, C. L., and Lührmann, R. (2016). ‘A spliceosome intermediate with loosely associated tri- snRNP accumulates in the absence of Prp28 ATPase activity’. *Nature Communications*, **5**(7), 11997.

Bono, F., Ebert, J., Unterholzner, L., Güttler, T., Izaurralde, E., and Conti, E. (2004). ‘Molecular insights into the interaction of PYM with the Mago-Y14 core of the exon junction complex’. *EMBO Reports*, ***5***(3), 304–310.

Burke, K. A., Janke, A. M., Rhine, C. L., and Fawzi, N. L. (2015). ‘Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II’. *Molecular Cell*, **60**(2), 231–241.

Capon, D. J., Chamow, S. M., Mordentit, J., Marsters, S. A., Gregory, T., Mitsuyat, H., Byrn, R. A., Lucas, C., Wurm, F. M., Groopman, J. E., Brodert, S., and Smith, D. H. (1984). ‘Designing CD4 immunoadhesins for AIDS therapy’. *Nature,* **337**(6207), 525-531.

Chen, H., and Pugh, B. F. (2021). ‘What do Transcription Factors Interact With?’. *Journal of Molecular Biology*. **433**(14), 166883.

Chen, I.-H. B., Sciabica, K. S., and Sandri-Goldin, R. M. (2002). ‘ICP27 Interacts with the RNA Export Factor Aly/REF To Direct Herpes Simplex Virus Type 1 Intronless mRNAs to the TAP Export Pathway’. *Journal of Virology*, **76**(24), 12877–12889.

Chen, M., and Manley, J. L. (2009). ‘Mechanisms of alternative splicing regulation: Insights from molecular and genomics approaches’. *Nature Reviews Molecular Cell Biology,* **10**(11), 741-754.

Cheng, H., Dufu, K., Lee, C. S., Hsu, J. L., Dias, A., and Reed, R. (2006). ‘Human mRNA Export Machinery Recruited to the 5′ End of mRNA’. *Cell*, **127**(7), 1389–1400.

Cho, S., and Dreyfuss, G. (2010). ‘A degron created by SMN2 exon 7 skipping is a principal contributor to spinal muscular atrophy severity’. *Genes and Development*, **24**(5), 438–442.

Cordaux, R., and Batzer, M. A. (2009). ‘The impact of retrotransposons on human genome evolution’. *Nature Reviews Genetics*, **10**(10), 691-703.

Core, L., and Adelman, K. (2019). ‘Promoter-proximal pausing of RNA polymerase II: A nexus of gene regulation’. *Genes and Development*, **33**(15-16), 960-982.

Dahodwala, H., and Lee, K. H. (2019). ‘The fickle CHO: a review of the causes, implications, and potential alleviation of the CHO cell line instability problem’. *Current opinion in Biotechnology*, **60,** 128-137.

de Moura, T. R., Mozaffari-Jovin, S., Szabó, C. Z. K., Schmitzová, J., Dybkov, O., Cretu, C., Kachala, M., Svergun, D., Urlaub, H., Lührmann, R., and Pena, V. (2018). ‘Prp19/Pso4 Is an Autoinhibited Ubiquitin Ligase Activated by Stepwise Assembly of Three Splicing Factors’. *Molecular Cell*, **69**(6), 979-992.

de Oliveira Mann, C. C., and Hornung, V. (2021). ‘Molecular mechanisms of nonself nucleic acid recognition by the innate immune system’. *European Journal of Immunology*. **51**(8), 1897-1910.

del Gatto-konczak, F., Olive, M., Gesnel, M., and Breathnach, R. (1999). ‘hnRNP A1 Recruited to an Exon In Vivo Can Function as an Exon Splicing Silencer’ *Molecular and Cellular Biology*. **19**(1), 251-260.

Diem, M. D., Chan, C. C., Younis, I., and Dreyfuss, G. (2007). ‘PYM binds the cytoplasmic exon-junction complex and ribosomes to enhance translation of spliced mRNAs’. *Nature Structural and Molecular Biology*, **14**(12), 1173–1179.

Douse, C. H., Tchasovnikarova, I. A., Timms, R. T., Protasio, A. v., Seczynska, M., Prigozhin, D. M., Albecka, A., Wagstaff, J., Williamson, J. C., Freund, S. M. V., Lehner, P. J., and Modis, Y. (2020). ‘TASOR is a pseudo-PARP that directs HUSH complex assembly and epigenetic transposon control’. *Nature Communications*, **11**(1). 4940.

Dufu, K., Livingstone, M. J., Seebacher, J., Gygi, S. P., Wilson, S. A., and Reed, R. (2010). ‘ATP is required for interactions between UAP56 and two conserved mRNA export proteins, Aly and CIP29, to assemble the TREX complex’. *Genes and Development*, **24**(18), 2043–2053.

