Functional Characterization of Naturally Occurring Mutants of the Human Guanine-rich RNA Sequence Binding Factor 1 (GRSF1) and Mechanistic Studies into the Molecular Basis of GRSF1-RNA Interaction.

Zur Erlangung des akademischen Grades des
Doktors der Naturwissenschaften (Dr. rer. nat)

eingereicht im Fachbereich Biologie, Chemie, Pharmazie
der Freien Universität Berlin

vorgelegt von
Sajad Ahmad Sofi
Biochemiker aus Budgam, Indien

Berlin 2017

Diese Arbeit wurde zwischen dem 01.01.2014 und 31.03.2017 am Institut für Biochemie der Medizinischen Fakultät der Charité – Universitätsmedizin Berlin unter der Leitung von Prof. Dr. sc. med. Hartmut Kühn angefertigt.

1. Gutachter: Prof. Dr. Hartmut Kühn
2. Gutachter: Prof. Dr. Rudolf Tauber

Disputation am 11. 09. 2017

ACKNOWLEDGEMENTS

This is by far the most exciting and hardest research project that I have ever completed. I would like to thank many unique people who helped me to accomplish this task.

First, I would like to thank my advisers, Prof. Dr. Hartmut Kühn and Prof. Dr. Rudolf Tauber for their supervision and support. My special thanks goes to Prof. Hartmut Kühn for giving me the opportunity to join his amazing group and for always being open for discussions and new ideas.

Massive thanks to Dr. Christoph Ufer ‘‘my academic dad’’ for being an inspiring role model as a person and as a scientist. Thank you Christoph for teaching me the techniques that were essential for GRSF1 work and inspiring me to be an independent scientist. It was a great privilege to be part of our GRSF1 team and thank you, for your insightful discussions and for providing endless intellectual and emotional support.

I would also like to thank Dr. Dagmar Heydeck for helping me to use and navigate different protein and genomic databases. Thank you to Dr. Astrid Borchert for her scientific advice and being so generous to me.

Many thanks to Ms. Sabine Stehling for her friendly conversations and for providing technical assistance related to my work.

I would also like to thank my colleagues especially Swathi Banthiya, Désirée Jähn, Bernhard Dumoulin, Sophia Roth, Simone Brütsch, Marlene Rademacher, Felix Karst and Dr. Eugenia Marbach for lunch breaks and making my stay in lab fun.

I thank my dear collaborators: Dr. Margitta Dathe for CD spectroscopy and Prof. Dr. Pallu Reddanna for 3D-structural protein modelling. I thank GRK 1673 for giving me the opportunity to take part in this graduate program. Thank you to Dr. Elisabeth Otto for smoothly coordinating this program and for being a good friend. My thanks goes to DFG and Sonnenfeld foundation for their financial support.

Finally I come to my family and closest friends. Special thanks to my Mom and Papa for your love and believing in me. The greatest thanks goes to my Mom for being a wonderful companion in my life and teaching me: to be dedicated, to respect people, gardening skills, and most importantly giving me a pet duck as a child. To my brother Ummer, for always being there whenever I needed him and to my sister Renu, she always inspires me how she accomplishes her dreams in spite of all challenges.

Andrea: I am so thankful and proud to have you as a friend in my life. You have been an emotional backbone for me throughout many ups and downs in my life in Berlin. The fact that you have lived through these experiences was a massive help to me. You have such a huge insight into human nature, you always had a clear understanding of my situation. I am going to continue our ‘Stammtisch’ regardless where I am. Also, thanks for

recommending me to drink ‘Grappa’ I mean which crazy person invented this?, it tastes like concentrated hydrochloric acid but less angry (please don’t ask me how do I know that).

Thank you to Hilal, for frequently visiting me in Berlin and cheering me up. You have such a great sense of humor. Many thanks to Shabir Hassan for always being there for me and hosting me in Zürich. I literally miss that ‘Salty tea’ in the breakfast. Thank you to Imran, for staying in touch.

Emma: You are the only woman that I am jealous of because you were already married☺. I absolutely cherish you as a friend in my life, I now have a life-long friend to give me professional help on what to try in different cities. Also, you remember your ‘book club meet ups’ I am definitely going to do this with my kids. Claire: You have been such a special friend in my life. Thank you for your emotional support and generosity. You are so kind, I love you so much. I should thank Mayur for supporting me emotionally and being there for me during tough times.

I should also thank ‘GRSF1 Club’ that was started by Désirée, Christoph and me in order to further understand the biology of GRSF1. Christoph: Vielen Dank für Kuchen und Kaffee.

Lastly, to ‘RNA’ for sparking my passion for science. It always gives me goose bumps thinking about how flexible and adaptable RNA is for almost any biological function.

TABLE OF CONTENTS

Abbreviations

ATP	Adenosine-5'-triphosphate
AD	Acidic Domain
AP	Alkaline Phosphatase Antibody
BA	Boric Acid
bp	base pair
CD	Circular Dichroism
cDNA	Complementary Deoxyribonucleic Acid
DNA	Deoxyribonucleic Acid
dNTPs	Deoxyribonucleotide triphosphates
DIG	Digoxygenin
DNase	Deoxyribonuclease
DTT	Dithiothreitol
Daz1	Deleted in azoospermia-like Protein 1
EDTA	Ethylenediamine –N,N,N,N'-tetraacetic acid
eIF	Eukaryotic Initiation Factor
F	Phenylalanine
G4	G-quadruplex
GRSF1	Guanine-Rich Sequence Binding Factor 1
GST	Glutathione-S-Transferase
GPx4	Glutathione Peroxidase 4
GRDs	Glycine-rich Domains
hnRNP	Heterogeneous Nuclear Ribonucleoprotein
HEPES	N-2-Hydroxyethylpiperazine-N-2-ethane sulphonic acid
IRES	Internal Ribosome Entry Site
IB	Immunoblot
Inr	Initiator-element
KH	K-Homology Domain
kDa	Kilodalton
LB-Medium	Luria-Bertani Medium
mRNA	Messenger RNA
MRGs	Mitochondrial RNA Granules
NP	Nucleoprotein
NMR	Nuclear Magnetic Resonance
nt	Nucleotide

ncRNA	Non-coding RNA
Oligo	Oligonucleotide
PCR	Polymerase Chain Reaction
Pre-mRNA	Precursor mRNA
qRRM	Quasi RNA Recognition Motif
QGRS	Quadruplex Forming G-Rich Seqeunces
RNA	Ribonucleic Acid
RBD	RNA-binding Domain
RBP	RNA-binding Protein
RNP	Ribonucleoprotein
RRM	RNA Recognition Motif
RNase	Ribonuclease
REMSA	RNA Electrophoretic Mobility Shift Assays
rpm	Rotations Per Minute
snRNA	Small Nuclear RNA
ssRNA	Single-Stranded RNA
SDS	Sodium Dodecyl Sulphate
snRNP	Small Nuclear Ribonucleoprotein
SNPs	Single Nucleotide Polymorphisms
tRNA	Transfer RNA
TBP	TATA-box Binding Protein
UTR	Untranslated Region
UTP	Uracil-5’-triphosphate
UV	Ultraviolet
V	Volt
W	Watt
WT	Wild-type

1. INTRODUCTION

1.1. Overview of gene expression and role of RNA modification

Differential gene expression enables the multi-cellular organisms to orchestrate development and to respond to environmental signals by precise regulation of the activity state of their genes (Ufer, 2012). Gene expression comprises the controlled conversion of genomic information into proteins (Crick FH, 1958). Gene expression is a highly intricate and multi-layered process that is tightly regulated on three different levels: transcriptional events, post-transcriptional events, and post-translational events (see **Figure 1.1**) (Jacob & Monod, 1960).

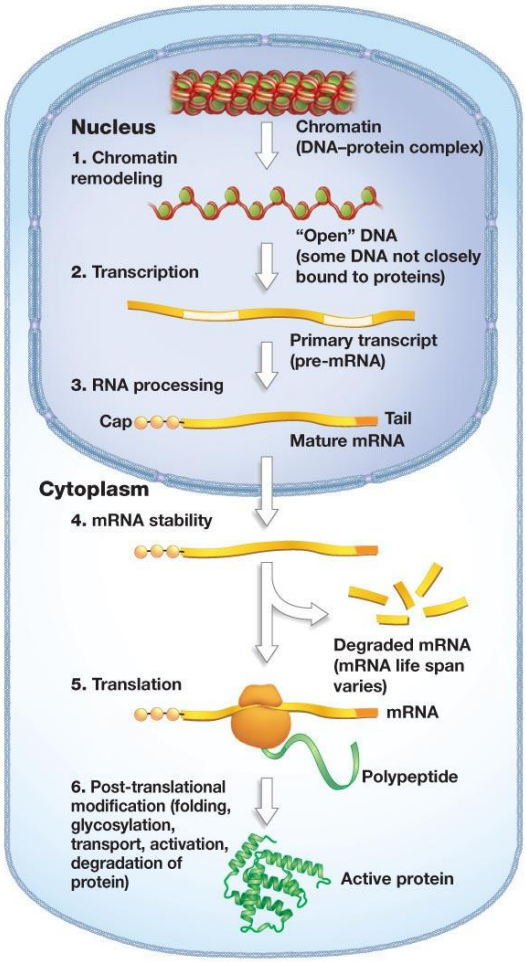

Figure 1.1. Layers of gene expression regulation in eukaryotes.
(Source: http://bio100.class.uic.edu/lectures/genetic_control.htm)

The first level comprises the mechanisms by which pre-mRNA transcripts are formed from genomic DNA. The second level of regulation (post-transcriptional control)

involves mechanisms that regulate the processing of primary transcripts before being translated. The final level comprises the mechanisms by which proteins are modified into their mature, functional forms (Struhl K, *et al*., 1999). The key player of post-transcriptional regulation of gene expression is the RNA. RNA is a highly versatile molecule encompassing both informational and catalytic functions in the cell (Gerstberger S, *et al*., 2014). RNA is never naked in the cell but assembles with proteins forming dynamic complexes termed ribonucleoproteins (RNPs) (Dreyfuss G, *et al*., 1988; Beckmann BM, *et al*., 2016). The RNPs regulate stability of the RNA, its translational activity and its subcellular localization establishing a separate level of gene expression regulation (Jansen RP, *et al*., 2014). Gene expression is precisely controlled at each of its three levels by many different factors, which include coding RNAs, non-coding RNAs as well as RNA-binding proteins (RBPs). These factors play a pivotal role during post-transcriptional processing. They include small nuclear RNAs (snRNAs) involved in splicing regulation, small nucleolar RNAs (snoRNAs) involved in ribosome biogenesis, as well as the ribosomal RNAs and transfer RNAs involved in translation (Guttman M, *et al*., 2012).

The human genome encodes about 1,500 RNA-binding proteins (RBPs) that control RNA metabolism from its synthesis to its final decay (Gerstberger S, *et al*., 2014;). First, they expand its repertoire of RBPs by producing new protein isoforms from given precursor mRNAs. Next, they regulate stability and translational efficiency of mRNA transcripts in the cytoplasm. Mutations in the genes that code for RBPs are frequently linked to diseases (Rinn JL, *et al*., 2014; Lenzken SC, *et al*., 2014). Despite growing insights into the mechanisms of RBP activity the biological role of most of them remains to be explored.

1.1.1. Gene expression regulation

The regulation of transcription involves the combined effects of both the chromatin structure and the interaction of regulatory proteins called transcription factors with DNA (Struhl K, *et al*., 1999). In eukaryotic cells the DNA is packaged into nucleosomes by proteins called histones. The nucleosomes are the basic units for the assembly of chromatin. The chromatin is a complex formed by DNA and histones (Richmond TJ, *et al*., 2003), in which the negatively charged DNA is twisted around the positively charged histone proteins (Nestler EJ, McGraw-Hill, 2001). The chromatin is organized into two distinct forms. The chromatin, which is more condensed is called heterochromatin and is transcriptionally inactive. In contrast, chromatin, which is less condensed is called euchromatin. This type of chromatin is transcriptionally active. Chromatin limits the access of transcription factors to their target DNA sequences, thereby preventing the recruitment

of RNA polymerase to the promoter of genes. The unavailability of promoter sequences for RNA polymerases results in repression of gene expression (Struhl K, *et al*., 1999). Nucleosome-mediated repression of gene expression is activated by activator proteins and multiple histone-modifying enzymes. The activator proteins alter the structure of nucleosomes on the DNA thereby altering the chromatin architecture of the genes. They are composed of multiple individual proteins and cofactors tethered together forming multifactorial regulatory protein assemblies. Histone modifying enzymes change the structure of the core histones of nucleosome thereby allowing chromatin to unwind to finally permit transcription to occur (Marmorstein R, *et al*., 2009).

In eukaryotes transcription occurs mainly in the nucleus but also inside mitochondria. Nuclear transcription is initiated by clearing of nucleosomes from the promoter paving the way for general transcription factors (GTFs) to access the promoter region of activated genes. GTFs form a complex of regulatory proteins bound by combinatorial molecular interactions with each other and with the promoter DNA (Fuda NJ, *et al*., 2009). GTFs operate by interacting either directly or indirectly with specific regulatory elements of the core promoters frequently localized near the transcriptional initiation site of each gene. Activators and repressors of transcription are specific transcription factors that interact with other regions (proximal or distal cis-regulatory elements) of the promoter and the coordinated association of these transcription factors forms an active transcription initiation complex. This complex recruits an appropriate RNA polymerase to transcribe the desired gene to yield a pre-mRNA transcript (Struhl K, *et al*., 1999). Within the nucleus, the newly synthesized pre-mRNA transcript undergoes extensive post-transcriptional processing before it is exported to cytoplasm for translation (Dreyfuss G, *et al*., 1988).

1.1.2. Role of RNA modification in gene expression regulation

Post-transcriptional processing of primary mRNA transcripts involves three major events: i) 5'-Capping, ii) pre-mRNA splicing, and iii) 3'-polyadenylation. These three events take place in the nucleus and occur co-transcriptionally (Proudfoot NJ, *et al*., 2002).

i) 5'-Capping: A cap structure is added to the 5' end of almost all eukaryotic mRNAs. The synthesis of the cap involves three different elementary reactions. The first reaction occurs after about 20-30 nucleotides have been transcribed (Proudfoot NJ, *et al*., 2002). An RNA 5' triphosphatase hydrolyzes the triphosphate of the first nucleotide to generate an RNA diphosphate. Next, a guanylyl transferase adds a GMP moiety from GTP to the first nucleotide of the RNA diphosphate forming an unusual 5'-5' triphosphate linkage. Finally, a methyltransferase methylates the transferred GMP at N^7 position to form

the m^7G cap (Shatkin and Manley 2000; Gu and Lima 2005). The m^7G cap binds to a specific cap binding complex (CBC), which plays an important role for translation via binding to the translational initiation factor, eIF-4E. The m^7G cap is believed to serve many functions from stabilizing the mRNA, enhancing its translation, to nuclear export of mRNAs via the nuclear pores (Ramanathan A, *et al*., 2016).

ii) Pre-mRNA splicing: The coding region of most genes in vertebrates is interrupted by non-coding sequences called introns. The introns are precisely excised from the pre-mRNA transcripts and the flanking exons are stitched together in a process called splicing. Splicing involves two transesterification reactions to generate a functional mRNA that is transported to the cytoplasm and translated into protein. First, the 2'-OH of a branch point adenosine acts as a nucleophile to attack the 5'-splice site to form a lariat structure. The reaction is completed when the 3'-OH of the freed 5'-exon acts as nucleophile to attack the intron-3' exon border releasing lariat structure and ligating the two exon sequences together (Sharp PA, 1994). A complex molecular machine called spliceosome facilitates splicing. This catalytic complex is composed of five small nuclear RNAs (snRNAs) (U1, U2, U4, U5, U6) and more than 100 proteins. These RNA-protein complexes called small nuclear ribonucleoprotein particles (snRNPs) are assembled at the pre-mRNA as short-living organelles (spliceosomes). They remove introns from the nascent RNA transcripts (Kramer, 1996).

iii) 3' polyadenylation: All eukaryotic protein coding mRNAs and long non-coding RNAs (lncRNAs), with the exception of replication-dependent histone mRNAs are processed to receive a uniform 3' poly (A) tails consisting of ~ 200 adenosine nucleotides. Poly (A) tail formation starts immediately after transcription has past the polyadenylation signal sequence. The reaction is directed by a complex polyadenylation machinery and by sequence elements present on the pre-mRNA. First, the RNA stand is specifically cleaved and the cleavage site in almost every mRNA is marked by two characteristic sequence elements: i) the highly conserved upstream AAUAAA hexamer and ii) a less conserved GU rich element, located downstream of the cleavage site. The polyadenylation reaction involves two steps. First, the AAUAAA element is recognized by a multi-protein complex that includes both, the polyadenylation special factor (CPSF, a 160 kDa protein) and a poly(A)-polymerase. The assembly of protein complexes on cleavage site generates the 3' end of the pre-mRNA to which the poly (A) tail is added by the catalytic activity of the poly (A) polymerase (Colgan DF, *et al*. 1997; Minvielle-Sebastia L, *et al*., 1999). The poly (A) tail is required for nuclear export, stability of mature transcripts, but also enhances the translational efficiency of mRNAs (Elkon R, *et al*., 2013).

In addition to these three major events of post-transcriptional RNA modification there are further mechanisms, which modify the RNA structure. RNA editing is one of these processes but the regulation and functional consequences of RNA editing has not been explored sufficiently. RNA editing is a unique process that generates a different RNA sequence by chemical modification of selected nucleotides. RNA editing enhances the diversity of products originating from a single gene. RNA editing occurs primarily in the nucleus and is catalyzed by nucleotide deaminases, such as adenosine deaminase or cytosine deaminase. These enzymes catalyze for instance the deamination of adenosine (A) to inosine (I), which alters the RNA sequence. In the majority of genes RNA editing occurs primarily in non-coding regions with the exception of a few genes, in which it occurs in their coding sequences (Glisovic, Bachorik, Yong, & Dreyfuss, 2008). One of the best examples is RNA editing of the apolipoprotein B100 mRNA, which gives rise to the apolipoprotein B48 mRNA, since editing introduces a pre-mature stop codon. *APOB* pre-mRNA is very long (14 kb) and has a very unusual structure. It consists of 29 exons and 28 introns (see **Figure 1.2**). ApoB naturally exists in two isoforms, apoB-48 (intestine) and apoB-100 (liver) (Chan L, 1993). Both isoforms share a common N-terminal part. The shorter apoB-48 protein is generated, when the long apoB mRNA is edited by the conversion of a CAA codon (exon 26) to a premature UAA termination codon (Khoo, Roca, Chew, & Krainer, 2007). This editing occurs in the intestine by cytosine deaminase. Thus, from the apoB mRNA two related proteins, the long apoB100 and the shorter apoB48 can be synthesized in a tissue-specific manner.

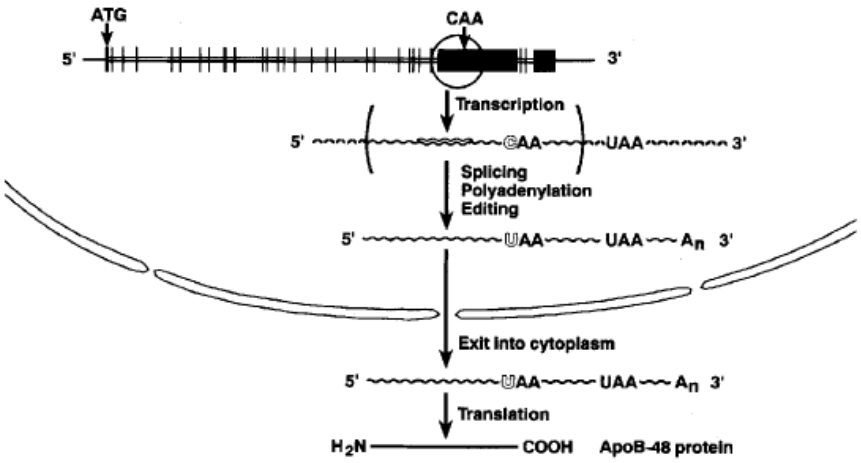

Figure 1.2. Schematic illustration of apoB mRNA editing (Chan L, 1993). The structure of apoB gene (shown at the top) and apoB-48 mRNA containing the mutated codon (CAA>UAA) is displayed at the bottom. The double lines and broad bars depict introns and exons, respectively.

Furthermore, non-coding regions of the human genome contain numerous sequences enriched in inverted Alu and long interspersed nuclear elements (LINE) that function as editing sites for adenosine deaminases. These repeat sequences are present in retrotransposons, which constitute mobile genetic elements that hop all over the genome. Expression of retrotransposons may be controlled by RNA editing.

Naked RNA is very unstable and does not function alone. In fact, in cells RNA is always present as ribonucleoprotein complex (RNPs) that are involved in the processing of newly synthesized RNA transcript. One of the processing events is RNA editing that results in addition or deletion of nucleotides or changes one nucleotide into another. The most prevalent RNA editing in eukaryotic genes is the conversion of adenosine (A) into inosine (I) in double-stranded RNA (dsRNA) through the action of adenosine deaminases acting on RNA (ADAR). The inosine nucleotide is read as guanosine (G) by the translational machinery thereby creating missense codons in mRNAs. The ADAR3 protein family consists of three members that are conserved in vertebrates. With the exception of ADAR3 all the members of ADAR protein family have enzymatic activity. The ADAR enzyme binds directly to the dsRNA substrates via dsRNA-binding domain. The dsRNA substrates form an RNA duplex formed by the interaction between the exon sequences containing adenosine with the downstream intronic sequence generating an editing site where editing occurs. The mechanism by which sites are recognized is highly specific and selective. Furthermore, ADAR1 and ADAR2 modify editing sites differently. Both of these enzymes are expressed in many tissues whereas, ADAR3 is present only in the brain. Most of the neuronal RNA transcripts encoding ion channels, G-protein coupled receptors (GPCRs) undergo A-I modifications (Nishikura, 2006). ADAR gene have been implicated in neuronal dysfunction and tumor malignancy revealing the importance of RNA modification in regulation of gene expression (Glisovic et al., 2008).

1.2. RNA-binding proteins

RNA in cells does not occur as naked polynucleotide but as ribonucleoprotein complexes. The RBPs play a crucial role in every event of posttranscriptional regulation of gene expression including RNA processing, nuclear export, cytoplasmic transport and cellular localization (Chen Y & Varani G, 2013). The RNA-protein interactions are main characteristics of RBPs and protein aberrations are often linked to diseases (Bensaid et al., 2009). There are, to date a total of ~ 1500 RBPs in humans and their remarkable diversity allows them to specifically interact with all known classes of RNAs (see **Figure 1.3**) (Gerstberger et al., 2014). The high versatility of RBPs allows them to fine-tune alternative splicing to create new proteins without affecting existing proteins (Rinn et al., 2014). The RBPs bind to RNAs with different RNA-sequence specificities and affinities.

Most of the RBPs contain well-characterized modular structures called RNA-binding domains (RBDs). RBDs consists of multiple repeats of few domains that combine in various arrangements to create four principle RNA-binding surfaces:

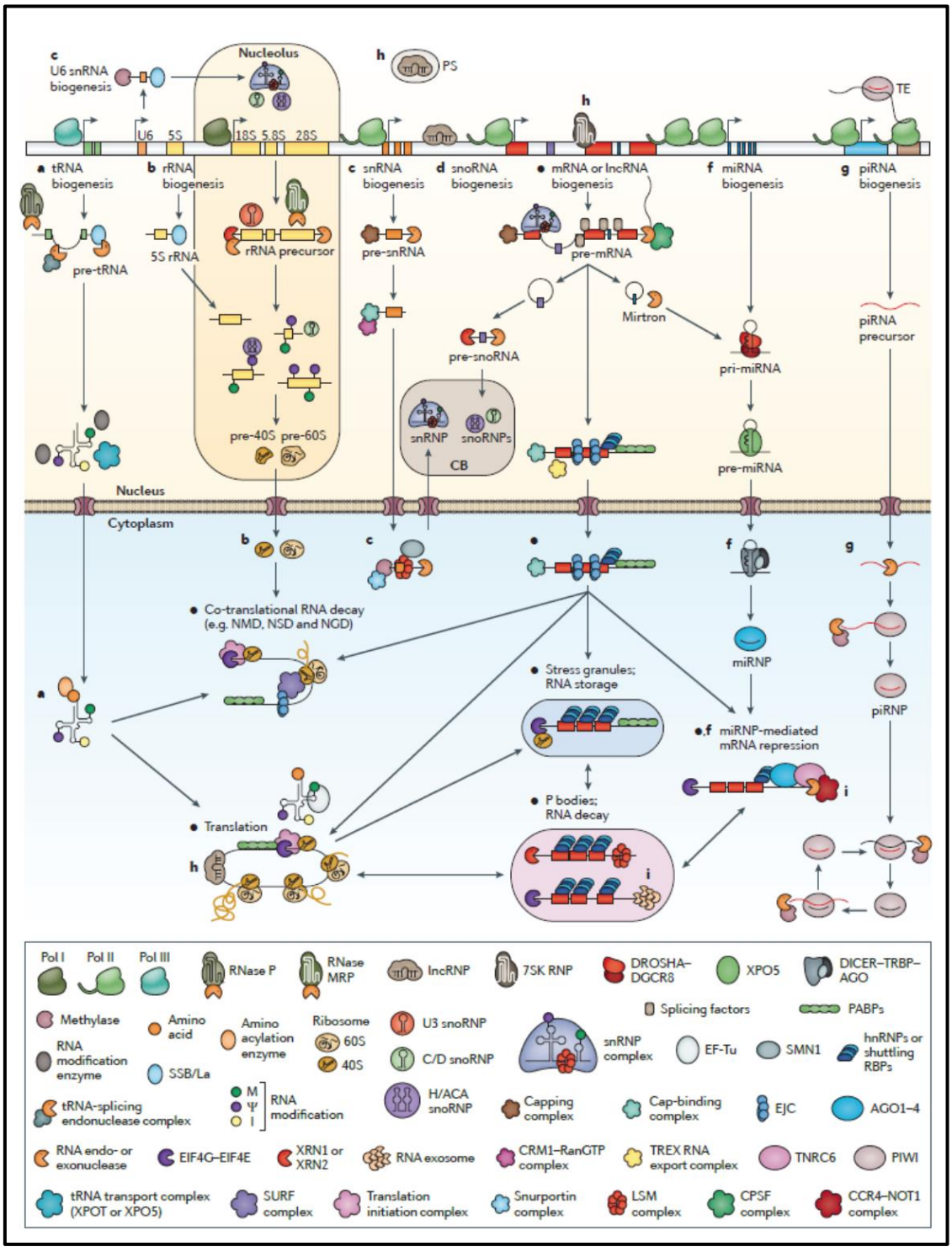

Figure 1.3. Outline of the major post-transcriptional gene regulation pathways in eukaryotes (Gerstberger et al., 2014).

i) RNA recognition motif (RRM), ii) the K-homology (KH) domain, iii) the zinc-finger domain, iv) the double stranded RNA-binding motif (dsRBM) (Beckmann BM, et al., 2016; Gerstberger et al., 2014).

1.2.1. The hnRNP F/H family of proteins

The heterogeneous nuclear ribonucleoproteins F/H (hnRNP F/H) belongs to the subfamily of the hnRNP proteins consisting of hnRNP H1, hnRNP H2, hnRNP H3 (2H9), hnRNP F and GRSF1. The hnRNP proteins were first identified as nuclear proteins that bind heterogeneous nuclear RNA (hnRNA) forming major components of the nucleus (Dreyfuss G, et al., 1988).The hnRNP F/H proteins are ubiquitously expressed proteins and their specific function is not precisely known. However, they are evolutionarily conserved in vertebrates but also occur in lower organisms. They are involved in mRNA capping, splicing, polyadenylation, RNA export, and translation (C. Dominguez, 2006; Cyril Dominguez, Fisette, Chabot, & Allain, 2010). The hnRNPs F and H have been implicated in the splicing of apoptotic *Bcl-X* m-RNA generating two different protein isoforms Bcl-X_L and Bcl-X_S with antagonistic function (C. Dominguez, 2006). The hnRNP F also regulates RNA modification by interacting with the nuclear cap binding protein complex (CBC). The hnRNP H1 and H2 function as part of the nuclear matrix and H3 is involved in splicing arrest induced by heat shock (Honoré et al., 1995). With the exception of GRSF1 that is located in the nucleus, the cytoplasm, and in the mitochondria most members of the hnRNP F/H family are primarily found in the nucleus. The hnRNP F/H proteins contain three conserved quasi-RNA recognition motif (qRRM) and two poorly characterized glycine-rich domains (GRDs). One of them (GYR) is located between qRRM2 and qRRM3 and the other one (GY) at the C terminus (Qian & Wilusz, 1994). The qRRMs recognize G-rich stretches of RNA (G-tract) and maintain the RNA in single-stranded conformation by remodeling secondary structures (G-quadruplexes) (Cyril Dominguez et al., 2010; Samatanga, Dominguez, Jelesarov, & Allain, 2013). The glycine-rich domain located between qRRM2 and qRRM3 interacts with Trn 1 import receptor to mediate intracellular shuttling of the hnRNP H and F (Van Dusen, Yee, McNally, & McNally, 2010).

1.2.2. The RNA recognition motif (RRM)

The RNA recognition motif (RRM) was originally discovered by biochemical characterization of the polyadenylate binding protein (PABP) and hnRNP C proteins. These proteins interact with heterogeneous nuclear RNAs (hnRNAs) and pre-mRNA in the nucleus (Dreyfuss G et al., 1988). The RRM is the most abundant structural motif in vertebrates present in about 2% of all human genes. The RRM domain is highly plastic

and does not only interact with RNA but also with DNA, proteins and also with lipids (Clingman et al., 2014). RRM is the most completely characterized class of RNA-binding domains (RBDs) with a domain length of 80-90 amino acids. The domain folds into four-stranded anti-parallel β-sheets packed against two α-helices that adopts a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology (see **Figure 1.4**)(Cléry & Allain, 2011). The conserved residues located in the two central β1 and β_3 strands are called RNP1 and RNP2, respectively. The RNP1 is a highly conserved eight amino-acids long sequence, which has been implicated in RNA interaction. The consensus sequence of RNP1 is Lys/Arg-Gly-Phe/Tyr-Gly/Ala-Phe/Tyr-Val/Ile/Leu-X-Phe/Tyr, where X can be any amino acid. The second six residue long sequence located at the N-terminus of the domain is called RNP2 having a consensus sequence Ile/Val/Leu-Phe/Tyr-Ile/Val/Leu-X-Asn-Leu (Cléry & Allain, 2011; Daubner, Cléry, & Allain, 2013; Ufer, 2012). The majority of the amino acid residues in RNP1 and RNP 2 are buried within the hydrophobic core of the protein formed by β-sheets and thus, they are not available for RNA-interaction. However, this arrangement of amino acids exposes four conserved residues that are mostly aromatic and positively charged residues to form a RNA-binding site on the surface of the central β-sheet. RNP1 and RNP2 involve typical RNA binding modules, which are present in all RRMs with the exception of representatives of the hnRNP F/H family of RBDs that contain a special binding domain designated the quasi-RNA recognition motif (qRRM) (C. Dominguez, 2006). This binding motif is defined by the lack of most of the conserved aromatic residues in the RNP1 and RNP2 regions (Chen Y & Varani G, 2013; Maris).

The secondary structure in RRMs is interrupted by non-conserved parts (loops). With the exception of loop 5 that adopts a small two-stranded β-sheet structure (β_3' and β_3') all the loops 1-4 are unstructured in their ligand-free form but become structured upon RNA-binding (Cléry & Allain, 2011). The RRM domain has the potential to adopt a wide range of different structures by altering their loops and position of secondary structures (Afroz, Cienikova, Cléry, & Allain, 2015). Furthermore the N- and C-terminal regions fold into secondary structure elements thereby extending the canonical RNA-binding surface. However in hnRNP F this extension is prevented by the C-terminal part that masks the β-sheet surface by adapting a α-helical structure. The RRM domain has most extensively been studied and numerous structures either in isolation (248 free RRMs) or RNA-bound (70 RRM-RNA complexes) have been solved (see **Table 1.1**) (Afroz et al., 2015).

However, the general code of RNA-recognition by the RRMs is not precisely understood (Afroz et al., 2015). The mode of RNA recognition by the RRMs is highly versatile, most of the RBPs contain RRMs that bind to the similar sequences in an independent manner (Cyril Dominguez et al., 2010).

Table 1.1. Structures of RRM domains in complex with RNA. The structures are deposited in Protein Data bank with corresponding PDB numbers (http://www.rcsb.org/pdb/home/home.do).

RRM PROTEIN STRUCTURE TITLE	PDB ID
Spliceosomal U2B″-U2A′ proteins bound to a fragment of the U2 snRNA	1A9N
U1A in complex with the RNA polyadenylation inhibition element	1AUD
U1A in complex with an RNA hairpin	1URN
U1A in complex with PIE RNA	1DZ5
Sex-lethal in complex with the tra mRNA precursor	1B7F
PABP in complex with a polyA tract RNA	1CVJ
Nucleolin RRM1 + 2 in complex with the SNRE RNA	1FJE
Nucleolin RRM1 + 2 in complex with a pre-rRNA target	1RKJ
HuC RRM1+2 in complex with a AU-rich element (ARE)	1FNX
HuD RRM1 + 2 in complex with a AU-rich element (class I ARE)	1FXL
HuD RRM1 + 2 in complex with an AU-rich element (class II ARE)	1G2E
PTB RRM1 in complex with the CUCUCU RNA	2AD9
PTB RRM2 in complex with the CUCUCU RNA	2ADB
PTB RRM3 + 4 in complex with the CUCUCU RNA	2ADC
Hrp1 RRM in complex with the UAUAUAUA RNA	2CJK
Fox-1 RRM in complex with the UGCAUGU RNA	2ERR
RBMY RRM in complex with a RNA stem loop	2FY1
U2AF65 RRM in complex with a polyuridine tract	2G4B
SRp20 RRM in complex the CAUC RNA	2I2Y
hnRNP F qRRM1 in complex with a G-tract RNA	2KFY
hnRNP F qRRM2 in complex with a G-tract RNA	2KG0
hnRNP F qRRM3 in complex with a G-tract RNA	2KG1
Prp24 RRM2 bound to a fragment of the U6 snRNA	2KH9
Tra2beta1 RRM in complex with the AAGAAC RNA	2KXN
Tra2beta1 RRM in complex with the GAAGAA RNA	2RRA
Nab3 RRM in complex with the UCUU RNA	2L41
La in complex with the UGCUGUUUU RNA	1ZH5
La in complex with the AUAUUUU RNA	2VOD
La in complex with the AUAAUUU RNA	2VON
La in complex with the UUUUUUUU RNA	2VOO
La in complex with the AUUUU RNA	2VOP
RNA15 RRM in complex with the GUUGU RNA	2X1A
Nab3 RRM bound to the UUCUUAUUCUUA RNA	2XNR
CUGBP1 RRM1 in complex with the GUUGUUUUGUUU RNA	3NNH
CUGBP1 RRM 1+2 in complex with UGUGUGUUGUGUG RNA	3NNC
CUGBP1 RRM3 in complex with the UGUGUG RNA	2RQC
DEAD box helicase YxiN in complex with a 23S rRNA fragment	3MOJ
CFIm68 RRM/CFIm25 in complex with RNA	3Q2T
Human Spliceosomal U1 snRNP	3CW1

The RRM recognizes short stretch of RNA (approx. 2-10 nucleotides in length) with low specificity and affinity. The high sequence specificity and affinity is achieved in multiple RRMs by cooperative binding that creates a large RNA-binding platform such as nucleolin (NCL), poly (A)-binding protein (PABP) (Beckmann BM, et al., 2016; Cléry & Allain, 2011).

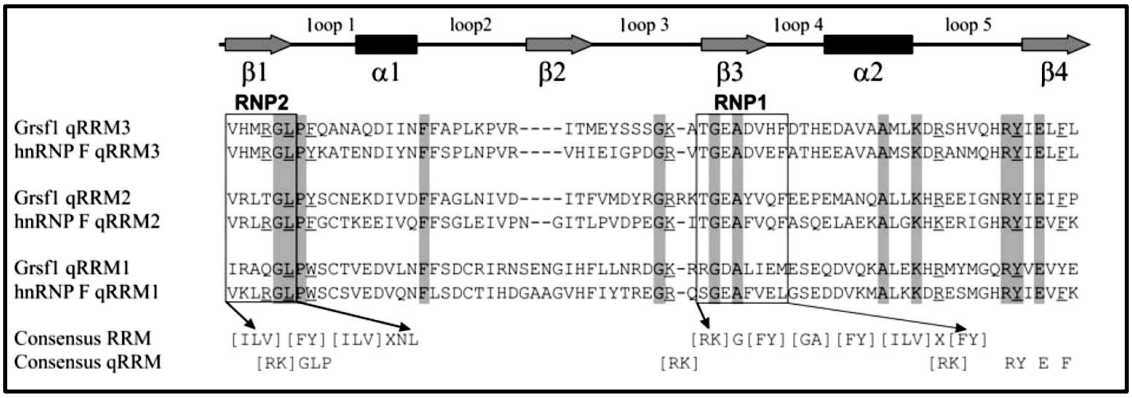

Figure 1.4. Alignment of qRRM of the murine hnRNP F and GRSF1 (Ufer, 2012). *Conserved amino acids are highlighted with gray shadow and amino acids residues involved in the recognition of A1-G2-G3-G4-A5-U6 RNA hexamer by hnRNP F are underlined.*

1.2.3. The quasi-RNA recognition motif (qRRM)

The quasi-RNA recognition motif (qRRM) was first reported in hnRNP F protein that belongs to the hnRNP F/H family of RBPs (C. Dominguez, 2006). The qRRM is unique in that its RNA-binding domains lack conserved aromatic residues in RNP1 and RNP2 (Honore et al., 1995). The essential length of the qRRM domain is undefined and the minimum target RNA sequence length is not known. All qRRMs adopt a three-dimensional structure that comprises of four anti-parallel beta-sheets with two interspersed α-helices forming a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology (Cléry & Allain, 2011). The qRRMs differ in their mechanism of substrate binding from classical RRMs that bind RNA via the canonical β-sheet. In contrast, the loop regions form the sites of RNA interaction in qRRMs. The three qRRMs of hnRNP F interact with a G-tract in an identical and unusual manner by maintaining the RNA in single-stranded forms (Dominguez et al., 2010; Samatanga et al., 2013). The amino acid residues involved in G-tract recognition in loop 1 and loop 5 are conserved among identified qRRMs suggesting the conserved mode of binding in qRRMs (Dominguez et al., 2010).

1.3. GRSF1 as RNA-binding protein: Characterization and RNA-binding properties

Guanine-rich RNA sequence binding factor 1 (GRSF1) is an RNA-binding protein involved in the regulation of post-transcriptional gene expression (Ufer et al., 2008). It is a

member of the heterogeneous nuclear ribonucleoprotein (hnRNP) F/H protein family that have been implicated in regulation of capping, splicing and polyadenylation of numerous cellular pre-mRNAs (see **Figure 1.5**)(C. Dominguez, 2006). It contains three quasi-RNA recognition motifs (qRRMs), an acidic domain in between qRRM2 and qRRM3 and a N-terminal alanine-rich domain found in some isoforms (see **Figure 1.5**). These qRRMs specifically recognize G-rich motifs ($AGGG^{A}/_{G}$) that are conserved sequences found in its target RNAs and control mRNA translation, RNA stability and maturation (Ufer, 2012). The acidic domain is rich in glutamate and proline residues and its function is not known. A second auxiliary domain is the alanine-rich domain. The exact function of this domain is not clear. GRSF1 is conserved in vertebrates (Antonicka, Sasarman, Nishimura, Paupe, & Shoubridge, 2013) but according to our database searches GRSF1-like sequences also occur in lower organisms. The qRRMs of GRSF1 show a high sequence identity with its related family member hnRNP F. Both, hnRNP F and GRSF1 contain three qRRM domains. When the qRRMs of these two proteins are compared, the two N-terminal domains qRRM1 and qRRM2 share over 50% amino acid sequence homology within their coding sequence and the C-terminal domain qRRM3 are even more similar (66% sequence homology). The RNP1 and RNP2 domains are poorly conserved in hnRNP F/H protein family but the amino acid residues in the linking regions, which is responsible for qRRM/RNA recognition and interaction, are mostly conserved in GRSF1 (Ufer, 2012).