Duivelshof, B. L., Murisier, A., Camperi, J., Fekete, S., Beck, A., Guillarme, D., and D'Atri, V. (2021). 'Therapeutic Fc-fusion proteins: Current analytical strategies'. *Journal of Separation Science*, **44**(1), 35-62.

Eckmann, C. R., Rammelt, C., and Wahle, E. (2011). 'Control of poly(A) tail length'. *Wiley Interdisciplinary Reviews: RNA*, **2**(3), 348-361.

Erkelenz, S., Mueller, W. F., Evans, M. S., Busch, A., Schöneweis, K., Hertel, K. J., and Schaal, H. (2013). 'Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms'. *RNA*, **19**, 96-102.

Fan, J., Kuai, B., Wu, G., Wu, X., Chi, B., Wang, L., Wang, K., Shi, Z., Zhang, H., Chen, S., He, Z., Wang, S., Zhou, Z., Li, G., and Cheng, H. (2017). 'Exosome cofactor hMTR 4 competes with export adaptor ALYREF to ensure balanced nuclear RNA pools for degradation and export' . *The EMBO Journal*, **36**(19), 2870–2886.

Feng, Y., Chen, M., and Manley, J. L. (2008). 'Phosphorylation switches the general splicing repressor SRp38 to a sequence-specific activator'. *Nature Structural and Molecular Biology*, **15**(10), 1040–1048.

Ferrara, N., Gerber, H. N., and LeCouter, J. (2003). 'The biology of VEGF and its receptors'. *Nature Medicine*, **9**, 669-676.

Fica, S. M., Oubridge, C., Galej, W. P., Wilkinson, M. E., Bai, X. C., Newman, A. J. and Nagai, K. (2017). 'Structure of a spliceosome remodelled for exon ligation'. *Nature,* **542**, 377-380.

Fratta, P., and Isaacs, A. M. (2018). 'The snowball effect of RNA binding protein dysfunction in amyotrophic lateral sclerosis'. *Brain*, **141**(5), 1236-1246.

Fribourg, S., Braun, I. C., Izaurralde, E., and Conti, E. (2001). 'Structural Basis for the Recognition of a Nucleoporin FG Repeat by the NTF2-like Domain of the TAP/p15 mRNA Nuclear Export Factor'. *Molecular Cell*, **8**(3), 645-656.

Friend, K., Lovejoy, A. F., and Steitz, J. A. (2007). 'U2 snRNP Binds Intronless Histone Pre-mRNAs to Facilitate U7-snRNP-Dependent 3′ End Formation'. *Molecular Cell*, **28**(2), 240–252.

Furuichi, Y. (2015). 'Discovery of m7G-cap in eukaryotic mRNAs'. *Proceedings of the Japan Academy Series B: Physical and Biological Sciences.* **91**(8), 394-409.

Galej, W. P., Wilkinson, M. E., Fica, S. M., Oubridge, C., Newman, A. J., and Nagai, K. (2016). 'Cryo-EM structure of the spliceosome immediately after branching'. *Nature*, **537**(7619), 197–201.

Garcia, J., Hurwitz, H. I., Sandler, A. B., Miles, D., Coleman, R. L., Deurloo, R., and Chinot, O. L. (2020). 'Bevacizumab (Avastin®) in cancer treatment: A review of 15 years of clinical experience and future outlook'.*Cancer Treatment Reviews,* **86**, 102017.

Garland, W., Müller, I., Wu, M., Schmid, M., Imamura, K., Rib, L., Sandelin, A., Helin, K., and Jensen, T. H. (2022). 'Chromatin modifier HUSH co-operates with RNA decay factor NEXT to restrict transposable element expression'. *Molecular Cell*, **82**(9), 1691-1707.

Garneau, N. L., Wilusz, J., and Wilusz, C. J. (2007). 'The highways and byways of mRNA decay'. *Nature Reviews Molecular Cell Biology*, **8**(2), 113-126.

Gatfield, D., le Hir, H., Schmitt, C., Braun, I. C., Kö, T., Wilm, M., and Izaurralde, E. (2001). 'The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in Drosophila'. *Current Biology*, **11**(21), 1716-1721.

Gehring, N. H., Lamprinaki, S., Kulozik, A. E., and Hentze, M. W. (2009). 'Disassembly of Exon Junction Complexes by PYM'. *Cell*, **137**(3), 536–548.

Geuens, T., Bouhy, D., and Timmerman, V. (2016). 'The hnRNP family: insights into their role in health and disease'. *Human Genetics,* **135**(8), 851-867.

Ghafouri-Fard, S., Abak, A., Baniahmad, A., Hussen, B. M., Taheri, M., Jamali, E., and Dinger, M. E. (2022). 'Interaction between non-coding RNAs, mRNAs and G-quadruplexes'. *Cancer Cell International*, **22**(1), 171.

Golovanov, A. P., Hautbergue, G. M., Tintaru, A. M., Lian, L. Y., and Wilson, S. A. (2006). 'The solution structure of REF2-I reveals interdomain interactions and regions involved in binding mRNA export factors and RNA'. *RNA*, **12**(11), 1933–1948.