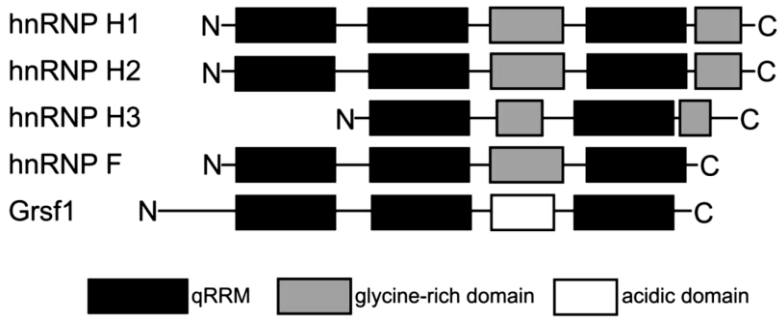

Figure 1.5. Diagrammatic representation of the domain organization of hnRNP F/H family proteins (Ufer, 2012).

1.3.1. G-tracts: Binding motif of hnRNP F/H family

The hnRNP F/H belongs to the family of hnRNP proteins, which are one of the most abundant nuclear proteins (Dreyfuss G, et al., 1988). These proteins are multifunctional that are not only implicated in polyadenylation, but are also responsible for regulating post-transcriptional processing of various pre-mRNAs. However, the detailed mechanisms by which hnRNP F/H proteins regulate post-transcriptional events are still not well understood. The hnRNP F/H proteins bind G-rich sequences of RNA known as G-

tracts. G-tract is a sequence of two, three or more consecutive guanines that are present in the surroundings of splice sites, of polyadenylation signals, but also in the untranslated regions (UTR) of mRNAs (Cyril Dominguez et al., 2010; Mukundan & Phan, 2013). These G-tracts have the potential of self-association forming non-canonical secondary structures called G-quadruplexes (Mukundan & Phan, 2013).

The hnRNP F/H family is unique in that its RNA-binding domains are less conserved within their coding sequence as found for other RRMs (Honore et al., 1995). The mechanism of hnRNP F binding has been elucidated in detail and a similar binding mechanism was proposed for GRSF1 (Ufer, 2012). Direct structural data for full-length GRSF1 is currently not available. However, the NMR structures of qRRM1 (PDB ID: 2HGL), qRRM2 (PDB ID: 2HGM), and qRRM3 (PDB ID: 2HGN) of hnRNP F have been determined (Cyril Dominguez et al., 2010). The members of hnRNP F/H family possess three quasi-RNA recognition motifs (qRRMs): qRRM1, qRRM2, and qRRM3 are capable of binding to RNA independently of each other (C. Dominguez, 2006; Cyril Dominguez et al., 2010). The qRRMs are folded into four anti-parallel beta-strands (ß1-ß4) with two interspersed alpha-helices (α_1-α_2) forming a canonical $ß_1\alpha_1ß_2ß_3\alpha_2ß_4$ topology.

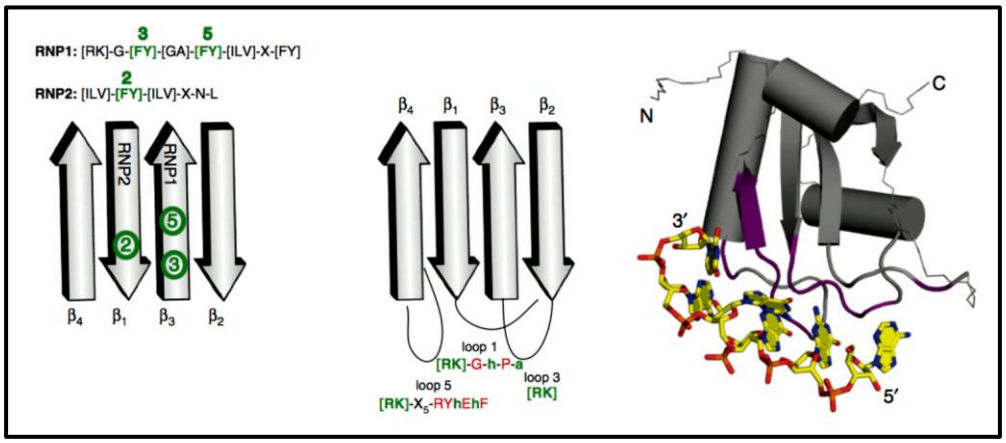

Figure 1.6. Comparison of qRRM and RRM domains and mode of RNA-binding in hnRNP F qRRMs (Daubner et al., 2013).

With the exception of hnRNP F where qRRM1 and qRRM2 have a β-hairpin (β3' and β3') between α_2 and β_4 they also have an additional α-helix (α3) at the C-terminal end. This additional α-helix forms a small hydrophobic core preventing amino acids in the β-sheet surface to participate in canonical RRM-RNA interactions (Dominguez and Allain, 2006). GRSF1 binds to a conserved G-rich motif ($AGGG^{A}/_{G}$) of RNAs via three qRRMs. The three well characterized GRSF1 substrates [influenza nucleoprotein (NP) mRNA (AGGGA), glutathione peroxidase 4 (GPx4) mRNA (AGGGGA), unconventional SNARE in the ER 1 homolog (Use1) mRNA (AGGGGA)] share the RNA binding sequence (Nieradka

et al., 2014; Park, Wilusz, & Katze, 1999; Ufer et al., 2008). The G-rich motif in *GRSF1* substrates is similar to the minimum RNA sequence of hnRNP F defined to be (AGGGAU). The determined RNA sequence (AGGGAU) in *Bcl-x* pre-mRNA folds into tetramolecular G-quadruplex that is recognized by hnRNP F in a novel way (see **Figure 1.6**). The hnRNP F qRRMs (qRRM1, qRRM2, qRRM3) encage the G-tract (GGG) and thus prevent G-tract from adopting other structures (Dominguez et al., 2010). A similar G-rich motif is found in the 5'-UTRs of *GRSF1* substrates (Ufer, 2012).

1.3.2. GRSF1 isoforms and localization

GRSF1 is encoded by a gene located on chromosome 4 (4q13.3) in humans and on chromosome 5 (88659448-88676171 bp) in mouse (Ufer, 2012). The murine *Grsf1* gene is organized into 10 exons and spans a region of approximately 15 kb (Banga SS, et al., 1996). Alternative exons 2-6 code for the two N-terminal RNA-binding domains, whereas downstream exons 7-9 code for the C-terminal RNA-binding domain. The acidic, α–helical domain located between RRM2 and RRM3 is coded by downstream exons 6 and 7 (Ufer, 2012). The gene locus of human GRSF1 has been correlated with familial mesial temporal lobe epilepsy (FMTLE). However, the involvement GRSF1 in the pathogenesis of this disorder has not been investigated in detail (Hedera, Blair, & Andermann, 2007). The alternative exons in conjunction with downstream exons give rise to four different isoproteins, which differ with respect to their N-terminal sequences (Qian & Wilusz, 1994; Ufer, 2012). The mechanism by which these isoforms are produced is not known. However, alternative splicing or alternative transcriptional initiation may be involved (Ufer, 2012; (Qian & Wilusz, 1994). With the exception of isoform 2 (NCBI Reference Sequence: NM_001098477.1), which lacks the N-terminal alanine-rich domain (amino acids 26-111), all GRSF1 isoforms contain three conserved RNA-binding domains (RBDs) (Ufer, 2012). The functionality of the alanine-rich domain of human GRSF1 is not known. However, this structural subunit was predicted to involve a mitochondrial targeting signal and thus, might be important for mitochondrial import (Antonicka et al., 2013). In fact, the GRSF1 isoforms lacking the alanine-rich domain are predominantly localized in the cytoplasm. The longest GRSF1 isoform (NCBI Reference Sequence: NM_002092.3) consists of 480 amino acids. It has been detected in mitochondrial RNA granules and may play a role in precursor RNA processing (Jourdain et al., 2013).

1.3.3. GRSF1 expression

Low level gene expression of *GRSF1* mRNA is detected in most mammalian tissues with the exception of spermatogenic cells, where high level expression are found

(Ufer et al., 2008). In the developing embryo *GRSF1* gene expression starts early on during embryonic development and the regulatory mechanisms for the switch on of GRSF1 expression are not known. Comparisons of the murine and human proximal promoter regions of the *GRSF1* gene indicated a number of structural similarities, particularly in the putative GC and initiator (Inr) regions. Both of these *cis*-regulatory elements are present in the human and mouse gene. *In silico* analysis predicted that the promoter region of both genes does neither contain a TATA nor a CCAAT box. However, a NF-κB binding site located 200 base pairs upstream of the translational start site is present in both genes. The functionality of this upstream regulatory region of the *GRSF1* gene has not been investigated (Ufer, 2012). NF-κB belongs to a family of transcription factors that are involved in the regulation of more than 100 genes ranging from pro-inflammatory cytokines, redox-regulated genes, and growth factors to gene involved in the regulation of apoptotic cell death (Hoffmann A, et al., 2006; Ufer et al., 2010). NF-kB was one of the first transcription factor identified to be redox-regulated in eukaryotic cells (Gloire G, et al., 2006). Expression of *GPx4* is regulated by TNFalpha (Hattori et al., 2007), which usually activates NFkB (Sneddon et al., 2003). However, the promoter region of the *GPx4* gene does not contain a NF-kB binding site and thus, the molecular mechanism by which TNFα induces *GPx4* expression is not well understood (Hattori et al., 2007; Sneddon et al., 2003). Indirect transcriptional and post-transcriptional events, which do not follow the canonical mechanistic pathway, may play a role in this process (Sneddon AA, et al., 2003). The promoter region of the *GRSF1* gene involves numerous *cis*-regulatory elements for other transcription factors such as Tcf/Lef (T-cell factor/Lymphoid enhancer factor). Tcf/Lef belongs to a family of transcription factors that are downstream targets of the canonical Wnt/β-catenin signaling pathway, which is involved in many cellular processes including differentiation, proliferation, gastrulation and axial embryo development (Sokol, 1999). Furthermore, the Wnt/β-catenin signaling pathway might indirectly require *GRSF1* since it may stabilize regulatory downstream components of the pathway. However, GRSF1 might not be directly involved in Wnt/β-catenin signaling (Lickert H, et al., 2005).

1.3.4. GRSF1 mode of action

As indicated above GRSF1 affects numerous post-transcriptional events (Jablonski & Caputi, 2009; Kash et al., 2002; Schaub, Lopez, & Caputi, 2007). However, so far only three RNA substrates have been identified as targets of GRSF1: i) Glutathione Peroxidase 4 (*GPx4*) mRNA, ii) unusual SNARE in the ER-1 (*Use1*) mRNA and iii) the influenza nucleoprotein (*NP*) mRNA (Nieradka et al., 2014a; Park et al., 1999; Ufer et al., 2008).

GRSF1 was first identified as a cytosolic RNA-binding protein that interacts with a conserved 14-nucleotide G-rich sequence (Qian & Wilusz, 1994).

One of the most interesting activities of GRSF1 is that it promotes selective translation of viral mRNA after infection of host cells with the influenza virus (Kash et al., 2002; Park et al., 1999). The influenza virus attenuates host cell protein synthesis and redirects the host cell protein synthesizing machinery towards the synthesis of influenza virus proteins. This process depends on the interaction between the conserved sequences present in the 5' UTR of the influenza virus mRNAs and recruited host cell proteins. The GRSF1 functions by binding to the conserved G-rich sequence (AGGGU) in the 5' UTR of the influenza virus RNA and this interaction leads to a strong upregulation of expression of the influenza nucleocaspid protein (NP) mRNA. GRSF1 does not function alone to stimulate translation, instead it may interact with other eukaryotic initiation factors to recruit ribosomes to NP mRNA and to stimulate mRNA translation (Kash, et al., 2002; Park et al., 1999). Moreover, GRSF1 has been reported to be involved in translation of mRNAs containing internal ribosomal entry sites (IRES) elements (Cobbold LC, et al., 2008).

Another example for regulation of mRNA translation by GRSF1 is *GPx4* gene expression. *GPx4* is a moonlighting selenoprotein and its mRNA is a target for GRSF1. GRSF1 up-regulates the expression of this RNA by interacting with a G-rich motif (AGGGGA) in the 5' UTR of the *GPx4* mRNA (Ufer et al., 2008). This complex along with other translation factors recruits *GPx4* mRNA to polysomes and activates translation. Expression silencing of GRFS1 by siRNA during in vitro mouse embryogenesis leads to defective embryonic development (Ufer et al., 2008). GRSF1 has also been implicated in processing of viral mRNAs including HIV mRNA, Herpes simplex virus type 1 (HSV-1) and regulates translation of micro-RNA (miR-346) by recruiting it to polysomes (Wang X, et al., 2016;Jablonski & Caputi, 2009).

The third example for the translational regulatory function of GRSF1 is its involvement in expression of the *Use1* mRNA. The Use1 protein has been implicated in retrograde transport of CopI coated vesicles, which shuttle chaperones and other ER-resident proteins back from the Golgi to the ER. GRSF1 binds to the *Use1* mRNA via a G-rich element (AGGGGA) in the 5'UTR of the *Use1* mRNA. The AGGGGA recognition sequence is a part of larger structure that can potentially fold into secondary structure G-quadruplex. Furthermore, RNA shift assays showed that deletion of surrounding $A(G)_3GGGA$ had a more pronounced impact on translation than deletion of the $A(G)_4A$ element itself (Nieradka et al., 2014).

GRSF1 is a multifunctional protein that not only binds to G-rich RNA sequences in the nucleus and the cytoplasm but also interacts with coding and long non-coding RNAs

(lncRNA) in the mitochondria (Antonicka & Shoubridge, 2015). In the mitochondrial context, GRSF1 was originally identified as RNA-binding protein that is required for oxidative phosphorylation (Bayona-Bafaluy et al., 2011). GRSF1 associates with nascent RNA in the mitochondrial matrix to form a dynamic structures called mitochondrial RNA granules (MRGs). MRGs are composed of translationally silent mRNA in complex with RNA-binding proteins. These ribonucleoparticles (RNPs) assemble themselves into microscopically visible mitochondrial RNA granules. Within MRGs the GRSF1 interacts with the RNase P complex and plays an essential role in RNA processing (Jourdain et al., 2013). Furthermore, GRSF1 binds to three mtRNAs including *ND6* mRNA and two lncRNAs (*cytb and ND5*) in mitochondrial RNA granules. These G-rich transcripts are transcribed from the light-strand promoter of mtDNA and each of these transcripts contains multiple GRSF1 consensus binding sites (AGGGD), where D is either A, U, or G. GRSF1 selectively interact with these G-rich binding sites and is involved in translation of mitochondrial RNAs and in ribosome assembly (Antonicka et al., 2013). In addition, GRSF1 has been suggested to bind to a nuclear DNA-encoded lncRNA called *RMRP* (the RNA component of the RNA processing endoribonuclease [RNase MRP]). Here GRSF1 binds to G-rich sequence (AGGGGA) in the 5'-UTR of *RMRP* lncRNA and plays an essential role in mediating its transport to mitochondria. However, the detailed mechanism has not well been elucidated (Noh et al., 2016).

1.3.5. Role of GRSF1 in cell signaling

Although the GRSF1 has been suggested to be involved as a downstream target of several intracellular signaling pathways [Wnt/β-catenin and mammalian target of rapamycin (mTOR) signaling] the biological function of GRSF1 in eukaryotic cells is still under discussion (Ufer, 2012). *GRSF1* knockout mice are currently not available and specific *GRSF1* inhibitors have not been developed. Other loss of function strategies such as RNAi-mediated expression silencing have been employed in several cellular *in vitro* systems but the results of these experiments cannot directly be extended to the *in vivo* situation.

The Wnt signaling pathway has been implicated in numerous biological events including cell differentiation, regulation of proliferation and embryonic development (Sokol, 1999). It is initiated by the binding of the secreted glycoprotein Wnt to the frizzled receptor on the cell surface of target cells. The Wnt receptor complex activates the cytoplasmic protein β-catenin. This β-catenin functions as an adaptor protein that forms a main component of cadherin-cell adhesion complex, but also works as an activator of gene transcription. The activated β-catenin translocates into the nucleus and activates the

Tcf/Lef transcription factors (Nelson & Nusse, 2004). It also cooperates with many other transcription factors to regulate expression of other specific target genes. RNAi-mediated gene knockout studies of *GRSF1* in developing mouse embryos showed impaired embryonic brain development (Ufer et al., 2008). Although the *GRSF1* gene promoter contains putative binding sites for Wnt-responsive transcription factors the precise role of GRSF1 in these processes remains to be explored (Ufer, 2012).

Mammalian target of rapamycin (mTOR) signaling regulates many cellular processes including cell growth, survival, as well as plays role in cancer (Hung, Garcia-Haro, Sparks, & Guertin, 2012). The mTOR is a key component that mediates both extracellular signals (hormones, mitogens, hypoxia, and trophic factors) and intracellular signals (DNA damage, oxidative stress, viral infection, and heat shock) and transduces them into changes of gene expression (Wullschleger, Loewith, & Hall, 2006). The mTOR protein is a serine-threonine kinase that belongs to the family of phospho-inositide 3-kinase (PI3K)-related kinases. It is ubiquitously expressed in mammalian cells and the mTOR pathway is employed by a variety of interconnected signaling cascades in mammalian cells. Activation of mTOR triggers the phosphorylation of the 4E-BP (4E binding protein), which leads to subsequent release of eIF4E (eukaryotic translation initiation factor 4E). eIF4E binds to the mRNAs containing cap structures and thereby forms a complex with other translation factors to initiate translation of mRNAs. Besides enhancing translational efficiency mTOR also regulates downstream target genes including *GRSF1*. The expression of *GRSF1* is indirectly regulated at translational levels by the RNA-binding protein Daz1 (Deleted in azoospermia-like 1) (Jiao X, et al., 2002). Daz1 is a germ cell-restricted RNA-binding protein that is highly expressed in germ cells and plays a key role in spermatogenesis (Yen, 2004). A similar tissue specific expression pattern has been reported for GRSF1, which is also expressed at high levels in spermatozoa. Daz1 regulates expression of multiple target mRNAs at various stages of sperm development. It binds directly to the 3'-UTR of *GRSF1* mRNA to enhance its translation. However the mechanism of translational activation is unclear (Ufer, 2012; Jiao X, et al., 2002).

GPx4 mRNA is one of the most intensively studied substrates of GRSF1. GPx4 protein performs multiple functions. It works as enzyme in anti-oxidative defense and also functions as regulator of gene expression and apoptosis (Ufer et al., 2008). In addition, it is a structural protein required for the formation of the mitochondrial capsule, which arrests the mitochondria in the mid piece of sperms (Brütsch SH, et al., 2015). The primary transcript of the *GPx4* gene undergoes alternative splicing to generate three distinct GPx4 isoforms, which are differentially localized in the cell: i) mitochondrial (m-GPx4), ii)

cytosolic (c-GPx4), and iii) nuclear (n-GPx4). GRSF1 binds to the G-rich motif within 5'-UTR of *m-GPx4* mRNA and regulates its translation. RNA gel shift assays showed that the minimum RNA binding motif that interacts with GRSF1 is 27-nt motif. Binding of GRSF1 to the mRNA of *mGPx4* up-regulates m-GPx4 expression and the molecular basis for this effects is that this interaction recruits *m-GPx4* mRNA to translationally active polysomes (Ufer et al., 2008). *In vitro* embryogenesis studies indicated that siRNA-mediated expression silencing of *GRSF1* in murine embryos induced structural alterations in the brain suggesting that GRSF1 may be essential for embryonic brain development. In detail, knockdown of *GRSF1* expression during *in vitro* embryogenesis impaired midbrain and hindbrain development at later stages of mouse embryo development by inducing lipid peroxidation and enhancing apoptosis. These effects are reversed upon overexpression of *GPx4*. This data suggested that GRSF1 modulates the redox state of the cell by controlling expression of the anti-oxidative *GPx4* gene (Ufer et al., 2008).

1.4. Fundamentals of G-quadruplex structures

Guanine-rich sequences in nucleic acids can self-associate to form unusual four-stranded helical structures called G-quadruplexes (G4s) (Brian O, Regan, 1991; Kim, Cheong, & Moore, 1991; Sen & Gilbert, 1988). These structures were first identified in telomeric DNA at the ends of chromosomes and were likely to be involved in chromosomal maintenance (Chen, 1992). Over the last two decades, extensive studies on G4s have revealed their importance in both DNA and RNA biology. These studies have indicated how these structures are formed, how they are stabilized and how they adopt various conformations to impact the biological function of the G4s.

1.4.1. G-quadruplex structures

The basic building block of G-quadruplex (G4) is a G-quartet (Gellert, Lipsett, & Davies, 1962) that consists of four coplanar guanines arranged in a tetragonal fashion and held together by four guanines hydrogen-bonded via Hoogsteen base-pairs (see **Figure 1.7**). Watson–Crick base-pair interaction do not play a major role. In Watson–Crick base-pair the hydrogen bonding network occurs between purines (Guanine, Adenine) and pyrimidines (Thymine, Cytosine, and Uracil) (Balagurumoorthy & Brahmachari, 1994). This hydrogen bond arrangement between bases gives rise to double helical structures, in which adenine pairs with thymine with only two hydrogen bonds (or uracil in RNA) and guanine pairs with cytosine with its three hydrogen bonds. Guanine and uracil can form a wobble pair in RNA but this interaction is less stable. The anti-parallel nature of strands in Watson–Crick base-pairing causes the N_7 and O_6 of the purine nucleobase not to be used

in the interaction. This is in contrast to Hoogsteen hydrogen bonding where N_7 and O_6 of each guanine forms four Hoogsteen base pairs with two other guanines, two N_1 - O_6 bonds and two N_2 - N_7 bonds (Balasubramanian, 2014). This arrangement generates a small central negative cavity that is stabilized by metal cations preferentially potassium and sodium and hence their formation is favored at physiological conditions (Parkinson et al, 2002; Wang & Patel, 1993). The Hoogsteen hydrogen bonding network is critical in the formation and stabilization of G-quartets (Neidle & Balasubramanian, 2006; Balasubramanian, 2014).

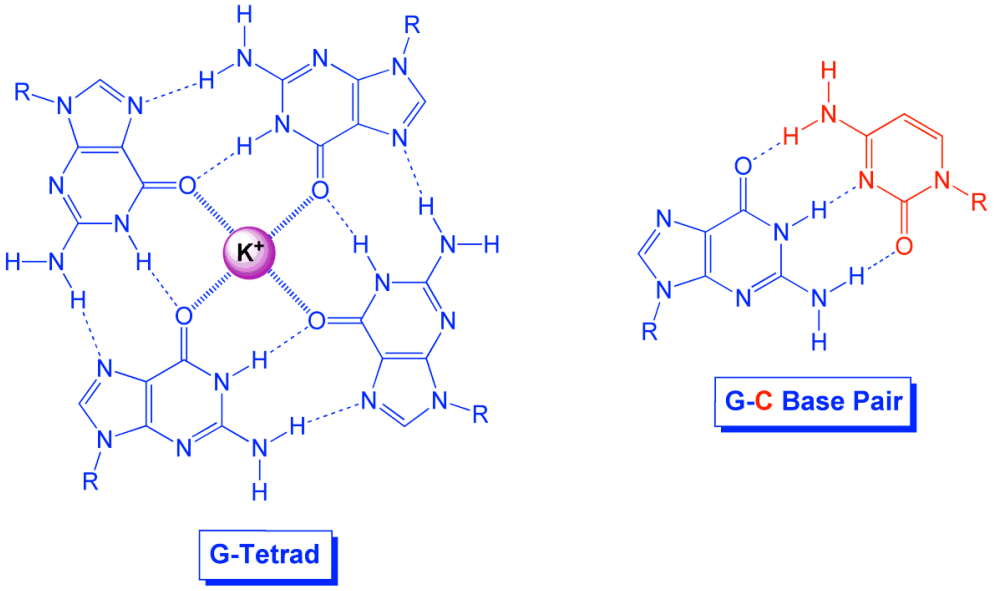

Fig 1.7. Schematic of G-quartet structure and its base pairing (Hoogsteen base versus Watson-Crick base pairing). (Source: http://www.chem.cmu.edu/groups/army/research/project1.html).

1.4.2. Folding and topology of G-quadruplexes

Depending on the orientation of the strands, G4s can be sub-grouped into either parallel, anti-parallel or mixed complexes (Tang & Shafer, 2006; Dai et al, 2007). The conformations arise due to different orientation of glycosidic bond between the pentose sugars and nucleobases of the nucleic acid. The two most common conformations observed in folded DNA structures are the *syn* and *anti* conformations (Tang & Shafer, 2006). The *syn* conformation is observed in DNA where arrangement of the sugar conformation is restricted to C2'-endo because of the absence of hydroxyl group at o2' position of the deoxyribose sugar. In RNA the sugar conformation is generally restricted to C3'-endo, which is partly due to steric hindrance, and partly to the hydrogen bonding associated with the o2' hydroxyl group on the sugar. The *anti* conformation is the

energetically favored conformation found in DNA (Neidle & Balasubramanian, 2006). As indicated above G-rich sequence in DNA can adopt both parallel and antiparallel G4s, while in RNA only parallel G4s are formed. In parallel G4s all the bonds are in the *anti* conformation and in anti-parallel G4s there are is a combination of *anti* and *syn* conformations (see **Figure 1.8**). However, the folding topology in quadruplexes is more complex because of the presence of loops and their different linking arrangements. These loops are of four different types: edgewise loops, diagonal loops, double-chain-reversal loops, and V-shaped loops (Phan, Kuryavyi, & Patel, 2006) (see **Figure 1.9**). In parallel G4s the strands are oriented in the same direction and two parallel strands are linked together by external loops (Phan et al, 2004; Seenisamy et al, 2004; Dai et al, 2007). In anti-parallel G4s the strands are oriented in opposite directions and a lateral loops connects similar anti-parallel strands and diagonal strands connect opposite anti-parallel strands (Agarwala, Pandey, & Maiti, 2015).

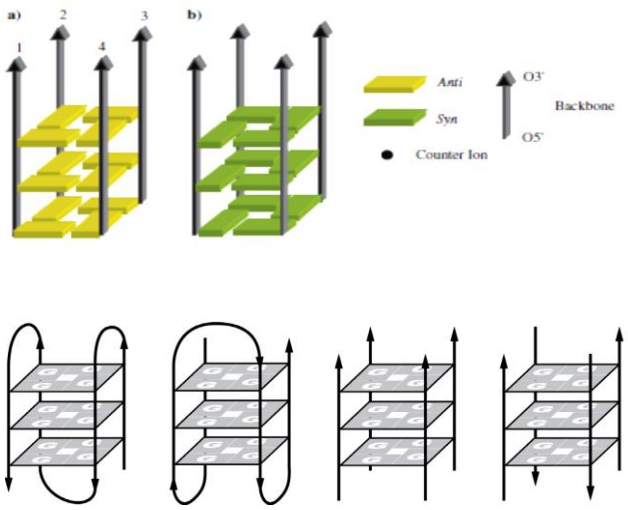

Figure 1.8. Schematic diagram showing G-quadruplex topologies strand polarity and glycosidic angles (Neidle & Balasubramanian, 2006): (**a**) Parallel with all anti glycosidic bond angles; (**b**) parallel with all syn glycosidic bond angles (**c**) Possible topologies for parallel and antiparallel G4s.

With RNA the single stranded nucleotides in G-quadruplexes are unconstrained by normal Watson-Crick base pairing giving them a great deal of flexibility and making them structurally more versatile (Neidle & Balasubramanian, 2006). G4s are highly polymorphic and can adopt a range of different conformations that adds to the diversity of G4s (Phan et al., 2006). These conformations involve intra- or inter-molecular folding of G-rich strands. Intra-molecular quadruplexes require self-association of four or more G-tract in one strand. Whereas inter-molecular quadruplexes arises from two or four associated strands resulting

into bimolecular and tetramolecular G4s (Tang & Shafer; Burge, Parkinson, Hazel, Todd, & Neidle, 2006).

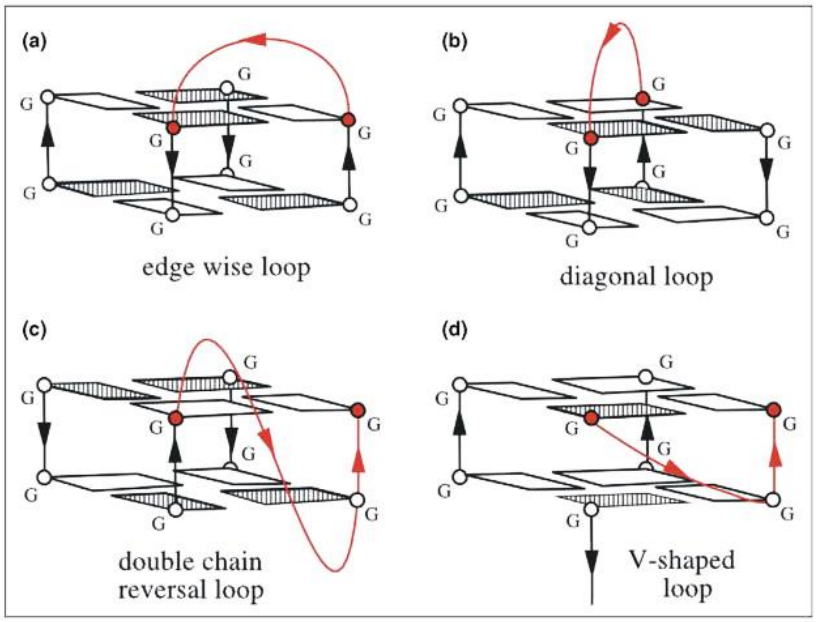

Figure 1.9. Cartoon illustration showing loop polymorphism and arrangement of the G-quadruplex (Phan et al., 2006): (a) Edgewise loop (joining two adjoining antiparallel strands) **(b)** Diagonal loop (joining two opposite antiparallel strands) **(c)** Double chain reversal loop (joining adjacent parallel strands) **(d)** V-shaped loop (Joining two corners of a G-quartet).

1.4.3. Methods used for identifying G-quadruplexes in nucleic acids

The human genome contains a large number of sequences that have the potential to form as many as 376000 G4s (Agarwala et al., 2015). In order to investigate the function of these G4s it is essential to identify the quadruplex forming sequences within various genes or pre-mRNA transcripts. Within the last decade a number of tools and techniques have been developed for the identification and characterization of these structures. These include: (1) Predictive tools to identify the G4-forming sequences; (2) Targeted synthesis of G4-forming sequences (DNA/RNA) and (3) Biophysical tests measuring the formation of G4 complexes. In addition to confirm the existence of G4s in cells and their biological function a number of strategies have been tested, which include the following methods: (1) Targeting G4 structures using synthetic ligands; (2) Using chemical compounds such as pyridostatin (PDS) to modulate G4 function (Rodriguez et al., 2008); (3) Antibody based intervention such as BG4 (structure-specific antibody) for the determination of G4 location in cells and determining the importance of nucleotides involved by mutation (Biffi, Di Antonio, Tannahill, & Balasubramanian, 2014).

The most commonly employed tool for studying G4s is a web program called Quadruplex forming G-Rich sequences (QGRS) Mapper http://bioinformatics.ramapo.edu/QGRS/index.php (Kikin, D'Antonio, & Bagga, 2006). QGRS mapper is used to predict the presence of putative G4s (composed of two or three G-quartets) using the following motif: $\mathbf{G}_x\mathbf{N}_{y1}\mathbf{G}_x\mathbf{N}_{y2}\mathbf{G}_x\mathbf{N}_{y3}\mathbf{G}_x$ where '*x*' is number of guanine tetrads in G4 and *'y_1, y_2* and *y_3'* are the lengths of the loops. The maximum length of sequence that can be analyzed is up to 45 bases (Kikin et al., 2006). However, the main limitation of the QGRS mapper is its inability to predict each and every putative G4 sequence in genome. The classical example is the determination of crystal structure of unusual G4 motif in RNA analog of GFP called spinach aptamer (Huang et al., 2014), which could not be predicted by QGRS mapper. For this reason, QGRS mapper, and similar tools like quadparser, should only be considered advisory (Huppert & Balasubramanian, 2005).

The G4s can be detected by using biophysical techniques such as UV melting, circular dichroism spectroscopy (CD), fluorescence resonance energy transfer (FRET), uv-vis spectroscopy, nuclear magnetic resonance spectrometry (NMR) and electro spray ionization mass spectroscopy (Agarwala et al., 2015). These techniques are reliable and provide valuable information about quadruplex formation. However, more recently the assay conditions (buffers, temperatures, presence and absence of other biomolecules) have become the subject of controversial discussions (Balasubramanian, 2014; Bates, Mergny, & Yang, 2007). Thus, G4s studies should always be carried out under close to physiological conditions.

Functional assays can be carried out using structure-specific antibodies and small synthetic ligands, which more or less specifically target G4s in the promoter region of genes and in different RNA species. Furthermore, the antibody BG4 recognizes and bind G4s with high affinity allowing visualization of G4s in both nucleus and the cytoplasm of cells (Biffi, Tannahill, McCafferty, & Balasubramanian, 2013). However, small synthetic molecules and antibodies may potentially interact with G4-forming sequences and induce them to adopt G4 conformation (Balasubramanian, 2014). Therefore, while using small molecule intervention or structural antibodies such considerations must be taken into account in order to better understand the properties of the G4.

1.4.4. Biological Role of RNA-quadruplexes

G-quadruplexes (G4s) are highly diverse structures that are distributed across the genome (Phan et al., 2006). They are located in various regulatory regions including promoters, telomeres and both 5'- and 3'- UTRs of mRNA transcripts. The biological

importance of G4s is not clearly elucidated however, a significant number of studies have gathered a huge body of evidence suggesting their involvement in important biological process such as mRNA splicing, translational repression, polyadenylation, telomere homeostasis, transcriptional termination and intracellular localization (Beaudoin, Jodoin, & Perreault, 2014; Beaudoin & Perreault, 2010b; Eddy & Maizels, 2006; Huppert & Balasubramanian, 2005). Moreover, G4s are also located in the promoters of various proto-oncogenes such as *c-MYC, C-Kit, c-myb* and *KRAS* (Rhodes & Lipps, 2015; (Beaudoin et al., 2014; Beaudoin & Perreault, 2010b) suggesting a connection with carcinogenesis. In fact, translational repression by RNA G4s was first reported in the 5' UTR of the human *NRAS* (Neuroblastoma RAS) proto-oncogene mRNA. Furthermore, in recent years a number of studies have reported a similar phenomenon of translational repression in different mRNA transcripts including *Zic-1* (Zinc finger of the cerebellum 1), *Bcl-2* (B-cell lymphoma 2) and *TRF-2* (Telomeric repeat-binding factor 2) mRNA (Agarwala et al., 2015; Arora et al., 2008; Gomez et al., 2010; Kumari, Bugaut, Huppert, & Balasubramanian, 2007). However, a small number of studies have also shown that G4s mediate translation in the internal ribosome entry site (IRES) of *FGF-2* and human *VEGF* mRNAs (Bonnal et al., 2003; Morris, Negishi, Pazsint, Schonhoft, & Basu, 2010). In addition G4s located in the 3' UTR of *IGF II* (insulin-like growth factor II) and *p53* mRNAs are involved in alternative polyadenylation (Cayrel, 2011; Christiansen, Kofod, & Nielsen, 1994). G4s also play an essential role in alternative splicing of various genes such as *hTERT* (human telomerase reverse transcriptase) and *FMR1* (Fragile X mental retardation 1) (Didiot et al., 2008; Gomez et al., 2004). More, recently RNA G4s were reported to occur in the cytoplasm of the cells using G4-specific antibody named BG4 (Biffi et al., 2014).

1.5. Genetic polymorphism of the human genome

The genetic variation in human populations was first studied in blood-group antigens where a single gene with three variants (alleles) gives rise to frequencies in ABO blood groups (Crow, 1993). Genetic polymorphism is a difference in DNA sequence among individuals, groups or population. Genetic polymorphisms may be induced by external agents or may result by chance. The genetic variations are frequently found as single nucleotide polymorphisms (SNPs) and here insertions, deletions or nucleotide exchanges occur. The human genome contains approximately 3.1 million SNPs and 90% of human genetic variation are ascribed to SNPs that occur at an allele frequency of >1% (Albert, 2011; Sachidanandam et al., 2001; Sunyaev, Ramensky, & Bork, 2000). SNPs are highly abundant in the human genome and occur at 1 out of every 1,000 bases. The

location of the SNPs plays an essential role for the phenotype of the allele carrier. Most of the physiologically relevant SNPs occur in the coding regions of the genes (exons) and have the potential to alter the structure and the function of the encoded protein. Although SNPs that occur in the non-coding regions (introns and untranslated regions) may also affect different steps of gene expression they are frequently not as severe as coding region SNPs (Syvänen, 2001).

1.5.1. Genetic polymorphism and single nucleotide exchanges

A single nucleotide polymorphism is a difference in single nucleotide of genomic DNA among individuals, groups or population. The genetic polymorphism is caused by random mutations in the genes and promotes diversity within the population. SNPs are the source of variation that arise by a single base mutation in the DNA sequence (Smith, 2002). Based on nucleotide base substitutions the SNPs are of two types:

- **Transition:** This is the most common type of substitution comprising two thirds of all SNPs. It is substitution that occurs between two-ring purines (A, G) or between one-ring pyrimidines (C, T).
- **Transversion:** This type of SNPs occur at lower frequency and persist as silent substitutions. In transversion a substitution occurs between a purine and a pyrimidine base.

The genetic polymorphism is mostly contributed by the SNPs (Smith, 2002) and the distribution of SNPs is uneven across the human genome. The majority of the SNPs are located in non-coding regions but they are also present in coding regions. Functional SNPs are associated with phenotype alterations. These SNPs can change the protein structure or change the amount or timing of protein biosynthesis. The SNPs in the non-coding regions may affect RNA splicing, stability, or translation (Albert, 2011).

1.5.2. Synonymous and non-synonymous SNPs

The point mutations (SNPs) occurring in the coding region of a gene are classified into two distinct types: Synonymous and Non-Synonymous SNPs

Synonymous SNPs: The synonymous mutations are silent in nature and change the nucleotide sequence of the gene without altering the protein sequence (Supek, et al., 2014). The synonymous mutations occur at silent positions in the gene and are generally nonfunctional. However, recent studies have shown that these synonymous sites are non-randomly distributed across the genome and indeed, synonymous sites are target of natural selection (Drummond & Wilke, 2008; Supek, et al., 2010). Some synonymous mutations are functionally active and affect many post-transcriptional and post-translational

events including rate of translation, mRNA folding, splicing, and protein folding (Zheng, Kim, & Verhaak, 2014).

Non-Synonymous SNPs: The non-synonymous mutations alter the amino acid sequence of the encoded protein by causing a change of a codon. These mutations may result in an amino acid exchange or in a premature stop codon. About 50% of all SNPs that occur in coding region result in codon changes (Smith, 2002). The frequency of non-synonymous mutations across the genome is less than one percent. Most of the non-synonymous mutations are neutral and do not alter the structure or function of the proteins. However, some mutations modify the protein structure and are associated with genetic diseases including cystic fibrosis and the Fragile X syndrome (Ferec & Cutting, 2012; Myrick et al., 2014).

1.5.3. Genetic polymorphism of GRSF1

GRSF1 consists of three qRRMs and two auxiliary domains (Park et al., 1999; Qian & Wilusz, 1994). These qRRMs are found in all vertebrates and domain organization of GRSF1 is conserved in various vertebrates suggesting that GRSF1 may serve a conserved purpose in vertebrates (Qian & Wilusz, 1994; Ufer, 2012). However the genetic variability of the human *GRSF1* gene has not been explored before and it remains to be investigated whether human diseases may be related to GRSF1 SNPs. Considering the size of the GRSF1 gene it is almost certain that SNPs will be found when searching the appropriate databases. However, to explore the functional consequences of such SNPs more complex research strategies are required.

1.6. Aim of the study

GRSF1 is a RNA-binding protein, which has been implicated in post-transcriptional RNA processing, RNA transport and ribosomal RNA translation. It was discovered many years ago but the detailed mechanism of its interaction with target RNA has not been explored in detail. Moreover, the genetic variability of human GRSF1 and the functional consequences of the naturally occurring GRSF1 mutants remained elusive. This project was initiated in order to investigate three major thematic priorities:

1. Mechanism of GRSF1-RNA interaction: To achieve this aim we first expressed wild-type human and murine GRSF1 as well as modified GRSF1 variants as recombinant proteins in E. coli and tested their *in vitro* RNA-binding activities. For this purpose, we established quantitative electrophoretic mobility RNA gel shift assays and quantified the binding constants of the recombinant protein constructs. In addition, we modified the putative RNA-binding sequence of human and murine GRSF1 and investigated their secondary structure by circular dichroism spectroscopy.