Gonatopoulos-Pournatzis, T., and Cowling, V. H. (2014). 'Cap-binding complex (CBC)'. *Biochemical Journal*, **457**(2), 231-242.

Greenberg, M. V. C., and Bourc'his, D. (2019). 'The diverse roles of DNA methylation in mammalian development and disease'. *Nature reviews. Molecular cell biology*, **20**(10), 590-607.

Gordon, M.S., Margolin, K., Talpaz, M., Sledge, G.W., Holmgren, E., Benjamin, R., Stalter, S., Shak, S., and Adelman, D. (2001). 'Phase 1 safety and pharmacokinetic study of recombinant human anti-vascular endothelial growth factor in patients with advanced cancer'. *Journal of clinical oncology*, **19**(3), 843-850.

Gustafsson, C., Govindarajan, S., and Minshull, J. (2004). 'Codon bias and heterologous protein expression'. *Trends in Biotechnology*, **22**(7), 346-353.

Hargous, Y., Hautbergue, G. M., Tintaru, A. M., Skrisovska, L., Golovanov, A. P., Stevenin, J., Lian, L. Y., Wilson, S. A., and Allain, F. H. T. (2006). 'Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8'. *EMBO Journal*, **25**(21), 5126–5137.

Harten, S. K., Bruxner, T. J., Bharti, V., Blewitt, M., Nguyen, T. M. T., Whitelaw, E., and Epp, T. (2014). 'The first mouse mutants of D14Abb1e (Fam208a) show that it is critical for early development'. *Mammalian Genome*, **25**(7–8), 293–303.

Hautbergue, G. M., Hung, M. L., Golovanov, A. P., Lian, L. Y. and Wilson, S. A. (2008). 'Mutually exclusive interactions drive handover of mRNA from export adaptors to TAP'. *Proceedings of the National Academy of Sciences of the United States of America*, **105**(13), 5154-5159.

Hautbergue, G. M., Hung, M. L., Walsh, M. J., Snijders, A. P. L., Chang, C. te, Jones, R., Ponting, C. P., Dickman, M. J., and Wilson, S. A. (2009). 'UIF, a New mRNA Export Adaptor that Works Together with REF/ALY, Requires FACT for Recruitment to mRNA'. *Current Biology*, **19**(22), 1918–1924.

Hay, N., and Sonenberg, N. (2004). 'Upstream and downstream of mTOR'. *Genes and Development,* **18**(16), 1926-1945.

Heath, C. G., Viphakone, N., and Wilson, S. A. (2016). 'The role of TREX in gene expression and disease'. *Biochemical Journal*, **473**(19), 2911-2935.

Hir, H. le, Saulière, J., and Wang, Z. (2016). 'The exon junction complex as a node of post-transcriptional networks'. *Nature Reviews Molecular Cell Biology*, **17**(1), 41-54.

Huang, Y., and Carmichael, G. G. (1997). ‘The mouse histone H2a gene contains a small element that facilitates cytoplasmic accumulation of intronless gene transcripts and of unspliced HIV-1-related mRNAs’. *Proceedings of the National Academy of Sciences of the United States of America*, **94**(19), 10104-10109.

Huang, Y., Gattoni, R., Sté, J., and Steitz, J. A. (2003). ‘Splicing Factors Serve as Adapter Proteins for TAP-Dependent mRNA Export’. *Molecular Cell,* **11**(3), 837-843.

Huang, Y., and Steitz, J. A. (2001). ‘Splicing Factors SRp20 and 9G8 Promote the Nucleocytoplasmic Export of mRNA’. *Molecular cell,* **7**(4), 899-905.

Hwang, H.J, Park, Y., and Kim, Y.K. (2021). ‘UPF1: From mRNA surveillance to Protein Quality Control’. *Biomedicines*, **9**(8), 995.

Irimia, M., and Roy, S. W. (2014). ‘Origin of spliceosomal Introns and Alternative Splicing’. *Cold Spring Harbor Perspectives in Biology,* **6**(6).

Ishigaki, S., Masuda, A., Fujioka, Y., Iguchi, Y., Katsuno, M., Shibata, A., Urano, F., Sobue, G., and Ohno, K. (2012). ‘Position-dependent FUS-RNA interactions regulate alternative splicing events and transcriptions’. *Scientific Reports*, **2**, 529.

Izaurralde, E., Lewis, J., Mcguigan, C., Jankowska, M., Darzynkiewicz, E., and Mattaj, W. (1994). ‘A Nuclear Cap Binding Protein Complex Involved in Pre-mRNA Splicing’. *Cell*, **78**(4), 657-668.

Jones, P. T., Dear, P. H., Foote, P. H., Neuberger, M. S., and Winter, G.(1986). ‘Replacing the complementarity-determining regions in a human antibody with those from a mouse’. *Nature*, **321**, 522-525.

Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G., and Ohno, S. (2006). ‘Binding of a novel SMG-1-upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay’. *Genes and development,* **20**(3), 355-367.

Katahira, J., Strä, K., ßer, S., Podtelejnikov, A., Mann, M., Jung, J. U., and Hurt, E. (1999). ‘The Mex67p-mediated nuclear mRNA export pathway is conserved from yeast to human’. *The EMBO Journal*, **18**(9), 2593-2609.

Kelley, B. (2024). ‘The history and potential future of monoclonal antibody therapeutics development and manufacturing in four eras’. *MABS,* **16**(1), 2373330.

Khan, M., Hou, S., Azam, S., and Lei, H. (2021). ‘Sequence-dependent recruitment of SRSF1 and SRSF7 to intronless lncRNA NKILA promotes nuclear export via the TREX/TAP pathway’. *Nucleic Acids Research*, **49**(11), 6420–6436.

Kim, J. Y., Kim, Y. G., and Lee, G. M. (2012). ‘CHO cells in biotechnology for production of recombinant proteins: Current state and further potential’. *Applied Microbiology and Biotechnology,* **93**(3), 917-930.

Kim, Y., Bjijrklund, S., Li, Y., Sayre, M. H., and Kornberg, R. D. (1994). ‘A Multiprotein Mediator of Transcriptional Activation and Its Interaction with the C-Terminal Repeat Domain of RNA Polymerase II’. *Cell*, **77**(4), 599-608.

Köhler, G., and Milstein, C. (1975). ‘Continous cultures of fused cells secreting antibody of predefined specificity’. *Nature,* **356**(5517), 495-497.

Kolakada, D., Campbell, A.E., Galvis, L.B., Li, Z., Lore, M., and Jagannathan, S. (2024). ‘A system of reporters for comparative investigation of EJC-independent and EJC-enhanced nonsense-mediated mRNA decay’. *Nucleic Acids Research,* **52**(6).

Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D. J., Luscombe, N. M., and Ule, J. (2011). ‘ICLIP - transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution’. *Journal of Visualized Experiments*, **50**, 2638.

Kotlajich, M. v., Crabb, T. L., and Hertel, K. J. (2009). ‘Spliceosome Assembly Pathways for Different Types of Alternative Splicing Converge during Commitment to Splice Site Pairing in the A Complex’. *Molecular and Cellular Biology*, **29**(4), 1072–1082.

Kühn, U., Gündel, M., Knoth, A., Kerwitz, Y., Rüdel, S., and Wahle, E. (2009). ‘Poly(A) tail length is controlled by the nuclear Poly(A)-binding protein regulating the interaction between Poly(A) polymerase and the cleavage and polyadenylation specificity factor’. *Journal of Biological Chemistry*, **284**(34), 22803–22814.

Lai, T., Yang, Y., and Ng, S. K. (2013). ‘Advances in mammalian cell line development technologies for recombinant protein production’. *Pharmaceuticals*, **6**(5), 579-603.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A.,

Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L.,

Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J., Szustakowki, J.; International Human Genome Sequencing Consortium. (2001). 'Initial sequencing and analysis of the human genome', *Nature,* **409**(6822), 860-921.

Lei, H., Dias, A. P., and Reed, R. (2011). 'Export and stability of naturally intronless mRNAs require specific coding region sequences and the TREX mRNA export complex'. *Proceedings of the National Academy of Sciences of the United States of America,* **108**(44), 17985-17990.

Lei, H., Zhai, B., Yin, S., Gygi, S., and Reed, R. (2013). 'Evidence that a consensus element found in naturally intronless mRNAs promotes mRNA export'. *Nucleic Acids Research*, **41**(4), 2517–2525.

Le Hir, H., Izaurralde, E., Maquat, L. E., and Moore, M. J. (2000).'The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions'. *The EMBO Journal,* **19**(24), 6860- 6869.

Le Hir, H., Saulière, J., and Wang, Z. (2016). 'The exon junction complex as a node of post-transcriptional networks'. *Nature reviews. Molecular cell biology*, **17**(1), 45-54.

Liang, W. W., and Cheng, S. C. (2015). 'A novel mechanism for Prp5 function in prespliceosome formation and proofreading the branch site sequence'. *Genes and Development*, **29**(1), 81–93.

Lingaraju, M., Johnsen, D., Schlundt, A., Langer, L. M., Basquin, J., Sattler, M., Heick Jensen, T., Falk, S., and Conti, E. (2019). 'The MTR4 helicase recruits nuclear adaptors of the human RNA exosome using distinct arch-interacting motifs'. *Nature Communications*, **10**(1), 3393.

Liu, B., Sun, L., Liu, Q., Gong, C., Yao, Y., Lv, X., Lin, L., Yao, H., Su, F., Li, D., Zeng, M., and Song, E. (2015). 'A Cytoplasmic NF-κB Interacting Long Noncoding RNA Blocks IκB Phosphorylation and Suppresses Breast Cancer Metastasis'. *Cancer Cell*, **27**(3), 370–381.