2. Evolutionary aspects of GRSF1: To achieve this aim we first screened publically available sequence databases for GRSF1-like sequences in viruses and living organisms of different levels of evolution and calculated the frequency of occurrence of GRSF1-related proteins in viruses, bacteria, acrchaea, fungi, lower and higher plants as well as in non-mammalian (fish, amphibia, reptiles, birds) and mammalian vertebrates and in mammals. Specifically we looked into selected model organisms, which represent living beings of different evolutionary levels, such as *E. coli* (bacteria), *S. cerevisiae* (fungi), *D. melanogaster* (insects), *C. elegans* (worms), *A. thaliana* (higher plants), *D. rerio* (bony fish), *X. tropicalis* (amphibia), *G. gallus* (birds), *H. neanderthalensis* (extinct primates) and *H. denisovan* (extinct primates).

3. Genetic multiplicity of human GRSF1 and functional consequences of naturally occurring mutations: To achieve this aim we first modeled the 3D-structure of the three RNA-binding domains of human GRSF1 on the basis of the NMR-structure of related RNA-binding proteins. Next, we searched a number of online genomic databases for naturally occurring human GRSF1 mutants and selected non-synonymous sequence alterations in the RNA-binding domains. These GRSF1 variants were expressed as recombinant proteins and their RNA-binding affinities were quantified by electrophoretic mobility RNA gel shift assays. Finally, the impact of these naturally occurring point mutations on global structure of GRSF1 was studied *in vitro* by monitoring the change in melting temperature (T_m) using thermal shift assays.

2. MATERIALS AND METHODS

2.1. MATERIALS

2.1.1. Laboratory equipment

Equipment	Manufacturer
Laminar Flow Cabinet	Steril S.p.a, Leipzig, Germany
Incubator	Cotech, Berlin, Germany
Water Bath	Haake, Karlsruhe, Germany
Vortex mixer	Janke & Kunkel GmbH, Staufen, Germany
Centrifuge 5417R	Eppendorf, Hamburg, Germany
Centrifuge 5804	Eppendorf, Hamburg, Germany
Centrifuge Sorvall TC; Rotor H400	Sorvall, Bad Homburg, Germany
Centrifuge Sorvall RC28S; Rotor GS3;	Sorvall, Bad Homburg, Germany
Dual-Action Shaker KL-2	Edmund Bühler GmbH, Hechingen, Germany
BioPhotometer	Eppendorf, Hamburg, Germany
Ultrasonicator	G. Heinemann, Schwabisch Gmund, Germany
Rotor-Gene RG-3000	Corbett Research, Sydney, Australia
Mighty Small Mini Vertical	Amersham Biosciences, Freiburg, Germany
Model S2 Electrophoresis Unit	Biometra, Göttingen, Germany
Blot-Chamber, Fast blot B44	Biometra, Göttingen, Germany
Incubator	Cotech, Berlin, Germany
Gel-Imager	Biometra, Göttingen, Germany
Thermomixer	Eppendorf, Hamburg, Germany
T3 Thermocycler	Biometra, Göttingen, Germany
Incubator T6120	Heraeus, Hanau, Germany
UV Transilluminator Ti5	Biometra, Göttingen, Germany
Power supply	Biometra, Göttingen, Germany
MilliQUF Plus	MilliPore, Bedford, USA
UV Crosslinker BLX-254	Vilber Lourmat, Lyon,France
UNO-Thermoblock	Biometra, Göttingen, Germany
Image Analyzer LAS-1000 CH	Fuji film, Tokyo, Japan
Spectropolarimeter J-720	JASCO, Gross-Umstadt, Germany
Spin-X® UF Concentrators	Corning, England, UK
Rotor-Gene RG-3000 real-time PCR	Corbett Research, Qiagen, Hilden, Germany
ÄKTA FPLC instrument	GE healthcare, Uppsala, Sweden

2.1.2. Chemicals

Chemicals	**Manufacturer**
Agarose	Promega, Mannheim, Germany
Acylamid/Bisacrylamid	Roth, Karlsruhe, Germany
Ammonium persulfate	Serva, Heidelberg, Germany
Acetic acid	Roth, Karlsruhe, Germany
Anti-Digoxigenin-AP antibody	Roche Diagnostics, Manheim, Germany
Anti-GST-Peroxidase	Sigma, Steinheim, Germany
ATP	Roche Diagnostics, Mannheim, Germany
Ampicillin	Roth, Karlsruhe, Germany
Boric acid	Roth, Karlsruhe, Germany
Bromophenol blue	Merck, Darmstadt, Germany
Chloramphenicol	Roth, Karlsruhe, Germany
EDTA	Sigma, Steinheim, Germany
Ethanol	Roth, Karlsruhe, Germany
Glucose	Roth, Karlsruhe, Germany
Glycine	Sigma, steinheim, Germany
Glutathione-Agarose	Machery-Nagel, Düren, Germany
Hydrochloric acid	Roth, Karlsruhe, Germany
Imidazole	Serva, Heidelberg, Germany
IPTG, Tween 20	Roth, Karlsruhe, Germany
Isopropanol (2-Propanol)	Merck, Darmstadt, Germany
Kanamycin	Sigma-Aldrich, Steinheim, Germany
Kanamycin sulfate	Roth, Karlsruhe, Germany
Methanol	VWR, Leuven, Netherlands
Maleic acid (1 M, pH 7.5)	Roche Diagnostics, Mannheim, Germany
Nickel-Agarose	Macherey-Nagel, Düren, Germany
N-nitroso-N-methylurea	Applichem, Darmstadt, Germany
Ponceau S	Sigma, St. Louis, USA
Sodium chloride	Roth, Karlsruhe, Germany
Sodium hydroxide	Roth, Karlsruhe, Germany
Sypro® Orange	Life Technologies, Carlsbad, USA
TCEP	Sigma, Dreisenhofen, Germany
TEMED	Serva, Heidelberg, Germany
Tris-Solution (1 M, pH 8.0)	Applichem, Darmstadt, Germany

2.1.3. Enzymes

Enzyme	Manufacturer
T4 DNA ligase	Roche Diagnostics, Mannheim, Germany
Turbo DNAse™	Ambion, Vilnius, Lithuania
T4 Polynucleotide Kinase	Promega, Mannheim, Germany
NcoI	Thermoscientific, Darmstadt, Germany
NotI	Thermoscientific, Darmstadt, Germany
SalI	Thermoscientific, Darmstadt, Germany
EcoRI	Thermoscientific, Darmstadt, Germany
HindIII	Thermoscientific, Darmstadt, Germany
SacI	Thermoscientific, Darmstadt, Germany
KpnI	Thermoscientific, Darmstadt, Germany
BamHI	Thermoscientific, Darmstadt, Germany
XhoI	Thermoscientific, Darmstadt, Germany
DpnI	Thermoscientific, Darmstadt, Germany
Alkaline-Phosphatase	Roche Diagnostics, Mannheim, Germany
Horseradish-Peroxidase	Sigma, Steinheim, Germany
PfuTurbo DNA Polymerase	Stratagene, La Jolla, USA

2.1.4. Buffers and media

The culture media for growing *E.coli* were purchased from following manufacturers.

Medium	Manufacturer
LB-Medium	Roth, Karlsruhe, Germany
LB-Agar	Roth, Karlsruhe, Germany
EnPresso® B tablet set	Biosilta, Berlin, Germany

LB-Medium

10 g/L Tryptone; 5 g/L Yeast extract; 5 g/L NaCl; pH 7.0

LB-Agar

10 g/L Tryptone; 5 g/L Yeast extract; 5 g/L NaCl; 15 g/L Agar-Agar; pH 7.0

SOC-Medium

20 g/L Tryptone; 5 g/L Yeast extract; 10 mM NaCl; 2.5 mM KCl; 10 mM $MgCl_2$; 10 mM $MgSO_4$; 20 mM Glucose

2.1.5. Plasmids and bacterial strains

The plasmids and bacterial strains were purchased from following manufacturers.

Plasmid	Manufacturer
pET-42a (+) Expression vector	Merck, Darmstadt, Germany
pET28b (+) Expression vector	Merck, Darmstadt, Germany
pCR2.1-TOPO® Cloning vector	Merck, Darmstadt, Germany
pGEX-4T-3 Expression vector	Amersham Biosciences, Freiburg, Germany

Bacterial strains	Manufacturer
XL1-Blue	Stratagene, La Jolla, USA
BL21 (DE3)	Life Technologies, Carlsbad, USA
BL21 (DE3)pLysS	Life Technologies, Carlsbad, USA
Rosetta (DE3)pLysS	Life Technologies, Carlsbad, USA

2.1.6. Commercial kits

Kit	Manufacturer
Advantage® 2 Polymerase Kit	Clontech, Palo Alto, USA
TOPO® TA Cloning Kit	Invitrogen, Karlsruhe, Germany
NucleoSpin® Gel and PCR Clean-up Kit	Macherey-Nagel, Düren, Germany
GeneJET Plasmid Miniprep Kit	Thermoscientific, Dreieich, Germany
LigaFast™ Rapid DNA Ligation System	Promega, Madison, USA
QuickChange® Site-Directed Mutagenesis Kit	Stratagene, La, Jolla, USA
NucleoBond® XtraMidiPlus EF	Macherey-Nagel, Düren, Germany
NucleoSpin® Plasmid EasyPure	Macherey-Nagel, Düren, Germany
MEGAshortscript™ T7 Transcription Kit	Ambion, Huntingdon, UK
Western Lightning® Chemilumniscence Reagent	Perkin Elmer, Boston, USA

2.1.7. Software

Software	Manufacturer
Image J	RSB; https://imagej.nih.gov/
MODELLER®9.14.	Accelrys Inc., San Diego, USA
PyMol v1.5	Schrondinger, LLC, New York, USA, 2012
Discovery Studio® Visualizer	BIOVIA, San Diego, USA
Inkscape	Free software foundation, Inc., Boston, USA
Adobe illustrator CS6	Adobe Systems Software Ireland Ltd.

2.2. METHODS

2.2.1. Preparative methods

2.2.1.1. Transformation of bacteria with plasmid DNA

Transformation of bacteria is a method for introducing foreign DNA into bacterial cells. Here the bacterial expression plasmids containing the coding region of the *GRSF1* gene (region 4q13.3) were transformed into different competent *E. coli* bacterial strains [e.g. BL21 DE3, BL21 (DE3) pLysS, Rossetta (DE3) pLysS, XL1-Blue], which were made competent by chemical treatment. Rossetta (DE3) strains were used for high level expression of eukaryotic proteins. These strains produce tRNAs for codons that are normally rarely used in *E. coli* allowing universal translation. BL21 (DE3) pLysS can also be used for high-level protein expression. These strains contain a T7 promoter, which reduces the basal level of expression of recombinant proteins in the absence of IPTG. XL1-Blue strains were used for routine cloning. The plasmid DNA (100 ng/100 µl competent cells) was incubated with competent bacterial cells on ice for about 20-30 minutes. After a heat shock at 42^{o}C for 45 sec, the cells were cooled down on ice for 2-3 minutes to allow plasmid DNA to enter the bacterial cells. Then, the cells were incubated in SOC medium at 37^{o}C for 1 hour under shaking. To isolate the bacteria harboring the recombinant plasmid from bacteria that do not contain it, transformed bacteria were plated on agar plates with different antibiotics. This allowed selection of only those bacteria that contain antibiotic resistance genes included in the recombinant plasmids. For BL21 DE3, BL21 (DE3) pLysS and XL1-Blue *E. coli* strains 50 µg/ml kanamycin were used. Similarly, for Rossetta (DE3) both 50 µg/ml kanamycin, and 35 µg/ml chloramphenicol were employed as selection marker. The agar plates were incubated overnight at 37^{o}C.

2.2.1.2. Cloning and ligation

Molecular cloning is a process aimed at isolating and amplifying the complementary DNA (cDNA) for a certain gene product, which can subsequently be introduced in a plasmid to be transformed into bacteria. It is a multi-step strategy that may involve different methodological approaches (Sambrook, et al., 1989). In this study the cDNA sequences encoding (full-length, ΔE1-hGRSF1, single and double GRSF1 truncation constructs) of human GRSF1 protein were amplified from an RNA extract of human embryonic kidney 293 (HEK293) and separately cloned using *GRSF1* reference cDNA sequence NM_002092.3 as a template. First, the DNA sequences with construct-specific amplification primers (see **Table 2.1**) were amplified in PCR using Advantage® 2 Polymerase Kit (Clontech, Palo Alto, USA) according to manufactures instructions. The PCR products were then cloned into the specific restriction sites (see **Table 2.1**) of pET-

42a expression vector (Merck, Darmstadt, Germany), which contains an N-terminal GST or His tag. The recombinant plasmid was later transformed into XL1-Blue competent cells (Stratagene, La Jolla, USA) and the plasmid DNA was extracted with NucleoSpin® Plasmid EasyPure (Macherey-Nagel, Düren, Germany) according to manufactures instructions. The plasmid DNA was quantitatively digested using construct specific restriction enzymes (see **Table 2.1**) and later analyzed on 1% agarose gel. Afterwards, the different *hGRSF1* inserts including; full-length (1460 bp), ΔE1-*hGRSF1* (1164 bp), ΔE1ΔR1 (915 bp), ΔE1ΔR2 (924 bp) and ΔE1ΔR3 (930 bp) were precisely excised from pET-42a plasmid and later purified using NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel, Düren, Germany). The DNA concentration was determined using Biophotometer (Eppendorf, Hamburg, Germany). The required amount of insert DNA (in ng) needed for ligation was calculated by using the following equation.

$$\frac{\text{ng of vector X bp size of insert}}{\text{bp size of vector}} \times \text{molar ratio of } \frac{\text{insert}}{\text{vector}} = \text{ng of insert}$$

Table 2.1. Primers used to amplify full-length, double-deletion and isolated qRRM constructs of *hGRSF1* in pET-42a and 2.1-Topo® vectors. Oligonucleotides were synthesized by Biotez, Berlin, Germany

Oligonucleotide	**Sequence 5'…..3'**
BamHI-*hGRSF1*1-up	CCC CGG ATC CAT TGG GCA CGG GAA CAA GGG AC
NotI–*hGRSF1*-do	CCC CGC GGC CGC TTA TTT TCC TTT AGG ACA TGA ATT TAG G
EcoRIΔE1-*hGRSF1*-up	GGG TCC ATG GAG GCC GAA TTC ATG GAG TCC AAA ACT ACT TAC CTG
HindIIIΔE1-*hGRSF1*-do	CAG GTA AGT AGT TTT GGA CTC CAT GAA TTC GGC CTC CAT GGA CCC
EcoRI-*hGRSF1*-ΔE1ΔR1-up	GGA AGT GGA TGA TGT CTT TCT C
HindIII-*hGRSF1*-ΔE1ΔR1-do	TGA CCT GCA AGC TCT TCA TTA AGA GAA AGA CAT CAT CCA CTT CC
EcoRI-*hGRSF1*- ΔE1ΔR2-up	GCC TGT GGT AAA TGA TGG TGT GCA TGT CGG TTC TTA TAA GGG AA
HindIII-*hGRSF1*-ΔE1ΔR2-up	TTC CCT TAT AAG AAC CGA CAT GCA CAC CAT CAT TTA CCA CAG GC
EcoRI-*hGRSF1*- ΔE1ΔR3-up	AAC TAC GTC TTC TCT GCA TTT TCA GGA TCC GTC GAC AAG CTT GC
HindIII-*hGRSF1*-ΔE1ΔR3-up	GCA AGC TTG TCG ACG GAT CCT GAA AAT GCA GAG AAG ACG TAG TT
BamHI-qRRM1-up	GGA TCC CCG TCC AAG TTA GAA GAG GA
XhoI-qRRM1-do	CTC GAG TTA AGG CGA AGA TTT GAC CTG CA
BamHI-qRRM2-up	GGA TCC CCT GTG GTA AAT GAT GGT GT
XhoI-qRRM1-do	CTC GAG TTA ACC GAC ATG TGT GTT CGA ACT T
BamHI-qRRM3-up	GGATCC CTG CAT TTT GTC CAC ATG AG
XhoI-qRRM3-do	CTC GAG CTG GAG CCC CTA GAG TCT TTA

By using the above equation suitable amount of vector and insert (in µl) was prepared by making a standard dilution. Furthermore, the purified DNA inserts were ligated into the pET42a expression plasmid using LigaFast Rapid DNA Ligation System (Promega, Madison, USA) according to manufactures instructions. The sequences of recombinant plasmids were verified by DNA sequencing and transformed for expression into Rosetta (DE3) pLysS strains (Life Technologies, Carlsbad, USA). 1 kb DNA ladder (New England Biolabs, Schwalbach, Germany) was used as molecular weight standard.

The DNA sequences encoding qRRM1 (amino acids 139-244), qRRM2 (amino acids 252-323) and qRRM3 (amino acids 400-480) were amplified from an RNA extract of HEK293 cells and cloned from cDNA of *hGRSF1* Nm_002092.3 as a template. The coding sequences of isolated qRRMs were amplified in PCR using the oligonucleotides (see **Table 2.1**) with Advantage® 2 Polymerase Kit Clontech (Clontech, Palo Alto, USA) according to manufactures instructions. The PCR products were cloned into the XhoI/BamHI site of the 2.1-TOPO® cloning vector (Merck, Darmstadt, Germany) using TOPO TA Cloning Kit (Invitrogen, Karlsruhe, Germany) according to the manufactures instructions. The recombinant plasmid was expressed in XL1-Blue competent cells and the plasmid DNA was extracted with NucleoSpin® Plasmid EasyPure (Macherey-Nagel, Düren, Germany) according to manufactures instructions. The plasmid was digested with XhoI and BamHI endonucleases and subsequently analyzed on 1% agarose gel. The cloned restriction fragments of size qRRM1 (318 bp), qRRM2 (216 bp) and qRRM3 (243 bp) were correctly cut from 2.1-TOPO® plasmid and purified using *NucleoSpin® Gel and PCR Clean-up kit (*Macherey-Nagel, Düren, Germany). Finally, the plasmid sequences were verified by DNA sequencing (Eurofins MWG Operon, Ebersberg, Germany). The purified qRRM inserts from each of the three qRRM domains were ligated to 2.1-TOPO® expression plasmid and expressed in BL21 (DE3) strains (Life Technologies, Carlsbad, USA) competent cells as a fusion protein with an N-terminal GST-tag. The molecular weight of DNA was measured with 1 kb DNA ladder (New England Biolabs, Schwalbach, Germany).

2.2.1.3. Site-directed mutagenesis

For the production of hGRSF1 point mutants, mutagenesis experiments were carried out using the QuickChange® Site-Directed Mutagenesis Stratagene (La Jolla, USA) following manufacturer's instructions. For this purpose a mutagenesis primers are designed carrying the corresponding amino acid exchanges. Plasmids containing a mutated and an unmutated strand are produced and in each of these plasmids the parental methylated DNA strand, which does not carry the mutation, is digested with the

restriction enzyme DpnI at 37ºC for 60 minutes. The mutated unmethylated plasmid DNA strand that is left over is doubled and the mutated double stranded plasmid is then transformed into chemically competent *E. coli* XL1-Blue cells.

2.2.1.4. Production and purification of recombinant GST-tagged proteins

The following materials are used for production and purification of recombinant GST-tagged recombinant GRSF1 constructs:

- LB medium (Roth®, Karlsruhe, Deutschland)
- LB Agar (Roth®, Karlsruhe, Deutschland)
- Antibiotics: 50 µg/ml kanamycin; 35 µg/ml chloramphenicol
- 1-fold PBS buffer: 140 mM NaCl; 2.7 mM KCl; 10 mM Na_2HPO_4; 1.8 mM KH_2PO_4; pH 7.3
- Wash buffer: 140 mM NaCl; 2.7 mM KCl; 10 mM Na_2HPO_4; 1.8 mM KH_2PO_4; pH 7.3
- Elution buffer: 50 mM Tris-HCl, pH 9.5; 10 mM L-glutathione reduced

For the expression of GST-fusion proteins, the recombinant plasmid DNA was transformed into chemically competent *E. coli* BL21 (DE3) bacterial strains. Expression of the GST-fusion proteins was carried out using EnPresso® B (Biosilta, Berlin, Germany) expression system according to the manufacturer's instructions. To start the pre-culture, 5-10 *E. coli* clones were picked from an agarose plate using sterile toothpicks and grown individually in 1 ml LB-medium containing 50 µg/ml kanamycin at 37ºC for 7-8 hours in shaker. Prior to main culture the optical density (OD) of bacterial pre-cultures was measured to be 0.150 using Eppendorf BioPhotometer®. The main culture was started by inoculating 50 ml of freshly prepared culture medium with one of the pre-cultures and this mixture was incubated overnight (15-18 h) at 30ºC at 250 r.p.m in shaker until the culture reached an OD_{600} of greater than 5. Then expression was induced by adding 1 mM isopropyl-*β-D*-thiogalacto-pyranoside (IPTG) to the bacterial main culture. The culture was grown overnight (15-18 h) at room temperature at 250 r.p.m in shaker until it reached an OD_{600} of 10-20. IPTG is an artificial inducer of recombinant protein expression that binds to Lac repressor and prevents it from binding to main operator (O_1) of the *lac* operon thereby allowing T7 RNA polymerase to transcribe T7 promoter controlled gene. After that bacteria were harvested by centrifuging at 4.000 r.p.m and 4ºC for 15 min. Cell pellets were resuspended in ice cold PBS buffer and sonicated twice (3 times; 10 sec; 20% maximal intensity) on ice using a Branson tip-sonifier (Heinemann, Schwabisch Gmund, Germany). For affinity chromatographic purification of the recombinant proteins the cell lysate was centrifuged at (20.000 x g at 4ºC for 20 min) and the supernatant containing soluble proteins were incubated with glutathione-coupled agarose at 4ºC for 1 hour with gentle

agitation. The GST-tagged proteins bind to the glutathione-agarose and unbound proteins were washed away four times with PBS washing buffer. Bound GST-tagged proteins were competitively eluted from the agarose beads with elution buffer containing glutathione. The elution fractions were combined and concentrated to the final concentration of 0.5-1 mg/ml using a protein concentrator (Corning, England, UK) with a cutoff limit of 5 kDa. The purified protein was either snap frozen in liquid nitrogen or mixed with 10% v/v glycerol, frozen and then stored at - 80^{o}C.

2.2.1.5. Production and purification of recombinant His-tagged proteins

The following materials are used for production and purification of recombinant His-tagged recombinant GRSF1 constructs:

- Wash buffer: 0.1 M Tris-HCl (pH 8.0); 300 mM NaCl;
- Wash buffer 1: 5 mM Tris-HCl; 15.8 mM NaCl; 10.52 mM imidazole
- Wash buffer 2: 14 mM Tris-HCl; 42.85 mM NaCl; 28.57 mM imidazole
- Elution buffer: 0.1 M Tris-HCl (pH 8.0); 300 mM NaCl; 200 mM imidazole

The proteins containing an N-terminal hexa-histidine tag were expressed using Enpresso® B growth system according to manufacturer's instructions (see **Section 2.2.1.4**). After induction of expression of the recombinant proteins with IPTG bacteria were harvested by centrifugation. Cell pellets were resuspended in ice cold PBS and centrifuged again at 4 000 r.p.m at 4^{o}C for 30 min. The suspended cells were sonicated twice (3 times; 10 sec; 20% maximal intensity) on ice using a Branson Digital Sonifier (G. Heinemann, Schwabisch Gmund, Germany). The cell lysate was cleared by centrifuging at 20.000 x g at 4^{o}C for 20 min. The supernatant containing soluble proteins was incubated with Ni-NTA agarose at 4^{o}C for 2 hours under agitation. The His-tagged proteins bind to the Ni^{2+}-ions and unbound proteins were washed off four times with wash buffer 1 and wash buffer 2, respectively. Bound hexahistidine-tagged proteins were competitively eluted with elution buffer containing 200 mM imidazole. Elution fractions were combined and concentrated to a final concentration of 0.5-1 mg/ml using protein concentrators with a cutoff limit of 5 kDa. The purified protein was either snap frozen in liquid nitrogen or mixed with 10% v/v glycerol, frozen and then stored at - 80^{o}C.

2.2.2. Analytical techniques

2.2.2.1. SDS–PAGE

The following materials are used for SDS-PAGE:

- 4% stacking gel: 1.3 ml Rotiphorese® NF-acrylamide/bis-acrylamide solution 30% (29:1) (Bio-Rad); 2.5 ml stacking buffer; 6.1 ml water; 10 µl TEMED; 100 µl 10% (w/v) ammonium persulfate.

- 10% resolving gel: 3.3 ml Rotiphorese® NF-Acrylamide/Bis-acrylamide solution 30% (29:1) (Bio-Rad); 2.5 ml resolving buffer; 4.1 ml water; 10 µl TEMED; 100 µl 10% (w/v) ammonium persulfate.
- stacking gel buffer: 0.5 M Tris; 0.4% (w/v) SDS; pH 6.7
- resolving gel buffer: 1.5 M Tris; 0.6% (w/v) SDS; pH 8.8
- 5-fold electrophoresis running buffer: 125 mM Tris; 100 mM glycine; 17 mM SDS
- Coomassie-Solution: 0.2 % (w/v) Coomassie brilliant blue; 10% (v/v) methanol; 20% (v/v) acetic acid.
- destaining solution: 40% (v/v) Methanol; 10% (v/v) acetic acid.

SDS-PAGE is the most widely used analytical method to separate proteins according to their molecular weight (Laemmli et al., 1970). SDS-PAGE gels contain a resolving gel with a small stacking gel above it. The protein extracts were mixed with loading buffer (4-fold Roti-Load 1; Roth, Karlsruhe, Germany) in a total volume of 14 µl and denatured at 95^{o}C for 10 min. The electrophoresis was carried out by running the gels (8x9 cm) in electrophoresis running buffer at a constant voltage of 150 V at room temperature until the staining front has reached the bottom of the gel. Next, the protein bands were visualized with Coomassie-solution and finally, the gels were destained with destaining fluid to allow visualization of protein bands. Precision Plus Unstained Protein™ Standards (Bio-Rad) was used as a standard molecular weight markers.

2.2.2.2. Immunoblot analysis

The following materials are used for Immunoblot analyses:

- Anode buffer I: 0.3 M Tris (pH 10.4); 20% (v/v) methanol
- Anode buffer II: 0.025 M Tris (pH 10.4); 20% (v/v) methanol
- Cathode buffer: 0.025 M Tris (pH 9.4); 40mM *ε-Aminocaproic acid*
- Anti-GST antibody: 10 ml 1X blocking buffer; 1:10000 anti-GST-Peroxidase antibody Sigma®, (Steinheim, Germany).
- 5-fold PBS (final volume 1 Liter): 7.8g Na_2HPO_4•$2H_2O$; 0.815 g KH_2PO_4; 43.83 g NaCl; pH 7.4
- PBS/Tween20: PBS; 0.1% (v/v) Tween 20

Immunoblotting is aimed at detecting proteins that have been separated by gel electrophoresis using a specific antibody. For this purpose the separated protein bands was transferred onto a nitrocellulose membrane (Amersham™ Protran™, Buckinghamshire, UK) by a semi-dry blotting method (constant voltage of 10 V) at room temperature for 1 hour by preparing gel sandwich. For visualization of protein bands the membrane was stained with Ponceau Red Sigma® (Sigma, St. Louis, USA) and incubated

at room temperature for 2 min under gentle agitation. Next, the membrane was rinsed with water for 3 min with shaking to wash off the excess dye. The membrane was then transferred into 5% (w/v) non-fat dry milk in blocking buffer for 30 min under agitation to prevent non-specific binding of the antibody to the membrane. Next, the blot was transferred to blocking buffer containing an appropriate dilution (1:10000) of anti-GST antibody coupled to peroxidase and was incubated for 1 hour under gentle agitation. This was followed by washing the membrane thrice with washing buffer (PBS/0.1% (v/v) Tween-20) for 5 min to wash off the unbound antibodies. Then, the blot was developed by adding Western Lightning® Chemiluminescence Reagent Plus (PerkinElmer, Boston, USA) onto the membrane at room temperature for 1 min with shaking following the manufacturer's instructions. Finally, to quantify the band intensities the membrane was exposed in a Luminescence Image Analyzer LAS-1000 CH (Fuji film, Tokyo, Japan). Precision Plus Prestained Protein™ Standards (Bio-Rad) were used as molecular weight markers.

2.2.2.3. Protein quantification

The Bradford assay (Bradford et al., 1976) is a method used for protein determination that employs Coomassie Brilliant Blue G-250. This dye unspecifically binds to all proteins. The protein determination of bacterial lysate was carried out using pre-composed Bradford solution (AppliChem, Darmstadt, Germany). The calibration of the BioPhotometer was carried out using a serum albumin standard solution of known concentration and the linear part of the calibration curve (0.5 – 10 mg/mL) was used to calibrate the readout scale. The elution fractions were mixed with Bradford dye and protein concentration was determined by measuring the absorbance of protein-dye complex at 595 nm using an Eppendorf BioPhotometer® (Eppendorf AG, Hamburg, Germany). Blank measurements were carried out by mixing distilled water and Bradford dye in a final volume of 1 ml in a 10 mm cuvette.

2.2.2.4. Size-exclusion chromatography

In order to further purify the recombinant GRSF1 fusion proteins, we performed size-exclusion chromatography on ÄKTA FPLC instrument (GE healthcare, Uppsala, Sweden). This method is used to separate and characterize proteins on the basis of molecular mass (Striegel, 2016). For this purpose 0.5 ml of affinity purified and concentrated GRSF1 fusion proteins were loaded onto a Superdex™ 75 10/300 GL column (GE healthcare, Uppsala, Sweden). The column was eluted with elution buffer (20 mM Tris-HCl; 100 mM NaCl) at a flow rate of 0.5 ml/min. Fractions of 0.5 ml were collected and the elution profile was recorded measuring the absorbance at 280 nm.

2.2.2.5. *In vitro* transcription

The following materials are used for in vitro transcription:

- TEN buffer: 110 mM Tris-HCl (pH 8.0), 1 mM EDTA, 1.25 M NaCl

In vitro transcription refers to the *in vitro* synthesis of an RNA transcripts from a linear DNA template containing the T7 promoter, which is located upstream the sequence to be transcribed. To synthesize RNA probes for RNA gel mobility shift assays (see **Section 2.2.2.8**), it was necessary to prepare appropriate DNA templates. These DNA templates were prepared following the protocol described before by Milligan et al. (Milligan et al., 1987). For this purpose, two complementary DNA oligonucleotides were incubated in TEN-buffer at 95ºC for 10 min in a water bath. Then the temperature was slowly lowered to room temperature and now the oligonucleotides anneal to make a double-stranded template. One single-stranded DNA oligonucleotide coded the bacteriophage T7 promoter. The second single-stranded DNA oligonucleotide involved the 5'-untranslated region of the *hGPx4* isoform and at its 3'-end the bacteriophage T7 promoter in reverse complementary orientation. The *in vitro* transcription was performed using the MEGAshortscript™ T7 Transcription kit (Ambion, Huntingdon, UK) following manufacturer's instructions. To obtain non-radioactive labelled RNA probes, 7.5 mM UTP; 1 mM DIG-UTP was added to transcription reaction to directly incorporate DIG-UTP into RNA transcript.

2.2.2.6. Purification of RNA probes

To remove nucleotides, proteins, and salts from RNA preparations spin column chromatography was used. For this purpose, Micro Bio-Spin 30 columns that contained a Bio-Gel® P polyacrylamide (P-30) matrix were used (Bio-Rad, California, USA) to separate RNAs according to their molecular weight from the other components of the synthesis mixture. Molecules smaller than the exclusion limit of the column are retained by the column. Spin columns were used following manufacturer's instructions.

2.2.2.7. Urea gel electrophoresis

The following materials are used for urea gel electrophoresis:

- 8 M Urea-polyacrylamide gel: 3.6 g urea; 0.75 ml 10-fold TBE; 0.95 ml Rotiphorese® NF-acrylamide/bis-acrylamide solution 40% (29:1); 15 ml DEPC-water ; 8 µl TEMED; 60 µl 10% (w/v) APS).
- DEPC-water: MilliQ-water; 0.05% (v/v) DEPC.
- TBE running buffer: 0.01 M Tris-base; 0.01 M boric acid; 0.097 µg/mL EDTA.

Denaturing urea polyacrylamide gel electrophoresis is a method for analyzing the integrity of *in vitro* transcribed RNAs (Summer H et al., 2009). RNAs mixed with gel loading buffer were denatured at 80^{o}C for 3 min. The electrophoresis was carried out by running pre-equilibrated urea gels (8x9 cm) in TBE running buffer at a constant voltage of 150 V at room temperature for 25 min. Next, the gels were stained with ethidium bromide, a fluorescent dye that interacts with nitrogenous bases of nucleic acids. Finally, the gels were exposed under a UV Transilluminator to visualize RNA. The RNA Century™ Marker mix (Ambion, Huntingdon, UK) was used as a standard to measure the molecular weight of RNA.

2.2.2.8. RNA electrophoretic mobility shift assays (REMSA)

The following materials are used for RNA electrophoretic mobility shift assays:

- 10-fold Maleic acid buffer: 1 M maleic acid; 1.5 M NaCl; pH 7.5
- 1-fold washing buffer: 10-fold maleic acid buffer; 0.3% (v/v) Tween 20
- 10-fold blocking buffer: maleic acid buffer; 10% (w/v) blocking reagent (Roche Diagnostics, Mannheim, Germany).
- 1-fold blocking buffer: 10-fold maleic acid buffer; 10% (v/v) 10-fold Blocking buffer (Roche Diagnostics, Mannheim, Germany).
- Anti-digoxigenin antibody: blocking buffer; 0.075 u/ml anti-digoxigenin antibody (alkaline phosphatase) (Roche Diagnostics, Mannheim, Germany).
- 1-fold detection buffer: 0.1 M Tris-HCl; 0.1 M NaCl; pH 9.5 (Roche Diagnostics, Mannheim, Germany).
- CSPD-Working Solution: detection buffer; 1% (v/v) CSPD (Roche Diagnostics, Mannheim, Germany).
- Binding buffer: 10 mM HEPES; 25 mM KCl; 1.563 mM EDTA; 40% Glycerol; 0.25 mM DTT.
- 5% native acrylamide gel: 2.5 ml acrylamide:bis-acrylamide solution 30% (37.5:1); 1.125 ml acrylamide 40%; 1.2 ml 10-fold TBE, 19.18 ml Nuclease free water, 20 µl TEMED, 240 µl 10% (w/v) APS).

RNA-protein interactions can be detected by RNA gel mobility shift assays. This method is based on the fact that RNA-protein complexes migrate slower through a native gel compared to free RNA since the molecular weight of the complex is higher than that of the free RNA (Fillebeen C, et al., 2014). Thus, for RNA/protein binding studies different amounts of protein were incubated with 0.5-1 pmol of DIG-labelled RNA probes at 30^{o}C for 20 min in binding buffer containing 1.33 µg/mL Heparin, 16.7 ng/mL yeast tRNA and 1.5 mM each of ATP, and GTP in a reaction volume of 15 µl on ice. DIG-labelled RNA was

denatured at 80°C for 3 min. Next, the RNA was allowed to renature on ice for 2 min. Then, renatured RNA was added to the reaction mixture. After that reaction mixture was further incubated at 30°C for 20 min. The reaction mixtures were loaded onto a pre-run native 5% polyacrylamide gel. The gels (8x9 cm) were run in TBE running buffer at a constant voltage of 150 V on ice for 40-44 min.

After electrophoresis gels were placed on positively charged nylon membrane and blotted for 40 min at room temperature. Next, the gels containing separated RNA probes and RNA-protein complexes were transferred to a blotting membrane (Tropilon Plus, Life Technologies, Carlsbad, USA) which was uniformly equilibrated before in 2-fold SSC buffer. Then, a shortwave UV light of 254 nm was applied to the membrane to crosslink the RNAs to the membrane and exposed for 0.120 mJ/cm^2. To wash away the unbound RNA-protein complexes the membrane was washed in washing buffer for 5 minutes under agitation. The membrane was then transferred to blocking buffer for 30 min under agitation to prevent non-specific binding of the antibody. To detect DIG-labelled RNA, the blots were transferred to blocking buffer for 30 min containing anti-digoxigenin antibody coupled to alkaline phosphatase. This was followed by washing the membrane twice with washing buffer for 15 min to wash away unbound antibody. Next, the blots were equilibrated in detection buffer for 5 minutes and then developed by adding 0.25 mM CSPD to the membrane at room temperature for 5 min and incubated at 37°C for 10 min to enhance the luminescent reaction. Finally the membrane was exposed in a Luminescence Imager Analyzer LAS-1000 CH (Fuji film,Tokyo, Japan).

The dissociation constants (K_D) were determined as previously described (Nieradka et al., 2014; Ufer et al., 2008). For this purpose, the intensities of the chemiluminescence of the free RNA band and the RNA-protein complex band were quantified using the Image J software (T. Ferreira, 2012) and the logarithmic values of the ratio of shifted versus free RNA were plotted as function of the logarithmic values of the molar concentration of recombinant protein present in the binding assay.

2.2.2.9. Circular dichroism (CD) spectroscopy

CD spectroscopy is a biophysical technique, which is used for determining the secondary structure or conformation of macromolecules. Chiral molecules absorb left- or right-circularly polarized light differently and the difference in absorption is measured as a function of wavelength. RNA stocks were prepared at 20 mM in Tris-HCl, pH 7.5, in the presence or absence of 100 mM of KCl or NaCl. CD experiments were performed at 10° C using a JASCO J-715 Spectropolarimeter (JASCO, Gross-Umstadt, Germany). RNA samples were heated to 70°C for 5 min and then slowly cooled down to 10°C at a rate of

0.01 K/s. RNAs were then diluted to 5 mM and CD spectra were recorded in 200 ml quartz cuvette with a path length of 1 mm in the wavelength range between 220-320 nm (measuring response of 2 sec, data pitch of 0.1 nm, bandwidth of 1 nm, scanning speed of 100 nm/min). Data were accumulated from 10 CD scans. Measurements were repeated three times independently.

2.2.2.10. Thermal shift assay

Thermal shift assay is a temperature-based assay to assess the stability of proteins by determining their melting temperatures. Purified hGRSF1 protein preparations were incubated with 200X fluorescent dye SYPRO® Orange dye (Life Technologies, Carlsbad, USA) in a reaction volume of 50 µl. The denaturation reaction was carried out in a Rotor-Gene RG-3000 real-time PCR machine (Corbett Research, Qiagen, Hilden, Germany). Samples were heated from 30^{o} to 95^{o}C in steps of 0.5^{o}C. The fluorescent dye binds to patches of hydrophobic amino acids in the protein core to minimize contact with water (See **Figure 2.1.** below). The increase in the temperature induces unfolding of the protein, which exposes hydrophobic core amino acids at the protein surface. The dye is able to bind to the surface exposed hydrophobic amino acids, which results in an increase in the protein fluorescence. The inflection point of the resulting thermal transition curve corresponds to the melting temperature (T_m) of the protein. T_m is determined by monitoring the increase in fluorescence depending on the temperature. The fluorescence decreases at higher temperatures (> 60–70^{o}C) since denatured proteins tend to aggregate, which reduces the binding of the dye. High T_m values suggest a stable protein structure that is resistant towards temperature induced denaturation. The fluorescent intensity was quantified using FAM/SYBR green filter. The melting temperature (T_m) determined for each GRFS1 mutant was representative of 3 independent experiments.

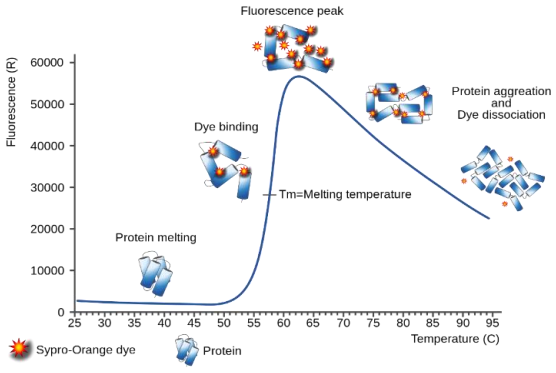

Figure 2.1. Principle of thermal stability assay include: (**A**) Protein denaturation and binding of dye to hydrophobic surfaces (**B**) Protein-dye complex fluoresces upon increase in temperature (**C**) Aggregation of denatured proteins and dissociation of dye at higher temperatures. The figure was taken from http://tinyurl.com/jcfr738.

2.2.3. *In silico* methods

2.2.3.1. *In silico* homology modeling

Homology protein modelling is a method developed by A. Sali and T. L. Blundell (A. Sali and T. L. Blundell, 1993). The method was intended to predict the most probable 3D structure for a protein in the absence of direct structural data. In order to perform homology modelling the 3D-structure of a related protein must be known (template) and the amino acid sequence homology between the template and the protein of interest should be as high as possible. Since, there is no experimentally determined 3D-structure available for GRSF1, the *in silico* structure of each qRRM domain of human GRSF1 was modeled using the MODELLER© software 9.14. (Accelrys Inc.,San Diego, USA) using the NMR structures of qRRM1 (PDB ID:2HGL), qRRM2 (PDB ID: 2HGM) and qRRM3 (PDB ID: 2HGN) of hnRNP F as a template. To improve the structural quality of predicted models a structural refinement method called $3D^{refine}$ (Bhattacharya et al., 2013) was used to optimize both the hydrogen bonding network as well as to minimize the atomic-level energy of the optimized model. On the basis of this overall model we extracted the 3D-structures of three quasi-domains (qRRMs) of hGRSF1 employing the PyMOL 1.3 visualization software (DeLano, 2002).

2.2.3.2. 3D visualization and structure analysis

The structural analysis of the *in silico* generated three qRRM domains were carried out using the PyMol visualization software 1.3 (Schrodinger, LLC, New York, USA, 2012). PyMol is a molecular graphic system for viewing homology models. (Seeliger D et al., 2010).

2.2.3.3. Statistic evaluations

The statistical analysis was carried out by using t-test function in Microsoft Excel. A t-test (Student, 1908) is a hypothesis test used to determine the statistical significance or non-significance between the two mean values in a set of data assuming a normal distribution of data in both groups. The test statistic was converted to a probability called the *p* value (Fisher, 1925) by t-test function. The *p* value was used as a parameter to measures the strength of evidence against the null hypothesis and observed statistic was ranked as; $p < 0.05$* (statistically significant), $p < 0.01$** (statistically moderately significant) and $p < 0.001$*** (statistically highly significant).

3. RESULTS

3.1. Recombinant expression and purification of full-length GRSF1 and its domains

3.1.1. Recombinant expression and purification of full-length human GRSF1 and its alanine-rich domain truncation mutant as N-terminal GST-fusion proteins

In order to elucidate the molecular mechanisms of GRSF1-RNA interactions we cloned different human *GRSF1* cDNAs (encoding for different GRSF1 constructs, such as wild-type full-length human GRSF1, truncated human GRSF1 species, and separate RNA-binding domains qRRM1-3), inserted the constructs into a bacterial expression vector and overexpressed the corresponding proteins in various *E. coli* strains. The recombinant proteins were subsequently purified in order to explore the following topics: i) Functional characterization of GRSF1-mRNA interaction. ii) Preparation of specific antibodies that can be used for immunoblotting and immunohistochemistry (expression profiling). iii) Protein crystallization and X-ray diffraction studies to resolve the 3D-structure of the protein(s). Such direct structural information is needed to understand the molecular basis of GRSF1-mRNA interaction.

One of the GRFS1 truncation mutants (GST-ΔE1-hGRSF1) lacked the N-terminal alanine-rich domain but involved the three RNA-binding domains qRRM1, qRRM2 and qRRM3. We constructed this mutant protein in order to shorten the coding sequence for better efficiency of recombinant expression. Since the Ala-rich domain of GRSF1 was suggested to serve as a mitochondrial targeting signal it may not be essential for RNA binding (Jourdain et al., 2013). Thus, the danger that this gene technical truncation would impact the RNA-binding properties was low. In fact, in preliminary RNA-binding studies using human *GPx4* mRNA as a substrate we obtained similar Kd-values (629 nM for full-length human GST-GRSF1 and 555 nM for GST-ΔE1-hGRSF1) for truncated and untruncated proteins.

For recombinant expression of wild-type full-length human GRSF1 and the Ala-rich domain truncation mutant (ΔE1-hGRSF1) the corresponding cDNA sequences were cloned between the BamHI and XhoI restriction sites of the bacterial expression plasmid pET-42a, which contained an N-terminal GST-tag. Competent *E. coli* BL21 DE3 cells were transformed with the recombinant plasmids and then cultured in 3 l (full length GRSF1) and 0.5 l (GST-ΔE1-hGRSF1) of kanamycin containing LB medium. Expression of the recombinant proteins was induced with 1 mM (final concentration) of isopropyl-*β-D*-thiogalacto-pyranoside (IPTG). The bacterial cells were harvested by centrifugation and the respective cell pellets were resuspended in 25 ml of PBS containing 0.5 M TCEP (reducing agent) and protease inhibitors (400 µl of protease inhibitor cocktail, SERVA Electrophoresis GmbH, Heidelberg, Germany per 25 ml). The reducing agent and protease

inhibitors were added to prevent oxidative protein aggregation and proteolysis of the recombinant proteins. The resuspended cells were disrupted by sonication and the resulting cell lysates were centrifuged in order to recover the lysis supernatant containing the soluble proteins. For affinity purification of the recombinant proteins, the lysis supernatants were incubated with glutathione-coupled agarose beads. The attached GST-fusion proteins were competitively eluted with an elution buffer containing reduced glutathione (10 mM). The different elution fractions were then analyzed by 10% SDS-PAGE (**Figure 3.1 A and B**).

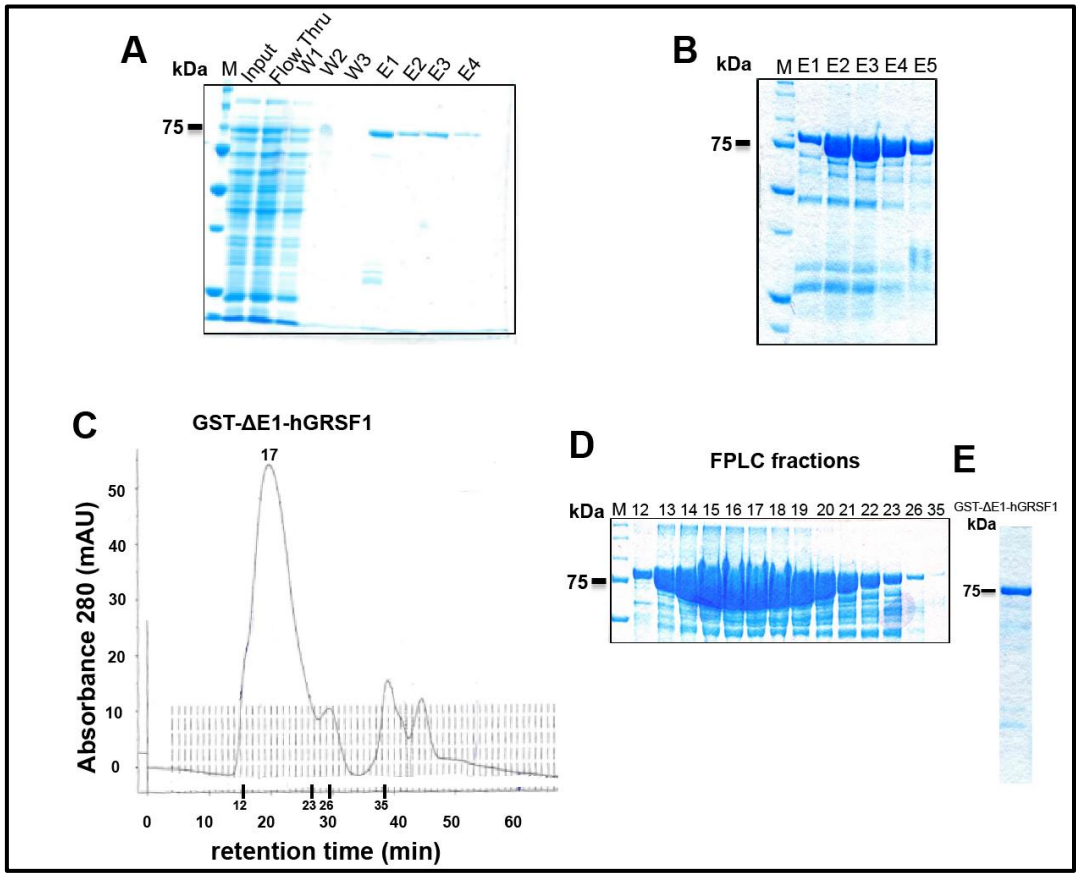

Figure 3.1 Expression and purification of recombinant full-length human GST-GRSF1 and its deletion mutant lacking the Ala-rich domain. A) Full length human GST-GRSF1 fusion protein was expressed in *E. coli* in a 3 l of a bacterial liquid culture. Cells were lysed and the 20,000 g the lysis supernatant was affinity chromatographed on a GSH-loaded agarose column, which retains the GST-tagged proteins. These proteins were then eluted from the column with a GSH-containing elution buffer and aliquots of different elution fractions were analyzed by SDS-PAGE. M: molecular weight markers. Input: lysis supernatant. Flow Thru: unbound proteins. W1+W2: wash fractions 1 and 2. E1-4: elution fractions 1-4. **B**) The truncated human GST-GRSF1 version, which lacks the Ala-rich domain, was expressed in E. coli in a 0.5 l bacterial liquid culture. Cells were lysed and the 20,000 g lysis supernatant was affinity chromatographed on a GSH-loaded agarose column, GST—tagged fusion proteins were eluted from the column with a GSH-containing elution buffer and aliquots of different elution fractions were analyzed by SDS-PAGE. M: molecular weight markers. E1-5: Elution fractions 1-5 (**C**) Size exclusion chromatography of pooled and concentrated GST-ΔE1-hGRSF1 elution fractions. The peak fraction 17 is labeled. **D**) SDS-PAGE of aliquots of different elution fractions of size exclusion chromatography (panel B). **E**) SDS-PAGE showing purified ΔE1-hGRSF1 protein of 75 kDa (lower panel).

When the elution fractions of the full-length GRSF1 construct were analyzed (**Figure 3.1 A**) we found that a major protein band migrating in the molecular weight range of 75 kDa was eluted in fractions E1, E2 and E3. There was hardly any protein eluted in the washing fractions W2 and W3. When the lysis supernatant of the GST-ΔE1-hGRSF1 construct was taken through the same experimental protocol we observed a similar elution pattern (**Figure 3.1 B**). Here again, the majority of the recombinant protein was eluted in fractions 1, 2 and 3 but elution fractions 4 and 5 also contained recombinant protein.

From Fig. 3.1. (panels A+B) it can be seen that the staining intensity of the protein bands was much higher for the truncated protein species although the volume of the bacterial culture was 6-fold lower (0.5 l vs. 3 l) and the samples were worked up identically. When we calculated the overall yield of protein expression and normalized these values to the same culture volume we found that for wild-type full length GRSF1 about 0.67 mg pure fusion protein was obtained from 1 l of liquid culture. In contrast, for the Ala-rich domain lacking truncation mutant an overall yield of 10 mg/l liquid culture was obtained. These calculations indicate that GST-ΔE1-hGRSF1 is expressed at significantly higher levels than full-length recombinant human GST-GRSF1. Furthermore, GST-ΔE1-hGRSF1 protein preparation yielded a protein preparation with a significantly higher degree of purity (89%). The elution fractions from both protein preparations were pooled and concentrated to a final volume of 500 µl (full-length GST-GRSF1) and 800 µl (GST-ΔE1-hGRSF1). The protein preparations were mixed with 10% v/v glycerol and stored at -80°C for further purification or later functional analysis.

The major conclusion of these expression studies was that the ala-rich domain truncation protein was expressed at higher levels when compared with the full-length wild-type human GRSF1. This data and the previous observation that the truncated protein exhibits similar RNA-binding affinities prompted us to use the GST-ΔE1-hGRSF1 construct for further functional and structural studies. Unfortunately, the degree of purity of the elution fractions of the affinity chromatography was not high enough for detailed structural investigations (X-ray crystallography) and antibody production. Thus, we decided to include a second step of chromatographic purification (size exclusion FPLC, SEC) into our purification protocol. For this purpose 0.5 ml of the affinity purified and concentrated GST-ΔE1-hGRSF1 fusion protein was loaded onto a gel-filtration FPLC column (Superdex 75 10/300 GL). The column was eluted with elution buffer (20 mM Tris-HCl; 100 mM NaCl) at a flow rate of 0.5 ml/min. Fractions of 0.5 ml were collected and the elution profile was recorded measuring the absorbance at 280 nm (**Figure 3.1 C**). The chromatogram shows one major protein peak, which was eluted between fractions 12 and 23. When aliquots of these elution fractions were analyzed by 10% SDS-PAGE (**Figure 3.1 D**) we detected one

major protein which migrated in the MW range of about 75 kDa. In addition, there were three minor protein peaks in the chromatogram but SDS-PAGE analysis of the corresponding elution fractions indicated the lack of any protein migrating in the 75 kDa MW range. Thus, GRSF1 does not contribute to these protein bands.

Next we pooled the elution fractions 13-21 and combined them with the corresponding fractions of a second chromatographic run. The combined elution pool was concentrated using an ultrafiltration concentrator (30 kDa MW cut-off) reaching a final protein concentration of about 5 mg/ml and the degree of purity was determined by densitometric evaluation of SDS-PAGE (**Figure 3.1 E**). We obtained a high degree of purity (93%) for the final GST-ΔE1-hGRSF1 protein preparation.

Summary: Full-length human GRSF1 and its alanine-rich domain truncation mutant were expressed in *E. coli* as N-terminal GST-tagged fusion proteins at levels of about 0.7 mg purified protein/l liquid culture and 10 mg purified protein/l liquid culture, respectively. The proteins were purified from the bacterial lysis supernatant to near homogeneity by consecutive affinity chromatography on a GSH-agarose column and gel filtration.

3.1.2. Specific proteolytic cleavage of the recombinant fusion protein and subsequent purification of the GRFS1 cleavage peptide

One of our aims for preparing large amounts of recombinant GRSF1 protein was to use it as immunogen for the generation of polyclonal anti-GRSF1 antibodies. For this purpose the use of the GST-GRSF1 fusion protein was not suitable since the resulting antibody was expected to strongly cross-react with glutathione transferase, which was used as tag for expression of the fusion protein. We therefore removed the GST-tag by proteolytic cleavage of the fusion protein and subsequent purification of the GRSF1 share from the cleavage mixture. For this purpose we took advantage of the coagulation factor Xa specific proteolytic cleavage site, which was present in the fusion protein between the N-terminal GST and the C-terminal GRSF1 domain. The proteolytic cleavage site was specifically introduced at this position by the producer of the expression vector. In order to optimize the cleavage conditions and the purification protocol for the GRSF1 part of the fusion protein an aliquot (30 mg of protein) of the pooled and concentrated size exclusion chromatography fractions was treated with 250 µl of coagulation factor Xa (activity 1.5 nkat/µg, concentration 1 mg/ml) in a reaction volume of 6 ml (buffer 50 mM Tris, pH 8.0; 100 mM NaCl; 6 mM $CaCl_2$) and an aliquot of the cleavage mixture was analyzed on 10% SDS-PAGE (**Figure 3.2 A**). Surprisingly, we observed three major cleavage peptides, which migrated in SDS-PAGE at 53 kDa, 31 kDa and 27 kDa. The theoretic MW weight of the GST share was 26.98 kDa, which is in good agreement with the experimental value

obtained for the 27 kDa cleavage peptide. For the GRSF1 share of the fusion protein a theoretical MW of 42 kDa was calculated on the basis of the amino acid composition. However, under our experimental conditions the putative GRSF1 cleavage peptide migrated with an apparent MW of 53 kDa. The disparity with the theoretical MW is discussed in more detail in discussion (see **Section 4.1**). To our surprise we also observed a prominent 31 kDa fragment in SDS-PAGE (**Figure 3.2 A, lane After Factor Xa**). The chemical identity of this peptide fragment remains unclear but possible explanations are elaborated in the Discussion (see **Section 4.1**).

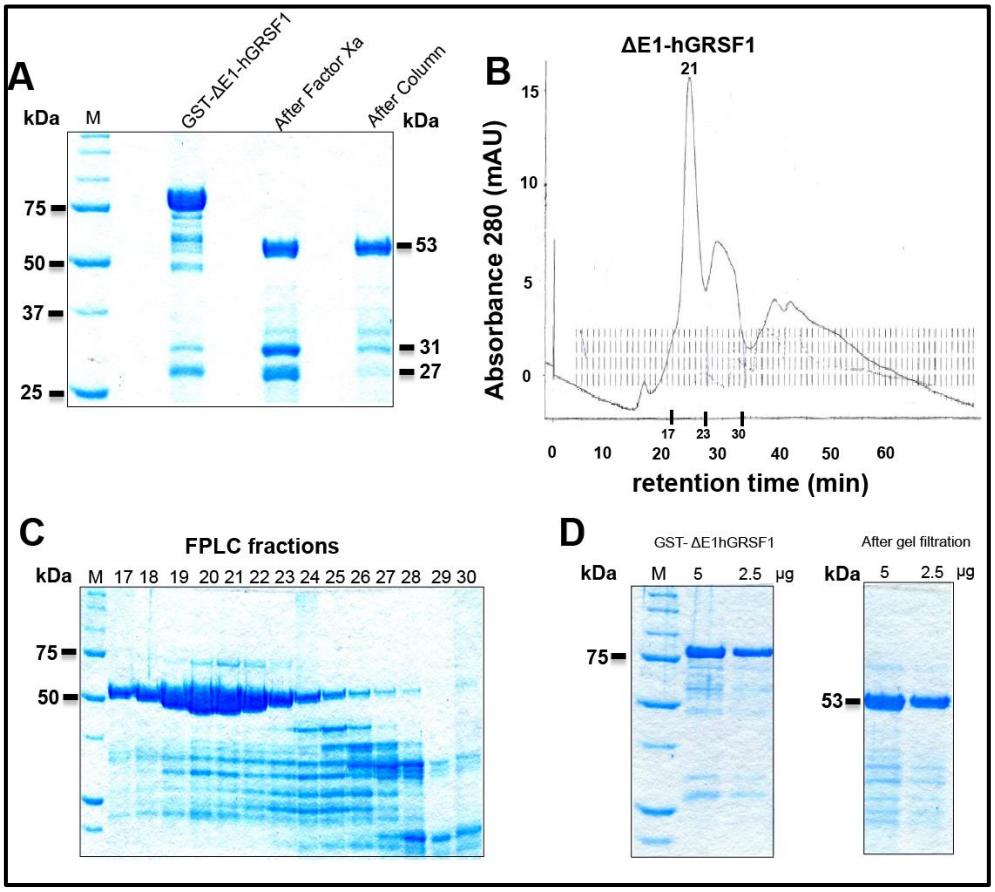

Figure 3.2 Removal of GST-tag and purification of recombinant Ala-deletion mutant lacking the Ala-rich domain. A) The purified GST- ΔE1-hGRSF1 fusion protein was cleaved with Factor Xa which separated the ~53 kDa recombinant ΔE1-hGRSF1 portion and the ~27 kDa GST portion (lane After Factor Xa). The digested protein mixture was then loaded on a GSH-loaded agarose column, which retains the GST-tag and the ~53 kDa ΔE1-hGRSF1 was collected in the flow through fraction (lane after column). **B**) Size exclusion chromatography of pooled and concentrated elution fractions of ΔE1-hGRSF1 protein. The peak fraction 21 is labeled. **C**) SDS-PAGE of aliquots of different elution fractions of size exclusion chromatography (panel B). **D**) SDS-PAGE of aliquots of different purified fractions (panel B). An aliquot of the full-length GRSF1 protein before size exclusion chromatography (5 and 2.5 µg, left) and size exclusion chromatography (5 and 2.5 µg, Right).

To remove GST share of the digested fusion protein the proteolysis mixture was rechromatographed by affinity chromatography on a GSH-agarose matrix. The flow

through fraction contained the GRSF1 share of the cleaved fusion protein whereas the majority of the GST share was retained on the GSH-matrix (data not shown). Next, the flow through fraction of the affinity column was concentrated, injected to size exclusion chromatography and elution fractions of 0.5 ml were collected (**Figure 3.2 B**). Here we observed one major protein peak, which was eluted between fractions 17-23 and a tale shoulder (elution fractions 24-30). In order to determine the protein composition in the different elution fractions we analyzed aliquots of these fractions by SDS-PAGE and found that fractions 17-23 mainly contain the GRSF1 share of the fusion protein. In fractions 24-30 (tale shoulder) other proteins were eluted. The chemical identities of these contaminating proteins have not been identified but the lower molecular weights suggested unspecific proteolytic cleavage products.

To analyze the degree of purity of the different enzyme preparations aliquots of the GST-GRSF1 fusion protein (before factor Xa cleavage) and after proteolytic cleavage and subsequent chromatographic purification were run on SDS-PAGE (**Figure 3.2 D**)**.** The electropherogram of the uncleaved GST-GRSF1 fusion protein (lane A+B) shows a single prominent protein band migrating with molecular weights of ~75 kDa (GST-ΔE1-hGRSF1). In contrast, after proteolytic cleavage and subsequent chromatographic purification one major protein band migrating with an apparent MW of 53 kDa was observed (ΔE1-hGRSF1 protein). The degree of purity of the final enzyme preparation (separated GRSF1 share) was 93 % as indicated by desitometric evaluation of the electropherogram.

Summary: The purified GST-Ala-rich domain truncation protein was proteolytically cleaved by Factor Xa into 3 fragments that include the 53 kDa recombinant GRSF1 portion, the 27 KDa GST portion and an unidentified 31 kDa protein. The GRSF1 cleavage peptide was purified to high purity (93%) and near homogeneity by employing GSH-agarose affinity column chromatography and gel filtration.

3.1.3. Bacterial expression and purification of the three RNA-binding domains of human GRSF1

The RNA-binding activities of the three quasi-RNA-recognition motifs (qRRMs) of hnRNP F (member protein of GRSF1) have been explored in detail (Cyril Dominguez et al., 2010; Samatanga et al., 2013). In contrast, the interactions between the three qRRM domains of GRSF1 and *GPx4* mRNA has not been characterized. In order to determine the dissociation constants (Kd) for the different RNA-binding domains, we separately overexpressed and purified these three domains [qRRM1 (residues 139-244), qRRM2 (residues 252-323) and qRRM3 (residues 400-480)]. The three qRRM domains were expressed and purified in a similar way as described earlier for full-length human GSRF1

and the alanine-rich domain deletion mutant. For this purpose 50 ml bacterial liquid cultures were set up, cells were harvested, resuspended in 10 ml PBS and lysed by sonication. The supernatants containing the soluble GST-tagged fusion proteins were applied to a glutathione-agarose affinity chromatography column and the attached proteins were competitively eluted with elution buffer containing reduced glutathione (10 mM). The elution fractions (1-4) from qRRM1, qRRM2 and qRRM3 fusion proteins were collected and analyzed on the 10% SDS-PAGE (**Figure 3.3 A and B**). This analysis showed one prominent band at ~40 kDa, ~38 kDa and ~37 kDa in the different liquid cultures representing the three RNA-binding domains qRRM1, qRRM2 and qRRM3, respectively. The electropherograms show that the three qRRM domains are well expressed (**Figure 3.3 A and B**) and the preparation procedure yielded the following highly concentrated protein solutions: i) 3 ml of 12 mg protein/ml for qRRM1, ii) 3 ml of 13 mg protein/ml for qRRM2, iii) 3 ml of 12.5 mg protein/ml for qRRM3). The degrees of purity of these preparations (94% purity for qRRM1, 74% purity for qRRM2, 75% for qRRM3) were sufficient for subsequent functional studies.

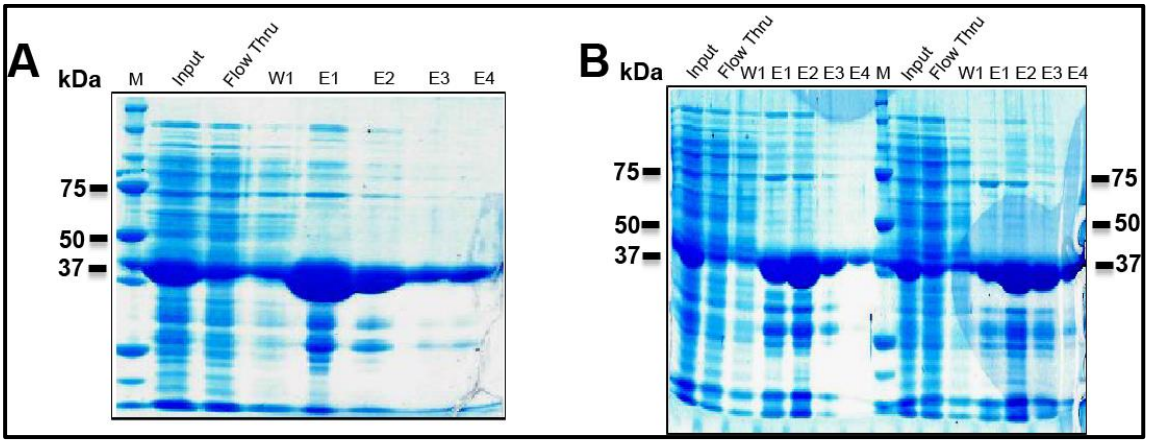

Figure 3.3 Expression and purification of the three RNA-binding domains of human GRSF1. The three domains were expressed as N-terminal GST-fusion proteins. The isolated GST-tagged human qRRM domains were expressed in *E. coli* in a 50 ml bacterial liquid culture. Cells were lysed and the 20,000 g the lysis supernatants were affinity chromatographed on a GSH-loaded agarose column. These proteins were then eluted from the column with a GSH-containing elution buffer and aliquots of different elution fractions were analyzed by SDS-PAGE. M: molecular weight markers. Input: lysis supernatant. Flow through: unbound proteins. W1: wash fractions 1. E1-4: elution fractions 1-4. **A**) qRRM-1, **B**) qRRM-2 and qRRM-3.

Summary: The recombinant RNA-binding domains (qRRM1, qRRM2, qRRM3) of human GRSF1 were overexpressed in E. coli as N-terminal GST fusion proteins and purified by affinity chromatography. The final yields were as follows: 12 mg/ml (3 ml qRRM1), 13.0 mg/ml (3 ml qRRM2) and 12.5 mg/ml (3 ml qRRM3).

3.1.4. Bacterial expression and purification of human GRSF1 truncation mutants lacking different RNA-binding domains

In order to quantify the relative contribution of different structural subunits for the RNA-binding activities of human GRFS1 we selectively deleted qRRM1 (ΔR1-hGRSF1), qRRM2 (ΔR2-hGRSF1), qRRM3 (ΔR3-hGRSF1) and the acidic domain (AD) of the ΔAla-hGRSF1 construct, which lacks the alanine-rich domain (**Figure 3.4**). In addition, we also created corresponding truncation mutants of full-length human GRSF1. However, here we only describe the expression results obtained for ΔAla-hGRSF1, in which Ala-rich domain is absent. These constructs were overexpressed and purified in a similar way as described earlier for full-length human GRSF1 and the alanine-rich domain deletion mutant.

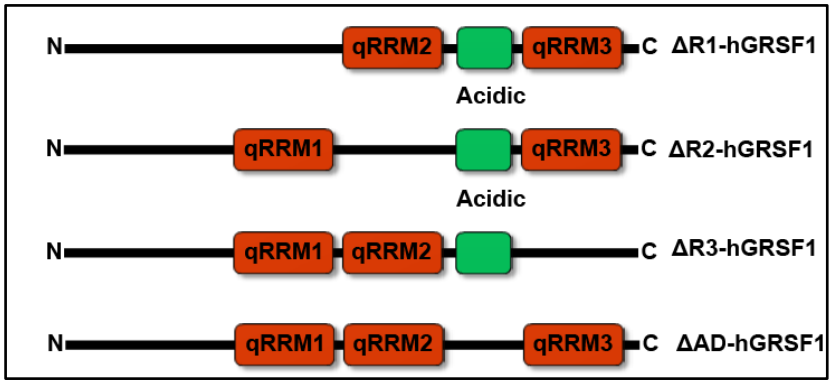

Figure 3.4 Diagram of single truncation mutants of ΔAla-hGRSF1 proteins. These protein mutants were prepared by deleting qRRM1 (ΔR1-hGRSF1), qRRM2 (ΔR2-hGRSF1), qRRM3 (ΔR3-hGRSF1), and acidic domain (ΔAD-hGRSF1).

Transformed bacteria were cultured in 50 ml cultures, harvested by centrifugation and resuspended in 10 ml of PBS. The supernatants containing the soluble GST-tagged fusion proteins were applied to a glutathione-agarose affinity column and attached proteins were competitively eluted with elution buffer containing reduced glutathione (10 mM). The different elution fractions were analyzed on 10% SDS-PAGE (**Figure 3.5**). In each lysis supernatant we detected a dominant protein band that migrated in the MW range between 65-70 kDa (**Figure 3.5 A and B**). These proteins correspond to the ΔR1-hGRSF1, ΔR2-hGRSF1, ΔR3-hGRSF1 and ΔAD-hGRSF1 of ΔAla-hGRSF1 shortened protein (**Figure 3.5 A and B**). The apparent molecular weights of these constructs were consistent with the theoretical values concluded from the amino acid composition. Furthermore, these expressions yielded significantly high amounts of concentrated protein preparations but the degree of purity was variable for the different proteins:

ΔR1-hGRSF1: 0.9 ml, 7.9. mg/ml, 58% purity

ΔR2-hGRSF1: 0.8 ml, 20.2 mg/ml, 57% purity

ΔR3-hGRSF1: 0.8 ml, 13.3 mg/ml, 79% purity

ΔAD-hGRSF1: 0.7 ml, 3.2 mg/ml, 85% purity.

These protein preparations were subsequently employed to determine the dissociation constants of RNA-binding.

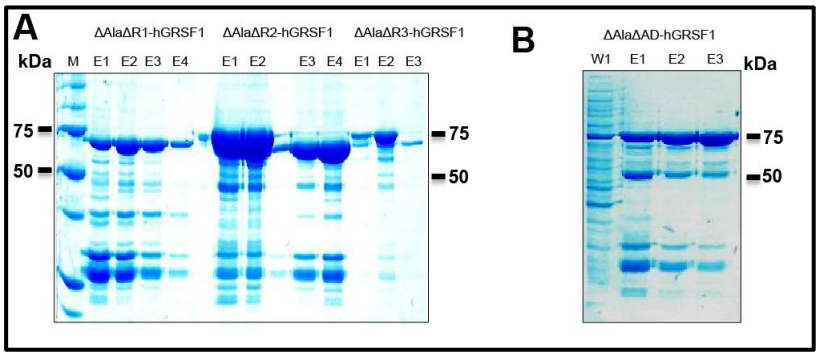

Figure 3.5 Expression and purification of the GST-tagged single truncation proteins of human ΔAla-GRSF1 from *E.coli*. The GST-tagged single deletion mutants of ΔAla-hGRSF1 (lacking Ala-rich domain) were expressed in *E. coli* in a 50 ml of a bacterial liquid culture. Cells were lysed and the 20,000 g the lysis supernatant was affinity chromatographed on a GSH-loaded agarose column, which retains the GST-tagged proteins. These proteins were then eluted from the column with a GSH-containing elution buffer and aliquots of different elution fractions were analyzed by SDS-PAGE. M: molecular weight markers. E1-4: elution fractions 1-4. **A**) Deletion of the different RNA-binding domains, **B**) deletion of the alanine-rich domain.

Summary: The single truncation mutants lacking the qRRM1 (ΔR1-hGRSF1), qRRM2 (ΔR2-hGRSF1), qRRM3 (ΔR3-hGRSF1) and acidic domain (ΔAD-hGRSF1) of the shortened version of ΔAla-hGRSF1 construct, which lacks the alanine-rich domain were overexpressed in *E.coli* as N-terminal GST fusion proteins. These truncation mutants were purified by passing bacterial lysis supernatant on a GSH-agarose affinity column and the following yields were obtained: i) 7.9 mg protein/ml (0.9 ml, ΔR1-hGRSF1), ii) 20.2 mg protein/ml (0.8 ml, ΔR2-hGRSF1), iii) 13.3 mg protein/ml (0.8 ml, (ΔR3-hGRSF1), iv) 3.2 mg protein/ml (0.8 ml (ΔAD-hGRSF1),.

3.1.5. Bacterial expression and purification of mouse GRSF1 truncation mutants lacking different RNA-binding domains

Similar to the human genome the mouse genome involves a single copy *Grsf1* gene and on the amino acid level the degree of homology between human and mouse GRSF1 is 91%. In order to explore the functionality of the different structural subunits of mouse GRSF1 we constructed similar GRSF1 truncation mutants as we did before for the human protein. The three mouse qRRM domains were expressed and purified in a similar way as described for the human protein but in order to save space we only report expression of the qRRM1 construct. The qRRM1 domain was cloned in pET-28b expression plasmid containing the N-terminal histidine tag (His-tag) and a 0.5 l bacterial liquid culture was grown. Cells were harvested by centrifugation and resuspended in 25 ml PBS. After sonication the supernatants containing the soluble his-tagged fusion proteins

were purified on a Ni-NTA affinity column and the attached his-tag fusion proteins were competitively eluted with elution buffer containing imidazole (200mM).

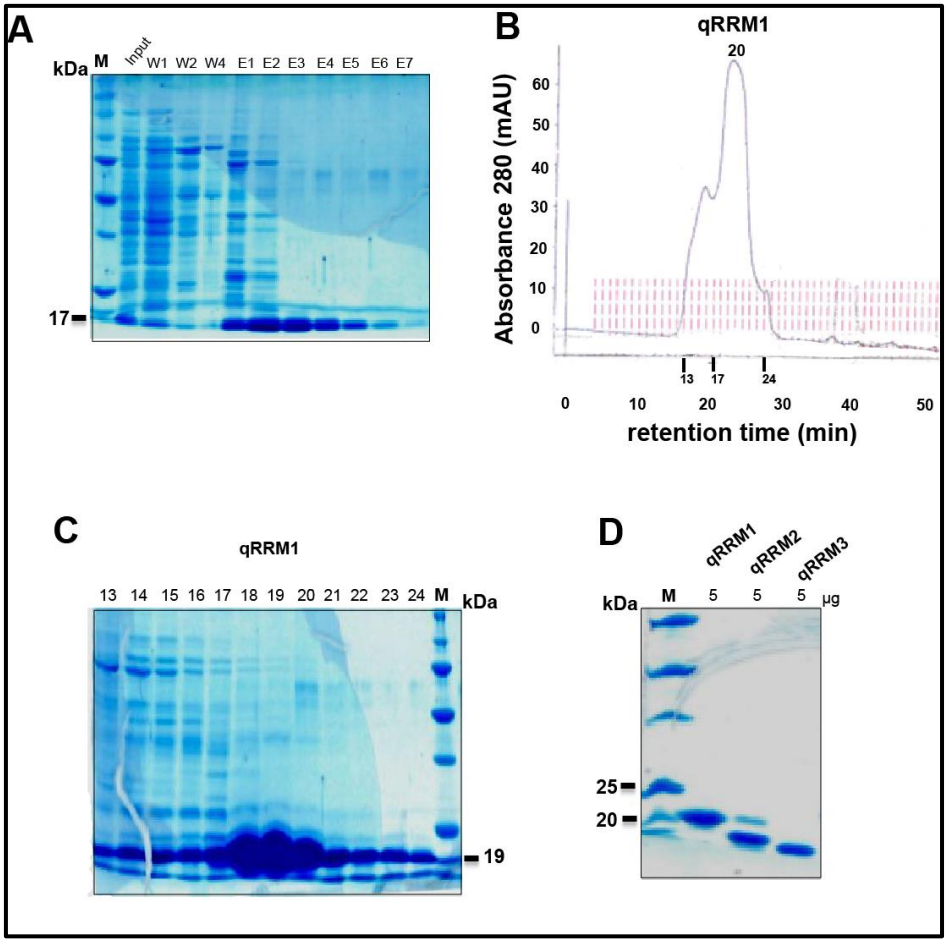

Figure 3.6 Expression and purification of recombinant his-tagged qRRM1 domain of mouse GRSF1. (**A**) His-tagged qRRM1 domain of mouse GRSF1 was expressed in *E. coli* in a 500 ml bacterial liquid culture. Cells were harvested, lysed and the 20,000 g the lysis supernatant was run over a Ni-NTA affinity column, which retains the his-tag fusion proteins. These proteins were then eluted from the column with an Imidazole-containing elution buffer and aliquots of different elution fractions were analyzed by SDS-PAGE. M: molecular weight markers. Input: lysis supernatant. W1+W2+W3: wash fractions 1, 2 and 3. E1-7: elution fractions 1-7. (**B**) Size exclusion chromatography of pooled and concentrated his-tag qRRM1 elution fractions obtained by affinity chromatography (panel A). The peak fraction 20 is labeled. (**C**) Aliquots (15 µl) of each gel filtration elution fraction were analyzed by 10% SDS-PAGE. (**D**) Aliquots (5 µg protein) of the final protein preparations (qRRM1, qRRM2, qRRM3) were analyzed by 10% SDS-PAGE.

The purified his-tagged fusion protein were then analyzed on 10% SDS-PAGE (**Figure 3.6 A**). Here we detected a prominent protein band at 19 kDa and this MW is consistent with the theoretical MW calculated for the qRRM1 domain of mouse GRSF1. SDS-PAGE also shows that the His-tag fusion protein is expressed efficiently and we calculated the following parameters for the different RNA-binding domains:

His-qRRM1: 0.2 ml, 60 mg/ml, 95% purity (after gel filtration)

His-qRRM2: 0.2 ml, 14 mg/ml, 95% purity (after gel filtration)

His-qRRM3: 0.2 ml, 17mg/ml, 95% purity (after gel filtration)

Although the affinity purified protein preparations already exhibited a high degree of purity we further purified his-tag fusion proteins by gel filtration. This purification was carried out in the same way as described for GST-ΔE1-hGRSF1 fusion protein. In brief, 0.5 ml of the affinity purified and concentrated his fusion proteins were loaded onto a gel-filtration FPLC column (Superdex 75 10/300 GL). The column was then eluted with elution buffer (20 mM Tris-HCl; 100mM NaCl) at a flow rate of 0.5 ml/min. Fractions of 0.5 ml were collected and the elution profile was recorded measuring the absorbance at 280 nm (**Figure 3.6 B**). The chromatogram shows one major protein peak, which is eluted between fractions 17 and 24 (**Figure 3.6 B**). When aliquots of these elution fractions were analyzed by 10% SDS-PAGE (**Figure 3.6 C**) we detected one major protein which migrated in the MW range of about 19 kDa.

Next we pooled the elution fractions 17-24 and combined them with the corresponding fractions of a second chromatographic run. The combined elution pool was concentrated using an ultrafiltration concentrator (5 kDa MW cut-off) reaching a final protein concentration of about 60 mg/ml. The degree of purity was determined by densitometric evaluation of SDS-PAGE. We achieved a very high degree of purity (95%) for the final his-tagged qRRM1 and this was also the case for qRRM2 and qRRM3 (**Figure 3.6 D**). The qRRM1-3 protein domains were snap frozen in liquid nitrogen and stored at -80°C.

Summary: The three mouse RNA-binding domains (qRRM1, qRRM2 and qRRM3) were overexpressed in *E. coli* as N-terminal His-tagged fusion proteins at levels of 0.2 ml of 60 mg protein/ml, 0.2 ml of 14 mg protein/ml and 0.2 ml of 17 mg protein/ml. These recombinant separated protein domains were first purified by GSH-agarose affinity column chromatography and afterwards by gel filtration in order to be used for 3D structure determination of GRSF1 protein by X-ray crystallography.

3.2. Evolutionary aspects of GRSF1

3.2.1. *In silico* search strategy for GRSF1-like sequences

GRSF1 has been suggested to be highly conserved in vertebrates, which are ranked in evolution above bony fish (Ufer, 2010). A corresponding gene was detected in zebrafish, which is frequently used as a model organism of bony fish (Ufer, 2012). Unfortunately, the occurrence of GRSF1 has not been explored in detail in lower model organisms. To investigate the occurrence of GRSF1-like sequences in viruses and across the three domains of terrestrial life (*Bacteria, Archaea* and *Eukarya*), we performed a computer search of the publically available protein sequences using the protein BLAST

program on the NCBI platform (https://blast.ncbi.nlm.nih.gov/Blast.cgi). We started our search using the amino acid sequence of full-length human GRSF1 (NP_002083.3) as a query template to find similar sequences in the different protein databases. To reduce the multiplicity of positive hits we only considered those sequences with an amino acid identity score higher than 20% (Altschul, Gish, Miller, Myers, & Lipman, 1990).