Lorson, C. L., Hahnen, E., Androphy, E. J., and Wirth, B. (1999). 'A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy'.

Proceedings of the National Academy of Sciences of the United States of America, **96**(11), 6307-63011.

Lonberg, N., Taylor, L. D., Harding, F. A., Trounstine, M., Higgins, K. M., Schramm, S. R., Kuo, C. C., Mashayekh, R., Wymore, K., McCabe, J. G., Munoz-O'Regan, D., O'Donell, S. L., Lapachet, E. S. G., Bengoechea, T., Fishwild, D. M., Carmack, C. E., Kay, R. M. and Huszar, D. (1994). 'Antigen-specific human antibodies from mice comprising four distinct genetic modifications'. *Nature*. **368**(6474), 856-859.

Loughlin, F. E., Lukavsky, P. J., Kazeeva, T., Reber, S., Hock, E. M., Colombo, M., von Schroetter, C., Pauli, P., Cléry, A., Mühlemann, O., Polymenidou, M., Ruepp, M. D., and Allain, F. H. T. (2019). 'The Solution Structure of FUS Bound to RNA Reveals a Bipartite Mode of RNA Recognition with Both Sequence and Shape Specificity'. *Molecular Cell*, **73**(3), 490-504.

Lu, R. M., Hwang, Y. C., Liu, I. J., Lee, C. C., Tsai, H. Z., Li, H. J., and Wu, H. C. (2020). 'Development of therapeutic antibodies for the treatment of diseases'. *Journal of Biomedical Science* **27**(1)

Lyu, X., Zhao, Q., Hui, J., Wang, T., Lin, M., Wang, K., Zhang, J., Shentu, J., Dalby, P. A., Zhang, H., and Liu, B. (2022). 'The global landscape of approved antibody therapies'. *Antibody Therapeutics,* **5**(4), 233-257.

Ma, X. M., Yoon, S. O., Richardson, C. J., Jülich, K., and Blenis, J. (2008). 'SKAR Links Pre-mRNA Splicing to mTOR/S6K1-Mediated Enhanced Translation Efficiency of Spliced mRNAs'. *Cell*, **133**(2), 303–313.

Maini, R. N., Elliott, M. J., Brennan, F. M., and Feldmann, M. (1995). 'Beneficial effects of tumour necrosis factor-alpha (TNF-α) blockade in rheumatoid arthritis (RA)'. *Clinical and Experimental Immunology*, **101**(2), 207-212.

Maréchal, A., Li, J. M., Ji, X. Y., Wu, C. S., Yazinski, S. A., Nguyen, H. D., Liu, S., Jiménez, A. E., Jin, J., and Zou, L. (2014). 'PRP19 Transforms into a Sensor of RPA-ssDNA after DNA Damage and Drives ATR Activation via a Ubiquitin-Mediated Circuitry'. *Molecular Cell*, **53**(2), 235–246.

Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N., and Reed, R. (2005). ‘Recruitment of the human TREX complex to mRNA during splicing’. *Genes and Development*, **19**(13), 1512–1517.

Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N., and Reed, R. (2005). ‘Recruitment of the human TREX complex to mRNA during splicing’. *Genes and Development*, **19**(13), 1512–1517.

Michlewski, G., Sanford, J. R., and Cáceres, J. F. (2008). ‘The Splicing Factor SF2/ASF Regulates Translation Intiiation by Enhancing Phosphorylation of 4E-BP1’. *Molecular Cell*, **30**(2), 179-189.

Mitoma, H., Horiuchi, T., Tsukamoto, H., and Ueda, N. (2018). ‘Molecular mechanisms of action of anti-TNF-α agents – Comparison among therapeutic TNF-α antagonists’. *Cytokine*, **101**, 56–63.

Mordstein, C., Savisaar, R., Young, R. S., Bazile, J., Talmane, L., Luft, J., Liss, M., Taylor, M. S., Hurst, L. D., and Kudla, G. (2020). ‘Codon Usage and Splicing Jointly Influence mRNA Localization’. *Cell Systems*, **10**(4), 351-362.

Morrison, S. L., Johnsont, M. J., Herzenbergt, L. A., and Oi, V. T. (1984). ‘Chimeric human antibody molecules: Mouse antigen-binding domains with human constant region domains’. *Proceedings of the National Academy of Sciences of the United States of America,* **81**, 6851-6855.

Moteki, S., and Price, D. (2002). ‘Functional Coupling of Capping and Transcription of mRNA’. *Molecular Cell,* **10**, 599-609.

Müller, I., and Helin, K. (2024). ‘Keep quiet: the HUSH complex in transcriptional silencing and disease’. *Nature Structural and Molecular Biology*, **31**(1), 11-22.

Nagy, P. D., and Bujarski, J. J. (1998). ‘Silencing Homologous RNA Recombination Hot Spots with GC-Rich Sequences in Brome Mosaic Virus’. *Journal of Virology*, **72**(2), 1122-1130.