3.2.2. GRSF1-like sequences in viruses

Viruses are small particles with a diameter as low as 15 nm. However, there are also giant viruses, which reach a size of up to 500 nm. Since viruses do not contain ribosomes they replicate only within a host cell. Because of their dependence on the host cell protein synthesizing machinery, they are excellent models to explore the regulation of mRNA translation (Park et al., 1999). The presence of GRSF1 in viruses has not been reported before and thus, we searched the virus protein databases of the NCBI platform for GRSF1-like sequences using human GRFS1 as template. Interestingly, our search revealed three hits in different viruses (**Table 3.1**).

Protein	Sequence ID	Virus	Sequence Identity
C protein	NP_958051.1	Mossman virus	50%
ORF056	YP_238617.1	Staphylococcus phage Twort	32%
Polyprotein	YP_605811.1	Banana streak virus Acuminata Yunnan	34%

Table 3.1 Amino acid sequence homology of GRSF1 protein in different viruses. The table shows the occurrence of GRSF1-like sequences in different viruses.

One GRSF1-like sequence was found in a single-stranded RNA (ssRNA) virus (Mossman virus, amino acid identity 50%) and a second one in the double-stranded DNA virus (Staphylococcus phage Twort, amino acid identity 32%). Finally, we found a GRSF1-like sequence in the banana streak virus Acuminata Yunnan. This retrovirus expresses a large protein (1900 amino acid), which shares an amino acid identity of 34% with human GRSF1. However, for neither of these proteins the functionality has been explored. Thus, it remains unclear whether the corresponding proteins exhibit RNA-binding capabilities. Although our database search indicated that GRSF1-like sequences occur in the genomes of viruses the occurrence frequency is rather low. We searched about 7000 publically available viral genomes and found only 3 hits. Thus, the occurrence frequency of GRSF1-like proteins is <0.05%. In the genomes of human pathogenic viruses such as HIV, HBV or HCV GRSF1 like sequences do not occur.

Summary: GRSF1-like sequences are rare in viruses (occurrence frequency <0.05%) and this data suggests that GRSF1 is not a typical viral protein. Viral GRSF1 may not be related to human virus infections.

3.2.3. GRSF1 like sequences in bacteria and archaea

Next, we investigated the occurrence of GRSF1-like sequences in bacteria and archaea. *Escherichia coli* (*E. coli*), a Gram-negative, rod-shaped bacterium, is frequently used as model organism for bacteria in molecular genetics (Taj, Samreen, Ling, Taj, & Yunlin, 2014). When we searched the *E. coli* reference genome we did not find any GRSF1-like sequences. Next, we searched for the presence of such sequences in three randomly selected bacterial groups that include purple bacteria, purple non-sulfur bacteria, and Gram-positive bacteria. Like in *E. coli*, we did not find GRSF1-like sequences in any of these bacterial species. Altogether we searched more than 40,000 publically available bacterial genomes (NCBI genome browser website as of March 06, 2017) and found no hits. These results prompted the conclusion that GRSF1 may not occur in bacteria.

Then we searched for GRSF1-like sequences in archaea. Here again, we did not find related sequence. At the search time (March 06, 2017) the NCBI genome database involved the sequence of more than 1000 archaeal genomes but we did not detected any GRSF1-related protein. Taken together, these data suggests that GRSF1 is not involved in bacteria and archaea physiology.

Summary: GRSF1-like sequences do not occur in bacteria and archaea and thus, this protein is not needed for bacterial or archaeal physiology.

3.2.4. GRSF1 like sequences in *Saccharomyces cerevisiae* and other fungi

Fungi are microorganisms that lack chlorophyll and are therefore heterotrophs. They play a role in alcoholic fermentation and are used in wine making, baking and beer brewing (Barnett, 1998). The main hallmark of fungi is that their cell walls are made up of chitin that separates these organisms from plants, bacteria and protists. *Saccharomyces cerevisiae* (Baker's yeast) is frequently used as a model organism for fungi. It constitutes a single-celled eukaryotic organism, which because of its simple maintenance and its small genome, has become an attractive model system to study the functionality of highly conserved genes (Botstein, Chervitz, & Cherry, 1997; Dorsey, Peterson, Bray, & Paquin, 1992). When we explored the presence of GRSF1-like sequences in baker's yeast, we did not find such sequence. To investigate whether such sequences occur in other Saccharomyces subspecies we applied our *in silico* search strategy to four randomly selected subspecies (*Saccharomyces capensis*, *Saccharomyces italicus*, *Saccharomyces oviformis*, *Saccharomyces uvarum var. melibiosus)*. Here again, we did not detect GRSF1-ike sequences.

Next, we searched the protein databases of other fungi for GRSF1-like sequences and retrieved five hits (**Table 3.2**). These proteins share a medium degree (23-35%) of

amino acid identity with human GRSF1 and some of them have been implicated in RNA-binding. However, here again the frequency of occurrence is rather low (0.25%) since only 5 hits were retrieved when we screened more than 2100 fungal genomes.

Protein	Sequence ID	Fungus	Sequence Identity
rna-binding protein prp24	XP_017993410.1	Malassezia pachydermatis	23%
THO complex subunit	NP_595161.1	Schizosaccharomyces pombe 972h-	35%
RNA recognition motif domain-containing protein	XP_018174889.1	Purpureocillium lilacinum	24%
Nucleotide-binding, alpha-beta plait, partial	XP_014576034.1	Metarhizium majus ARSEF 297	24%
polyadenylate-binding protein 2	XP_011319224.1	Fusarium graminearum PH-1	24%

Table 3.2 GRSF1-like proteins in different fungal species. The table shows the degree of amino acid identity of GRSF1-like sequences with human GRSF1 in GRSF1-positive fungi.

Summary: GRSF1-like sequences are rare in fungi (occurrence frequency <0.3%) and this data suggests that GRSF1 is not a typical fungal protein.

3.2.5. GRSF1-like sequences in *Arabidopsis thaliana* and other plants

Arabidopsis thaliana is frequently used as a model organism for higher plants (Mitchell-Olds, Thomas, December 2001). Its genome is relatively small and only comprises 135 Mbp. It was the first plant genome that was completely sequenced (The Arabidopsis Genome Initiative, 2000) and since then *A. thaliana* is a popular research tool for exploring the molecular biology and physiology of plants in general (Koornneef & Meinke, 2010). When we screened the proteome of *A. thaliana* for the presence of GRSF1-like sequences we detected a single RNA-binding protein (RBP) called RNA-binding family protein (NP_201402.3). This protein shares an amino acid identity of 34% with human GRSF1. In addition, we explored the occurrence of GRSF1-like sequences in other randomly selected *Arabidopsis* subspecies and found similar proteins in *A. lyrata* (XP_002866778.1) and in *A. salsuginea* (XP_006406335.1). These two proteins share an amino acid identity of 33% with human GSRF1. However, no such protein was detected in *A. heynh*. This result does not necessarily mean that GRSF1-like sequences are absent in this *Arabidopsis* subspecies since the negative outcome could be related to the low quality of the genomic information on this plant species.

Next, we applied our search strategy to detect GRSF1-like sequences in other plants. Until now 58 completely annotated plant genomes are available on NCBI platform

(https://www.ncbi.nlm.nih.gov/genome/annotation_euk/all/). 56 of these genomes were blastable and thus, could be used for protein blast searching. When we searched these 56 genomes using the amino acid sequence of full-length human GRSF1, we obtained 55 hits suggesting that GRSF1-like proteins are present in almost all plants (occurrence frequency 98.21%). Unfortunately, we could not find GRSF1-like proteins in the predicted proteome of *Zea mays*. However, when we blasted another NCBI database (https://www.ncbi.nlm.nih.gov/genome/browse/), which contains annotated as well as unannotated genomes, we found a GRSF1-like protein in the *Zea mays (*XP_008647847.1, amino acid identity 34%).

When we carried out more specific searching strategies in selected lower and higher plant species we found *a hypothetical protein (*XP_001771489.1*) in* model moss *Physcomitrella patens*, which represents lower plants. This protein shares an amino acid identity of 30% with human GRSF1. In addition, we found a GRSF1-like protein (NP_777440.1, amino acid identity 45%) in the hornwort species *Anthoceros angustus*, which also represents lower plants. Finally, in order to explore whether GRSF1 like sequences are present in higher plants, we selected by chance five different species of higher plants. Here we found GRSF1-like sequences in all selected species, such as *Oryza sativa Japonica* (XP_015617185.1, amino acid identity 34%), *Zea mays (*XP_008647847.1, amino acid identity 34%), *Brassica rapa* (XP_009112266.1, amino acid identity 35%), *Triticum aestivum* (GenBank: AAB38974.1, amino acid identity 21%) and in *Helianthus annuus (*GenBank: AAF02776.1, amino acid identity 45%).

3.2.6. GRSF1-like sequences in lower animals

3.2.6.1. *Drosophila melanogaster* and other insects

Drosophila melanogaster (fruit fly) is an insect model organism, which is widely used for biological research in studies of genetics, physiology, microbial pathogenesis, and evolution (Jennings, 2011). It is a common pest in homes, restaurants, and other places where food is served. It is frequently used in research because it is an animal species that is easy to maintain, has a relatively small genome (only four pairs of chromosomes), breeds quickly, and lays many eggs. To test whether this species expresses GRSF1-like proteins we searched its proteome, which was predicted on the basis of its reference genome, with the amino acid sequence of human GRSF1. Here we found an RNA-binding protein called fusilli isoform G (NP_001163161.1), which shares an amino acid identity of 29% with human GRSF1. Applying the same searching strategy we then investigated whether other subspecies of *Drosophila* contain GRSF1-like proteins. Here we found an uncharacterized protein called Dsimw501_GD20630, isoform A (XP_016035086.1), which shares an amino acid identity of 43% with human GRSF1 in *D. simulans*. Furthermore, we

found a hnRNP H3 isoform X1 (XP_017118850.1) and a hnRNP A3 homolog 2 isoform X2 (XP_017083817.1) with amino acid identities of 42% to human GRSF1 in *D. elegans* and in *D. eugracilis* respectively. In addition, we found a hnRNP F protein (XP_017062200.1) with an amino acid identity of 43% with GRSF1 in *D. ficusphila.* These data suggests that GRSF1-like proteins occur in *Drosophila* subspecies but little is known on the biological relevance of these proteins.

Next, we selected by chance four different insect species and found GRSF1-like sequences in *Apis mellifera* (XP_006568316.1, amino acid identity 44%), *Anopheles gambiae* str. PEST (XP_320791.4, amino acid identity 31%), *Papilio machaon* (XP_014370498.1, amino acid identity 39%) and in *Musca domestica* (XP_005185725.1, amino acid identity 45%). This data indicates that GRSF1 like sequences occur in insects but owing to the limited number of specific searching results a general statement on the occurrence frequency of GRSF1 in insects can hardly be made.

Summary: GRSF1-like proteins occur in *Drosophila melanogaster*, in other *Drosophila* subspecies and in all four by chance selected insect species

3.2.6.2. *Caenorhabditis elegans* and other worms

As an additional representative of lower animals we searched the predicted proteome of *Caenorhabditis elegans* for the occurrence of GRSF1-like proteins. *C. elegans* is a soil-dwelling nematode that is frequently used as non-vertebrate model organism in developmental and neurobiology. It is predominantly hermaphroditic (self-fertilizing) and transparent, which allows direct structural and functional characterization of virtually each cell in the organism (Corsi, Wightman, & Chalfie, 2015). When we applied our searching strategy we identified the RNA-binding protein sym-2 (NP_495960.2). This protein involves 618 amino acids and shares a 31% amino acid identity with human GRSF1. To explore whether GRSF1-like proteins are present across different *Caenorhabditis* subspecies, three related subspecies were selected by chance. Our homology analysis detected a RNA-binding protein (CRE-TWK-4, XP_003094466.1) in *C. remanei* and a similar protein in *C. briggsae* (CBR-HRPF-1 protein, XP_002639457.1). The CRE-TWK-4 protein is very long (1026 amino acids) and shares 30% amino acid identity with human GRSF1. The CBR-HRPF-1 protein is shorter (556 amino acids) and shares a 35% amino acid identity with human GRSF1. Finally, when we screened the predicted *C. vulgaris* proteome for the GRSF1-like sequences we found a similar RNA-binding protein (CRE-TWK-4, XP_003094466.1), which shares a 30% amino acid identity with human GRSF1. Taken together these results suggest that GRSF1 occurs in different *Caenorhabditis* subspecies.

When we explored the occurrence of GRSF1-like sequences in the proteomes of other roundworms we detected such proteins in: i) *Necator americanus* (XP_013297876.1, 28% amino acid identity), ii) *Loa loa* (XP_003142925.1, 29% amino acid identity), iii) *Trichinella spiralis* (XP_003374467.1, 28% amino acid identity), iv) *Brugia malayi* (XP_001893871.1, 29% amino identity). Altogether, we searched the proteomes of 105 roundworm that are currently available in the NCBI database but only retrieved 4 hits. Thus, the occurrence frequency is lower than 4 %.

Summary: GRSF1-like sequences occur in *Caenorhabditis elegans and other Caenorhabditis* subspecies. However, in the predicted proteomes of other roundworms GRSF1-like proteins are rare.

3.2.6.3. GRSF1 like sequences in vertebrates including mammals

It has been suggested before that GRSF1 is highly conserved in lower and higher vertebrates (Ufer, 2012). For this study we searched the proteomes of randomly selected vertebrate species to explore the occurrence frequency of GRSF1-like sequences.

Zebrafish and other bony fishes: We first analyzed the predicted zebra fish (*Danio rerio*) proteome. The zebrafish is a valuable vertebrate model organism that is well suited to study developmental biology (Dooley & Zon, 2000). The eggs are fertilized outside the organism, the embryos are transparent so that embryogenesis can easily be followed. Our searching strategy indicated that a GRSF1 related protein (NP_001039317.1) is present in the zebrafish proteome. This protein consists of 301 amino acids and shares an amino acid identity of 40% with human GRSF1. We then randomly selected four other fish species and detected a GRSF1 related protein in *Cyprinus carpio* (XP_018935608.1, amino acid identity of 46%), *Salmo salar* (NP_001135339.1, amino acid identity of 48%), *Latimeria chalumnae* (XP_006003791.1, amino acid identity of 54%) and in *Oreochromis niloticus* (XP_003444523.1, amino acid identity of 47%). This data indicates that GRSF1-like proteins are apparently widely distributed in fish.

Amphibia: Next we searched for occurrence frequency of GRSF-like proteins in amphibia. Unfortunately, there is only scattered information on the genomes of different amphibia in the NCBI database but the complete genomes of *Xenopus tropicalis* (clawed frog), *Nanorana parkeri* (Tibetan frog) (Sun et al., 2015), *Ambystoma mexicanum* (Axolotl) and *Xenopus laevis* are available. *Xenopus tropicalis* is frequently used as model organism for amphibia because of its relatively small genome and its short regeneration time (Akkers, Jacobi, & Veenstra, 2012). We found a GRSF1-like protein (XP_012825659.1, amino acid identity of 41% with human GRSF1) in the predicted

reference proteome of the clawed frog. In addition, a GRSF1-like protein was detected in the proteome of *Nanorana parkeri* (XP_018409058.1, amino acid identity 50%) and in *Xenopus laevis* (NP_001121205.2, amino acid identity 49%). These data suggest that GRSF1-like proteins are present in amphibians but owing to the low number of available genomes a more comprehensive statement on the occurrence frequency in amphibia cannot be made.

Reptiles: We then searched the presence of GRSF1-like proteins in reptiles. Among reptiles *Anolis carolinensis* (anole lizard) is frequently used as a model organism because of the low cost of breeding. It has a genome with a size of 1.78 Gb that contains higher number of mobile elements than any other sequenced amniote genome (Alföldi et al., 2011). We found a GRSF1-like protein (XP_003221902.1, amino acid identity 62%) in the reference proteome of anole lizard. In addition, we searched for GRSF1-like proteins in other by chance selected reptile species and found such proteins in *Chelonia mydas* (XP_007065973.1, sequence homology 67%), *Chrysemys picta bellii (*XP_005304175.1, sequence homology 66%), *Crocodylus porosus* (XP_019389516.1, sequence homology 60%) and in *alligator mississippiensis (*XP_014462839.2, sequence homology 60%). Thus, GRSF1-like proteins occur in reptiles but again but owing to the low number of available genomes a more comprehensive statement on the occurrence frequency in reptiles can hardly be made.

Birds: The domestic chicken (*Gallus gallus) is frequently* used as a model organism in embryology and phylogenetics (Burt, 2007). When we searched the predicted *G. gallus* proteome for GRSF1-like proteins we found that the putative chicken ortholog of human GRSF1 (XP_015131904.1) shares a 50% amino acid identity with the human protein. Then we analyzed the predicted protein sequences of four by chance selected other birds. We found a GRSF1-like protein in the proteome of *Corvus brachyrhynchos* (XP_017594433.1, amino acid identity 54%), *Columba livia* (XP_013223032.1, amino acid identity 56%), *Melopsittacus undulates* (XP_005142797.2, amino acid identity 55%), and *Falco cherrug (*XP_014132261.1, amino acid identity 57%). These data suggest that the GRSF1-like proteins occur in birds but owing to the low number of specific searches a more comprehensive statement on the occurrence frequency in birds can hardly be made.

Mammals: To explore the occurrence of GRSF1-like proteins in mammals we analyzed a number of by chance selected mammalian predicted proteomes. Mice are extensively used mammalian model organism (Waterston et al., 2002). The putative murine ortholog (NP_848815.2) of human GRSF1 shares a high (91%) degree of amino acid conservation with human GRSF1. In addition, we explored the frequency of GRSF1 protein in other randomly selected mammalian species and found GRSF1 related proteins

in the proteomes of *Rattus norvegicus* (NP_001094360.1, amino acid identity 90%), *Bos taurus* (NP_001071439.1, amino acid identity 93%), *Canis familiaris* (XP_013974145.1, amino acid identity 95%), *Felis catus* (XP_003985346.1, amino acid identity 95 %) *Oryctolagus cuniculus* (XP_008265989.2, amino acid identity 94 %), *Sus scrofa (*XP_003129120.1, amino acid identity 92%), *Capra hircus (*XP_017905038.1, amino acid identity 93%), *Loxodonta Africana (*XP_003414210.1, amino acid identity 92%), *Panthera tigris altaica (*XP_007086988.1, amino acid identity 95%), and in *Ursus maritimus (*XP_008691413.1, amino acid identity 96%). Thus, from the 10 randomly selected species of higher mammals we found putative GRSF1 orthologs in all predicted proteomes.

In addition, we found GRSF1-like proteins in lower mammals including *Monodelphis domestica* (gray short-tailed opossum *(*XP_001364439.2, amino acid identity 80%) and *Sarcophilus harrisii* (Tasmanian devil) *(*XP_012407379.1, amino acid identity 83%). Surprisingly, we did not find GRSF1-like proteins in *Phascolarctos cinereus* (koala), *macropus eugenii* (tammar wallaby) or in *Vombatus ursinus* (common wombat). Thus from the five by chance selected species of lower mammals we only detected GRSF1-like proteins in two species. This result does not necessarily mean that a GRSF1-like proteins are absent in these lower mammals. The negative outcome of our search might be related to the lower quality of the deposited genomic sequences.

Non-human primates: Next, we investigated the presence of GRSF1 related proteins in non-human primates. Chimpanzee (*Pan troglodytes*) is the closest living relative of modern humans and the chimpanzee genome was the first non-human primate genome that was sequenced (Consortium, 2005). We found a GRSF1-like protein (XP_016807098.1) in the reference genome of chimpanzee that shares a sequence identity of 99% with human GRSF1. When we applied our searching strategy to four other non-human primate species, we found GRSF1-related protein in *Gorilla gorilla* (XP_004038836.1, sequence identity 99%), *Pongo abelii* (XP_009238334.1, sequence identity 93%) and *Papio anubis* (XP_009205236.1, sequence identity 97%). In addition, we found a GRSF1-related protein in the predicted proteome of *Callithrix jacchus* (XP_002745793.1, sequence identity 95%), which represents a lower non-human primate. These data suggests that GRSF1 proteins are present in most non-human primates.

Modern and extinct humans: Finally, we explored the occurrence of GRSF1 in modern humans (*Homo sapiens*) and in two extinct human subspecies (*H. neanderthalensis* and *H. denisovan*). The H. sapiens reference genome involves a single GRSF1 gene, which is localized on the long arm of chromosome 4 (4q,13-3). It encodes for a 480 amino acid protein and is divided into 10 exons and 9 introns. In the proteomes of *H. neanderthalensis* and *H. denisovan*, which were predicted on the basis of the

genomic sequences of these extinct human subspecies, we also found single GRSF1 orthologs. Both proteins share a >99% sequence identity with the *H. sapiens* protein and an amino acid alignment is given in **Figure 3.7**

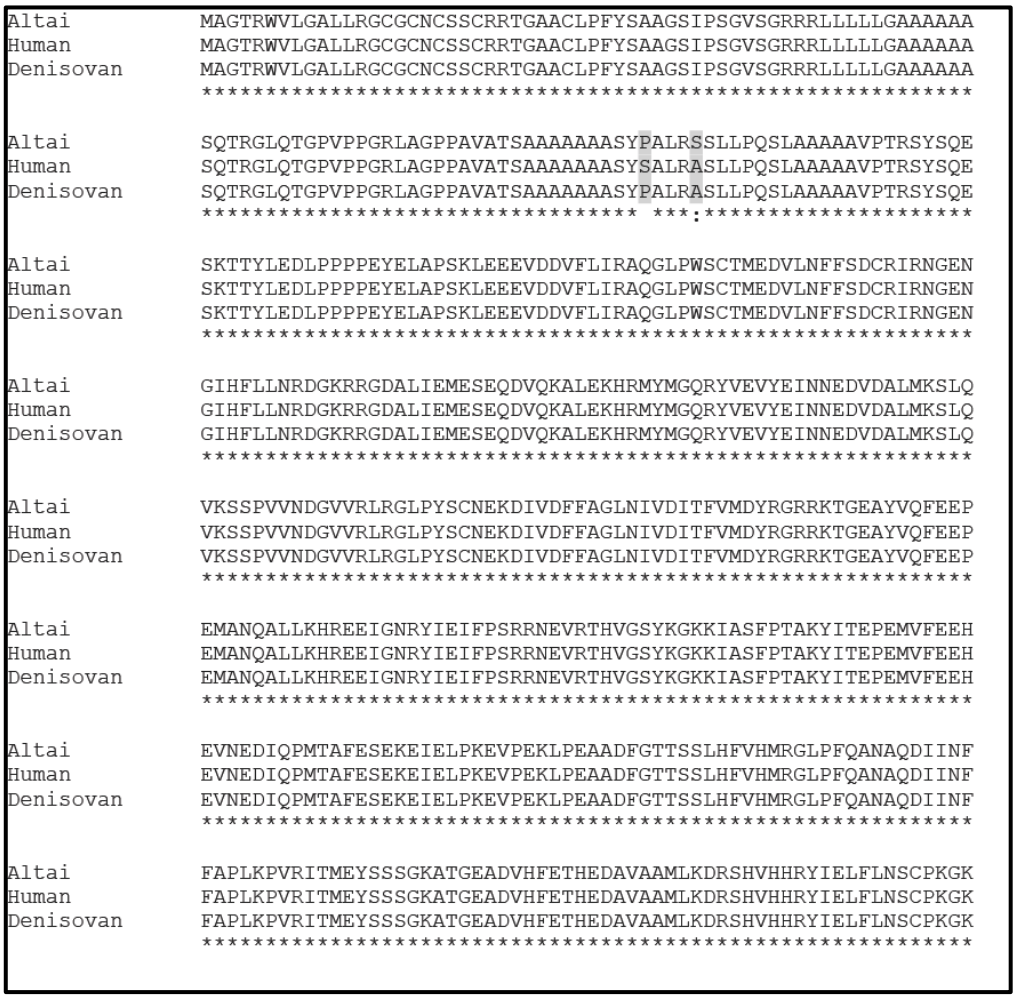

Figure 3.7. Sequence alignment of GRSF1 protein among *Homo sapiens H. neanderthalensis* and *H. denisovan*. The GRSF1 protein sequence of *H. neanderthalensis* and *H. denisovan* share >99% sequence identity with human GRSF1. In addition sequence alignment revealed two single nucleotide variants (Ser to Phe and Ala to Ser) in *H. neanderthalensis* and one in *H. denisovan* (Ser to Phe). The strictly conserved amino acid residues are indicated by asterisks.

Summary: GRSF1 frequently occurs in higher vertebrates including mammals. As modern humans (*H. sapiens*) the genomes of the extinct human subspeciies *H. neanderthalensis* and *H. Denisovans* involve a single GRSF1 gene and the corresponding protein share a >99% amino acid identity with the H. sapiens ortholog.

3.2.7. Structural conservation of GRSF1 protein in vertebrates

To infer whether the structure of the GRSF1 protein has been conserved during vertebrate development, we compared in detail the amino acid sequences of full-length human GRSF1 with the sequence of selected putative orthologs from different vertebrate

species (**Figure 3.8**). The domain architecture of GRSF1 is strictly conserved among the different species representing evolutionary stages. All GRSF1 proteins involve the three RNA-binding domains (qRRM1, qRRM2 and qRRM3), an Ala-rich domain and an acidic domain (**Figure 3.8**).

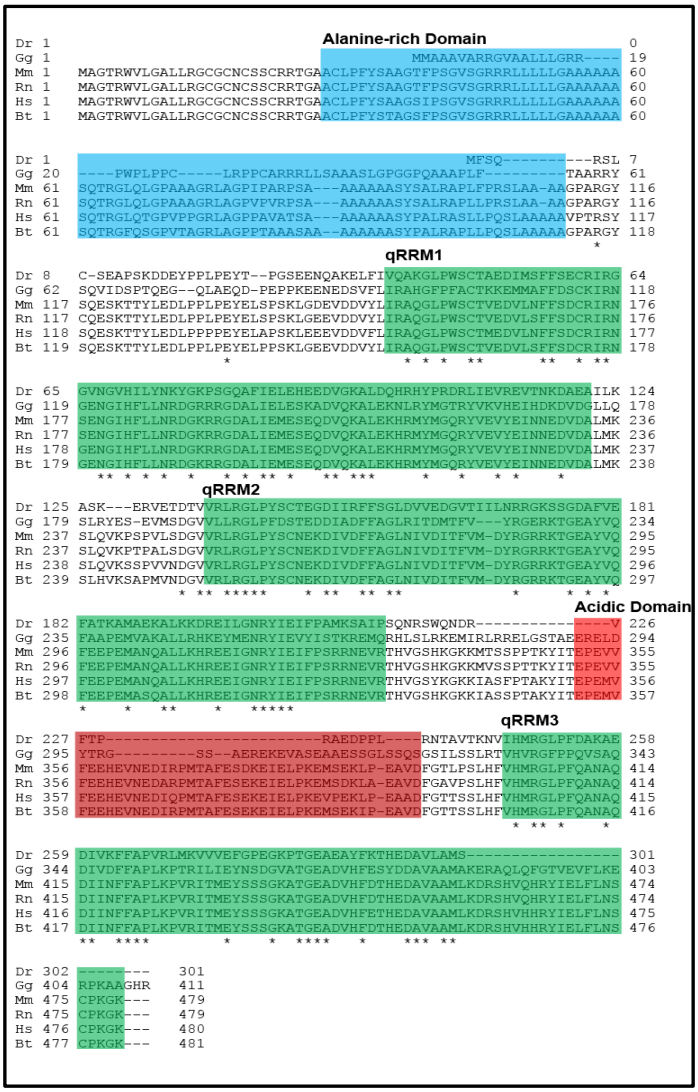

Figure 3.8. Alignment of GRSF1 proteins in various vertebrates. The domain organization of the GRSF1 is conserved in various vertebrate species such as *Danio rerio (Dr), Gallus gallus (Gg), Mus musculus (Mm), Rattus norvegicus (Rn), Homo sapeins (Hs)* and *Bos taurus (Bt).* The three qRRMs (highlighted in green) are conserved in all species. Alanine-rich domain (highlighted in blue) and acidic domain (highlighted in red) can also be found in all species. The strictly conserved amino acid residues are indicated by asterisks.

The three RNA-binding domains (qRRM1-3) share a medium degree of amino acid conservation across these species. In fact, the qRRM1 domains share a sequence identity of 37% over its 82 amino acid region. Similarly, the qRRM2 and qRRM3 domains share 34% and 29% sequence identity over 79 and 77 amino acids regions, respectively. These data suggest that the overall structures of RNA-binding elements are conserved during evolution and this is likely to be the case for the functional properties. In addition to these

structural similarities we observed a high degree of amino acid conservation in the Ala-rich-domain and in the acidic domain among cattle, mouse, rat and human GRSF1 (**Figure 3.8**). In contrast, the degree of amino acid conservation in these structural subunits between higher vertebrates (cattle, mouse, rat, human) and lower vertebrates (zebrafish, chicken) is limited (**Figure 3.8**). These data suggest that during vertebrate evolution a high evolutionary pressure selectively prevented structural alterations in the RNA-binding domains but not in the Ala-rich and the acidic domains.

Summary: The global structural architecture of GRSF1 has been conserved during vertebrate evolution. This amino acid conservation is high for the RNA-binding domains but considerably lower for the auxiliary domains (Ala-rich domain, acidic domain).

3.3. Biophysical characterization of the G-quadruplex structures in the 5'-UTR of GRSF1 substrates

In addition to the sequence of RNA substrates the RNA secondary structure may impact protein binding. This is mainly due to the fact that RNA easily folds into a multitude of secondary structures. In order to obtain more information on the characteristics of GRSF1 substrates, we investigated their secondary structure by circular dichroism (CD) spectroscopy (see **Section 2.2.2.9**). This experimental set-up allows the *in vitro* detection of G-quadruplex structures (G4) in RNA transcripts. For this purpose RNAs were transcribed *in vitro* and purified using size-exclusion chromatography, which is described in detail in Materials and Methods.

3.3.1. *In silico* prediction of potential G-quadruplexes in GRSF1 RNA substrates

It has been previously shown that the G-rich RNA substrates of hnRNP F can fold into G4 structure *in vitro* and the binding of hnRNP F qRRMs to the G-tract RNA induces melting of these structures (Cyril Dominguez et al., 2010). In contrast to that very little is known in this respect for GRSF1 substrates. To predict the ability of GRSF1 substrates to fold into G4 structures we analyzed the three RNA transcripts representing the 5'-UTR of mouse *GPx4*, human *GPx4* and mouse *Use1* using a web server called QGRS (quadruplex forming G-rich sequences) at http://bioinformatics.ramapo.edu/QGRS/index.php. This program generates information about the composition and the distribution of potential QGRS and provides a G-score as a readout parameter for the formation probability of G-quadruplexes (G4s) (Kikin et al., 2006; Menendez, Frees, & Bagga, 2012). Following this rationale, all three GRSF1 target mRNA sequences form G4 structures as indicated by their G-scores (**Figure 3.9 Panel A**). The

mouse *GPx4* RNA can form one G4 consisting of two G-tetrads with a G-score of 17 (**Figure 3.9 Panel A**). In contrast, the human *GPx4* mRNA forms one G4 with two G-tetrads and its G-score is 19 (**Figure 3.9 Panel A**). The mouse *Use1* RNA can form a single G4 with a similar number of G-tetrads and same G-score.

Summary: *In silico* analysis predicted G-quadruplex structures in the 5'-UTR regions of mouse *GPx4 mRNA*, human *GPx4* mRNA and mouse *Use 1 mRNA*.

3.3.2. Ability of the respective RNA sequences to fold into G-quadruplex structures *in vitro*

In order to verify whether or not the predicted 5'-UTR G-rich sequences adopted a G4 structure *in vitro*, we applied CD spectroscopy.

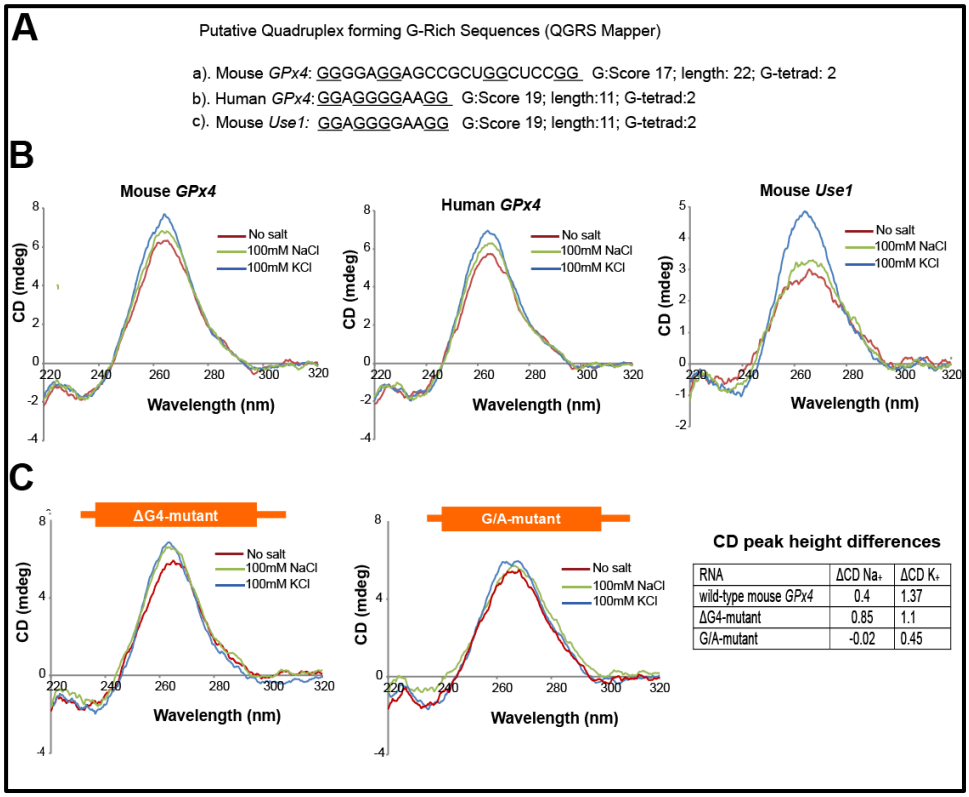

Figure 3.9. ***In vitro* characterization of RNA G-quadruplexes of GRSF1 substrates.** (**A**) Predicted G-quadruplex structures in the 5'-UTR of mouse *GPx4* (a), human *GPx4* (b) and mouse *Use1* (c) mRNA by QGRS mapper. Potential G4 forming nucleotides are underlined (**B**) CD spectroscopy measurements were performed using a 54-nt RNA probe of mouse *GPx4* wt sequence (5'-GGC-CUC-GCG-CGU-CCA-UUG-GUC-GGC-UGC-GUG-AGG-GGA-GGA-GCC-GCU-GGC-UCC-GGC-3'),human *GPx4* wt sequence (5'-GCC-GAC-GCG-CGU-CCA-UUG-GUC-GGC-UGG-ACG-AGG-GGA-GGA-GCC-GCU-GGC-UCC-AG 3') and 40-nt long mouse *Use1* 5'-UTR as previously described (Nieradka et al., 2014) in the absence or presence of salts (**Panel B and C**) at 10°C. (**C**) CD spectra of ΔG4 deletion mutant and G/A-mutant, these mutated RNA probes were synthesized by *in vitro* transcription and purified by size-exclusion chromatography (see Materials and methods for details). Comparison of peak height differences of CD spectra of wild-type, ΔG4 deletion mutant and G/A-mutant (**Panel C bottom right**). Peak height was calculated by measuring the CD wavelength at 264 nm (positive peak) and the obtained value was subtracted from the background (H_20).

For this purpose we first generated three RNA probes [mouse *GPx4* (54-nt), human *GPx4* (54-nt) and mouse *Use1* (40-nt)]. These RNA probes represented the sequence of the 5'-UTRs of respective mRNAs. They were *in vitro* transcribed and purified using size-exclusion chromatography (see Materials and Methods for details). We then subjected these synthesized RNA probes to CD spectroscopic analysis to determine the presence of G4 structures in these substrates. This technique has been very successfully used to detect the G4 structures possessing the typical spectrum caused by the topology of G4 structures (Beaudoin et al., 2014; Beaudoin & Perreault, 2010a; Jodoin et al., 2014). The RNA adopts a parallel G4 structure that provokes the appearance of positive peak at 264 nm and a negative peak at 240 nm (Malgowska, Czajczynska, Gudanis, Tworak, & Gdaniec, 2016). We monitored the CD spectra of each RNA probe in the absence of salts or in the presence of 100 mM of either Na^+ or K^+ (**Figure 3.9 Panel B**). These two monovalent cations are known to stabilize the G4 structures (Balasubramanian, 2014). Indeed all the cellular substrates [*GPx4* (mouse and human) and *Use1*] of GRSF1 showed CD spectra that represents parallel G4 formation (positive peak at 264 nm and a negative peak at 240 nm) (**Figure 3.9 Panel B**). In addition, these G4 structures are stabilized by K^+ and Na^+ ions. This stabilizing effect can be concluded from higher (264 nm) peak intensities in the presence of K^+ and Na^+ (**Figure 3.9 Panel B**). In contrast to that a reduction in peak intensities were observed in the absence of salts (**Figure 3.9 Panel B**). This effect was more pronounced for the *Use1* substrate (**Figure 3.9 Panel B**). Taken together, our *in vitro* data confirms the *in silico* prediction suggesting the presence of G4 structures within the 5'-UTR regions of mouse *GPx4*, human *GPx4* and mouse *Use1* mRNA substrates. In addition our data clearly shows that these G4 structures are stabilized by monovalent cations especially in presence of K^+ ion.

Summary: Our CD spectroscopic measurements indicate that the GRSF1 substrates (mouse *GPx4*, human *GPx4* and mouse *Use1*) fold into G-quadruplex structures.

3.3.3. Spectroscopic determination of G4 structures in mutant RNA constructs of GRSF1

We previously showed that the *Use1* RNA adopted a G4 structure and the central $A(G)_4A$ element is part of this structure (Nieradka et al., 2014). Next, we investigated which nucleotides within the *GPx4* RNA substrate are involved in G4 formation. For this purpose in addition to the wild-type (wt) RNA (**Figure 3.9 Panel B Top left**), a ΔG4 deletion mutant and a G-to-A- exchange mutant were constructed (**Figure 3.9 Panel C Top left** and **Right**). These RNAs were derived from the 5'-UTR of *GPx4* substrate and were *in vitro* transcribed. The ΔG4-mutant was generated by the precise deletion of the central four guanines of the G-rich motif. In contrast, the G-to-A exchange mutant was constructed by

replacing the six key guanines by adenines. To further investigate, the effect of the ΔG4 deletion and the G-to-A exchange on G4 formation, we analyzed the wt and mutant constructs by CD spectroscopy. The CD spectrum of wt showed a positive peak at 265 nm and a negative peak at 240 nm (**Figure 3.9 Panel B Top left**) indicative of a parallel G-quadruplex. Surprisingly, no significant differences in peak heights were observed between the CD spectra of wt and ΔG4-mutant (**Figure 3.9 Panel C bottom right**). This data suggest that the central four guanines of the A(G)$_4$A element do not participate in G4 formation. In contrast, the CD spectra of the G-to-A-exchange mutant differed significantly from that of the wt construct (**Figure 3.9 Panel C bottom middle**). Here a significant increase the CD-signal was found (**Figure 3.9 Panel C bottom right**). This finding suggests that the G-to-A mutations of the central G-rich motif prevent the G4 formation.

Summary: The central G-rich motif A(G)$_4$A may not be essential for G-quadruplex formation. However, G-to-A exchange impacts G4 formation.

3.3.4. Characterization of interactions between RNA mutants and GRSF1 protein

Encouraged by the results indicating that the G-to-A substitutions impaired G4 formation *in vitro* in the *GPx4* RNA substrate probe, we next measure the binding parameters. For this purpose mutated RNA substrates (ΔG4-mutant and G-to-A exchange mutant) were used as probes and we compared the binding affinities of wild-type and mutant RNA probes to recombinant GRSF1 employing our RNA electrophoretic mobility shift assays. We first used the wt RNA probe (54-nt) containing the A(G)$_4$A motif and determined a Kd-value of 66 nM. This data suggests a high binding affinity (**Figure 3.10 Panel A Top left**).