Nott, A., le Hir, H., and Moore, M. J. (2004). ‘Splicing enhances translation in mammalian cells: An additional function of the exon junction complex’. *Genes and Development*, **18**(2), 210–222.

Pabis, M., Neufeld, N., Steiner, M. C., Bojic, T., Shav-Tal, Y., and Neugebauer, K. M. (2013). 'The nuclear cap-binding complex interacts with the U4/U6·U5 tri-snRNP and promotes spliceosome assembly in mammalian cells'. *RNA*, **19**(8), 1054–1063.

Palmiter, R. D., Sandgrent, E. P., Avarbockt, M. R., Diane Allen, D., and Brinstert, R. L. (1991). 'Heterologous introns can enhance expression of transgenes in mice'. *Proceedings of the National Academy of Sciences of the United States of America,* **88**, 478-482.

Passmore, L. A., and Coller, J. (2022). 'Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression'. *Nature Reviews Molecular Cell Biology*, **23**(2), 93-106.

Patel, A. A., and Steitz, J. A. (2003). 'Splicing double: insights from the second spliceosome'. *Nature Reviews. Molecular Cell Biology*, **4**(12), 960-970.

Philip Benoit Bouvrette, L., Cody, N. A. L., Bergalet, J., Lefebvre, F. A., Diot, C., Wang, X., Blanchette, M., and Lécuyer, E. (2018). 'CeFra-seq reveals broad asymmetric mRNA and noncoding RNA distribution profiles in Drosophila and human cells'. *RNA,* **24**, 98-113.

Prigozhin, D. M., Douse, C. H., Farleigh, L. E., Albecka, A., Tchasovnikarova, I. A., Timms, R. T., Oda, S. I., Adolf, F., Freund, S. M. V., Maslen, S., Lehner, P. J., and Modis, Y. (2020). 'Periphilin self-association underpins epigenetic silencing by the HUSH complex'. *Nucleic Acids Research*, **48**(18), 10313–10328.

Puck, E. T., Cieciura, S. J., and Robinson, R. (1958). 'GENETICS OF SOMATIC MAMMALIAN CELLS'. *The Journal of experimental medicine*, **108**(6), 945–956

Ramanathan, A., Robb, G. B., and Chan, S. H. (2016). 'mRNA capping: Biological functions and applications'. *Nucleic Acids Research,* **44**(16), 7511-7526.

Reed, R., and Hurt, E. (2002). 'A Conserved mRNA Export Machniery Coupled to pre-mRNA splicing'. *Cell*, **108**, 523-531.

Reinhart, D., Damjanovic, L., Kaisermayer, C., Sommeregger, W., Gili, A., Gasselhuber, B., Castan, A., Mayrhofer, P., Grünwald-Gruber, C., and Kunert, R. (2019). 'Bioprocessing of Recombinant CHO-K1, CHO-DG44, and CHO-S: CHO Expression Hosts Favor Either mAb Production or Biomass Synthesis'. *Biotechnology Journal*, **14**(3). 1700686.

Romig, H., Fackelmayer, F. O., Renz, A., Ramsperger, U., and Richter, A. (1992). 'Characterization of SAF-A, a novel nuclear DNA binding protein from HeLa cells with high affinity for nuclear matrix/scaffold attachment DNA elements'. *The EMBO Journal*, **11**, 3431-3440.

Sanford, J. R., Gray, N. K., Beckmann, K., and Cáceres, J. F. (2004). 'A novel role for shuttling SR proteins in mRNA translation'. *Genes and Development*, **18**(7), 755–768.

Saulière, J., Murigneux, V., Wang, Z., Marquenet, E., Barbosa, I., le Tonquèze, O., Audic, Y., Paillard, L., Crollius, H. R., and le Hir, H. (2012). 'CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex'. *Nature Structural and Molecular Biology*, **19**(11), 1124–1131.

Schaal, T. D., and Maniatis, T. (1999). 'Multiple Distinct Splicing Enhancers in the Protein-Coding Sequences of a Constitutively Spliced Pre-mRNA'. *Molecular and Cellular Biology*, **19**(1), 261-273.

Schellenberg, J., Nagraik, T., Wohlenberg, O. J., Ruhl, S., Bahnemann, J., Scheper, T., and Solle, D. (2022). 'Stress-induced increase of monoclonal antibody production in CHO cells'. *Engineering in Life Sciences*, **22**(5), 427–436.

Schimke, R. T. (1984). 'Gene Amplification in Cultured Animal Cells Review'. *Cell,* **37**, 705-713.

Schlautmann, L. P., and Gehring, N. H. (2020). 'A day in the life of the Exon junction complex'. *Biomolecules*, **10**(6), 1-17.

Schroeder, H. W., and Cavacini, L. (2010). 'Structure and function of immunoglobulins'. *Journal of Allergy and Clinical Immunology*, **125**(202), 41-52.

Schwartz, J. C., Ebmeier, C. C., Podell, E. R., Heimiller, J., Taatjes, D. J., and Cech, T. R. (2012). 'FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2'. *Genes and Development*, **26**(24), 2690–2695.

Seczynska, M., Bloor, S., Cuesta, S. M., and Lehner, P. J. (2022). 'Genome surveillance by HUSH-mediated silencing of intronless mobile elements'. *Nature*, **601**(7893), 440–445.

Seczynska, M., and Lehner, P. J. (2023). 'The sound of silence: mechanisms and implications of HUSH complex function'. *Trends in Genetics,* **39**(4), 251-267.

Shabalina, S. A., Ogurtsov, A. Y., Spiridonov, A. N., Novichkov, P. S., Spiridonov, N. A., & Koonin, E. v. (2010). 'Distinct patterns of expression and evolution of intronless and intron-containing mammalian genes'. *Molecular Biology and Evolution*, **27**(8), 1745–1749.

Shaul, O. (2017). How introns enhance gene expression. *International Journal of Biochemistry and Cell Biology*, **91**, 145-155.

Shepard, P. J., and Hertel, K. J. (2009). 'The SR protein family'. *Genome biology*, **10**(10). 242.

Shorrock, H.K., Gillingwater, T.H., and Groen, E.J.N. (2018). Overview of Current Drugs and Molecules in Development for Spinal Muscular Atrophy Therapy, *Drugs*, **78**, 293-305.

Singh, G., Kucukural, A., Cenik, C., Leszyk, J. D., Shaffer, S. A., Weng, Z., and Moore, M. J. (2012). 'The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus'. *Cell*, **151**(4), 750–764.

Singh, N. N., and Singh, R. N. (2011). 'Alternative splicing in spinal muscular atrophy underscores the role of an intron definition model'. *RNA Biology,* **8**(4), 600-606.

Spencley, A. L., Bar, S., Swigut, T., Flynn, R. A., Lee, C. H., Chen, L. F., Bassik, M. C., and Wysocka, J. (2023). 'Co-transcriptional genome surveillance by HUSH is coupled to termination machinery'. *Molecular Cell*, **83**(10), 1623-1639.

Takahama, K., Takada, A., Tada, S., Shimizu, M., Sayama, K., Kurokawa, R., and Oyoshi, T. (2013). 'Regulation of telomere length by G-quadruplex telomere DNA- and TERRA-binding protein TLS/FUS'. *Chemistry and Biology*, **20**(3), 341–350.

Tange, T., Shibuya, T., Jurica, M. S., and Moore, M. J. (2005). 'Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core'. *RNA*, **11**(12), 1869–1883.

Tarun, S. Z., and Sachs, A. B. (1996). 'Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G'. *The EMBO Journal*, **15**(24), 7168-7177.

Tavanez, J. P., Madl, T., Kooshapur, H., Sattler, M., and Valcárcel, J. (2012). 'HnRNP A1 Proofreads 3' Splice Site Recognition by U2AF'. *Molecular Cell*, **45**(3), 314–329.

Tchasovnikarova, I. A., Timms, R. T., Matheson, N. J., Wals, K., Antrobus, R., Göttgens, B., Dougan, G., Dawson, M. A., and Lehner, P. J. (2015). 'Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells'. *Science*, **348**(6242), 1481–1485.

Tholen, J., and Galej, W. P. (2022). 'Structural studies of the spliceosome: Bridging the gaps'. *Current Opinion in Structural Biology*, **77**, 102461.

Thompson, V. F., Wieland, D. R., Mendoza-Leon, V., Janis, H. I., Lay, M. A., Harrell, L. M., and Schwartz, J. C. (2023). 'Binding of the nuclear ribonucleoprotein family member FUS to RNA prevents R-loop RNA:DNA hybrid structures'. *Journal of Biological Chemistry*, **299**(10). 105237.

Tihanyi, B., & Nyitray, L. (2020). Recent advances in CHO cell line development for recombinant protein production. *Drug Discovery Today: Technologies,* **38**, 25-34.

Tischer, E., Mitchell, R., Hartman, T., Silva, M., Gospodarowicz, D., Fiddes, J.C., and Abraham, J.A. (1991). 'The human gene for vascular endothelial growth factor. Multiple proteins forms are encoded though alternative exon splicing'.*Journal of biological chemistry*, **266**(18), 11947- 11954.

Todd, P. A., and Brogden, R. N. (1989). 'Muromonab CD3, A review of its pharmacology and Therapeutic Potential'. *Durg Evaluation*, **37**, 871-899.

Tunbak, H., Enriquez-Gasca, R., Tie, C. H. C., Gould, P. A., Mlcochova, P., Gupta, R. K., Fernandes, L., Holt, J., van der Veen, A. G., Giampazolias, E., Burns, K. H., Maillard, P. v., and Rowe, H. M. (2020). 'The HUSH complex is a gatekeeper of type I interferon through epigenetic regulation of LINE-1s'. *Nature Communications*, **11**(1), 5387.