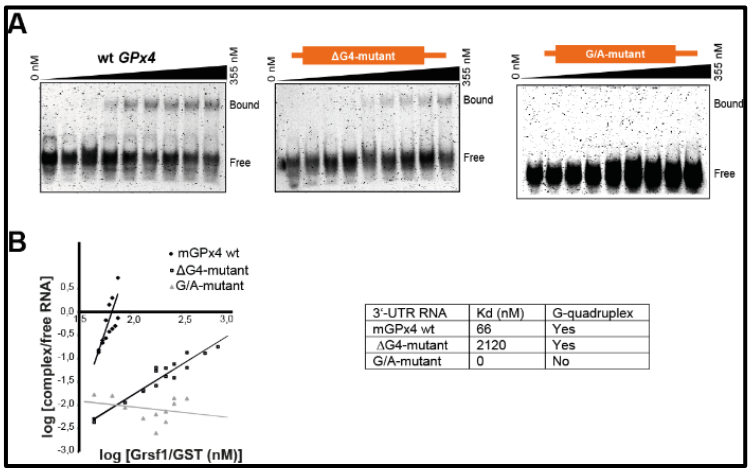

Figure 3.10. Determining the binding affinity of GRSF1 with different *GPx4* RNA probes. (**A**) The wt RNA probe harboring A(G)$_4$A motif, *GPX4* ΔG4 probe lacking central four guanines in the A(G)$_4$A element and G/A-mutant containing six G-to-A substitutions were used to measure the binding affinities (Kd) with the recombinant full-length GRSF1 protein. For this purpose these digoxigenin labeled RNA probes was incubated *in vitro* with different amounts of purified recombinant GRSF1-GST full-length. Aliquots of this incubation

mixture (15 µl) were loaded on a 5% polyacrylamide gel (native conditions) and northern blots of separated protein-RNA complexes were stained for digoxigenin to visualize the RNA band-shift signal as shown. The blots were visualized as described in the Materials and Methods. (see **Section 2.2.2.8**). (**B**) Quantification of GRSF1-GST binding to the wt RNA probe, *GPX4* ΔG4 probe and G/A-mutant RNA probe. The ratio of signal intensities of the shifted complex/free probe was plotted against the GRSF1-GST concentration. The intercept with the X-axis represent the Kd (logarithmic scale) (Nieradka et al., 2014).

Next, we used the ΔG4 deletion mutant as probe, which still forms a G4 structures according to our CD measurements (**Figure 3.9 Panel C Middle left**). Here we determined the Kd-value of 2120 nM for this RNA construct. Thus, deletion of the cognate GGGG binding motif in the $A(G)_4A$ element decrease RNA-binding by 30-fold (**Figure 3.10 Panel A Top middle**). We then used the RNA probe (G-to-A-mutant), in which six guanines were replaced by adenines. Interestingly, with this substrate GRSF1 binding was completely abolished (**Figure 3.10 Panel A Top right**). These findings indicate that both, deletion of the central G-rich sequence and the G-to-A exchange strongly impact GRSF1 binding.

Summary: Deletion of the central G-rich sequence in the $A(G)_4A$ motif and the G-to-A exchange strongly impact GRSF1 binding.

3.4. Molecular mechanisms of GRSF1-RNA interactions

The molecular mechanisms of the interaction of GRSF1 with RNA have not been characterized so far in detail. In order to shed light on the structural basis of RNA-binding by GRSF1, we first created two sets of GRSF1 variants: i) To judge the relative contributions of the three different RNA-binding domains (qRRM1, qRRM2 and qRRM3) we expressed these domains individually as separated GST-tag fusion proteins (Section 3.1.3) ii) To explore the impact of the different GRSF1 subunits we expressed a series of truncation mutants that lacked individual RNA-binding domains. For instance, human GRSF1, which lacks the qRRM1 domain was termed ΔR1-hGRSF1 (Δ stands for deletion). Similarly, the construct lacking the qRRM2 domain was named ΔR2-hGRSF1. Finally, the construct lacking the qRRM3 domain was named ΔR3-hGRSF1 (**Figure 3.12 Panel B**). Moreover, we designed a truncation construct that lacked alanine-rich domain (ΔAla-hGRSF1) and an additional double truncated construct lacking both, the alanine-rich (Ala-rich) and the acidic domain (AD) (ΔAlaΔAD-hGRSF1) (**Figure 3.12 Panel B**). All these constructs were recombinantly expressed as GST-tag fusion proteins in *E.coli* and purified to near homogeneities by affinity chromatography on a glutathione agarose column (see the Section 3.1 for details). The purified proteins were used to test their RNA-binding affinities, which were determined by quantitative RNA electrophoretic mobility shift assays

(qREMSAs). For this purpose, we used a digoxigenin-labeled RNA probe (54-nt) that represents the 5'-UTR of human *GPx4* mRNA and contained the G-rich element $A(G)_4A$ motif (**Figure 3.12 Panel A**). The corresponding RNA probes were *in vitro* transcribed and purified using a micro spin column-based purification procedure (see Materials and Methods for details).

3.4.1. Characterization of RNA substrates of GRSF1 protein

Although a few different RNAs have been shown to bind to GRSF1 only two of them have been characterized in more detail. One of these substrates is the mRNA of murine *GPX4* (Ufer et al., 2008). For this RNA a Kd of 40 nM has previously been determined (Ufer et al., 2008). The other RNA is derived from the *Use1* gene. The Kd of GRSF1 for this RNA is somewhat higher (527 nM) suggesting a lower binding affinity (Nieradka et al., 2014). Since *GPx4* mRNA is the better substrate for GRSF1, we used the human ortholog for our experiments that were focused on understanding the molecular mechanisms of GRSF1-mRNA interactions. Thus, we initially measured the Kd of human GRSF1 for a 54 nucleotide fragment of the human *GPx4* mRNA that contains the central AGGGGA motif (**Figure 3.12 Panel A**). Interestingly, we determined a Kd of about 555 nM (**Table 3.3**), which is more than 10-fold higher than the Kd-value reported for murine *GPx4* mRNA (40 nM) (Ufer et al., 2008). The variable binding affinities of human GRSF1 to similar RNA substrates that involve the central $A(G)_4A$ motif was rather surprising. This data therefore indicates that other substrate characteristics affect GRSF1 binding and these parameters are discussed in more detail in discussion (see **Discussion 4.4**).

Summary: GRSF1 protein binds to G-rich (AGGGGA) RNA substrates with different affinities suggesting that in addition to the G-rich motif other substrate characteristics affect GRSF1 binding.

3.4.2. RNA-binding properties of individually expressed qRRM1, qRRM2 and qRRM3 domains of human GRSF1

To gain functional insights into the molecular mechanisms of RNA-binding by human GRSF1 we first individually expressed and purified the three qRRM domains [qRRM1 (residues 139-244), qRRM2 (residues 252-323) and qRRM3 (residues 400-480)] of human GRSF1 and tested the ability of each qRRM domain to bind *GPx4* RNA using qREMSA. For this purpose, a purified sequence of human *GPx4* mRNA (54-nt) containing the central AGGGGA element (**Figure 3.12 Panel A**) was used as substrate probe. We first, measured the binding affinity of the qRRM1 domain. Surprisingly, we did not observe any shift signal and these data suggested that there is no high affinity interaction between qRRM1 and the substrate probe (**Figure 3.11**). We next separately tested the ability of the

qRRM2 and qRRM3 domains of human GRSF1 to bind the RNA probe. Here again, we did not observe any shift signal (data not shown). From these results one may conclude that the three RNA-binding domains of human GRSF1 do not exhibit a high RNA-binding affinity when separately expressed.

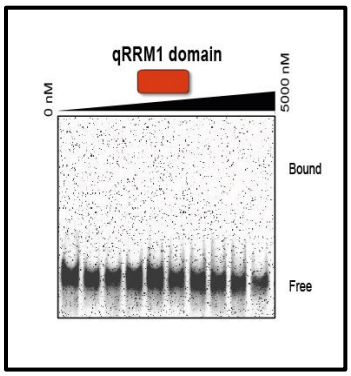

Figure 3.11. RNA-binding affinity of qRRM1 domain of GRSF1 with human *GPx4* RNA probe (54-nt). The individual qRRM1 domain (amino acids 139-244) of human GRSF1 was used to measure the binding affinity (Kd) with the 5'-UTR of human *GPx4* RNA probe using RNA electrophoretic mobility shift assays. For this purpose a digoxigenin labeled RNA probe harboring $A(G)_4A$ element was incubated *in vitro* with different amounts of purified recombinant qRRM1 domain and aliquots of this incubation mixture (15 µl) were loaded on a 5% polyacrylamide gel (native conditions) and northern blots of separated protein-RNA complexes were stained for digoxigenin to visualize the RNA band-shift signal as shown.

Summary: The three separated qRRM domains (qRRM1-3) of human GRSF1 do not bind RNA individually.

3.4.3. RNA-binding activities of the different GRSF1 truncation constructs

GRSF1 has been shown previously to bind with high affinity to mRNA substrates containing G-rich elements (Kash et al., 2002; Nieradka et al., 2014a; Ufer et al., 2008). However the relative contributions of the different structural subunits (domains) remained elusive. In order to evaluate the contribution of different structural domains for RNA-binding we created a series of truncation constructs (**Figure 3.12 Panel B**) and tested the RNA-binding affinities of these constructs employing qREMSAs (**Figure 3.12 Panel C**). First, we investigated the binding properties of the full-length human GRSF1 to the human *GPx4* mRNA probe. We determined a Kd of 629 nM (**Table 3.3** and **Figure 3.12 Panel C Top left**) suggesting high affinity binding. We next tested the N-terminal deletion construct lacking only the alanine-rich domain (amino acids 26-111). In this construct, the three RNA-binding domains (qRRM1-3) and acidic domain (AD) (**Figure 3.12 Panel B**) were preserved. This truncated construct was denoted as ΔAla-hGRSF1. The N-terminal Ala-

rich domain of human GRSF1 was recently suggested to serve as a mitochondrial targeting signal (Jourdain et al., 2013) and thus, this domain may not contribute to RNA-binding. Here we obtained a Kd-value of 555 nM for this truncated version (**Table 3.3** and **Figure 3.12 Panel C Top middle**). Thus, the mRNA binding affinity of this truncation construct is very similar to that of full-length human GRSF1 (629 nM). These results demonstrate that deletion of the Ala-rich domain has almost no effect on RNA-binding and therefore this domain may not be involved in RNA- binding. Thus, our results support the previous suggestion that the Ala-rich domain does not play a major role for the RNA interaction (Jourdain et al., 2013).

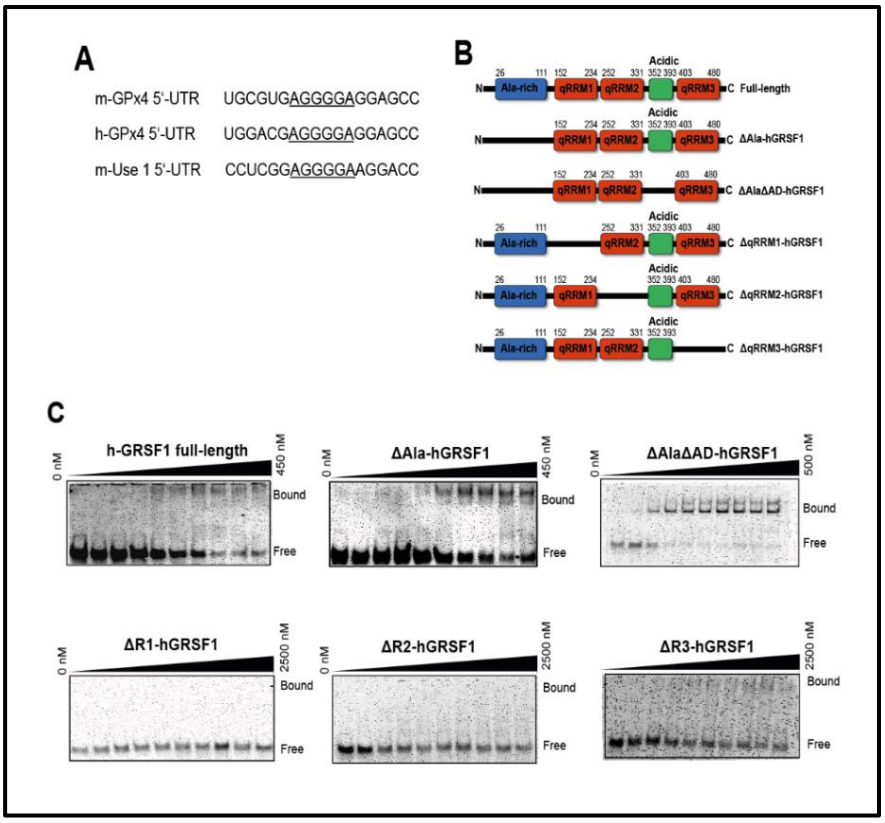

Figure 3.12. RNA-binding affinities of different domains of GRSF1 with human *GPx4* RNA probe (54-nt). (**A**) The A(G)$_4$A element of the mouse *GPx4*, human *GPx4* and mouse *Use1*. (**B**) Schematic representation of the GST-tagged full-length GRSF1 and its truncation constructs. The domains and amino acid coordinates are indicated. (**C**) The indicated full-length and additional constructs, were used to measure the binding affinities (Kd) with the 5'-UTR of human *GPx4* RNA probe using RNA electrophoretic mobility shift assays. For this purpose a digoxigenin labeled RNA probe harboring A(G)$_4$A element was incubated *in vitro* with different amounts of purified recombinant GRSF1-GST full-length and truncation proteins as indicated. Aliquots of this incubation mixture (15 µl) were loaded on a 5% polyacrylamide gel (native conditions) and northern blots of separated protein-RNA complexes were stained for digoxigenin to visualize the RNA band-shift signal as shown. The blots were visualized as described in the Materials and Methods. (see **Section 2.2.2.8**).

We next wanted to explore whether acidic domain (amino acids 352-393) might play any role in RNA-binding. This domain is located near the C-terminal region of GRSF1

and interconnects the RNA-binding domain qRRM2 with qRRM3 (**Figure 3.12 Panel B**). Although, the exact function of this domain has not be explored in detail (Qian & Wilusz, 1994; Ufer, 2012) this structural element has been implicated in protein-protein interactions (Qian & Wilusz, 1994). To gain further insights into the function of AD, we created a deletion construct that lacked both AD and Ala-rich domain but retained the three qRRM domains (qRRM1-3). This construct was denoted as ΔAlaΔAD-hGRSF1 (**Figure 3.12 Panel B**). Surprisingly, we obtained a Kd of 194 nM for this construct (**Table 3.3** and **Figure 3.12 Panel C Top right**). This data suggests that the double deletion protein exhibits a 3-fold higher RNA-binding affinity, when compared with full-length human GRSF1 (629 nM). From these results one may conclude that the AD domain downregulates the RNA-binding affinity of human GRSF1.

Protein	Kd (nM)
Full-length GRSF1	629
ΔAla-hGRSF1	555
ΔAlaΔAD-hGRSF1	194
ΔqRRM1-hGRSF1	n.d.
ΔqRRM2-hGRSF1	n.d.
ΔqRRM3-hGRSF1	2695

Table 3.3 RNA-binding affinities of the full-length GRSF1 and its truncation constructs with human *GPx4* RNA probe (54-nt). Quantitative RNA electrophoretic mobility shift assays (qREMSA) were carried out to determine the dissociation constant (Kd) for the full-length and different truncation constructs as shown in the (**Fig. 3.12 Panel B**), n.d. signifies not detected. For this purpose different amounts of purified recombinant GST-GRSF1 and its truncation constructs were separately incubated *in vitro* with a digoxigenin labeled RNA probe (**Fig. 3.12 Panel C**) that represents the 5'-UTR of human *GPx4* mRNA containing the A $(G)_4$A motif (**Fig. 3.12 Panel A**). The RNA-binding affinities were estimated as shown previously (Nieradka et al., 2014a), by quantifying (free RNA versus bound RNA) using Image J quantification software.

The most notable feature of the GRSF1 protein is the presence of the three RNA-binding domains (RBDs) (qRRM1, qRRM2 and qRRM3) (Ufer, 2012). However, the relative contributions of these structural subunits to RNA-binding have not been explored. In quest to identify which of these qRRM domains may contributes to RNA-binding and which might not, we constructed different truncation mutants lacking the individual RBDs qRRM1 (ΔR1-hGRSF1), qRRM2 (ΔR2-hGRSF1) and qRRM3 (ΔR3-hGRSF1) (**Figure 3.12 Panel B**). Deletion of the qRRM1 and qRRM2 domains completely abolished the RNA-binding activity (**Table 3.3** and **Figure 3.12 Panel C Bottom left** and **middle**) of human GRSF1. No shift signals could be detected in the RNA shift mobility assays. Deletion of the qRRM3 domain increased the apparent Kd more than 4-fold (**Table 3.3** and **Figure 3.12**

Panel C Bottom right). Collectively, these data suggests that all three qRRM domains (qRRM1, qRRM2 and qRRM3) are needed for high affinity RNA-binding.

Summary: The Ala-rich domain of human GRSF1 is not involved in RNA-binding but acidic domain negatively regulates this enzyme property. All three RNA-binding domains (qRRM1-3) contribute to RNA-binding and qRRM1 and qRRM2 are essential.

3.4.4. Defining the minimum RNA sequence motif required for GRSF1 binding

In order to resolve the minimal *GPx4* mRNA sequence required for GRSF1 binding, we constructed three truncated RNA probes representing the region that is located close to the translational initiation site of human *GPx4* mRNA (**Figure 3.13 Panel A**).

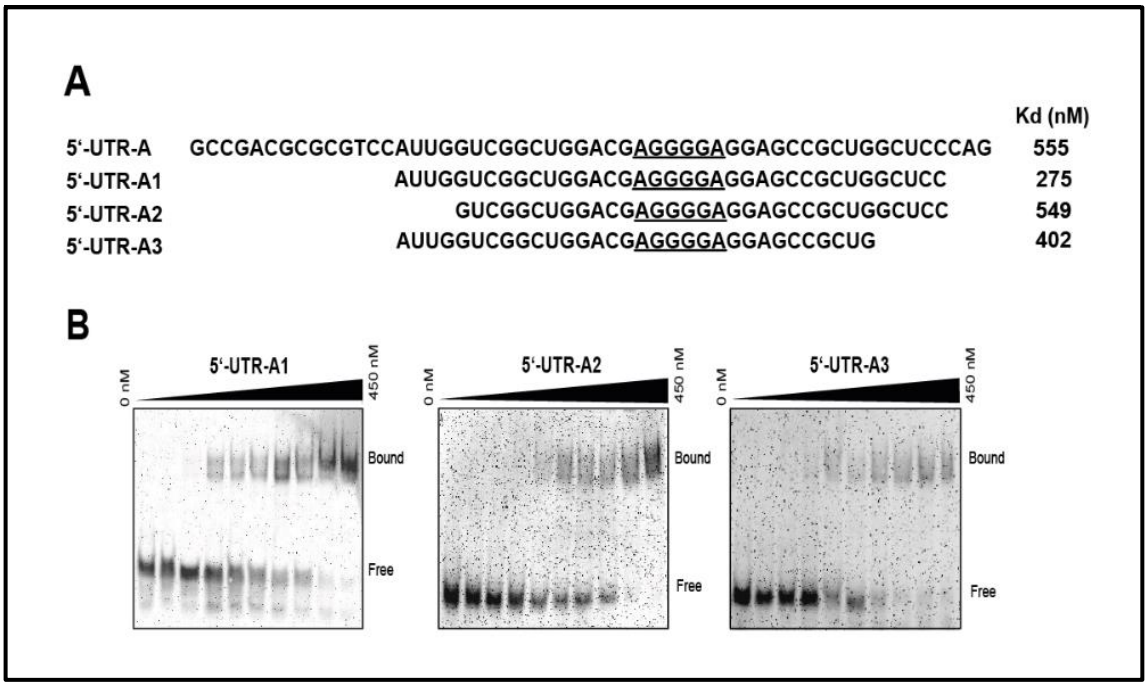

Fig. 3.13. Truncation of the GRSF1 binding motif in the 5'-UTR of the human *GPx4* RNA. (**A**) To define the GRSF1 binding motif different mutant RNA probes were designed. These probes that contain the $A(G)_4A$ motif ($A(G)_4A$ is underlined) were incubated *in vitro* with purified wild-type ΔAla-hGRSF1-GST and the shift signals were analyzed as described in **Fig 3.12 Panel C**. The RNA-binding affinities (Kd-values) of these RNA mutants were estimated as shown previously (Nieradka et al., 2014), by quantifying (free RNA versus bound RNA) using Image J quantification software. (**B**) Gel shift pattern of the different probes using purified recombinant ΔAla-hGRSF1-GST fusion protein.

These RNA probes were generated by *in vitro* transcription (see Materials and Methods for details) and their binding to the Ala-rich domain deficient variant of human GRSF1 (ΔAla-hGRSF1) was tested by qREMSAs (**Figure 3.13 Panel B**). When we first used the RNA fragment (named as 5'UTR-A1 [37-nt]), we obtained a Kd of 275 nM (**Figure 3.13 Panel A**), which indicates that this RNA fragment is a suitable substrate for

human GRSF1 (**Figure 3.13 Panel B Top left**). Next, we employed two shorter RNA probes (named as 5'UTR-A2 [33-nt] and 5'UTR-A3 [32-nt] respectively) and here we calculated similar Kd-values (**Figure 3.13 Panel A** and **Panel B middle** and **Top right**). From these results one may conclude that all RNA probes, which involved the central G-rich sequence, where suitable GRSF1 substrates. The flanking sequences of this motif may regulate the binding affinity but are obviously not essential.

Summary: Subtle truncation of the RNA probes (**Figure 3.13**) did not dramatically impact GRSF1 binding. Thus, the flanking sequences of the central G-rich motif may not be of major functional importance.

3.5. Functional characterization of naturally occurring genetic variations of the human GRSF1

3.5.1. Structural models of the RNA-binding domains of human GRSF1 and identification of RNA recognizing residues

The 3D structure of full-length human GRSF1 is not known and the structural basis for GRSF1-RNA interaction has not been characterized so far. However, the structure of the related protein hnRNP F has been studied in detail (C. Dominguez, 2006; Cyril Dominguez et al., 2010). Since the RNA-binding domains of human GRSF1 and hnRNP F share a medium degree of amino acid sequence identity homology modeling of the GRSF1 RNA-binding domains was carried out. Both proteins contain three independent RNA-binding domains. In addition to the RNA-binding domains GRSF1 contains an N-terminal alanine-rich domain, which is missing in hnRNP F. For our functional RNA-binding studies this structural subunit was deleted, since this peptide sequence does not impact RNA-binding (Jourdain et al., 2013; Ufer et al., 2008). Moreover, the glycine-rich domain separating qRRM2 and qRRM3 in hnRNP F is replaced in GRSF1 by an acidic domain. In order to study the relative importance of the different domains of human GRSF1 for RNA-protein interaction we employed comparative protein structure modeling (**Figure 3.14**).

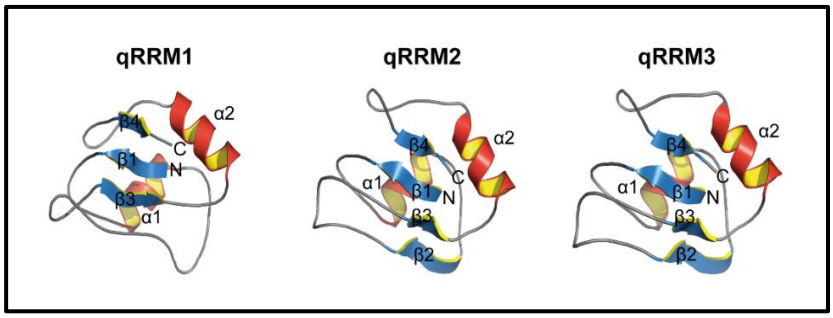

Figure 3.14. Structural models of the three human GRSF1 RNA-binding domains. Structural models were established for the three RNA-binding domains of human GRSF1 on the basis of the NMR structure of the corresponding motifs of hnRNP F. Cartoon representation of the structures of qRRM1, qRRM2 and qRRM3 are given and the secondary structural elements are annotated.

The modeled structures of the three GRSF1 RNA-binding domains are well defined and adopt the canonical $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$-fold, which consists of four anti-parallel β-sheets (β_1-β_4) packed against two α-helices (α_1, α_2) (**Figure 3.14**). Loop regions (loops 1-5) of variable lengths interconnect the α-helices and β-sheets. Our modeling data revealed that the qRRM2 and qRRM3 domains adopt a rather stable 3D conformation. In contrast, qRRM1 was predicted to adopt a range of different conformations and no β_2 sheet could be identified using our approach.

In order to identify the amino acid residues that are important for RNA recognition, we performed sequence alignments of the RNA-binding domains of GRSF1 and hnRNP F. As previously reported hnRNP F binds RNA in a unique way involving aromatic amino acid residues located in the connecting loops 1 (W20, F120, Y298), positively charged residues in loop 3 (R52, K150, R326), aromatic residues in loop 5 (Y82, Y180, Y356) and finally an arginine (R) located in β1 (R16 in qRRM1, R116 in qRRM2, and R294 in qRRM3) (Cyril Dominguez et al., 2010). Most of the RNA-binding amino acids are strictly conserved in GRSF1 or are conservatively replaced by similar amino acid (**Figure 3.15**). In addition, a number of adjacent amino acids are also involved in stabilizing protein-RNA interactions and these residues are also highly conserved (**Figure 3.15**). However, despite the striking similarities between the two proteins there are remarkable structural differences. For instance, the charged arginine in loop 1 is replaced by a neutral but polar glutamine (Q155) in qRRM1 of GRSF1. Furthermore, the polar Thr found in loop 1 in all three RNA-binding domains of hnRNP F is replaced by Asn in qRRM2 (N262) and qRRM3 (N413) of human GRSF1.

qRRM1	GRSF1	Q_{155}	L_{157}	W_{159}	***K*** $_{191}$	R_{214}	R_{220}	Y_{221}	E_{223}	***Y*** $_{225}$
	hnRNP F	R_{16}	L_{18}	W_{20}	***R*** $_{52}$	R_{75}	R_{81}	Y_{82}	E_{84}	***F*** $_{86}$
qRRM2	GRSF1	R_{255}	L_{257}	N_{262}	***Y*** $_{259}$	***R*** $_{287}$	***R*** $_{311}$	R_{317}	Y_{318}	E_{320}	F_{322}
	hnRNP F	R_{116}	L_{118}	T_{123}	***F*** $_{120}$	***K*** $_{150}$	***K*** $_{173}$	R_{179}	Y_{180}	E_{182}	F_{184}
qRRM3	GRSF1	R_{406}	L_{408}	***F*** $_{410}$	N_{413}	***K*** $_{438}$	R_{461}	R_{467}	Y_{468}	F_{472}	
	hnRNP F	R_{294}	L_{296}	***Y*** $_{298}$	T_{301}	***R*** $_{326}$	R_{349}	R_{355}	Y_{356}	F_{360}	

Figure 3.15. Conserved RNA-binding amino acids in GRSF1 and hnRNP F. Amino acid involved in RNA-binding of hnRNP F are indicated and the corresponding residues in human GRSF1 are aligned. Strictly conserved amino acids are indicated in bold, conservative mutations in italic bold.

Summary: The three qRRMs (qRRM1-3) of GRSF1 adopt $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$-fold however, β_2 sheet is not present in qRRM1 domain. Moreover, the amino acid residues that recognize RNA are present in the loop 1, 3 and 5 and not on the β sheets.

3.5.2. Identification of non-synonymous mutations in the RNA-binding domains of human GRSF1

In order to explore how the genetic variability in the *GRSF1* gene impacts the RNA-binding capacity of the corresponding protein we first identified naturally occurring mutations in the GRSF1 gene by screening the genomic sequences deposited in various online databases (http://www.internationalgenome.org, http://exac.broadinstitute.org, https://esp.gs.washington.edu/drupal/, https://www.ncbi.nlm.nih.gov/SNP). These searches identified a total of 294 nucleotide mutations. These nucleotide exchanges were distributed over the entire coding region of the *GRSF1* gene. However, an allele frequency of >1%, which would classify these mutations as single nucleotide polymorphisms, was only observed for a single amino acid exchange and this polymorphism causes a Ser95Pro exchange within the N-terminal alanine-rich domain. All other amino acid exchanges detected must thus be classified as rare mutations. We next quantified the amino acid polymorphisms in each of the RNA-binding domains and found a total of 44, 32 and 37 polymorphisms in the qRRM1 (amino acids 152-234), the qRRM2 (amino acids 252-331) and the qRRM3 (amino acids 403-480) domains, respectively. Moreover, the alanine-rich domain (amino acids 26-111) and the acidic-domain (amino acids 352-393) located between qRRM2 and qRRM3 contain a total of 11 and 26 distinct amino acid exchanges, respectively. Next, we screened the amino acid residues, which have been implicated in RNA recognition, for the presence of amino acid polymorphisms. Here we identified two of such amino acid exchanges in the qRRM2 domain. In the qRRM1 and qRRM3 domains a total of four and five polymorphisms were identified, respectively. None of the amino acid mutations in GRSF1 have been related to the pathogenesis of any major human disease.

Summary: *In silico* genomic database screening identified a total of 294 naturally occurring genetic variations, however with the exception of Ser95Pro none of the other variations had the allele frequency >1%.

3.5.3. Functional consequences of amino acid exchanges in the three qRRM domains of human GRSF1

Next, we tested the functional significance of amino acid variations in the RNA-binding domains of human GRSF1. For more than 40% of all amino acid residues located in the three RNA-binding domains (26 amino acids) amino acid variations could be identified (11 amino acid exchanges). These mutations were: i) Q155R, T162S, R188L and R214H in qRRM1, ii) Y318C and F322S in qRRM2 and iii) F410S, K438N, R461L, Y468C and F472L in qRRM3. Since these mutations can modify the RNA-binding capacity of GRSF1 we tested the impact of these amino acid exchanges on the RNA-binding

capabilities of recombinant human GRSF1. For this purpose we expressed wild-type human GRSF1 lacking the N-terminal alanine-rich domain (ΔE1-hGRSF1) and the corresponding mutants in *E.coli*, purified the recombinant proteins by affinity chromatography on GSH-agarose and employed them for *in vitro* RNA-binding assays (RNA electrophoretic mobility shift assays).

First, we explored the impact of Q155R exchange on the RNA-binding capacity of GRSF1. When we mutated the neutral but polar glutamine-155, which is located according to homology modeling in loop 1 of the qRRM1 RNA-binding domain, to a negatively charged arginine (non-conservative amino acid exchange), we observed a decrease in the RNA-binding capacity (**Figure 3.15**).

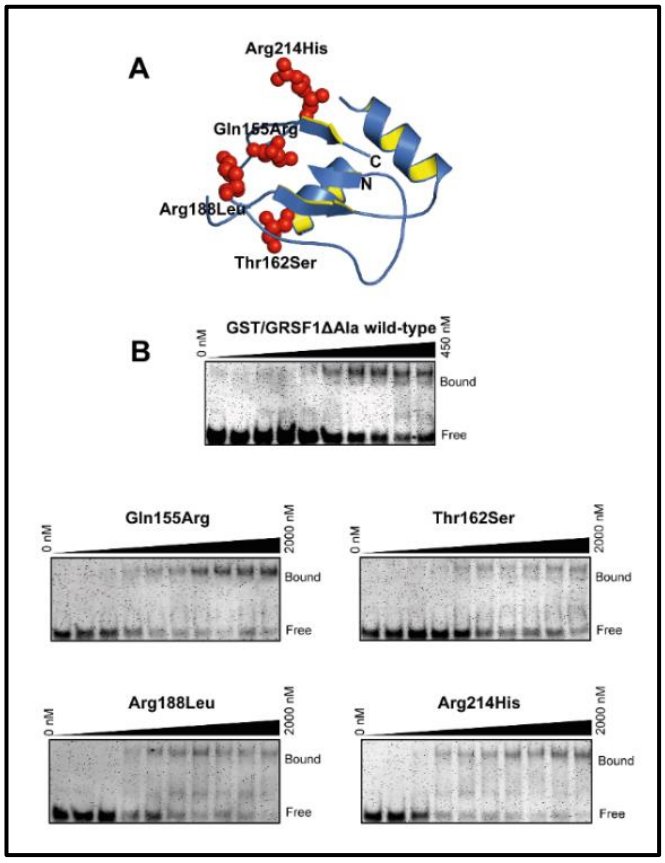

Figure 3.15 Impact of naturally occurring mutants in the qRRM1 RNA-binding domain of human GRSF1 on RNA-binding affinity. (**A**) Localization of the mutated amino acids in the 3D model of the qRRM1 RNA-binding domain. The mutated residues are shown as red spheres. (**B**) Quantitative RNA electrophoretic mobility shift assays (REMSA) were carried out to determine the dissociation constant (Kd) for the GRSF1-probe complex. For this purpose different amounts of purified recombinant wild-type and mutant protein was incubated with a digoxigenin labeled probe representing the 5'-UTR of human *GPx4* containing the $A(G)_4A$ motif. Each gel is one representative experiment of at least three independent experiments.

In fact, the Kd-value increased from 555 nM for wild-type GRSF1 to 1336 nM for the mutant protein. These data suggest that the positive charge of R155 may be relevant for RNA-binding. Next, we explored the impact of the T162S exchange on the RNA-binding

capacity of GRSF1. T162 is also located in loop 1 of the qRRM1 RNA-binding domain (**Figure 3.15 A**) and constitutes an uncharged but polar amino acid, which carries a free OH-group. As hydrogen donor it may be involved in the formation of hydrogen bridges. T162S mutation drastically reduces the affinity of GRSF1 to the RNA probe (Kd of 2649 vs. 555 nM of the wild-type protein) (**Table 3.4**). In the naturally occurring T162S mutant Thr-162 is replaced by a serine residue that also carry an OH-group and thus may also function as hydrogen donor. Conservation of this amino acid property together with the drop of RNA-binding affinity suggest that the formation of a hydrogen bridge between this amino acid and the RNA substrate may not play a major role in protein-RNA interaction. Taken together, our mutagenesis data at Q155 and T162 implicate these two amino acids in protein-RNA interaction of human GRSF1. This conclusion is consistent with corresponding NMR data on hnRNP F implicating R16 in protein-RNA interaction (Cyril Dominguez et al., 2010).

Next, we tested the impact of the two other mutations (R188L and R214H) localized in the qRRM1 RNA-binding domain. These two positively charged arginine residues are located in loop 3 and loop 5 of this RNA-binding motif. The R188L exchange represents a strongly non-conservative mutation since a positively charged amino acid is replaced with a neutral residue. In contrast, the R214H exchange is somewhat more conservative since the positive charge is retained. For the R188L mutant the functional RNA binding assays revealed that the mutant proteins exhibit similar binding affinity as wild-type GRSF1 (**Figure 3.15**). This data suggests that R188L exchange may not alter RNA-binding capacity of GRSF1 (**Table 3.4**) despite the fact that a positive charge is removed.

Domain	Variation	Variation ID	Kd (nM)	SD	p	n
	Wild-type		555	297		14
qRRM1	Gln155Arg	rs775148299	1336	1108.6	0.006	7
	Thr162Ser	rs115135136	2649	1004.7	< 0.001	5
	Arg188Leu	rs761151698	890	413	0.067	6
	Arg214His	rs368531595	632	277.7	0.689	3
qRRM2	Tyr318Cys	rs771152420	1121	484.4	0.010	4
	Phe322Ser	rs747049473	2564	1221.5	<0.001	4
qRRM3	Phe410Ser	rs779804453	760	294	0.295	3
	Lys438Asn	rs186328559	793	239.7	0.218	3
	Arg461Leu	rs773565301	757	401.7	0.282	4
	Tyr468Cys	rs369348124	1108	392.8	0.003	6
	Phe472Leu	rs762126036	755	218.8	0.295	3

Table 3.4. Binding affinity of naturally occurring human GRSF1 variants to a human *GPx4* mRNA. GRSF1 variants were expressed in *E.coli* and purified as described in Materials and Methods. The Kd-values

for the different recombinant proteins were determined by quantitative electrophoretic mobility RNA shift assays. In each experiment the Kd values were calculated from 8-10 different experimental data points and each experiment was carried out at least in triplicate. Statistic significance (p, 2-sided) for the comparison of the Kd values of the mutant proteins with wild-type GRSF1 is given.

Thus, it may be concluded that the charge of R188 may not be important for RNA-protein interaction. R214H exchange did neither alter the RNA-binding capacity and thus, no functional consequences are expected even for homozygous carriers of this mutant allele. However, in this case it might still be possible that a less conservative mutation, which inverses the charge of this residue (e.g. H214D or H214E), would have an impact on RNA-binding affinity.

Functional consequences of amino acid exchanges in the qRRM2 domain of human GRSF1– Our database searches revealed two naturally occurring mutations in the qRRM2 domain (Y318C, F322S) of human GRSF1. Y318 is an aromatic non-polar residue located in loop 5 of qRRM2 (**Figure 3.16**).

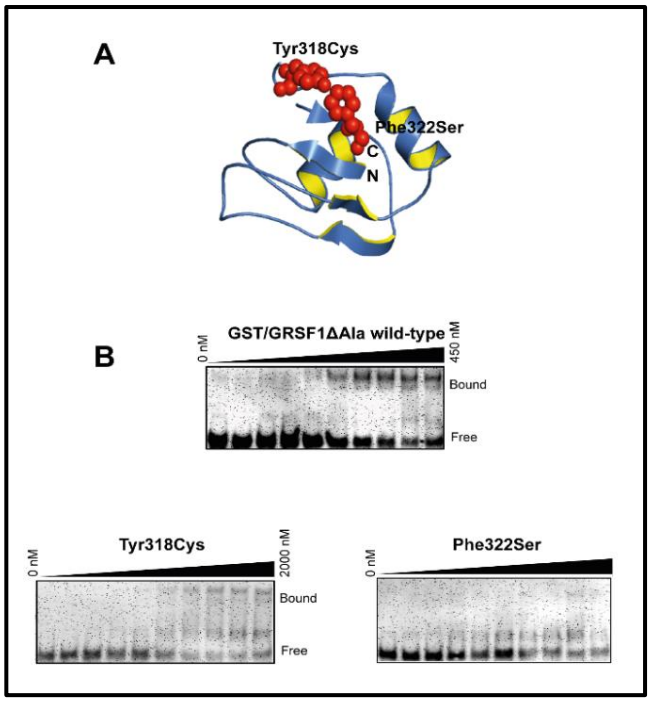

Figure 3.16. Impact of naturally occurring mutants in the qRRM2 RNA-binding domain of human GRSF1 on RNA-binding affinity. (**A**) Localization of the mutated amino acids in the 3D model of the qRRM2 RNA-binding domain. The mutated residues are shown as red spheres. (**B**) Quantitative RNA electrophoretic mobility shift assays (REMSA) were carried out to determine the dissociation constant (Kd) for the GRSF1-probe complex. For this purpose different amounts of purified recombinant wild-type and mutant protein was incubated with a digoxigenin labeled probe representing the 5'-UTR of human *GPx4* containing the $A(G)_4A$ motif. Each gel is one representative experiment of at least three independent experiments.

This amino acid is conserved in all the qRRMs and we found that the RNA-binding affinity of human GRSF1 decreases when this non-polar residue was replaced with a

redox sensitive cysteine. In fact, the Kd value of RNA-binding of the mutant protein was twice as high as for wild-type GRSF1 (**Table 3.4**). An even stronger reduction of the RNA-binding affinity was observed for the F322S exchange. Here the Kd-value of RNA-binding was increased to 2564 nM. F322 is an aromatic residue that was mapped to the β4 sheet of the qRRM2 domain. Taken together, these data suggest that the naturally occurring mutants Y318C and F322S impair the RNA-binding affinity of human GRSF1.

Functional consequences of amino acid exchanges in the qRRM3 domain of human GRSF1 - We then examined the impact of the five amino acid exchanges (F410S, K438N, R461L, Y468C, F472L) in the qRRM3 domain (**Figure 3.17**).

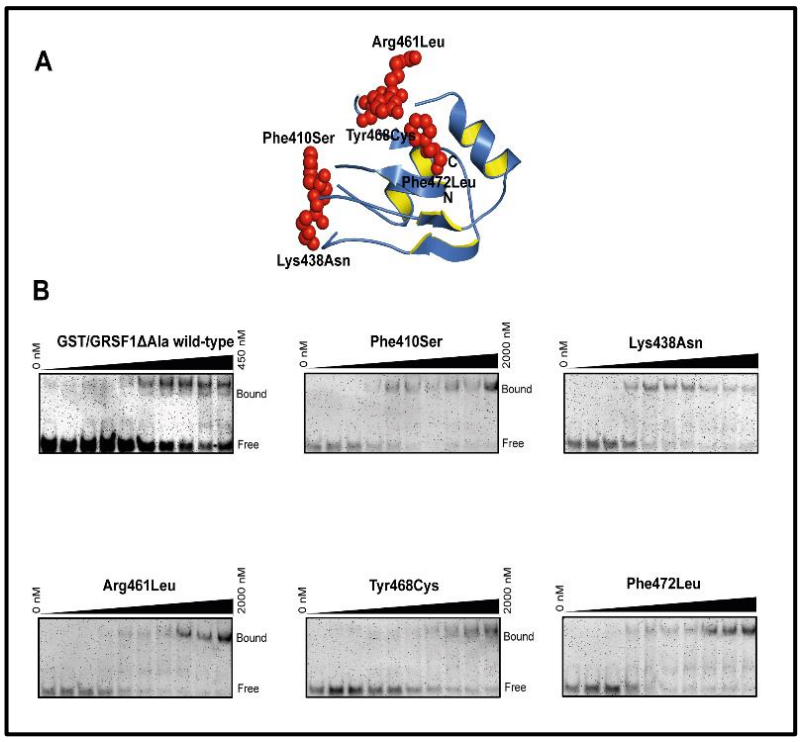

Figure 3.17. Impact of naturally occurring mutants in the qRRM3 RNA-binding domain of human GRSF1 on RNA-binding affinity. (**A**) Localization of the mutated amino acids in the 3D model of the qRRM3 RNA-binding domain. The mutated residues are shown as red spheres. (**B**) Quantitative RNA electrophoretic mobility shift assays (REMSA) were carried out to determine the dissociation constant (Kd) for the GRSF1-probe complex. For this purpose different amounts of purified recombinant wildtype and mutant protein was incubated with a digoxigenin labeled probe representing the 5'-UTR of human *GPx4* containing the $A(G)_4A$ motif. Each gel is one representative experiment of at least three independent experiments.

Here we found that the non-conservative amino acid exchanges F410S (bulky non-polar Phe is mutated to a smaller polar Ser), K438N (positively charged Lys is mutated to an uncharged polar asparagine), R461L (positively charged Arg is mutated to a non-polar Leu) did not alter the RNA-binding capacity of human GRSF1 as indicated by the similar Kd-values (**Table 3.4**). These data suggest that the naturally occurring amino acid exchanges should not impact the protein-RNA interaction *in vivo*. Thus, this genetic

variability is likely to be without functional consequence. Similarly, the conservative amino acid exchange F472L (non-polar Phe is mutated to non-polar Leu) did hardly impact the RNA-binding affinity (**Table 3.4**). In contrast, the non-conservative Y468C (bulky Tyr carrying an aromatic OH-group is mutated to a SH-group containing Cys) reduced the RNA-binding capacity (**Table 3.4**) as indicated by the two-fold increased Kd-value (555 nM for wild-type GRSF1 vs. 1108 nM for the mutant protein). Homozygous allele carriers are likely to express a GRSF1 variant with defective functionality.

Summarizing the functional data obtained for the naturally GRSF1 mutants one can conclude that the Q155R, T162S (located in qRRM1), Y318C, F322S (located in qRRM2) and Y468C (located in qRRM3) mutants exhibit defective RNA-binding properties. In contrast, the other mutants were without major functional impact although some of them were strongly non-conservative.

Summary: Our RNA gel experiments demonstrated that Q155R and T162S in qRRM1, T318C and F322S in qRRM2, T468C and F472L in qRRM3 show defective RNA-binding capacities.

3.5.4. Mechanistic investigations on functionally defective GRSF1 variants

T162S exchange reduced the RNA-binding activity of GRSF1 almost five-fold as concluded from the Kd-values (**Table 3.4**). Since both amino acid carry an aliphatic OH-group the major difference between these residues is their size (van der Waals volume of 93 Å^3 for Thr *vs.* 73 Å^3 for Ser). To test the relative contributions of size and OH-group for the drop in RNA-binding capacity of this naturally occurring GRSF1 mutant we first replaced the Thr with a Val. Both amino acids have similar van der Waals volumes (105 Å^3) but the Val lacks the hydroxyl group. Here we observed an almost three-fold reduction in the RNA-binding activity (**Table 3.5**). Next, we tested how the introduction of a bulky residue carrying a hydroxyl group (T162Y) impacts the RNA-binding activity. Here we found that the T162Y mutant exhibited a similar Kd-values (566 nM) as wild-type GRSF1 (555 nM). Taken together this data suggest that the presence of the OH-group is vital for the RNA-binding activity of GRSF1, but the size of the side chain at this position may also play a role.

Next, we analyzed the Y318C exchange, which reduced the RNA-binding capacity of GRSF1 by a factor of 2 (**Table 3.5**). Here a bulky side chain (van der Waals volume of 141 Å^3) carrying an aromatic hydroxyl group was replaced with a small amino acid (86 Å^3) involving a thiol group. To test the relative contributions of the two structural properties to the observed drop an RNA-binding capacity we created the Y318S (conservation of the and the Y318F (removal of the OH-group but conservation of the side chain volume)

mutants and tested their RNA-binding affinities. For the Y318S exchange we were not able to detect any RNA-binding activity of this mutant protein (**Table 3.5**). In contrast, the Y318F did not improve RNA-binding capabilities since its Kd value (**Table 3.5**) was not significantly different from the Y318C mutant (**Table 3.5**). These data suggest that both, the bulky aromatic ring as well as the OH-group are important for the RNA-binding properties of GRSF1.

Domain	Variation	Kd (nM)	SD	Relative Kd	p	n
	wild-type	555	297	1		14
qRRM1	Thr162Val	1479	865	2.7	0.004	3
	Thr162Tyr	566	284	1.0	0.956	3
qRRM2	Tyr318Ser	n.d.	n.a.	n.a.	n.a.	4
	Tyr318Phe	1900	976	3.4	<0.001	3
qRRM3	Phe322Tyr	n.d.	n.a.	n.a.	n.a.	4
	Phe322Met	n.d.	n.a.	n.a.	n.a.	4

Table 3.5. Binding affinity of artificial human GRSF1 variants to human *GPx4* mRNA. GRSF1 variants were expressed in *E.coli* and purified as described in Materials and Methods. The Kd-values for the different recombinant proteins were determined by quantitative electrophoretic mobility RNA shift assays. The relative Kd was defined as ratio of the mutant Kd / wild-type Kd. Ratios <1 indicate improved binding affinity. In each experiment the Kd values were calculated from 8-10 different experimental data points and each experiment was carried out at least in triplicate. Statistic significance (p, 2-sided) for the comparison of the Kd values of the mutant proteins with wild-type GRSF1 is given.

Finally, we explored the molecular basis for the drop in the RNA-binding affinity of the F322S mutant. For this variant we observed a five-fold lower RNA-binding affinity when compared with wild-type GRSF1 (**Table 3.5**). The F322S exchange removes bulk at this position but introduced an additional OH-group. To explore the mechanistic basis for the observed drop in RNA-binding affinity we first created the F322M mutant, which removed the OH-group of the Ser but conserves the side chain size (Ser and Met have similar site chain geometry). For this mutant we did not detect any RNA-binding activity (**Table 3.5**). Next, we conserved the OH-group of the Ser of the naturally occurring mutant but added more bulk at this position (F322Y). Again, no RNA-binding activity was observed for this mutant protein species (**Table 3.5**). Taken together, these data suggest that both, removal of bulk and introduction of the OH-group in the naturally occurring mutant (F322S) are involved in the dropdown in RNA-binding affinity.

Summary: Using mutagenesis experiments and RNA gel shift assays, we show that chemistry and geometry of crucial amino acid side chains impact the RNA-binding behavior of human GRSF1.

3.5.5. Thermostability of GRSF1 mutants

Site directed mutagenesis is always problematic since subtle point mutations might induce global structural changes, which might impact the functionality of proteins. In order to exclude that the amino acids exchanges carried out here induced major alterations in the protein structure we performed thermostability assays on all wild-type and mutant proteins expressed in this study. For this purpose proteins were incubated with a fluorescence probe and gradually heated to induce thermal denaturation. Here we found (**Figure 3.18**) that the denaturation behavior was similar for all recombinant proteins. Most of the proteins showed a melting temperature (T_m) of around 54°C. These data suggest that the mutations did not dramatically alter the overall structure and the thermostability of the recombinant proteins.

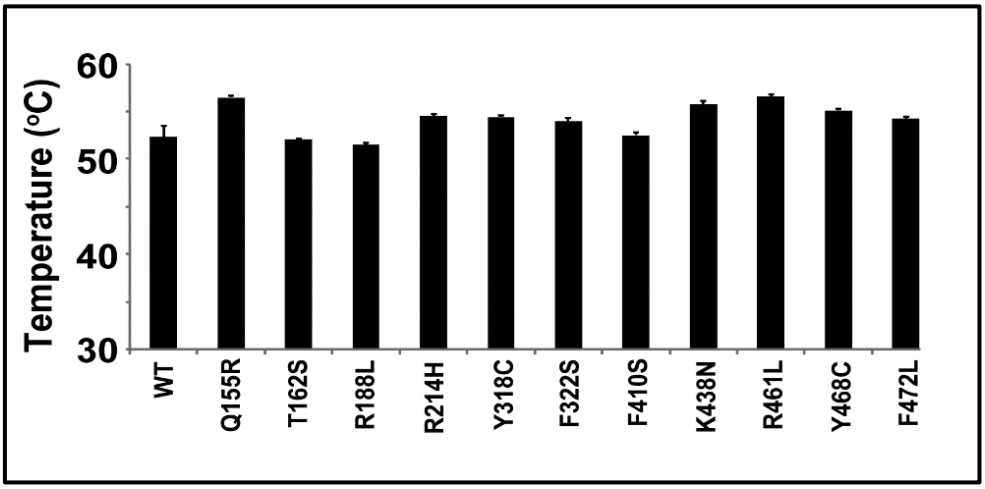

Figure 3.18. Thermal stability of naturally occurring and artificial human GRSF1 mutants. The melting temperatures for wild-type human GRSF1 and its mutants are plotted. Purified proteins were incubated with SYPRO® Orange dye and melting temperature (T_m) of the proteins was determined by monitoring the increase in fluorescence.

Summary: Using Thermal shift assays, we found that none of the naturally occurring genetic variations affect the overall protein structure and stability of GRSF1.

4. DISCUSSION

4.1. Expression and purification of GRSF1 and its different domains in E.coli

Although there is no crystal structure for GRSF1 other structural properties of this protein have well been characterized (Qian & Wilusz, 1994). Homology modeling, which has been carried out on the basis of the 3D-structure of other RNA-binding proteins, suggested that it consists of three conserved qRRM motifs, which have been identified as RNA-binding domains (RBDs) and of two other auxiliary domains (Ufer, 2012). One of these additional domains, which is rich in alanine residues (alanine-rich domain), is located at the N-terminus of the protein. The second auxiliary domain, which is rich in glutamate residues (acidic domain) is located between the qRRM2 and the qRRM3 domains (Qian & Wilusz, 1994; Ufer, 2012). In contrast to our structural knowledge the functional properties of GRSF1 have not been well characterized. For instance, it has not been explored, which role the different RNA-binding domains may play during protein-RNA interaction and the impact of point mutations on the RNA-binding capacity of the protein has neither been investigated. To study these topics effective recombinant expression systems are needed to prepare sufficient amounts of recombinant proteins, which are required for comprehensive functional studies.

Here we expressed and purified different human and mouse GRSF1 constructs as his- or GST-tagged fusion proteins in mg amounts and these constructs include full-length wild-type human GRSF1 (residues 1-480) and a truncated enzyme mutant, which lacks the alanine-rich domain (ΔAla-hGRSF1). In addition, the three RNA-binding domains (qRRM1, qRRM2, qRRM3) of mouse and human GRSF1 were separately expressed and have been made available for subsequent functional characterization. When we expressed the ΔAla-hGRSF1 construct and cleaved off the GST-fusion tag we found that the resulting ΔAla-hGRSF1 (no tag) migrated in SDS-PAGE with the apparent molecular weight of 53 kDa. In contrast, a theoretical MW of 42 kDa was predicted on the basis of the amino acid composition of this construct. Although the molecular basis for this discrepancy remains unclear the data is consistent with an early study on the electrophoretic mobility properties of human GRSF1. In this study the authors reported a similar electrophoretic mobility for native human GRSF1, which runs in SDS-PAGE at a significantly higher molecular weight (Jourdain et al., 2013).

To remove the GST-tag from the GRSF1 share of the fusion protein we subjected the purified GST-ΔAla-hGRSF1 fusion construct (without Ala-domain) to Factor Xa proteolysis. Here we observed an unexpected cleavage peptide that migrated at ~31 kDa in SDS-PAGE. To explain the chemical identity of this peptide we considered the possibility that this protein band might represent Factor Xa. However, neither the MW nor

the intensity of the 31 kDa band was consistent with this conclusion. Next, we searched the amino acid sequence of the recombinant fusion protein for an additional Factor Xa proteolytic cleavage site (Ile-Glu-Gly-Arg), but we did not find such sequence in the ΔE1-hGRSF1 fusion protein. Thus, an additional specific proteolytic cleavage could be ruled out. Thus, we concluded that that the 31 kDa cleavage peptide might have been formed via atypical proteolytic cleavage of the fusion protein by factor Xa. However, the exact cleavage site and the underlying mechanism remain to be determined. N-terminal amino acid sequencing of the 31 kDa cleavage peptide might help to precisely identify the cleavage site.

When we attempted to overexpress and purify full-length human GRSF1 in various prokaryotic overexpression systems (different *E.coli* strains) we consistently observed low level expression (**Figure 3.1 A**). To improve the expression level we tried to express a truncated version of GRSF1, which lacks the alanine-rich domain (ΔAla-hGRSF1). Here we observed a more than 10-fold higher expression level **(Figure 3.1 B)**. Although the molecular basis for the low-level expression of full-length human GRSF1 has not been clarified it may be related to the protein-chemical characteristics of the alanine-rich domain. This domain involves a large share of hydrophobic amino acids, which might cause substantial aggregation when proteins are expressed at high levels. As further possible explanation one may discus that the alanine-rich domain may cause misfolding of the recombinant protein if suitable eukaryotic folding catalysts (chaperons) are not present in sufficient amounts.

As indicated above the X-ray structure of GRSF1 has not been solved. As major reason for this lack in direct structural information the unavailability of sufficient amounts of highly purified protein may be discussed. There is no native source for the preparation of large amounts of highly pure human or murine GRSF1 and suitable recombinant overexpression systems have not been described in the literature. Here we worked out a highly efficient prokaryotic overexpression system for human GRSF1, which also works for different murine GRSF1 constructs. Moreover, we developed a relatively simple protein purification strategy, which involves both, affinity chromatography and size exclusion chromatography. The combination of these two methods allowed the preparation of mg amounts of recombinant protein, which have subsequently been used for functional characterization, crystal trials and for the preparation of antibodies. However, despite many attempts unfortunately we did not obtain crystals for any of the protein constructs.

Our initial expression strategy was based on expression of a GST-tagged fusion protein constructs. These constructs were suitable for functional studies since the GST-domain of the fusion protein does not impact the RNA-binding properties. However, for

crystal trials and antibody production these fusion proteins are not well suited because of the large size (30 kDa) of the tag-domain. To prepare GST-free GRSF1 constructs there are several ways, which include the following two scenarios: i) Proteolytic cleavage of the GST-GRSF1 constructs employing the factor Xa cleavage site (**Figure 3.2 A**). This procedure works well but it is rather laborious and suffers from low efficiency. Protein recovery is limited and one obtains an unidentified 31 kDa cleavage fragment. This fragment can be removed by binding to the GSH-matrix but its formation lowers the overall yield of this procedure. ii) Expression of GRSF1 constructs as his-tag fusion proteins. We tested this option for the expression of the individual RNA-binding domains of mouse GRSF1. These constructs include (**Figure 3.4**) the mouse qRRM1 domain (residues 120-238, 19 kDa), the qRRM2 domain (residues 240-336, 17 kDa) and the qRRM3 domain (residues 389-480, 16 kDa). These constructs were expressed at high yields (12 mg, 2.8 mg and 3.4 mg pure protein per 0.5 l bacterial liquid culture) and the final protein preparations exhibited a high (>95%) degree purity. They are suitable for direct structural investigations by X-ray crystallography and/or NMR-spectroscopy.

4.2. Phylogenetic analysis indicated the occurrence of GRSF1 like sequences at low frequency in lower living organisms and in viruses

It has previously been suggested that the gene encoding human GRSF1 is highly conserved in higher and lower vertebrates, but does not regularly occur in lower organisms (Antonicka et al., 2013; Ufer, 2012). Unfortunately, the occurrence of GRSF1-like sequences in frequently employed model organism has not been studied and it remained unclear whether GRSF1-like sequences occur in viruses. Moreover, more and more genomic sequences are currently deposited each day in the publically available databases and thus, we felt it was about time to carry out a new database search. When we searched the viral, bacterial and archaeal proteomes, which were predicted on the basis of the genomic sequences we did not found any GRSF1-like proteins and these data suggest that this protein is not a typical virus constituent and may neither be important for bacterial and archaeal physiology. However, we detected GRSF1-like sequences in fungi, and in plants. However, in these species GRSF1-like proteins occur not very frequently and thus, there is no systematic evolution of this protein in these lower organisms.

Although, GRSF1 has been previously suggested to have no evolutionary conserved homologs beyond the ray-finned bony fishes (Antonicka et al., 2013) (**Figure 4.1**), our homology searches indicated the existence of a GRSF1-like protein (uniprot ID E1JH76) in *D. melanogaster* and in *C.elegans*. Both proteins have been suggested to function as RBPs share a reasonable sequence identities of 29% (Fusilli isoform G, *D.*

melanogaster) and 31% (sym-2, *C. elegans*) with human GRSF1 at the protein level. Moreover, the domain organization of Fusilli isoform G and sym-2 protein is strikingly similar to that of human GRSF1. Both proteins contain three RNA-binding domains (RBDs).

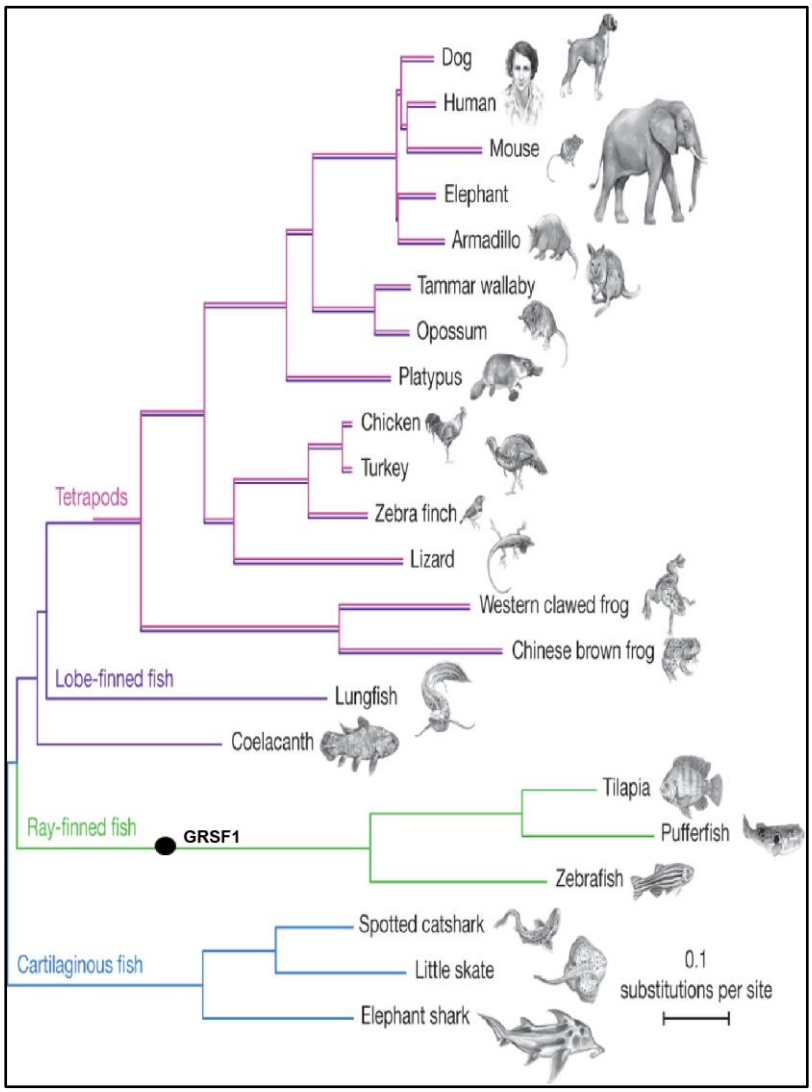

Figure 4.1 A modified representation of phylogenetic tree of jawed vertebrates showing introduction of GRSF1 protein (black circle) in ray-finned fishes during vertebrate evolution (Amemiya et al., 2013).

However, the *D. melanogaster* protein (Fusilli isoforms) carries a 3'-5' exonuclease domain, which is replaced by Ala-rich domain in human GRSF1. In addition, previous reports have suggested that some Fusilli isoforms are closely related to hnRNP F and both proteins have been suggested to share similar RNA substrates (Wakabayashi-Ito, Belvin, Bluestein, & Anderson, 2001). Furthermore, the three RBDs of the sym-2 protein (*C-elegans* protein) have been implicated in regulation of alternative splicing and thus may play a key role in embryogenesis of *C. elegans* (Barberan-Soler & Zahler, 2008). Interestingly, the involvement in RNA splicing has also been highlighted for the hnRNP F/H

protein family including GRSF1. These proteins regulate splicing of HIV mRNA (Jablonski & Caputi, 2009), numerous pre-mRNA species such as the Bcl-x (Garneau, Revil, Fisette, & Chabot, 2005), rat β-tropomyosin (Chen, Kobayashi, & Helfman, 1999), the HIV type 1 tat (Jacquenet et al., 2001) and the Rous sarcoma virus NRS (Fogel & McNally, 2000). Collectively, in the light of these findings we suspect that the Fusilli protein G and sym-2 may be regarded as GRSF1 homologs. However, more investigations need to be carried out to verify the relationship between these proteins and GRSF1.

GRSF1-like proteins do not frequently occur in lower plants but our systematic search in higher plant genomes indicated a high occurrence frequency in plants. In fact, all blastable higher plant genomes involve GRSF1-like sequences but unfortunately, little is known about the functional properties of these GRSF1-like proteins.

GRSF1 is evolutionary conserved in various vertebrates including birds, amphibia, mammals as well as fishes (Ufer, 2012) (see **Figure 4.1**). We investigated the conservation of GRSF1 in vertebrates by exploring the occurrence of GRSF1-like sequences in selected representatives of vertebrates such as zebrafish, chicken, mouse, rat, and cattle. The sequence alignment of GRSF1 protein in human, H. Neanderthal and H. denisovan revealed that GRSF1 protein in H. Neanderthal and H. denisovan shares a sequence identity of >99% with human GRSF1 (see **Figure 3.7**). The second interesting observation we made was that the protein sequences in H. Neanderthal and H. denisovan contain single nucleotide variants (SNVs). Two of these SNVs were found in primary sequence of H. Neanderthal (Ser to Phe (rs3780903) and Ala to Ser) (see **Figure 3.7**). Interestingly, the same SNV (Ser to Phe variation (rs3780903) was observed in H. denisovan (Ser to Phe rs3780903) (see **Figure 3.7**). Thus we observed a total of two SNVs in protein sequence of H. Neanderthal and H. denisovan.

We found that all the analyzed species contain a GRSF1-like protein with complete set of domains that includes qRRM1, qRRM2 and qRRM3 as well as acidic domain and ala-rich domain (see **Figure 3.8**). This conservation of the GRSF1 structure was also found in chimpanzee and dog (Ufer, 2012). Moreover, the GRSF1 found in chimpanzee carries an extended N-terminal Ala-rich domain (Jourdain et al., 2013; Ufer, 2012). In addition, the multiple sequence alignment of various GRSF1 sequences from different vertebrate species revealed that the three qRRMs (qRRM1-3) share a reasonable degree of homology amongst themselves, A high degree of amino acid conservation was also observed when the amino acid sequences of the Ala-rich and acidic domains of mouse, rat, humans and cattle GRSF1 were compared (see **Figure 3.8**). In contrast, the degree of amino acid conservation in the auxiliary domains (Ala-rich domain, acidic domain) in zebrafish GRSF1 is much lower. The low similarity in primary sequence of each

of the three qRRMs suggest that GRSF1 protein may show different RNA-binding activities. These findings are consistent with our results that suggest different RNA-binding affinities (Kd-values) for GRSF1 that harbor these domains (unpublished data). Furthermore, previous study have shown that proteins containing non-identical domains show dissimilar functions (Maris et al., 2005). Taken together, these data suggests that structure of GRSF1 protein is more conserved than that of other protein sequence and therefore may serve a similar purpose throughout vertebrate kingdom. Moreover, in the absence of any experimental and evolutionary evidence, it is tempting to speculate that conservation of GRSF1 structure despite low primary sequence similarity implies that there is an evolutionary pressure to conserve the structure and function of GRSF1 protein across vertebrate kingdom.

4.3. GRSF1 RNA substrates fold into parallel G-quadruplex structures

RNA G-quadruplexes (G4s) have come into the focus of scientific research only in recent years (Ji et al., 2011; Millevoi, Moine, & Vagner, 2012), although the first solution structure of a parallel RNA G-quadruplex (G4) has already been published in 1992 (Cheong & Moore, 1992). The existence of these structure has been the subject of intense investigation, but only recently they have been proven to occur *in vivo* in the human cells (Biffi et al., 2014). Several studies have suggested their involvement in mRNA processing, and mRNA translation (König, Evans, & Huppert, 2010; Millevoi et al., 2012). Moreover, G4-forming sequences are widespread throughout the genome. On the DNA level corresponding sequences are highly enriched in telomeres and in gene promoters (Maizels & Gray, 2013; Rhodes & Lipps, 2015). Since the formation of G4-structures is difficult to assay *in vivo* an *in silico* prediction methods for G4-complex formation from the primary RNA structure was developed (Kikin et al., 2006). This program calculates a G-score as measure for the likelihood of an RNA sequence to form stable G4 structures. The highest possible score obtained using QGRS mapper is 105 (Kikin et al., 2006). A recent study has reported that RNA sequences (ARPC2 and MMP16) with a G-score in the range of 39-41 fold into stable G4 structures (Von Hacht et al., 2014). In our study we calculated a G-score of 19 for human *GPx4* mRNA, a G-Score of 17 for mouse *GPx4* mRNA and a G-score of 19 for), mouse *Use 1* mRNA. These data suggest that these RNA sequences fold into G4 structures with moderate stability, which is consistent with previous data (Nieradka et al., 2014). However, it is important to consider that this type of predictive *in silico* search is likely to overestimate the prevalence of G4 structures in the 5'-UTR of a cellular transcriptome. The only safe way to show the formation of such structures is the solution of the crystal structure (Huang et al., 2014).

To confirm the presence of G4 structures in the GRSF1 RNA substrates, we separately monitored the CD spectra of individual RNA probes. Our CD analysis confirmed the presence of parallel G4 structures. Thus, our data further supports the *in silico* data suggesting the presence of G4 structure within the 5'-UTR of *GPx4* (human and mouse) and *Use1*. These findings are consistent with our previous data on *Use1* RNA that adopts a parallel G4 structure (Nieradka et al., 2014). Furthermore, all the CD measurements were performed at RNA concentration of 5 µM and close to these concentrations RNA has been suggested to form biologically relevant unimolecular G4 structures (Aher, Erande, Fernandes, & Kumar, 2015; Lane, Chaires, Gray, & Trent, 2008). Previous study on the formation of G4 structures were carried out at physiologically irrelevant (25 mM) RNA concentrations (Cyril Dominguez et al., 2010). Unimolecular G4 structures have been suggested to occur *in vivo* (Lane et al., 2008). Thus, our experimental conditions and experimental setup is comparable to above reported study (Lane et al., 2008).

We also analyzed the RNA mutants for the formation of G4s. Our CD analysis suggest that the G-tract (GGGG) within $A(G)_4A$ element may not be essential for G4 formation but the G-to-A mutations impair this process.

Only a small number of proteins that bind to RNA G4 structures have been characterized (Brázda, Hároníková, Liao, & Fojta, 2014). A comparison of the binding affinities of ΔG4 deletion mutant and mutated G-to-A exchange construct revealed the Kd-values of 2120 nM and 0 nM respectively (**Figure 3.13 Panel B**). The mutated G-to-A mutant does not form a G4 structure and no binding of this mutant to GRSF1 was observed. In addition, we have shown previously that G4 forming *Use 1* substrate binds GRSF1 with even weaker affinity than wild-type *GPx4* substrate (Nieradka et al., 2014b) (Kd-value of 527 nM vs 66 nM for wild-type mouse *GPx4* RNA). Moreover, we obtained a Kd-value of 4.4 µM for *Use1* probe without $A(G)_4A$ motif (Nieradka et al., 2014) that is higher than Kd-value of *GPx4* probe lacking $A(G)_4A$ element (2120 nM). This diversity in binding affinities may result from the different sequence and structure requirements of RNA substrates (**Section 4.4 for details**). Although direct structural data on GRSF1-RNA interactions is lacking, GRSF1 may potentially bind these G4 structures in a similar way as described for the related protein hnRNP F (Ufer, 2012).

More recent structural studies on the binding of isolated RNA-binding domains (qRRMs) of human hnRNP F to its RNA substrates indicated a more complex binding mechanism. In the absence of hnRNP F the RNA substrates adopt G4-structures (Cyril Dominguez et al., 2010; Samatanga et al., 2013). However, when the NMR solution structure of an isolated qRRMs in complex with its substrate RNA was solved it became evident that protein binding dissolves the G4 structure. In fact, in the RNA-protein complex

the RNA was detected in single-stranded conformation. Thus, protein binding apparently unfolds the G4 structures (C. Dominguez, 2006; Cyril Dominguez et al., 2010). Although similar mechanistic studies have not been carried out for GRSF1 it might be suspected that GRSF1 follows a similar RNA-binding mechanism, which involves unfolding of the G4 structures. Unfolding of mRNA secondary structures has been shown to enhance translation efficiency (Samatanga et al., 2013) making this process less dependent on ATP (Guo & Bartel, 2016). Therefore NMR spectroscopy of GRSF1 in complex with RNA is needed to obtain such structural information in future studies.

4.4. Functional insights into GRSF1-RNA interactions

GRSF1 is a key regulator of the post-transcriptional gene expression (Ufer, 2012) and specifically binds to RNAs containing G-rich sequences (AGGGGA) (Qian & Wilusz, 1994). However, little is known about the molecular mechanisms of GRSF1/mRNA interaction. Furthermore, the minimal RNA sequence required for GRSF1 binding has not precisely been defined. If one compares the nucleotide sequences of the two major GRSF1 substrates explored so far (*GPx4* mRNA and *Use1* mRNA) one can conclude that both of them involve the central G-rich motif [$A(G)_4A$]. However, the sequences surrounding the G-rich motif are very different (Nieradka et al., 2014). These data suggest that the G-rich motif flanking sequences may impact the binding affinities. Unfortunately, our limited truncation studies did not provide more detailed insight into the regulatory relevance of the flanking sequences.

Another important factor, which is likely to impact the intensity of GRSF1-mRNA interaction, is the stability of RNA secondary structure. Our CD spectroscopy data obtained for *Use1* and *GPx4* RNAs confirmed that these RNA substrates fold into four stranded G-quadruplexes and the central $A(G)_4A$ element is part of these structures (Nieradka et al., 2014). Since direct structural information on GRSF1-RNA interaction is lacking (Nieradka et al., 2014), the precise location of the $A(G)_4A$ element in the G-quadruplex structure is not known. Moreover, using RNA gel shift experiments we showed that GRSF1 bind to G-quadruplex structure in *GPx4* RNA (see **Section 3.3.4**). These structures might even be more important for GRSF1 binding than the cognate $A(G)_4A$ motif. We therefore suspect that the stability of G-quadruplex structures, might be a key property, which impacts the intensity of interaction of GRSF1 with different mRNAs. Thus GRSF1 may bind with high affinity to stable G-quadruplexes and with low affinity to unstable G-quadruplexes. This is consistent with recent study that demonstrates that the RNA-binding protein nucleolin binds with high affinity at RNA G-quadruplexes (Von Hacht et al., 2014).

Despite our detailed knowledge on RNA-protein interaction of the human hnRNP F protein, little is known on the corresponding features of GRSF1 (Ufer, 2012). As GRSF1 the hnRNP F protein contains three RNA-binding domains (qRRMs) and for this protein each qRRM domain was shown to bind to its target RNA independently of the other domains (Cyril Dominguez et al., 2010). In contrast, our truncation studies suggested that the three RNA-binding domains of human GRSF1 do not bind the substrate RNA independently of each other. However, our results do not necessarily mean that the separated RNA-binding domains do not bind to substrate RNAs but their binding affinities are considerably lower than that of the full-length protein The biological function of multiple RNA-binding domains in a single protein has been suggested to improve the RNA-binding affinity as well as the sequence-specificity (Antoine Cléry & Allain, 2013; Maris et al., 2005). In fact, the three RNA-binding domains of hnRNP K have been suggested to bind cooperatively, which is likely to improve the RNA-binding affinity (Paziewska, Wyrwicz, Bujnicki, Bomsztyk, & Ostrowski, 2004). Moreover, direct structural studies on two consecutive RNA-binding domains complexed with RNA in several RNA-binding proteins (Handa et al., 1999;, (Wang, Tanaka Hall, & Hall, 2001) Deo, Bonanno, Sonenberg, & Burley, 1999; Allain, Bouvet, Dieckmann, & Feigon, 2000; Johansson et al., 2004) have confirmed the cooperative nature of RNA-binding. In all of these cases the multiple RNA-binding domains cooperatively bind to substrate RNA, which optimizes binding affinity and binding specificity. Therefore it is plausible that the multiple RNA-binding domains of human GRSF1 also bind their RNA substrates in a cooperative manner.

The functionality of the different structural units of human GRSF1 has not been studied in detail (Ufer, 2012). Of course, the three qRRM domains have been implicated in RNA-binding because of their structural similarity to corresponding structural subunits of other RNA-binding proteins but the functionality of the Ala-rich domain and the acidic domain (AD) has not been explored in the past. Here we showed that the Ala-rich domain is not required for RNA-binding and that the AD-domain negatively regulates RNA-binding (see **Section 3.4.3 for details**).

The minimum length of *GPx4* RNA required for binding to human GRSF1 has not been defined. Our RNA gel shift experiments using mutant RNA probes of different length indicated that these mutant RNAs do not significantly affect RNA-binding (see **Section 3.4.4 for details**). In addition, we could define the minimal RNA sequence of 32 nucleotides and this sequence did not affected RNA-binding. Further experiments are therefore needed to define the minimal RNA sequence required for binding to GRSF1.

4.5. Functional investigations of genetic variations in human GRSF1

4.5.1. Structural modeling of GRSF1

The structure of hnRNP F, a functional relative of GRSF1, is well defined and the mechanisms involved in RNA-binding of this protein have been comprehensively characterized (Cyril Dominguez et al., 2010; Samatanga et al., 2013). In contrast, the 3D structure of human GRSF1 has not been solved and the molecular basis of the RNA-binding capabilities of this protein has not been explored in detail. The RRM fold is one of the most abundant RNA-binding domains in vertebrates and more than 200 structural variations of this motif have been reported (Afroz et al., 2015). Interestingly, RRMs not only interact with RNA but also with DNA, proteins and even lipids (Clingman et al., 2014). A typical RRM domain comprises 80-90 amino acids and consists of four anti-parallel β-sheets packed against two α-helices. The overall protein adopts a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology (Antoine Cléry & Allain, 2013). GRSF1 has been discovered as RNA-binding protein, which interacts with guanine-rich nucleotide sequences and the three qRRM domains have been suggested to function as RNA-binding domains (Ufer, 2012). Importantly, RNA substrates of GRSF1 have been shown to fold into G-quadruplexes (Nieradka et al., 2014a), which has previously been demonstrated for a Bcl-x substrate of hnRNP F (Cyril Dominguez et al., 2010). In order to obtain structural information on the three qRRM domains of GRSF1, we carried out amino acid sequence alignments and structural homology modeling on the basis of the solution structure of the corresponding structural elements of hnRNP F. The results of this homology modeling suggested that all three RNA-binding domains adopt the canonical $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ fold, which is characteristic for RNA-binding proteins (**Figure 3.14**). In hnRNP F the RNA-binding domains qRRM1 and qRRM2 involve secondary structural elements between α2 and β4 (C. Dominguez, 2006) but corresponding structures have not been detected in the RNA-binding domains of GRSF1.

The qRRMs of hnRNP F recognize RNA in a unique way, and positively charged amino acids located in loop 1, 3, and 5 as well as in the β1 strand have been implicated in RNA-binding (Cyril Dominguez et al., 2010). To identify the amino acids that are involved in RNA-binding of GRSF1 we carried out amino acid alignments and found a high degree of conservation of the responsible amino acids in human GRSF1 suggesting similar mechanisms of RNA-binding for hnRNP F and GRSF1. However, there are a number of differences and the most striking difference is that qRRM2 and qRRM3 contain an Asn (N) at positions 262 and 413. In hnRNP F these positions, which have been implicated in protein-RNA interaction, are occupied by threonine residues (T) at positions 123 and 301. However, neither of these residues was affected by any of the naturally occurring GRSF1

mutations. A second interesting difference is the presence of a neutral Gln (Q) at position 155 in qRRM1 of GRSF1. In hnRNP F a positively charged Arg (R) is present at this position and this residue has been implicated in protein-RNA interaction (Cyril Dominguez et al., 2010). Interestingly, one of the naturally occurring mutants of GRSF1 constitutes a back-mutation of this structural difference between hnRNP F and GRSF1.

4.5.2. Functional alterations induced by naturally occurring GRSF1 mutations

Aromatic and positively charged amino acids in the RNA-binding domains of hnRNP F have been implicated in protein-RNA interaction (C. Dominguez, 2006; Cyril Dominguez et al., 2010). Most of these amino acids are conserved in GRSF1 and naturally occurring mutations of these residues are likely to impair the RNA-binding affinity. Furthermore, mutations affecting these crucial residues in hnRNP F and SRSF2 alter the RNA-binding properties (C. Dominguez, 2006; Zhang et al., 2015). Here we quantified the genetic variability of the human *GRSF1* gene and tested the functional consequences of naturally occurring mutations in the three RNA-binding domains of human GRSF1. Our functional data indicated that some of the naturally occurring amino acid exchanges in qRRM1 (Q155R, T162S) have functional consequences since they impair the RNA-binding capability of GRSF1. Comparable functional alterations have been detected for the naturally occurring mutations in qRRM2 (T318C, F322S) and this was also the case for the T468C and F472L exchanges in qRRM3. On the other hand, R188L and R214H exchange (in qRRM1) and F410S, K438N and R461L mutations (in qRRM3) were without significant functional impact although some of the amino acid exchanges (R188L, R461L) were strongly non-conservative. Strikingly, most amino acids that induce defective RNA-binding are not localized in the β-sheet (only one residue located in β4-sheet) but are present in the loop regions (loop 1, 3 and 5). This is in sharp contrast to classical RRM domains the protein-RNA interactions of which are mediated by the β-sheets. Thus, our mutagenesis results confirm the previous suggestion that the ß-sheet connecting loop regions are more important for protein-RNA interaction in qRRMs.

Our genomic database searches indicated that only one naturally occurring mutation (Ser95Pro) in human GRSF1 has an allele frequency of >1%. We hypothesize that the low allele frequencies of naturally occurring GRSF1 mutations suggests that there is an evolutionary pressure on the GRSF1 gene, which prevents the accumulation of functional inactive alleles within the human population. This is of course a tempting speculation, which needs to be supported by additional experimental data. However, our genomic database searches suggested that GRSF1-like sequences are highly conserved in vertebrates and these findings suggests that this protein is needed in higher animals.

4.5.3. Mechanism of RNA-binding

The mechanism of RNA-binding by hnRNP F involves hydrogen bonds between the nucleotide bases and amino acid residues but also $\pi-\pi$ interactions between nucleotide bases and aromatic amino acid side chains (Cyril Dominguez et al., 2010; Samatanga et al., 2013). To explore the RNA-binding mechanisms of human GRSF1 we created a number of mutant GRSF1 variants and tested their RNA-binding affinities as suitable measure for GRFS1 functionality. Our results characterized the molecular mechanisms of GRSF1-RNA interaction and explained the loss of RNA-binding capacity by some rare non-synonymous nucleotide exchanges [T162S (loop 1), Y318C and F322S (loop 5) in qRRM1 and qRRM2 respectively]. Analysis of the RNA gel shift assays provide mechanistic insights into the functional mechanism of six mutants of GRSF1 including T162V and T162Y in qRRM1, Y318S, Y318F, F322Y and F322M in qRRM2. We observed the highest perturbation in RNA-binding for the T162V, Y318S, Y318F, F322Y and F322M mutants. T162V exchange may impair RNA-binding by limiting peptide backbone flexibility (Betts & Russell, 2007), which might in turn decrease the RNA-binding affinity. However, in the absence of direct structural data characterizing the molecular details of protein-RNA interaction (co-crystallization studies of GRSF1 variants with RNA probes or NMR studies on GRSF1-RNA complexes) further mechanistic speculations may not be productive.