Valencia, P., Dias, A. P., and Reed, R. (2008). 'Splicing promotes rapid and efficient mRNA export in mammalian cells'. *Proceedings of the National Academy of Sciences of the United States of America,* **105**(9), 3386-3391.

Viphakone, N., Hautbergue, G. M., Walsh, M., Chang, C. te, Holland, A., Folco, E. G., Reed, R., and Wilson, S. A. (2012). 'TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export'. *Nature Communications*, **3**, 1006.

Viphakone, N., Sudbery, I., Griffith, L., Heath, C. G., Sims, D., and Wilson, S. A. (2019). 'Co-transcriptional Loading of RNA Export Factors Shapes the Human Transcriptome'. *Molecular Cell*, **75**(2), 310-323.

Wang, K., Yin, C., Du, X., Chen, S., Wang, J., Zhang, L., Wang, L., Yu, Y., Chi, B., Shi, M., Wang, C., Reed, R., Zhou, Y., Huang, J., and Cheng, H. (2019). 'A U2-snRNP–independent role of SF3b in promoting mRNA export'. *Proceedings of the National Academy of Sciences of the United States of America*, **116**(16), 7837–7846.

Wang, Y., Liu, J., Huang, B., Xu, Y.-M., Li, J., Huang, L.-F., Lin, J., Zhang, J., Min, Q.-H., Yang, W.-M., and Wang, X.-Z. (2015). 'Mechanism of alternative splicing and its regulation'. *Biomedical Reports*, **3**(2), 152–158.

Wang, Z., Murigneux, V., and le Hir, H. (2014). 'Transcriptome-wide modulation of splicing by the exon junction complex'. *Genome Biology*, **15**(12), 551.

Wilkinson, M. E., Charenton, C., and Nagai, K. (2020). 'RNA Splicing by the Spliceosome'. *Annual Review of Biochemistry,* **89**, 359-388.

Woodward, L. A., Mabin, J. W., Gangras, P., and Singh, G. (2017). 'The exon junction complex: a lifelong guardian of mRNA fate'. *Wiley Interdisciplinary Reviews: RNA,* **8**(3).

Xu, W. J., Lin, Y., Mi, C. L., Pang, J. Y., and Wang, T. Y. (2023). 'Progress in fed-batch culture for recombinant protein production in CHO cells'. *Applied Microbiology and Biotechnology*, **107**(4), 1063-1075.

Yan, C., Wan, R., and Shi, Y. (2019). 'Molecular mechanisms of pre-mRNA splicing through structural biology of the spliceosome'. *Cold Spring Harbor Perspectives in Biology*, **11**(1).

Yoon, D.-W., and Lee, H. (1997). 'Tap: A Novel Cellular Protein That Interacts with Tip of Herpesvirus Saimiri and Induces Lymphocyte Aggregation'. *Immunity*, **6**, 571-582.

Yu, Y., Maroney, P. A., Denker, J. A., Zhang, X. H. F., Dybkov, O., Lührmann, R., Jankowsky, E., Chasin, L. A., and Nilsen, T. W. (2008). 'Dynamic Regulation of Alternative Splicing by Silencers that Modulate 5′ Splice Site Competition'. *Cell*, **135**(7), 1224–1236.

Zeh, N., Schmidt, M., Schulz., P., and Fischer, S. (2024). ‘The new frontier in CHO cell line development: From random to targeted transgene integration technologies’. *Biotechnology Advances*, **75**, 108402.

Zhan, X., Yan, C., Zhang, X., Lei, J., & Shi, Y. (2018). ‘Structure of a human catalytic step I spliceosome’. *Science*, **359**, 537-545.

Zhang, L., Song, D., Zhu, B., and Wang, X. (2019). ‘The role of nuclear matrix protein HNRNPU in maintaining the architecture of 3D genome’. *Seminars in Cell and Developmental Biology*, **90**, 161-167.

Zhao, S., Mysler, E., and Moots, R. J. (2018). ‘Etanercept for the treatment of rheumatoid arthritis’. *Immunotherapy*, **10**(6), 433–445.

Zheng, K., Yarmarkovich, M., Bantog, C., Bayer, R., and Patapoff, T. W. (2014). ‘Influence of glycosylation pattern on the molecular properties of monoclonal antibodies’. *mAbs*, **6**(3), 649–658.

Zhu, J. (2012).’ Mammalian cell protein expression for biopharmaceutical production’. *Biotechnology Advances*, **30**(5), 1158–1170.

Zuckerman, B., Ron, M., Mikl, M., Segal, E., and Ulitsky, I. (2020). ‘Gene Architecture and Sequence Composition Underpin Selective Dependency of Nuclear Export of Long RNAs on NXF1 and the TREX Complex’. *Molecular Cell*, **79**(2), 251-267.