4.5.4. GRSF1 point mutations do not alter the global protein structure and thermostability

Subtle point mutations can significantly impact the global structure of proteins and thus the results of *in vitro* mutagenesis studies may be misleading (Kumar, Satish, Patel, & Panchaldr, 2016). To quantify the impact of naturally occurring point mutations on the global structure of GRSF1 we performed thermal stability assay on all GRSF1 variants used in this study. Here we observed similar thermal unfolding curves suggesting that the introduced point mutations did hardly impact the global protein structure. In addition, our protein stability data revealed that all protein variants exhibit a melting point of around 54 °C. Taken together these data suggest that the drop in RNA-binding affinity observed for several naturally occurring GRFS1 mutants may not be related to severe disturbance of the global protein structure but rather to more subtle and local alterations of specific protein-RNA binding forces.

5. ZUSAMMENFASSUNG/SUMMARY

GRSF1 is a ubiquitously occurring RNA-binding protein (RBP) that contains three quasi-RNA recognition motifs (qRRMs). These domains bind G-rich RNA sequences and the minimal RNA targeting sequence has been narrowed down to an $A(G)_4A$ sequence. Before this project the molecular mechanisms of GRSF1-RNA interaction and the functional consequences of naturally occurring mutations in human GRSF1 were not known. Here we expressed different human and mouse GRSF1 constructs, purified them to near electrophoretic homogeneity and employed these proteins for RNA-binding studies.

GRSF1 frequently occurs in mammals but its distribution in other living organisms has not been studied in detail. Here we performed comprehensive *in silico* searches for GRSF1-like proteins and found that such sequences do not frequently occur in viral, bacterial, archaeal and fungal proteomes. However, related sequences were found in higher frequency in mosses, higher plants and lower non-mammalian animals.

To explore the RNA-binding mechanism of GRSF1 we modified both, the RNA substrates and the RNA-binding protein. First, we analyzed RNA constructs representing GRSF1 substrates by circular dichroism spectroscopy and found that these oligonucleotides fold into parallel G-quadruplex secondary structures. Next, we functionally characterized the different structural domains of GRSF1 and determined their binding constants with a labeled RNA probe representing the 5'-UTR of human *GPx4* mRNA. Our results indicate that the N-terminal Ala-rich domain is not essential for RNA-binding. In contrast, the three canonical RNA-binding domains contribute to high affinity RNA-binding.

Finally, we functionally characterized naturally occurring genetic variations of human GRSF1. For this purpose we first modeled the 3D structure of the qRRM domains on the basis of the NMR structure of hnRNP F and identified putative RNA interacting amino acids. The models indicated that the qRRM domains adopt the canonical $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$-fold, which is characteristic for RNA-binding proteins. To explore the genetic variability of human GRSF1 we searched different genomic databases and found a total of 294 genetic variations. However, except for the S95P exchange none of them has an allele frequency >1%. Exploring the functional consequences of selected non-synonymous nucleotide exchanges we found that the following mutants exhibited impaired RNA-binding capabilities: Q155R and T162S in qRRM1, T318C and F322S in qRRM2, T468C and F472L in qRRM3. To investigate the molecular basis for this impairment we created a number of additional GRSF1 mutants and observed that chemistry and geometry of critical amino acid site chains impact the RNA-binding behavior of human GRSF1. To exclude that our mutations have altered the global structure of GRSF1 we performed thermal shift assays and found that the naturally occurring mutations did neither impact the global protein structure nor protein stability.

GRSF1 ist ein ubiquitär vorkommendes RNA-bindendes Protein (RBP), das drei quasi-RNA-Erkennungsmotive (qRRMs) enthält. Diese Domänen binden G-reiche RNA-Sequenzen und die minimale RNA-Zielsequenz wurde auf das $A(G)_4A$-Sequenzmotiv eingeengt. Vor Beginn dieses Projektes waren die molekularen Mechanismen der GRSF1-RNA-Wechselwirkungen und die funktionellen Konsequenzen von natürlich vorkommenden Mutationen im humanen GRSF1-Gen nicht bekannt. Um diese Themen zu erforschen, haben wir zunächst verschiedene humane und murine GRSF1-Konstrukte als rekombinante Proteine exprimiert und weitgehend aufgereinigt. Diese Proteine wurden anschließend für mechanistische Untersuchungen zur RNA-Bindungsfähigkeit eingesetzt.

Um die Evolution von GRSF1 zu erforschen, haben wir öffentlich zugängige Sequenzdatenbanken nach GRSF1-ähnlichen Sequenzen durchsucht. Dabei konnten wir festgestellen, dass solche Proteine in viralen, pro-karyotischen und Pilzproteomen wenig verbreitet sind. Im Gegensatz dazu kommen GRSF1-ähnliche Sequenzen in Moosen, höheren Pflanzen und bei niederen Tieren weiter verbreitet vor.

Um den RNA-Bindungsmechanismus von GRSF1 besser zu verstehen, haben wir sowohl die RNA Substrate als auch das Bindungsprotein (GRSF1) zielgerichtet modifiziert. Zuerst wurden dabei Messungen des Zirkulardichroismus an potentiellen GRSF1 Substraten durchgeführt. Aus diesen Daten wurde geschlussfolgert, dass GRSF1 Substrate sich in parallele G-quadruplex Strukturen falten, die als Erkennungsstrukturen dienen. Anschließend wurde die funktionelle Bedeutung der verschiedenen GRSF1 Domänen untersucht. Die Ala-reiche Domäne hat für die RNA Bindung kaum Bedeutung. Demgegenüber tragen alle drei qRRM Domänen zur hochaffinen RNA-Bindung bei.

Um die funktionellen Auswirkungen natürlich vorkommender Mutationen im humanen GRSF1 Gen analysiert. Dafür wurden zuerst 3D-Strukturmodelle für die drei RNA-Bindungsdomänen des humanen GRSF1 erstellt. Diese Modellierungen ergaben, dass sich die GRSF1 qRRMs in das klassische $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$-Motiv falten, welches charakteristisch für RNA bindende Proteine ist. Bei unserer Suche nach natürlich vorkommenden Mutationen im humanen GRSF1 Gens identifizierten wir einen SNP (single nucleotide polymorphism) und knapp 300 seltenen Mutationen. Von diesen wiesen die folgenden Aminosäureaustausche funktionelle Defizite auf: Q155R, T162S in qRRM1, T318C, F322S in qRRM2 und T468C, F472L in qRRM3. Mechanistische Untersuchungen lassen darauf schließen, dass die Chemie und die Geometrie kritischer Aminosäureseitenketten für die RNA-Bindungsfähigkeit bedeutsam sind. Vergleichende Fluoreszenzmessungen zeigten, dass alle hergestellten GRSF1 Mutanten keine gravierenden strukturellen Unterschiede zum Wildtypenzym aufwiesen.

6. BIBLIOGRAPHY

Afroz, T., Cienikova, Z., Cléry, A., & Allain, F. H. T. (2015). One, Two, Three, Four! How Multiple RRMs Read the Genome Sequence, *558*, 235–278. http://doi.org/10.1016/bs.mie.2015.01.015

Agarwala, P., Pandey, S., & Maiti, S. (2015). The tale of RNA G-quadruplex. *Organic & Biomolecular Chemistry*, *13*(20), 5570–85. http://doi.org/10.1039/c4ob02681k

Aher, M. N., Erande, N. D., Fernandes, M., & Kumar, V. A. (2015). Unimolecular antiparallel G-quadruplex folding topology of 2'-5'-isoTBA sequences remains unaltered by loop composition. *Organic & Biomolecular Chemistry*, *13*(48), 11696–703. http://doi.org/10.1039/c5ob01923k

Akkers, R. C., Jacobi, U. G., & Veenstra, G. J. C. (2012). Xenopus Protocols. *Xenopus Protocols: Post-Genomic Approaches Methods in Molecular Biology*, *917*, 279–292. http://doi.org/10.1007/978-1-61779-992-1

Albert, P. R. (2011). What is a functional genetic polymorphism? Defining classes of functionality. *Journal of Psychiatry and Neuroscience*, *36*(6), 363–365. http://doi.org/10.1503/jpn.110137

Alföldi, J., Di Palma, F., Grabherr, M., Williams, C., Kong, L., Mauceli, E., … Lindblad-Toh, K. (2011). The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature*, *477*(7366), 587–91. http://doi.org/10.1038/nature10390

Allain, F. H. T., Bouvet, P., Dieckmann, T., & Feigon, J. (2000). Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO Journal*, *19*(24), 6870–6881. http://doi.org/10.1093/emboj/19.24.6870

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–10. http://doi.org/10.1016/S0022-2836(05)80360-2

Amemiya, C. T., Alföldi, J., Lee, A. P., Fan, S., Philippe, H., Maccallum, I., … Lindblad-Toh, K. (2013). The African coelacanth genome provides insights into tetrapod evolution. *Nature*, *496*(7445), 311–6. http://doi.org/10.1038/nature12027

Antonicka, H., Sasarman, F., Nishimura, T., Paupe, V., & Shoubridge, E. A. (2013). The Mitochondrial RNA-Binding Protein GRSF1 Localizes to RNA Granules and Is Required for Posttranscriptional Mitochondrial Gene Expression. *Cell Metabolism*, *17*(3), 386–398. http://doi.org/10.1016/j.cmet.2013.02.006

Antonicka, H., & Shoubridge, E. A. (2015). Mitochondrial RNA Granules Are Centers for Posttranscriptional RNA Processing and Ribosome Biogenesis. *Cell Reports*, *10*(6), 920–932. http://doi.org/10.1016/j.celrep.2015.01.030

Arora, A., Dutkiewicz, M., Scaria, V., Hariharan, M., Maiti, S., & Kurreck, J. (2008). Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif. *RNA (New York, N.Y.)*, *14*(7), 1290–1296. http://doi.org/10.1261/rna.1001708

Balagurumoorthy, P., & Brahmachari, S. K. (1994). Structure and stability of human telomeric sequence. *The Journal of Biological Chemistry*, *269*(34), 21858–21869.

Balasubramanian, S. (2014). Chemical biology on the genome. *Bioorganic and Medicinal Chemistry*, *22*(16), 4356–4370. http://doi.org/10.1016/j.bmc.2014.05.016

Barberan-Soler, S., & Zahler, A. M. (2008). Alternative splicing regulation during C. elegans development: Splicing factors as regulated targets. *PLoS Genetics*, *4*(2). http://doi.org/10.1371/journal.pgen.1000001

Barnett, J. A. (1998). A history of research on yeasts 1: Work by chemists and biologists 1789-1850. *Yeast*, *14*(16), 1439–1451. http://doi.org/10.1002/(SICI)1097-0061(199812)14:16<1439::AID-YEA339>3.0.CO;2-Z

Bates, P., Mergny, J.-L., & Yang, D. (2007). Quartets in G-major. The First International Meeting on Quadruplex DNA. *EMBO Reports*, *8*(11), 1003–1010. http://doi.org/10.1038/sj.embor.7401073

Beaudoin, J. D., Jodoin, R., & Perreault, J. P. (2014). New scoring system to identify RNA G-quadruplex folding. *Nucleic Acids Research*, *42*(2), 1209–1223. http://doi.org/10.1093/nar/gkt904

Beaudoin, J. D., & Perreault, J. P. (2010a). 5???-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Research*, *38*(20), 7022–7036. http://doi.org/10.1093/nar/gkq557

Beaudoin, J. D., & Perreault, J. P. (2010b). 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Research*, *38*(20), 7022–7036. http://doi.org/10.1093/nar/gkq557

Beckmann, B. M., Castello, A., & Medenbach, J. (2016). The expanding universe of ribonucleoproteins : of novel RNA-binding proteins and unconventional interactions. *Pflügers Archiv - European Journal of Physiology*. http://doi.org/10.1007/s00424-016-

1819-4

Bensaid, M., Melko, M., Bechara, E. G., Davidovic, L., Berretta, A., Catania, M. V., … Bardoni, B. (2009). FRAXE-associated mental retardation protein (FMR2) is an RNA-binding protein with high affinity for G-quartet RNA forming structure. *Nucleic Acids Research*, *37*(4), 1269–79. http://doi.org/10.1093/nar/gkn1058

Betts, M. J., & Russell, R. B. (2007). Amino-Acid Properties and Consequences of Substitutions. *Bioinformatics for Geneticists: A Bioinformatics Primer for the Analysis of Genetic Data: Second Edition*, *4*, 311–342. http://doi.org/10.1002/9780470059180.ch13

Biffi, G., Di Antonio, M., Tannahill, D., & Balasubramanian, S. (2014). Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nature Chemistry*, *6*(1), 75–80. http://doi.org/10.1038/nchem.1805

Biffi, G., Tannahill, D., McCafferty, J., & Balasubramanian, S. (2013). Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature Chemistry*, *5*(3), 182–6. http://doi.org/10.1038/nchem.1548

Bonnal, S., Schaeffer, C., Créancier, L., Clamens, S., Moine, H., Prats, A. C., & Vagner, S. (2003). A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. *Journal of Biological Chemistry*, *278*(41), 39330–39336. http://doi.org/10.1074/jbc.M305580200

Botstein, D., Chervitz, S. a., & Cherry, J. M. (1997). Yeast as a Model Organism. *Science (New York, N.Y.)*, *227*(5330), 1259–1260. http://doi.org/10.1126/science.277.5330.1259

Brázda, V., Hároníková, L., Liao, J., & Fojta, M. (2014). DNA and RNA Quadruplex-Binding Proteins. *International Journal of Molecular Sciences*, *15*(10), 17493–17517. http://doi.org/10.3390/ijms151017493

Brian O, Regan, M. G. (1991). © 19 9 1 Nature Publishing Group. *Letters To Nature*.

Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., & Neidle, S. (2006). Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Research*, *34*(19), 5402–5415. http://doi.org/10.1093/nar/gkl655

Burt, D. W. (2007). Emergence of the chicken as a model organism: implications for

agriculture and biology. *Poultry Science*, *86*(7), 1460–1471. http://doi.org/10.1093/ps/86.7.1460

Cayrel, A. (2011). Essential role for the interaction between hnRNP H / F and a G quadruplex in 3 9 -end processing and function during DNA damage, 220–225. http://doi.org/10.1101/gad.607011.interaction

Chen, C. D., Kobayashi, R., & Helfman, D. M. (1999). Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat ??-tropomyosin gene. *Genes and Development*, *13*(5), 593–606.

Cheong, C., & Moore, P. B. (1992). Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structures. *Biochemistry*, *31*(36), 8406–8414. http://doi.org/10.1021/bi00151a003

Christiansen, J., Kofod, M., & Nielsen, F. C. (1994). A guanosine quadruplex and two stable hairpins flank a major cleavage site in insulin-like growth factor II mRNA. *Nucleic Acids Research*, *22*(25), 5709–16. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=310137&tool=pmcentrez&rendertype=abstract\nhttp://www.ncbi.nlm.nih.gov/pubmed/7838726\nhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC310137

Cléry, A., & Allain, F. H.-T. (2013). From Structure To Function of Rna Binding Domains, (January).

Cléry, A., & Allain, H. F. (n.d.). RNA RECOGNITION MOTIFS (RRMs) A Simple Fold Binding a Large Panel of RNA Sequences and Structures.

Clingman, C. C., Deveau, L. M., Hay, S. a., Genga, R. M., Shandilya, S. M. D., Massi, F., & Ryder, S. P. (2014). Allosteric inhibition of a stem cell RNA-binding protein by an intermediary metabolite. *eLife*, *2014*, 1–26. http://doi.org/10.7554/eLife.02848

Colgan, D. F., & Manley, J. L. (1997). Mechanism and regulation of mRNA polyadenylation, (212), 2755–2766.

Consortium, T. C. S. and A. (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, *437*(7055), 69–87. http://doi.org/10.1038/nature04072

Corsi, A. K., Wightman, B., & Chalfie, M. (2015). A transparent window into biology: A primer on Caenorhabditis elegans. *Genetics*, *200*(2), 387–407.

http://doi.org/10.1534/genetics.115.176099

Crow, J. F. (1993). Felix Bernstein and the first human marker locus. *Genetics*, *133*(1), 4–7.

Daubner, G. M., Cléry, A., & Allain, F. H. T. (2013). RRM-RNA recognition: NMR or crystallography...and new findings. *Current Opinion in Structural Biology*, *23*(1), 100–108. http://doi.org/10.1016/j.sbi.2012.11.006

DeLano, W. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography*, *700*.

Deo, R. C., Bonanno, J. B., Sonenberg, N., & Burley, S. K. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell*, *98*(6), 835–845. http://doi.org/10.1016/S0092-8674(00)81517-2

Didiot, M. C., Tian, Z., Schaeffer, C., Subramanian, M., Mandel, J. L., & Moine, H. (2008). The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. *Nucleic Acids Research*, *36*(15), 4902–4912. http://doi.org/10.1093/nar/gkn472

Dominguez, C. (2006). NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: a novel mode of RNA recognition. *Nucleic Acids Research*, *34*(13), 3634–3645. http://doi.org/10.1093/nar/gkl488

Dominguez, C., Fisette, J.-F., Chabot, B., & Allain, F. H.-T. (2010). Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. *Nature Structural & Molecular Biology*, *17*(7), 853–61. http://doi.org/10.1038/nsmb.1814

Dooley, K., & Zon, L. I. (2000). Zebrafish: a model system for the study of human disease. *Current Opinion in Genetics & Development*, *10*(3), 252–256. http://doi.org/10.1016/S0959-437X(00)00074-5

Dorsey, M., Peterson, C., Bray, K., & Paquin, C. E. (1992). Spontaneous amplification of the ADH4 gene in Saccharomyces cerevisiae. *Genetics*, *132*(4), 943–950. http://doi.org/10.1371/journal.pone.0016015

Drummond, D. A., & Wilke, C. O. (2008). Mistranslation-Induced Protein Misfolding as a Dominant Constraint on Coding-Sequence Evolution. *Cell*, *134*(2), 341–352. http://doi.org/10.1016/j.cell.2008.05.042

Eddy, J., & Maizels, N. (2006). Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Research*, *34*(14), 3887–3896. http://doi.org/10.1093/nar/gkl529

Ferec, C., & Cutting, G. R. (2012). Assessing the disease-liability of mutations in CFTR. *Cold Spring Harbor Perspectives in Medicine*, *2*(12), 1–14. http://doi.org/10.1101/cshperspect.a009480

Fogel, B. L., & McNally, M. T. (2000). A cellular protein, hnRNP H, binds to the negative regulator of splicing element from Rous sarcoma virus. *Journal of Biological Chemistry*, *275*(41), 32371–32378. http://doi.org/10.1074/jbc.M005000200

Garneau, D., Revil, T., Fisette, J. F., & Chabot, B. (2005). Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. *Journal of Biological Chemistry*, *280*(24), 22641–22650. http://doi.org/10.1074/jbc.M501070200

Gellert, M., Lipsett, M. N., & Davies, D. R. (1962). Helix Formation By Guanylic Acid. *Proceedings of the National Academy of Sciences*, *48*(12), 2013–2018. http://doi.org/10.1073/pnas.48.12.2013

Gerstberger, S., Hafner, M., & Tuschl, T. (2014). A census of human RNA-binding proteins. *Nature Publishing Group*, *15*(12), 829–845. http://doi.org/10.1038/nrg3813

Glisovic, T., Bachorik, J. L., Yong, J., & Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. *FEBS Letters*, *582*(14), 1977–86. http://doi.org/10.1016/j.febslet.2008.03.004

Gomez, D., Gu??din, A., Mergny, J. L., Salles, B., Riou, J. F., Teulade-Fichou, M. P., & Calsou, P. (2010). A G-quadruplex structure within the 5???-UTR of TRF2 mRNA represses translation in human cells. *Nucleic Acids Research*, *38*(20), 7187–7198. http://doi.org/10.1093/nar/gkq563

Gomez, D., Lamarteleur, T., Lacroix, L., Mailliet, P., Mergny, J. L., & Riou, J. F. (2004). Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic Acids Research*, *32*(1), 371–379. http://doi.org/10.1093/nar/gkh181

Guo, J. U., & Bartel, D. P. (2016). RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science (New York, N.Y.)*, *353*(6306), aaf5371–aaf5371. http://doi.org/10.1126/science.aaf5371

Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., … Yokoyama, S. (1999). Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. *Nature*, *398*(6728), 579–585. http://doi.org/10.1038/19242

Hattori, H., Imai, H., Kirai, N., Furuhama, K., Sato, O., Konishi, K., & Nakagawa, Y. (2007). Identification of a responsible promoter region and a key transcription factor, CCAAT/enhancer-binding protein epsilon, for up-regulation of PHGPx in HL60 cells stimulated with TNF alpha. *Biochemical Journal*, *408*, 277–286. http://doi.org/Doi 10.1042/Bj20070245

Hedera, P., Blair, M. a, & Andermann, E. (2007). Familial mesial temporal lobe epilepsy maps, 1–5. http://doi.org/10.1212/01.wnl.0000261246.75977.89

Hung, C. M., Garcia-Haro, L., Sparks, C. A., & Guertin, D. A. (2012). mTOR-dependent cell survival mechanisms. *Cold Spring Harbor Perspectives in Biology*, *4*(12), 1–17. http://doi.org/10.1101/cshperspect.a008771

Huppert, J. L., & Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. *Nucleic Acids Research*, *33*(9), 2908–2916. http://doi.org/10.1093/nar/gki609

Jablonski, J. a, & Caputi, M. (2009). Role of cellular RNA processing factors in human immunodeficiency virus type 1 mRNA metabolism, replication, and infectivity. *Journal of Virology*, *83*(2), 981–992. http://doi.org/10.1128/JVI.01801-08

Jacob, F., & Monod, J. (1960). Genetic Regulatory Mechanisms in the Synthesis. http://doi.org/10.1016/S0022-2836(61)80072-7

Jacquenet, S., Méreau, A., Bilodeau, P. S., Damier, L., Stoltzfus, C. M., & Branlant, C. (2001). A Second Exon Splicing Silencer within Human Immunodeficiency Virus Type 1 tat Exon 2 Represses Splicing of Tat mRNA and Binds Protein hnRNP H. *Journal of Biological Chemistry*, *276*(44), 40464–40475. http://doi.org/10.1074/jbc.M104070200

Jansen, R., Niessing, D., Baumann, S., & Feldbru, M. (2014). mRNA transport meets membrane traffic, *30*(9). http://doi.org/10.1016/j.tig.2014.07.002

Jennings, B. H. (2011). Drosophila-a versatile model in biology & medicine. *Materials Today*, *14*(5), 190–195. http://doi.org/10.1016/S1369-7021(11)70113-4

Ji, X., Sun, H., Zhou, H., Xiang, J., Tang, Y., & Zhao, C. (2011). Research progress of RNA quadruplex. *Nucleic Acid Therapeutics*, *21*(3), 185–200.

http://doi.org/10.1089/oli.2010.0272

Jodoin, R., Bauer, L., Garant, J.-M., Mahdi Laaref, A., Phaneuf, F., & Perreault, J.-P. (2014). The folding of 5'-UTR human G-quadruplexes possessing a long central loop. *RNA (New York, N.Y.)*, *20*, 1129–41. http://doi.org/10.1261/rna.044578.114

Johansson, C., Finger, L. D., Trantirek, L., Mueller, T. D., Kim, S., Laird-Offringa, I. A., & Feigon, J. (2004). Solution structure of the complex formed by the two N-terminal RNA-binding domains of nucleolin and a pre-rRNA target. *Journal of Molecular Biology*, *337*(4), 799–816. http://doi.org/10.1016/j.jmb.2004.01.056

Jourdain, A. A., Koppen, M., Wydro, M., Rodley, C. D., Lightowlers, R. N., Chrzanowska-Lightowlers, Z. M., & Martinou, J.-C. (2013). GRSF1 Regulates RNA Processing in Mitochondrial RNA Granules. *Cell Metabolism*, *17*(3), 399–410. http://doi.org/10.1016/j.cmet.2013.02.005

Kash, J. C., Cunningham, D. M., Smit, M. W., Park, Y., Fritz, D., Wilusz, J., … Katze, M. G. (2002). Selective Translation of Eukaryotic mRNAs : Functional Molecular Analysis of GRSF-1 , a Positive Regulator of Influenza Virus Protein Synthesis Selective Translation of Eukaryotic mRNAs : Functional Molecular Analysis of GRSF-1 , a Positive Regulator of I. *Journal of Virology*, *76*(20), 10417–10426. http://doi.org/10.1128/JVI.76.20.10417

Khoo, B., Roca, X., Chew, S. L., & Krainer, A. R. (2007). Antisense oligonucleotide-induced alternative splicing of the APOB mRNA generates a novel isoform of APOB. *BMC Molecular Biology*, *8*, 3. http://doi.org/10.1186/1471-2199-8-3

Kikin, O., D'Antonio, L., & Bagga, P. S. (2006). QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Research*, *34*(Web Server), W676–W682. http://doi.org/10.1093/nar/gkl253

Kim, J., Cheong, C., & Moore, P. B. (1991). Tetramerization of an RNA oligonucleotide containing a GGGG sequence. *Nature*, *351*(6324), 331–332. http://doi.org/10.1038/351331a0

König, S. L. B., Evans, A. C., & Huppert, J. L. (2010). Seven essential questions on G-quadruplexes. *Biomolecular Concepts*, *1*(2), 197–213. http://doi.org/10.1515/bmc.2010.011

Koornneef, M., & Meinke, D. (2010). The development of Arabidopsis as a model plant. *Plant Journal*, *61*(6), 909–921. http://doi.org/10.1111/j.1365-313X.2009.04086.x

Kumar, P., Satish, A., Patel, P., & Panchaldr, H. (2016). Genomics Data Mutation-based structural modi fi cation and dynamics study of amyloid beta peptide (1 – 42): An in - silico-based analysis to cognize the mechanism of aggregation, *7*, 189–194. http://doi.org/10.1016/j.gdata.2016.01.003

Kumari, S., Bugaut, A., Huppert, J. L., & Balasubramanian, S. (2007). An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nature Chemical Biology*, *3*(4), 218–221. http://doi.org/10.1038/nchembio864

Lane, A. N., Chaires, J. B., Gray, R. D., & Trent, J. O. (2008). Stability and kinetics of G-quadruplex structures. *Nucleic Acids Research*, *36*(17), 5482–5515. http://doi.org/10.1093/nar/gkn517

Maizels, N., & Gray, L. T. (2013). The G4 Genome. *PLoS Genetics*, *9*(4). http://doi.org/10.1371/journal.pgen.1003468

Malgowska, M., Czajczynska, K., Gudanis, D., Tworak, A., & Gdaniec, Z. (2016). Overview of the RNA G-quadruplex structures, *63*(4), 609–621. http://doi.org/10.18388/abp.2016_1335

Maris, C., Dominguez, C., & Allain, F. H. T. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS Journal*, *272*(9), 2118–2131. http://doi.org/10.1111/j.1742-4658.2005.04653.x

Marmorstein, R., & Trievel, R. C. (2009). Biochimica et Biophysica Acta Histone modifying enzymes : Structures , mechanisms , and speci fi cities. *BBA - Gene Regulatory Mechanisms*, *1789*(1), 58–68. http://doi.org/10.1016/j.bbagrm.2008.07.009

Menendez, C., Frees, S., & Bagga, P. S. (2012). QGRS-H Predictor: A web server for predicting homologous quadruplex forming G-rich sequence motifs in nucleotide sequences. *Nucleic Acids Research*, *40*(W1), 96–103. http://doi.org/10.1093/nar/gks422

Millevoi, S., Moine, H., & Vagner, S. (2012). G-quadruplexes in RNA biology. *Wiley Interdisciplinary Reviews: RNA*, *3*(4), 495–507. http://doi.org/10.1002/wrna.1113

Morris, M. J., Negishi, Y., Pazsint, C., Schonhoft, J. D., & Basu, S. (2010). An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES. *Journal of the American Chemical Society*, *132*(50), 17831–17839. http://doi.org/10.1021/ja106287x

Mukundan, V. T., & Phan, A. T. (2013). Bulges in G-quadruplexes: Broadening the definition of G-quadruplex-forming sequences. *Journal of the American Chemical Society*, *135*(13), 5017–5028. http://doi.org/10.1021/ja310251r

Myrick, L. K., Nakamoto-Kinoshita, M., Lindor, N. M., Kirmani, S., Cheng, X., & Warren, S. T. (2014). Fragile X syndrome due to a missense mutation. *European Journal of Human Genetics : EJHG*, *22*(10), 1185–9. http://doi.org/10.1038/ejhg.2013.311

Nelson, W. J., & Nusse, R. (2004). Convergence of Wnt, beta-catenin, and cadherin pathways. *Science (New York, N.Y.)*, *303*(5663), 1483–7. http://doi.org/10.1126/science.1094291

Nieradka, A., Ufer, C., Thiadens, K., Grech, G., Horos, R., van Coevorden-Hameete, M., … von Lindern, M. (2014a). Grsf1-Induced Translation of the SNARE Protein Use1 Is Required for Expansion of the Erythroid Compartment. *PLoS ONE*, *9*(9), e104631. http://doi.org/10.1371/journal.pone.0104631

Nieradka, A., Ufer, C., Thiadens, K., Grech, G., Horos, R., van Coevorden-Hameete, M., … von Lindern, M. (2014b). Grsf1-induced translation of the SNARE protein Use1 is required for expansion of the erythroid compartment. *PloS One*, *9*(9), e104631. http://doi.org/10.1371/journal.pone.0104631

Nishikura, K. (2006). Editor meets silencer: RNA editing and RNA interference. *Nature Reviews Molecular Cell Biology*, *7*(12), 919–931. http://doi.org/10.1038/nrm2061.Editor

Park, Y. W., Wilusz, J., & Katze, M. G. (1999). Regulation of eukaryotic protein synthesis: selective influenza viral mRNA translation is mediated by the cellular RNA-binding protein GRSF-1. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(12), 6694–9. http://doi.org/10.1073/pnas.96.12.6694

Paziewska, A., Wyrwicz, L. S., Bujnicki, J. M., Bomsztyk, K., & Ostrowski, J. (2004). Cooperative binding of the hnRNP K three KH domains to mRNA targets. *FEBS Letters*, *577*(1-2), 134–140. http://doi.org/10.1016/j.febslet.2004.08.086

Phan, A. T., Kuryavyi, V., & Patel, D. J. (2006). DNA architecture: from G to Z. *Current Opinion in Structural Biology*, *16*(3), 288–298. http://doi.org/10.1016/j.sbi.2006.05.011

Proudfoot, N. J., Furger, A., Dye, M. J., & William, S. (2002). Integrating mRNA Processing with Transcription, *108*, 501–512.

Qian, Z., & Wilusz, J. (1994). GRSF-1: a poly(A)+ mRNA binding protein which interacts with a conserved G-rich element. *Nucleic Acids Research*, *22*(12), 2334–2343.

Rhodes, D., & Lipps, H. J. (2015). G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Research*, *43*(18), 8627–37. http://doi.org/10.1093/nar/gkv862

Richmond, T. J., & Davey, C. A. (2003). The structure of DNA in the nucleosome core, 145–150.

Rinn, J. L., Ule, J., Attar, N., Borsenberger, V., Crowe, M., Lehbauer, J., … Wessels, L. (2014). 'Oming in on RNA–protein interactions. *Genome Biology*, *15*(1), 401. http://doi.org/10.1186/gb4158

Rodriguez, R., Müller, S., Yeoman, J. A., Trentesaux, C., Riou, J. F., & Balasubramanian, S. (2008). A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *Journal of the American Chemical Society*, *130*(47), 15758–15759. http://doi.org/10.1021/ja805615w

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., … Altshuler, D. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, *409*(6822), 928–933. http://doi.org/10.1038/35057149

Samatanga, B., Dominguez, C., Jelesarov, I., & Allain, F. H.-T. (2013). The high kinetic stability of a G-quadruplex limits hnRNP F qRRM3 binding to G-tract RNA. *Nucleic Acids Research*, *41*(4), 2505–2516. http://doi.org/10.1093/nar/gks1289

Schaub, M. C., Lopez, S. R., & Caputi, M. (2007). Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes. *Journal of Biological Chemistry*, *282*(18), 13617–13626. http://doi.org/10.1074/jbc.M700774200

Schlaich, N. (2010). Regulation of Gene Expression Regulation of Gene Expression, *136*(3), 2013–2015. http://doi.org/10.1016/B978-012095440-7/50028-7

Sen, D., & Gilbert, W. (1988). Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, *334*(6180), 364–6. http://doi.org/10.1038/334364a0

Smith, K. (2002). Genetic Polymorphism and SNPs, 1–13.

Sneddon, A. A., Wu, H. C., Farquharson, A., Grant, I., Arthur, J. R., Rotondo, D., …

Wahle, K. W. J. (2003). Regulation of selenoprotein GPx4 expression and activity in human endothelial cells by fatty acids, cytokines and antioxidants. *Atherosclerosis*, *171*(1), 57–65. http://doi.org/10.1016/j.atherosclerosis.2003.08.008

Sokol, S. Y. (1999). Wnt signaling axis specification in vertebrates. *Genetics & Development*, (Sokol 2011), 405–410. http://doi.org/10.1016/S0959-437X(99)80061-6

Striegel, A. M. (2016). Viscometric Detection in Size-Exclusion Chromatography: Principles and Select Applications. *Chromatographia*, *79*(15-16), 945–960. http://doi.org/10.1007/s10337-016-3078-0

Sun, Y.-B., Xiong, Z.-J., Xiang, X.-Y., Liu, S.-P., Zhou, W.-W., Tu, X.-L., … Zhang, Y.-P. (2015). Whole-genome sequence of the Tibetan frog Nanorana parkeri and the comparative evolution of tetrapod genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(11), E1257–62. http://doi.org/10.1073/pnas.1501764112

Sunyaev, S., Ramensky, V., & Bork, P. (2000). Towards a structural basis of human non-synonymous single nucleotide polymorphisms. *Trends in Genetics*, *16*(5), 15–17. http://doi.org/10.1016/S0168-9525(00)01988-0

Syvänen, a C. (2001). Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Reviews. Genetics*, *2*(12), 930–42. http://doi.org/10.1038/35103535

T. Ferreira, W. R. (2012). ImageJ User Guide IJ 1.46r. *IJ 1.46r*, 185. http://doi.org/10.1038/nmeth.2019

Taj, M. K., Samreen, Z., Ling, J. X., Taj, I., & Yunlin, W. (2014). Escherichia coli as a Model Organism. *International Journal of Engineering Research and Science & Technology*, *3*(April), 1–10.

The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, *408*(6814), 796–815. http://doi.org/10.1038/35048692

Ufer, C. (2012). The biology of the RNA binding protein guanine-rich sequence binding factor 1. *Current Protein & Peptide Science*, *13*(4), 347–57. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22708492

Ufer, C., Wang, C. C., Borchert, A., Heydeck, D., Editors, R., Coffman, J. A., … Sato, H.

(2010). Redox Control in Mammalian Embryo Development, *13*(6).

Ufer, C., Wang, C. C., Fähling, M., Schiebel, H., Thiele, B. J., Billett, E. E., … Borchert, A. (2008). Translational regulation of glutathione peroxidase 4 expression through guanine-rich sequence-binding factor 1 is essential for embryonic brain development. *Genes & Development*, *22*(13), 1838–50. http://doi.org/10.1101/gad.466308

Van Dusen, C. M., Yee, L., McNally, L. M., & McNally, M. T. (2010). A glycine-rich domain of hnRNP H/F promotes nucleocytoplasmic shuttling and nuclear import through an interaction with transportin 1. *Molecular and Cellular Biology*, *30*(10), 2552–2562. http://doi.org/10.1128/MCB.00230-09

Von Hacht, A., Seifert, O., Menger, M., Schütze, T., Arora, A., Konthur, Z., … Kurreck, J. (2014). Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic Acids Research*, *42*(10), 6630–6644. http://doi.org/10.1093/nar/gku290

Wakabayashi-Ito, N., Belvin, M. P., Bluestein, D. a, & Anderson, K. V. (2001). fusilli, an essential gene with a maternal role in Drosophila embryonic dorsal-ventral patterning. *Developmental Biology*, *229*(1), 44–54. http://doi.org/10.1006/dbio.2000.9954

Wang, X., Tanaka Hall, T. M., & Hall, T. M. T. (2001). letters Structural basis for recognition of AU-rich element RNA by the HuD. *Nature Structural Biology*, *8*(2), 141–145. http://doi.org/10.1038/84131

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., … Mouse Genome Sequencing, C. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, *420*(6915), 520–562. http://doi.org/10.1038/nature01262

Wullschleger, S., Loewith, R., & Hall, M. N. (2006). TOR signaling in growth and metabolism. *Cell*, *124*(3), 471–484. http://doi.org/10.1016/j.cell.2006.01.016

Yen, P. H. (2004). Putative biological functions of the DAZ family. *International Journal of Andrology*, *27*(3), 125–129. http://doi.org/10.1111/j.1365-2605.2004.00469.x

Zhang, J., Lieu, Y. K., Ali, A. M., Penson, A., Reggio, K. S., Rabadan, R., … Manley, J. L. (2015). Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(34), E4726–34. http://doi.org/10.1073/pnas.1514105112

Zheng, S., Kim, H., & Verhaak, R. G. W. (2014). Silent mutations make some noise. *Cell*, *156*(6), 1129–1131. http://doi.org/10.1016/j.cell.2014.02.037

7. PUBLICATIONS & SCIENTIFIC CONTRIBUTIONS

Publications

Nieradka,A., Ufer,C., Thiadens,K., Grech,G., Horos,R., van Coevorden-Hameete,M., van den Akker,E., **Sofi,S**., Kuhn,H. and von Lindern,M. (2014) Grsf1-Induced Translation of the SNARE Protein Use1 Is Required for Expansion of the Erythroid Compartment. *PLoS One*, **9**, e104631.

Sofi, S., *et al*. (2017).''Functional characterization of naturally occurring genetic variations of the human Guanine-rich RNA sequence binding factor 1 (GRSF1)''. (Manuscript submitted).

Sofi, S., *et al*. (2017).''Recombinant expression, purification and functional characterization of the full-length RNA-binding protein GRSF1 and several truncation mutants including its RNA-binding domains ''. (Manuscript under preparation).

Sofi, S., *et al*. (2017).''Characterization of the RNA substrates of RNA-binding protein GRSF1 ''. (Manuscript under preparation).

Posters

Sofi, **S**., Kühn, H., Ufer, C., “Guanine-rich sequence binding factor 1 binds to G-quadruplex structures in RNA’’. *Functional Molecular Infection Epidemiology, 6. April 2016, RKI, Berlin, Germany (Oral Presentation).*

Sofi, **S**., Kühn, H., Ufer, C., “Guanine-rich sequence binding factor 1 binds to G-quadruplex structures in RNA’’. *The Non-Coding Genome Conference, 18-21 October 2015, EMBL, Heidelberg, Germany.*

Sofi, **S**., Kühn, H., Ufer, C., “Guanine-rich sequence binding factor 1 binds to G-quadruplex structures in RNA’’. *40th FEBS Congress, 4-9 July 2015, Berlin, Germany.*

Sofi, **S**., Kühn, H., Ufer, C., “Guanine-rich sequence binding factor 1 binds to G-quadruplex structures in RNA’’. *Pathogen Biology and Genomics, IRTG-GRK 1673 workshop, 23-27 February 2015, University of Hyderabad, India.*

Sofi, **S**., Kühn, H., Ufer, C., “Guanine-rich sequence binding factor 1 binds to G-quadruplex structures in RNA’’. *Conference on Computational RNA Biology, 11-13 November 2014, Cambridge, UK.*

Sofi, **S**., Kühn, H., Ufer, C., “Guanine-rich sequence binding factor 1 binds to G-quadruplex structures in RNA’’. *Frontiers of Parasitology, ZIBI summer symposium, 29 June-01 July 2014, Berlin, Germany.*

8. SELBSTSTÄNDIGKEITSERKLÄRUNG

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbständig angefertigt habe. Ich versichere, dass ich ausschließlich die angegebenen Quellen und Hilfen in Anspruch genommen habe.

Sajad Ahmad Sofi
Berlin, den 24.04.17